

UNIVERSIDAD DE GRANADA

FACULTAD DE CIENCIAS



DEPARTAMENTO DE QUÍMICA FÍSICA

**PROTEÍNAS ANCESTRALES:
RECONSTRUCCIÓN Y CARACTERIZACIÓN
ENERGÉTICA, FUNCIONAL Y
ESTRUCTURAL**

Álvaro Inglés Prieto

TESIS DOCTORAL

PROGRAMA DE DOCTORADO EN QUÍMICA

GRANADA

2012

Editor: Editorial de la Universidad de Granada
Autor: Álvaro Inglés Prieto
D.L.: GR 1065-2013
ISBN: 978-84-9028-485-8

UNIVERSIDAD DE GRANADA
FACULTAD DE CIENCIAS
DEPARTAMENTO DE QUÍMICA FÍSICA

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

Fdo: Álvaro Inglés Prieto
Licenciado en Bioquímica por la Universidad de Salamanca

Granada, a 5 de Noviembre de 2012

Vº Bº
Directores de la tesis:

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

Fdo.: Beatriz Ibarra Molero
Profesora titular
Departamento de Química
Física
Universidad de Granada

Fdo.: José Antonio Gavira Gallardo
Científico titular CSIC
Laboratorio de estudios
cristalográficos
Instituto andaluz de ciencias de la
tierra (CSIC)

Los directores de la tesis, José Manuel Sánchez Ruiz, Beatriz Ibarra Molero, José Antonio Gavira Gallardo y el doctorando, Álvaro Inglés Prieto, garantizamos, al firmar esta tesis doctoral, que el trabajo ha sido realizado por el doctorando bajo la dirección de los directores de la tesis y hasta donde nuestro conocimiento alcanza, en la realización del trabajo, se han respetado los derechos de otros autores a ser citados, cuando se han utilizado sus resultados o publicaciones.

Granada, a 5 de Noviembre de 2012

Directores de la Tesis:

Fdo.: José Manuel Sánchez Ruiz

Fdo.: Beatriz Ibarra Molero

Fdo.: José Antonio Gavira Gallardo

Doctorando:

Fdo.: Álvaro Inglés Prieto

A mi madre y a mi padre

AGRADECIMIENTOS

Por suerte, tengo muchas personas a las que agradecer el apoyo que me han dado a lo largo de esta tesis.

Quiero agradecer, en primer lugar, a mis directores de tesis: José Manuel Sánchez Ruiz por haberme dado la oportunidad de trabajar en este grupo y en este proyecto, además de por todo lo que he aprendido de él. A Beatriz Ibarra Molero, quiero agradecerle su ayuda y sus sugerencias, ofreciéndome su apoyo en todo momento. A José Antonio Gavira Gallardo (Gavi) quiero agradecerle el haber conseguido siempre motivarme y hacer que tratara de superarme cada día.

Gracias a mis compañeros del laboratorio por su ayuda y sus consejos: Valeria, Diego, Sisi, Héctor, Asun y Mari Carmen. A los compañeros del otro grupo: Sara, Ana Mari, Javi Murciano, Salva, Adela, David, Crunchi, Fadia y Fátima. De manera especial quiero agradecer a José Luis, Lorena, Roci, Fran, Carles, Noel, Ángel, Bertri y David porque además de ser mis compañeros, he tenido la suerte de que sean mis amigos. No tengo palabras para decíros lo mucho que agradezco las tardes (y noches) de cañas y tapeo o a las subidas a la sierra.

Quiero agradecer también a George Makhatadze (Rensselaer Polytechnic Institute, Troy, USA), a Ana Fernández Sesma (Mount Sinai School of Medicine, NY, USA) y a Julio Fernández (Columbia University, NY, USA) por hacerme dado la oportunidad de realizar las estancias breves en sus laboratorios.

Agradezco también el apoyo que siempre he recibido de mi familia, pues, a pesar de la distancia, nunca han dejado de animarme y siempre han tenido palabras de consuelo en los momentos más difíciles.

Quisiera agradecer a mis amigos de Graná: Nata, Carlos, Luis y Alba; gracias por hacer que me sienta como en casa. A mis amigos de Cáceres, Salamanca y Madrid, especialmente a Carla, Josete, Mjo, Genis, Jaime, Javi, Fran y Alex; quiero agradecerles su constancia y su amistad a lo largo de todos estos años.

A Inma, gracias por todos los momentos que hemos compartido en estos años y los que nos quedan por compartir. Gracias por todo tu apoyo y comprensión, por aceptarme como soy y por completarme. Gracias por estar ahí.

"Dícebamus hesterna día..."

(Fray Luís de León)

ÍNDICE

1. INTRODUCCIÓN Y OBJETIVOS.....	17
1.1. Introducción general a la resurrección de proteínas ancestrales.....	19
1.2. Implicaciones evolutivas de la resurrección de proteínas ancestrales.....	20
1.3. Implicaciones en ingeniería de proteínas de la resurrección de proteínas ancestrales.....	21
1.4. Planteamiento general de esta tesis doctoral: resurrección y caracterización biofísica, estructural y energética de tiorredoxinas resucitadas.....	23
1.5. Objetivos concretos.....	25
1.6. Bibliografía.....	26
2. MATERIALES Y MÉTODOS.....	29
2.1. Materiales.....	31
2.1.1. Productos.....	31
2.1.2. Instrumentación.....	32
2.2. Métodos.....	35
2.2.1. Mutagénesis dirigida.....	35
2.2.2. Transformación de bacterias y expresión de tiorredoxinas.....	37
2.2.2.1. Transformación en bacterias <i>XL1-BLUE</i>	39
2.2.2.2. Aislamiento de plásmidos.....	39
2.2.2.3. Transformación en bacterias <i>BL21(DE3)</i>	40
2.2.2.4. Pruebas de sobreexpresión de tiorredoxinas.....	41
2.2.3. Purificación de tiorredoxinas.....	42
2.2.3.1. Purificación de tiorredoxinas <i>His-tagged</i>	43
2.2.3.2. Purificación de tiorredoxinas sin <i>His-tag</i>	45
• Cromatografía de intercambio iónico.....	45
• Cromatografía de filtración en gel.....	47
2.2.3.3. Preparación de la muestra de tiorredoxina.....	49
• Cambio de la solución tamponante.....	49
• Determinación de la concentración de tiorredoxina.....	51
• Procedimiento de concentración de tiorredoxina.....	52
2.2.4. Construcción y caracterización de bibliotecas combinatoriales de variantes de tiorredoxinas.....	52
2.2.5. Ensayo de actividad oxido-reductasa de tiorredoxinas por precipitación de insulina.....	54
2.2.6. Calorimetría diferencial de barrido.....	57
2.2.6.1. Descripción del microcalorímetro.....	57
2.2.6.2. Descripción del experimento de DSC.....	59
2.2.6.3. Análisis basado en la termodinámica del equilibrio: modelo de dos estados.....	61
2.2.7. Determinación estructural de tiorredoxinas por cristalografía de rayos-X.....	67
2.2.7.1. Cristalización de tiorredoxinas ancestrales.....	68
• Métodos de cristalización de proteínas.....	69
• Búsqueda de condiciones de cristalización.....	76
2.2.7.2. Difracción de rayos-X.....	79

2.2.7.3. Métodos de resolución estructural.....	82
• Método de reemplazamiento molecular.....	83
2.2.7.4. Construcción y refinamiento del modelo.....	84
2.2.7.5. Validación del modelo estructural.....	87
2.2.8. Análisis bioinformático de secuencias y estructuras de tiorredoxinas.....	87
2.2.8.1. Alineamiento de secuencias de tiorredoxinas.....	87
2.2.8.2. Alineamiento estructural de tiorredoxinas.....	88
2.2.8.3. Asignación de elementos de estructura secundaria.....	88
2.2.9. Análisis energético-estructural de tiorredoxinas.....	89
2.2.9.1. Cálculos de área superficial accesible al disolvente.....	89
2.2.9.2. Estimación de cavidades internas.....	90
2.2.9.3. Estimación de los enlaces de hidrógeno.....	91
2.2.9.4. Cálculos Electrostáticos.....	92
2.2.9.5. Estimación de los puentes salinos.....	93
2.2.9.6. Cálculos de propensión de elementos de estructura secundaria.....	94
2.2.9.7. Determinación de los contactos entre residuos.....	95
2.2.10. Análisis por regresión de mínimos cuadrados parciales de las bibliotecas combinatorias de variantes de tiorredoxinas.....	95
2.2.10.1. Introducción al análisis multivariante.....	96
2.2.10.2. Procedimiento de regresión por mínimos cuadrados parciales.....	98
2.2.11. Bibliografía.....	102
3. RESURRECCIÓN DE PROTEÍNAS ANCESTRALES, CARACTERIZACIÓN BIOFÍSICA Y ESTUDIOS DE CATÁLISIS ENZIMÁTICA MEDIANTE TÉCNICAS DE MOLÉCULA INDIVIDUAL.....	109
3.1. Introducción.....	111
3.2. Procedimiento Experimental.....	113
3.2.1. Análisis filogenético y reconstrucción de secuencias de tiorredoxinas ancestrales.....	113
3.2.2. Expresión y purificación de tiorredoxinas.....	113
3.2.3. Medidas de la actividad enzimática de las tiorredoxinas ancestrales con ensayos bioquímicos convencionales.....	113
3.2.4. Experimentos de calorimetría diferencial de barrido.....	114
3.2.5. Experimentos de microscopía de fuerza atómica.....	115
3.2.6. Análisis de los datos de AFM.....	116
3.3. Resultados.....	116
3.3.1. Reconstrucción de tiorredoxinas ancestrales.....	116
3.3.2. Estabilidad térmica de las tiorredoxinas ancestrales.....	120
3.3.3. Cinética química de reducción de puentes disulfuro dependiente de la fuerza.....	123
3.3.4. Actividad de las tiorredoxinas ancestrales en condiciones ácidas.....	132
3.4. Discusión.....	135
3.5. Bibliografía.....	136
ARTÍCULO: <i>Single-molecule paleoenzymology probes the chemistry of resurrected enzymes</i>	141
4. OBTENCIÓN DE LAS ESTRUCTURAS DE TIORREDOXINAS PRECÁMBRICAS POR CRISTALOGRAFÍA DE RAYOS-X Y ANÁLISIS EVOLUTIVO A LO LARGO DE 4 MIL MILLONES DE AÑOS	157
4.1. Introducción	159
4.2. Determinación de las estructuras de tiorredoxinas precámbricas por cristalografía	

de rayos-X.....	162
4.2.1. Cristalización de tiorredoxinas precámbricas.....	163
4.2.2. Optimización de condiciones de cristalización.....	169
4.2.3. Recogida de datos de difracción.....	171
4.2.4. Resolución del modelo estructural.....	175
4.3. Análisis de la evolución de la estructura de tiorredoxinas a lo largo de 4 mil millones de años.....	180
4.4. Bibliografía.....	188
5. ANÁLISIS ENERGÉTICO-ESTRUCTURAL DE TIORREDOXINAS PRECÁMBRICAS.....	191
5.1. Introducción.....	193
5.2. Resultados y Discusión.....	194
5.2.1. El efecto hidrofóbico.....	194
5.2.2. El número de enlaces de hidrógeno.....	198
5.2.3. La distribución de carga superficial.....	201
5.2.4. El número de puentes salinos.....	204
5.2.5. Propensión de aminoácidos en elementos de estructura secundaria.....	206
5.2.6. El número de prolina y glicinas.....	209
5.3. Bibliografía.....	213
6. ESTUDIO EXPERIMENTAL DE LA ESTABILIDAD TÉRMICA DE LA TIORREDOXINA ANCESTRAL LPBCA.....	217
6.1. Introducción.....	219
6.2. Resultados y Discusión.....	220
6.2.1. Evaluación del papel de las prolina y glicinas en la estabilidad de tiorredoxina de <i>E. coli</i> y LPBCA.....	220
6.2.2. Evaluación del papel de la mutación no conservativa K90L en la estabilidad de tiorredoxina de <i>E. coli</i> y LPBCA.....	228
6.3. Bibliografía.....	240
7. CONCLUSIONES.....	241
8. RESUMEN.....	245
9. ABREVIATURAS Y SÍMBOLOS.....	251
10. APÉNDICE: otros artículos publicados durante esta tesis doctoral.....	253
11. CURRICULUM VITAE.....	287

1. INTRODUCCIÓN Y OBJETIVOS

1.1. Introducción general a la resurrección de proteínas ancestrales

Hace unos cincuenta años, Linus Pauling y Emile Zuckerkandl sugirieron la posibilidad de reconstruir secuencias de proteínas ancestrales a partir del análisis comparativo de secuencias de esas proteínas en organismos actuales (Pauling y Zuckerkandl 1963). El campo de la Paleobioquímica que ellos vaticinaron cobró vida unos 25 años después, gracias al aumento del número de secuencias conocidas, al desarrollo de procedimientos estadísticos fiables para la reconstrucción de secuencias y a los avances en ingeniería genética. Todos estos avances permitieron “resucitar” y estudiar *in vitro* proteínas pertenecientes a organismos extintos (Malcolm, Wilson et al. 1990). Desde entonces, se han publicado varios estudios de caracterizaciones experimentales de proteínas resucitadas. En un artículo de revisión de 2007, Benner y colaboradores recopilan diecinueve ejemplos de proteínas resucitadas hasta ese año (Benner, Sassi et al. 2007), aunque varios estudios adicionales han sido publicados después de dicha revisión: (Ortlund, Bridgham et al. 2007; Gaucher, Govindarajan et al. 2008; Bridgham, Ortlund et al. 2009; Bridgham, Eick et al. 2010; Carroll, Ortlund et al. 2011; Eick y Thornton 2011; Finnigan, Hanson-Smith et al. 2011; Perez-Jimenez, Ingles-Prieto et al. 2011; Finnigan, Hanson-Smith et al. 2012).

Los resultados de estos trabajos sugieren que la resurrección ancestral de proteínas tiene dos importantes implicaciones. Por un lado tiene implicaciones evolutivas, dado que estos estudios han contribuido a ampliar la información sobre el entorno físico que rodeaba a los organismos extintos, y sobre cómo sus proteínas (y los propios organismos) se adaptaron a los cambios en dicho entorno (Benner, Sassi et al. 2007). Por otro lado, tiene implicaciones en ingeniería de proteínas, ya que las proteínas ancestrales estudiadas a menudo presentan unas propiedades alteradas en términos de estabilidad o actividad catalítica (Gaucher, Govindarajan et al. 2008; Perez-Jimenez, Ingles-Prieto et al. 2011). Estas dos implicaciones se discuten brevemente en las secciones siguientes.

1.2. Implicaciones evolutivas de la resurrección de proteínas ancestrales

Podemos describir la historia de la vida en la Tierra como 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 los organismos interactúan con su entorno y se adaptan a los cambios en el mismo. En este contexto, la paleobioquímica experimental nos brinda la oportunidad de investigar en el laboratorio la historia molecular de los organismos modernos (Benner, Sassi et al. 2007), permitiendo una correlación entre los registros paleontológico, geológico y molecular de la historia de la vida en la Tierra. Esto implica que el estudio de proteínas resucitadas puede corroborar hipótesis que no podrían ser comprobadas solamente con el estudio de restos fósiles.

A pesar de los numerosos ejemplos de resurrección de proteínas ancestrales, la mayoría de las proteínas resucitadas existieron hace unos pocos millones de años. Sin embargo, se han publicado recientemente varios ejemplos de reconstrucción de proteínas pertenecientes a organismos del periodo Precámbrico (hace 4.5-0.5 mil millones de años antes del presente) (Gaucher, Govindarajan et al. 2008; Perez-Jimenez, Ingles-Prieto et al. 2011). Estos trabajos han permitido hacer estimaciones muy atrás en el tiempo sobre el entorno de la Tierra primitiva. Por ejemplo, en un estudio publicado en *Nature* (Gaucher, Govindarajan et al. 2008), Eric Gaucher y colaboradores determinaron la termoestabilidad de 25 factores de elongación resucitados que cubrían una escala temporal de unos 3.5 a 0.5 miles de millones de años antes del presente. Los resultados de este estudio apoyaron la idea de que los océanos primitivos eran cálidos y que se han ido enfriando a lo largo de la evolución, lo cual es consistente con las estimaciones de la temperatura de los océanos primitivos, derivadas de los estudios de deposición de isótopos de oxígeno en rocas. Este resultado ha sido corroborado por el trabajo presentado en esta tesis doctoral (Perez-Jimenez, Ingles-Prieto et al. 2011).

Desde el punto de vista molecular, el estudio de proteínas ancestrales resucitadas puede esclarecer cómo han cambiado las propiedades de las proteínas tales como la función o la estabilidad, a lo largo de la evolución. Por ejemplo, se sabe poco acerca de cómo se ha originado la función de las proteínas y cómo han afectado las condiciones ambientales a su evolución química (Zalatan y Herschlag 2009). Existen varios ejemplos de enzimas actuales que catalizan específicamente una

reacción y que están relacionadas evolutivamente con enzimas promiscuas que catalizan dicha reacción. Esta observación plantea una cuestión obvia: ¿el antepasado común de estas enzimas presentaba promiscuidad?. Gracias a la resurrección de proteínas ancestrales estamos más cerca de conocer la respuesta a esta cuestión. Por ejemplo, al estudiar la actividad de unión a ligandos en una familia de receptores de esteroides, se observó que el receptor ancestral resucitado presentaba una actividad de unión a ligando promiscua por un sustrato que surgió millones de años después que el receptor ancestral (Dean y Thornton 2007). Dado el estrecho vínculo entre la unión de ligandos y la catálisis enzimática, este resultado tiene claras implicaciones en la promiscuidad en enzimas.

Por otro lado, poco se conoce acerca de la evolución de la estructura de las proteínas, a pesar del gran número de secuencias depositadas en el *Protein Data Bank*. De hecho, existen bastantes dificultades para clasificar los diferentes tipos de plegamientos (*fold*) que presentan las estructuras y discernir qué caracteres estructurales son originales y cuales son derivados en las estructuras de proteínas. En esta tesis doctoral, proponemos que la resurrección de proteínas ancestrales puede constituir una herramienta muy útil para estudiar la evolución de la estructura de las proteínas.

1.3. Implicaciones en ingeniería de proteínas de la resurrección de proteínas ancestrales

La capacidad de las proteínas para llevar a cabo las funciones moleculares con enorme rapidez y especificidad hace que tengan potencialmente numerosas aplicaciones biomédicas (Lazar, Marshall et al. 2003) y biotecnológicas (Kirk, Borchert et al. 2002). La mayoría de estas aplicaciones requiere, sin embargo, modificaciones en la proteína de interés para modular alguna propiedad (estabilidad, actividad catalítica, inmunogenicidad, farmacodinámica, etc.). Por tanto, el término ingeniería de proteínas se refiere, en general, a la preparación de proteínas con propiedades alteradas mediante la modificación de la secuencia proteica.

En este contexto, algunos estudios de resurrección de proteínas del período Precámbrico han revelado que las proteínas ancestrales son muy estables y presentan mayor actividad catalítica que las proteínas actuales (Gaucher,

Govindarajan et al. 2008; Perez-Jimenez, Ingles-Prieto et al. 2011), lo cual es muy relevante desde el punto de vista de la ingeniería de proteínas.

Esta alteración de las propiedades puede explicarse en términos matemáticos, ya que la ingeniería de proteínas puede verse como el movimiento de un punto, que representa la molécula, en un espacio de propiedad frente a la secuencia proteica (Fig. 1.1). Así, las mutaciones puntuales en la secuencia conducen típicamente a leves modulaciones en las propiedades (el punto se mueve poca distancia en el espacio propiedad/secuencia), mientras que la resurrección de proteínas ancestral, y más aun si extendemos la resurrección a proteínas de periodo precámbrico, implica grandes cambios en la secuencia, que pueden ocasionar grandes modulaciones en la propiedad de las proteínas (el punto se mueve una gran distancia en el espacio propiedad/secuencia). Sin embargo, a pesar de los grandes cambios en la secuencia, la resurrección de proteínas precámbricas conduce a proteínas bien plegadas, estables y activas (Gaucher, Govindarajan et al. 2008; Perez-Jimenez, Ingles-Prieto et al. 2011). Por este motivo, la resurrección de proteínas ancestrales puede convertirse en una de las principales fuentes de la próxima generación de proteínas con aplicaciones biotecnológicas y biomédicas.

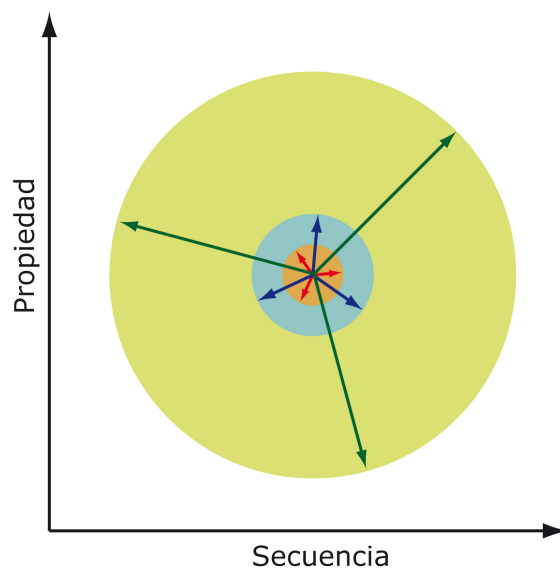


Figura 1.1. Relación entre el rango de modulación en una propiedad de la proteína en función del número de mutaciones realizadas en su secuencia. Con mutaciones puntuales se produce un efecto pequeño en la propiedad (área naranja), con varias mutaciones se produce un efecto mayor en la propiedad (área azul), mientras que con la resurrección de proteínas se producen grandes cambios en la secuencia de una proteína, alcanzándose grandes rangos de modulación en la propiedad (área verde).

1.4. Planteamiento general de esta tesis doctoral: resurrección y caracterización biofísica, estructural y energética de tiorredoxinas resucitadas

El trabajo abordado en la presente memoria ha consistido en la resurrección de tiorredoxinas del período precámbrico y su caracterización energética, funcional y estructural. La consecución de este trabajo ha sido posible gracias a la colaboración de nuestro grupo en el departamento de Química Física (Universidad de Granada), que lidera el Dr. José Manuel Sánchez Ruiz, con el grupo del Dr. Eric Gaucher (Georgia Tech., Atlanta), el grupo del Dr. Julio Fernández (Universidad de Columbia, Nueva York) y grupo del Dr. Juan Manuel García Ruiz (Laboratorio de Estudios Cristalográficos, CSIC).

Como proteína modelo, hemos empleado la tiorredoxina. Las tiorredoxinas son enzimas que catalizan la reducción de puentes disulfuro en todos los organismos vivos (Holmgren 1985). Al ser una proteína ubicua y estar implicada en algunas enfermedades en humanos (Powis y Montfort 2001), los estudios acerca del origen evolutivo de las funciones asociadas con tiorredoxina son de enorme interés. Sin embargo, lo que añade una dimensión especial a este estudio es que el grupo del Dr. Julio Fernández, en colaboración con nuestro grupo, puso a punto hace unos años un procedimiento para estudiar la actividad de varias tiorredoxinas actuales a nivel de molécula individual por microscopía de fuerza atómica (Wiita, Perez-Jimenez et al. 2007). Este estudio reveló dos mecanismos para la reducción de puentes disulfuro catalizada por tiorredoxina, los cuales parecen claramente diferenciados desde un punto de vista químico e incluso evolutivo. Así, uno de los mecanismos se asemeja a la reducción mediante un agente químico convencional y cabe especular que sea el mecanismo original (en términos evolutivos) de las tiorredoxinas. El otro mecanismo implica reconocimiento molecular entre la tiorredoxina y el sustrato proteico y podría tratarse de una adquisición posterior en el transcurso de la evolución.

Para tratar de aclarar el origen evolutivo de estos mecanismos de catálisis hemos empleado la resurrección de tiorredoxinas ancestrales. Con este objetivo surgió la colaboración entre el grupo del Dr. Gaucher, el del Dr. Fernández y nuestro grupo. Para abordar la resurrección y caracterización de las tiorredoxinas ancestrales, en una primera etapa, hemos realizado en nuestro laboratorio la expresión y purificación (resurrección) de las tiorredoxinas ancestrales a partir de las

secuencias reconstruidas por el Dr. Eric Gaucher. A continuación, hemos determinado la estabilidad térmica mediante calorimetría diferencial de barrido, así como la actividad oxido-reductasa de las diferentes tiorredoxinas ancestrales. Una vez purificadas, las hemos enviado al laboratorio del Dr. Julio Fernández donde han estudiado su catálisis enzimática mediante microscopía de fuerza atómica. A parte del origen evolutivo de los mecanismos de catálisis de estas enzimas, este estudio puede esclarecer la manera en que se han adaptado los organismos a los cambios en las condiciones del entorno. Los resultados de este trabajo están recopilados en el capítulo tercero de la presente memoria.

En una segunda etapa, hemos determinado las estructuras tridimensionales de las tiorredoxinas ancestrales resucitadas mediante cristalografía de rayos-X. Este trabajo lo hemos llevado a cabo en el laboratorio del grupo del Dr. Juan Manuel García Ruiz. Además, hemos realizado un análisis comparativo de las estructuras de tiorredoxinas precámbricas con tiorredoxinas actuales de *E. coli* y humana. Indudablemente, el análisis de estas estructuras nos permitirá expandir el alcance y las implicaciones de los estudios de proteínas resucitadas. Cabe destacar que estas son las primeras proteínas precámbricas cuya estructura ha sido determinada experimentalmente. Por ello, nos encontramos en una posición única para entender a nivel molecular la evolución del *fold* de estas enzimas. Los resultados de este trabajo están recopilados en el capítulo cuarto de la presente memoria.

En una tercera etapa, hemos realizado un exhaustivo estudio computacional comparativo de las tiorredoxinas precámbricas y las tiorredoxinas actuales de *E. coli* y humana, centrado principalmente en la evaluación de las distintas contribuciones energéticas a estabilidad de las proteínas: área superficial accesible al disolvente, interacciones hidrofóbicas, número de enlaces de hidrógeno, número de puentes salinos, propensión de residuos de situarse en elementos de estructura secundaria y residuos con efecto sobre la entropía conformacional. Los resultados de este trabajo están recopilados en el capítulo quinto de la presente memoria.

A partir de los resultados obtenidos del estudio anterior, hemos realizado un análisis mutacional experimental de la estabilidad de la tiorredoxina más estable (LPBCA) comparándola con la tiorredoxina menos estable (*E. coli*). Para ello, hemos utilizado un enfoque de bibliotecas combinatoriales de mutantes de tiorredoxina, junto con el método de regresión de mínimos cuadrados parciales para reconstruir el espacio mutacional completo de dichas bibliotecas. Este estudio podrían desvelar

los mecanismos moleculares de las tiorredoxinas para adaptarse a los cambios en el entorno a lo largo de la evolución. Los resultados de este trabajo están recopilados en el capítulo sexto de la presente memoria.

1.5. Objetivos concretos

En resumen, los objetivos concretos que nos hemos fijado al abordar el presente trabajo son los siguientes:

1. Expresión y purificación de las tiorredoxinas ancestrales, cuyas secuencias han sido reconstruida por el Dr. Eric Gaucher.
2. Determinación de la actividad oxido-reductasa de las tiorredoxinas resucitadas empleando técnicas bioquímicas convencionales (ensayo de agregación de insulina).
3. Determinación de la estabilidad térmica de las tiorredoxinas resucitadas mediante calorimetría diferencial de barrido.
4. Determinación de la estructura tridimensional de las tiorredoxinas resucitadas mediante cristalografía de rayos-X.
5. Estudio comparativo de las estructuras tridimensionales de las tiorredoxinas ancestrales y las tiorredoxinas actuales de *E. coli* y humana.
6. Estudio comparativo energético-estructural de las tiorredoxinas ancestrales y actuales de *E. coli* y humana para evaluar los factores relacionados con la estabilidad de proteínas.
7. Análisis mutacional experimental de la hiperestabilidad térmica de la tiorredoxina LPBCA con un enfoque de bibliotecas combinatoriales junto con el método de regresión de mínimos cuadrados parciales.

1.6. Bibliografía

- Benner, S. A., S. O. Sassi and E. A. Gaucher (2007). "Molecular Paleoscience: Systems Biology from the Past." Adv Enzymol Relat Areas Mol Biol **75**: 1-132, xi.
- Bridgham, J. T., G. N. Eick, C. Larroux, K. Deshpande, M. J. Harms, M. E. Gauthier, E. A. Ortlund, B. M. Degnan and J. W. Thornton (2010). "Protein Evolution by Molecular Tinkering: Diversification of the Nuclear Receptor Superfamily from a Ligand-Dependent Ancestor." PLoS biology **8**(10).
- Bridgham, J. T., E. A. Ortlund and J. W. Thornton (2009). "An Epistatic Ratchet Constrains the Direction of Glucocorticoid Receptor Evolution." Nature **461**(7263): 515-519.
- Carroll, S. M., E. A. Ortlund and J. W. Thornton (2011). "Mechanisms for the Evolution of a Derived Function in the Ancestral Glucocorticoid Receptor." PLoS genetics **7**(6): e1002117.
- Dean, A. M. and J. W. Thornton (2007). "Mechanistic Approaches to the Study of Evolution: The Functional Synthesis." Nature reviews. Genetics **8**(9): 675-688.
- Eick, G. N. and J. W. Thornton (2011). "Evolution of Steroid Receptors from an Estrogen-Sensitive Ancestral Receptor." Molecular and cellular endocrinology **334**(1-2): 31-38.
- Finnigan, G. C., V. Hanson-Smith, B. D. Houser, H. J. Park and T. H. Stevens (2011). "The Reconstructed Ancestral Subunit a Functions as Both V-Atpase Isoforms Vph1p and Stv1p in *Saccharomyces Cerevisiae*." Molecular biology of the cell **22**(17): 3176-3191.
- Finnigan, G. C., V. Hanson-Smith, T. H. Stevens and J. W. Thornton (2012). "Evolution of Increased Complexity in a Molecular Machine." Nature **481**(7381): 360-364.
- Gaucher, E. A., S. Govindarajan and O. K. Ganesh (2008). "Palaeotemperature Trend for Precambrian Life Inferred from Resurrected Proteins." Nature **451**(7179): 704-707.
- Holmgren, A. (1985). "Thioredoxin." Annu Rev Biochem **54**: 237-271.
- Kirk, O., T. V. Borchert and C. C. Fuglsang (2002). "Industrial Enzyme Applications." Current opinion in biotechnology **13**(4): 345-351.
- Lazar, G. A., S. A. Marshall, J. J. Plecs, S. L. Mayo and J. R. Desjarlais (2003). "Designing Proteins for Therapeutic Applications." Current Opinion in Structural Biology **13**(4): 513-518.
- Malcolm, B. A., K. P. Wilson, B. W. Matthews, J. F. Kirsch and A. C. Wilson (1990). "Ancestral Lysozymes Reconstructed, Neutrality Tested, and Thermostability Linked to Hydrocarbon Packing." Nature **345**(6270): 86-89.

- Ortlund, E. A., J. T. Bridgham, M. R. Redinbo and J. W. Thornton (2007). "Crystal Structure of an Ancient Protein: Evolution by Conformational Epistasis." Science **317**(5844): 1544-1548.
- Pauling, L. and E. Zuckerkandl (1963). "Chemical Paleogenetics: Molecular "Restauration Studies" of Extinct Forms of Life." Acta. Chem. Scand. A **17**: 7.
- Perez-Jimenez, R., A. Ingles-Prieto, Z. M. Zhao, I. Sanchez-Romero, J. Alegre-Cebollada, P. Kosuri, S. Garcia-Manyes, T. J. Kappock, M. Tanokura, A. Holmgren, J. M. Sanchez-Ruiz, E. A. Gaucher and J. M. Fernandez (2011). "Single-Molecule Paleoenzymology Probes the Chemistry of Resurrected Enzymes." Nature structural & molecular biology **18**(5): 592-596.
- Powis, G. and W. R. Montfort (2001). "Properties and Biological Activities of Thioredoxins." Annu Rev Biophys Biomol Struct **30**: 421-455.
- Wiita, A. P., R. Perez-Jimenez, K. A. Walther, F. Grater, B. J. Berne, A. Holmgren, J. M. Sanchez-Ruiz and J. M. Fernandez (2007). "Probing the Chemistry of Thioredoxin Catalysis with Force." Nature **450**(7166): 124-127.
- Zalatan, J. G. and D. Herschlag (2009). "The Far Reaches of Enzymology." Nature chemical biology **5**(8): 516-520.

2. MATERIALES Y MÉTODOS

2.1. MATERIALES

2.1.1. Productos

Los reactivos que se han utilizado para realizar los experimentos son:

- DIFCO:

Extracto de levadura (Ref. 212750).

Triptona (Ref. 211705)

Agar (Ref. 214010).

- MERCK:

Cloruro sódico. PM=58.44 g/mol (Ref. 1.06404.1000).

Azida sódica. PM=55.01 g/mol (Ref. 1.06688.1000).

Acetato sódico. PM=82.03 g/mol (Ref. 1.06268.1000).

NaH₂PO₄·H₂O. PM=137.99 g/mol (Ref. 1.06346.1000).

Na₂HPO₄·2H₂O. PM=177.99 g/mol (Ref. 1.06580.1000).

Hidróxido sódico. PM=40 g/mol (Ref. 6498.1000).

Imidazol. PM=68.08 g/mol (Ref. #1.04716.0250).

- SIGMA:

EDTA (Ácido EtilenDiaminoTetrasódico). PM=372.2 g/mol(Ref. E-5134).

Trizma Base. PM=121.14 g/mol (Ref. T-6066).

Brilliant Blue G. PM=826.0 g/mol (Ref. R-250).

Sal sódica de HEPES. PM=260.29 g/mol (Ref. H-3784).

HEPES ≥99.5%. PM= 238.31 g/mol (Ref. H-4034).

Poliacrilamida. PM=71.08 g/mol (Ref. K-31503209/324).

IPTG. PM=238.31 g/mol (Ref. I6758).

Kanamicina A. PM=582.6 g/mol (Ref. K-1377).

Glicerol. PM=92.09 g/mol (Ref. G-7893).

- PANREAC:

Ácido clorhídrico 37%. PM=36.48 g/mol (Ref. 131020.1611).

- J.T. BAKER:

Ácido acético glacial. PM=60.05 g/mol (Ref. 6052).

Para realizar la mutagénesis dirigida se utilizaron los siguientes materiales:

- AGILENT TECHNOLOGIES:

Kit de mutagénesis dirigida *QuickChange Lightning* (Ref. #210518).

Células Supercompetentes XLI-Blue. (Ref. #200236).

Células Supercompetentes BL21-DE3 (Ref. #69387-3).

- BECTON-DICKINSON:

Tubos Falcon[®] 2059 de polipropileno (Ref.#352059).

- NOVAGEN:

Vector pET30A(+).

- QIAGEN:

Kit de purificación de ADN QIAprep[®] Spin Miniprep (Ref.#27104).

Vector pQE80L.

Los materiales que se usaron para la cristalización de proteínas son:

- HAMPTON RESEARCH:

Kits de muestreo de cristalización PEGRx 1 (Ref. # HR2-082) y PEG/Ion screen (Ref. # HR2-126).

- TRIANA SCIENCE & TECHNOLOGY:

Capilares de cristalización de 0.1 mm de diámetro (Ref.# CP-01-50), 0.2 mm de diámetro (Ref.# CP-02-50) y de 0.3 mm diámetro (Ref.# CP-03-50).

Kit de muestreo de cristalización por contradifusión GCB-Dominó 24 condiciones (Ref.# GCB-CSK-24).

Kit de optimización de cristalización por contradifusión GCB-D PEG 400-4000-8000 (MixPEG) (Ref.# KIT-PEG448-49).

Kit de optimización de cristalización por contradifusión GCB-D de Sulfato de Amonio (Ref.# KIT-AS-49).

En la purificación de las tiorredoxinas se utilizó:

- SPECTRA/POR:

Membranas de diálisis con un punto de corte entre 6000-8000 Daltons (diámetro: 23 mm, Ref.#132650; diámetro:50 mm, Ref.#132665).

- GE HEALTHCARE:

Filtros de 0.20 (Ref. #6780-1302) y 0.45 μ m de tamaño de poro (Ref. #6780-1304).

- BIORAD

Marcador de peso molecular para proteínas de rango amplio (Ref. 161-0317).

2.1.2. Instrumentación.

Para el desarrollo de este trabajo se han utilizado los siguientes instrumentos:

Autoclave "Autester-D". (SELECTA)

Estufa de cultivos "Incubat". (SELECTA)

Para realizar la mutagénesis dirigida se utilizó:

Termociclador modelo *Matercycler gradient*. (EPPENDORF)

Los experimentos de cristalización de alto rendimiento de procesamiento se han hecho en:

Sistema de cristalización Hamilton Star Plus. (HAMILTON)

La determinación de la concentración de proteínas y los ensayos de actividad enzimática con insulina se ha realizado en:

Espectrofotómetro UV-visible modelo 8453. (AGILENT TECHNOLOGIES)

Para la purificación de las proteínas se usaron los siguientes instrumentos:

Centrífuga Modelo H-251, con rotor M1. (KOKUSAN)

Centrífuga modelo Roto Super 40, con rotor A.69. (HETTICH)

Ultracentrífuga Optima™ LE-80K, con rotor 45Ti. (BECKMAN COULTIER)

Microcentrífuga modelo 5424. (EPPENDORF)

pH-metro modelo Digit-501, de sensibilidad 0.01 unidades con dos tipos de electrodos de vidrio modelos 52-03 "clásico" para volúmenes grandes (diámetro de bulbo 10 mm) y 52-09 "micro" para volúmenes pequeños (diámetro de bulbo de aproximadamente 5 mm). (CRISON)

Dispositivos de concentración por centrifugación *Vivaspin* de tamaño de poro 3000 Da de 20 ml (Ref. #VS2091), de 6 ml (Ref. #VS0691). (SARTORIUS)

Sistema de electroforesis Mini-PROTEAN® Tetra. (BIORAD)

Fuente eléctrica PowerPac™ Basic. (BIORAD)

Columna de gel-filtración modelo Superdex 75, 16/60 *prep. grade* (GE-HEALTHCARE).

Columna de intercambio iónico modelo XK-50 50/20 rellena de resina Fractogel® EMD DEAE (M) (MERCK).

Columna de gel-filtración para eliminación de sales PD-10 (GE-HEALTHCARE).

Columna de afinidad de Níquel His GraviTrap con 1ml de resina (GE-HEALTHCARE).

Sistema de FPLC (Fast Protein Liquid Chromatography) ÄKTA Prime plus (GE-HEALTHCARE)

Para los experimentos de Calorimetría Diferencial de Barrido se ha utilizado:

Microcalorímetro modelo VP-DSC (MICROCAL INC).

Dos microcalorímetros modelo VP-DSC capilar con sistema de procesamiento de alto rendimiento (robot) (MICROCAL INC)

El análisis de los resultados se ha realizado con:

Ordenador MacBook Pro con sistema operativo Lion y Microsoft Windows XP. (APPLE)

Ordenador Amilo M1451G con sistema operativo Microsoft Windows XP (FUJITSU-SIEMENS).

2.2. MÉTODOS

2.2.1. Mutagénesis dirigida

La mutagénesis dirigida *in vitro* es una técnica de gran importancia que nos ayuda a caracterizar la compleja relación entre la estructura y la función de las proteínas. Esta técnica nos permite realizar mutaciones individuales y múltiples en un plásmido de doble cadena teniendo la posibilidad de cambiar, insertar o eliminar uno o varios aminoácidos de la secuencia de cualquier proteína. Para llevar a cabo este procedimiento, hemos usado el kit de mutagénesis dirigida *QuickChange[®] Lightning* (Qiagen). Este kit no requiere vectores específicos, sitios de restricción únicos, transformaciones múltiples ni ninguna etapa de metilación *in vitro*. Además, está optimizado para vectores de un tamaño menor de 14 kb e incluye una mezcla de ADN polimerasas modificadas, basadas en la ADN polimerasa de *Pirococcus furiosus* o *PFU*, que poseen una altísima fidelidad. La duración de todo el proceso es de dos horas aproximadamente y obtenemos el producto listo para transformarlo en bacterias (Fig. 2.1.).

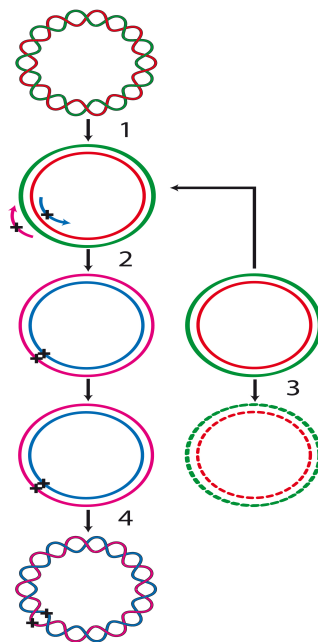


Figura 2.1. Representación esquemática del método de mutagénesis dirigida *QuickChange[®] Lightning*. El paso 1 consiste en la desnaturalización térmica del plásmido parental (verde-rojo). A continuación se muestra el anillamiento a las cadenas del plásmido parental de los oligonucleótidos promotores portando la mutación (magenta-azul). El paso 2 es la extensión de las cadenas de ADN por la ADN-polimerasa. El paso 3 consiste en la digestión enzimática de la cadena parental metilada y hemimetilada con *DpnI*. En el paso 4, el plásmido está listo para la transformación en bacterias.

Para llevar a cabo la mutagénesis dirigida, necesitamos un plásmido de ADN de doble cadena, que tenga clonado el gen que expresa la proteína que queremos mutar (plásmido parental), y dos oligonucleótidos cebadores sintéticos de cadena sencilla, que contengan la mutación deseada. Para diseñar los oligonucleótidos cebadores, hemos utilizado el programa *QuickChange® Primer Design* (<https://www.genomics.agilent.com>). Para ejecutarlo, tenemos que introducir la secuencia de nucleótidos de la proteína que queremos mutar, seleccionar la mutación deseada y la posición donde queremos dicha mutación. A continuación, el programa genera las secuencias óptimas de los oligonucleótidos directo y reverso que contienen dicha mutación. Estos oligonucleótidos han sido sintetizados por la compañía MWG Biotech (www.mwg-biotech.com).

Una vez que tenemos los oligonucleótidos cebadores, el plásmido parental y el kit de mutagénesis, podemos llevar a cabo la reacción. Para realizar la mezcla de reacción, hay que tener en cuenta que la cantidad de oligonucleótidos cebadores estén en exceso respecto a la cantidad de plásmido parental; así, nos aseguramos de que no existirán problemas durante la reacción. La mezcla de reacción consta de los siguientes reactivos:

- 40 μ l de agua bidestilada.
- 5 μ l de 10x tampón de reacción.
- 0.5 μ l plásmido parental (100ng/ μ l).
- 1 μ l oligonucleótido directo (125ng/ μ l).
- 1 μ l oligonucleótido reverso (125ng/ μ l).
- 1 μ l de mezcla de dNTPs.
- 1.5 μ l de reactivo *QuickSolution®*.
- 1 μ l de Mezcla de Enzimas ADN Polimerasa *QuickChange® Lightning*.

La reacción de mutagénesis se lleva a cabo en un termociclador y consta de varias etapas (Fig. 2.1.). La primera etapa es la desnaturalización térmica del vector. Tras la desnaturalización, comienza la etapa de anillamiento, en la cual los oligonucleótidos cebadores hibridan con la cadena de ADN complementaria del plásmido parental. En la etapa de elongación, la enzima DNA polimerasa se encarga de la adición de nucleótidos a ambas cadenas usando como molde el plásmido parental. Esta etapa se realiza con una alta fidelidad sin desplazar los oligonucleótidos cebadores. Los parámetros que hemos programado en el

termociclador, siguiendo las instrucciones del kit de mutagénesis, son los siguientes:

Tabla 2.1. Parámetros del termociclador para el método de mutagénesis dirigida empleando el kit *QuickChange® Lightning*:

Etapa	Nº de ciclos	Temperatura	Tiempo
Desnaturalización	1	95°C	2 minutos
		95°C	20 segundos
Anillamiento	18	60°C	10 segundos
Elongación	1	68°C	30 segundos/kb de plásmido
		68°C	5 minutos

Una vez finalizado el programa del termociclador, tratamos la mezcla de reacción con 2 μ l de la enzima endonucleasa *Dpn I* y la incubamos 5 minutos a 37°C, Esta enzima es específica para ADN metilado y hemimetilado; por tanto, digierá el ADN parental procedente de *E. coli* para seleccionar solamente el plásmido que contiene la mutación. En este momento, tenemos el producto listo para transformar las células supercompetentes *XL1-Blue* como se indica en el apartado 2.2.2.1.

2.2.2. Transformación de bacterias y expresión de tiorredoxinas

La transformación consiste en introducir un plásmido externo dentro de la bacteria. Esta técnica se basa en la función original de un plásmido, que es transferir información genética vital para la supervivencia de la célula.

Los genes que codifican las tiorredoxinas ancestrales están clonadas entre los sitios *BamHI* y *KpnI* del plásmido de expresión pQE80L (Qiagen) (Fig. 2.2. A). Este vector está diseñado para permitir la expresión controlada de proteínas en cualquier cepa de *E. coli*. Para facilitar la purificación posterior, las proteínas tienen una cola de 6 histidinas en el extremo N-terminal. El vector pQE80L combina un potente promotor de fago T5 (reconocido por la RNA polimerasa de *E. coli*) con un sistema de represión del operón *lac*. Esto provoca que la expresión basal de proteína antes de la inducción por isopropil- β -D-1-tiogalactopiranosido (IPTG) sea muy baja. El gen de resistencia a ampicilina que contiene el vector permite la selección de los clones de bacterias que tienen el plásmido inserto.

Para realizar los experimentos de cristalización, hemos usado proteínas sin cola de histidinas. En este caso, los genes están clonados entre los sitios *NdeI* y *XhoI* del plásmido de expresión pET30A(+) (Novagen) (Fig. 2.2. B). Este plásmido emplea el sistema promotor T7, que junto al sistema de represión *lacI* hace que la expresión de proteínas recombinantes sea muy eficiente. La cola de histidinas en el extremo N-terminal que incluye el plásmido se elimina al clonar entre los sitios de restricción utilizados. Este vector tiene un gen de resistencia a kanamicina, que permite la selección de los clones de bacterias con el plásmido inserto.

Por otro lado, los genes codificantes de las variantes de las bibliotecas combinatoriales están clonados entre los sitios *NdeI* y *XhoI* del vector pET24(+) (Novagen) (Fig. 2.2. C). Este plásmido es similar al pET30A(+), ya que también posee el sistema promotor T7, el sistema de represión *lacI* y un gen de resistencia a kanamicina. En esta ocasión, se ha conservado la cola de histidinas en el extremo N-terminal de la proteína.

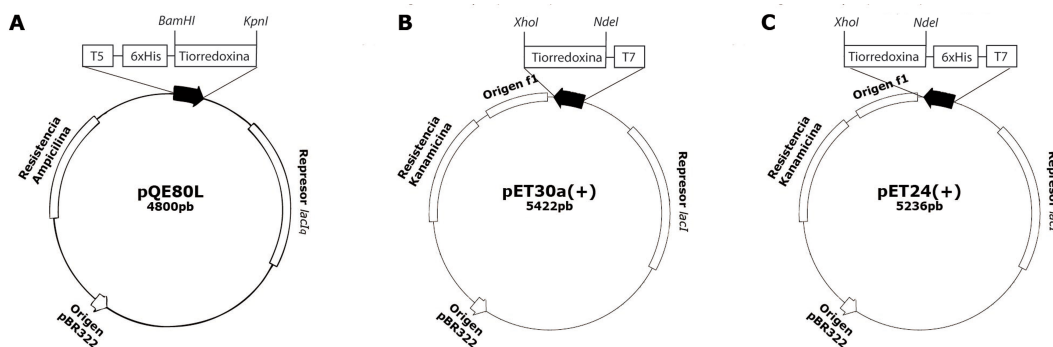


Figura 2.2. Esquemas de los plásmidos utilizados en este trabajo con el gen de tiorredoxina de *E. coli* clonado. A: plásmido pQE80L. B: plásmido pET30a(+). C: plásmido pET24(+)

Para que la transformación tenga éxito, usamos cepas de bacterias modificadas genéticamente denominadas **cepas supercompetentes**. Para la amplificación del plásmido insertado hemos utilizado células *XL1-Blue* (Agilent Technologies). Las células elegidas para expresar las tiorredoxinas son *BL21 (DE3)* (Agilent Technologies). Ambas son cepas de bacterias modificadas de la especie *Escherichia coli*.

2.2.2.1. Transformación en bacterias *XL1-Blue*

Hemos utilizado células *XL1-Blue* para la amplificación del plásmido inserto. Estas células son deficientes en endonucleasa (*endA*), lo cual mejora la calidad del ADN aislado. Además, tiene mutado el gen responsable de la recombinación (*recA*), lo que incrementa la estabilidad del inserto. La mutación *hsdR* previene la rotura del ADN por el sistema de endonucleasa *EcoK*.

La transformación la realizamos siguiendo el protocolo comercial proporcionado por *Agilent Tech.* con algunas modificaciones: mezclamos 20 μ l de células supercompetentes con 50-100 ng de ADN de interés. Incubamos la mezcla 30 minutos en hielo y suministramos un choque térmico de 30 segundos a 42°C. Este paso es fundamental para que la transformación tenga éxito. A continuación, incubamos 2 minutos en hielo y añadimos 500 μ l de medio NZY⁺ precalentado a 42°C. Después de incubar a 37°C el producto de transformación, transferimos 180 μ l a una placa de LB-agar suplementado con el antibiótico correspondiente (30 μ g/ml si es kanamicina o 100 μ g/ml si es ampicilina) e incubamos durante 16 horas a 37°C.

2.2.2.2. Aislamiento de plásmidos

Con el fin de extraer y purificar el plásmido que se ha amplificado en el paso anterior, se lleva a cabo un aislamiento de plásmido o *miniprep* con el kit de *QIAprep spin* (Quiagen).

Las columnas incluidas en el kit contienen una membrana de gel de sílice a la que se une el ADN plasmídico en presencia de una alta concentración de sal caotrópica. Esto permite la elución en un pequeño volumen de tampón de bajo contenido en sal o agua. El ADN plasmídico que se obtiene es de una alta pureza y está listo para ser utilizado.

El protocolo de la purificación del plásmido sigue un simple proceso de lavado-elución. En todos los casos, hemos seguido el protocolo facilitado por el fabricante: en primer lugar, lisamos los cultivos bacterianos y los centrifugamos. Aplicamos el sobrenadante a la columna *QIAprep* donde se adsorbe el plásmido a la membrana de gel de sílice. Las impurezas son arrastradas por los sucesivos lavados. Por último, extraemos el ADN puro de la columna con 50 μ l de tampón de elución o

agua. El kit de miniprep *QIAprep* permite la purificación de hasta 20 µg de plásmido en grado de biología molecular listo para usarse inmediatamente en la secuenciación o en procedimientos de clonaje.

Para asegurarnos de que la secuencia de ADN de la proteína está correctamente insertada en el plásmido y no hay mutaciones indeseadas, debemos secuenciarlo. Este servicio lo realiza la compañía alemana *Eurofins MWG Operon* (www.eurofinsdna.com). La preparación de la muestra consiste en introducir en un tubo de 1.5 ml, 15 µl de ADN purificado a una concentración de 50-100 ng/µl. Si la concentración es menor, secamos unos 30 µl de la muestra a 60°C. Antes de enviar las muestras, debemos identificarlas con unas etiquetas y seleccionar en el servidor web de la compañía el promotor de nuestro plásmido.

2.2.2.3. Transformación en bacterias *BL21 (DE3)*

Una vez hemos comprobado que la secuencia de la proteína es la correcta, el siguiente paso es transformar bacterias supercompetentes especializadas en sobreexpresión de proteínas. Las que hemos utilizado son una cepa de *Escherichia coli* denominada *BL21 (DE3)*. Esta cepa nos garantiza un elevado nivel de sobreexpresión de proteína y son fáciles de inducir mediante IPTG. Al ser cepas modificadas *E. coli* B no tienen ni proteasa *lon* ni la proteasa de la membrana externa *ompT*; capaces de degradar la proteína recombinante durante la purificación.

La transformación la hemos realizado siguiendo el protocolo comercial con ciertas modificaciones: Añadimos 20 µl de células supercompetentes y 50-100 ng de ADN. Incubamos la mezcla 30 minutos en hielo y damos un choque térmico de 30 segundos a 42°C. Luego, incubamos 2 minutos en hielo y añadimos 500 µl de medio NZY⁺ precalentado a 42°C. Después, incubamos a 37°C con agitación durante una hora y transferimos 180 µl del producto de transformación a una placa de LB-agar suplementado con el antibiótico correspondiente. Por último, incubamos las placas durante 16 horas a 37°C.

2.2.2.4. Pruebas de sobreexpresión de tiorredoxinas

Si el proceso de transformación es exitoso, crecerán colonias en la placa. Antes de empezar la purificación a gran escala, es necesario probar el nivel de sobreexpresión de algunas de estas colonias, ya que puede haber diferencias en la cantidad de proteína producida.

El test de expresión es un protocolo diseñado para comprobar la sobreexpresión de genes bajo el control de promotores inducibles por IPTG en escala analítica. De esta forma, podemos saber rápidamente cuales son las colonias que expresan la proteína y en qué cantidad. El protocolo es el siguiente:

En primer lugar, inoculamos las colonias en 2 ml de LB más el antibiótico correspondiente y las incubamos a 37°C, 200 r.p.m. hasta una densidad óptica a 600nm (D.O.₆₀₀) de 0.5-0.8 U.A. Una vez alcanzada este valor, inoculamos 5 ml de nuevo medio LB con el antibiótico correspondiente con 40 µl del preinóculo y lo incubamos a 37°C, durante 2 a 3 horas hasta obtener una D.O.₆₀₀ de 0.5-0.6. En este momento, transferimos una alícuota de 0.5 ml del cultivo para el control negativo (no inducido) y otra de 1 ml para hacer los gliceroles y poder conservar esta colonia (en caso de que se exprese la proteína). Al cultivo restante, le añadimos IPTG a una concentración final de 0.4 mM para inducir la sobreexpresión de proteína recombinante, y lo incubamos durante 2 o 3 horas a 37°C. A continuación, transferimos 0.5 ml a un tubo que será la muestra inducida. Para lisar las células, tomamos 15 µl de muestra inducida y no inducida y añadimos 15 µl de *buffer* de carga (que contiene beta-mercaptoetanol). Nos aseguramos de mezclar bien, pipeteando arriba y abajo; tras lo cual, hervimos las muestras durante 5 minutos y centrifugamos 10 min para precipitar el material insoluble de las bacterias. Por último, cargamos 15 µl de las muestras y el marcador de peso molecular en un gel de 15% de poliacrilamida y llevamos a cabo una electroforesis en gel de poliacrilamida con dodecilsulfato sódico (SDS-PAGE).

En el gel de poliacrilamida (Fig. 2.3.) observamos si se ha producido la sobreexpresión de proteínas de forma correcta. Para que así sea, en la muestra inducida deberá aparecer una banda más intensa con el peso molecular aproximado de la proteína (usamos marcadores proteicos de peso molecular conocido), mientras que en la muestra no inducida, no debería aparecer ninguna banda o ser muy débil a ese mismo peso molecular.

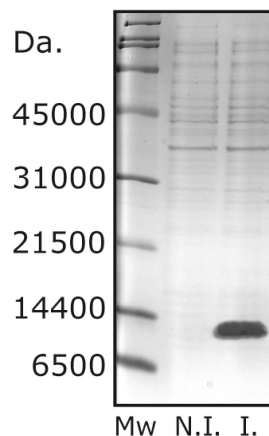


Figura 2.3. Gel SDS-PAGE obtenido de la prueba de expresión de la tiorredoxina de *E. coli* His-tagged (~12000 Da). Mw son los marcadores de peso molecular, N.I. es la muestra no inducida e I es la muestra inducida con IPTG.

El clon que presente la mayor sobreexpresión; es decir, cuya banda sea la más intensa y tenga el peso molecular correspondiente, debemos conservarlo. Para ello, añadimos 1 ml de glicerol al 60% (actúa como crioprotector de las células) a 1 ml de cultivo sin inducir (que habíamos reservado), lo mezclamos bien y lo congelamos a -80°C . Más adelante, podremos partir de este glicerol para la purificación de la proteína a mayor escala sin necesidad de volver a transformar.

2.2.3. Purificación de tiorredoxinas

En este momento, tenemos un reservorio congelado de bacterias con la secuencia correcta insertada en un plásmido de expresión (comprobado mediante secuenciación) y que sobreexpresan gran cantidad de proteína. Por tanto, podemos realizar la purificación de la tiorredoxina a mayor escala. En este trabajo, hemos realizado la purificación utilizando diferentes métodos: para purificar las proteínas que tienen anclada una cola de histidinas, hemos usado el método de afinidad empleando columnas de níquel-NTA (matriz de agarosa derivatizada con ácido nitrilotriacético que actúa como quelante de los iones de níquel divalente Ni^{2+}), y para purificar las proteínas sin cola de histidinas, hemos utilizado el método de cromatografía de intercambio iónico seguida de una cromatografía de filtración en gel.

En todos los casos, hemos comprobado la pureza de la tiorredoxina por inspección óptica, mediante SDS-PAGE. La proporción de poli(acrilamida) de todos los geles

fabricados en este trabajo es del 15% en el gel separador y del 5% en el gel apilador.

2.2.3.1. Purificación de tiorredoxina *His-tagged*

Como ya hemos mencionado el apartado 2.2.1, con el fin de facilitar la purificación de las proteínas recombinantes, hemos clonado los distintos genes en el vector que codifica una cola de 6 histidinas consecutivas en el extremo N-terminal.

El grupo de histidinas tiene capacidad para quelar cationes divalentes como el níquel (Fig. 2.4). Como consecuencia, si pasamos un lisado bacteriano por una resina con estos iones inmovilizados, la proteína recombinante quedará adherida específicamente mientras que el resto de componentes atravesará la columna.

Para comenzar la purificación, tomamos un pequeño volumen de cada glicerol raspando con una punta de pipeta, y lo inoculamos en 10 ml de LB con 100 $\mu\text{g/ml}$ de ampicilina o 30 $\mu\text{g/ml}$ de kanamicina. Dejamos crecer las bacterias durante toda la noche a 37°C y con agitación.

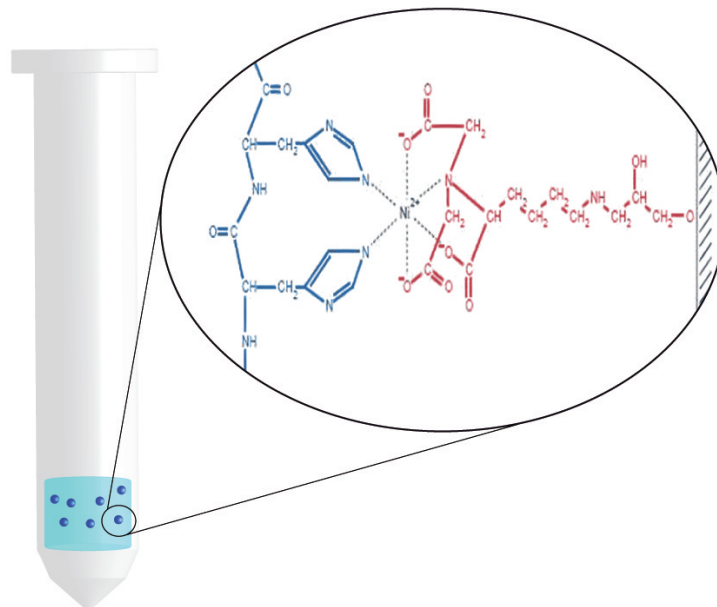


Figura 2.4. Columna de Níquel con la proteína (bolas azules) unida a la matriz Ni-NTA (verde). La ampliación muestra la interacción entre las cadenas laterales de las histidinas (azul) y la matriz de Ni²⁺-NTA (rojo). Figura de la ampliación tomada de www.gelifesciences.com.

Al día siguiente, vertemos el preinóculo en 1 L de LB suplementado con el antibiótico correspondiente y lo incubamos a 37°C hasta alcanzar una D.O.₆₀₀ de 0.6 a 0.8 unidades de absorbancia. Una vez que las bacterias han alcanzado este grado de crecimiento, se centrifuga el cultivo a 7.000 r.p.m., durante 10 minutos a una temperatura de 4°C. A partir de este momento, es fundamental mantener las muestras a una temperatura de 4°C para inhibir a las proteasas celulares.

A continuación, resuspendemos el *pellet* en 60 ml de *binding buffer* frío (20mM Fosfato Sódico, 500mM NaCl y 20mM de Imidazol, pH 7.4). Este *buffer* mantiene las histidinas con un nivel de protonación adecuado para que se unan fuertemente al catión divalente, dado que el pH es un parámetro especialmente importante. La elevada concentración de sal del tampón impide las interacciones electrostáticas inespecíficas que podrían darse entre otras proteínas y la matriz. Para facilitar la lisis posterior, podemos congelar la solución de bacterias en esta etapa.

Para romper las células y liberar la proteína producida, hemos lisado las células mediante sonicación durante 6 ciclos, suministrando pulsos de 30 segundos a un 65% de potencia y esperando 40 segundos para evitar el calentamiento de la muestra. Como precaución, hemos mantenido los tubos en hielo durante todo el proceso. Tras la lisis, centrifugamos el lisado de bacterias en la ultracentrífuga a 20.000 r.p.m. durante 20 minutos a 4°C.

Posteriormente, filtramos el sobrenadante que resulta y lo introducimos en las columnas de afinidad de níquel *His GraviTrap* que operan por gravedad. El volumen de resina es de 1 ml, lo cual nos permite purificar hasta un máximo de aproximadamente 40 mg de proteína. Hay que recoger una alícuota del residuo que fluye a través de la columna para someterlo a electroforesis.

Tras pasar el volumen completo de sobrenadante, lavamos con 10 ml de *binding buffer*, que arrastrará muchos de los contaminantes que se adhieren inespecíficamente a la resina; también tomamos una alícuota del eluido de la columna para electroforesis. En este momento, idealmente, solo estará pegada a la resina nuestra proteína. Para extraerla, añadimos 3 ml de *elution buffer* (20 mM Fosfato Sódico, 500 mM NaCl y 500 mM de Imidazol, pH 7.4). El imidazol competirá con nuestra proteína por adherirse a los cationes divalentes y así la desplazará, provocando su expulsión de la columna. Tomamos una alícuota del eluido para SDS-PAGE y conservamos el resto a 4°C.

2.2.3.2. Purificación de tiorredoxina sin *His-tag*

Este tipo de purificación se ha utilizado para las tiorredoxinas sin cola de histidinas clonadas en el vector pET-30A(+). Las tiorredoxinas resultantes serán utilizadas para la cristalización; por tanto, necesitamos una gran cantidad. El proceso es algo más laborioso, ya que se lleva a cabo empleando dos métodos cromatográficos consecutivos y un cuidadoso proceso de concentración y diálisis.

El cultivo de bacterias y la inducción de la sobreexpresión de proteína es parecido al que hemos explicado en el apartado anterior: primero, inoculamos un raspado del glicerol de la proteína en 10 ml de medio LB con 30 µg/ml de kanamicina y lo cultivamos a 37°C durante la noche. Al día siguiente, transferimos el preinóculo a 2 L de LB suplementado con kanamicina y lo cultivamos hasta una D.O.₆₀₀ de 0.6 a 0.8 unidades de absorbancia. Luego, inducimos la sobreexpresión de proteína con 0,4 mM de IPTG e incubamos 5 horas a 37°C. Transcurrido este tiempo, centrifugamos las células a 7000 r.p.m. durante 10 min a 4°C; las resuspendemos en 50 ml de *buffer* TE (30 mM Tris, 1 mM EDTA pH 8.3), que contiene 25 unidades por mililitro de la nucleasa *Benzonasa*[®] (*Sigma*) para romper el ADN y un comprimido del cóctel de inhibidores de proteasas (*Roche*), para evitar la acción de las proteasas bacterianas. Para conseguir la lisis de las células procedemos como se describe en el apartado anterior.

A partir de este momento, el protocolo que seguimos es diferente al practicado en la purificación con colas de histidinas: incubamos el lisado durante 15 minutos a temperatura ambiente para que pueda actuar la nucleasa y poder obtener una muestra sin contaminación por ADN. Además, la viscosidad será menor y ello contribuirá a que los sucesivos pasos de centrifugación y filtrado sean más rápidos. Centrifugamos el lisado a 20000 r.p.m. durante 20 minutos a 37°C para que los componentes celulares insolubles precipiten. Por último, filtramos el sobrenadante con filtros de 45 µm de tamaño de poro.

- **Cromatografía de intercambio iónico**

En la siguiente etapa, llevamos a cabo una primera purificación del lisado mediante **cromatografía de intercambio iónico**. En este tipo de cromatografía, la fase estacionaria tiene grupos cargados positivamente dentro de la resina (*Fractogel*[®] EMD-DEAE M), que interaccionan electrostáticamente con las regiones cargadas negativamente de la superficie de las proteínas. Esta resina está indicada para la

purificación de proteínas ácidas o neutras y para la eliminación de ácidos nucleicos. Las proteínas son retenidas de forma eficiente en la resina cuando el pH del *buffer* es, al menos, una unidad superior al punto isoeléctrico de la proteína. Por tanto, la fuerza de esta unión depende del sistema tamponante; el pH del *buffer*, que determina la carga de la superficie de la proteína; el grado de ionización de los grupos de la fase estacionaria, la concentración de contraiones y la capacidad de unión de la resina.

La columna está conectada a un sistema de *Fast Protein liquid Chromatography* (FPLC). Este sistema está compuesto por una bomba peristáltica que introduce, mediante un sistema de válvulas, la solución de proteína en la columna a una velocidad de flujo constante. Además, nos permite monitorizar el proceso de purificación porque está conectado a un ordenador, que registra el espectro de absorción UV, la conductividad, el flujo y el pH gracias a las diferentes sondas que tiene el instrumento. De esta manera, cuando pasa proteína a través del espectrofotómetro integrado en el sistema, observaremos un pico en la gráfica de absorbancia a 280 nm frente a tiempo/volumen (cromatograma). Además, consta de un colector de fracciones, que recoge y separa en distintos tubos todo lo que fluye de la columna.

El volumen de resina dentro de la columna es de 180 ml, más que suficiente para que no se colapse con la solución de proteína que inyectamos. Antes de cargar la proteína en la columna, esta debe estar equilibrada con *buffer* TE. Entonces, podemos inyectar el sobrenadante filtrado a través del FPLC a una velocidad de flujo de 2 ml/min. Posteriormente, procedemos al lavado de la columna con 250 ml de TE para eliminar todo aquel material celular que se haya pegado a la resina inespecíficamente.

Para extraer la proteína, es necesario pasar una elevada concentración de iones, ya que el equilibrio de intercambio iónico puede considerarse como una competencia entre los iones que se desplazan por la fase móvil por unirse a los grupos cargados del intercambiador. Tenemos que eluir de manera gradual mediante un gradiente iónico, así provocaremos un debilitamiento paulatino de las interacciones con los grupos cargados de la resina y obtendremos nuestra proteína más pura, debido a que los contaminantes se van eliminando a diferentes concentraciones de sal.

Este gradiente de concentración de sal se consigue mediante un sistema de válvulas y dos reservorios desde los cuales el FPLC succiona las distintas disoluciones. Uno de los reservorios (A) contiene *buffer* TE y el otro (B) TE con 1 M NaCl, ambos a pH 8.3. A este pH, la tiorredoxina está cargada positivamente; por tanto, interactuará con la resina de forma eficiente. En este punto, programamos en el FPLC un gradiente de 0% a 100% de volumen del *buffer* B (TE+1 M NaCl), hasta un volumen total de 600 ml y una velocidad de flujo de 2 ml/min. Además, establecemos que recoja fracciones de 8 ml para poder aislar la proteína de interés.

Como ya hemos señalado, el FPLC va a medir la absorbancia a 280 nm del eluido y la va registrar en un cromatograma, con unos picos que corresponderán a las proteínas contenidas en la muestra (Fig. 2.5.). Debido a que cada proteína interactúa de forma diferente con la resina, no sabemos en qué fracciones se encuentra la tiorredoxina. Para detectarla, realizamos un *SDS-PAGE* con las fracciones de los distintos picos.

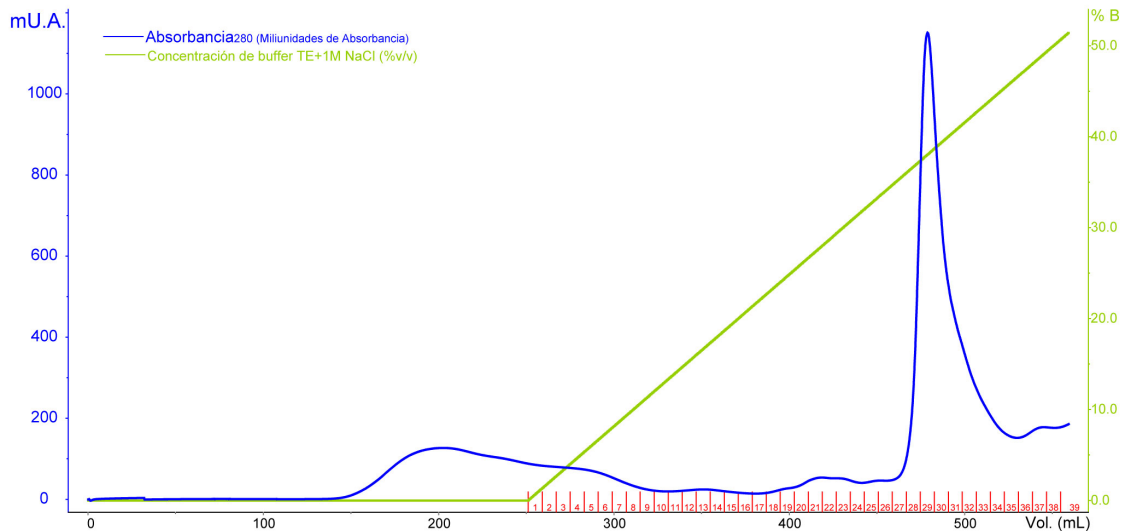


Figura 2.5. Cromatograma de la purificación de la tiorredoxina de *E. coli* mediante cromatografía de intercambio iónico. En azul se muestra la Absorbancia₂₈₀ en función del volumen. En verde se muestra la concentración del *buffer* TE+1M NaCl en función del volumen y en rojo se muestran las diferentes fracciones recogidas.

- **Cromatografía de filtración en gel.**

En la siguiente etapa, purificamos la proteína mediante cromatografía de filtración en gel. Este último paso es fundamental para obtener una proteína con un grado de pureza muy elevado. La filtración en gel puede ser aplicada para separar una amplia gama de moléculas, incluyendo proteínas y enzimas, polisacáridos y ácidos

nucleicos de acuerdo a su tamaño. Hay dos categorías principales de cromatografía de filtración en gel: **separación en grupos** (*desalting*) y **fraccionamiento**. En la separación en grupos, las muestras se separan en dos grupos principales. Se utiliza, por ejemplo, para eliminar las sales y otros contaminantes de bajo peso molecular que se quedan atrapadas en las esferas de la resina mientras que de las proteínas de mayor tamaño son expulsadas. El método de fraccionamiento se utiliza para la separación de macromoléculas de diferentes tamaños. El rango de separación de la resina define la gama de pesos moleculares que son divisibles. Las aplicaciones típicas del fraccionamiento utilizando filtración en gel incluyen la purificación y determinación del peso molecular de proteínas, péptidos y ácidos nucleicos.

La resina que hemos empleado es *HiLoad™ Superdex™ 75 (GE Healthcare)* de grado preparativo. Está compuesta por una matriz de dextrano y agarosa altamente reticulada. La selectividad del dextrano unido a la estabilidad fisicoquímica de la agarosa posibilitan una separación de alta resolución (rango de separación 3000-70000 Da).

Para llevar a cabo la inyección de proteína en la columna, primero reunimos las fracciones que contienen tiorredoxina (comprobadas por SDS-PAGE) y las concentramos por centrifugación en un *Centricon (Sartorius)* hasta 5 ml. Esta solución de proteína se filtra con filtros de 45 μm de tamaño de poro. A continuación, inyectamos la proteína mediante el FPLC en la columna a un flujo de 1 ml/min. La resina debe estar previamente equilibrada con *buffer* HEPES 10 mM, 150 mM de NaCl pH 7.0. Una vez inyectada, programamos el FPLC para que extraiga la proteína con este *buffer* a un flujo de 1 ml/min y recoja fracciones de 5 ml.

Como en el caso anterior, se genera un cromatograma (ABS_{280} frente a volumen/tiempo) (Fig. 2.6A) donde observamos una serie de picos que corresponden a proteínas con distintos pesos moleculares. Las tiorredoxina de *E. coli* y las tiorredoxinas ancestrales reconstruidas se extraen a un volumen de 80-90 ml. Comprobamos las fracciones correspondientes a este pico por SDS-PAGE (Fig. 2.6 B) y reunimos solo las más puras e intensas y las guardamos a 4°C.

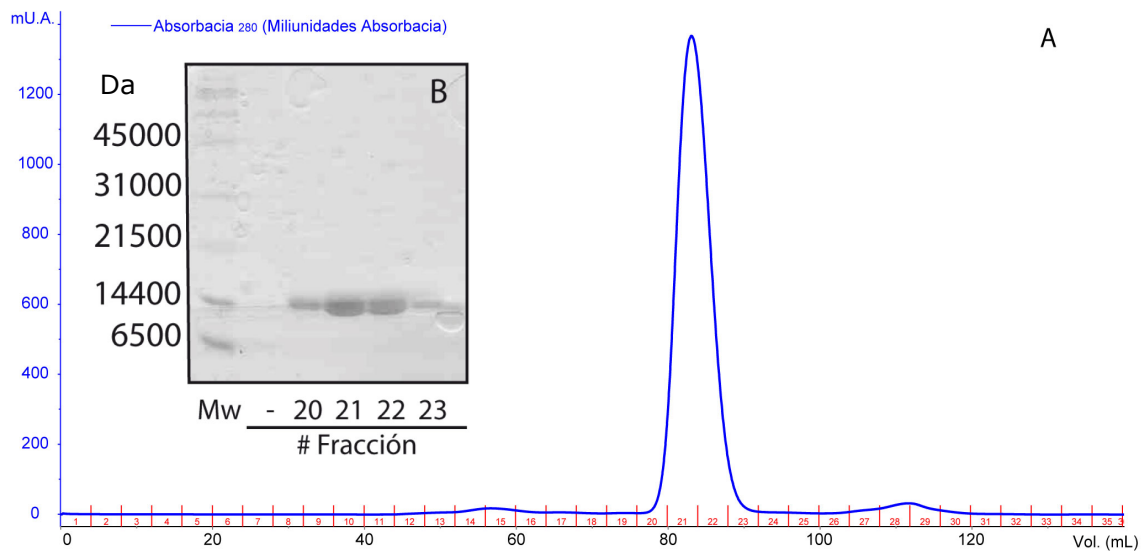


Figura 2.6. Purificación con el método de cromatografía de **filtración en gel**. A: cromatograma de la purificación de la **tioredoxina** de *E. coli*. En azul se muestra la Absorbancia₂₈₀ en función del volumen y en rojo se muestran las diferentes fracciones recogidas. B: muestra el gel de las fracciones correspondientes al pico principal. Mw se refiere a los marcadores de peso molecular. Las fracciones 20, 21 22 y 23 corresponden a las fracciones del pico principal en el cromatograma.

2.2.3.3. Preparación de la muestra de **tioredoxina**

Antes de poder utilizar la proteína, tenemos que disolverla en el tampón correspondiente, ya que la muestra resultante de la purificación tiene gran cantidad de sales que pueden interferir en los diferentes ensayos.

- **Cambio de la solución tamponante**

El siguiente paso consiste en la preparación del buffer donde irá disuelta la **tioredoxina** para realizar los diferentes experimentos. El procedimiento general es preparar el *buffer* a partir de la mezcla de dos disoluciones, una ácida y otra básica, de igual concentración, e ir midiendo con el pH-metro hasta obtener el pH deseado. Por ejemplo, para la preparación del tampón de ácido 2-4-(2-Hidroxietil)-1-Pireraciniil-Etanosulfónico (HEPES) 50mM a pH 7.0, elaboramos dos disoluciones: una de sal sódica de HEPES y otra de ácido HEPES. Pesamos los gramos necesarios, los disolvemos en agua bidestilada y ajustamos el volumen en sendos matraces aforados. A continuación, añadimos la disolución de sal sódica de HEPES a la de ácido HEPES hasta obtener el pH deseado y filtramos la disolución con un filtro de 0.45 μ m de tamaño de poro. Hemos llevado a cabo el cambio de *buffer* de la muestra con dos métodos: **cromatografía de filtración en gel** y **diálisis**.

Hemos usado la cromatografía de filtración en gel para cambiar el *buffer* de las muestras purificadas de las variantes de las bibliotecas combinatoriales. Hemos empleado este procedimiento en lugar de la diálisis porque para un primer muestreo no es necesario dializar exhaustivamente la muestra. Además, teniendo en cuenta que podemos llegar a purificar hasta 15 mutantes en un solo día, es un método que ahorra mucho tiempo. La cromatografía de filtración en gel es de tipo separación de grupos para la eliminación de sal. Para ello, hemos empleado las columnas desechables PD-10 (*GE Healthcare*). Estas columnas tienen un tamaño de poro que permite la separación de moléculas de elevado peso molecular (mayor de 5000 Da) de las sustancias de bajo peso molecular (menor de 1000 Da).

El procedimiento es el siguiente: en primer lugar, tenemos que equilibrar las columnas con 30 ml del *buffer* en el que queremos disolver la proteína. Tras lo cual, agregamos 2.5 ml de solución de proteína a la columna. La proteína purificada avanza más rápido y sale de la columna ya disuelta en el tampón, mientras que los componentes del tampón previo se quedarán retenidos en la columna.

Para llevar a cabo experimentos más rigurosos, como los de calorimetría diferencial de barrido, es necesaria una diálisis exhaustiva de la muestra. Para realizar la diálisis, hemos introducido un volumen de solución de proteína en una membrana de diálisis de 14.6 mm de diámetro de la marca *Spectra/Por* (*Spectrum labs*), con un tamaño de poro entre 6000 a 8000 Da. Este tamaño de poro permite el paso de moléculas de peso molecular menor al de tiorredoxina. Cerramos la membrana con unas pinzas (*Spectrum labs*), la introducimos en un recipiente que contiene un volumen 500 veces mayor al de la muestra y lo dejamos en agitación durante 8 h horas a 4°C. Transcurrido este tiempo, reemplazamos el *buffer* de la diálisis por *buffer* nuevo. Este procedimiento se repite una vez más, guardando el *buffer* del último cambio para usarlo como referencia.

Por último, con objeto de eliminar cualquier residuo insoluble de la muestra, centrifugamos dicha disolución en la microcentrífuga a 14000 r.p.m, 20 min a 4°C y la filtramos con filtros de 0.45 µm de tamaño de poro.

• Determinación de la concentración de tiorredoxina

La concentración de tiorredoxina en disolución la hemos determinado espectrofotométricamente registrando el espectro ultravioleta-visible. Para registrar el espectro de la tiorredoxina (Fig. 2.7.), tomamos una alícuota de la disolución de proteína y la diluimos de manera que su máximo de absorbancia sea menor a 1.0 U.A. De esta forma, nos aseguramos de estar en el rango lineal de la dependencia de la absorbancia con la concentración. Antes de medir el espectro de la proteína tenemos que hacer el blanco en el instrumento con el *buffer* de la diálisis.

El valor que tomamos para calcular la concentración es el máximo de absorbancia a 280 nm. Tomamos este valor porque todas las tiorredoxinas analizadas tienen al menos un triptófano y varias tirosinas. Estos aminoácidos aromáticos absorben luz a una longitud de onda de 280 y 274 nm respectivamente. El coeficiente de extinción teórico utilizado para la determinación de la concentración, ha sido calculado a partir de las secuencias de aminoácidos mediante el programa *ProtParam Tool* que se encuentra en la web: <http://www.expasy.org/>. Una vez determinados estos parámetros y aplicando la ley de Lambert-Beer, podremos obtener la concentración de tiorredoxina.

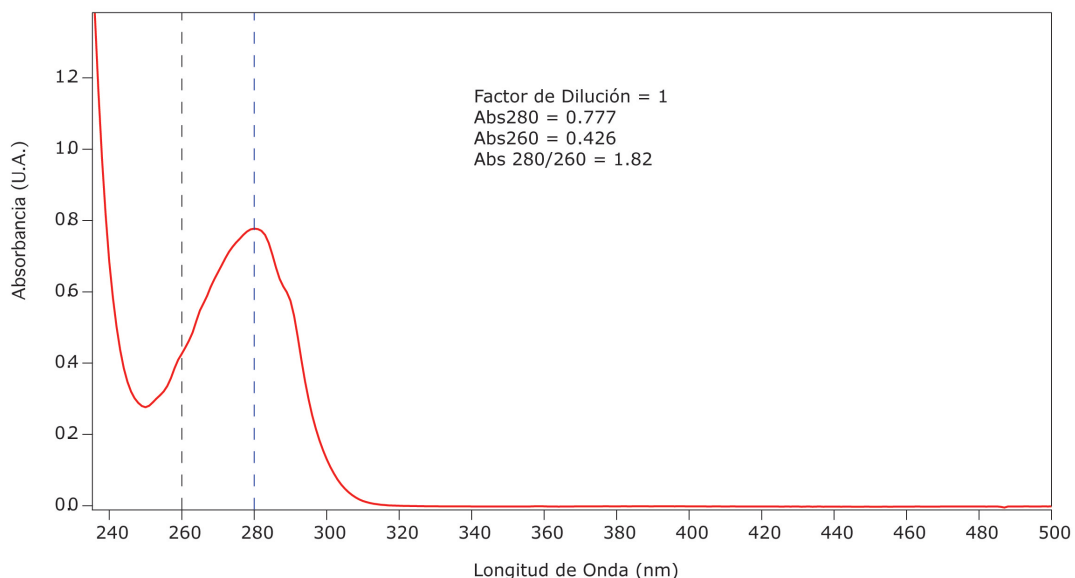


Figura 2.7. Espectro de absorción de la tiorredoxina de *E. coli*. Como guía visual, se muestra en línea discontinua, la prolongación de las longitudes de onda 280 y 260 nm. La muestra de proteína no parece presentar contaminación por ácidos nucleicos porque la relación 280/260 es mayor que 1.7.

Para que la medida sea fidedigna, debemos fijarnos en que no haya dispersión de luz en la zona del visible del espectro y en el cociente entre la absorción a 280 y 260 nm. El cociente $A_{280/260}$ se suele utilizar para evaluar la contaminación por ADN de la muestra de proteína. Si el valor es menor de 1.7, quiere decir que nuestra muestra de proteína posiblemente contenga ADN (máximo de absorción de luz a 260 nm).

Con el fin de eliminar la contaminación por ácidos nucleicos (si la hubiera), realizamos una diálisis exhaustiva de la muestra de durante 36 horas frente a 20mM de Glicina-HCl a pH 2 (3 cambios).

- **Procedimiento de concentración de tiorredoxina**

Para realizar los experimentos de cristalización es necesario partir de una muestra de proteína suficientemente concentrada (normalmente en el rango de los 10 mg/ml). Por ello, hemos concentrado la muestra de proteína por centrifugación usando tubos *Centricon* de un tamaño de poro de 1000 Da. La evolución de la concentración la hemos seguido a diferentes tiempos para asegurarnos de que la proteína no esta agregando y/o precipitando.

2.2.4. Construcción y caracterización de bibliotecas combinatoriales de variantes de tiorredoxinas.

En este trabajo hemos estudiado dos bibliotecas combinatoriales de variantes de tiorredoxinas. Esta aproximación se basa en obtener un conjunto de variantes de un gen que contenga todas las combinaciones posibles de una serie de mutaciones que queremos estudiar.

Las dos bibliotecas combinatoriales de variantes de tiorredoxina se han construido utilizando el método de *gene assembly mutagenesis* (Stemmer, Crameri et al. 1995; Bessette, Mena et al. 2003). Este método se basa en la capacidad que tiene la ADN polimerasa de reconstruir genes o plásmidos completos a partir de un conjunto de oligonucleótidos cortos y solapados (Stemmer, Crameri et al. 1995). Las diferentes variantes genéticas pueden ser generadas fácilmente por reemplazamiento de un oligonucleótido sin mutar por uno mutado (más su complementario y solapado). El

proceso se divide en tres etapas (Fig. 2.8). La primera etapa es una PCR de reconstrucción: en ella la ADN polimerasa liga aleatoriamente los oligonucleótidos solapados. Tras la etapa de reconstrucción, se amplifican los genes completos mediante PCR empleando los oligonucleótidos cebadores de los extremos del gen, produciendo así genes completos con diferentes secuencias. La última etapa es la digestión de los oligonucleótidos generados con las enzimas de restricción *NdeI* y *XhoI* y el clonaje de los genes en el plásmido pET24(+).

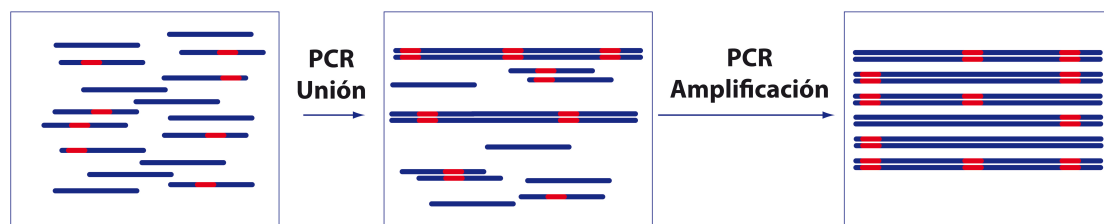


Figura 2.8. Representación esquemática del método de *gene assembly mutagenesis* (Bessette, Mena et al. 2003). En la primera etapa, se unen un conjunto de variantes génicas a partir de oligonucleótidos cortos y solapados, que codifican ambas cadenas del gen y contienen las bases mutadas en las posiciones deseadas (en rojo). La siguiente etapa consiste en la amplificación de las diferentes variantes génicas empleando los oligonucleótidos cebadores de los extremos.

Este método tiene una limitación técnica ya que puede introducir mutaciones no deseadas en el gen o que éste no se clone en el vector, lo cual baja la eficiencia de la técnica. Debido a esta razón, tenemos que realizar un control de calidad secuenciando varios clones elegidos aleatoriamente como paso previo a la caracterización de las variantes. La compañía *TOP Gene* ha sido la encargada de la construcción de las dos bibliotecas combinatoriales de variantes de tiorredoxinas y de realizar los controles de calidad de las mismas.

El número de variantes de la biblioteca que vamos a caracterizar experimentalmente lo hemos fijado en torno al 10% del tamaño de las bibliotecas. Dado que hemos mutado 8 posiciones en las secuencias de tiorredoxinas, las dos bibliotecas contendrán 256 variantes diferentes con todas las combinaciones posibles de mutación ($2^8=256$). La elección de estas posiciones se detalla en el capítulo 6 de la presente memoria.

El primer paso de la caracterización de mutantes es transformar las bacterias supercompetentes *BL21(DE3)* con ADN de la biblioteca. Para ello seguimos el protocolo indicado en el apartado 2.2.2.3.

Como mencionamos en el apartado anterior, esta técnica no tiene una efectividad del 100%. Por ello, es necesario realizar una prueba de expresión (apartado 2.2.2.4) previa a la purificación de las variantes. El objetivo es detectar qué colonias no expresan tiorredoxina y poder desecharlas. El número de colonias que vamos a utilizar para el estudio dependerá de los resultados del control de calidad y del tamaño de muestra que queramos estudiar. A modo de ejemplo, si nos hemos planteado estudiar un 10% de variantes diferentes y el resultado del control de calidad es del 80% de variantes que expresan proteína, tomaremos al menos 45 colonias de la placa.

Una vez que hemos averiguado las colonias que expresan tiorredoxina, procedemos a la expresión y la purificación de las variantes de tiorredoxina (apartado 2.2.3.1 de este capítulo). El siguiente paso es el cambio de *buffer* de la muestra a HEPES 50 mM pH 7.0, que se lleva a cabo mediante columnas PD-10 (apartado 2.2.3.3). Para terminar, centrifugamos a 14000 r.p.m a 4°C las muestras y medimos la concentración de proteína.

En este momento, solo nos falta averiguar qué mutaciones tienen las diferentes variantes caracterizadas (y si alguna secuencia se repite). Para ello hemos realizado una *miniprep* a partir de las colonias congeladas en glicerol y secuenciado el plásmido de cada variante siguiendo el protocolo descrito en el apartado 2.2.2.2 del presente capítulo.

2.2.5. Ensayo de actividad oxido-reductasa de tiorredoxinas por precipitación de insulina.

El presente ensayo fue desarrollado por Arne Holmgren (Holmgren 1979) y se lleva a cabo con ligeras modificaciones. En este experimento, vamos a medir espectrofotométricamente la turbidez de una disolución de insulina en función del tiempo. La insulina es una proteína compuesta de dos cadenas polipeptídicas (A y B) unidas entre sí por dos puentes disulfuro. Durante el tiempo que dura reacción, la tiorredoxina acelera la reducción de los puentes disulfuro presentes en la insulina por parte del ditiotreitól de acuerdo con la siguiente reacción:

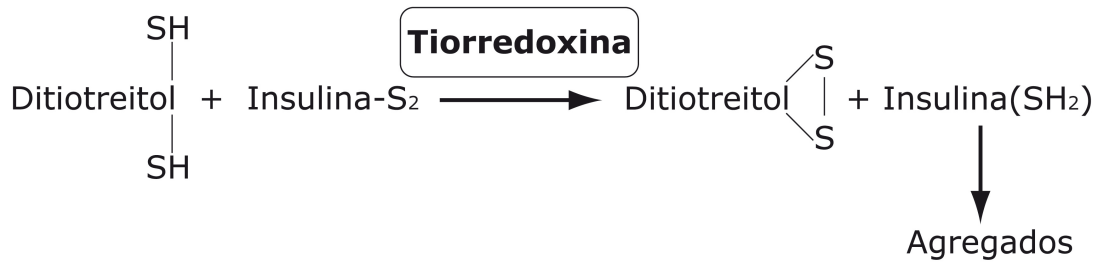


Figura 2.9. Esquema de la reacción de reducción de insulina por DTT. La reacción está catalizada por tiorredoxina. La agregación de la insulina provoca un aumento de la absorbancia medida a 650 nm.

Tras la reducción, los puentes disulfuro se vuelven a formar de manera desordenada entre las cadenas ocasionando la agregación. Cuando estos agregados alcanzan un tamaño determinado, dispersan la luz, lo que resulta en un aumento de la absorbancia a 650 nm. La actividad enzimática de la tiorredoxina está relacionada con la velocidad a la que se forman estos agregados.

La preparación del experimento se lleva a cabo en una cubeta de plástico de 1 cm de paso óptico, en la cual introducimos 784 μl de *buffer* 0.1M fosfato potásico, 2mM EDTA, pH 6.5. El pH debemos medirlo cuidadosamente, ya que los valores de actividad dependen fuertemente del mismo. A continuación, añadimos 166 μl de tiorredoxina (a una concentración determinada) disuelta en *buffer* 50 mM HEPES a pH 7, al cual añadimos 50 μl de 10 mg/ml de insulina disuelta en *buffer* 50 mM Tris-HCl pH 8.3. Una vez tengamos todos los componentes bien mezclados en la cubeta, la introducimos en el espectrofotómetro y la dejamos 10 minutos a 37°C para que su temperatura se equilibre. Hemos usado un baño termostatzado que recircula agua alrededor de las celdas para controlar la temperatura del experimento en todo momento. Para iniciar la reacción, añadimos a la cubeta 10 μl de 0.1 M ditiotreitól (DTT), mezclamos rápidamente con una pipeta *pasteur* e iniciamos la medida de la absorbancia a una longitud de onda de 650 nm durante unos 40 minutos.

El espectrofotómetro registra una curva que representa la absorbancia frente al tiempo. Se observa que hay una etapa inicial, donde no se observa agregación (*lag time*); un tramo donde hay una dependencia aproximadamente lineal de la absorbancia con el tiempo y una tercera etapa estacionaria, donde no hay variación (Fig. 2.10.A). A partir de la curva, calculamos la primera derivada, de la cual obtenemos el valor máximo de velocidad en términos de variación de absorbancia

en función del tiempo (Fig. 2.10.B). Cuando este valor es superior a 0.2-0.25 U.A. \cdot min⁻¹ se pierde la linealidad con la concentración de tiorredoxina; por esta razón, debemos trabajar con una concentración final de tiorredoxina en el rango de 1 a 3 μ M. La actividad enzimática (U.A. \cdot min⁻¹ \cdot μ mol⁻¹) se calcula a partir de la dependencia lineal de la velocidad máxima (Abs/t) frente a la concentración de tiorredoxina.

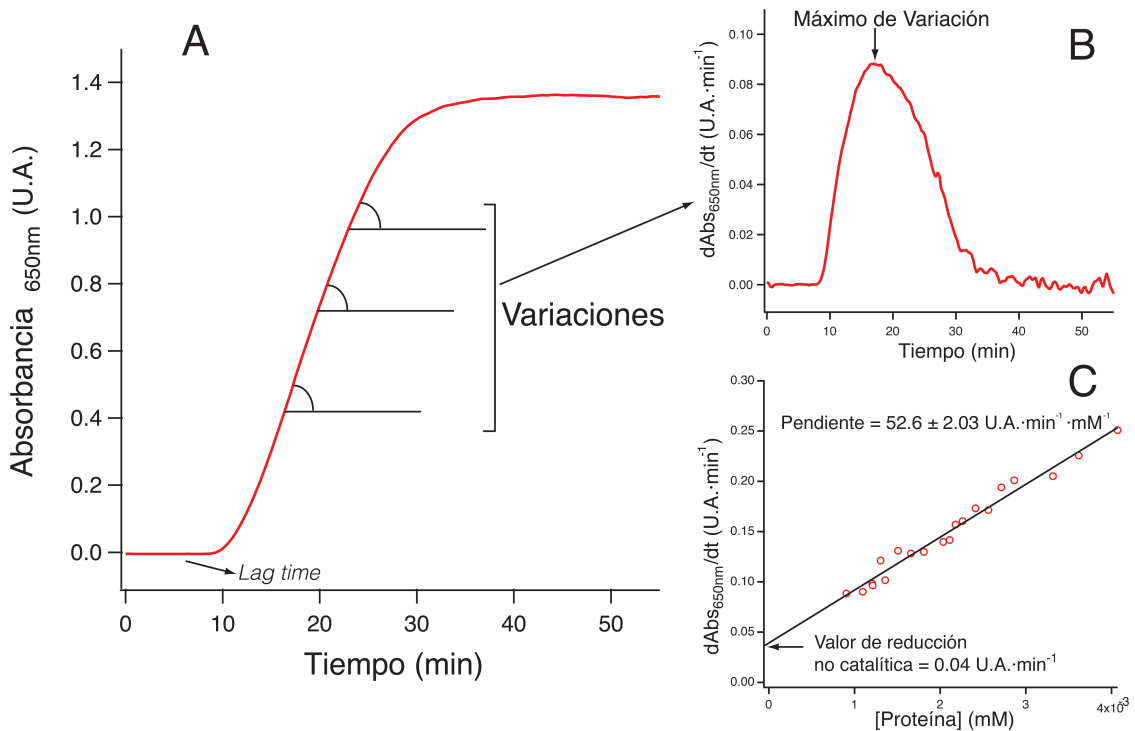


Figura 2.10. Diferentes etapas de la determinación de la actividad oxido-reductasa de la tiorredoxina mediante el ensayo de agregación de insulina. A: datos registrados de absorbancia a 650 nm en función del tiempo a una concentración de tiorredoxina de 1.48 μ M. B: Cálculo de la derivada de la curva registrada. Se indica el máximo de variación (dAbs/dt), relacionado con la velocidad de la reacción. C: representación del máximo de variación Abs/t en función de la concentración de tiorredoxina (mM). Se ha realizado un ajuste lineal, a partir del cual es posible determinar el valor de la actividad oxido-reductasa de tiorredoxina expresado en U.A. \cdot min⁻¹ \cdot μ mol⁻¹.

El componente no catalítico de la reacción se obtiene a partir del ajuste anterior por extrapolación a concentración cero de tiorredoxina (Fig. 2.10.C). Lo realizamos de esta manera porque el tiempo para la reducción no catalítica (sin trx) es demasiado largo y acaban apareciendo artefactos que distorsionan los resultados. Como control, paralelamente a cada nuevo ensayo de actividad, hemos realizado varias determinaciones de la actividad de tiorredoxina de *E. coli* a diferentes concentraciones de proteína y hemos cotejado los resultados obtenidos con los de un juego de datos de referencia para esta tiorredoxina.

2.2.6. Calorimetría diferencial de barrido.

La Calorimetría Diferencial de Barrido (*Differential Scanning Calorimetry* o DSC, por sus siglas en inglés) es una técnica que permite la caracterización energética de los cambios conformacionales inducidos por temperatura en polímeros biológicos y agregados supramoleculares. Existen diversas revisiones sobre los aspectos más interesantes de esta técnica y su aplicación a la ingeniería de proteínas (Privalov 1979; Privalov 1980; Privalov 1982; Sanchez-Ruiz y Mateo 1987; Chowdhry y Cole 1989; Privalov 1989; Freire, van Osdol et al. 1990; Cooper y Johnson 1994; Sanchez-Ruiz 1995; Ibarra-Molero y Sanchez-Ruiz 2006; Ibarra-Molero y Sanchez-Ruiz 2008; Sanchez-Ruiz 2010; Sanchez-Ruiz 2011)

Mediante un experimento de calorimetría diferencial de barrido es posible obtener la capacidad calorífica molar aparente de una proteína en función de la temperatura (termograma). El posterior análisis de esta dependencia hace posible una caracterización termodinámica completa de la transición lograda por efecto de la temperatura. Un termograma típico de DSC tiene forma de pico, correspondiente a un proceso endotérmico de acuerdo con el Segundo Principio de la Termodinámica (supuesto el proceso de equilibrio). Los experimentos de DSC, analizados según la termodinámica de equilibrio, nos permiten conocer la capacidad calorífica parcial absoluta de una molécula y la población de estados intermedios y los parámetros termodinámicos tales como el cambio de entalpía (ΔH), cambio de entropía (ΔS) y cambio de la capacidad calorífica (ΔC_p) asociados a la transición inducida por la temperatura.

2.2.6.1. Descripción del Microcalorímetro.

Un microcalorímetro básico consta de dos células, una de referencia y otra de muestra; una camisa adiabática que envuelve a las células; un dispositivo *peltier*, que calienta o enfría la camisa adiabática; una serie de precisos sensores de temperatura y un sistema para producir sobrepresión. El funcionamiento del aparato consiste en calentar las dos células simultáneamente a una velocidad de calentamiento constante en un ambiente adiabático. De tal manera que la señal registrada por el instrumento es proporcional a la diferencia de capacidad calorífica entre las dos células. Hemos empleado dos modelos distintos de calorímetros: el VP-DSC (MicroCal) y el VP-DSC capilar (Microcal).

El modelo VP-DSC (MicroCal) tiene células construidas en tántalo con forma de moneda cuyo volumen efectivo es de 0.5 ml aproximadamente. Se accede a ellas a través de un fino capilar de 1.5 mm de diámetro. El calorímetro opera entre -10°C y 120°C , a velocidades de barrido que van de 0 a $90^{\circ}\text{C}/\text{hora}$ y con sobrepresión para evitar la ebullición del líquido.

El otro modelo, es un calorímetro VP-DSC capilar (MicroCal) (Fig. 2.11) que está basado en el modelo anterior. La diferencia fundamental la encontramos en la estructura de las células, en este caso son capilares en espiral de aproximadamente 0.130 ml de volumen interno útil. Esta estructura exige a que cada una de las células tenga dos accesos, de manera que las células se pueden limpiar y cargar simplemente por desplazamiento del líquido. En este modelo, el rango de temperaturas va de -10°C a 130°C (140°C si aumentamos la sobrepresión) y puede operar a velocidades de barrido de 0 a $250^{\circ}\text{C}/\text{hora}$.

La principal ventaja del VP-DSC capilar es tiene implementado un sistema de procesamiento de alto rendimiento. Las muestras de proteína y el *buffer* de referencia se colocan en placas de 96 pocillos (MicroCal) y las almacenamos dentro de un depósito termoregulado a 6°C . Un brazo robótico se encarga de tomar las muestras correspondientes e introducir las dentro de las células a través de un sistema de válvulas. Conectado a este sistema de válvulas, se encuentra una entrada de nitrógeno gas que suministra la sobrepresión.

Ambos modelos están conectado a un ordenador donde, gracias al programa *VPviewer*, solo tenemos que programar los barridos que vamos a realizar (temperatura inicial y final, velocidad de barrido, etc.). El programa del modelo VP-DSC capilar nos permite programar algunas opciones más, como número de lavados entre barridos, número total de barridos o las muestras a las que aplicaremos un segundo barrido.

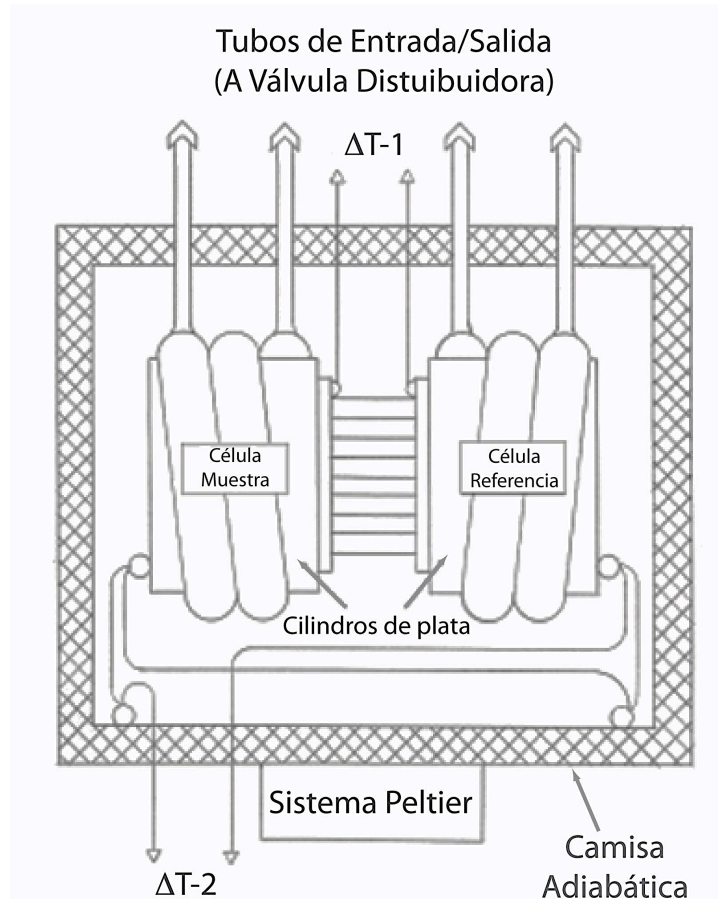


Figura 2.11: Diagrama esquemático del núcleo térmico del microcalorímetro VP-DSC capilar. Se indican las células en espiral de muestra y de referencia, cada una con los tubos capilares de entrada y salida; la camisa adiabática; el sistema *peltier*, que calienta o enfría la camisa; el sensor de temperatura $\Delta T-1$, entre las células de muestra y referencia; y el sensor de temperatura $\Delta T-2$, entre las células y la camisa. Imagen adaptada de www.microcal.com.

2.2.6.2. Descripción del experimento de DSC.

Para realizar el experimento, necesitamos muestra de proteína, cuya concentración ha sido determinada minuciosamente, y el *buffer* de referencia, que debe ser idéntico a aquél en que se encuentra disuelta la proteína. Por esta razón, es necesario realizar una diálisis exhaustiva de nuestras muestras y usar *buffer* del último cambio de la diálisis en la célula de referencia.

Una vez cargadas las muestras y programada la rutina de trabajo, debemos estabilizar el instrumento haciendo de 4 a 6 barridos llenando con el *buffer* de referencia ambas células hasta conseguir reproducibilidad en la señal. Esta línea base *buffer-buffer* se le restará posteriormente al perfil proteína-*buffer* con objeto de eliminar la contribución a la señal debida a aspectos instrumentales. Una vez que observamos que la historia térmica del instrumento se ha establecido, es decir,

que la línea base *buffer-buffer* es reproducible ($\pm 2 \mu\text{cal/grado}$), será posible cambiar la disolución de la célula de muestra por nuestra disolución de proteína (esto se realiza de forma automática en el VP-DSC capilar).

En el primer barrido de la proteína, el microcalorímetro registra un termograma que representa la capacidad calorífica molar aparente de la proteína en función de la temperatura. Un termograma típico de DSC muestra un pico o transición que se atribuye a la absorción de calor asociada a su desnaturalización, mientras que los valores de la señal en la pre-transición y post-transición reflejan las capacidades caloríficas parciales de los estados nativo y desnaturalizado de la proteína respectivamente (Fig. 2.12).

Antes de terminar el experimento, dejamos que el aparato enfríe las muestras y realice otro barrido sobre las mismas. A este segundo barrido se denomina *reheating* y nos ofrece información sobre si el proceso de desnaturalización es caloriméricamente reversible. Generalmente, la reversibilidad de una muestra se expresa en términos del porcentaje de área bajo la curva del segundo barrido de la misma. Éste es un aspecto muy importante ya que no se puede aplicar un análisis basado en la termodinámica de equilibrio si la proteína presenta una transición irreversible.

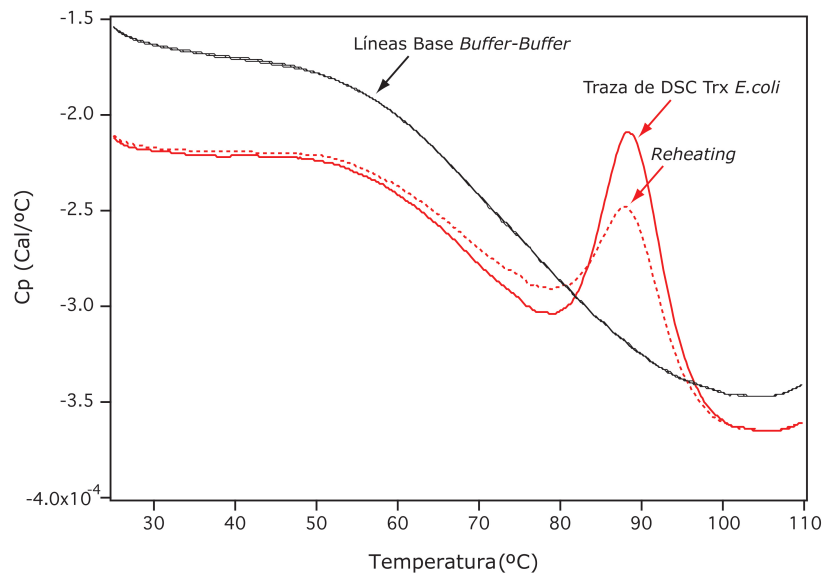


Figura 2.12. Termogramas registrados por el microcalorímetro en un experimento típico de DSC. En línea negra se muestran tres barridos *buffer-buffer* (línea base) anterior al realizado para la proteína. En línea continua roja se muestra la traza de DSC para la tiorredoxina de *E. coli* a una concentración de 0.32 mg/ml . En línea discontinua roja se muestra el *reheating* de la misma muestra de trx.

2.2.6.3. Análisis basado en la termodinámica del equilibrio: modelo de dos estados

Antes de proceder al análisis del termograma de proteína, es necesario restar la línea base *buffer-buffer* anterior al barrido de la proteína para eliminar las asimetrías inherentes a las células del aparato. Posteriormente, normalizamos los datos por la velocidad de barrido y por la concentración de proteína. Hemos realizado este cálculo utilizando el programa Origin 7 (*OriginLab*) con un complemento para análisis de DSC (*MicroCal*).

Para poder realizar el análisis de los datos según la termodinámica del equilibrio, debemos suponer que se establece el equilibrio químico entre los estado más poblados (estado nativo y desnaturalizado) y además, que el perfil de DSC refleje de forma precisa un proceso de desnaturalización en el equilibrio. Generalmente, se acepta que un proceso transcurre en equilibrio cuando existe reversibilidad calorimétrica (Sanchez-Ruiz 1995; Plaza del Pino, Ibarra-Molero et al. 2000).

Una vez comprobado que un sistema es caloriméricamente reversible, intentamos ajustar el modelo termodinámico más sencillo posible a los termogramas obtenidos.

El modelo más simple de desnaturalización es aquel en que solo están significativamente poblados el estado de la proteína nativo (N) y el desnaturalizado (D) y sus cantidades relativas a una temperatura dada están determinadas por el valor de la constante de equilibrio (K) a esa temperatura. Esta relación podemos expresarla mediante las siguientes ecuaciones:



$$K = \frac{[U]}{[N]} \quad (2.2)$$

En general, los cambios en un parámetro termodinámico producidos por la desnaturalización, $\Delta_N^D J$, se definen como las diferencias entre los valores de dicho parámetro en el estado desnaturalizado y el nativo:

$$\Delta_N^D J = J(D) - J(N) \quad (2.3)$$

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 estado de referencia el estado nativo. Así, la entalpía de exceso y la capacidad calorífica de exceso vienen dadas por:

$$\langle \Delta H \rangle = x_D \cdot \Delta_N^D H = \frac{K}{(1+K)} \cdot \Delta_N^D H \quad (2.4)$$

$$C_p^{ex} = \frac{\partial \langle \Delta H \rangle}{\partial T} = \frac{(\Delta_N^D H)^2}{RT^2} \cdot \frac{K}{(1+K)^2} + x_D \cdot \Delta_N^D C_p \quad (2.5)$$

donde x_D es la fracción de proteína que se encuentra en estado desplegado y $\Delta_N^D C_p = C_p^D - C_p^N$ donde C_p^D es la capacidad calorífica del estado desplegado y C_p^N la del estado nativo, y 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:

$$\frac{\partial \ln K}{\partial T} = \frac{\Delta_N^D H}{RT^2} \quad (2.6)$$

$$\frac{\partial \Delta_N^D H}{\partial T} = \Delta_N^D C_p \quad (2.7)$$

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 2.5, 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 (tiene forma de pico), mientras que el segundo término representa la capacidad calorífica promedio de la proteína y corresponde a la línea base química (forma sigmoidal). El área encerrada entre la transición y la línea base química es igual al cambio total de entalpía en el proceso de desnaturalización y se conoce como *entalpía calorimétrica*, $\Delta_N^D H$ (Fig. 2.13). Esta depende de la temperatura según la ecuación de Kirchoff y del cambio de la capacidad calorífica en el proceso de desnaturalización ($\Delta_N^D C_p$). Si la transición es simétrica y estrecha, no se comete mucho error si se desprecia esta dependencia

dentro del rango estrecho de la transición y se asigna la entalpía calorimétrica a la temperatura de desnaturalización: $\Delta_N^D H(T_m)$.

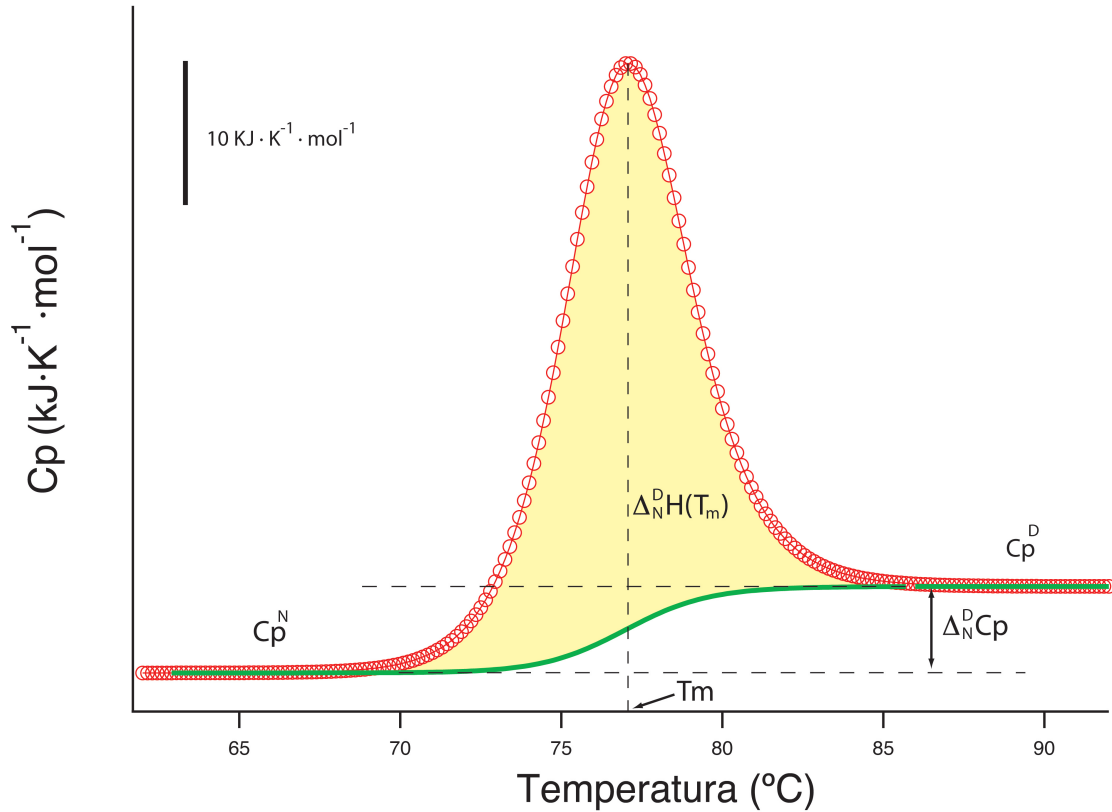


Figura 2.13. Curva de capacidad calorífica en función de la temperatura calculada para una proteína hipotética que se despliega según el modelo de dos estados (parámetros utilizados para la simulación son $T_m = 350\text{K}$, $\Delta H = 400\text{kJ/mol}$ y $\Delta C_p = 6\text{kJ} \cdot \text{K}^{-1} \cdot \text{mol}^{-1}$). En rojo se muestra el termograma de la proteína. El máximo corresponde a la T_m de la proteína. C_p^N representa la capacidad calorífica del estado nativo y C_p^D la del estado desplegado. $\Delta_N^D C_p$ es el cambio de la capacidad calorífica en el proceso de desnaturalización. En línea verde se representa la línea base química (línea que conecta de un modo suave los valores de capacidad calorífica pre- y posttransicionales y que representa la capacidad calorífica media de la proteína). En amarillo se representa el área encerrada entre las dos curvas equivale a la entalpía calorimétrica $\Delta_N^D H(T_m)$.

Como a $T=T_m$ el cambio de energía de Gibbs es cero, el cambio de entropía a la T_m [$\Delta_N^D S(T_m)$] se puede calcular de esta manera:

$$\Delta_N^D S(T_m) = \frac{\Delta_N^D H(T_m)}{T_m} \quad (2.8)$$

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 usando ecuaciones termodinámicas habituales:

$$\Delta_N^D H(T) = \Delta_N^D H(T_m) + \int_{T_m}^T \Delta_N^D C_p \cdot dT \quad (2.9)$$

$$\Delta_N^D S(T) = \frac{\Delta_N^D H(T_m)}{T_m} + \int_{T_m}^T \frac{\Delta_N^D C_p}{T} \cdot dT \quad (2.10)$$

$$\Delta_N^D G(T) = \Delta_N^D H(T_m) - T \cdot \Delta_N^D S(T) \quad (2.11)$$

Si suponemos que el cambio de capacidad calorífica permanece constante con la temperatura, las ecuaciones 2.9, 2.10, 2.11 pasarían a ser:

$$\Delta_N^D H(T) = \Delta_N^D H(T_m) + \Delta_N^D C_p \cdot (T - T_m) \quad (2.12)$$

$$\Delta_N^D S(T) = \frac{\Delta_N^D H(T_m)}{T_m} + \Delta_N^D C_p \cdot \ln\left(\frac{T}{T_m}\right) \quad (2.13)$$

$$\Delta_N^D G(T) = \Delta_N^D H(T_m) \cdot \left[1 - \frac{T}{T_m}\right] + \Delta_N^D C_p \cdot \left[T - T_m - T \cdot \ln\left(\frac{T}{T_m}\right)\right] \quad (2.14)$$

La dependencia de ΔG con la temperatura se conoce como **curva de estabilidad** y un ejemplo está representado en la figura 2.14.A. Las propiedades fundamentales de la curva de estabilidad han sido descritas por Becketl y Shellman (Becketl y Schellman 1987; Schellman 1987). A continuación comentamos las más significativas:

- La curvatura viene dada por, $\partial^2 \Delta G / \partial T^2 = -\Delta C_p / T$ y siempre adquiere valores negativos ya que el cambio de capacidad calorífica de desnaturalización es siempre positivo.

- La pendiente de la curva es igual a $-\partial \Delta_N^D G / \partial T = \Delta_N^D S$ y muestra un extremo, concretamente un máximo, a la temperatura T_S , a la que $\Delta S = 0$ (Fig. 2.14.B). La temperatura T_H , a la que la entalpía de desnaturalización es cero, es ligeramente más baja que T_S . La proteína nativa es estable en el rango de temperatura en que $\Delta_N^D G > 0$. Según podemos ver en la figura 2.14.B, la curva de estabilidad cruza la

línea de $\Delta_N^D G = 0$ en dos puntos diferentes, siendo la temperatura más alta la correspondiente a la temperatura de desnaturalización (T_m) usada en el cálculo de la curva de estabilidad. El punto de corte correspondiente a la temperatura menor es el resultado de la extrapolación de la curva e indica que es posible desnaturalizar la proteína también por frío (desnaturalización fría), idea predicha por Brandts (Brandts 1964) y demostrada con posterioridad en estudios experimentales que sugieren que la desnaturalización fría es una propiedad común de las proteínas (Privalov y Potekhin 1986).

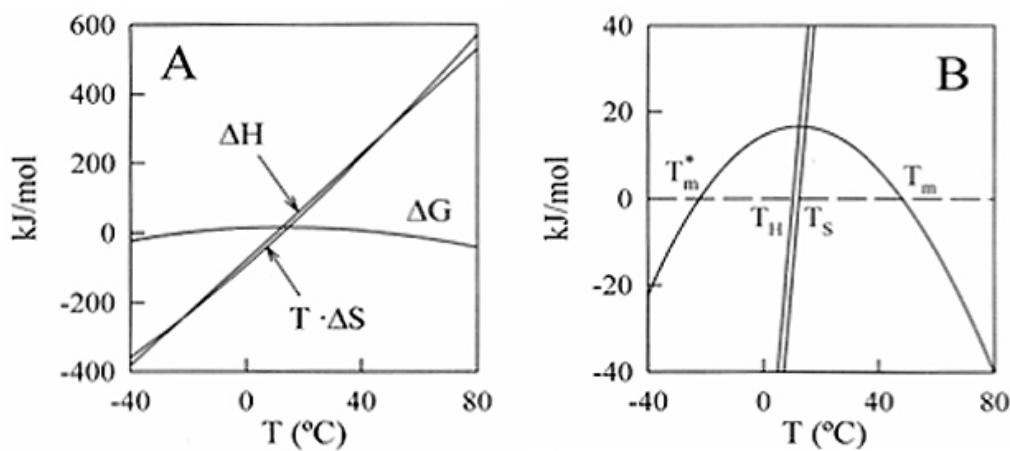


Figura 2.14. A: Efecto de la temperatura sobre los cambios de entalpía, entropía y energía libre de Gibbs para la desnaturalización de una proteína hipotética. B: ampliación que muestra las características principales de la curva de estabilidad de proteínas (la representación de $\Delta_N^D G$ frente a T). Se han indicado las temperaturas de 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 ΔH y ΔS se hacen cero, respectivamente. Los valores de $\Delta_N^D H$, $\Delta_N^D S$ y $\Delta_N^D G$ se calcularon usando las ecuaciones 2.12, 2.13 y 2.14, y valores de $T_m=48.3^\circ\text{C}$, $\Delta_N^D H(T_m)=290 \text{ kJ}\cdot\text{mol}^{-1}$ y $\Delta_N^D C_p=7.6 \text{ KJ}\cdot\text{K}^{-1}\cdot\text{mol}^{-1}$. Figura tomada de (Sanchez-Ruiz 1995).

Así pues, si tenemos los valores de T_m , $\Delta_N^D H(T_m)$ y $\Delta_N^D C_p$ determinados a partir del termograma de DSC, disponemos de una descripción termodinámica completa del proceso de desnaturalización de la proteína en estudio. Este formalismo es sólo aplicable en el caso de que el modelo de equilibrio de dos estados describa adecuadamente el proceso de desnaturalización en estudio. Es posible determinar la aplicabilidad de dicho modelo a partir de un análisis directo de la transición de DSC. Si realizamos simulaciones dando diferentes valores al cambio de entalpía de desnaturalización en la ecuación 2.4, veremos que esta determina el área bajo la transición así como su anchura. La entalpía de desnaturalización calculada a partir de la forma (anchura) de la transición se conoce como entalpía aparente o de van't

Hoff (ΔH^{vH}). Se puede derivar una expresión para esta entalpía a partir del término debido al desplazamiento del equilibrio de desnaturalización por efecto de la temperatura en la ecuación 2.5 y considerando $T=T_m$ (y por tanto $K=1$):

$$\Delta H^{vH} = 4RT_m^2 \cdot \left[\frac{\Delta C_{p,m}}{\Delta_N^D H(T_m)} \right] \quad (2.15)$$

donde $\Delta C_{p,m}$ es la capacidad calorífica de exceso a la temperatura T_m medida sobre la línea base química. El cociente $\Delta C_{p,m}/\Delta_N^D H(T_m)$ no depende de la cantidad de proteína presente en la célula del calorímetro, pero sí es sensible a la forma (anchura) de la transición.

Obviamente, no puede haber dos cambios diferentes de entalpía para el mismo proceso. La ecuación 2.15 se obtiene suponiendo un modelo de dos estados, de manera que si realmente el proceso de desnaturalización en estudio transcurre según este modelo, encontraremos que las entalpías de van't Hoff y la calorimétrica son iguales dentro del error experimental. Si existen diferencias significativas entre estos dos valores, se concluye que el proceso de desnaturalización no sigue un modelo de dos estados.

Los resultados de validación del modelo de dos estados se suelen dar como la relación entre las entalpías calorimétrica y de van't Hoff: $r = \Delta_N^D H(T_m)/\Delta H^{vH}$. Así, si estados intermedios llegan a poblarse significativamente durante el proceso, entonces la transición de DSC será más ancha que la predicha por el modelo de dos estados y por consiguiente, la entalpía de van't Hoff será menor que la calorimétrica. Otro caso que puede conducir a un parámetro r aparentemente distinto de la unidad, es la presencia de desnaturalización fría en la traza calorimétrica. En este caso, la entalpía de van't Hoff y la calorimétrica podrían estar mal determinadas con las ecuaciones habituales. Privalov y colaboradores han encontrado una relación de $r = 1.05 \pm 0.03$ para diversas proteínas globulares pequeñas (Privalov 1979), indicando una baja proporción de estados intermedios y por tanto, una buena concordancia con el modelo de dos estados.

2.2.7. Determinación estructural de tiorredoxinas por cristalografía de rayos-X

Para determinar la estructura de las proteínas hay varias aproximaciones: la Resonancia Magnética Nuclear (RMN), la difracción de rayos-X y/o neutrones sobre cristales de proteína o la microscopía electrónica. Sin embargo sólo la RMN y la cristalografía generan un modelo atómico detallado y fidedigno necesario para analizar y comparar la estructura tridimensional de las proteínas

La RMN nos permite determinar las posiciones de cada uno de los átomos que constituyen la molécula de proteína cuando sometemos la muestra a un campo magnético. Una de las ventajas de esta técnica es que la proteína se encuentra en disolución acuosa. Estas condiciones representan mejor el entorno de las proteínas en el interior de las células. El inconveniente es que, por el momento, solo podemos utilizar proteínas de un tamaño máximo de 60 a 85 KDa. Además, el resultado no es una estructura única, sino que se obtienen estructuras moleculares que se encuentran en un equilibrio conformacional.

La cristalografía de rayos-X es una técnica experimental basada en que los cristales difractan los rayos-X debida a la interacción entre esta radiación y la nube electrónica de los átomos. Los rayos-X tienen la longitud de onda (Ångströms, $\sim 10^{-10}$ m) del orden de la distancia interatómica de las moléculas. A partir del patrón de difracción obtenido de esta dispersión, se puede reconstruir la densidad electrónica y obtener un modelo único de la estructura tridimensional de la proteína. En contraste con la RMN, no hay limitación de tamaño para la molécula o complejo que puede ser estudiado. La principal desventaja es que necesitamos obtener un cristal, que actúa como elemento amplificador de la señal, de cuyo tamaño y calidad cristalina dependerá la calidad final de los datos de difracción y, por tanto, la precisión del modelo obtenido.

Estas dos técnicas pueden considerarse complementarias y podemos elegir una u otra según la solubilidad de la proteína, la facilidad o dificultad de cristalizarla y su tamaño. La técnica que hemos elegido para determinar las estructuras de siete tiorredoxinas resucitadas del periodo Precámbrico ha sido la cristalografía de rayos-X. Estas tiorredoxinas están relacionadas con distintos nodos del árbol filogenético que son interesantes desde el punto de vista evolutivo (véase el capítulo 3 de la presente memoria). En este apartado se explicarán los procedimientos que hemos

seguido para cristalizar las distintas tiorredoxinas resucitadas y para la determinación de los distintos modelos estructurales.

2.2.7.1. Cristalización de tiorredoxinas ancestrales

Para llevar a cabo un experimento de difracción es necesario conseguir cristales de proteína de suficiente calidad y tamaño. Un cristal es un sólido ordenado en el que se repite una unidad organizativa básica denominada **Unidad Asimétrica**. Los cristales de proteínas son similares a los cristales de cualquier otro tipo de material orgánico o inorgánico con la salvedad del tamaño y la consecuente distribución de cargas. Aunque el tipo y fuerzas de las interacciones que unen estas moléculas son similares, la heterogeneidad de las proteínas hacen que el número de interacciones sea mucho menor, dejando grandes huecos que se rellenan con agua no ordenada y de ahí su mayor inestabilidad. Por otra parte, las moléculas no son exactamente iguales debido a que pueden presentar diferentes conformaciones de las cadenas laterales (confórmeros), que son distinguibles a resoluciones del orden de los 2.5 Å o inferiores. Todos estos elementos se oponen al principio de uniformidad y ello dificulta la obtención de cristales.

La cristalización de proteínas es un ensayo de prueba y error en el que se busca la precipitación controlada de la macromolécula de forma ordenada, en estado cristalino. Hay muchos factores que controlan e influyen en la solubilidad de la proteínas pero, como regla general, es la disposición de muestras de elevada pureza. A continuación se detallan los cuatro pasos esenciales a tener en cuenta a la hora de cristalizar *de novo* una proteína (Drenth 2007):

1. Para determinar la **pureza** de la proteína sometemos a la muestra final (dializada y centrifugada) a un SDS-PAGE. El grado de pureza debe ser al menos del 95%. Normalmente, cargamos dos muestras de cada proteína. Para ver la presencia de pequeñas impurezas, la primera muestra se carga a una concentración alta, de manera que la banda observada sea muy intensa. La segunda muestra se carga a baja concentración para determinar el tamaño y comprobar si la muestra es homogénea. El resultado será una banda muy perfilada correspondiente a nuestra proteína.

Además de la pureza, una solución de proteína debe reunir otras características (McPherson 1999): La **concentración** de proteína ha de ser suficientemente alta

para permitir el crecimiento de cristales. Otra característica importante a tener en cuenta es la **homogeneidad** de la muestra. Normalmente, la heterogeneidad suele aparecer por la manipulación experimental (grados distintos de oxidación, desnaturalización, proteolisis, etc.).

2. La proteína ha de estar disuelta en un solvente salino u orgánico adecuado, que puede incluir sales u otros aditivos, donde su solubilidad sea óptima.

3. Reducir la solubilidad de proteína para alcanzar el nivel de **sobresaturación** adecuada para inducir la nucleación. En esta etapa, se forman pequeños agregados que pueden actuar como núcleos para el crecimiento cristalino. La formación espontánea del núcleo cristalino requiere el suministro de energía (un valor de sobresaturación suficientemente alto) (Weber 1991). Este fenómeno es conocido como nucleación homogénea si no hay partículas extrañas en el medio. La presencia de impurezas en el medio puede hacer que se formen núcleos a valores de sobresaturación inferiores (nucleación heterogénea) (Feher y Kam 1985).

4. Una vez que se ha formado el núcleo (hemos pasado la barrera energética), el crecimiento puede proseguir a niveles de sobresaturación medios o bajos hasta alcanzar el equilibrio (Weber 1991).

- **Métodos de cristalización de proteínas.**

En resumen, para cristalizar una proteína es necesario reducir su solubilidad de la proteína hasta que alcance el estado de sobresaturación. Existen varias maneras de llevar una solución de proteínas al estado sobresaturado (Messerschmidt 2007), la más sencilla es cambiar la temperatura o la concentración de agente precipitante. Estos agentes pueden ser de naturaleza salina (sulfato amónico), solventes orgánicos (2-metil-2,4-pentadiol o MPD) o los precipitantes poliméricos (polietilenglicoles o PEGs). Otras variables que afectan a la solubilidad son: el pH, el empleo de ligandos naturales o artificiales, y un amplio rango de aditivos.

La figura 2.15 muestra un diagrama de fases de una proteína en función de la concentración de precipitante. El diagrama está dividido en dos regiones: una región sobresaturada y una región subsaturada cuyo límite es la curva de solubilidad. La curva de solubilidad que es de carácter termodinámico y está determinada por la cantidad de proteína que permanece en solución en equilibrio

con el cristal a una determinada concentración de agente precipitante. Dentro de la región sobresaturada, podemos distinguir a su vez dos zonas: la zona inestable y la zona metaestable, ambas separadas por la curva de metaestabilidad. Esta curva es de carácter cinético y su localización depende de la rapidez con la que ingresemos en la región sobresaturada desde la subsaturada.

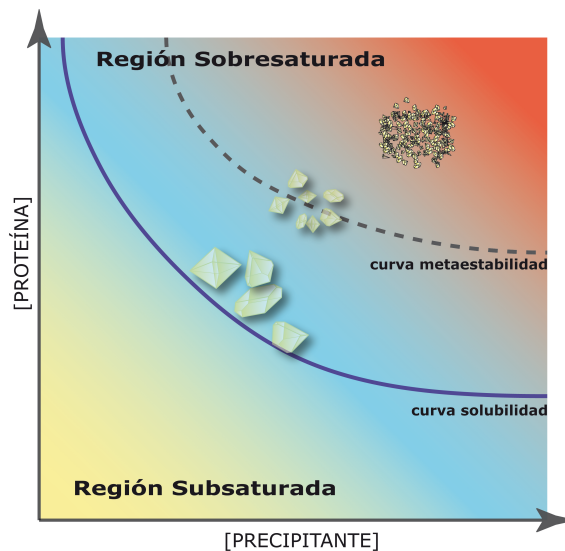


Figura 2.15: Diagrama de fases que ilustra la variación de concentración de proteína en función de la concentración de agente precipitante. En amarillo se muestra la región subsaturada. En azul se muestra la zona metaestable y en rojo la zona inestable. La curva de solubilidad (línea continua azul) separa la región subsaturada de la región sobresaturada. Dentro de la región sobresaturada, la curva de metaestabilidad (línea discontinua negra) separa la zona metaestable de la zona inestable.

Durante el proceso de cristalización, la nucleación ocurre en la región inestable, en las condiciones próximas a la curva de metaestabilidad, ingresando posteriormente en la región de metaestabilidad en la que los núcleos formados pueden continuar creciendo (si aún queda suficiente proteína en solución) hasta alcanzar el equilibrio. Sin embargo, si se alcanza un nivel de sobresaturación elevada muy rápidamente, se puede producir la precipitación masiva de la proteína o que los cristales formados sean de muy baja calidad por la elevada velocidad de crecimiento, inutilizándolos para los experimentos de difracción de rayos-X. Por tanto es necesario un buen control no solo de la nucleación, sino también del crecimiento de los cristales para generar un material útil para la difracción.

Existen diferentes técnicas para conseguir que se den las condiciones óptimas para la cristalización de proteínas como se recoge en la excelente revisión realizada por Carter y Sweet (Carter y Sweet 1997). A continuación hacemos una breve reseña

de las mas empleadas, haciendo especial hincapié en el método de contradifusión y el método de difusión de vapor, empleados en este trabajo.

A. Método de mezcla directa (*batch*)

Se consigue la formación de cristales al mezclar la solución de proteínas y el agente precipitante directamente. Una variante es el *microbatch*, donde una gota que contiene tanto proteína como precipitante se sumerge en un aceite inerte que impide la evaporación de las gotas. Este método se utiliza normalmente para la técnica de sembrado o *seeding*. En este caso, introducimos un cristal ya formado para que aumente su tamaño.

B. Método de diálisis

Este método consiste en unas pequeñas cámaras o botones de diálisis donde se coloca la proteína en un hueco y se cubre con una membrana semipermeable que solo permite el tránsito de pequeñas moléculas. Al sumergir los botones en la solución de agente precipitante, la difusión de este nos permite alcanzar el estado de sobresaturación deseado.

C. Método de contradifusión

Esta técnica consiste en obtener cristales de proteína cuando se ponen en contacto el agente precipitante con solución de proteína introducida en una cámara, de manera que el transporte de masa esté controlado por difusión. Esto se consigue con el uso de capilares cuyo diámetro es muy pequeño en relación a la longitud del contenedor del agente precipitante (Otalora, Gavira et al. 2009).

Al contrario que en otras técnicas de cristalización que buscan encontrar condiciones iniciales próximas al equilibrio, la contradifusión busca altos valores iniciales de sobresaturación, provocando incluso la formación de precipitados amorfos en las primeras etapas del experimento. Para evitar valores de sobresaturación demasiado altos, que se alcanzarían al poner en contacto la solución de proteína con el agente precipitante, se añade una tercera capa entre estas. Esta capa suele ser de un material inerte como gel de agarosa o sílice. Debido a que la proteína se encuentra en un contenedor alargado se produce, por el efecto de la difusión, un gradiente de sobresaturación a lo largo del capilar donde se pueden apreciar los diferentes estadios de cristalización (Fig. 2.16) (García-Ruiz 2003).

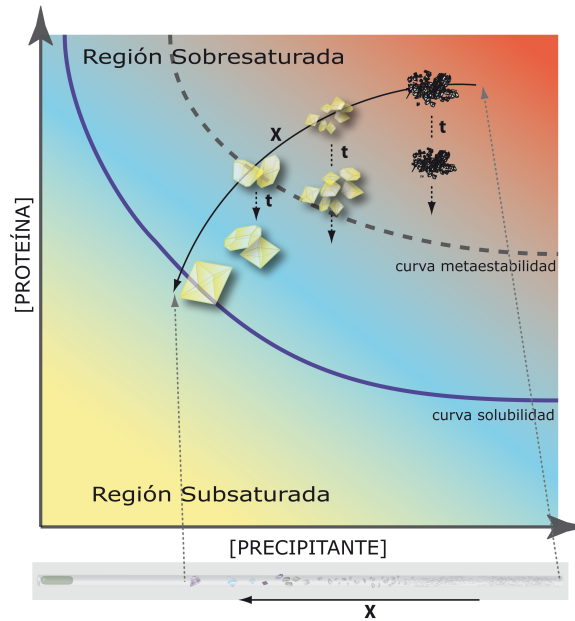


Figura 2.16. Rango de la sobresaturación dentro de un capilar empleando la técnica de contradifusión. Mientras el agente precipitante difunde en la solución de proteína (confinada en el interior del capilar), el efecto de la precipitación y el transporte de masa generan una onda de sobresaturación (X) que viaja a lo largo del capilar a medida que pasa el tiempo (t). El diagrama inferior muestra la evolución teórica de la sobresaturación de una solución de proteína, desde una alta densidad de nucleación a una menor densidad de nucleación.

Hemos empleado este método para la cristalización de tiorredoxinas debido a que presenta grandes ventajas: i) en un experimento de cristalización por contradifusión se recorre una amplia zona del diagrama de fases, lo que permite encontrar automáticamente las condiciones de cristalización óptimas y en consecuencia cristales de elevada calidad (Otalora, Gavira et al. 2009); ii) gracias al empleo de capilares, la cantidad de proteína que usamos es muy baja (entre 0.3 nL y 10 μ L dependiendo del diámetro y longitud del capilar) y iii) al emplear capilares con baja absorción de los rayos-X, es posible difractar los cristales sin necesidad de extraerlos del capilar, evitando así el estrés al que se someten los cristales durante su manipulación (Gavira, Toh et al. 2002).

Existen tres configuraciones para realizar los experimentos de contradifusión, aunque para la cristalización de tiorredoxina solo hemos empleado el método de 3 capas y la *Granada Crystallization Box*.

- **Configuración de 3 capas (3 Layers o 3L):** En esta configuración usamos un capilar tipo *Lindeman* (normalmente empleado en difracción de polvo). La principal peculiaridad de este capilar es que tiene un diámetro mayor por

uno de los extremos. La longitud y diámetro del mismo varía dependiendo del tipo de experimento que vayamos a realizar.

El experimento se realiza del siguiente modo: llenamos la parte de menor diámetro del capilar con la solución de proteína y sellamos el extremo con cera líquida. Esto constituye la capa de proteína. Por el extremo de mayor diámetro, añadimos gel de agarosa (0.5% p/v) fundido y lo dejamos solidificar unos minutos, lo que constituye la capa separadora. La misión de esta capa es retrasar el proceso de difusión inicial y evitar la mezcla inmediata que ocurriría si añadimos el precipitante directamente sobre la capa de proteína. Se puede optar por prescindir de esta capa si la solución de proteína está gelificada (**método de 2 capas**). Por último, añadimos el volumen de la solución precipitante deseada, formando así la tercera capa de agente precipitante y sellamos este extremo con cera (Fig. 2.17 A).

- **Granada Crystallization Box (GCB[®], Triana S&T)**: La GCB es un dispositivo que consta de dos partes: una caja estrecha de poliestireno abierta por un extremo y una tapa de goma para cerrar la caja. La caja es transparente y estrecha para permitir la observación de los experimentos con el microscopio óptico. Las cajas se rellenan con la solución de agente precipitante, sobre la cual colocamos una capa de gel de agarosa. Estas cajas permiten acomodar hasta 6 capilares de un diámetro variable (de 0.1 a 1.5 mm).

Para realizar el experimento, en primer lugar, llenamos el capilar con la solución de proteína y lo sellamos con plastilina por un extremo. A continuación, introducimos el capilar en la GCB por el extremo abierto de tal manera que se quedará una pequeña parte taponándolo al atravesar el gel de agarosa. De esta forma, gran parte del capilar queda sumergido en la solución de agente precipitante, sujeto por la capa de gel. En este sistema, también se forman tres capas. La capa de proteína se encuentra dentro del capilar, la capa de agente precipitante en el interior de la caja y la capa separadora está en el extremo sumergido del capilar (Fig. 2.17 B).

- **Método de acupuntura en gel**: Para este sistema se utilizan también las GCB como contenedor, aunque se puede emplear cualquier otro dispositivo que permita un sellado hermético. La diferencia con la configuración anterior es que el gel de agarosa se encuentra en el fondo de la caja y encima se vierte la solución de agente precipitante tras haber colocado los capilares conteniendo la solución de proteína.

Para realizar el experimento de este modo, los capilares rellenos de solución de proteína se sellan en la parte superior con plastilina y se puntean en la capa de gel. Acto seguido se añade la solución de agente precipitante que difunde por el gel e ingresa en el capilar. Esta configuración permite alterar la composición de la solución de precipitante o sustituirla por completo (Fig. 2.17 C).

Las diferencias entre estas tres configuraciones responden a aspectos cuantitativos del experimento, como el rango de sobresaturación barrido o el tiempo de nucleación. Sin embargo, todos ellos comparten los mismos principios de operación cualitativos, es decir, el desarrollo dinámico de un gradiente de sobresaturación heterogéneo producido en la cámara de la proteína.

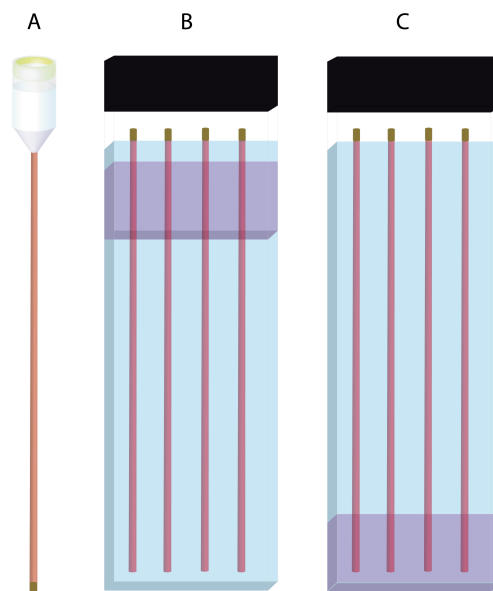


Figura 2.17. Representación esquemática de las tres configuraciones más usadas del método de contradifusión en capilares. A: configuración de tres capas en un capilar de *Lindeman*. B: la configuración *Granada Crystallization Box* (GCB). C: la configuración de acupuntura en gel. En las tres configuraciones, el capilar está relleno con la solución de proteína (rojo). El gel de agarosa (magenta) separa la proteína de la solución precipitante (azul) y/o sirve como soporte físico de los capilares en las configuraciones GCB (B) y Acupuntura en gel (C). La cera para sellar los capilares aparece en amarillo.

D. Método de Difusión de Vapor

En este método, una gota que contiene, por lo general, partes iguales de proteína y agente precipitante se coloca en una cámara hermética con un reservorio que contiene el agente precipitante a mayor concentración que en la gota de tal forma que no entren en contacto. La gota perderá poco a poco agua hasta que se igualen las concentraciones de agente en el reservorio y en la gota. Hay dos variantes: **gota colgante**, la gota se sitúa en un cubreobjetos encima de la solución; y **gota**

sentada (Fig. 2.18 B), la gota se coloca sobre un pedestal. Este método permite la implementación de técnicas de procesamiento de alto rendimiento.

Hemos realizado algunos experimentos de cristalización de tiorredoxina empleando el método de difusión de vapor en modo gota sentada mediante un robot de cristalización modelo *Hamilton Star Plus* (Fig 2.18 A). El robot de cristalización ha permitido que el proceso sea más rápido y además mucho más eficiente, ya que reduce el volumen de muestra empleado de microlitros a nanolitros (100 a 300 nL en nuestro caso). Esto incrementa enormemente la cantidad de condiciones de cristalización que podemos ensayar con la misma cantidad de muestra.

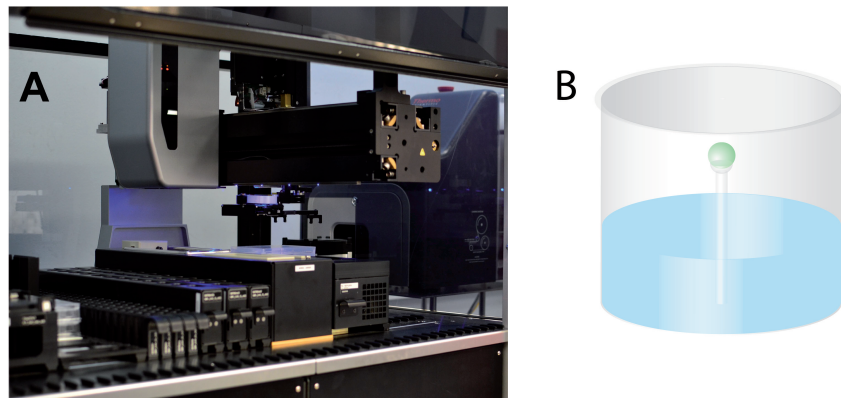


Figura 2.18. Método de gota sentada empleando un robot de cristalización. A: muestra el sistema nanodispensador del robot. B: representación esquemática del método de difusión de vapor en modo gota sentada (que usa el robot) en un pocillo. La solución de agente precipitante (azul) está en el fondo del pocillo. La solución de proteína mezclada con el precipitante (verde) se sitúa sobre un pedestal. Una vez introducidos todos los componentes se sella la placa.

El robot consta de un nanodispensador (Fig. 2.18 A), que toma las muestras de proteína y solución precipitante; un reservorio, donde colocamos las soluciones de las diferentes condiciones; un sistema de lectura de códigos de barras y un dispositivo de sellado de placas. Nuestro sistema puede ensayar hasta 192 condiciones de forma simultánea. Normalmente, las distintas soluciones de agente precipitante están divididas en bloques de 96 condiciones en el reservorio del robot (placas *DeepWell*). Para realizar los ensayos, utilizamos placas de cristalización de 96 pocillos identificadas con un código de barras. El sistema dispensador transporta cada solución de cristalización desde el reservorio del robot al reservorio de cada pocillo en la placa (50 microlitros). A continuación, toma 200 nL de proteína, la deposita en el pedestal del pocillo y la mezcla con 200 nL de la condición de

precipitación. Por último, el robot sella herméticamente las placas. En este momento, las retiramos y las guardamos a la temperatura deseada.

• **Búsqueda de condiciones de cristalización**

El experimento de cristalización se suele llevar a cabo en dos etapas: primero se realiza una búsqueda inicial o muestreo primario de las condiciones de cristalización seguida de una optimización de estas condiciones.

El muestreo primario de condiciones se lleva a cabo utilizando kits de condiciones que exploran el espacio de cristalización. Existen numerosas compañías (Hampton, Triana S&F, etc.) que venden kits comerciales basadas en el rastreo de condiciones que han dado resultados positivos para muchas proteínas en el pasado (Jancarik y Kim 1991). Una vez tengamos cristales (y nos aseguramos de que son cristales de proteína), podemos llevar a cabo una mejora de la calidad y el tamaño de los cristales por medio de un sistema de muestreo secundario u optimización.

En nuestro trabajo, hemos realizado el muestreo primario de las condiciones de cristalización de las tiorredoxinas ancestrales utilizando el método de contradifusión con GBCs y las 3 capas, este último solo se ha utilizado en el caso de las tiorredoxinas con cola de histidinas. Para el muestreo inicial, los capilares utilizados tienen un diámetro de 0.1 mm. Hemos empleado diferentes concentraciones de proteína en cada caso, aunque siempre superior a 30 mg/ml. Los experimentos de muestreo primario los hemos mantenido a 4°C y 20°C en incubadoras. La composición de las condiciones de muestreo primario se detalla a continuación:

Tabla 2.2. Composición de las condiciones empleadas en el muestreo primario de la cristalización de tiorredoxinas ancestrales.

Soluciones precipitantes empleadas en configuración 3 capas:	
Condición #	Agente precipitante
1	MPD 60% (v,v) en NaAc 15 mM pH 3.5, Ac ₂ Cu 1 mM
2	MPD 60% (v,v) en NaAc 15 mM pH 4.0, Ac ₂ Cu 1 mM
3	MPD 60% (v,v) en NaAc 15 mM pH 4.4, Ac ₂ Cu 1 mM
4	MPD 60% (v,v) en NaAc 15 mM pH 5.0, Ac ₂ Cu 1 mM
5	MPD 60% (v,v) en NaAc 15 mM pH 5.5, Ac ₂ Cu 1 mM
6	MPD 60% (v,v) en HEPES 15 mM pH 6.0, Ac ₂ Cu 1 mM
7	MPD 60% (v,v) en HEPES 15 mM pH 6.5, Ac ₂ Cu 1 mM
8	MPD 60% (v,v) en HEPES 15 mM pH 7.0, Ac ₂ Cu 1 mM
9	MPD 60% (v,v) en HEPES 15 mM pH 8.0, Ac ₂ Cu 1 mM
10	MPD 60% (v,v) en Tris-HCl 15 mM pH 8.5, Ac ₂ Cu 1 mM

Kit de Muestreo de Cristalización por Contradifusión en Granada Crystallization Box (GCB-CSK) de 24 condiciones

Condición #	Agente precipitante
1	30% PEG 4000, 0.2M MgCl ₂ , 0.1M Tris-HCl pH 8.50
2	30% PEG 4000, 0.2M AcNH ₄ , 0.1M NaAc pH 4.60
3	20% PEG 8K, 0.2M MgAc ₂ , 0.1M Na-Caco pH 6.50
4	1.25M Na Citrato, 0.1M Na-Hepes pH 7.50
5	1.7M NH ₄ Sulfato, 3.5% PEG 400 0.1M Hepes pH 7.50
6	30% PEG 1500
7	2.0M NH ₄ Sulfato, 0.1M Tris-HCl pH 8.50
8	26% PEG 4000, 0.32M Li Sulfato, 0.1M Tris-HCl pH 8.50
9	30% PEG 8000, 0.2M NH ₄ Sulfato
10	20% PEG 8000, 0.1M Tris-HCl pH 8.50
11	10% Isopropanol, 20% PEG 4000, 0.1M Hepes pH 7.50
12	20% PEG 8000, 0.2M Zn Acetato, 0.1M Na Caco. pH 6.50
13	30% MPD, 0.02M Ca Cloruro, 0.1M AcNa pH 4.60
14	1.5M NH ₄ Fosfato, 0.1M Citrato-Na pH 5.60
15	30% PEG 400, 0.2M Na Citrato, 0.1M Tris-HCl pH 8.50
16	30% PEG 400, 0.2M CaCl ₂ , 0.1M Hepes pH 7.50
17	2.0M Li Sulfato, 0.1M Hepes pH 7.50
18	25% PEG 4000, 0.2M NH ₄ Sulfato, 0.1M AcNa pH 4.60
19	30% MPD, 0.2M Mg Acetato, 0.1M Na Caco. pH 6.50
20	30% PEG 8000, 0.1M NaAc, 0.1M Na Caco. pH 6.50
21	6.0M Na Formato
22	2.0M Na Formato, 0.1M AcNa pH 4.60
23	0.82M K Fosfato, 0.82M Na Fosfato, 0.1M Hepes pH 7.50
24	20% PEG 8000, 0.05M K Fosfato

Kit de Muestreo de Cristalización GCB-CSK Mix-PEGs (pH 4-9)

Condición #	Agente precipitante
1	NaAc 0.1M, PEG 400 20%, PEG 4000 15%, PEG 8000 10%; pH 4.0
2	NaAc 0.1M, PEG 400 20%, PEG 4000 15%, PEG 8000 10%; pH 5.0
3	Tris-HCl 0.1M, PEG 400 20%, PEG 4000 15%, PEG 8000 10%; pH 6.0
4	Tris-HCl 0.1M, PEG 400 20%, PEG 4000 15%, PEG 8000 10%; pH 7.0
5	Tris-HCl 0.1M, PEG 400 20%, PEG 4000 15%, PEG 8000 10%; pH 8.0
6	Tris-HCl 0.1M, PEG 400 20%, PEG 4000 15%, PEG 8000 10%; pH 9.0

Kit de Muestreo de Cristalización GCB-CSK de 30% PEG 1500 (pH 4-9)

Condición #	Agente precipitante
1	NaAc 0.1M, PEG 1500 30%, pH 4.0
2	NaAc 0.1M, PEG 1500 30%, pH 5.0
3	Tris-HCl 0.1M, PEG 1500 30%, pH 6.0
4	Tris-HCl 0.1M, PEG 1500 30%, pH 7.0
5	Tris-HCl 0.1M, PEG 1500 30%, pH 8.0
6	Tris-HCl 0.1M, PEG 1500 30%, pH 9.0

En caso de no obtener ningún cristal con el método de contradifusión, hemos utilizado el método de difusión de vapor en modo gota sentada con el robot de cristalización. Hemos probado 96 condiciones de kit comerciales de la compañía *Hampton Research*. El volumen de las gotas de proteína ha sido entre 200 y 300 nL. Las placas con las gotas las hemos depositado a 4º y 20ºC.

El proceso de optimización lo hemos llevado a cabo por el método de contradifusión en la configuración GCB en capilares de 0.2 ó 0.3 mm de diámetro. En general, este proceso supone ampliar el rango de pH y la concentración de precipitante o de aditivo alrededor de la condición que dio los mejores cristales. Por ejemplo, si la mejor condición que ha dado cristales contiene PEG, podemos probar a ampliar el rango de masa molecular del mismo. Respecto a la solución de proteína, podemos probar diferentes concentraciones o añadirles aditivos que ayuden a la cristalización. Por ejemplo, en una sola caja podemos introducir hasta seis capilares con diferente concentración de proteína, disuelta en distintos tampones y con distintos aditivos.

El seguimiento de la evolución espacial y temporal del crecimiento cristalino se ha realizado con un microscopio óptico de transmisión y una lupa binocular equipados con una cámara digital. La cámara está conectada a un ordenador que permite la adquisición de imágenes en varios estadios del crecimiento. La observación de los capilares la realizamos tras colocarlos, a los tres días, a la semana y pasados tres meses. La figura 2.19 muestra varios cristales de tiorredoxinas ancestrales obtenidos mediante el método de contradifusión.

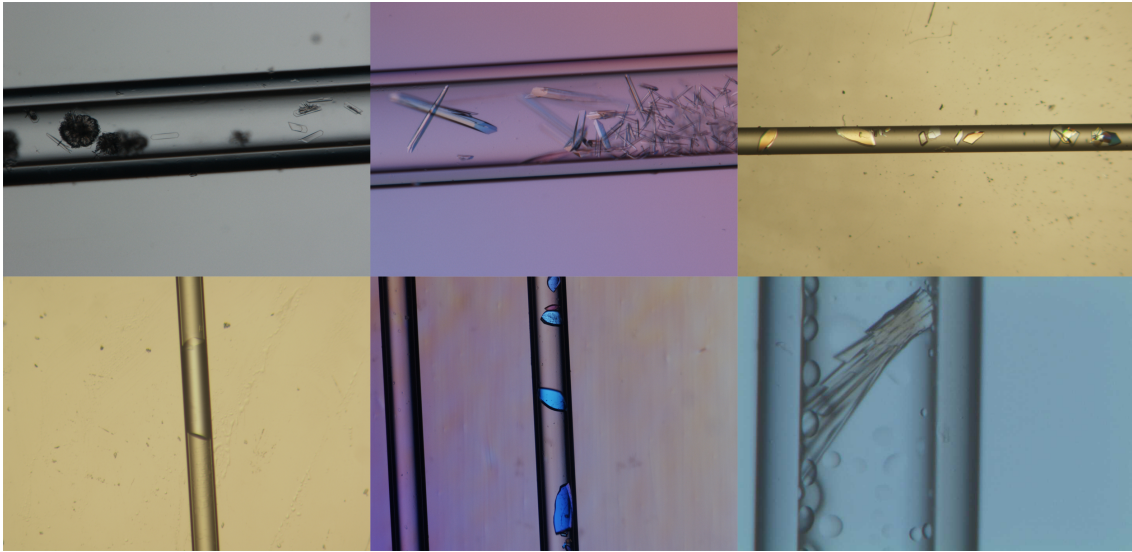


Figura 2.19. Ejemplos de cristales de diferentes tiorredoxinas ancestrales resucitadas obtenidos mediante el método de contradifusión en capilares.

2.2.7.2. Difracción de rayos-X

Cuando un haz de rayos-X incide sobre un cristal de proteína, éste interacciona con los electrones de los átomos que componen dicho cristal haciéndolos vibrar. Los electrones oscilantes emiten a su vez radiación X. Este fenómeno se conoce como **dispersión**. Cuando los átomos de un cristal están ordenados de forma periódica y regular, puede ocurrir que estas emisiones interfieran y se cancelen o que se combinen y refuercen dando lugar a la dispersión cooperativa o **difracción**.

En un experimento de difracción, se hace incidir un haz monocromático sobre un cristal de proteína. Este proceso produce calor y radicales libres en la molécula, por lo que necesitamos preparar el cristal para que pueda soportar estas condiciones durante el máximo período de tiempo posible. Una manera de evitar la destrucción del cristal es realizando los experimentos a una temperatura de -173°C . Sin embargo, antes es necesario sustituir el solvente que recubre y empapa el cristal por una solución que contenga cualquier producto que evite la formación de cristales de hielo (crioprotector). A este procedimiento se le denomina **crioprotección**. Existe una amplia gama de agentes crioprotectores y de metodologías asociadas a su uso, aunque el más empleado es el glicerol, que se añade al cristal mediante un baño rápido en la solución crioprotectora. Para realizar la crioprotección, hemos extraído el cristal del capilar muy cuidadosamente y lo hemos sumergido en una mezcla de 90% de solución de cristalización y 10% de glicerol (solución crioprotectora). A continuación, hemos recogido el cristal de

tioredoxina con un dispositivo de lazo de nylon muy fino (*loop*) y lo hemos introducido en nitrógeno líquido para sobre-enfriarlo, evitando la formación de hielo. Una ventaja del uso de capilares es que, ocasionalmente, podemos utilizarlos directamente para el experimento de difracción sin necesidad de extraer el cristal.

Una vez que tenemos el cristal preparado para soportar las condiciones de difracción, necesitamos montar el cristal en el sistema de difracción. Estos sistemas constan de un generador de rayos-X; una cabeza goniométrica, donde se fija el cristal y permite rotarlo en varias direcciones; un sistema de enfriamiento y un detector de rayos-X. De esta forma, los rayos-X inciden sobre el cristal a la vez que gira y el detector recoge la intensidad de los haces difractados. Un ejemplo del patrón de difracción obtenido se muestra en la figura 2.20.

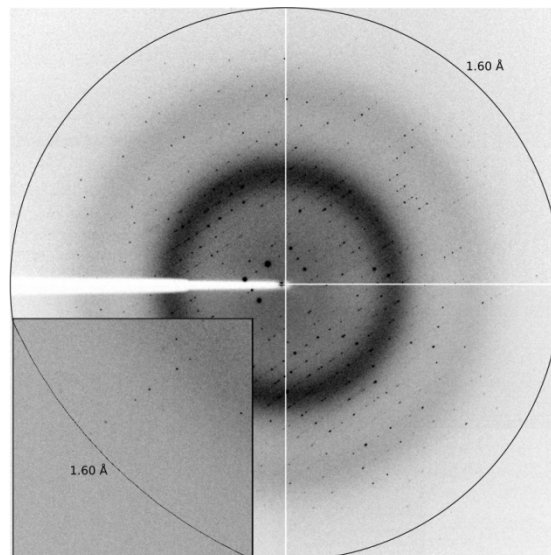


Figura 2.20: Patrón de difracción obtenido de un cristal de tioredoxina ancestral. El inserto muestra una ampliación de los puntos de difracción a alta resolución.

Para realizar los test de difracción hemos utilizado un generador de rayos-X de ánodo rotatorio *Microstar* de BRUKER *Microfocus* (Montel Optics) junto con un detector CCD (*Charged Coupled Device*) Proteum X8 y un sistema de enfriamiento de flujo laminar de nitrógeno gas (*Krioflex*). El haz de rayos-X se genera con una lámpara de ánodo rotatorio de cobre con filtro de níquel que permite seleccionar la radiación $K\alpha$ del cobre ($\lambda=1.54 \text{ \AA}$). Para realizar el test, hemos expuesto el cristal a la radiación durante un tiempo de exposición de 15-180 segundos. Hemos tomado imágenes separadas 90° entre sí y analizado la relación señal/ruido en el límite de difracción del cristal. Hemos llevado a cabo algunos experimentos con el cristal en condiciones crioprotectoras y otros difractando el cristal dentro del capilar

directamente a temperatura ambiente. Con la ayuda del programa *Proteum*, indexamos los datos y definimos la celdilla unidad y el grupo espacial.

Asimismo, para difractar los cristales de proteína, hemos utilizado radiación sincrotrón con longitudes de onda de 1 Å generadas en el acelerador de partículas ESRF (*European Synchrotron Radiation Facilities*) de Grenoble, Francia. Se han utilizado varias líneas con diferentes valores de intensidad y modelo de detectores CCD. Para realizar el experimento, colocamos los cristales en la cabeza goniométrica. Una vez montado el cristal deseado, realizamos un test difracción, para evaluar la calidad del cristal tomando una imagen a 0° y otra a 90°. A continuación, tomamos un juego de datos fijando un tiempo de exposición en el cual, el deterioro del cristal sea el mínimo posible. Fijamos el ángulo de giro en ϕ en función al tiempo de exposición, teniendo en cuenta la mosaicidad del cristal (relacionada con el grado de desorden cristalino). La distancia cristal-detector la definimos en función de la longitud de onda y del límite de resolución observado en el test de difracción, tratando de obtener la máxima resolución posible. De esta forma, hemos tomado las suficientes imágenes para obtener una completitud de datos mayor del 95%

Del análisis de las imágenes del juego de datos (posición e intensidad de los puntos) podemos derivar la geometría de la celdilla unidad y su grupo espacial, que junto con la intensidad asociada a cada punto recíproco es toda la información que nos ofrece un experimento de difracción. Una vez determinada la celdilla unidad y grupo puntual del cristal se puede asignar a cada punto de difracción el plano cristalino al que corresponde (indexado), integrar las intensidades y escalarlas de acuerdo a las condiciones experimentales. En nuestro caso, las imágenes fueron indexadas e integradas con el paquete XDS (Kabsch 2010) o DENZO (paquete HKL2000) y escaladas con el programa SCALA (Evans 2006) del paquete de programas CCP4 (Collaborative Computational Project 1994) o con SCALEPACK del paquete de programas HKL2000 (Otwinowski y Minor 1997).

Existen varios parámetros importantes para caracterizar y determinar la calidad los datos de difracción. Estos se recogen en la tabla de estadísticas del procesado de datos y son: **grupo espacial** al cual pertenece el cristal, geometría de la **celdilla unidad** (en Å), **resolución** mínima y máxima del juego de datos (en Å), **$I/\sigma(I)$** es el cociente entre el promedio de intensidad y la desviación estándar media de la intensidad. Su valor no debería ser menor que 2 para el rango de resolución. El

R_{sym} (%) es una medida del acuerdo entre múltiples medidas de una misma reflexión incluyendo las equivalentes por simetría. El valor de R_{sym} total suele ser en torno a 5 % para un buen juego de datos. Además, se muestran en la tabla la **Completitud** (%) del juego de datos en el rango de resolución escogido, cuyo valor debería ser entre el 85-100%, y por último la **multiplicidad**, que es el promedio del número de veces que se ha medido independientemente cada intensidad de reflexión. La multiplicidad media de un juego de datos es normalmente 4, aunque cuanto mayor sea, más preciso será el juego de datos.

2.2.7.3. Métodos de resolución estructural

El resultado de la difracción de un cristal es una serie de imágenes con un patrón de intensidades. La información estructural está concentrada en dicho patrón de difracción en forma de los denominados factores de estructura, $F(h,k,l)$, asociados a cada intensidad, $I(h,k,l)$. El objetivo del estudio estructural es la determinación de la densidad electrónica, $\rho(x,y,z)$ para cada punto de la celdilla unidad partiendo de la información que nos ofrecen los espectros de difracción. Por tanto, necesitamos establecer una relación directa entre la estructura cristalina (espacio cristalino) y el patrón de difracción (espacio recíproco). Esta relación se establece mediante una función denominada transformada de Fourier (ecuación 2.16).

$$\rho(x,y,z) = \left(\frac{1}{V}\right) \sum_h \sum_k \sum_l |F(hkl)| \cos 2\pi(hx+ky+lz-\phi(hkl)) \quad (2.16)$$

Esta expresión relaciona la densidad electrónica, $\rho(x,y,z)$, en cada punto (x,y,z) del cristal, con los factores de estructura $F(h,k,l)$ definidos en cada punto del espacio recíproco (h,k,l) del espectro de difracción. Los Factores de estructura son ondas, por ello, para poder resolver la ecuación es necesario disponer no sólo de los módulos que obtenemos del experimento de difracción, $|F(h,k,l)|$, sino también de sus fases asociadas, $\phi(hkl)$, cuya información se pierde durante el experimento por que carecemos de una referencia. Esto se conoce comúnmente como "el problema de la fase" en cristalografía.

Si disponemos de las fases podemos reconstruir el mapa de densidad electrónica del cristal aplicando la transformada de Fourier (Ec. 2.16). Se han desarrollado distintas estrategias para poder determinar las fases en experiencias paralelas, entre ellas cabe destacar: el **método de reemplazamiento isomorfo múltiple**

(*Multiple Isomorphous Replacement* o *MIR*); el **método de dispersión anómala múltiple** (*Multi-wavelength Anomalous Diffraction* o *MAD*) ó el **método de reemplazamiento molecular** (*Molecular Replacement* o *MR*). En nuestro caso, y puesto que disponemos de modelos estructurales, hemos empleado el método de reemplazamiento molecular.

- **Método de reemplazamiento molecular.**

Para llevar a cabo este método, debemos utilizar como modelo una proteína homóloga cuya estructura haya sido resuelta previamente. Esta sirve como punto de partida en el cálculo de las fases iniciales. Esta técnica fue descrita inicialmente por Michael Rossmann y David Blow (Rossmann y Blow 1962). El posicionamiento de la molécula conocida (A) en la celdilla unidad de la proteína desconocida (B) requiere determinar su orientación correcta y su posición precisa. Ambas operaciones se realizan empleando las funciones de rotación (R) y traslación (T) (Fig. 2.21). Matemáticamente, se expresa mediante los coeficientes de correlación de las funciones de Patterson del experimento y la calculada con el modelo. Si el coeficiente es alto, quiere decir que existe una buena correlación entre los espectros de difracción experimental y el calculado con la proteína conocida. Cuando se encuentra la mejor orientación y traslación para el modelo, podemos calcular el mapa de densidad electrónica empleando las fases del modelo y las amplitudes medidas experimentalmente.

Hemos empleado como modelo de reemplazamiento la tiorredoxina de *Escherichia coli* (PDB: **2trx**), la tiorredoxina humana (PDB: **1erv**) o una tiorredoxina ancestral que hubiéramos resuelto previamente, todas ellas sin moléculas de agua. En el capítulo 4 se detalla cuáles de estos modelos hemos empleado para la asignación de las fases de cada tiorredoxina ancestral. El programa elegido para llevar a cabo el método de reemplazamiento molecular ha sido *Molrep* (Vagin y Teplyakov 1997), basado en el método de Patterson, del paquete de programas CCP4.

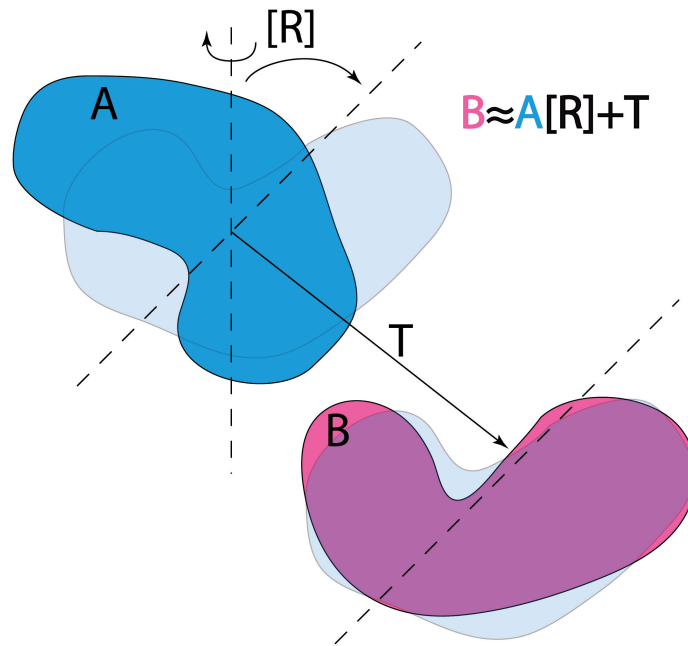


Figura 2.21. Representación esquemática del método de Reemplazamiento Molecular. Las operaciones que permiten superponer el modelo (A) con los datos experimentales (B), son las de rotación, representada por la matriz de rotación [R] y traslación, representada por el vector T.

2.2.7.4. Construcción y refinamiento del modelo

Una vez asignadas las fases y resuelta la transformada de Fourier, hemos obtenido como resultado el **mapa de densidad electrónica** (Fig. 2.21). El mapa mostrará zonas con densidad electrónica positiva, donde podemos trazar de forma automática el modelo estructural y zonas con densidad electrónica negativa, que representa posiciones ocupadas por el modelo pero no presentes en el cristal. Este modelo se define por las coordenadas atómicas (x, y, z) y un factor de temperatura (B) para cada átomo, que expresa la movilidad térmica de dicho átomo alrededor de su posición de equilibrio. La calidad del trazado de la cadena de aminoácidos dentro de los mapas de densidad depende de la exactitud de los datos de difracción y sobre todo de la resolución de los mismos. El programa *Molrep* inserta automáticamente la cadena polipeptídica de la proteína modelo. Por lo tanto, es necesario cambiar la secuencia de aminoácidos a la de nuestra proteína. Este paso, lo realizamos en *COOT*, truncando las cadenas laterales de los aminoácidos a alanina, para luego introducir la verdadera secuencia de la tiorredoxina en estudio.

El modelo inicial normalmente no se adecúa correctamente a las observaciones experimentales como se deduce del mapa de densidad electrónica calculado como

$F_0 - F_c$ (Fig. 2.22), donde F_c corresponde a los factores de estructura calculados a partir del modelo inicial y F_0 corresponde a los factores de estructura experimentales. Por ello, es necesario ejecutar varios ciclos de **refinamiento** en el que se ajustan los parámetros estructurales (x, y, z, B) para optimizar el acuerdo entre la predicción y las observaciones.

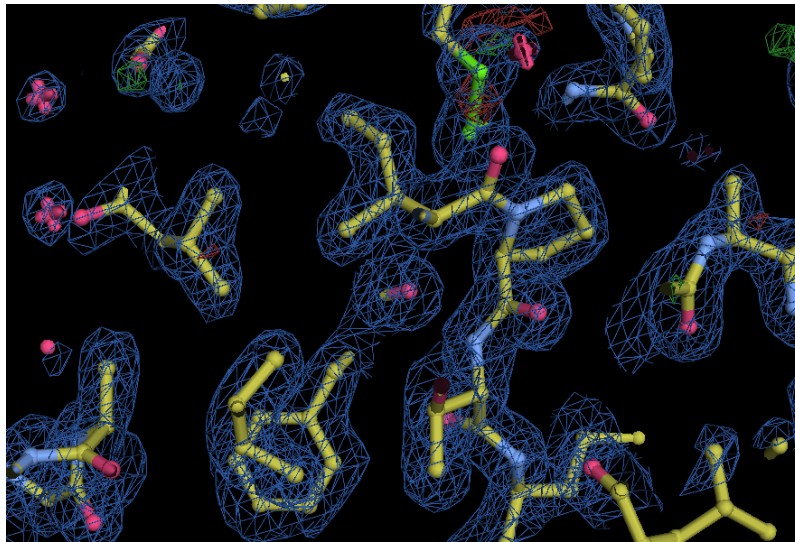


Figura 2.22: Mapa de densidad electrónica de una tioredoxina ancestral. En contorno azul, se representa $2F_0 - F_c$; en contorno verde se representa $F_0 - F_c$; en rojo, zonas de densidad electrónica negativa. En amarillo se representa el esqueleto del modelo. Las moléculas de agua, que se ajustan frente a $2F_0 - F_c$, están coloreadas en rosa.

La calidad del modelo cristalográfico se mide con el factor de desacuerdo (factor R) (Ec. 2.17). El factor R mide las diferencias entre el espectro observado experimentalmente $|F_0|$ y el modelo inicial $|F_c|$. Por tanto, para mejorar el modelo trataremos de minimizar R, hasta un valor constante cuando tengamos estructuras a alta resolución (Wlodawer, Minor et al. 2008).

$$R = \frac{\sum |F_0| - |F_c|}{\sum |F_0|} \quad (2.17)$$

Se ha comprobado sin embargo, que un valor bajo de R no tiene necesariamente que corresponder a un buen modelo final ya que podemos disminuir de forma arbitraria el valor de R si incrementamos el número de parámetros ajustables empleados para describir el modelo (Kleywegt y Brunger 1996). Por ello, se introdujo el cálculo estadístico adicional del factor R_{free} (Kleywegt y Brunger 1996), que corresponde a una aplicación del método estadístico de validación cruzada (Stone 1974). Este método consiste en separar una cierta fracción de los datos de

difracción para la validación cruzada. Se establecerán así dos conjuntos de datos: datos de trabajo (90-95%), que se utilizan para el refinamiento y conjunto de prueba (10-5%), que no se introducen en el proceso de refinamiento. El R_{free} mide el grado en que el modelo predice los datos del conjunto de prueba de tal manera que los cambios en el modelo que no mejoren la capacidad del mismo para predecir este conjunto de datos, harán que el R_{free} no baje. Por tanto, existe una alta correlación entre R_{free} y la precisión del modelo atómico.

Existen varios métodos para realizar el refinamiento (minimizar y convergencia de R y R_{free}). Normalmente, en la primera etapa del refinamiento, aplicamos el refinamiento de cuerpo rígido, en el cual se mantiene la estructura de la proteína como un todo y sólo aplicamos rotaciones y traslaciones. A continuación, usamos el método de *Simulated Annealing* basado en técnicas Monte-Carlo y que equivale al proceso físico de calentar la proteína para luego enfriarla lentamente y obtener la estructura cristalina. Ambos métodos, entre otros, están implementado en el programa *Phenix.refine* del paquete de programas *Phenix* (Adams, Afonine et al. 2010) que hemos usado para el refinado de las estructuras. Seguidamente, procedemos a una reconstrucción manual y asignación de moléculas de agua en el programa *COOT* (Emsley y Cowtan 2004). Llevamos a cabo varios ciclos de refinamiento de este tipo, hasta que R y R_{free} converjan.

En este momento, incluimos en el refinamiento una agrupación de átomos definidos a partir de las librerías de TLS (*Titration-Liberation-Screw*). Este paso consiste en dividir la estructura en unos pocos grupos de átomos basándonos en su flexibilidad (factor B). Para llevar a cabo la agrupación automática, hemos recurrido a *phenix.refine* o al servidor *TLS motion determination* (Painter y Merritt 2006). De nuevo, se realizan varios ciclos de refinamiento y reconstrucción en *COOT* hasta alcanzar la convergencia de los valores de R y R_{free} . Este método nos permite mejorar el modelo debido a que introduce solo 10 parámetros de restricción y realiza un refinamiento de cuerpo rígido de cada grupo, disminuyendo así el número de variables.

Las últimas etapas del refinamiento las realizamos de forma manual en *COOT*. Estas incluyen el refinamiento de los ligandos moleculares e iones diferentes a las moléculas de agua. Además, hay que asignar conformaciones alternas de las cadenas laterales (rotámeros) que no están bien definidas dentro del mapa de densidad electrónica y minimizar la colisión entre átomos.

2.2.7.5. Validación del modelo estructural

El proceso de refinamiento debe proporcionar un modelo fehaciente de la proteína, muy ajustado a las observaciones experimentales. El modelo final puede ser considerado aceptable cuando presenta unos valores que se ajustan al mapa de densidad electrónica y presenta la mayoría de torsiones de enlace dentro de los valores permitidos del diagrama de Ramachandran (Ramachandran, Ramakrishnan et al. 1963). Además, debe haber una ausencia de choques entre cadenas laterales, una distancia y unos ángulos de enlaces adecuados y parámetros térmicos razonables. Hemos chequeado todos estos parámetros mediante los programas *Procheck* (Laskowski, MacArthur et al. 1993) y *Molprobit* (Chen, Arendall et al. 2010) y hemos realizado las correcciones necesarias en el modelo de forma manual con el programa de visualización *COOT*.

Una vez que hemos finalizado este proceso, se guarda la información en un fichero con extensión *pdb*, que contiene las coordenadas de cada átomo, los factores térmicos y cualquier otra información relevante (programas empleados, autores, etc.). Además se genera un segundo fichero con formato *mtz* ó *cif* que contiene la información cristalográfica completa (los factores de estructura calculados). Estos ficheros se depositan en la base de datos *Protein Data Bank* (<http://www.ebi.ac.uk/pdbe/>) donde se revisan, validan y se identifican con un código único para poner esta información a disposición de la comunidad científica.

2.2.8. Análisis bioinformático de alineamientos de secuencias y estructuras de tiorredoxinas

2.2.8.1. Alineamiento de secuencias de tiorredoxinas

El alineamiento de las secuencias de aminoácidos de tiorredoxinas lo hemos realizado con el programa *CLC Sequence Viewer* (*CLC Bio*). El alineamiento nos permite identificar regiones de la secuencia de aminoácidos similares que pueden estar relacionadas con aspectos funcionales, estructurales o evolutivos. Hemos ejecutado el algoritmo de alineamiento introduciendo los parámetros por defecto: coste de apertura de *gap* = 10, extensión de *gap* = 0, coste de *gap* final = como cualquier otro.

2.2.8.2. Alineamiento estructural de tiorredoxinas

Para comparar las diferentes estructuras de tiorredoxinas ancestrales y actuales es necesario realizar un alineamiento estructural. Este alineamiento nos puede ayudar a identificar diferencias en las estructuras de las tiorredoxinas ancestrales y actuales que pueden estar relacionadas con aspectos funcionales, estructurales o evolutivos. Para realizar el alineamiento estructural, hemos utilizado el servidor web **DALI pairwise comparison** (Hasegawa y Holm 2009) en la dirección http://ekhidna.biocenter.helsinki.fi/dali_server/start. Para realizar el cálculo, necesitamos introducir los archivos con las coordenadas atómicas de las dos proteínas a comparar. El programa calcula el alineamiento estructural óptimo y subóptimos entre las distintas cadenas de dos proteínas. El algoritmo DALI (*Distance matrix ALIgment*) utiliza matrices de distancia entre los carbonos alfa (CA) que calcula a partir de las coordenadas de los átomos de las proteínas que compara. Para optimizar la similitud entre los parámetros comparados, se emplea el método de iteración de Monte Carlo (Holm y Sander 1993), que produce varios resultados de alineamiento y los clasifica de acuerdo a las distancias intramoleculares. El resultado que obtenemos es una tabla con los parámetros del alineamiento estructural de una cadena de la proteína con otra, donde se muestra el *Z-score*, que valora la calidad del alineamiento; el número de residuos alineados; la desviación cuadrática media (R.M.S.D.), que es la distancia de desviación media entre los carbonos alfa alineados, y la identidad de secuencia entre dos cadenas. Además, permite la visualización del alineamiento generado a partir del alineamiento estructural (incluyendo la asignación de estructura secundaria con DSSP).

2.2.8.3. Asignación de elementos de estructura secundaria

Para llevar a cabo la asignación de las hélices alfa y de las hojas beta a los modelos estructurales de tiorredoxinas actuales y ancestrales hemos utilizado el programa **DSSP** (Kabsch y Sander 1983; Joosten, te Beek et al. 2011). Este programa asigna estos elementos de estructura secundaria a partir del archivo con las coordenadas de los átomos de la estructura (archivo .pdb) y el cálculo de la energía de los puentes de hidrógeno entre átomos. Los dos mejores enlaces de hidrógeno para cada átomo se usan para determinar el tipo de estructura secundaria al que

pertenece cada residuo de la proteína. Este programa está disponible en la siguiente dirección web: [http:// swift.cmbi.ru.nl/gv/dssp/](http://swift.cmbi.ru.nl/gv/dssp/).

Para identificar qué residuos están en los giros beta hemos utilizado el programa **Promotif** (Hutchinson y Thornton 1996) junto con el programa **DSSP**. Ha sido necesario usar ambos programas para una determinación más precisa de la posición de este elemento. Esto es debido a que *Promotif* asigna algunos residuos a los giros beta cuando estaban dentro de otros elementos predichos por DSSP. El programa *Promotif* define un giro beta como 4 aminoácidos consecutivos (anotados como i , $i+1$, $i+2$ e $i+3$), siendo la distancia entre el carbono alfa del residuo i y el carbono alfa del residuo $i+3$ menor de 7 Å. Además, los dos residuos $i+1$ e $i+2$ no pueden estar formando parte de una hélice. Los giros beta son asignados a una de las 9 clases basándose en los valores de los ángulos ϕ y ψ de los residuos $i+1$ e $i+2$.

2.2.9. Análisis energético-estructural de tiorredoxinas

En este apartado, describiremos las herramientas bioinformáticas que hemos usado para realizar el análisis energético-estructural de las diferentes tiorredoxinas. Para estos cálculos, hemos empleado tanto la secuencia de aminoácidos como los modelos estructurales de las tiorredoxinas ancestrales obtenidos mediante cristalografía de rayos-X y los modelos estructurales de las tiorredoxinas actuales descargados del *Protein Data Bank* (PDB).

2.2.9.1. Cálculos de área superficial accesible al disolvente

Los cálculos de área superficial accesible al disolvente (ASA) de las tiorredoxinas se realizaron con el programa **ASA-ENG** desarrollado en nuestro grupo (por José Manuel Sánchez Ruiz). Este programa emplea una modificación del algoritmo de Shrake-Rupley (Shrake y Rupley 1973) que sitúa aleatoriamente 2000 puntos alrededor de la esfera de Van der Waals que representa cada átomo. Los puntos se colocan a una distancia de las esferas de van der Waals que está determinada por el radio de una esfera (sonda) que "barre" la superficie de la molécula simulando al disolvente. Para ejecutarlo, hemos utilizado el fichero de las coordenadas atómicas

de la estructura de la proteína en cuestión sin aguas y sin rotámeros (eliminado los átomos con configuraciones alternas de menor ocupación). Los parámetros que hemos introducido en el programa han sido un radio de 1.4 Å para la sonda del disolvente y los valores de radios de van der Waals para los átomos de la proteína propuestos por Chothia (Chothia 1975). El programa ASA-ENG calcula los valores de área accesible al disolvente (en Å²) para cada átomo en la estructura nativa. Para calcular el ASA del estado desplegado el programa emplea el modelo tripéptido. Este modelo calcula el ASA para cada residuo (X) situándolo en el tripéptido Gly-X-Gly con la misma conformación que en la proteína nativa. Además del modelo tripéptido del estado desnaturalizado de la proteína, hemos recurrido a los modelos de límite inferior y límite superior propuestos por Rose y colaboradores (Creamer, Srinivasan et al. 1995; Creamer, Srinivasan et al. 1997) en los cuales los valores de ASA del estado desplegado están determinados para el residuo completo, para su cadena lateral o para su cadena principal. Estos modelos establecen los límites inferior y superior del valor del área superficial accesible al solvente de proteínas desplegadas de tal manera que el valor de ASA de las residuos en estado desplegado se sitúa entre estos dos límites extremos. Estos modelos simulan el estado desplegado de la proteína mediante modelos de *hard-spheres* (límite superior) y en modelos con fragmentos de proteínas (límite inferior)

El ASA de la proteína (en Å²) se puede calcular como el cociente en tanto por ciento (% ASA) o la diferencia (Δ ASA) del ASA total de la proteína en estado nativo y el ASA total de la proteína del estado desplegado. Además, con este cálculo podemos obtener el ASA total, de las cadenas laterales y de la cadena principal de cada residuo.

2.2.9.2. Estimación de cavidades internas

El empaquetamiento de los residuos en el interior de las proteínas no siempre es perfecto y pueden aparecer huecos dentro de la molécula. Desde el punto de vista computacional, una cavidad es el espacio vacío que no es accesible al solvente (representado como una esfera de radio conocido o sonda) La identificación de las cavidades internas de las proteínas ha sido posible gracias al programa MSroll que emplea el método de Connolly (Connolly 1983). Este consiste en determinar el volumen y el área de las cavidades internas mediante una sonda que recorre el interior de la cavidad. De esta manera, la superficie de la cavidad (en Å²) será el área recorrida por la esfera en el interior de la cavidad, mientras que el volumen de

la cavidad (\AA^3) será el volumen total que puede ser ocupado por una esfera de un determinado radio. Para ejecutar el cálculo hemos introducido el archivo con las coordenadas de la proteína sin moléculas de agua y sin rotámeros y un valor de sonda esférica de 1.2 \AA . Hemos escogido este radio basándonos en los cálculos de cavidades que realizaron el Dr. Brian Matthews y colaboradores (Eriksson, Baase et al. 1992). En este estudio, los autores han llevado a cabo minuciosamente la determinación del volumen y el área de cavidades diseñadas en la lisozima del fago T₄. Además, hemos utilizado otros programas para la determinación de las cavidades como **VOIDOO** (Kleywegt y Jones 1994) o **Castp** (Dundas, Ouyang et al. 2006).

2.2.9.3. Estimación de los enlaces de hidrógeno

Los enlaces de hidrógeno se forman cuando un átomo dona el átomo de hidrógeno unido covalentemente a él (donador) a un átomo electronegativo (aceptor). En las proteínas, los donadores típicos son el oxígeno de grupos -OH (Ser, Thr, Tyr) o del agua y el nitrógeno en $-\text{NH}_3^+$ (como en Lys, Arg) o en -NH- (como en el enlace peptídico o en Trp, His, Arg). Los pares electrónicos aislados de estos mismos donadores pueden servir como aceptores de enlaces de hidrógeno. Igualmente pueden ser aceptores los pares de los oxígenos carbonílicos =O (como el del enlace peptídico) y los de los nitrógenos con tres enlaces covalentes =N- (como en His o Trp). Estos últimos no pueden ser donadores, pues carecen de hidrógenos.

Para estimar el número de enlaces de hidrógeno intramoleculares en las estructuras de las tiorredoxinas hemos empleado el programa **PFIS v.3.20** (*PDB File Information Software*) (Pace, Horn et al. 2001), proporcionado generosamente por el Dr. Pace (Texas A&M University). Este programa identifica los enlaces de hidrógeno en varios pasos: en primer lugar, calcula las distancias entre todos los donadores y aceptores posibles, y solo selecciona aquellos enlaces en los que la distancia entre el átomo de hidrógeno y el átomo aceptor es menor de 2.6 \AA . También selecciona aquellos en los que la distancia entre el donador y el átomo aceptor que forman el enlace de hidrógeno es menor de 3.6 \AA . En el siguiente paso, el programa evalúa la geometría de los enlaces de hidrógeno seleccionados en el paso anterior. Se tienen en cuenta dos ángulos diferentes:

A) Uno de ellos es el ángulo del conjunto de átomos donador-hidrógeno-aceptor (ángulo DHA), que debe tener un valor entre 90° y 180°.

B) Los valores aceptables para el segundo ángulo, formado por el conjunto de átomos hidrógeno-aceptor-precedente al aceptor (ángulo HAAA, de sus siglas en inglés) depende del tipo de aceptor implicado en el enlace de hidrógeno:

- Para aceptores con orbitales híbridos tipo sp^2 , el carbonilo del esqueleto peptídico, el oxígeno aceptor de las cadenas laterales de Asn, Gln, y Tyr, y el grupo aceptor de His, el ángulo debe estar entre 90° y 180° .
- Para aceptores con orbitales híbridos tipo sp^3 y los grupos hidroxilo de Ser y Thr, el ángulo debe estar entre 60° y 180° .

Un enlace de hidrógeno es aceptado por el programa cuando tanto el ángulo DHA como el ángulo HAAA están dentro de los rangos definidos anteriormente. Antes de ejecutar el programa, es necesario añadir los átomos de hidrógeno a la estructura, ya que no es posible determinar su posición mediante cristalografía de rayos-X. Con tal fin, hemos usado el programa **Chimera** (Pettersen, Goddard et al. 2004). Para ejecutar el programa hemos introducido el archivo con las coordenadas sin moléculas de agua.

Además, hemos utilizado programa **Optimal Hydrogen Bonding Network** (Hooft, Sander et al. 1996) del servidor **What If** (Vriend 1990) (<http://swift.cmbi.ru.nl/servers/html/index.html>). A diferencia del programa anterior, que contabiliza todos los enlaces de hidrógeno posibles o potenciales, esta herramienta calcula el número de enlaces de hidrogeno que estarían formando una red óptima. Hay que tener en cuenta que podría no ser la mejor red estática, sino que son los enlaces de hidrógeno que podrían estar contribuyendo en mayor medida a la estabilidad de la proteína. El programa estima la importancia relativa de unos enlaces de hidrógeno respecto de otros mediante un cálculo de campo de fuerza (Hooft, Sander et al. 1996). Así, a cada uno le asigna una "puntuación" (de 0.0 hasta 1.0) siendo 1.0 la calificación de un enlace de hidrógeno perfecto.

2.2.9.4. Cálculos Electroestáticos

Para realizar los cálculos electrostáticos hemos usado el programa **SATK-ENG**, también desarrollado en nuestro grupo de investigación (Ibarra-Molero, Loladze et al. 1999; Sanchez-Ruiz y Makhatadze 2001). Este programa calcula la energía de interacción entre cargas situadas en las zona de protonación de los grupos ionizables empleando para ello el modelo de Tanford-Kirkwood (Tanford y Kirkwood 1957). De esta manera, obtenemos la carga promedio de cada grupo ionizable, el

valor de pK_a efectivo y el valor de la energía de interacción carga-carga, usando el procedimiento propuesto por Tanford y Roxby (Tanford y Roxby 1972). En esta aproximación se asume que cada grupo ionizable percibe la carga promedio de todos los otros grupos ionizables. Además, el programa nos proporciona los valores de energía total debida a interacciones carga-carga y la carga neta de la proteína a un pH determinado.

Para realizar los cálculos es necesario introducir los datos de ASA para todos los grupos ionizables y la matriz de distancias entre los átomos cargados de estos grupos. Estos datos se obtienen del programa ASA-ENG. Además, introducimos el valor teórico del peso molecular y del volumen específico de la proteína, calculados con la aplicación *Protein Properties Calculator* (Makhatadze, Medvedkin et al. 1990). Por último, fijamos el valor de temperatura en 298°K, la constante dieléctrica del solvente y de la proteína en 78.5 y 4, respectivamente; la fuerza iónica en 0.05 M y damos un valor de pH para el que queremos que se realicen los cálculos.

2.2.9.5. Estimación de los puentes salinos

Para la determinación de los puentes salinos presentes en las proteínas hemos usado el servidor ESBRI: <http://bioinformatica.isa.cnr.it/ESBRI> (Costantini, Colonna et al. 2008). El programa reconoce un puente salino si el grupo negativo (átomo de oxígeno del carboxilo de la cadena lateral de un Asp o Glu) y el grupo positivo (átomo de nitrógeno de la cadena lateral de una Arg, Lys o His) se encuentran a una distancia menor de 4 Å (Kumar y Nussinov 1999; Sarakatsannis y Duan 2005). Para ejecutar el programa introducimos el archivo con las coordenadas de la estructura sin moléculas de agua.

Para comprobar los resultados, hemos utilizado también el programa **Salt Bridges** del paquete de programas **What If** (Vriend 1990). En este caso, el programa reconoce un puente salino si el grupo negativo y el grupo positivo se encuentran a una distancia menor de 7 Å. Para cotejar los resultados de este programa con el anterior, solo hemos tenido en cuenta los que están a menos de 4.0 Å de distancia.

Una vez tenemos la lista completa de puentes salinos para cada proteína, hemos calculado la energía de interacción entre los dos residuos que forman el puente con el programa SATK-ENG como explicamos en el apartado anterior.

2.2.9.6. Cálculos de propensión de elementos de estructura secundaria

En las estructuras de proteínas se puede observar que determinados residuos aparecen con más frecuencia en unos elementos de estructura secundaria que en otros, por ejemplo, la alanina aparece con mucha frecuencia en las hélices alfa. Además se ha demostrado que este fenómeno puede tener algún efecto sobre algunas propiedades de las proteínas, como la estabilidad (Zhang, Baase et al. 1992; Munoz, Blanco et al. 1995). Así, la presencia de tal o cual residuo se ha podido relacionar con la estabilidad del elemento de estructura del que forma parte. Normalmente, el procedimiento empleado en estos estudios ha sido sustituir los residuos con menor propensión para un elemento de estructura secundaria por otro con mayor propensión para dicho elemento y determinar la estabilidad de los mutantes. Hay muchos grupos de investigación que han estudiado esta importante cuestión y han conseguido generar escalas que cuantifican la contribución de cada aminoácidos a la estabilidad de proteínas (Minor y Kim 1994; Munoz y Serrano 1995; Munoz y Serrano 1995; Luque, Mayorga et al. 1996; Rohl, Chakrabartty et al. 1996; Pace y Scholtz 1998; Guruprasad y Rajkumar 2000; Trevino, Schaefer et al. 2007).

En este trabajo hemos determinado la propensión de los residuos de las hélice alfa, las hojas beta y los giros beta de las tiorredoxinas ancestrales. Para ello, partiendo de la asignación de elementos de estructura secundaria (apartado 2.2.8.3) hemos asignado el valor correspondiente a partir de las escalas de propensión para cada aminoácido dentro de un motivo estructural determinado. Las escalas utilizadas pueden encontrarse en los siguientes artículos:

- **Hélices alfa:** Hemos empleado el programa AGADIR (Munoz y Serrano 1995; Munoz y Serrano 1995; Pace y Scholtz 1998) del servidor web <http://agadir.crg.es/about.jsp>. Este programa usa un algoritmo para predecir el contenido de hélices en péptidos. Para ejecutar el programa hemos introducido la secuencia de aminoácidos y los parámetros de fuerza iónica 0.05 M, temperatura 298°K y pH 7. El programa calcula el valor de "helicidad" para cada residuo de la proteína. Además, hemos comparado los resultados de AGADIR con otras escalas de propensión de aminoácidos de hélices alfa (Luque, Mayorga et al. 1996; Rohl, Chakrabartty et al. 1996).

- **Hojas beta:** Hemos empleado la escala termodinámica de propensión de aminoácidos en hojas beta propuesta por el Dr. Peter Kim (Minor y Kim 1994). Para elaborarla se mutaron una serie de residuos en los péptidos modelo y se midió su termoestabilidad, de esta manera obtienen la energía relativa en $\Delta\Delta G$ respecto a alanina para cada aminoácido.

- **Giros beta:** Hemos seguido la misma estrategia para determinar los diferentes tipos de giros beta en nuestras proteínas que el Dr. N. Pace (Trevino, Schaefer et al. 2007). Para estimar la propensión de aminoácidos de situarse en cada tipo de giro beta en términos de potencial de posición, hemos empleado una escala estadística basada en la posición que ocupa cada aminoácido en los diferentes tipos de giros beta. (Guruprasad y Rajkumar 2000). Para esta estadística se emplearon 426 estructuras de proteínas diferentes.

2.2.9.7. Determinación de los contactos entre residuos

Para identificar los residuos que están en contacto dentro de un radio determinado hemos utilizado el programa **Esfera2**, desarrollado en nuestro laboratorio. Para ejecutar el programa necesitamos el archivo con las coordenadas atómicas, el archivo con los valores de ASA para cada aminoácido (archivo de salida del programa ASA-ENG), la posición del aminoácido para el cual queremos determinar los contactos y el valor del radio de la esfera. El resultado será una lista de los átomos que están en contacto con todos los átomos del residuo de interés dentro del radio introducido (en Å).

2.2.10. Análisis por regresión de mínimos cuadrados parciales de las bibliotecas combinatoriales de variantes de tiorredoxinas.

Como hemos mencionado en el apartado 2.2.4, hemos construido y caracterizado dos bibliotecas combinatoriales de variantes de tiorredoxinas. Estas bibliotecas nos contienen hasta 256 posibles variantes cada una. Aunque no es un número muy elevado de variantes, la caracterización de toda la población de la biblioteca es demasiado ardua y tediosa. Por esta razón hemos recurrido a un procedimiento computacional-experimental. Este procedimiento consiste en caracterizar una

pequeña población de variantes de cada biblioteca y seguidamente reconstruir el espacio mutacional completo de la biblioteca mediante un método estadístico. En nuestro estudio contamos con numerosas variables, por ello, hemos empleado un método estadístico de análisis multivariante denominado Regresión de Mínimos Cuadrados Parciales (PLS-R, de sus siglas en inglés). Este procedimiento computacional-experimental ha sido utilizado con éxito recientemente por nuestro grupo (García-Seisdedos, Ibarra-Molero et al. 2012).

2.2.10.1. Introducción al análisis multivariante

En muchas ocasiones, en un experimento científico necesitamos explicar el comportamiento de una variable dependiente (\mathbf{y}) a partir de una variable independiente (\mathbf{x}). Sin embargo, no siempre se produce esta situación, siendo necesaria más de una variable independiente. Para determinar el efecto que tienen un grupo de variables \mathbf{x} (en nuestro caso, una matriz de K variables y N muestras) sobre una o varias variables \mathbf{y} (matriz de M variables y N muestras), utilizamos los métodos de análisis multivariante (Fig. 2.23). Los modelos más utilizados son los siguientes:

- **Regresión Lineal Múltiple (MLR):** Es el modelo más sencillo, fue propuesto en 1960 por Sternberg (Sternberg, Stillo et al. 1960). Calcula la relación lineal entre las variables aplicando el método de mínimos cuadrados (Draper y Smith 1998). Este modelo tiene algunas limitaciones: para poder utilizar el modelo, el número de muestras debe ser mayor que el número de variables, obteniendo infinitas soluciones en caso de no ser así. Además, el número de variables independientes debe ser pequeño y estas deben ser linealmente independientes entre ellas, de lo contrario, el modelo sería ineficaz debido a la colinealidad (diferentes variables independientes aportan la misma información, ya que están correlacionadas) (Geladi y Kowalski 1986). Una manera de evitar estos inconvenientes es trabajar con un número pequeño de variables seleccionadas (no correlacionadas) y eliminar las demás.

- **Métodos basados en la reducción de variables:** Las variables originales se combinan para dar lugar a unas pocas variables linealmente independientes (ortogonales), evitando de esta forma la colinealidad. Por tanto, en este tipo de métodos se realiza una reducción de los datos previa al análisis. Los métodos de reducción de variables más utilizados son la Regresión por Componentes Principales

(PCR de sus siglas en inglés) y la Regresión por Mínimos Cuadrados Parciales (PLS-R).

a) Regresión en Componentes Principales (PCR): Para la reducción de datos, la PCR realiza un análisis de componentes principales (Ringner 2008) en las variables \mathbf{x} , a partir de las cuales y mediante combinación lineal encontrará un menor número de variables ortogonales denominadas **componentes principales**, que se usan como nuevas variables \mathbf{x} para tratar de explicar la variable \mathbf{y} . De esta forma, elimina la información redundante y la contribución del ruido. El principal inconveniente de este método es que el subconjunto elegido de componentes principales para representar las variables \mathbf{x} no tiene porqué ser óptimo para la predicción de la variable \mathbf{y} .

b) Regresión por Mínimos Cuadrados Parciales (PLS-R): Este método utiliza la información contenida tanto en la matriz de variables \mathbf{x} como la contenida en la matriz de variables \mathbf{y} para calcular unas nuevas variables denominadas **variables latentes**, que son parecidas a los componentes principales. La principal diferencia con el modelo anterior es que las variables latentes explican la mayor parte de la varianza en la variable independiente \mathbf{y} , además, se calculan para explicar la covarianza con las variables dependientes tanto como sea posible. Por ello, este modelo es el más adecuado cuando el número de variables independientes sea más elevado que el número de muestras y que variables dependientes.

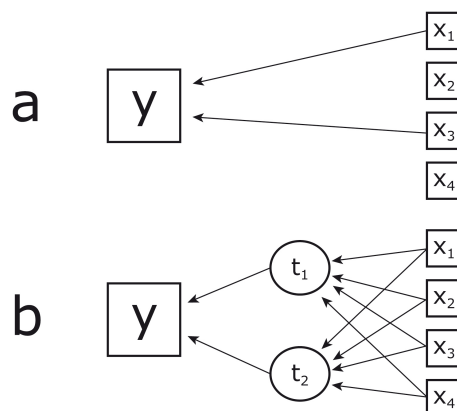


Figura 2.23: Diagrama conceptual de los dos tipos de métodos más populares en el tratamiento de la colinealidad. En "a" se seleccionan cuidadosamente unas pocas variables, prescindiendo del resto, para construir el modelo mediante MLR. En "b" las variables "x" se combinan linealmente para generar otro tipo de variables (t_1 , t_2) que son menos en número que las originales y linealmente independientes entre sí, de manera que estas nuevas variables inferidas se relacionarán con una ecuación de regresión (Naes, Isaaksson et al. 2004).

2.2.10.2. Procedimiento de regresión por mínimos cuadrados parciales

El método de PLS-R fue desarrollada por el economista H. Wold en 1975 (Wold 1975). Originalmente fue concebida para su uso en el campo de la econometría, aunque luego se convirtió en una técnica muy popular en el campo de la quimiometría (Wold 1966; Geladi y Kowalski 1986; Naes, Isaaksson et al. 2004), y hoy en día se ha extendido a otras muchas áreas del conocimiento. En concreto, se ha usado esta herramienta en el campo de ingeniería de proteínas para el estudio de bibliotecas combinatoriales de variantes (Fox, Roy et al. 2003; Fox 2005; Fox, Davis et al. 2007).

En este trabajo hemos tratado de explicar el comportamiento de la estabilidad (T_m) relacionándola con el efecto de varias mutaciones realizadas en la secuencia de las tiorredoxinas. La T_m constituirá pues la variable dependiente, mientras que la información relativa a la presencia o ausencia de mutación, así como el efecto de no aditividad en el efecto de estas mutaciones serán las variables independientes.

En relación a la variable dependiente, antes de comenzar el análisis, es necesario hacer un centrado y escalado de los datos. El centrado consiste en sustraer a cada valor, el promedio de los valores de T_m . De esta forma, todos los valores estarán referidos al valor medio. Además, hacemos un escalado dividiendo los datos centrados por su desviación estándar. Tras lo cual, obtendremos la matriz de variables dependientes formadas por los valores normalizados de T_m de los mutantes.

Para realizar el análisis de los datos hemos usado el programa Unscrambler X (*CAMO software*). Gracias al cual, hemos ajustado cada una de las réplicas generadas al siguiente modelo:

$$y = \sum_k x_k \cdot p_k + \sum_k \sum_{l \neq k} x_{kl} \cdot p_{kl} \quad (2.18)$$

El significado de los símbolos en la ecuación 2.18 es el siguiente: y es la variables dependiente (T_m normalizada). El primer término de la ecuación 2.18. está relacionada con el efecto de las mutaciones, de tal manera que x_k es la variable independiente, que toma valores de 0 ó 1 en función de la ausencia o presencia de

mutación en la posición k , respectivamente, y p_k es un coeficiente de regresión que mide el efecto de la mutación en la posición k en la T_m . El segundo término de la ecuación 2.18. está relacionado con el efecto de no aditividad de las mutaciones, de tal manera que $x_{kl}=x_k \cdot x_l$ es una variable independiente que toma el valor 1 cuando las mutaciones en la posición k y l ocurren simultáneamente (y tomará valor 0 en caso contrario) y p_{kl} es un coeficiente de regresión que mide el efecto de acoplamiento en la T_m entre mutaciones en las posiciones k y l .

A modo de resumen, la ecuación 2.18 representa un modelo simple en el que se ven involucradas un gran número de variables independiente, que incluyen el efecto de mutaciones independientes (p_k) así como la posibilidad de que los efectos de las mutaciones sean no aditivos ($p_{kl} \neq 0$). Como podemos observar, tenemos mayor número de parámetros ajustables que variables dependientes. Por esta razón, es necesario reducir el número de variables mediante el método de PLS-R. Una vez calculadas las variables latentes mediante el algoritmo NIPALS (Wold 1966), se ajusta el modelo a las nuevas variables latentes mediante la ecuación 2.18. A partir del ajuste, podremos predecir de forma sencilla los valores de temperatura de desnaturalización para toda la población de las bibliotecas de variantes de tiorredoxina.

Dado que para la construcción del modelo solo disponemos de los datos experimentales de 20-30 mutantes de 256 que contiene la biblioteca combinatorial, este estará expuesto a cierto grado de incertidumbre, debida a la presencia de posibles errores experimentales o datos anómalos, que no son fáciles de detectar. Para abordar esta incertidumbre, hemos realizado un remuestreo del conjunto de datos utilizando la técnica de *Bootstrapping* (Efron 1979). El remuestreo de los datos con este método se realiza de forma aleatoria permitiendo reemplazamientos, resultando el mismo tamaño de muestra. Para cada conjunto de datos experimentales de cada una de las bibliotecas, hemos obtenido 20 réplicas y las hemos analizado individualmente con el método PLS-R

Con el fin de averiguar si el modelo es apropiado para explicar los datos experimentales o predecir nuevos datos, tenemos que hacer una validación del mismo y evaluar su capacidad predictiva. El método que hemos utilizado ha sido la validación cruzada completa (Wold 1978), ya que es el más indicado cuando la población de muestras es relativamente pequeño (Geladi y Kowalski 1986). La validación cruzada consiste en separar el juego de datos en una serie de segmentos.

En el caso de emplear la validación cruzada completa, el número de segmentos es igual al número de muestras, de tal manera que se construye el modelo con el conjunto de las muestras excepto una, que se deja fuera del cálculo. Esta muestra aislada será predicha por el modelo, permitiendo de esta manera la validación interna del mismo.

La validación se realizará primero para una variable latente, luego se repetirá al introducir la segunda variable y así sucesivamente, de tal forma que todos los segmentos participan en la calibración y validación interna del modelo. El proceso se realiza para cada variable latente calculando el **Error Cuadrático Medio** (MSE de sus siglas en inglés) para cada segmento y acumulándolo, de forma que obtenemos una buena predicción el poder predictivo mediante el Error Cuadrático Medio de la Predicción (MSEP de sus siglas en inglés) es:

$$\text{MSEP} = \frac{\sum_{i=1}^j (\hat{y}_i - y_i)^2}{j} \quad (2.19)$$

Donde \hat{y} son los valores predichos para cada muestra i ; y los valores experimentales para cada muestra i ; j es el número de segmentos utilizados en la valoración, en nuestro caso igual al número de muestras (N).

Otro punto importante es averiguar el número óptimo de variables latentes que hay que seleccionar para que formen parte del modelo y este tenga una importante capacidad predictiva. Por ello, no podemos seleccionar ni demasiados, ya que supondrá un sobre-ajuste del modelo y consecuentemente una pérdida de capacidad predictiva, ni muy pocos, ya que el modelo no será capaz de explicar parte de la varianza de los datos. El número de variables que se han seleccionado es el valor óptimo sugerido por el programa, basándose en el Error Cuadrático Medio de la Predicción (MSEP). Este número de variables ha sido diferente en cada una de las 20 réplicas. Sin embargo, en todos los casos ha sido muy inferior al número de variables originales (entre 3 y 9 nuevas variables).

Por último, con el fin de reconstruir la biblioteca combinatorial, hemos realizado la predicción de la variable dependiente para toda la biblioteca a partir del método PLS-R generado para cada una de las 20 réplicas. Una vez obtenidos los resultados, representamos la T_m normalizada predicha para cada réplica frente a la secuencia del mutante correspondiente. Esto nos permite observar qué variantes tiene mayor

T_m y relacionarla con su secuencia correspondiente. A continuación, generamos en el laboratorio las mejores variantes por mutagénesis dirigida (apartado 2.2.1) y las caracterizamos experimentalmente (apartado 2.2.4), lo cual nos permite una validación de esta estrategia. Las nuevas variantes serán incorporadas al conjunto de datos experimentales formando un nuevo juego de datos sobre el que podremos repetir el proceso mencionado anteriormente. De este modo, obtendremos un mejor modelo que se ajuste más a la realidad y pueda realizar predicciones más precisas.

2.2.11. Bibliografía

- Adams, P. D., P. V. Afonine, G. Bunkoczi, V. B. Chen, I. W. Davis, N. Echols, J. J. Headd, L. W. Hung, G. J. Kapral, R. W. Grosse-Kunstleve, A. J. McCoy, N. W. Moriarty, R. Oeffner, R. J. Read, D. C. Richardson, J. S. Richardson, T. C. Terwilliger and P. H. Zwart (2010). "Phenix: A Comprehensive Python-Based System for Macromolecular Structure Solution." Acta crystallographica. Section D, Biological crystallography **66**(Pt 2): 213-221.
- Becktel, W. J. and J. A. Schellman (1987). "Protein Stability Curves." Biopolymers **26**(11): 1859-1877.
- Bessette, P. H., M. A. Mena, A. W. Nguyen and P. S. Daugherty (2003). "Construction of Designed Protein Libraries Using Gene Assembly Mutagenesis." Methods in molecular biology **231**: 29-37.
- Brandts, J. F. (1964). "The Thermodynamics of Protein Denaturation. I. The Denaturation of Chymotrypsinogen." Journal of the American Chemical Society **86**(20): 4291-4301.
- Carter, C. W. and R. M. Sweet (1997). Methods in Enzymology. New York, Academic Press.
- Chen, V. B., W. B. Arendall, 3rd, J. J. Headd, D. A. Keedy, R. M. Immormino, G. J. Kapral, L. W. Murray, J. S. Richardson and D. C. Richardson (2010). "Molprobity: All-Atom Structure Validation for Macromolecular Crystallography." Acta Crystallogr D Biol Crystallogr **66**(Pt 1): 12-21.
- Chothia, C. (1975). "Structural Invariants in Protein Folding." Nature **254**(5498): 304-308.
- Chowdhry, B. Z. and S. C. Cole (1989). "Differential Scanning Calorimetry: Applications in Biotechnology." Trends in Biotechnology **7**(1): 11-18.
- Collaborative Computational Project, N. (1994). "The Ccp4 Suite: Programs for Protein Crystallography." Acta crystallographica. Section D, Biological crystallography **50**(Pt 5): 760-763.
- Connolly, M. L. (1983). "Solvent-Accessible Surfaces of Proteins and Nucleic Acids." Science **221**(4612): 709-713.
- Cooper, A. and C. M. Johnson (1994). "Introduction to Microcalorimetry and Biomolecular Energetics." Methods in molecular biology **22**: 109-124.
- Costantini, S., G. Colonna and A. M. Facchiano (2008). "Esbri: A Web Server for Evaluating Salt Bridges in Proteins." Bioinformatics **3**(3): 137-138.
- Creamer, T. P., R. Srinivasan and G. D. Rose (1995). "Modeling Unfolded States of Peptides and Proteins." Biochemistry **34**(50): 16245-16250.
- Creamer, T. P., R. Srinivasan and G. D. Rose (1997). "Modeling Unfolded States of Proteins and Peptides. Ii. Backbone Solvent Accessibility." Biochemistry **36**(10): 2832-2835.

- Draper, N. R. and H. Smith (1998). Applied Regression Analysis, Wiley.
- Drenth, J. (2007). Principles of Protein X-Ray Crystallography New York, Springer Advanced Texts in Chemistry.
- Dundas, J., Z. Ouyang, J. Tseng, A. Binkowski, Y. Turpaz and J. Liang (2006). "Castp: Computed Atlas of Surface Topography of Proteins with Structural and Topographical Mapping of Functionally Annotated Residues." Nucleic Acids Res **34**(Web Server issue): W116-118.
- Efron, B. (1979). "Bootstrap Methods: Another Look at the Jackknife." Annals Of Statistics **7**(1): 1-26.
- Emsley, P. and K. Cowtan (2004). "Coot: Model-Building Tools for Molecular Graphics." Acta Crystallogr D Biol Crystallogr **60**(Pt 12 Pt 1): 2126-2132.
- Eriksson, A. E., W. A. Baase, X. J. Zhang, D. W. Heinz, M. Blaber, E. P. Baldwin and B. W. Matthews (1992). "Response of a Protein Structure to Cavity-Creating Mutations and Its Relation to the Hydrophobic Effect." Science **255**(5041): 178-183.
- Evans, P. (2006). "Scaling and Assessment of Data Quality." Acta Crystallographica Section D **62**(1): 72-82.
- Feher, G. and Z. Kam (1985). "Nucleation and Growth of Protein Crystals: General Principles and Assays." Methods in enzymology **114**: 77-112.
- Fox, R. (2005). "Directed Molecular Evolution by Machine Learning and the Influence of Nonlinear Interactions." Journal of theoretical biology **234**(2): 187-199.
- Fox, R., A. Roy, S. Govindarajan, J. Minshull, C. Gustafsson, J. T. Jones and R. Emig (2003). "Optimizing the Search Algorithm for Protein Engineering by Directed Evolution." Protein engineering **16**(8): 589-597.
- Fox, R. J., S. C. Davis, E. C. Mundorff, L. M. Newman, V. Gavrilovic, S. K. Ma, L. M. Chung, C. Ching, S. Tam, S. Muley, J. Grate, J. Gruber, J. C. Whitman, R. A. Sheldon and G. W. Huisman (2007). "Improving Catalytic Function by Prosar-Driven Enzyme Evolution." Nature biotechnology **25**(3): 338-344.
- Freire, E., W. W. van Osdol, O. L. Mayorga and J. M. Sanchez-Ruiz (1990). "Calorimetrically Determined Dynamics of Complex Unfolding Transitions in Proteins." Annu Rev Biophys Biophys Chem **19**: 159-188.
- Garcia-Ruiz, J. M. (2003). "Counterdiffusion Methods for Macromolecular Crystallization." Methods in enzymology **368**: 130-154.
- Garcia-Seisdedos, H., B. Ibarra-Molero and J. M. Sanchez-Ruiz (2012). "Probing the Mutational Interplay between Primary and Promiscuous Protein Functions: A Computational-Experimental Approach." PLoS computational biology **8**(6): e1002558.
- Gavira, J. A., D. Toh, J. Lopez-Jaramillo, J. M. Garcia-Ruiz and J. D. Ng (2002). "Ab Initio Crystallographic Structure Determination of Insulin from Protein to

- Electron Density without Crystal Handling." Acta crystallographica. Section D, Biological crystallography **58**(Pt 7): 1147-1154.
- Geladi, P. and B. R. Kowalski (1986). "Partial Least-Square Regression: A Tutorial." Analytica chimica acta **185**: 1-17.
- Guruprasad, K. and S. Rajkumar (2000). "Beta-and Gamma-Turns in Proteins Revisited: A New Set of Amino Acid Turn-Type Dependent Positional Preferences and Potentials." Journal of biosciences **25**(2): 143-156.
- Hasegawa, H. and L. Holm (2009). "Advances and Pitfalls of Protein Structural Alignment." Current Opinion in Structural Biology **19**(3): 341-348.
- Holm, L. and C. Sander (1993). "Protein Structure Comparison by Alignment of Distance Matrices." Journal of molecular biology **233**(1): 123-138.
- Holmgren, A. (1979). "Thioredoxin Catalyzes the Reduction of Insulin Disulfides by Dithiothreitol and Dihydrolipoamide." J Biol Chem **254**(19): 9627-9632.
- Hooft, R. W., C. Sander and G. Vriend (1996). "Positioning Hydrogen Atoms by Optimizing Hydrogen-Bond Networks in Protein Structures." Proteins **26**(4): 363-376.
- Hutchinson, E. G. and J. M. Thornton (1996). "Promotif--a Program to Identify and Analyze Structural Motifs in Proteins." Protein science : a publication of the Protein Society **5**(2): 212-220.
- Ibarra-Molero, B., V. V. Loladze, G. I. Makhatadze and J. M. Sanchez-Ruiz (1999). "Thermal Versus Guanidine-Induced Unfolding of Ubiquitin. An Analysis in Terms of the Contributions from Charge-Charge Interactions to Protein Stability." Biochemistry **38**(25): 8138-8149.
- Ibarra-Molero, B. and J. M. Sanchez-Ruiz (2006). Differential Scanning Calorimetry of Proteins: An Overview and Some Recent Advances. Advanced Techniques in Biophysics. J. L. Arrondo y A. Alonso, Elsevier: 27-48.
- Ibarra-Molero, B. and J. M. Sanchez-Ruiz (2008). Statistical Differential Scanning Calorimetry: Probing Protein Folding-Unfolding Ensembles. Protein Folding, Misfolding and Aggregation. V. Muñoz, RSC Publishing: 85-103.
- Jancarik, J. and S.-H. Kim (1991). "Sparse Matrix Sampling: A Screening Method for Crystallization of Proteins." Journal of Applied Crystallography **24**(4): 409-411.
- Joosten, R. P., T. A. te Beek, E. Krieger, M. L. Hekkelman, R. W. Hooft, R. Schneider, C. Sander and G. Vriend (2011). "A Series of Pdb Related Databases for Everyday Needs." Nucleic acids research **39**(Database issue): D411-419.
- Kabsch, W. (2010). "Xds." Acta Crystallogr D Biol Crystallogr **66**(Pt 2): 125-132.
- Kabsch, W. and C. Sander (1983). "Dictionary of Protein Secondary Structure: Pattern Recognition of Hydrogen-Bonded and Geometrical Features." Biopolymers **22**(12): 2577-2637.

- Kleywegt, G. J. and A. T. Brunger (1996). "Checking Your Imagination: Applications of the Free R Value." Structure **4**(8): 897-904.
- Kleywegt, G. J. and T. A. Jones (1994). "Detection, Delineation, Measurement and Display of Cavities in Macromolecular Structures." Acta Crystallogr D Biol Crystallogr **50**(Pt 2): 178-185.
- Kumar, S. and R. Nussinov (1999). "Salt Bridge Stability in Monomeric Proteins." Journal of molecular biology **293**(5): 1241-1255.
- Laskowski, R. A., M. W. MacArthur, D. S. Moss and J. M. Thornton (1993). "Procheck: A Program to Check the Stereochemical Quality of Protein." Journal of Applied Crystallography **26**(2): 283-291.
- Luque, I., O. L. Mayorga and E. Freire (1996). "Structure-Based Thermodynamic Scale of Alpha-Helix Propensities in Amino Acids." Biochemistry **35**(42): 13681-13688.
- Makhatadze, G. I., V. N. Medvedkin and P. L. Privalov (1990). "Partial Molar Volumes of Polypeptides and Their Constituent Groups in Aqueous Solution over a Broad Temperature Range." Biopolymers **30**(11-12): 1001-1010.
- McPherson, A. (1999). Crystallization of Biological Macromolecules. New York, Cold Spring Harbor.
- Messerschmidt, A. (2007). X-Ray Crystallography of Biomacromolecules: A Practical Guide Wiley-VCH.
- Minor, D. L., Jr. and P. S. Kim (1994). "Measurement of the Beta-Sheet-Forming Propensities of Amino Acids." Nature **367**(6464): 660-663.
- Munoz, V., F. J. Blanco and L. Serrano (1995). "The Hydrophobic-Staple Motif and a Role for Loop-Residues in Alpha-Helix Stability and Protein Folding." Nat Struct Biol **2**(5): 380-385.
- Munoz, V. and L. Serrano (1995). "Elucidating the Folding Problem of Helical Peptides Using Empirical Parameters. II. Helix Macrodipole Effects and Rational Modification of the Helical Content of Natural Peptides." Journal of molecular biology **245**(3): 275-296.
- Munoz, V. and L. Serrano (1995). "Elucidating the Folding Problem of Helical Peptides Using Empirical Parameters. III. Temperature and Ph Dependence." Journal of molecular biology **245**(3): 297-308.
- Naes, T., T. Isaaksson, T. Fearn and T. Davies (2004). A User-Friendly Guide to Multivariate Calibration and Classification. Chichester, UK, NIR Publications.
- Otalora, F., J. A. Gavira, J. D. Ng and J. M. Garcia-Ruiz (2009). "Counterdiffusion Methods Applied to Protein Crystallization." Prog Biophys Mol Biol **101**(1-3): 26-37.
- Otwinowski, Z. and W. Minor (1997). Processing of X-Ray Diffraction Data Collected in Oscillation Mode. Methods in Enzymology. C. W. Carter y R. M. Sweet. New York, Academic Press. **276: Macromolecular Crystallography, part A**: 307-326.

- Pace, C. N., G. Horn, E. J. Hebert, J. Bechert, K. Shaw, L. Urbanikova, J. M. Scholtz and J. Sevcik (2001). "Tyrosine Hydrogen Bonds Make a Large Contribution to Protein Stability." J Mol Biol **312**(2): 393-404.
- Pace, C. N. and J. M. Scholtz (1998). "A Helix Propensity Scale Based on Experimental Studies of Peptides and Proteins." Biophysical journal **75**(1): 422-427.
- Painter, J. and E. A. Merritt (2006). "Tlsm Web Server for the Generation of Multi-Group Tls Models." Journal of Applied Crystallography **39**(1): 109-111.
- Pettersen, E. F., T. D. Goddard, C. C. Huang, G. S. Couch, D. M. Greenblatt, E. C. Meng and T. E. Ferrin (2004). "Ucsf Chimera--a Visualization System for Exploratory Research and Analysis." Journal of computational chemistry **25**(13): 1605-1612.
- Plaza del Pino, I. M., B. Ibarra-Molero and J. M. Sanchez-Ruiz (2000). "Lower Kinetic Limit to Protein Thermal Stability: A Proposal Regarding Protein Stability in Vivo and Its Relation with Misfolding Diseases." Proteins **40**(1): 58-70.
- Privalov, P. L. (1979). "Stability of Proteins: Small Globular Proteins." Advances in protein chemistry **33**: 167-241.
- Privalov, P. L. (1980). "Scanning Microcalorimeters for Studying Macromolecules." Pure and Applied Chemistry **52**(2): 479-497.
- Privalov, P. L. (1982). "Stability of Proteins. Proteins Which Do Not Present a Single Cooperative System." Advances in protein chemistry **35**: 1-104.
- Privalov, P. L. (1989). "Thermodynamic Problems of Protein Structure." Annual review of biophysics and biophysical chemistry **18**: 47-69.
- Privalov, P. L. and S. A. Potekhin (1986). "Scanning Microcalorimetry in Studying Temperature-Induced Changes in Proteins." Methods in enzymology **131**: 4-51.
- Ramachandran, G. N., C. Ramakrishnan and V. Sasisekharan (1963). "Stereochemistry of Polypeptide Chain Configurations." Journal of molecular biology **7**: 95-99.
- Ringner, M. (2008). "What Is Principal Component Analysis?" Nature biotechnology **26**(3): 303-304.
- Rohl, C. A., A. Chakrabarty and R. L. Baldwin (1996). "Helix Propagation and N-Cap Propensities of the Amino Acids Measured in Alanine-Based Peptides in 40 Volume Percent Trifluoroethanol." Protein science : a publication of the Protein Society **5**(12): 2623-2637.
- Rossmann, M. G. and D. M. Blow (1962). "The Detection of Sub-Units within the Crystallographic Asymmetric Unit." Acta Crystallographica **15**(1): 24-31.
- Sanchez-Ruiz, J. M. (1995). "Differential Scanning Calorimetry of Proteins." Subcell Biochem **24**: 133-176.

- Sanchez-Ruiz, J. M. (1995). "Differential Scanning Calorimetry of Proteins." Sub-cellular biochemistry **24**: 133-176.
- Sanchez-Ruiz, J. M. (2010). "Protein Kinetic Stability." Biophysical chemistry **148**(1-3): 1-15.
- Sanchez-Ruiz, J. M. (2011). "Probing Free-Energy Surfaces with Differential Scanning Calorimetry." Annual review of physical chemistry **62**: 231-255.
- Sanchez-Ruiz, J. M. and G. I. Makhatadze (2001). "To Charge or Not to Charge?" Trends Biotechnol **19**(4): 132-135.
- Sanchez-Ruiz, J. M. and P. L. Mateo (1987). "Differential Scanning Calorimetry of Membrane Proteins." Revis Biol Celular **11**: 15-45.
- Sarakatsannis, J. N. and Y. Duan (2005). "Statistical Characterization of Salt Bridges in Proteins." Proteins **60**(4): 732-739.
- Schellman, J. A. (1987). "The Thermodynamic Stability of Proteins." Annual review of biophysics and biophysical chemistry **16**: 115-137.
- Shrake, A. and J. A. Rupley (1973). "Environment and Exposure to Solvent of Protein Atoms. Lysozyme and Insulin." Journal of molecular biology **79**(2): 351-371.
- Stemmer, W. P., A. Cramer, K. D. Ha, T. M. Brennan and H. L. Heyneker (1995). "Single-Step Assembly of a Gene and Entire Plasmid from Large Numbers of Oligodeoxyribonucleotides." Gene **164**(1): 49-53.
- Sternberg, J. C., H. S. Stillo and R. H. Schwendeman (1960). "Spectrophotometric Analysis of Multicomponent Systems Using Least Squares Method in Matrix Form. Ergosterol Irradiation System." Analytical Chemistry **32**(1): 84-90.
- Stone, M. (1974). "Cross-Validation and Multinomial Prediction." Biometrika **61**(3): 509-515.
- Tanford, C. and J. G. Kirkwood (1957). "Theory of Protein Titration Curves. I. General Equations for Impenetrable Spheres." Journal of the American Chemical Society **79**(20): 5333-5339.
- Tanford, C. and R. Roxby (1972). "Interpretation of Protein Titration Curves. Application to Lysozyme." Biochemistry **11**(11): 2192-2198.
- Trevino, S. R., S. Schaefer, J. M. Scholtz and C. N. Pace (2007). "Increasing Protein Conformational Stability by Optimizing Beta-Turn Sequence." J Mol Biol **373**(1): 211-218.
- Vagin, A. and A. Teplyakov (1997). "Molrep: An Automated Program for Molecular Replacement." Journal of Applied Crystallography **30**(6): 1022-1025.
- Vriend, G. (1990). "What If: A Molecular Modeling and Drug Design Program." Journal of molecular graphics **8**(1): 52-56, 29.
- Weber, P. C. (1991). "Physical Principles of Protein Crystallization." Advances in protein chemistry **41**: 1-36.

- Wlodawer, A., W. Minor, Z. Dauter and M. Jaskolski (2008). "Protein Crystallography for Non-Crystallographers, or How to Get the Best (but Not More) from Published Macromolecular Structures." The FEBS journal **275**(1): 1-21.
- Wold, H. (1966). Estimation of Principal Components and Related Models by Iterative Least Squares. New York, Academic Press.
- Wold, H. (1975). Soft Modeling by Latent Variables; the Non-Linear Partial Least Squares Approach. London, Academic Press.
- Wold, S. (1978). "Cross-Validatory Estimation of the Number of Components in Factor and Principal Components Models." Technometrics **20**(4): 397-405.
- Zhang, X. J., W. A. Baase and B. W. Matthews (1992). "Multiple Alanine Replacements within Alpha-Helix 126-134 of T4 Lysozyme Have Independent, Additive Effects on Both Structure and Stability." Protein science : a publication of the Protein Society **1**(6): 761-776.

**3. RESURRECCIÓN DE PROTEÍNAS
ANCESTRALES, CARACTERIZACIÓN BIOFÍSICA
Y ESTUDIOS DE CATÁLISIS ENZIMÁTICA
MEDIANTE TÉCNICAS DE MOLÉCULA
INDIVIDUAL**

3.1. Introducción

Este trabajo se enmarca dentro de la colaboración entre los grupos de investigación del Dr. Eric Gaucher (*Georgia Institute of Technology, Atlanta*), que realizó el análisis filogenético y la reconstrucción de secuencias ancestrales de tiorredoxinas, el grupo del Dr. Julio Fernández (*Universidad de Columbia, Nueva York*), que llevó a cabo los experimentos de catálisis enzimática mediante microscopía de fuerza atómica; y nuestro grupo de la Universidad de Granada, donde hemos llevado a cabo la obtención y caracterización biofísica de estas enzimas. En concreto, mi contribución a este trabajo ha sido la de expresar y purificar las siete tiorredoxinas ancestrales, así como la determinación de la estabilidad y la actividad oxidoreductasa mediante el ensayo de precipitación de insulina.

La paleobioquímica experimental nos brinda la oportunidad de investigar en el laboratorio la historia molecular de los organismos modernos (Thornton 2004; Hall 2006; Benner, Sassi et al. 2007; Liberles 2007). El estudio de proteínas resucitadas puede revelar información importante sobre la adaptación al clima y a las alteraciones fisiológicas y ecológicas de formas extintas de vida (Chang, Jonsson et al. 2002; Thornton, Need et al. 2003; Thomson, Gaucher et al. 2005; Boussau, Blanquart et al. 2008). Sin embargo, a pesar de los numerosos ejemplos experimentales de resurrección de proteínas ancestrales (Chang, Jonsson et al. 2002; Benner, Sassi et al. 2007; Shimizu, Yokobori et al. 2007), la mayoría existieron hace unos pocos millones de años (Gaucher, Thomson et al. 2003; Thornton 2004; Benner, Sassi et al. 2007). Por consiguiente, muchas hipótesis sobre la vida primitiva están aun sin comprobar, especialmente en periodos de tiempo asociados con cambios rápidos en los sistemas biológicos (Nisbet y Sleep 2001). Este viaje en el tiempo está muy limitado por la ambigüedad en los modelos históricos usados para deducir secuencias ancestrales. Por ejemplo, son fuentes comunes de ambigüedad las incertidumbres en las bases de datos, en los alineamientos de secuencias y en la construcción de árboles filogenéticos, así como los fallos en las teorías evolutivas (Benner, Sassi et al. 2007). No obstante, se han empleado varios enfoques para superar estas limitaciones y se han publicado varios ejemplos de reconstrucción de proteínas pertenecientes al Precámbrico (hace 4.5-0.5 mil millones de años o gigaaños (Ga) en términos geológicos) (Benner, Sassi et al. 2007; Pollock 2007). Uno de esos casos es la reconstrucción de una familia de genes de un factor de elongación de bacterias ancestrales de 0.5-3.5 mil millones de años de antigüedad (Gaucher, Thomson et al. 2003; Gaucher, Govindarajan et

al. 2008). El estudio de estas proteínas mostró que las bacterias primitivas vivían en ambientes cálidos y también reveló una correlación entre la termoestabilidad de la vida primitiva y la temperatura de los océanos primitivos, como se deduce de los registros geológicos.

Estos estudios de reconstrucción de secuencias y la posterior resurrección de las proteínas han despejado el camino a los investigadores a la hora de formular cuestiones sobre los organismos primitivos y sus biomoléculas. Por ejemplo, se sabe poco acerca de cómo ha surgido la química de enzimas primitivas y cómo las condiciones ambientales han afectado a su evolución química (Zalatan y Herschlag 2009). Estas preguntas no pueden responderse directamente al examinar los registros fósiles. En este trabajo, hemos obtenido y estudiado tiorredoxinas de organismos extintos que vivían en el Precámbrico en un esfuerzo para entender la evolución de las reacciones enzimáticas. Las tiorredoxinas pertenecen a una amplia familia de enzimas oxido-reductasas presentes en todos los organismos vivos (Holmgren 1985).

Anteriormente, se han estudiado en detalle los mecanismos químicos de la reducción de puentes disulfuro por tiorredoxinas a una escala inferior al Ångström (Wiita, Perez-Jimenez et al. 2007; Perez-Jimenez, Li et al. 2009), empleando la espectroscopía de molécula individual en su variante de *force-clamp*. La combinación de la espectroscopía de fuerza atómica a nivel de molécula única y la resurrección de tiorredoxinas ancestrales podría aportar nuevos conocimientos acerca de la actividad oxido-reductasa de estas enzimas.

Los resultados de este trabajo han sido publicados en la revista ***Nature Structural & Molecular Biology*** en el año 2011, con el título "*Single-molecule paleoenzymology probes the chemistry of resurrected enzymes*". Autores: Pérez-Jiménez, R.; Inglés-Prieto, A.; Zhao, Z.M.; Sánchez-Romero, I.; Alegre-Cebollada, J.; Kosuri, P.; García-Manyes, S.; Kappock, T.J.; Tanokura, M.; Holmgren, A.; Sánchez-Ruiz J.M.; Gaucher, E.A. y Fernández J.M. El artículo junto con el material suplementario se ajuntan al final de este capítulo.

3.2. Procedimiento experimental

3.2.1. Análisis filogenético y reconstrucción de secuencias de tiorredoxinas ancestrales

El grupo del Dr. Eric Gaucher ha reconstruido 203 secuencias de tiorredoxina actuales pertenecientes a los tres dominios de la vida del *GenBank* (www.ncbi.nlm.nih.gov/genbank); los números de identificación de las secuencias se encuentran en el material suplementario del artículo adjunto. Dichas secuencias se alinearon usando el programa *MUSCLE* (Edgar 2004), realizándose correcciones manuales posteriormente. El análisis filogenético se llevó a cabo mediante el programa *PAUP* 4.0 beta* (*Sinauer Associates*), empleando el criterio de mínima distancia de evolución con 1000 réplicas de muestreo (*bootstrapping*). Las secuencias de las tiorredoxinas ancestrales se reconstruyeron empleando el programa *PAML 3.14* (Yang 1997) e incorporando la distribución *gamma* para las tasas de sustitución variable a lo largo de las posiciones en la secuencia. Para un nodo dado en el árbol, se calculó el valor de probabilidad *a posteriori* para los 20 aminoácidos considerando cada sitio de la secuencia deducida. El aminoácido con la mayor probabilidad se asignó a cada posición de la secuencia.

3.2.2. Expresión y purificación de tiorredoxinas

Los genes de las tiorredoxinas ancestrales se han sintetizados y optimizados para la expresión en células de *Escherichia coli* por la compañía *Epoch Biolabs*. A continuación, el grupo del Dr. Fernández ha clonado estas secuencias en el vector pQE80L (Qiagen). En nuestro laboratorio, hemos expresado y purificado las proteínas como se indica en el apartado 2.2.3.1. de Materiales y Métodos. Al finalizar la purificación de las proteínas, las hemos dializado exhaustivamente frente a 50 mM HEPES, pH 7.0. Por último, hemos comprobado la pureza de la proteína mediante SDS-PAGE.

3.2.3. Medidas de la actividad enzimática de las tiorredoxinas ancestrales con ensayos bioquímicos convencionales

La actividad oxido-reductasa en disolución para las tiorredoxinas ancestrales la hemos determinado usando el ensayo de precipitación de insulina tal como se

indica en el apartado 2.2.5 de Materiales y Métodos (Holmgren 1979; Shimizu, Yokobori et al. 2007; Perez-Jimenez, Wiita et al. 2008). El laboratorio del Dr. Fernández ha analizado la actividad de estas enzimas a pH 5, usando DTNB (ácido 5,5'-ditiobis(2-nitrobenzoico)) como sustrato. En este ensayo, las tiorredoxinas se pre-activaban mediante la incubación con 1mM DTT. La reacción se iniciaba al añadir tiorredoxina activa hasta una concentración final de 4 μ M a una cubeta que contenía 1 mM DTNB en *buffer* 20 mM acetato sódico, pH 5.0. Se ha observado el cambio en la absorbancia a 412 nm debido a la formación de TNB durante 1 min y la actividad se ha determinado a partir de la pendiente $d\Delta A_{412}/dt$. Como control, han realizado un experimento sin tiorredoxina, que se ha restado a los experimentos como línea base.

3.2.4. Experimentos de calorimetría diferencial de barrido.

En nuestro laboratorio, hemos empleado la calorimetría diferencial de barrido para el estudio de la estabilidad térmica de las diferentes tiorredoxinas ancestrales (explicado en detalle en el apartado 2.2.6 del capítulo Materiales y Métodos). La estabilidad térmica de las tiorredoxinas ancestrales y actuales la hemos medido con un calorímetro *VP-DSC* capilar (MicroCal). Los barridos los hemos realizado a una velocidad de 1.5 K/min. Para equilibrar el calorímetro realizamos varias líneas base *buffer-buffer* con el tampón de la diálisis antes del experimento. Las concentraciones de proteínas eran entre 0.3-0.7 mg/ml. Estos valores los hemos determinado espectrofotométricamente a 280 nm, usando coeficientes de extinción molar y pesos moleculares teóricos. Los termogramas experimentales los hemos analizado según el modelo termodinámico de dos estados (apartado 2.2.6.3 de Materiales y métodos).

3.2.5. Experimentos de microscopía de fuerza atómica

El microscopio de fuerza atómica (AFM, de sus siglas en inglés) utilizado por el grupo del Dr. Fernández para el estudio de la catálisis enzimática es un modelo artesanal que está diseñado y fabricado en su laboratorio (Fernandez y Li 2004). El *cantilever* que han empleado es un modelo MLCT de nitrito de silicona (Veeco). El *cantilever* se calibra usando el teorema de equipartición dando como resultado una

constante de elasticidad de 0.02 N m^{-1} (Florin, Rief et al. 1995). El AFM trabaja en modo *Force-Clamp* con una resolución de 0.5 nm. La respuesta (*feedback*) puede alcanzar los 0.5 s. El *buffer* que se ha empleado en estos experimentos ha sido 10 mM HEPES, pH 7.2, 150 mM NaCl, 1 mM EDTA y 2 mM NADPH. La tiorredoxina se añadió a una concentración determinada. El *buffer* también contiene tiorredoxina reductasa (50 nM), para mantener las tiorredoxinas en su estado reducido. La tiorredoxina reductasa de *E. coli* reduce las tiorredoxinas de origen bacteriano, mientras que la tiorredoxina reductasa eucariota reduce tiorredoxinas de arquea y eucariotas. Sin embargo, se obtuvieron resultados similares usando 200 μM DTE, lo cual prueba que las tiorredoxinas reductasas actuales mantienen completamente reducidas las enzimas tiorredoxinas ancestrales. Para los experimentos a pH ácido se empleó el buffer 20 mM acetato sódico pH 5.0 y 200 μM DTE.

Para llevar a cabo el experimento de *force-clamp*, se depositan de 3-6 μl de sustrato con un puente disulfuro ($\sim 0.1 \text{ mg/ml}$) sobre un cubreobjetos recubierto de oro. Como sustrato, se ha empleado una poliproteína diseñada compuesta de 8 repeticiones del dominio I27 de la titina cardiaca humana. Para crear el puente disulfuro, se han mutado los aminoácidos en las posiciones 35 y 75 a cisteínas (I27_{G32C-A75C})₈. Las cisteínas se oxidan de forma espontánea, de manera que forman un puente disulfuro enterrado en la forma plegada de cada monómero I27. A continuación, se añade una gota que contiene $\sim 100 \mu\text{l}$ de la solución de tiorredoxina. El procedimiento en modo *force-clamp* consiste en tres pulsos de fuerza. En el primer pulso, la punta del *cantilever* es presionada contra la superficie a 800pN durante 2 s. En el segundo pulso, la proteína (I27_{G32C-A75C})₈ adherida se estira a 175-185 pN durante 0.2-0.3 s. El tercer pulso, supone el test de fuerza donde se registran los eventos de reducción. Este pulso se aplica a diferentes fuerzas (30-500pN) el número de veces suficientes para capturar todos los posibles eventos de reducción.

Las trazas de AFM se recogieron y analizaron mediante un programa (desarrollado en el laboratorio del Dr. Fernández) ejecutado en Igor Pro 6.03 (*Wavemetrics*). Las trazas que presentan 7-8 eventos de reducción a cada fuerza se suman, se normalizan y se ajustan a un modelo exponencial simple. De este ajuste, se obtiene una constante de relajación τ , a partir de la cual se calcula la velocidad de reducción (r) a una fuerza dada ($r=1/\tau$). Se ha empleado el método de *bootstrapping* para calcular el error de la velocidad de reducción. El *bootstrapping*

se ejecutó 1000 veces para cada velocidad de reducción con el fin de obtener una distribución desde la cual se pueda calcular el error estándar de la media.

3.2.6. Análisis de los datos de AFM

El grupo del Dr. Fernández ha ajustado Los datos obtenidos del AFM al modelo cinético de tres estados (Wiita, Perez-Jimenez et al. 2007; Perez-Jimenez, Li et al. 2009). Brevemente, en este modelo se tienen en cuenta tres mecanismos químicos diferentes. Las constantes de velocidad que se emplea en el modelo cinético son las siguientes:

$$k_{01} = \alpha_0 [\text{Trx}]$$

$$k_{12} = \beta_0 \exp(F\Delta x_{12}/k_B T) + \lambda_0$$

$$k_{02} = \gamma_0 [\text{Trx}] \exp(F\Delta x_{02}/k_B T) + \lambda_0$$

$$k_{10} = \delta_0$$

Las constantes k_{01} y k_{02} dependen de la concentración de proteína de forma lineal. k_{12} y k_{02} dependen de la fuerza de forma exponencial. El modelo cinético se resuelve usando análisis de matrices y los parámetros α_0 , β_0 , λ_0 , γ_0 , δ_0 , x_{02} pueden obtenerse para cada enzima ancestral. Los parámetros cinéticos óptimos para cada una de las tiorredoxinas ancestrales se han calculado con optimización numérica usando el método *downhill simplex* (Nelder 1965).

3.3. Resultados

3.3.1. Reconstrucción de tiorredoxinas ancestrales

A partir de las numerosas secuencias disponibles de tiorredoxinas actuales, el grupo del Dr. Eric Gaucher ha construido un árbol filogenético que incorpora 203 secuencias distintas de tiorredoxina pertenecientes a los tres dominios de la vida (Fig. 3.1). De este árbol, hemos seleccionado varios nodos biológicamente relevantes para reconstruir su secuencia y resucitar las proteínas en el laboratorio. El tiempo de divergencia de los nodos del árbol lo hemos estimado asumiendo que la raíz del mismo se encuentra entre el dominio de bacterias y el ancestro común de arqueas y eucariotas (Hedges 2009).

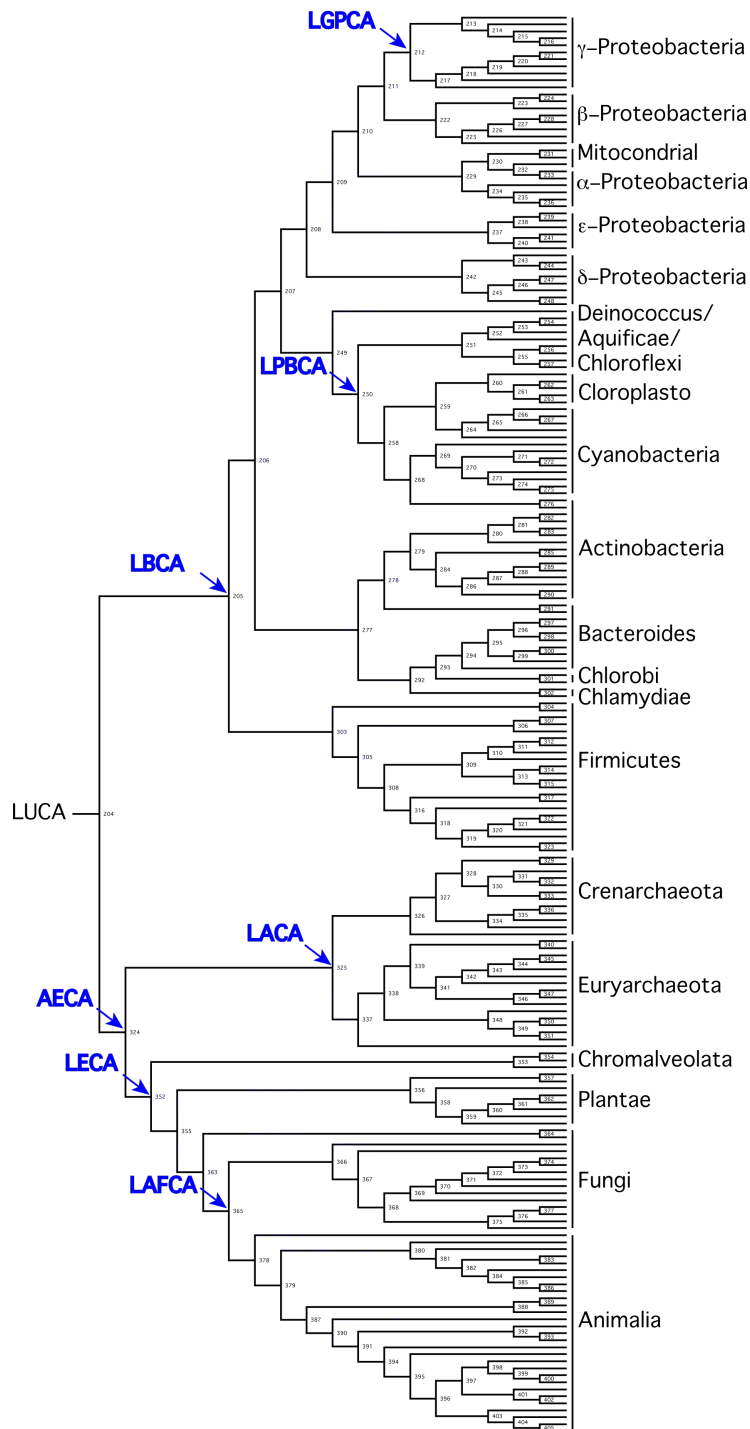


Figura 3.1. Árbol filogenético usado para la reconstrucción de las secuencias de tiorredoxinas ancestrales. Se han usado un total de 203 secuencias de tiorredoxinas actuales. Los nodos de interés resucitados se indican con flechas: el último ancestro común de bacterias (LBCA), el último ancestro común de arquea (LACA), el ancestro común de arquea y eucariotas (AECA), el último ancestro común de eucariotas (LECA), el último ancestro común de las cianobacterias y los grupos de Deinococcus y thermus (LPBCA), el último ancestro común de γ -proteobacteria (LGPCA) y el último ancestro común de animales y hongos (LAFCA). Los números que aparecen corresponden a los diferentes nodos.

Concretamente, hemos resucitado las tiorredoxinas pertenecientes al último ancestro común de bacterias (LBCA), el último ancestro común de arqueas (LACA) y el ancestro común de arquea y eucariotas (AECA). Se considera que estos organismos habitaron la tierra hace 4.2 - 3.5 miles de millones de años , después de divergir del último ancestro común universal (LUCA) (Boussau, Blanquart et al. 2008; Hedges 2009). También se ha seleccionado el nodo correspondiente al último ancestro común de eucariotas (LECA), que vivió en el Proterozoico, hace unos 1.60 Ga. Hemos seleccionado otros dos nodos internos en el linaje de bacterias, el último ancestro común de las cianobacterias y los grupos de deinococcus y thermus (LPBCA), que existió hace aproximadamente 2.5 Ga y representa el origen de las bacterias fotosintéticas, y el último ancestro común de γ -proteobacteria (LGPCA), el cual existió hace aproximadamente 1.61 Ga. Por último, se ha escogido también el último ancestro común de animales y hongos (LAFCA), que vivió hace unos 1.37 Ga (Fig. 3.2 A).

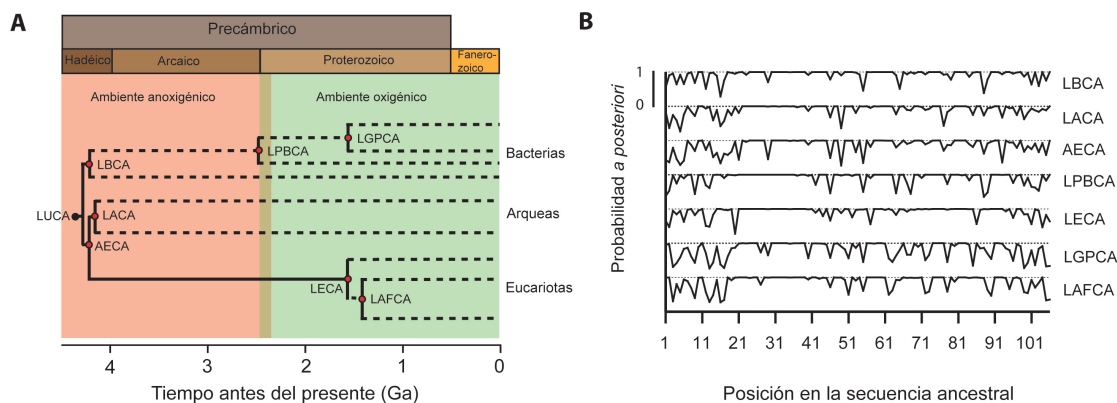


Figura 3.2. Análisis filogenético de tiorredoxinas y reconstrucción de secuencias ancestrales. A: Árbol filogenético esquemático que muestra el tiempo geológico en el que vivieron diferentes organismos extintos. Las líneas discontinuas muestran bifurcaciones posteriores. Los tiempos de divergencia se han recopilado de múltiples fuentes y se resumen en el libro de S. Hedges (Hedges 2009). La figura indica el ambiente global, aunque en los ambientes actuales se encuentran organismos tanto aeróbicos como anaeróbicos. B: Distribución de probabilidad *a posteriori* de los aminoácidos deducidos a lo largo de las 106 posiciones para los nodos internos de interés. El aminoácido deducido para cada posición es el residuo con la mayor probabilidad *a posteriori*.

Aunque hubiera sido interesante estudiar la transición química entre LUCA y el último ancestro común de bacterias y arqueas/eucariotas, la secuencia de la tiorredoxina de LUCA no puede ser inferida con exactitud, ya que habría que hacer suposiciones sobre la vida primitiva muy atrás en el tiempo. Por ejemplo, el último ancestro común de bacterias representa una población primitiva, así como el último ancestro común de arqueas/eucariotas representa otra población primitiva. Es

lógico deducir que LUCA se sitúa en algún lugar en la rama que conecta ambas poblaciones ancestrales. Sin embargo, no se puede identificar su posición exacta del nodo ancestral sin un grupo de referencia (*outgroup*) que permita determinar el tiempo de divergencia de LUCA. Existen razones para creer que LUCA existió en esta rama (tomando como base los eventos de duplicación de genes primitivos, que dieron lugar a los parálogos aún presentes en todos los organismos actuales), pero no se sabe en qué punto de la rama existió. Esta razón nos impidió reconstruir la secuencia de tiorredoxina de LUCA.

Para reconstruir las secuencias de tiorredoxinas ancestrales, el grupo del Dr. Eric Gaucher ha empleado métodos estadísticos basados en máxima verosimilitud (Gaucher, Thomson et al. 2003; Liberles 2007). Para cada nodo del árbol, ha calculado los valores de probabilidad *a posteriori* para los 20 aminoácidos en cada posición de la secuencia ancestral inferida. Estos valores representan la probabilidad de que un cierto residuo ocupe una posición específica en la secuencia en un punto particular en la filogenia. Las probabilidades *a posteriori* se han calculado en base a una matriz de sustitución de aminoácidos (Yang, Kumar et al. 1995). Tras esta etapa, ha reconstruido la secuencia ancestral con mayor probabilidad (M-PAS, según sus siglas en inglés) en cada nodo, asignando a cada posición de la secuencia el residuo con la mayor probabilidad *a posteriori* (Fig. 3.3). Por último, el grupo del Dr. Gaucher ha determinado la distribución de probabilidad *a posteriori* de los aminoácidos deducidos a lo largo de las 106 posiciones en las secuencias seleccionadas (Fig. 3.2 B).

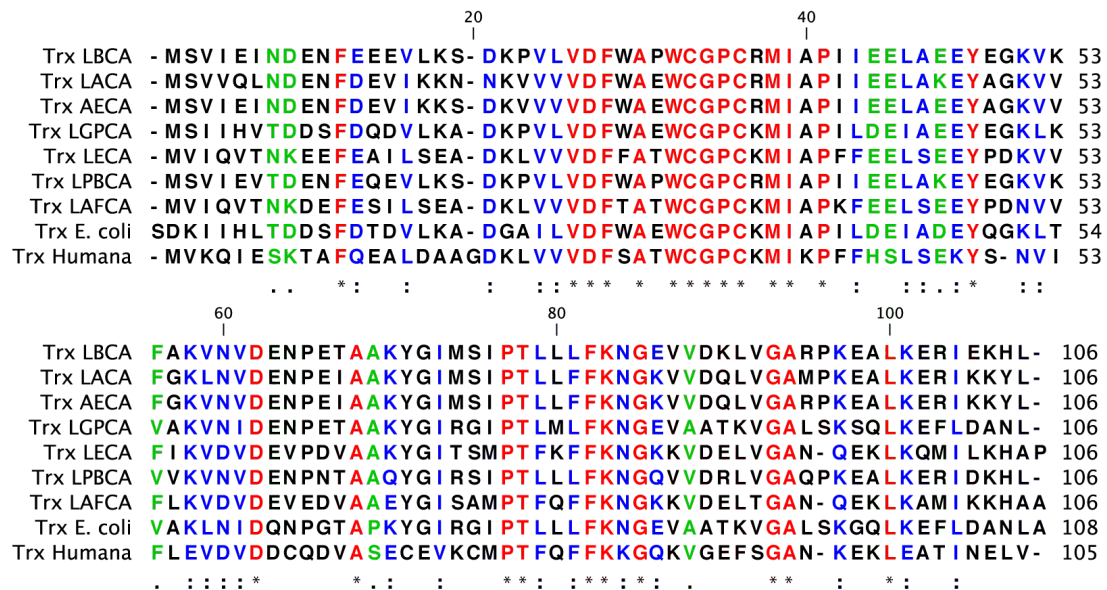


Figura 3.3. Alineamiento múltiple de las secuencias de tiorredoxinas ancestrales de mayor probabilidad (M-PAS, de sus siglas en inglés) pertenecientes a organismos representativos extintos. También se han incluido las secuencias de las tiorredoxinas humana y de *E. coli* en el alineamiento. Se indican en rojo los residuos idénticos, también marcados con un asterisco. Los residuos en azul y verde, también indicados por dos puntos y un punto, respectivamente, representan mutaciones conservativas. El centro activo CGPC está altamente conservado.

3.3.2. Estabilidad térmica de las tiorredoxinas ancestrales

En nuestro laboratorio, como primer paso en el estudio de las propiedades fisicoquímicas de las enzimas resucitadas, hemos empleado la calorimetría diferencial de barrido (DSC) para medir su estabilidad térmica (Fig. 3.4). La temperatura de desnaturalización (T_m) puede proporcionar una idea del rango de temperatura en el que operan las proteínas. Las tres tiorredoxinas más antiguas (LBCA, AECA y LACA) muestran una T_m similar, alrededor de 113°C. LBCA mantiene un estado nativo muy poblado hasta aproximadamente 105°C, donde comienza la transición térmica (Fig. 3.4 G y H). Por el contrario, las T_m de tiorredoxinas actuales de *E. coli* y humana son 88.8 (Fig. 3.4 H) y 93.3°C, respectivamente. La diferencia de temperatura entre las tiorredoxinas más antigua y la más actual es de aproximadamente 25°C, un valor similar al obtenido para el Factor de Elongación bacteriano (Gaucher, Govindarajan et al. 2008), lo que corrobora la hipótesis de la naturaleza termófila de LBCA, AECA y LACA (Boussau, Blanquart et al. 2008).

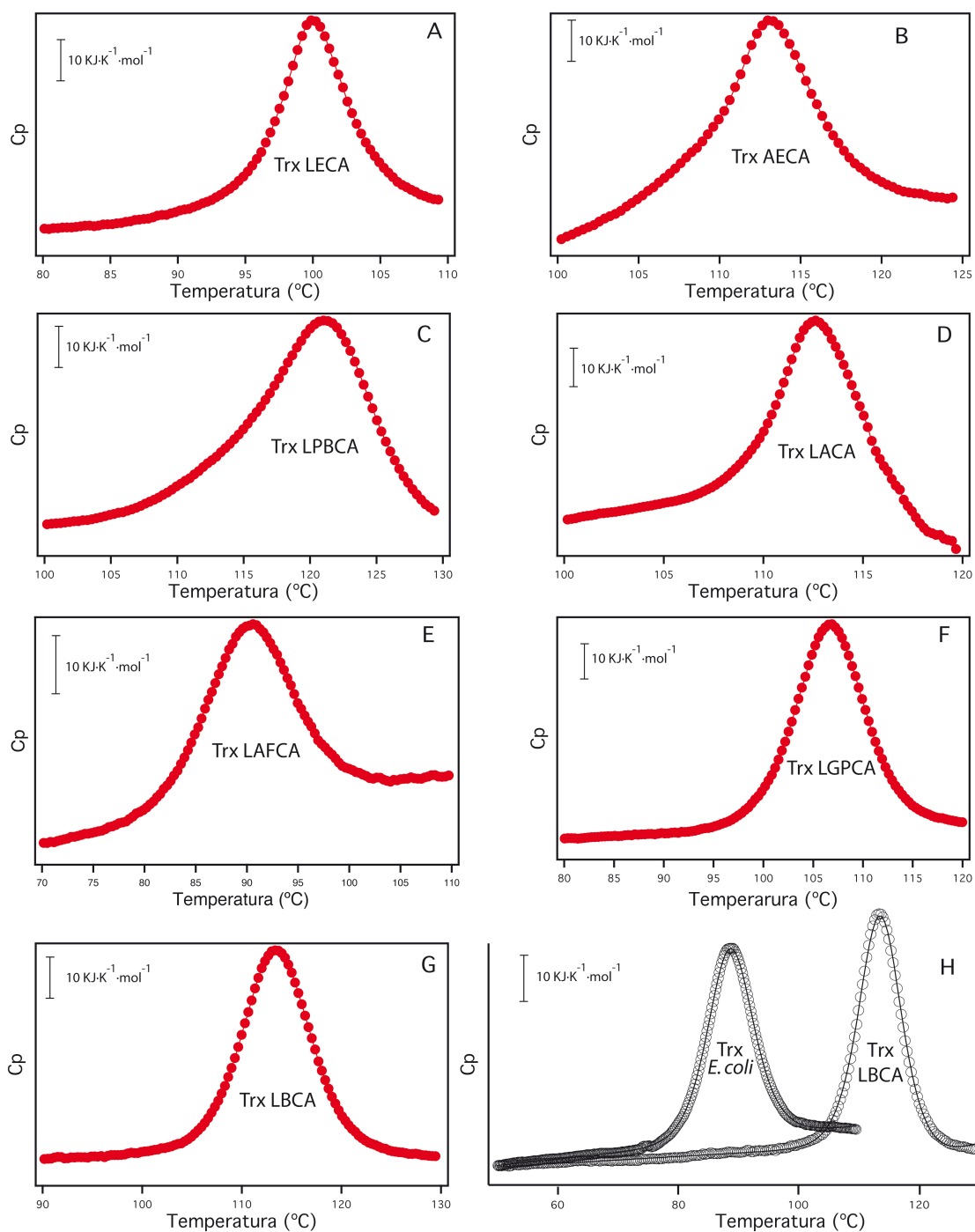


Figura 3.4: A-G: Termogramas de DSC experimentales de las siete tiorredoxinas precámbricas resucitadas. Se han restado las líneas base buffer-buffer a las trazas y se han normalizado por la concentración de tiorredoxina. El máximo corresponde aproximadamente a la T_m de las tiorredoxinas ancestrales resucitadas. H: Termogramas de DSC experimentales para la tiorredoxina actual de *E. coli* y la tiorredoxina ancestral de LBCA. Los experimentos de DSC se llevaron a cabo a una concentración de tiorredoxina de 0.3-0.7 mg/mL, en 50 mM HEPES a pH 7.0 y a una velocidad de barrido de 1.5 K/min. La incertidumbre instrumental es < 0.5 °C.

Hemos comparado la T_m obtenida para cada una de las proteínas resucitadas con su correspondiente tiempo geológico (Fig. 3.5). A partir de estos datos, hemos determinado una tendencia de descenso de la paleotemperatura de 5.8 ± 1.8 K/Ga (Fig. 3.5). Aunque las temperaturas de desnaturalización determinadas para las tiorredoxinas ancestrales siguen una tendencia de enfriamiento similar a la de los océanos primitivos, los valores son alrededor de 50°C más altos que los de las temperaturas de dichos océanos deducidas mediante la deposición de isótopos de oxígeno sobre rocas (Gaucher, Govindarajan et al. 2008) (Fig. 3.5). Nuestra hipótesis es que la evolución opera fundamentalmente sobre la estabilidad cinética de tiorredoxina y que este fenómeno se pudiese reflejar en estabilidad termodinámica. Sin embargo, aparte de la obvia pérdida de función debido a la desnaturalización, no está claro cómo se relaciona exactamente el valor de T_m con el *fitness* de las tiorredoxinas.

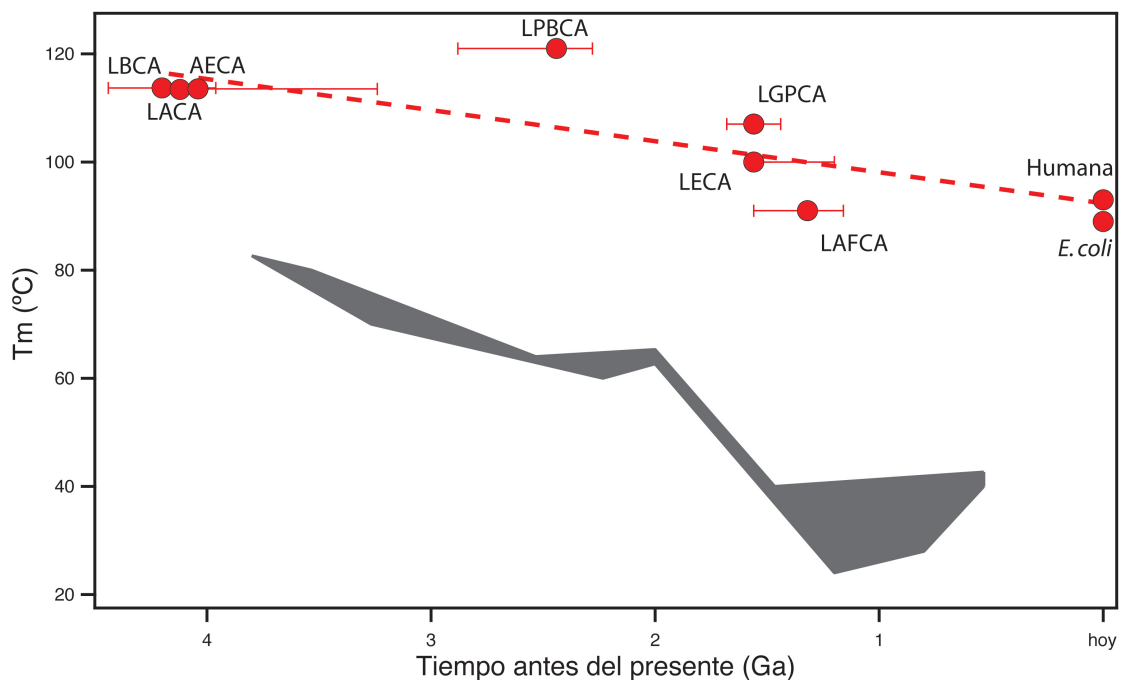


Figura 3.5. Temperatura de desnaturalización (T_m) frente a tiempo geológico estimado para las tiorredoxinas ancestrales. Se indican las tiorredoxinas actuales de *E. coli* y humana. La línea discontinua representa el ajuste lineal a los datos mostrando una tendencia de la paleotemperatura con un decrecimiento en T_m de 5.8 ± 1.8 K/Ga. El área gris representa la temperatura de los antiguos océanos inferida a partir de estudios de isótopos de oxígeno en rocas (Gaucher, Govindarajan et al. 2008).

3.3.3. Cinética química de reducción de puentes disulfuro dependiente de la fuerza.

El grupo del Dr. Julio Fernández ha estudiado los mecanismos químicos de la reducción de puentes disulfuro empleados por las tiorredoxinas resucitadas. Dada la antigüedad de estas enzimas, siendo algunas de ellas anteriores a la acumulación de oxígeno en la atmósfera, parecía lógico pensar que su química catalítica podría ser similar a la de las moléculas que contienen sulfuro. Estas últimas usan un mecanismo de reducción de sustitución nucleofílica bimolecular (S_N2) mediante una colisión dirigida (Russell y Hall 1997). Por el contrario, las tiorredoxinas usan una compleja mezcla de mecanismos químicos, incluyendo una reacción con una unión del sustrato y reorganización crucial que justifica el vasto incremento de eficiencia de tiorredoxina sobre los compuestos sulfuro más simples que estaban disponibles en la geoquímica temprana (Wiita, Perez-Jimenez et al. 2007; Perez-Jimenez, Li et al. 2009).

Recientemente, el grupo del Dr. Fernández ha desarrollado un ensayo, basado en la espectroscopía de fuerzas de molécula individual, para estudiar los mecanismos químicos de reducción de puentes disulfuro de las tiorredoxinas. En este ensayo, se mide el efecto en la velocidad de reducción de un sustrato (con un puente disulfuro) por un nucleófilo mientras se aplica sobre el sustrato una fuerza controlada de forma precisa. Este ensayo distingue rápidamente la química simple S_N2 de nucleófilos, como el hidróxido, glutatión y L-cisteína, de químicas de reducción más complejas de las tiorredoxinas, haciéndolo ideal para el estudio de la química de las tiorredoxinas ancestrales resucitadas (Fig. 3.6) (Wiita, Ainavarapu et al. 2006; Wiita, Perez-Jimenez et al. 2007; Koti Ainavarapu, Wiita et al. 2008; Garcia-Manyes, Liang et al. 2009; Liang y Fernandez 2009; Perez-Jimenez, Li et al. 2009).

Empleando un microscopio de fuerza atómica, se estira una única poliproteína en la misma solución que contiene el nucleófilo deseado. En primer lugar se aplica un pulso de fuerza constante (175-185 pN, 0.2-0.3 s) a la poliproteína (I27_{G32C-A75C})₈, lo que despliega rápidamente los módulos I27_{G32C-A75C} (apartado 3.2.5). Los eventos individuales de desplegamiento se registran como escalones de ~ 11 nm por módulo (Fig 3.6 A y Fig. 3.7). Después de este primer pulso, los puentes disulfuro quedan expuestos al solvente. Seguidamente, se aplica un segundo pulso de fuerza para observar las reducciones individuales de los puentes disulfuro por las

tiorredoxinas, que se registran como una segunda serie de escalones de ~ 14 nm por dominio (Fig. 3.6. A y B, Fig. 3.7. y Fig. 3.8).

Todas las enzimas ancestrales resucitadas han producido peldaños de eventos de reducción (Fig. 3.6. A y B, Fig. 3.7. y Fig. 3.8.), lo que indica que todas ellas son activas. Para medir la velocidad de reducción, se suman 15-80 trazas similares a las que aparecen en la figura 3.6 B, y se ajusta el promedio resultante a una exponencial simple (Fig. 3.6. C). Del ajuste se obtiene la constante de relajación (τ) a partir de la cual calculamos la velocidad de reducción del puente disulfuro a una fuerza dada (Fig. 3.6. D). El método *bootstrapping* se emplea para obtener el error de la velocidad de reducción. Este proceso se repite a diferentes fuerzas de estiramiento (Fig. 3.6. C). El conjunto de datos resultante es una medida de la dependencia con la fuerza de la velocidad de reducción de los puentes disulfuro.

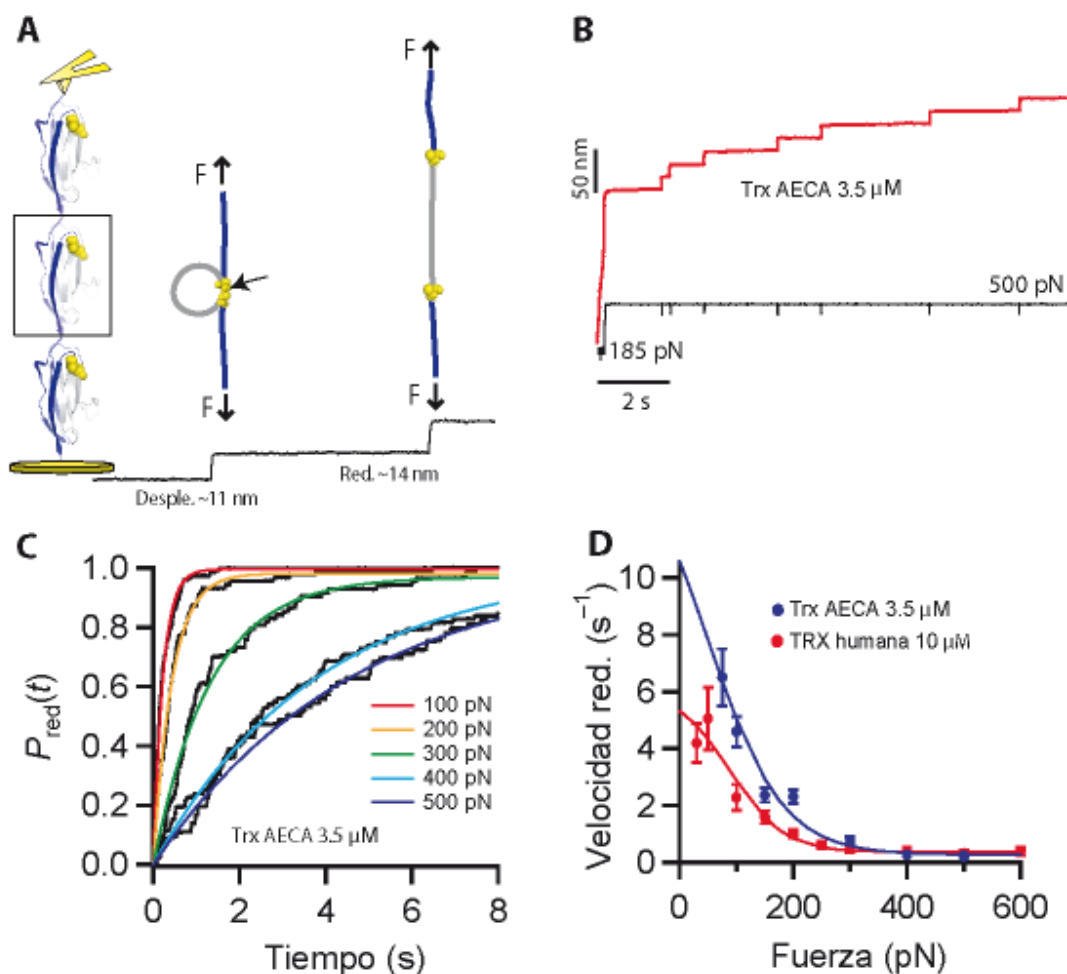


Figura 3.6. Ensayo de molécula individual de reducción de puentes disulfuro. A: representación esquemática del ensayo. Un primer pulso de fuerza despliega rápidamente los dominios I27_{SS} (escalones de 11 nm). Cuando el puente disulfuro queda expuesto al solvente una única molécula de tiorredoxina puede reducirlo (escalón de 14 nm). B: Trazas experimentales de *force-clamp* mostrando reducciones de puentes disulfuro de una poliproteína (I27_{G32C-A75C})₈. El pulso de desplegamiento se estableció en 185 pN 0.2 s, y el pulso de fuerza para las reducciones a 500 pN. C: Probabilidad de reducción ($P_{red}(t)$) resultante de sumar y normalizar las trazas con las reducciones a diferentes fuerzas para la tiorredoxina de AECA (3.5 μ M). D: dependencia de la velocidad de reducción con la fuerza para la tiorredoxina de AECA y la tiorredoxina humana. Ambas enzimas muestran un patrón similar: dependencia negativa de la velocidad de reducción con la fuerza, entre 30-200 pN, consistente con un mecanismo Michaelis-Menten y un mecanismo independiente de la fuerza (por encima de 200 pN), que podría describirse mediante una reacción SET (Perez-Jimenez, Li et al. 2009). Nótese la mayor actividad de la tiorredoxina AECA (3.5 μ M para tiorredoxina de AECA frente a los 10 μ M para la humana). Las líneas representan los ajustes al modelo cinético de tres estados. Las barras de error vienen dadas por el error estándar de la media obtenido usando el método de *bootstrapping*.

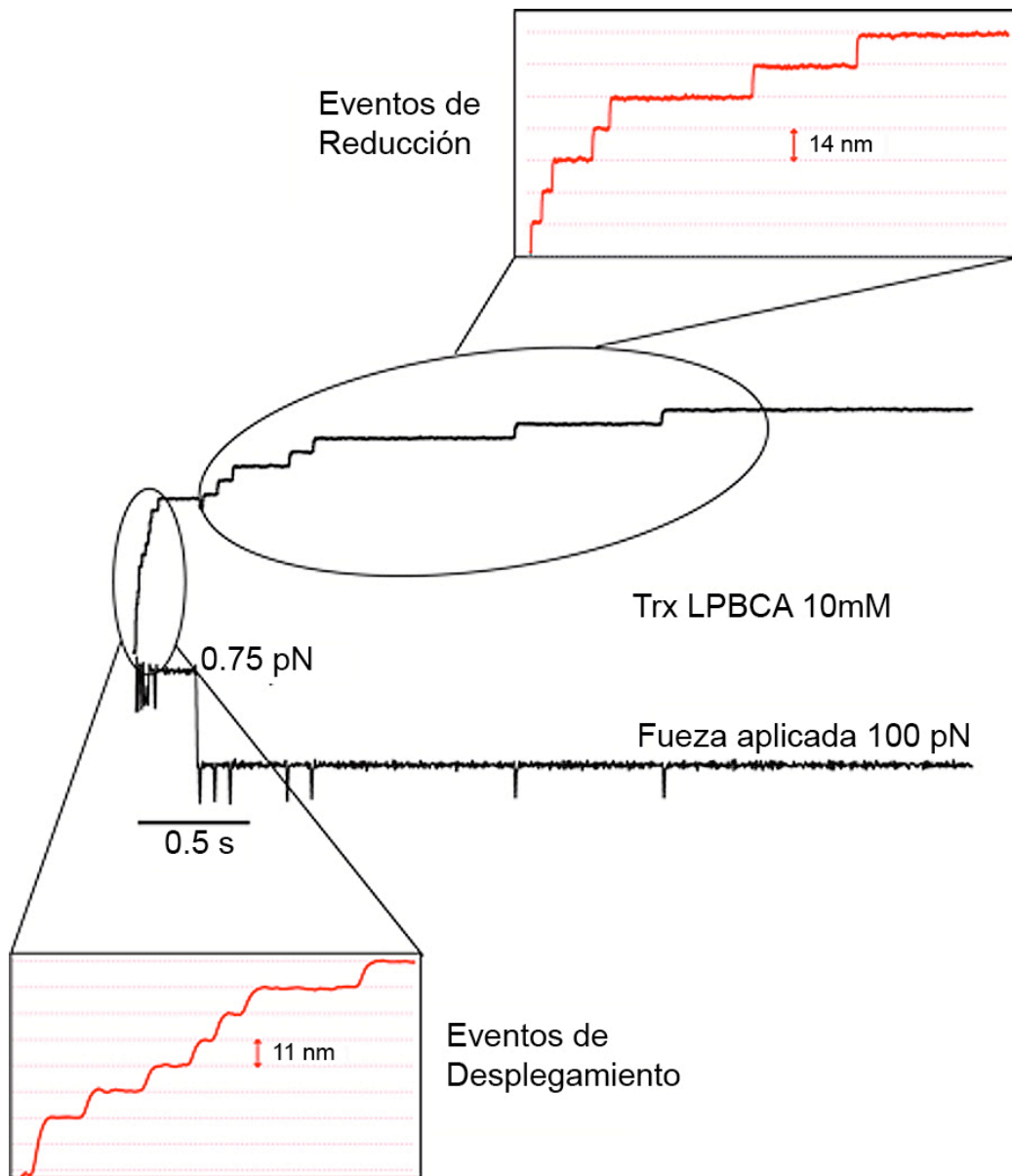


Figura 3.7. Experimento de *force-clamp* para la detección de eventos de reducción de puentes disulfuro. Un primer pulso de fuerza (175 pN, 0.3 s) despliega los dominios I27_{G32C-A75C}, excepto el lazo formado por el puente disulfuro. Los eventos de desplegamiento se pueden registrar como series de escalones de ~11 nm por módulo (panel inferior). Un segundo pulso de fuerza (100 pN) es aplicado para observar las reducciones individuales de los puentes disulfuro por las enzimas tioredoxina. En este caso, la liberación de los residuos atrapados en el lazo formado por el puente disulfuro da lugar a incrementos de longitud de ~14 nm por módulo (panel superior).

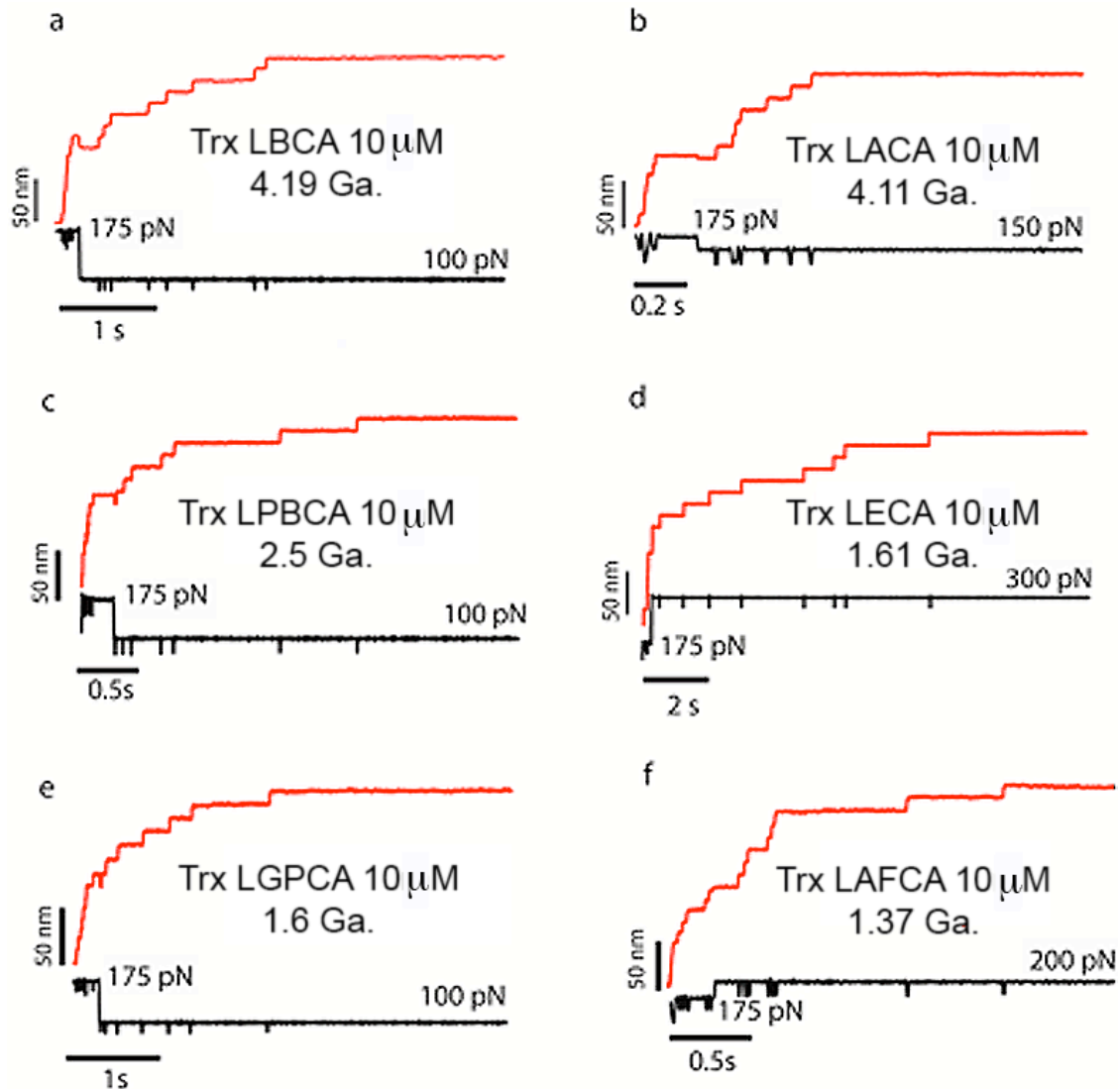


Figura 3.8. Trazas experimentales de reducción de puentes disulfuro mediante tiorredoxinas ancestrales. Se muestran tanto el pulso de desplegamiento (175 pN) como el pulso a diferentes fuerzas realizado para que se produzcan las reacciones. Se emplean numerosas trazas como estas (entre 15-80) a cada fuerza para completar la dependencia con la fuerza de la reducción de puentes disulfuro mediante tiorredoxinas ancestrales (Fig. 3.7).

El mecanismo químico de reducción de puentes disulfuro se puede distinguir según su sensibilidad a la fuerza aplicada al sustrato (Perez-Jimenez, Li et al. 2009). Los agentes reductores simples de grupos tiol muestran una dependencia con la fuerza en la que la velocidad de reducción aumenta siempre con la fuerza aplicada al sustrato (Wiita, Ainavarapu et al. 2006; Koti Ainavarapu, Wiita et al. 2008). Por el contrario, las tiorredoxinas actuales muestran una dependencia negativa con la fuerza en el rango comprendido entre 30-200 pN (Perez-Jimenez, Li et al. 2009) (Fig. 3.9). Este mecanismo es consistente con una reacción de unión de Michaelis-Menten seguida por una reorientación inhibida por la fuerza del puente disulfuro del sustrato que es necesaria para que la reacción S_N2 tenga lugar (Wiita, Perez-

Jimenez et al. 2007). En un segundo mecanismo, la velocidad de reducción aumenta exponencialmente a fuerzas por encima de los 200 pN. Este mecanismo puede describirse por una reacción simple S_N2 y sólo se encuentra en tiorredoxina de origen bacteriano. Además, en todas las tiorredoxinas hay una pequeña contribución de un mecanismo de reducción independiente de la fuerza que podría atribuirse a una reacción de transferencia de un único electrón (SET) (Perez-Jimenez, Li et al. 2009). Este último mecanismo de reducción puede observarse en todas las enzimas ancestrales (Fig. 3.9).

Estas enzimas poseen patrones similares a los encontrados en las tiorredoxinas actuales (Fig. 3.6 D y Fig. 3.9). Se puede ajustar la dependencia de velocidad de reducción con la fuerza de las enzimas resucitadas usando el modelo cinético de tres estados, que ha sido previamente empleado con tiorredoxinas actuales (Wiita, Perez-Jimenez et al. 2007; Perez-Jimenez, Li et al. 2009) (Tabla 3.1).

Tabla 3.1. Parámetros cinéticos para las tiorredoxinas ancestrales y para las actuales de *E. coli* y humana. Los parámetros se han obtenido usando el modelo cinético de tres estados (Wiita, Perez-Jimenez et al. 2007; Perez-Jimenez, Li et al. 2009). Estos parámetros son el resultado de la optimización numérica de los ajustes globales usando el método *downhill simplex*. Los errores corresponden a la desviación estándar.

Trx	α_0 ($\mu\text{M}^{-1}\cdot\text{s}^{-1}$)	β_0 (s^{-1})	γ_0 ($\mu\text{M}^{-1}\cdot\text{s}^{-1}$)	k_{10} (s^{-1})	Δx_{12} (Å)	Δx_{02} (Å)	λ_0 (s^{-1})
LBCA	0.47±0.08	30±2	0.004±0.001	5.8±0.7	- 0.74±0.06	0.19±0.02	0.09±0.02
LACA	8.2±0.2	43±3	-	3.8±1	- 0.76±0.04	-	0.38±0.05
AECA	4.2±0.3	25±2	0.019±0.004	3.8±0.6	- 0.84±0.05	0.19±0.02	0.21±0.04
LPBCA	0.47±0.04	30±3	0.017±0.002	4.9±0.5	- 0.71±0.01	0.17±0.02	0.19±0.02
LECA	0.76±0.08	38±2	-	4.2±0.7	- 0.80±0.03	-	0.18±0.01
LGPCA	0.48±0.02	34±2	0.012±0.002	3.8±0.4	- 0.83±0.02	0.17±0.02	0.35±0.02
LAFCA	0.81±0.10	37±3	-	4.6±0.8	- 0.74±0.03	-	0.06±0.02
<i>E. coli</i>	0.25±0.02	24±2	0.012±0.002	4.7±0.5	- 0.74±0.05	0.16±0.01	0.08±0.02
Humana	0.52±0.05	33±2	-	3.1±0.9	- 0.71±0.05	-	0.35±0.02

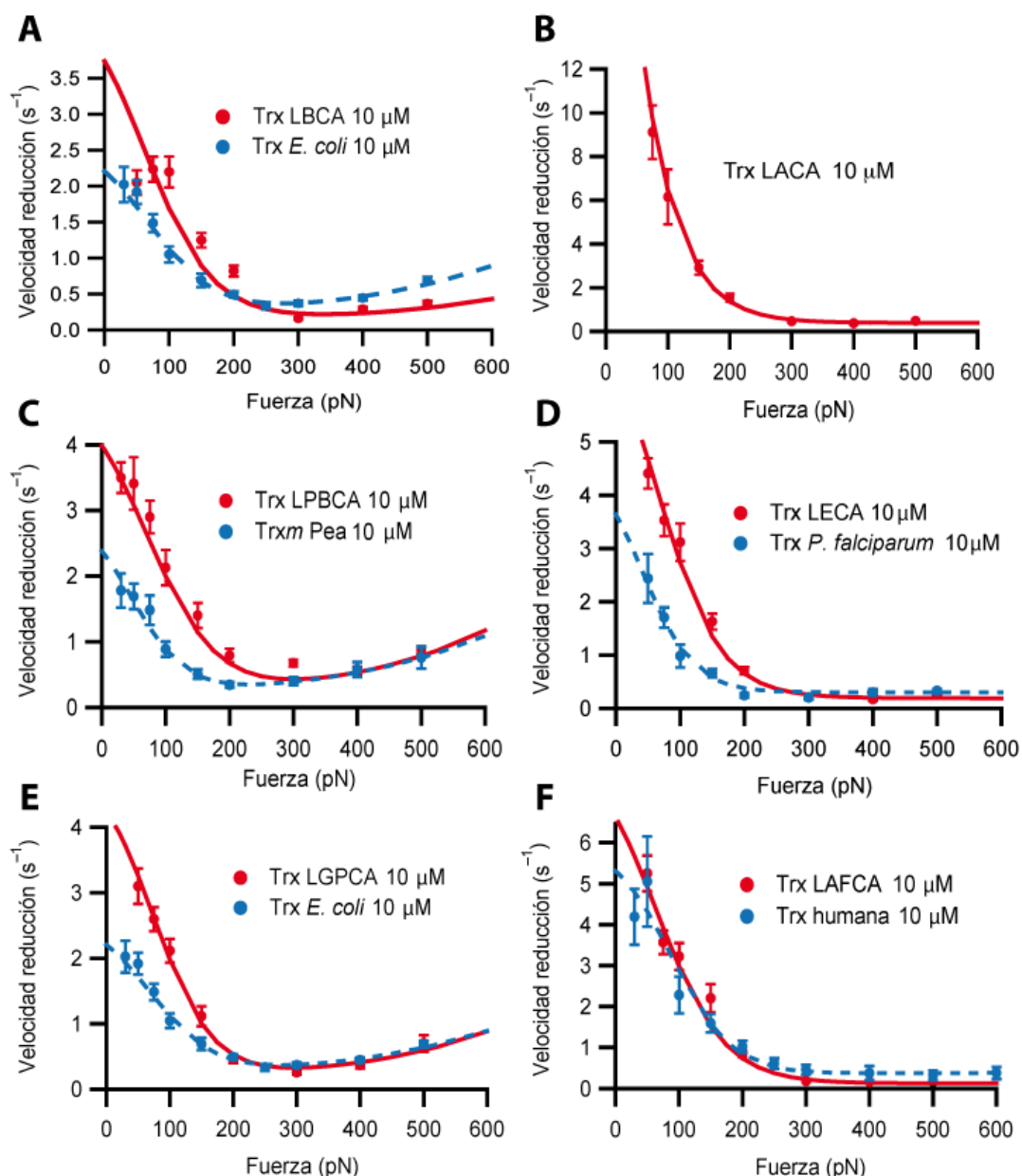


Figura 3.9. Dependencia de la velocidad de reducción de puentes disulfuro con la fuerza para las tiorredoxinas ancestrales. A-F: la velocidad de reducción a una fuerza dada se obtiene al sumar las trazas obtenidas (entre 15-80 trazas), promediarlas y ajustarlas a una exponencial simple, como se muestra en la figura 3.8 B. Las líneas continuas son los ajustes del modelo cinético de tres estados. Los círculos en gris y las líneas discontinuas representan la velocidad vs fuerza para tiorredoxinas actuales: *trxm* de cloroplasto de guisante (C), tiorredoxina de *P. falciparum* (D), tiorredoxina de *E. coli* (A,E) y la tiorredoxina humana (F) (Perez-Jimenez, Li et al. 2009). Estas enzimas actuales son descendientes de las ancestrales del mismo gráfico. Las barras de error representan el error estándar de la media obtenido usando el método de *bootstrapping*.

Cabría esperar que las tiorredoxinas primitivas tuvieran mecanismos químicos menos desarrollados. En este contexto, cabe destacar que uno de los principales factores que controla la química de la catálisis de tiorredoxina es la geometría del surco de unión (*groove*). En tiorredoxinas bacterianas actuales, el surco de unión es

menos pronunciado que en las tiorredoxinas de eucariota (Perez-Jimenez, Li et al. 2009). Esta diferencia estructural podría ser responsable de los diferentes comportamientos químicos observados en tiorredoxinas eucariotas y bacterianas. Parece razonable que las enzimas ancestrales puedan haber tenido un surco menos estructurado, haciendo que su química sea más similar a la de agentes reductores simples como la L-cisteína o el TCEP (Koti Ainarapu, Wiita et al. 2008). Sin embargo, la química de las tiorredoxinas parece haber sido establecida temprano en la evolución, hace 4 mil millones de años. Esta observación sugiere que el paso de simples compuestos reductores a enzimas bien estructuradas y funcionales ocurrió pronto en la evolución molecular (Nisbet y Sleep 2001).

No obstante, varias características de los mecanismos catalíticos de algunas tiorredoxinas ancestrales son dignas de mencionar: por ejemplo, la elevada actividad de las tiorredoxinas más antiguas, AECA y LACA, cuando el sustrato se estira a fuerzas por debajo de los 200 pN (Fig. 3.8 D y Fig. 3.9 B). A partir del ajuste del modelo cinético de tres estados a los datos de la velocidad de reducción frente a la fuerza, si extrapolamos a fuerza cero resultan constantes de velocidad de $30 \times 10^5 \text{ M}^{-1} \text{ s}^{-1}$ para la tiorredoxina de AECA y $29 \times 10^5 \text{ M}^{-1} \text{ s}^{-1}$ para la de LACA. La extrapolación a fuerza cero en el resto de las tiorredoxinas ancestrales predice constantes de velocidad desde $3.7 \times 10^5 \text{ M}^{-1} \text{ s}^{-1}$ hasta $6.6 \times 10^5 \text{ M}^{-1} \text{ s}^{-1}$ (Fig. 3.10). Aunque estos últimos valores están dentro del rango de los de las tiorredoxinas actuales empleando AFM (Fig. 3.9 y ref. (Perez-Jimenez, Li et al. 2009) y experimentos en disolución (Holmgren 1979), hay una tendencia en las enzimas resucitadas de mostrar mayores velocidades de reducción a fuerzas por debajo de los 200 pN (Fig. 3.9). Se especula que esta tendencia podría estar relacionada con la especificidad del sustrato a las enzimas. Las enzimas ancestrales podrían ser menos específicas al sustrato que las actuales y por lo tanto podrían ser más eficientes con sustratos genéricos como los usados en este trabajo.

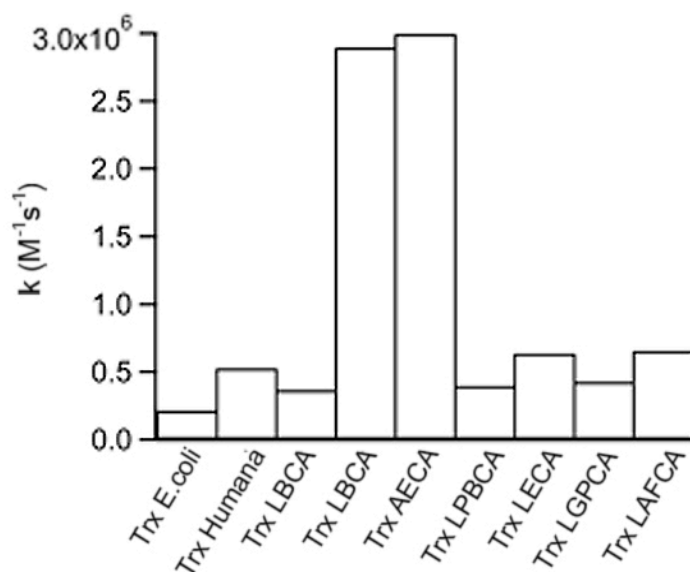


Figura 3.8. Constantes de velocidad para la reducción de puentes disulfuro por tiorredoxinas ancestrales y actuales. Estos valores se han obtenido extrapolando a fuerza cero el ajuste a los datos de la velocidad de reducción vs. fuerza (Fig. 3.7) al modelo cinético de tres estados.

Otra característica notable es la pequeña pendiente ascendente observada a bajas fuerzas para la tiorredoxina de LBCA, con un máximo a ~ 100 pN (Fig. 3.7 A). Una posible explicación es que la unión entre el sustrato y la enzima no sea óptima a fuerza cero, ya que parece que se consigue una mejor conformación aplicando una fuerza al sustrato.

Además, en nuestro grupo, hemos medido la actividad de las enzimas ancestrales usando el ensayo de agregación de insulina (Fig. 3.11) (apartado 2.2.5 de Materiales y Métodos). Los valores de las velocidades de precipitación de insulina obtenidos con este ensayo son similares a las determinadas previamente para la tiorredoxina de *E. coli* (Holmgren 1979; Shimizu, Yokobori et al. 2007).

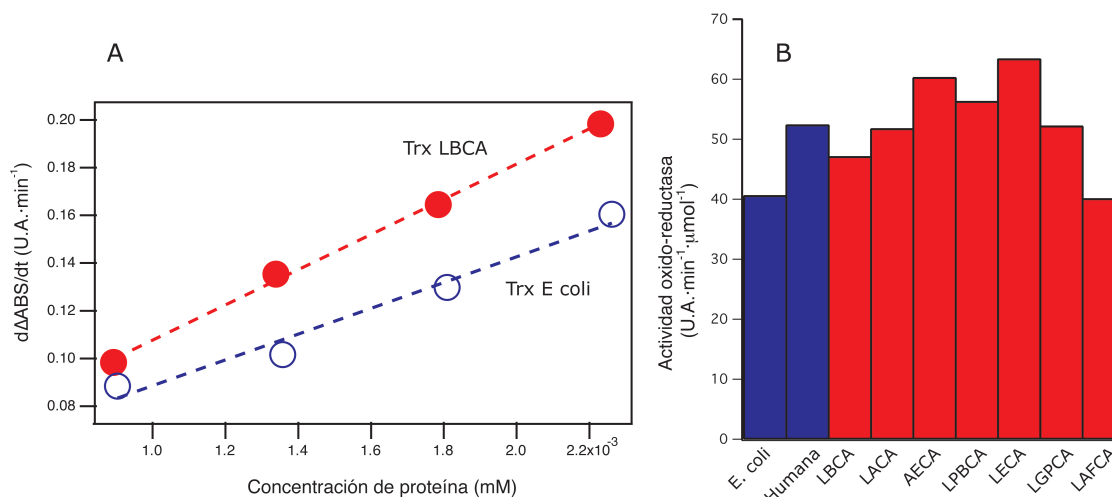


Figura 3.11. Actividad oxido-reductasa de las tiorredoxinas determinada mediante el ensayo enzimático en disolución de la turbidez de la insulina (Holmgren 1979). A: Representación del máximo de variación Abs/t en función de la concentración de tiorredoxina de *E. coli* (azul) y de la tiorredoxina precámbrica LBCA (rojo). Se ha realizado un ajuste lineal (línea discontinua), a partir del cual es posible determinar el valor de la actividad oxido-reductasa de tiorredoxina. B: Valores de actividad oxido-reductasa mediante el ensayo de precipitación de la insulina para tiorredoxinas ancestrales y actuales de *E.coli* y humana. El ensayo de turbidez es menos sensible que el AFM detectando diferencias en actividad entre las diferentes enzimas. Este ensayo no puede usarse para investigar la actividad de las enzimas a pH 5.0 debido a la precipitación de la insulina a pHs inferiores a 6.0 (Holmgren 1979; Holmgren 1979).

3.3.4. Actividad de las tiorredoxinas ancestrales en condiciones ácidas.

El grupo del Dr. Fernández ha comprobado la actividad de las tiorredoxinas ancestrales bajo condiciones ácidas. LBCA, AECA y LACA habitaron en un ambiente anaerobio, que era probablemente rico en compuestos de sulfuro y CO₂, mientras que LPBCA, LECA, LGPCA y LAFCA vivieron en un ambiente aerobio (Nisbet y Sleep 2001) (Fig. 3.2 A). Los altos niveles de CO₂ en el eón Hadeico eran en parte responsables de los bajos valores propuestos de pH de los océanos (~ 5.5) (Walker 1983; Russell y Hall 1997). Por lo tanto, siguiendo las hipótesis de que la vida primitiva vivía en el agua del mar, el hábitat natural en el que vivían LBCA, AECA y LACA es probable que fuese ácido, además de caliente. Esto es especialmente importante porque la reactividad de las tiorredoxinas es debida, en parte, a su bajo valor de pK_a de la cisteína catalítica: ~ 6.5 comparado con 8.0 para la cisteína libre (Holmgren 1985). Este bajo valor de pK_a es una consecuencia de las complejas interacciones electrostáticas entre varios residuos que estabilizan la forma deprotonada de la cisteína reactiva (Dyson, Jeng et al. 1997); por ello, la

actividad de tiorredoxina es muy sensible al pH. Las enzimas actuales no trabajan bien a pH ácido, debido a que el tiol catalítico podría protonarse e inactivarse.

El grupo del Dr. Fernández ha comparado la reactividad de las enzimas LACA, AECA y LBCA con las enzimas tiorredoxina actuales humana y de *E. coli* a pH 5.0. En estos experimentos han percibido claramente la capacidad de las enzimas resucitadas de operar en ambientes de pH ácido. Además, han medido la dependencia de la velocidad de reducción con la fuerza para LACA, AECA y LBCA a pH 5.0 en el rango de fuerzas de 50-150 pN (Fig. 3.12A).

Para la tiorredoxina de AECA, una extrapolación a fuerza cero da una constante de la velocidad de reducción de $19 \times 10^5 \text{ M}^{-1} \text{ s}^{-1}$. De manera similar, se ha estimado para LACA una constante de velocidad de $6.2 \times 10^5 \text{ M}^{-1} \text{ s}^{-1}$, mientras que para la tiorredoxina de LBCA las velocidades de reducción observadas a pH 5.0 eran similares a las obtenidas a pH 7.2 (Fig. 3.12 A). Estos elevados valores son similares a los medidos para algunas tiorredoxinas actuales a pH neutro (Perez-Jimenez, Li et al. 2009). También han comparado la velocidad de reducción medida a 100 pN y pH 5.0 para LACA, AECA y LBCA con las velocidades de algunas enzimas tiorredoxina actuales (Fig. 3.12 B). Las velocidades obtenidas para las enzimas ancestrales son mucho mayores que las de las tiorredoxinas de *E. coli* y humana.

Debido a la precipitación espontánea de la insulina a pHs por debajo de 6, se ha empleado DTNB como sustrato para la reducción de puentes disulfuro para confirmar la capacidad de las enzimas ancestrales de trabajar a pH 5.0 (Fig. 3.12 C). Este análisis de las enzimas reconstruidas indica que las tiorredoxinas ancestrales estaban bien adaptadas para funcionar bajo condiciones ácidas y que las tiorredoxinas podrían mantener constantes de velocidad de reducción similares cuando evolucionaron en ambientes más alcalinos.

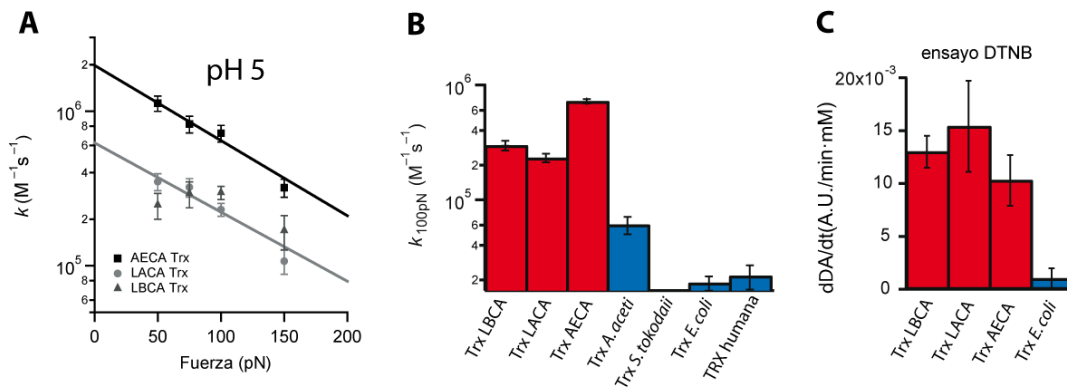


Figura 3.10. Velocidad de reducción de puentes disulfuro a pH 5.0. A: Alta actividad de las tioredoxinas de AECA (cuadrados) y LACA (círculos) a pH 5 cuando el sustrato es estirado a bajas fuerzas (50-150 pN). La tioredoxina de LBCA (triángulos) mostró una actividad similar a la obtenida a pH 7.2, con una tendencia similar (Fig. 3.9 A). Las líneas continuas son los ajustes exponenciales a los datos experimentales. B: Las constantes de velocidad para la reducción de puentes disulfuro por tioredoxinas ancestrales (rojas) a pH 5.0 son mayores que las obtenidas para las tioredoxinas actuales (azules) de *E. coli* y humana. La tioredoxina del acidófilo *A. acetii* muestra actividad a pH 5.0; la tioredoxina del termófilo *S. Tokodaii* no presenta actividad de reducción detectable al mismo pH. Todos los experimentos se realizaron a una fuerza de estiramiento de 100 pN. Las barras de error representan el error estándar de la media obtenido usando el método de *bootstrapping*. C: Actividad de tioredoxinas ancestrales (rojas) y de la actual (azul) de *E. coli* medidas usando DTNB como sustrato a pH 5.0 y determinada siguiendo espectrofotométricamente la formación de TNB a 412 nm. Las barras de error representan las desviaciones estándar para tres medidas diferentes.

Una característica de la familia de enzimas tioredoxina es que muchas de ellas son secretadas al ambiente extracelular donde se encuentran la mayoría de las proteínas con puentes disulfuro (Windle, Fox et al. 2000; Xu, Sukumar et al. 2008). Desde esta perspectiva, las tioredoxinas son quizás uno de los pocos tipos de enzimas para los que se puede establecer una correlación entre su sensibilidad al pH y las condiciones ambientales encontradas fuera de las células (Windle, Fox et al. 2000; Xu, Sukumar et al. 2008). La comparación de la tolerancia ácida de las enzimas resucitadas con la de enzimas pertenecientes a organismos extremófilos actuales puede reportarnos gran información. Por ejemplo, la tioredoxina de *Sulfolobus tokodaii* (arquea termofílica (Ming, Kato et al. 2007)), con una temperatura de desnaturalización de 122 °C, es activa a pH 7.0 ($0.12 \times 10^5 M^{-1} s^{-1}$ a 50 pN) pero no muestra actividad detectable a pH 5.0 (Fig. 3.10 B), lo cual no es sorprendente ya que *Sulfolobus* regula su pH citosólico (Baker-Austin y Dopson 2007). Por el contrario, la tioredoxina de *Acetobacter acetii* (bacteria acidofílica que crece a pH 4.0 (Starks, Francois et al. 2007)) es activa a pH 5.0 ($0.6 \times 10^5 M^{-1} s^{-1}$ a 100 pN), reflejando la acidez de su citosol (Menzel 1985; Starks, Francois et al. 2007).

3.4. Discusión

Los resultados del presente trabajo de investigación muestran que los mecanismos químicos observados en las tiorredoxinas actuales estaban presentes en las tiorredoxinas resucitadas de organismos del Precámbrico. Los datos obtenidos apoyan la idea de que las tiorredoxinas ancestrales pertenecientes a LBCA, AECA y LACA, que existieron entre mediados y finales del Eón Hadeico, eran muy resistentes a la temperatura y activas en condiciones relativamente ácidas. Los resultados obtenidos respaldan, además, la opinión de que hace 4 mil millones de años las tiorredoxinas estaban presentes en ambientes calientes y que estos ambientes se enfriaron progresivamente hasta hace 0.5 Ga (Gogarten-Boekels, Hilario et al. 1995; Nisbet y Sleep 2001; Knauth 2003; Gaucher, Govindarajan et al. 2008). Sin embargo, también es posible que una tierra primitiva más fría estuviese poblada por psicrófilos, mesófilos y termófilos y que éstos podrían haber sido los únicos supervivientes de los cataclismos (por ejemplo, las glaciaciones globales en la tierra primitiva (Gogarten-Boekels, Hilario et al. 1995; Nisbet y Sleep 2001). Los hallazgos obtenidos también apoyan la idea de que muchas rutas bioquímicas importantes en la biosfera actual se establecieron hace 3.5 mil millones de años (Nisbet y Sleep 2001). Importantes rutas como la producción de energía, degradación de azúcares, biosíntesis de cofactores y procesamiento de aminoácidos están muy conservados desde bacterias a humanos y, probablemente, estaban presentes en LUCA (Peregrin-Alvarez, Tsoka et al. 2003). La evolución opera claramente a múltiples niveles de organización biológica; sin embargo, los mecanismos enzimáticos que acompañan a los cambios de adaptación parecen estar muy conservados. La capacidad de las enzimas de mantener reactividad química y mecanismos específicos en ambientes dispares es necesaria para la diversificación de la vida. Como se muestra en este trabajo, esta capacidad está ejemplificada por las tiorredoxinas, pero también podría ser el caso de proteínas universales, tales como la ubiquitina, RNasa, ATPasa u otras enzimas metabólicas que se han manteniendo en casi todos los organismos a lo largo de la historia de la vida. Por lo tanto, la resurrección experimental de ancestros de estas proteínas universales junto con la sensibilidad de las técnicas de molécula individual podrían constituir una poderosa herramienta para entender el origen y la evolución de la vida en la Tierra.

3.5. Bibliografía

- Baker-Austin, C. and M. Dopson (2007). "Life in Acid: Ph Homeostasis in Acidophiles." Trends in microbiology **15**(4): 165-171.
- Benner, S. A., S. O. Sassi and E. A. Gaucher (2007). "Molecular Paleoscience: Systems Biology from the Past." Advances in enzymology and related areas of molecular biology **75**: 1-132, xi.
- Boussau, B., S. Blanquart, A. Necșulea, N. Lartillot and M. Gouy (2008). "Parallel Adaptations to High Temperatures in the Archaean Eon." Nature **456**(7224): 942-945.
- Chang, B. S., K. Jonsson, M. A. Kazmi, M. J. Donoghue and T. P. Sakmar (2002). "Recreating a Functional Ancestral Archosaur Visual Pigment." Molecular biology and evolution **19**(9): 1483-1489.
- Dyson, H. J., M. F. Jeng, L. L. Tennant, I. Slaby, M. Lindell, D. S. Cui, S. Kuprin and A. Holmgren (1997). "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 **36**(9): 2622-2636.
- Edgar, R. C. (2004). "Muscle: Multiple Sequence Alignment with High Accuracy and High Throughput." Nucleic acids research **32**(5): 1792-1797.
- Fernandez, J. M. and H. Li (2004). "Force-Clamp Spectroscopy Monitors the Folding Trajectory of a Single Protein." Science **303**(5664): 1674-1678.
- Florin, E. L., M. Rief, H. Lehmann, M. Ludwig, C. Dornmair, V. T. Moy and H. E. Gaub (1995). "Sensing Specific Molecular Interactions with the Atomic Force Microscope." Biosensors and Bioelectronics **10**(9-10): 895-901.
- Garcia-Manyes, S., J. Liang, R. Szoszkiewicz, T. L. Kuo and J. M. Fernandez (2009). "Force-Activated Reactivity Switch in a Bimolecular Chemical Reaction." Nature chemistry **1**(3): 236-242.
- Gaucher, E. A., S. Govindarajan and O. K. Ganesh (2008). "Palaeotemperature Trend for Precambrian Life Inferred from Resurrected Proteins." Nature **451**(7179): 704-707.
- Gaucher, E. A., J. M. Thomson, M. F. Burgan and S. A. Benner (2003). "Inferring the Palaeoenvironment of Ancient Bacteria on the Basis of Resurrected Proteins." Nature **425**(6955): 285-288.
- Gogarten-Boekels, M., E. Hilario and J. P. Gogarten (1995). "The Effects of Heavy Meteorite Bombardment on the Early Evolution--the Emergence of the Three Domains of Life." Origins of life and evolution of the biosphere : the journal of the International Society for the Study of the Origin of Life **25**(1-3): 251-264.
- Hall, B. G. (2006). "Simple and Accurate Estimation of Ancestral Protein Sequences." Proceedings of the National Academy of Sciences of the United States of America **103**(14): 5431-5436.

- Hedges, S. B. K., S. (2009). The Timetree of Life Oxford Univ. Press, Oxford.
- Holmgren, A. (1979). "Reduction of Disulfides by Thioredoxin. Exceptional Reactivity of Insulin and Suggested Functions of Thioredoxin in Mechanism of Hormone Action." The Journal of biological chemistry **254**(18): 9113-9119.
- Holmgren, A. (1979). "Thioredoxin Catalyzes the Reduction of Insulin Disulfides by Dithiothreitol and Dihydrolipoamide." The Journal of biological chemistry **254**(19): 9627-9632.
- Holmgren, A. (1985). "Thioredoxin." Annual review of biochemistry **54**: 237-271.
- Knauth, L. P. L., D.R. (2003). "High Archean Climatic Temperature Inferred from Oxygen Isotope Geochemistry of Cherts in the 3.5 Ga Swaziland Supergroup, South Africa. ." Geol. Soc. Am. Bull. **115**: 566-580.
- Koti Ainavarapu, S. R., A. P. Wiita, L. Dougan, E. Uggerud and J. M. Fernandez (2008). "Single-Molecule Force Spectroscopy Measurements of Bond Elongation During a Bimolecular Reaction." Journal of the American Chemical Society **130**(20): 6479-6487.
- Liang, J. and J. M. Fernandez (2009). "Mechanochemistry: One Bond at a Time." ACS nano **3**(7): 1628-1645.
- Liberles, D. A. (2007). Ancestral Sequence Reconstruction Oxford Univ. Press, Oxford, New York.
- Menzel, U. G., G. (1985). "The Internal Ph of *Acetobacterium Wieringae* and *Acetobacter Aceti* During Growth and Production of Acetic Acid." Arch. Microbiol. **143**: 47-51.
- Ming, H., Y. Kato, K. Miyazono, K. Ito, M. Kamo, K. Nagata and M. Tanokura (2007). "Crystal Structure of Thioredoxin Domain of St2123 from Thermophilic Archaea *Sulfolobus Tokodaii* Strain7." Proteins **69**(1): 204-208.
- Nelder, J. A. M., R.A. (1965). "Simplex Method for Function Minimization." Comput. J. **7**: 308-313.
- Nisbet, E. G. and N. H. Sleep (2001). "The Habitat and Nature of Early Life." Nature **409**(6823): 1083-1091.
- Peregrin-Alvarez, J. M., S. Tsoka and C. A. Ouzounis (2003). "The Phylogenetic Extent of Metabolic Enzymes and Pathways." Genome research **13**(3): 422-427.
- Perez-Jimenez, R., J. Li, P. Kosuri, I. Sanchez-Romero, A. P. Wiita, D. Rodriguez-Larrea, A. Chueca, A. Holmgren, A. Miranda-Vizueté, K. Becker, S. H. Cho, J. Beckwith, E. Gelhaye, J. P. Jacquot, E. A. Gaucher, J. M. Sanchez-Ruiz, B. J. Berne and J. M. Fernandez (2009). "Diversity of Chemical Mechanisms in Thioredoxin Catalysis Revealed by Single-Molecule Force Spectroscopy." Nature structural & molecular biology **16**(8): 890-896.

- Perez-Jimenez, R., A. P. Wiita, D. Rodriguez-Larrea, P. Kosuri, J. A. Gavira, J. M. Sanchez-Ruiz and J. M. Fernandez (2008). "Force-Clamp Spectroscopy Detects Residue Co-Evolution in Enzyme Catalysis." The Journal of biological chemistry **283**(40): 27121-27129.
- Pollock, D. D. C., B.S.W. (2007). Ancestral Sequence Reconstruction Oxford Univ. Press, Oxford, New York.
- Russell, M. J. and A. J. Hall (1997). "The Emergence of Life from Iron Monosulphide Bubbles at a Submarine Hydrothermal Redox and Ph Front." Journal of the Geological Society **154**(3): 377-402.
- Shimizu, H., S. Yokobori, T. Ohkuri, T. Yokogawa, K. Nishikawa and A. Yamagishi (2007). "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 **369**(4): 1060-1069.
- Starks, C. M., J. A. Francois, K. M. MacArthur, B. Z. Heard and T. J. Kappock (2007). "Atomic-Resolution Crystal Structure of Thioredoxin from the Acidophilic Bacterium Acetobacter Aceti." Protein science : a publication of the Protein Society **16**(1): 92-98.
- Thomson, J. M., E. A. Gaucher, M. F. Burgan, D. W. De Kee, T. Li, J. P. Aris and S. A. Benner (2005). "Resurrecting Ancestral Alcohol Dehydrogenases from Yeast." Nature genetics **37**(6): 630-635.
- Thornton, J. W. (2004). "Resurrecting Ancient Genes: Experimental Analysis of Extinct Molecules." Nature reviews. Genetics **5**(5): 366-375.
- Thornton, J. W., E. Need and D. Crews (2003). "Resurrecting the Ancestral Steroid Receptor: Ancient Origin of Estrogen Signaling." Science **301**(5640): 1714-1717.
- Walker, J. C. G. (1983). "Possible Limits on the Composition of the Archean Ocean." Nature **302**: 518-520.
- Wiita, A. P., S. R. Ainarapu, H. H. Huang and J. M. Fernandez (2006). "Force-Dependent Chemical Kinetics of Disulfide Bond Reduction Observed with Single-Molecule Techniques." Proceedings of the National Academy of Sciences of the United States of America **103**(19): 7222-7227.
- Wiita, A. P., R. Perez-Jimenez, K. A. Walther, F. Grater, B. J. Berne, A. Holmgren, J. M. Sanchez-Ruiz and J. M. Fernandez (2007). "Probing the Chemistry of Thioredoxin Catalysis with Force." Nature **450**(7166): 124-127.
- Windle, H. J., A. Fox, D. Ni Eidhin and D. Kelleher (2000). "The Thioredoxin System of Helicobacter Pylori." The Journal of biological chemistry **275**(7): 5081-5089.
- Xu, S. Z., P. Sukumar, F. Zeng, J. Li, A. Jairaman, A. English, J. Naylor, C. Ciurtin, Y. Majeed, C. J. Milligan, Y. M. Bahnasi, E. Al-Shawaf, K. E. Porter, L. H. Jiang, P. Emery, A. Sivaprasadarao and D. J. Beech (2008). "Trpc Channel Activation by Extracellular Thioredoxin." Nature **451**(7174): 69-72.

- Yang, Z. (1997). "Paml: A Program Package for Phylogenetic Analysis by Maximum Likelihood." Computer applications in the biosciences **13**(5): 555-556.
- Yang, Z., S. Kumar and M. Nei (1995). "A New Method of Inference of Ancestral Nucleotide and Amino Acid Sequences." Genetics **141**(4): 1641-1650.
- Zalatan, J. G. and D. Herschlag (2009). "The Far Reaches of Enzymology." Nature chemical biology **5**(8): 516-520.

Single-molecule paleoenzymology probes the chemistry of resurrected enzymes

Raul Perez-Jimenez¹, Alvaro Inglés-Prieto², Zi-Ming Zhao³, Inmaculada Sanchez-Romero², Jorge Alegre-Cebollada¹, Pallav Kosuri^{1,4}, Sergi Garcia-Manyes¹, T Joseph Kappock⁵, Masaru Tanokura⁶, Arne Holmgren⁷, Jose M Sanchez-Ruiz², Eric A Gaucher^{3,8,9} & Julio M Fernandez¹

It is possible to travel back in time at the molecular level by reconstructing proteins from extinct organisms. Here we report the reconstruction, based on sequence predicted by phylogenetic analysis, of seven Precambrian thioredoxin enzymes (Trx) dating back between ~1.4 and ~4 billion years (Gyr). The reconstructed enzymes are up to 32 °C more stable than modern enzymes, and the oldest show markedly higher activity than extant ones at pH 5. We probed the mechanisms of reduction of these enzymes using single-molecule force spectroscopy. From the force dependency of the rate of reduction of an engineered substrate, we conclude that ancient Trxs use chemical mechanisms of reduction similar to those of modern enzymes. Although Trx enzymes have maintained their reductase chemistry unchanged, they have adapted over 4 Gyr to the changes in temperature and ocean acidity that characterize the evolution of the global environment from ancient to modern Earth.

Experimental paleogenetics and paleobiochemistry provide an opportunity to investigate in the laboratory the molecular history of modern organisms^{1–4}. The study of resurrected proteins can also reveal valuable information about the adaptation of extinct forms of life to climatic, ecological and physiological alterations^{5–8}. Despite numerous experimental examples of protein resurrection in the laboratory¹, most of the resurrected proteins have reconstructed enzymes thought to exist a few million years (Myr) ago^{1,2,9}. Consequently, many hypotheses about ancient life remain untested, especially for time periods associated with rapid changes in biological systems¹⁰. Such time traveling is largely limited by ambiguity in the historical models used for inferring ancestral sequences. For instance, uncertainties in database and sequence alignments, failures in evolutionary theories and uncertainty in the construction of phylogenetic trees are common sources of ambiguity¹. Nonetheless, several approaches are commonly used to overcome these limitations, and some examples of reconstructed proteins from the Precambrian (4.5–0.5 Gyr ago) have been reported^{1,11}. One such instance is the resurrection of an elongation factor gene family from ancient bacteria 0.5–3.5 Gyr old^{9,12}. The ancestral elongation factors showed that ancient bacteria lived in a hot environment and also revealed a correlation between the thermostability of ancient life and the temperature of ancient oceans as inferred from geological records.

These reconstruction and resurrection studies have paved the way for researchers to formulate interesting questions about ancient

organisms and the biomolecules that supported them. For instance, little is known about how the chemistry of primitive enzymes arose and how the environmental conditions affected the evolution of their chemistry¹³. These questions cannot be answered directly by examining fossil records. In an effort to understand the evolution of enzymatic reactions, we have reconstructed and tested Trx enzymes from extinct organisms that lived in the Precambrian.

Thioredoxins belong to a broad family of oxidoreductase enzymes that are ubiquitous in all living organisms¹⁴. The archetypical active site (CXXC) and the Trx fold have been conserved throughout evolution, indicating that Trx enzymes were probably present in primitive forms of life. Single-molecule force-clamp spectroscopy has been used to examine in detail the chemical mechanisms of disulfide reduction by Trx enzymes at the sub-Ångström scale^{15,16}. The combination of single-molecule force spectroscopy and the resurrection of ancestral proteins might reveal new insights into the reductase activity of these sulfur-based enzymes.

RESULTS

Reconstruction of ancestral Trx enzymes

Using the large number of extant Trx sequences available, we have constructed a highly articulated phylogenetic tree encompassing more than 200 diverse Trx sequences from the three domains of life (**Supplementary Fig. 1** and **Supplementary Note** for GI numbers).

¹Department of Biological Sciences, Columbia University, New York, New York, USA. ²Facultad de Ciencias, Departamento de Química-Física, Universidad de Granada, Granada, Spain. ³Georgia Institute of Technology, School of Biology, Atlanta, Georgia, USA. ⁴Department of Biochemistry and Molecular Biophysics, Columbia University, New York, New York, USA. ⁵Department of Biochemistry, Purdue University, West Lafayette, Indiana, USA. ⁶Department of Applied Biological Chemistry, Graduate School of Agricultural and Life Sciences, The University of Tokyo, Tokyo, Japan. ⁷Division of Biochemistry, Department of Medical Biochemistry and Biophysics, Karolinska Institutet, Stockholm, Sweden. ⁸Georgia Institute of Technology, School of Chemistry, Atlanta, Georgia, USA. ⁹Georgia Institute of Technology, Parker H. Petit Institute for Bioengineering and Bioscience, Atlanta, Georgia, USA. Correspondence should be addressed to R.P.-J. (raulpjc@biology.columbia.edu) or J.M.F. (jfernandez@columbia.edu).

Received 9 August 2010; accepted 24 January 2011; published online 3 April 2011; doi:10.1038/nsmb.2020

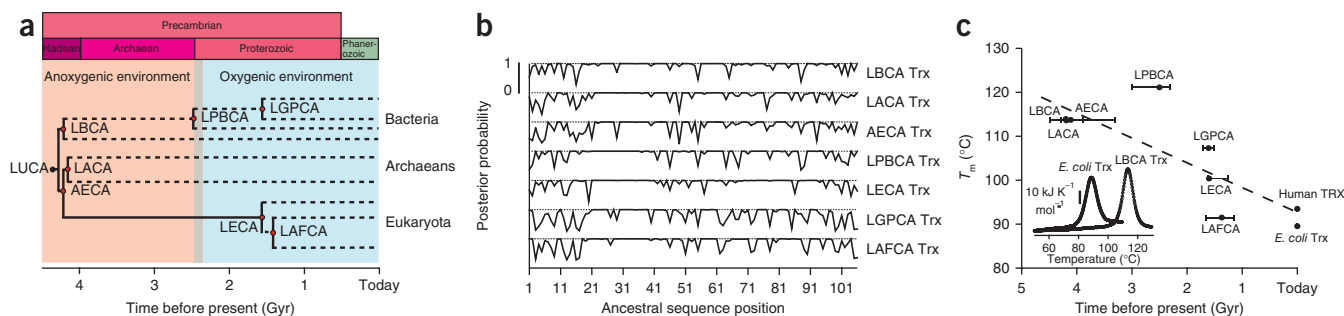


Figure 1 Phylogenetic analysis of Trx enzymes and ancestral sequence reconstruction. (a) Schematic phylogenetic tree showing the geological time in which different extinct organisms lived (see text). Dashed lines represent further bifurcations. Divergence times are compiled from multiple sources and are summarized in ref. 17. The figure indicates the global environment, although both aerobic and anaerobic organisms are found in modern environments. (b) Posterior probability distribution of the inferred amino acids across 106 sites for the interested internal nodes. The inferred amino acid at each site for the interested internal node is the residue with the highest posterior probability. (c) Denaturation temperatures (T_m) versus geological time for ancestral Trx enzymes. Modern *E. coli* and human Trx enzymes are also indicated. The dashed line represents a lineal fit to the data. Inset, experimental DSC thermograms for *E. coli* Trx and LBCA Trx. The instrumental uncertainty of DSC measurements is <0.5 °C.

From this tree we sampled several biologically relevant nodes for sequence reconstruction and laboratory resurrection. We applied estimates of divergence dates to nodes in the tree assuming that the root of the tree lies between bacteria and the common ancestor of archaea and eukaryotes¹⁷.

We resurrected Trx enzymes belonging to the last bacterial common ancestor (LBCA), the last archaeal common ancestor (LACA) and the archaeal-eukaryotic common ancestor (AECA; **Fig. 1**). These organisms are thought to have inhabited Earth 4.2–3.5 Gyr ago (**Fig. 1a**) after diverging from the last universal common ancestor (LUCA)^{7,17}. We also selected a node corresponding to the last eukaryotic common ancestor (LECA) that lived in the Proterozoic, about 1.60 Gyr ago. We selected two other internal nodes in the bacterial lineages; the last common ancestor of the cyanobacterial and deinococcus and thermus groups (LPBCA), which existed around 2.5 Gyr ago and represents the origin of photosynthetic bacteria, and the last common ancestor of γ -proteobacteria (LGPCA), which is around 1.61 Gyr old. Finally, we also chose the last common ancestor of animals and fungi (LAFCA), which lived around 1.37 Gyr ago (**Fig. 1a**). Although it would be interesting to observe the chemical transition between LUCA and the last common ancestors of bacteria, and archaea and eukaryotes, the reconstruction of Trx from LUCA cannot be inferred without

making far-reaching assumptions about ancient life. For instance, the last common ancestor of bacteria represents an ancient population, and the last common ancestor of archaea and eukaryotes represents another ancient population. It is practical to infer that LUCA lies somewhere along the branch that connects these two ancient populations. However, we cannot identify the position of LUCA (the ancestral node) without an outgroup. We have reason to believe that LUCA existed on this branch (on the basis of ancient gene duplication events that gave rise to paralogs still present in all modern organisms), but we do not know where on the branch LUCA existed. This prevents us from inferring the sequence of LUCA (*pace* midpoint rooting).

We reconstructed the sequences of the ancestral Trx enzymes using statistical methods based on maximum likelihood^{3,9}. For a given node in the tree, we calculated posterior probability values for all 20 amino acids considering each site of the inferred sequence. These values represent the probability that a certain residue occupied a specific position in the sequence at a particular point in the phylogeny. The posterior probabilities were calculated on the basis of an amino acid replacement matrix¹⁸. We then reconstructed the most probabilistic ancestral sequence (M-PAS) at a specific node by assigning to each site the residue with the highest posterior probability. We determined the posterior probability distribution of the inferred amino acids across 106 sites for the selected sequences (**Fig. 1b**). The M-PASs of interest are summarized in **Supplementary Figure 2**. We synthesized the genes that encode these sequences, expressed the proteins from *Escherichia coli* cells and purified them.

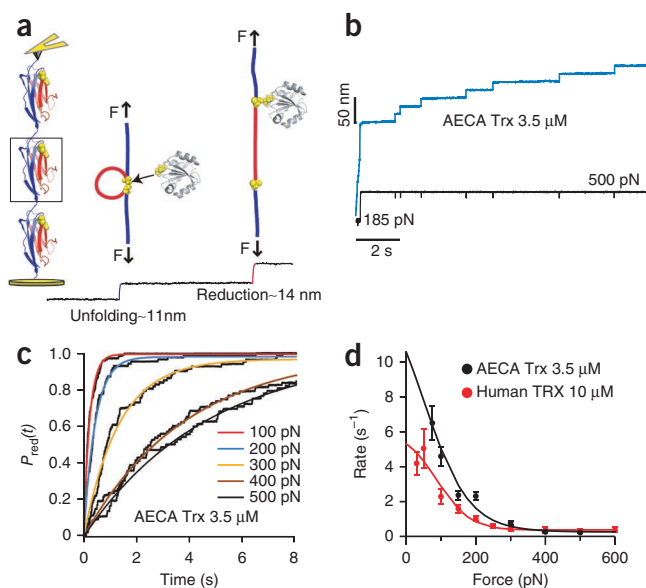


Figure 2 Single-molecule disulfide reduction assay. (a) Schematic representation of the single-molecule disulfide reduction assay. A first pulse of force rapidly unfolds the I27_{G32C-A75C} domains (11-nm step). When the disulfide bond is exposed to the solvent a single Trx molecule can reduce it (14-nm step). (b) Experimental force-clamp trace showing single disulfide reductions of a (I27_{G32C-A75C})₈ polypeptide. The unfolding pulse was set at 185 pN for 0.2 s and the test-pulse force at 500 pN. (c) Probability of reduction ($P_{red}(t)$) resulted from summing and normalizing the reduction test pulse at different forces for AECA Trx (3.5 μ M). (d) Force dependency of disulfide reduction by AECA Trx; human TRX is also shown for comparison. Both Trx enzymes show a similar pattern: a negative force dependency of the reduction rate, from 30–200 pN, consistent with a Michaelis-Menten mechanism and a force-independent mechanism, from 200 pN and up, described by an electron transfer reaction¹⁶. Notice the higher activity for AECA Trx (3.5 μ M for AECA Trx versus 10 μ M for human TRX). The lines represent fittings to the kinetic model (see Online Methods). The error bars are given by the s.e.m. obtained using the bootstrap method.

Figure 3 Force dependence of disulfide reduction by ancestral Trx enzymes. (a–f) The reduction rate at a given force is obtained by summing, averaging and fitting to single exponential numerous traces (15–80) like the one shown in **Figure 2b**. The solid lines are fitting to the kinetic model. The gray circles and dashed lines represent the rate versus force dependence for modern Trxs: Pea Trxm from chloroplast (c), *P. falciparum* Trx (d), *E. coli* Trx (a,e) and human TRX (f; all extracted from ref. 16). These modern Trxs are descendants of the ancestral Trxs in the same plot. The error bars represent the s.e.m. obtained using the bootstrap method.

Thermal stability of ancient Trx enzymes

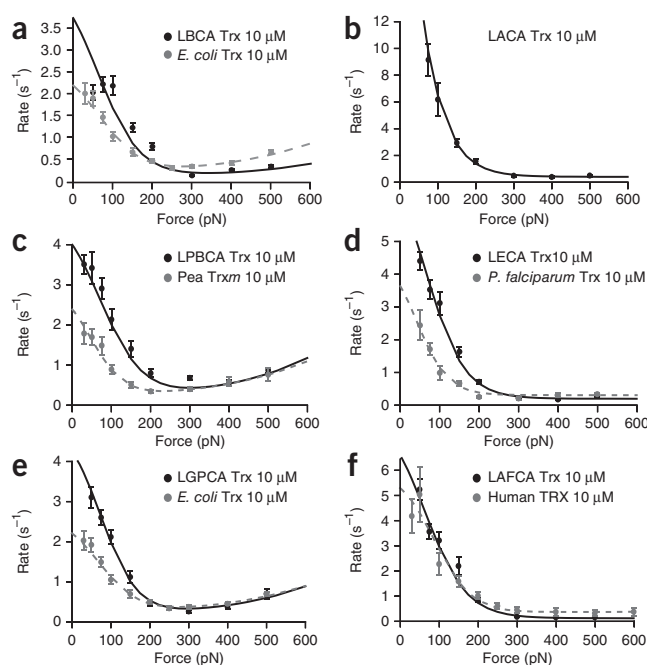
As a first step toward investigating the physicochemical properties of these resurrected enzymes, we used differential scanning calorimetry (DSC) to measure their thermal stabilities. The denaturation temperature (T_m) can provide an idea of the temperature range in which the proteins operate. A plot of the T_m of the resurrected enzymes against geological time is shown in **Figure 1c**. The three oldest Trx enzymes from LBCA, AECA and LACA showed a similar T_m of around 113 °C. LBCA Trx maintained a highly populated native state up to ~105 °C, where the thermal transition began (**Fig. 1c**, inset). By contrast, we determined T_m for modern *E. coli* and human Trxs of 88.8 and 93.3 °C, respectively. The ΔT_m between the oldest and modern Trx is around 25 °C, a value similar to that determined for bacterial elongation factor¹², which corroborates the hypothesis of the thermophilic nature of LBCA, AECA and LACA⁷. From the data in **Figure 1c**, we determined a paleotemperature trend yielding a decrease in the T_m of 5.8 ± 1.8 K Gyr⁻¹.

Although the thermodynamic denaturation temperatures determined for the ancestral Trxs follow a cooling trend similar to that of ancient oceans, the values are about 50 °C higher than the ocean temperatures inferred from maximum $\delta^{18}\text{O}$ (ref. 12). It is possible that Trx evolution operated primarily on kinetic stability, and this could be reflected in thermodynamic stability¹⁹. However, other than the obvious loss of function upon denaturation, it is unclear exactly how the value of T_m is related to Trx enzyme fitness.

Force-dependent chemical kinetics of disulfide reduction

We investigated the chemical mechanisms of disulfide bond reduction used by the resurrected enzymes. Given the ancient origin of the reconstructed Trx enzymes, with some of them predating the buildup of atmospheric oxygen, it seemed logical to assume that their catalytic chemistry might be closer to that of simple sulfur-based molecules that use a straightforward collision-driven substitution nucleophilic bimolecular (S_N2) mechanism of reduction²⁰. By contrast, Trx enzymes use a complex mixture of chemical mechanisms, including a crucial substrate binding and rearrangement reaction that accounts for the vast increase in the efficiency of Trx over the simpler sulfur compounds that were available in early geochemistry^{15,16}.

Recently, we developed an assay based on single-molecule force spectroscopy that measures the effect of applying a well-controlled force to a disulfide-bonded substrate on its rate of reduction by a nucleophile. This assay readily distinguishes the simple S_N2 chemistry of nucleophiles such as hydroxide, glutathione and L-cysteine from the more complex reduction chemistry of the Trx enzymes^{15,16,21–24}, making it ideal for probing the chemistry of the resurrected enzymes (**Fig. 2**). As a substrate, we use an engineered polyprotein made of eight repeats of the I27 immunoglobulin-like protein modified by mutating amino acids 32 and 75 to cysteine (I27_{G32C-A75C})₈. The cysteines oxidize spontaneously, forming disulfide bonds that are hidden in each folded I27 domain in the chain. We pick up and stretch single polyproteins using an atomic force microscope (AFM) in solutions containing the desired nucleophile. In a typical experiment,



we first apply a constant force to the polyprotein (175–185 pN, 0.2–0.3 s), and this rapidly unfolds the I27_{G32C-A75C} modules up to the disulfide bond. The unfolding events result in a stepwise increase in the length of the polyprotein where each module contributes with ~11 nm in length (**Fig. 2a** and **Supplementary Fig. 3**). After unfolding occurs, every disulfide bond is exposed to the solvent. If active Trx enzymes are present in the solution, single reduction events of ~14 nm per module can be readily observed (**Fig. 2a,b** and **Supplementary Figs. 3,4**). All the ancestral enzymes that we resurrected could trigger staircases of reduction events (**Fig. 2b** and **Supplementary Figs. 3,4**), indicating that they were all active. To measure the reduction rate, we summed 15–80 reduction staircases similar to the one shown in **Figure 2b** and fit the resulting average with a single exponential. We repeated this procedure for different pulling forces (**Fig. 2c**). The resulting set of data measured the force dependency of the rate of reduction of the disulfide bond (**Fig. 2d**).

The chemical mechanisms of disulfide reduction can be distinguished by their sensitivity to the force applied to the substrate¹⁶. Simple thiol-reducing agents show a force dependency in which the rate always increased exponentially with the applied force^{21,22}. By contrast, modern Trx enzymes show a negative force dependency in the range of 30–200 pN (ref. 16). This mechanism is consistent with a Michaelis-Menten binding reaction followed by a force-inhibited reorientation of the substrate disulfide bond that is necessary for an S_N2 reaction to occur¹⁵. In a second mechanism, the rate of reduction increases exponentially at forces above 200 pN. This mechanism can be described by a simple S_N2 reaction and is found only in Trx enzymes of bacterial origin. In all Trx enzymes there is a force-independent mechanism of reduction that can be ascribed to a single electron transfer reaction¹⁶.

The same three reduction mechanisms can be observed in the ancient enzymes with similar patterns to those found in extant Trxs (**Figs. 2d** and **3**). We could easily fit the force dependency of the reduction rate measured from the resurrected enzymes using the three-state kinetic model that was previously used with modern Trxs^{15,16} (see Online Methods and **Supplementary Table 1**). One might expect that Trx enzymes from primitive forms of life should have less developed chemical mechanisms. For instance, one of the main factors that controls the chemistry of Trx catalysis is the geometry of the

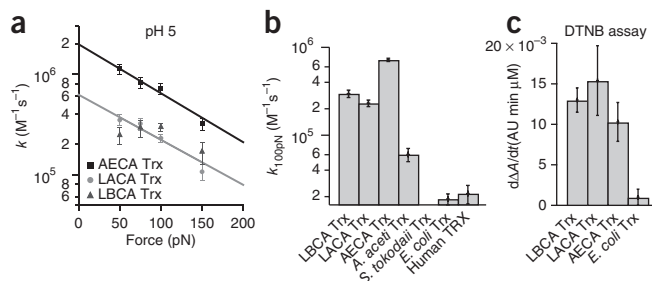


Figure 4 Rate constants of disulfide bond reduction at pH 5. **(a)** High activity for AECA (squares) and LACA (circles) Trxs at pH 5 when the substrate was pulled at low forces (50–150 pN). LBCA Trx (triangles) showed similar activity to that at pH 7.2 with a similar trend (**Fig. 3a**). The solid lines are exponential fit to the experimental data. **(b)** Rate constants for disulfide reduction by ancestral Trx enzymes at pH 5 are higher than for modern *E. coli* and human Trx. Thioredoxin from the acidophile *A. acetii* shows activity at pH 5; enzymes from the thermophilic *S. tokodaii* do not show a detectable rate of reduction at the same pH. All experiments were conducted at a pulling force of 100 pN. Error bars represent s.e.m. obtained using the bootstrap method. **(c)** Activity of ancestral Trxs and modern *E. coli* Trx measured using DTNB as substrate at pH 5 and determined by monitoring spectrophotometrically the formation of TNB at 412 nm. Error bars represent s.d. from three different measurements.

binding groove. In modern Trxs from bacteria, the binding groove is less pronounced than in eukaryotic Trxs¹⁶. This structural difference is responsible for the different chemical behavior observed in eukaryotic and bacterial Trxs. It seems reasonable that ancient enzymes would have had a less structured groove, making their chemistry more similar to that of simple reducing agents such as L-cysteine or TCEP²². However, the chemistry of Trx enzymes seems to have been established early in evolution, about 4 Gyr ago, in the same manner as is observed today. This observation suggests that the step from simple reducing compounds to well-structured and functional enzymes occurred early in molecular evolution¹⁰.

Nonetheless, several features of the catalytic mechanisms of some ancestral Trxs are noteworthy: for instance, the high activity of AECA and LACA Trxs when the substrate was pulled at forces below 200 pN (**Figs. 2d** and **3b**). From the fitting of the reduction rate versus force data to the three-state kinetic model, an extrapolation to zero force yielded rate constants of $30 \times 10^5 \text{ M}^{-1} \text{ s}^{-1}$ for AECA Trx and $29 \times 10^5 \text{ M}^{-1} \text{ s}^{-1}$ for LACA Trx. The extrapolation to zero force in the rest of the ancestral Trxs predicted rate constants from $3.7 \times 10^5 \text{ M}^{-1} \text{ s}^{-1}$ to $6.6 \times 10^5 \text{ M}^{-1} \text{ s}^{-1}$ (**Supplementary Fig. 5**). Although these latter values are within the range of those found in extant Trx enzymes using AFM (**Fig. 3** and ref. 16) and bulk experiments²⁵, there was a trend in the reconstructed enzymes to show higher reduction rates at forces below 200 pN (**Fig. 3**). We speculate that this trend might be related to the substrate specificity of the enzymes. Ancient enzymes may be less substrate specific than modern ones and therefore might be more efficient with generic substrates such as those used here.

Another notable feature is the small upward slope observed at low forces for LBCA Trx with a maximum at ~ 100 pN (**Fig. 3a**). Although structural information would be needed to fully address this point, it seems possible that the binding between substrate and enzyme is not optimum at zero force. A better conformation appears to be achieved by applying force to the substrate.

We also measured the activity of the ancestral enzymes using the conventional insulin assay (**Supplementary Fig. 6**). The values of insulin precipitation rates obtained with this assay are similar to those previously determined for *E. coli* Trx^{26,27}.

Activity of ancestral Trxs under acidic conditions

LBCA, AECA and LACA thrived in an anoxygenic environment that was probably rich in sulfur compounds and CO_2 , whereas LPBCA, LECA, LGPCA and LAFCA lived in a global oxygenic environment¹⁰ (**Fig. 1a**). The high level of CO_2 in the Hadean was partly responsible for the proposed low pH of the ancient oceans (~ 5.5)^{28,29}. Therefore, following the hypothesis that early life lived in seawater, the natural habitat in which LBCA, AECA and LACA lived was likely to have been acidic as well as hot. This is especially important because the reactivity of Trx enzymes is due, in part, to the low pK_a value of the reactive cysteine: ~ 6.5 compared with 8.0 for free cysteine¹⁴. This low pK_a value is a consequence of complex electrostatic interactions between several residues that stabilize the deprotonated form of the reactive cysteine³⁰. Thus, Trx activity is highly sensitive to pH, and modern enzymes would not work well at low pH because the catalytic thiol would be protonated and inactive.

We compared the reactivity of LACA, AECA and LBCA enzymes with the extant human and *E. coli* Trx enzymes at pH 5. The ability of the resurrected enzymes to operate in low pH environments was apparent from these experiments. We measured the force dependency of reduction for AECA, LACA and LBCA at pH 5 over the 50–150 pN force range (**Fig. 4a**). For AECA Trx, an extrapolation to zero force gave a reduction rate constant of $19 \times 10^5 \text{ M}^{-1} \text{ s}^{-1}$ (**Fig. 4a**, solid line); similarly, for LACA we estimated a rate constant of $6.2 \times 10^5 \text{ M}^{-1} \text{ s}^{-1}$, whereas for LBCA Trx the reduction rates observed at pH 5 were similar to those measured at pH 7.2 (**Fig. 4a**). These are high values similar to those measured for some modern Trx enzymes at neutral pH (ref. 16). We also compared the rate of reduction measured at 100 pN for LBCA, LACA and AECA with the rates of some modern Trx enzymes also measured at pH 5 (**Fig. 4b**).

Owing to spontaneous precipitation of insulin at $\text{pH} < 6$, we used DTNB as a substrate for disulfide reduction to verify the ability of the oldest enzymes to work at pH 5 (**Fig. 4c**). This analysis of reconstructed enzymes indicated that that ancient Trx enzymes were well adapted to function under acidic conditions and that Trx enzymes could maintain similar reduction rate constants as they evolved in more alkaline environments. A feature of the thioredoxin family of enzymes is that many of them are secreted to the extracellular environment where most disulfide-bonded proteins are found^{31,32}. From this perspective, thioredoxin enzymes are perhaps one of the few types of enzymes for which a correlation can be established between their pH sensitivity and the environmental conditions found outside cells^{31,32}. It is informative to compare the acid tolerance of the resurrected enzymes with that of enzymes from extant extremophiles. For example, Trx from *Sulfolobus tokodaii* (thermophilic archaea³³), with a melting temperature of 122 °C (**Supplementary Fig. 7**), is active at pH 7 ($0.12 \times 10^5 \text{ M}^{-1} \text{ s}^{-1}$ at 50 pN) but does not show detectable activity at pH 5 (**Fig. 4b**), which is not surprising given that *Sulfolobus* regulates its cytosolic pH³⁴. By contrast, Trx from *Acetobacter acetii* (acidophilic bacteria³⁵ that grows at pH 4) is active at pH 5 ($0.6 \times 10^5 \text{ M}^{-1} \text{ s}^{-1}$ at 100 pN), reflecting its acidic cytosol^{35,36}.

DISCUSSION

Our results show that the chemical mechanisms observed in modern Trx enzymes were present in reconstructed Trxs predicted to be present in Precambrian organisms. Our data support the view that ancestral Trx enzymes from LBCA, AECA and LACA, which lived in the mid to late Hadean, were highly resistant to temperature and active in relatively acidic conditions. Our results support the view that in early life Trx enzymes were present in hot environments and that these environments progressively cooled from 4 to 0.5 Gyr ago^{10,12,37,38}.

However, it is also possible that a much cooler early Earth was populated by psychrophiles, mesophiles and thermophiles and that the latter could have been the only survivors of cataclysmic events (for example, the late heavy bombardment or global glaciations on early Earth^{10,39}). Our findings also support the idea that many important biochemical pathways in the modern biosphere were already established by 3.5 Gyr ago¹⁰. For instance, metabolism is one of the most conserved cellular processes. Important pathways such as energy production, sugar degradation, cofactor biosynthesis and amino acid processing are highly conserved from bacteria to human and were probably present in LUCA⁴⁰. Evolution clearly operates at multiple levels of biological organization; however, enzymatic mechanisms accompanying adaptive changes seem to be highly conserved. The ability of enzymes to maintain specific chemical reactivity and mechanisms in disparate environments is necessary for the diversification of life. As we show here, this ability is exemplified by Trx enzymes, and it could also be the case for universal proteins, such as ubiquitin, RNase, ATPase or other metabolic enzymes that have been maintained in nearly all organisms throughout the history of life. We propose that the experimental resurrection of ancestors of these universal proteins together with the sensitivity of single-molecule techniques can be a powerful tool for understanding the origin and evolution of life on Earth.

METHODS

Methods and any associated references are available in the online version of the paper at <http://www.nature.com/nsmb/>.

Note: Supplementary information is available on the Nature Structural & Molecular Biology website.

ACKNOWLEDGMENTS

Supported by the US National Institutes of Health (HL061228 and HL066030 to J.M.F.); the Spanish Ministry of Science and Innovation (J.M.S.-R.); NASA Astrobiology (Georgia Institute of Technology); NASA Exobiology (E.A.G.); Fundación Ibercaja and Fundación Caja Madrid (R.P.-J. and S.G.-M.); and Fundación Alfonso Martín Escudero (J.A.-C.). We thank B. Ibarra-Molero (University of Granada) for assistance with bulk enzymatic assays.

AUTHOR CONTRIBUTIONS

R.P.-J., J.M.S.-R., E.A.G. and J.M.F. designed the research; Z.-M.Z. and E.A.G. conducted the phylogenetic analysis and sequence reconstruction; A.I.-P. and J.M.S.-R. expressed and purified the ancestral enzymes and conducted the calorimetric measurements and analysis; T.J.K. provided *A. acetii* Trx; M.T. provided *S. tokodaii* Trx; A.H. provided human TRX; R.P.-J., I.S.-R., J.A.-C., P.K. and S.G.-M. performed AFM experiments; R.P.-J. and I.S.-R. analyzed AFM data; R.P.-J., E.A.G. and J.M.F. wrote the paper; all authors participated in revising the manuscript.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

Published online at <http://www.nature.com/nsmb/>.

Reprints and permissions information is available online at <http://npg.nature.com/reprints/index.html>.

- Benner, S.A., Sassi, S.O. & Gaucher, E.A. Molecular paleoscience: systems biology from the past. *Adv. Enzymol. Relat. Areas Mol. Biol.* **75**, 1–132 (2007).
- Thornton, J.W. Resurrecting ancient genes: experimental analysis of extinct molecules. *Nat. Rev. Genet.* **5**, 366–375 (2004).
- Liberles, D.A. (ed.) *Ancestral Sequence Reconstruction* (Oxford Univ. Press, Oxford, New York, 2007).
- Hall, B.G. Simple and accurate estimation of ancestral protein sequences. *Proc. Natl. Acad. Sci. USA* **103**, 5431–5436 (2006).
- Thornton, J.W., Need, E. & Crews, D. Resurrecting the ancestral steroid receptor: ancient origin of estrogen signaling. *Science* **301**, 1714–1717 (2003).
- Thomson, J.M. *et al.* Resurrecting ancestral alcohol dehydrogenases from yeast. *Nat. Genet.* **37**, 630–635 (2005).
- Boussau, B., Blanquart, S., Necsulea, A., Lartillot, N. & Gouy, M. Parallel adaptations to high temperatures in the Archaean eon. *Nature* **456**, 942–945 (2008).

- Chang, B.S., Jonsson, K., Kazmi, M.A., Donoghue, M.J. & Sakmar, T.P. Recreating a functional ancestral archosaur visual pigment. *Mol. Biol. Evol.* **19**, 1483–1489 (2002).
- Gaucher, E.A., Thomson, J.M., Burgan, M.F. & Benner, S.A. Inferring the palaeoenvironment of ancient bacteria on the basis of resurrected proteins. *Nature* **425**, 285–288 (2003).
- Nisbet, E.G. & Sleep, N.H. The habitat and nature of early life. *Nature* **409**, 1083–1091 (2001).
- Pollock, D.D. & Chang, B.S.W. in *Ancestral Sequence Reconstruction* (ed. Liberles, D.A.) 85–94 (Oxford Univ. Press, Oxford, New York, 2007).
- Gaucher, E.A., Govindarajan, S. & Ganesh, O.K. Palaeotemperature trend for Precambrian life inferred from resurrected proteins. *Nature* **451**, 704–707 (2008).
- Zalatan, J.G. & Herschlag, D. The far reaches of enzymology. *Nat. Chem. Biol.* **5**, 516–520 (2009).
- Holmgren, A. Thioredoxin. *Annu. Rev. Biochem.* **54**, 237–271 (1985).
- Wiita, A.P. *et al.* Probing the chemistry of thioredoxin catalysis with force. *Nature* **450**, 124–127 (2007).
- Perez-Jimenez, R. *et al.* Diversity of chemical mechanisms in thioredoxin catalysis revealed by single-molecule force spectroscopy. *Nat. Struct. Mol. Biol.* **16**, 890–896 (2009).
- Hedges, S.B. & Kumar, S. *The Timetree of Life* (Oxford Univ. Press, Oxford, 2009).
- Yang, Z., Kumar, S. & Nei, M. A new method of inference of ancestral nucleotide and amino acid sequences. *Genetics* **141**, 1641–1650 (1995).
- 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. *J. Mol. Biol.* **362**, 966–978 (2006).
- Kice, J.L. Nucleophilic substitution at different oxidation states of sulfur. in *Progress in Inorganic Chemistry* (ed. Edwards, J.O.) 147–206 (John Wiley & Sons, 2007).
- Wiita, A.P., Ainavaru, S.R., Huang, H.H. & Fernandez, J.M. Force-dependent chemical kinetics of disulfide bond reduction observed with single-molecule techniques. *Proc. Natl. Acad. Sci. USA* **103**, 7222–7227 (2006).
- Koti Ainavaru, S.R., Wiita, A.P., Dougan, L., Uggerud, E. & Fernandez, J.M. Single-molecule force spectroscopy measurements of bond elongation during a bimolecular reaction. *J. Am. Chem. Soc.* **130**, 6479–6487 (2008).
- Garcia-Manyes, S., Liang, J., Szoszkiewicz, R., Kuo, T.L. & Fernandez, J.M. Force-activated reactivity switch in a bimolecular chemical reaction. *Nat. Chem.* **1**, 236–242 (2009).
- Liang, J. & Fernandez, J.M. Mechanochemistry: one bond at a time. *ACS Nano* **3**, 1628–1645 (2009).
- Holmgren, A. Reduction of disulfides by thioredoxin. Exceptional reactivity of insulin and suggested functions of thioredoxin in mechanism of hormone action. *J. Biol. Chem.* **254**, 9113–9119 (1979).
- Suarez, M. *et al.* Using multi-objective computational design to extend protein promiscuity. *Biophys. Chem.* **147**, 13–19 (2010).
- Holmgren, A. Thioredoxin catalyzes the reduction of insulin disulfides by dithiothreitol and dihydroliipoamide. *J. Biol. Chem.* **254**, 9627–9632 (1979).
- Walker, J.C.G. Possible limits on the composition of the archaean ocean. *Nature* **302**, 518–520 (1983).
- Russell, M.J. & Hall, A.J. The emergence of life from iron monosulphide bubbles at a submarine hydrothermal redox and pH front. *J. Geol. Soc. Lond.* **154**, 377–402 (1997).
- 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* **36**, 2622–2636 (1997).
- Xu, S.Z. *et al.* TRPC channel activation by extracellular thioredoxin. *Nature* **451**, 69–72 (2008).
- Windle, H.J., Fox, A., Ni Eidhin, D. & Kelleher, D. The thioredoxin system of *Helicobacter pylori*. *J. Biol. Chem.* **275**, 5081–5089 (2000).
- Ming, H. *et al.* Crystal structure of thioredoxin domain of ST2123 from thermophilic archaea *Sulfolobus tokodaii* strain 7. *Proteins* **69**, 204–208 (2007).
- Baker-Austin, C. & Dopson, M. Life in acid: pH homeostasis in acidophiles. *Trends Microbiol.* **15**, 165–171 (2007).
- Starks, C.M., Francois, J.A., MacArthur, K.M., Heard, B.Z. & Kappock, T.J. Atomic-resolution crystal structure of thioredoxin from the acidophilic bacterium *Acetobacter acetii*. *Protein Sci.* **16**, 92–98 (2007).
- Menzel, U. & Gottschalk, G. The internal pH of *Acetobacterium wieringae* and *Acetobacter acetii* during growth and production of acetic acid. *Arch. Microbiol.* **143**, 47–51 (1985).
- Knauth, L.P. & Lowe, D.R. High Archean climatic temperature inferred from oxygen isotope geochemistry of cherts in the 3.5 Ga Swaziland Supergroup, South Africa. *Geol. Soc. Am. Bull.* **115**, 566–580 (2003).
- Schulte, M. The emergence of life on earth. *Oceanography (Wash. D.C.)* **20**, 42–49 (2007).
- Gogarten-Boekels, M., Hilario, E. & Gogarten, J.P. The effects of heavy meteorite bombardment on the early evolution—the emergence of the three domains of life. *Orig. Life Evol. Biosph.* **25**, 251–264 (1995).
- Peregrin-Alvarez, J.M., Tsoka, S. & Ouzounis, C.A. The phylogenetic extent of metabolic enzymes and pathways. *Genome Res.* **13**, 422–427 (2003).

ONLINE METHODS

Phylogenetic analysis and ancestral sequence reconstruction. We retrieved 203 thioredoxin sequences from the three domains of life from GenBank. For GI numbers, see **Supplementary Note**. Sequences were aligned using MUSCLE⁴¹ and further corrected manually. The multiple sequence alignment is available upon request. The phylogenetic analysis was carried out by the minimum evolution distance criterion with 1,000 bootstrap replicates using PAUP* 4.0 beta (Sinauer Associates). Ancestral sequences were reconstructed using PAML version 3.14 and incorporated the gamma distribution for variable replacement rates across sites⁴². For each site of the inferred sequences, posterior probabilities were calculated for all 20 amino acids. The amino acid residue with the highest posterior probability was then assigned at each site.

Protein expression and purification. Genes encoding the ancestral Trx enzymes were synthesized and codon-optimized for expression in *E. coli* cells (Epoch Biolabs). The genes were cloned into pQE80L vector (Qiagen) and transformed in *E. coli* BL21 (DE3) cells (Invitrogen). Cells were incubated overnight in LB medium at 37 °C, and protein expression was induced with 1 mM IPTG. Cell pellets were sonicated and the His₆-tagged proteins were loaded onto His GraviTrap affinity column (GE Healthcare). The purified protein was verified by SDS-PAGE. The proteins were then loaded into PD-10 desalting column (GE Healthcare) and finally dialyzed against 50 mM HEPES, pH 7.0 buffer. The preparation of (I27_{G32C-A75C})₈ was carried out as described^{15,43}.

DSC experiments. Thermal stabilities of ancestral and modern Trx enzymes were measured with a VP-Capillary DSC (MicroCal). Protein solutions were dialyzed into a buffer of 50 mM HEPES, pH 7. The scan speed was set to 1.5 K min⁻¹. Several buffer-buffer baselines were first obtained for proper equilibration of the calorimeter. Concentrations were 0.3–0.7 mg ml⁻¹ and were determined spectrophotometrically at 280 nm using theoretical extinction coefficients and molecular weights. The experimental traces were analyzed following the two-state thermodynamic model⁴⁴.

AFM experiments. The AFM used is a custom-made design previously described⁴⁵. Cantilever model MLCT of silicon nitride were used (Veeco). We calibrated the cantilever using the equipartition theorem⁴⁶ giving rise to a typical spring constant of 0.02 N m⁻¹. The AFM works in the force-clamp mode with length resolution 0.5 nm. The feedback response can reach 5 ms. The buffer used in the experiment was 10 mM HEPES, pH 7.2, 150 mM NaCl, 1 mM EDTA and 2 mM NADPH. We add Trx enzymes to the desired concentration. The buffer also contains Trx reductase (50 nM) to keep Trx enzymes in their reduced state. *E. coli* Trx reductase works well with bacterial-origin Trx enzymes, whereas eukaryotic Trx reductase works with archaea and eukaryote Trx enzymes; however, similar results were obtained using DTE 200 μM to keep Trx enzymes reduced, thus showing that modern Trx reductases maintain fully reduced ancestral Trx enzymes. For the experiments at pH 5, we used 20 mM sodium acetate buffer and 200 μM DTE.

To perform the experiment, we deposited 3–6 μl of substrate ~0.1 mg ml⁻¹ on a gold-covered coverslide. A drop of ~100 μl containing the Trx solution was then added. The force-clamp protocol consists of three pulses of force. In the first pulse the cantilever tip was pressed against the surface at 800 pN for 2 s. In the second pulse the attached (I27_{G32C-A75C})₈ was stretched at 175–185 pN for 0.2–0.3 s. The third pulse was the test force where the reduction

events were captured. This pulse was applied at different forces (30–500 pN) enough times to capture all the possible reduction events.

The traces were collected and analyzed using custom-written software in Igor Pro 6.03 (Wavemetrics). The traces containing the reduction events at each force were summed, normalized and fitted with a single exponential. From the fitting we obtained a time constant τ and thus the reduction rate at a given force ($r = 1/\tau$). We used the bootstrapping method to obtain the error of the reduction rates. The bootstrapping was run 1,000 times for each reduction rate to obtain a distribution from which the s.e.m. could be calculated.

Thioredoxin bulk enzymatic measurements. Bulk-solvent oxidoreductase activity for ancestral thioredoxins was determined using the insulin precipitation assay as described^{26,27,43}. To further verify the activity of ancestral Trxs enzymes at acidic pH, we used DTNB (5,5'-dithiobis-(2-nitrobenzoic acid)) as substrate at pH 5. In this assay, Trxs enzymes were preactivated by incubation with 1 mM DTT. The reaction was initiated by adding active Trx to a final concentration of 4 μM to the cuvette containing 1 mM DTNB in 20 mM sodium acetate buffer, pH 5. We followed the change in absorbance at 412 nm due to the formation of TNB during 1 min. Activity was determined from the slope $d\Delta A_{412}/dt$. A control experiment lacking Trx was registered and subtracted as baseline.

AFM data analysis. The data were fitted following a three-state kinetic model as described^{15,16}. In this model three different chemical mechanisms are taken into account. The rate constants used in the kinetic model are

$$\begin{aligned}k_{01} &= \alpha_0 [\text{Trx}] \\k_{12} &= \beta_0 \exp(F\Delta x_{12}/k_B T) + \lambda_0 \\k_{02} &= \gamma_0 [\text{Trx}] \exp(F\Delta x_{02}/k_B T) + \lambda_0 \\k_{10} &= \delta_0\end{aligned}$$

The rate constants k_{01} and k_{02} depend on Trx concentration in a linear manner. k_{12} and k_{02} exponentially depend on force. The kinetic model is solved using matrix analysis and parameters α_0 , β_0 , Δx_{12} , γ_0 , Δx_{02} , λ and δ_0 can be obtained for each ancestral enzyme. The optimal kinetic parameters are calculated by numerical optimization using the downhill simplex method⁴⁷ (**Supplementary Table 1**). An extensive explanation of the different chemical mechanisms can be found in ref. 16, where the origin of this chemical diversity was explained on the basis of the structural features of the binding groove.

Igor Pro software was used for data analysis. All figures were generated using Igor Pro and Adobe Illustrator CS4.

- Edgar, R.C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32**, 1792–1797 (2004).
- Yang, Z.H. PAML: a program package for phylogenetic analysis by maximum likelihood. *Comput. Appl. Biosci.* **13**, 555–556 (1997).
- Perez-Jimenez, R. *et al.* Force-clamp spectroscopy detects residue co-evolution in enzyme catalysis. *J. Biol. Chem.* **283**, 27121–27129 (2008).
- Ibarra-Molero, B., Loladze, V.V., Makhatazde, G.I. & Sanchez-Ruiz, J.M. 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 (1999).
- Fernandez, J.M. & Li, H. Force-clamp spectroscopy monitors the folding trajectory of a single protein. *Science* **303**, 1674–1678 (2004).
- Florin, E.L. *et al.* Sensing specific molecular interactions with the atomic-force microscope. *Biosens. Bioelectron.* **10**, 895–901 (1995).
- Nelder, J.A. & Mead, R.A. Simplex method for function minimization. *Comput. J.* **7**, 308–313 (1965).

Supplementary Information for:

Single-molecule paleoenzymology probes the chemistry of resurrected enzymes

Raul Perez-Jimenez^{1*}, Alvaro Inglés-Prieto², Zi-Ming Zhao³, Inmaculada Sanchez-Romero², Jorge Alegre-Cebollada¹, Pallav Kosuri^{1,4}, Sergi Garcia-Manyes¹, T. Joseph Kappock⁵, Masaru Tanokura⁶, Arne Holmgren⁷, Jose M. Sanchez-Ruiz², Eric A. Gaucher^{3,8,9} & Julio M. Fernandez^{1*}.

¹*Department of Biological Sciences, Columbia University, New York, New York, USA.*

²*Facultad de Ciencias, Dpto. de Química-Física, Universidad de Granada, Granada, Spain.*

³*Georgia Institute of Technology, School of Biology, Atlanta, Georgia, USA.*

⁴*Department of Biochemistry and Molecular Biophysics, Columbia University, New York, New York, USA.*

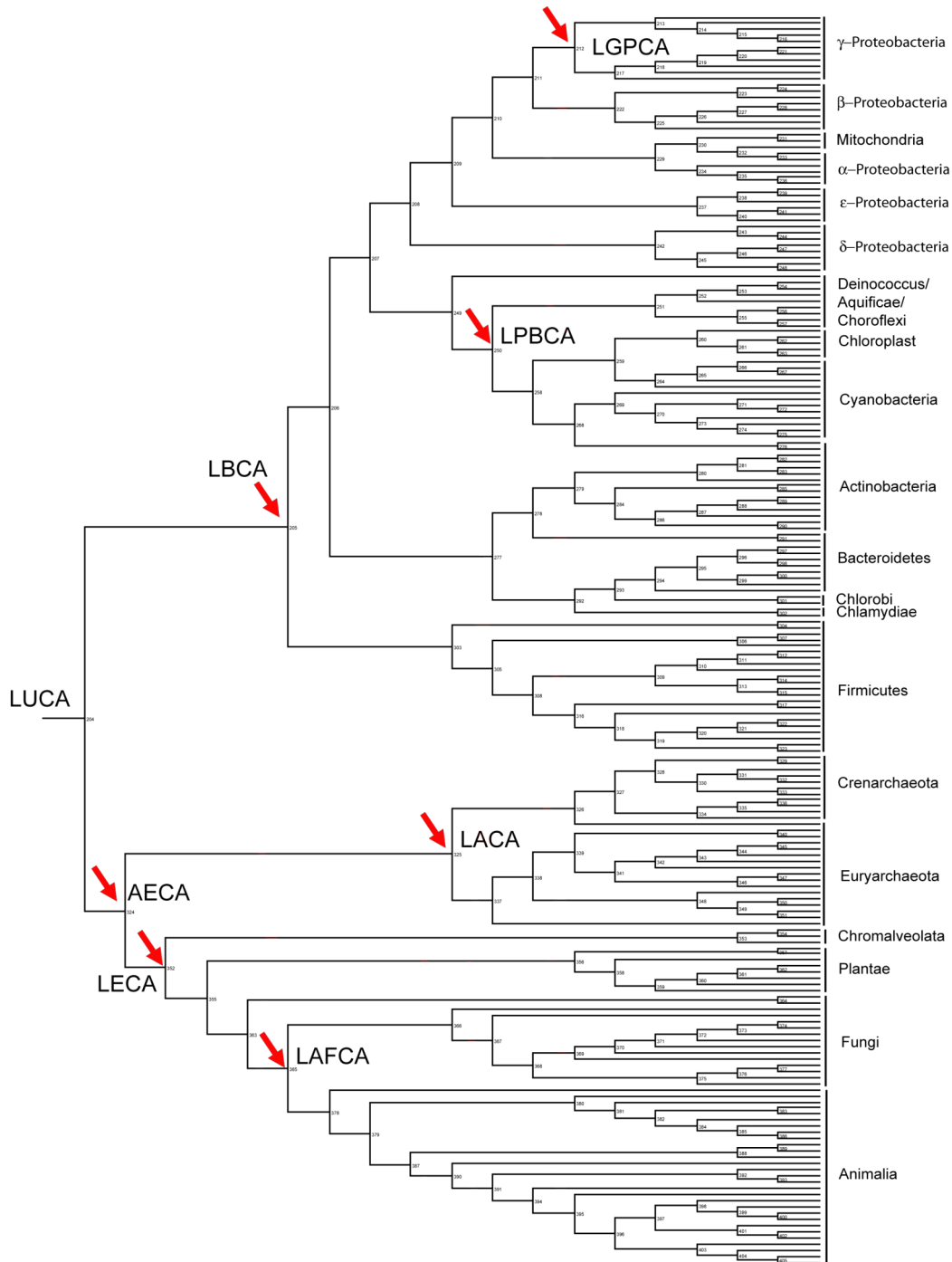
⁵*Department of Biochemistry, Purdue University, West Lafayette, Indiana 47907-2063.*

⁶*Department of Applied Biological Chemistry, Graduate School of Agricultural and Life Sciences, The University of Tokyo, Tokyo.*

⁷*Division of Biochemistry, Department of Medical Biochemistry and Biophysics, Karolinska Institutet, SE-171 77, Stockholm, Sweden.*

⁸*Georgia Institute of Technology, School of Chemistry, Atlanta, Georgia, USA.*

⁹*Georgia Institute of Technology, Parker H. Petit Institute for Bioengineering and Bioscience, Atlanta, Georgia, USA.*



Supplementary Fig.1. Phylogenetic Tree used for the ancestral sequence reconstruction of Trx enzymes. A total of 203 sequences were used (see list below). The nodes of interest are indicated with red arrows. Last bacterial common ancestor (LCBA), last archaeal common ancestor (LACA), archaea/eukaryota common ancestor (AECA), last common ancestor of cyanobacterial and deinococcus/thermus groups (LPBCA, origin of photosynthetic bacteria); last eukaryotic common ancestor (LECA), last common ancestor of γ -proteobacteria (LGPCA) and last common ancestor of animals and fungi (LAFCA).

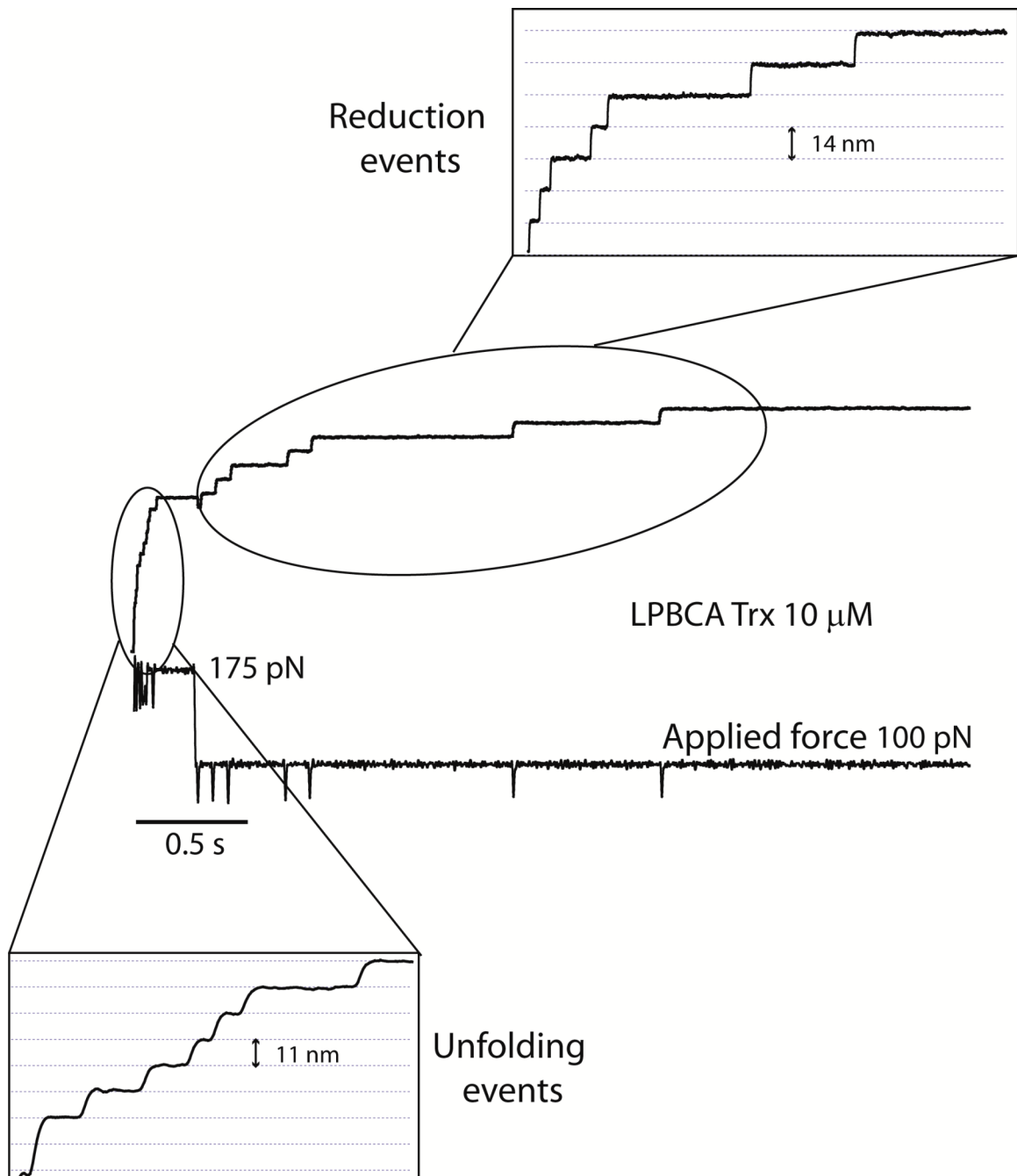
```

LBCA Trx      -MSVIEINDENFEEEV LKS-DKPVLVDFWAPWCGPCRMIAPIIEELAE EYEGKVK
LACA Trx      -MSVVQLNDENFDEVIKKN-NKVVVDFWAEWCGPCRMIAPIIEELAKE YAGKVV
AECA Trx      -MSVIEINDENFDEVIKKS-DKVVVDFWAEWCGPCRMIAPIIEELAE EYAGKVV
LGPCA Trx     -MSIIHVTDDSF DQDVLKA-DKPVLVDFWAEWCGPCKMIAPILDEIAE EYEGKLLK
LECA Trx      -MVIQVTNKEEFEAILSEA-DKLVVVDFATWCGPCKMIAPFFEEELSEE YPDKVV
LPBCA Trx     -MSVIEVTDENFEQEV LKS-DKPVLVDFWAPWCGPCRMIAPIIEELAKE YEGKVK
LAFCA Trx     -MVIQVTNKDEFESILSEA-DKLVVVDFATWCGPCKMIAPKFEELSEE YPDNVV
E.coli Trx    SDKIIHLTDDSF DTDVLKA-DGAILVDFWAEWCGPCKMIAPILDEIAE YQGKLT
Human TRX     -MVKQIESKTAFQEALDAAGDKLVVVDFSATWCGPCKMIKPF FHSLSSEKYS-NVI
              .. * : : : : : : * * * * * : * * : . . . . . : * : :

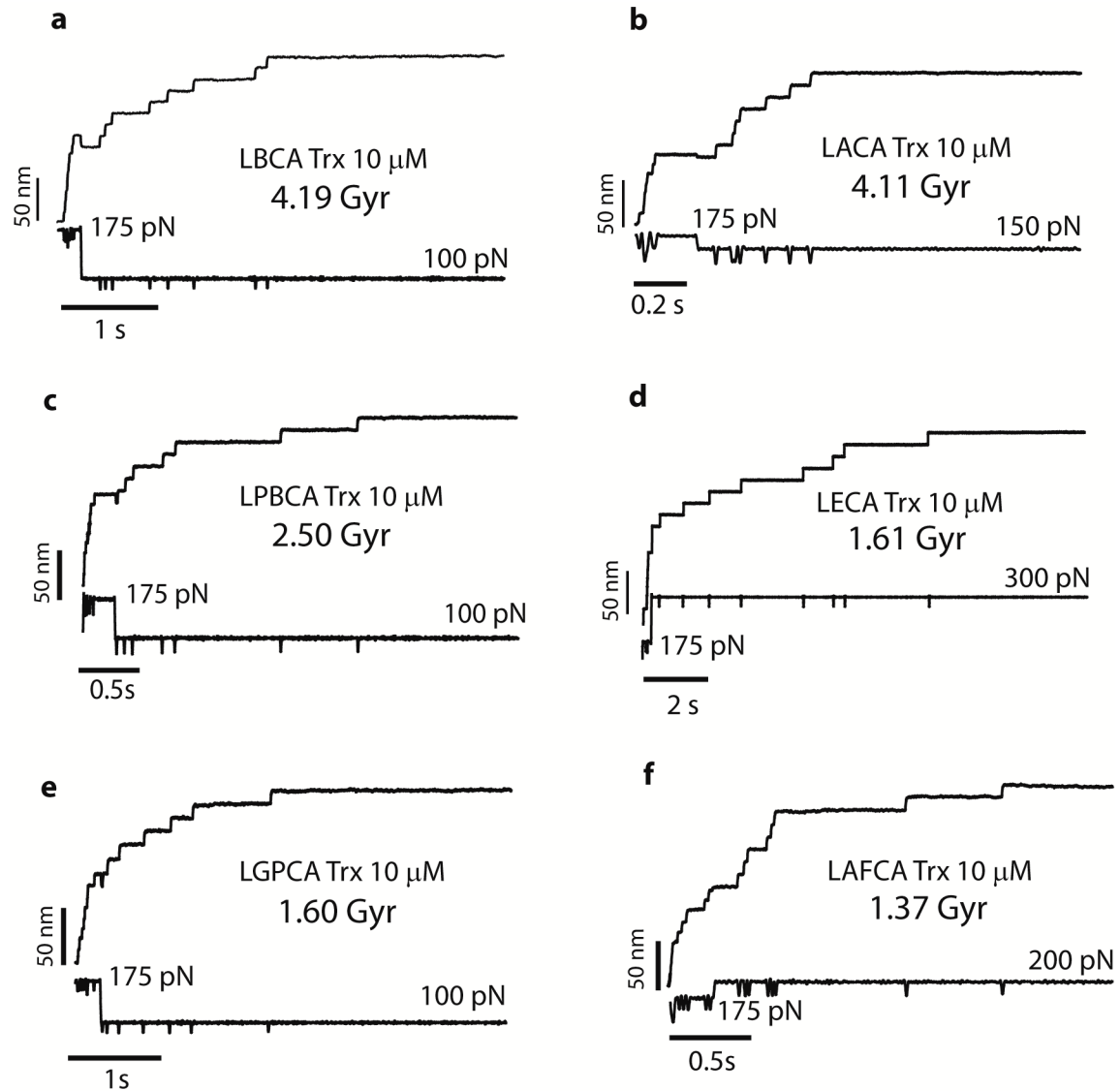
LBCA Trx      FAKVNV DENPETAAKYGIMS IPTLLL FKNGEVVDKLVGARPKEALKERIEKHL-
LACA Trx      FGKLVN DENPEIAAKYGIMS IPTLLF FKN GKVVDQLVGAMPKEALKERIKKYL-
AECA Trx      FGKLVN DENPEIAAKYGIMS IPTLLF FKN GKVVDQLVGARPKEALKERIKKYL-
LGPCA Trx     VAKVNI DENPETAAKYGIRGI PTLML FKNGEVAATKVGALS KSQLKEFLDANL-
LECA Trx      FIKVDV DEVPDVA AKYGI TSMP TFKFFKN GKKVDELV GAN- QEKLKQ MILKHAP
LPBCA Trx     VVKVNV DENPNTAAQYGI RSIP TLLL FKN GQVVDRLV GAQPKEALKERIDKHL-
LAFCA Trx     FLKVDV DEVEDVA AEYGI SAMP TFOFFKN GKKVDEL TGAN- QEKLKAMIKKHAA
E.coli Trx    VAKVNI DQNPGTAPKYGI RGI PTLLL FKN GEVAATKVGALS KSQLKEFLDANLA
Human TRX     FLEV DVDCQDVASECEVKCTP TFOFFKKGQKVGEFSGAN- KEKLEATINELV-
              . : : : : * : * . : : * : : * : * . * * : * : :

```

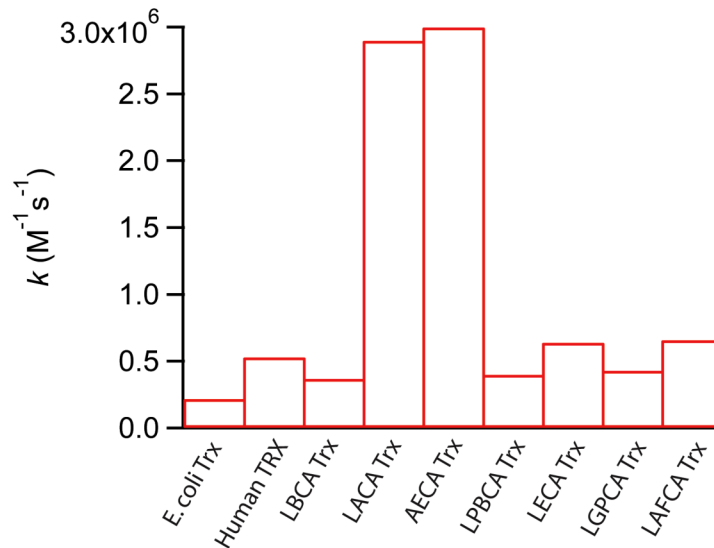
Supplementary Fig.2. M-PASs for Trx enzymes belonging to representative extinct organisms. The sequences are calculated using maximum likelihood methods (see main text and methods section). We also included *E. coli* and human Trx sequences for comparative purposes. The active site, CGPC, is highly conserved.



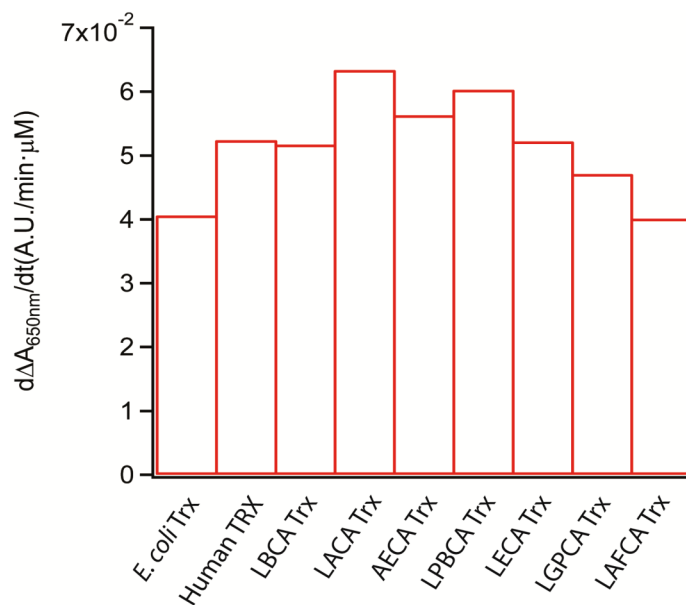
Supplementary Fig.3. Force-clamp experiment for detection of single disulfide reduction events. A first pulse of force (175 pN, 0.3 s) unfolds the I27_{G32C-A75C} domains up to the disulfide bond. The unfolding events can be monitored as a series of step of ~11 nm per module (bottom panel). A second pulse of force (100 pN) is applied to monitor single disulfide reduction by Trx enzymes. In this case the release of the trapped residues behind the disulfide bond gives rise to a length increment of ~14 nm per module (top panel).



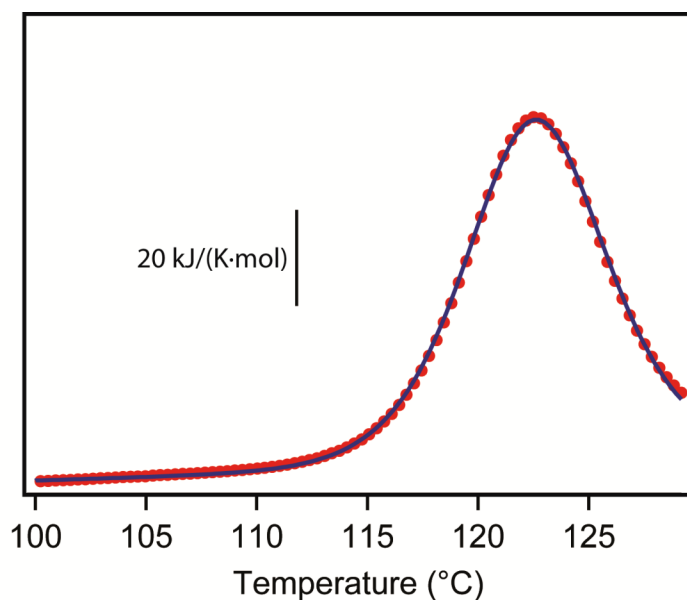
Supplementary Fig.4. Experimental traces of single disulfide reductions by ancestral Trxs. Both, the unfolding pulse (175 pN) and the test pulse at different forces are shown. Individual reduction events can be observed in the test-pulse force. Numerous traces like these (15-80) are used at every force to complete the full force-dependency of disulfide bond reduction by Trx enzymes, as it is shown in **Fig.3** in the main text.



Supplementary Fig.5. Rate constants for disulfide bond reduction by ancestral and modern Trxs enzymes. These values are obtained by extrapolating to zero force the fitting of the reduction rate vs. force data (Fig.3 in main text) to the three-state kinetic model described in the methods section.



Supplementary Fig.6. Insulin activity assay for ancestral and modern Trx enzymes. Activity determined with the turbidity insulin bulk enzymatic assay¹. The turbidity assay is less sensitive in detecting differences in activity amongst the different enzymes. This assay cannot be used to probe the activity of the enzymes at pH 5 due to the precipitation of insulin at pH below 6 (ref¹⁻²).



Supplementary Fig. 7. Experimental DSC thermogram for *Sulfolobus tokodaii* Trx (Archaea). The solid blue line represents fit to the two-state thermodynamic model³. A T_m of 122.6° C is obtained from the fit.

Enzyme	α_0 ($\mu\text{M}^{-1}\cdot\text{s}^{-1}$)	β_0 (s^{-1})	γ_0 ($\mu\text{M}^{-1}\cdot\text{s}^{-1}$)	k_{10} (s^{-1})	Δx_{12} (Å)	Δx_{02} (Å)	λ_0 (s^{-1})
LBCA Trx	0.47±0.08	30±2	0.004±0.001	5.8±0.7	-0.74±0.06	0.19±0.02	0.09±0.02
LACA Trx	8.2±0.2	43±3	-	3.8±1	-0.76±0.04	-	0.38±0.05
AECA Trx	4.2±0.3	25±2	0.019±0.004	3.8±0.6	-0.84±0.05	0.19±0.02	0.21±0.04
LPBCA Trx	0.47±0.04	30±3	0.017±0.002	4.9±0.5	-0.71±0.01	0.17±0.02	0.19±0.02
LECA Trx	0.76±0.08	38±2	-	4.2±0.7	-0.80±0.03	-	0.18±0.01
LGPCA Trx	0.48±0.02	34±2	0.012±0.002	3.8±0.4	-0.83±0.02	0.17±0.02	0.35±0.02
LAFCA Trx	0.81±0.10	37±3	-	4.6±0.8	-0.74±0.03	-	0.06±0.02
<i>E. coli</i> Trx1*	0.25±0.02	24±2	0.012±0.002	4.7±0.5	-0.74±0.05	0.16±0.01	0.08±0.02
Human TRX1*	0.52±0.05	33±2	-	3.1±0.9	-0.71±0.05	-	0.35±0.02

Supplementary Table 1. Kinetic parameters for ancestral Trxs. The parameters were obtained using the kinetic model previously described (see methods section and references 14 and 15 in the main text). They are the result of numeric optimization of the global fit using the downhill simplex method. The errors correspond to the standard deviation. (*) *E. coli* and human Trxs are also included, obtained from previous publications⁴⁻⁵.

Supplementary Note.

Thioredoxin sequences used for ancestral sequences reconstruction. GI numbers were accessed from GenBank. The names of the hosting organisms are also provided:

57164261 Ovis	167537844 Monosiga
27806783 Bos	67479051 Entamoeba
47523692 Sus	165988451 Dictyostelium
126352340 Equus	15236327 Arabidopsis
6755911 Mus	15232567 Arabidopsis
16758644 Rattus	154721452 Limonium
146291083 Rabbit	162461510 Zea
135773 Human	157335070 Vitis
67461921 Ponab	145351136 Ostreococcus
267126 Macmu	53801490 Helicosporidium
13560979 Callithrix	1620905 Fagopyrum
126339826 Monodelphis	46226985 Cryptosporidium
149412981 Ornithorhynchus	68350806 Theileria
45382053 Gallus	148804689 Plasmodium
29373131 Melopsittacus	11498883 Archaeoglobus
12958636 Ophiophagus	116754023 Methanosaeta
194332745 Xenopus	91773622 Methanococcoides
47215756 Tetraodon	154149646 Candidatus
9837585 Ictalurus	88603734 Methanospirillum
50539990 Danio	48477193 Picrophilus
194160556 Drosophila	150401020 Methanococcus
17648013 Drosophila	124485138 Methanocorpusculum
194141429 Drosophila	116754438 Methanosaeta
48104680 Apis	76802488 Natronomonas
91084205 Tribolium	110667588 Haloquadratum
148298796 Bombyx	55380304 Haloarcula
90819972 Graphocephala	76802694 Natronomonas
169639275 Litopenaeus	16120325 Halobacterium
30580603 Geocy	11499727 Archaeoglobus
115401922 Aspergillus	13541608 Thermoplasma
119479067 Neosartorya	119720035 Thermofilum
40746887 Aspergillus	159040636 Caldivirga
115401518 Aspergillus	70607552 Sulfolobus
150951554 Pichia	15899007 Sulfolobus
46441186 Candida	15922449 Sulfolobus
126213085 Pichia	124027987 Hyperthermus
50309357 Kluyveromyces	118431868 Aeropyrum
151943486 Saccharomyces	146304377 Metallosphaera
50291653 Candida	70607229 Sulfolobus
151941211 Saccharomyces	15897303 Sulfolobus
19114764 Schizosaccharomyces	126465005 Staphylothermus

118431901	Aeropyrum	21223797	Streptomyces
15894111	Clostridium	15611050	Mycobacterium
20808289	Thermoanaerobacter	72163508	Thermobifida
16079205	Bacillus	21222296	Streptomyces
16077522	Bacillus	16329883	Synechocystis
15901736	Streptococcus	17229833	Nostoc
29377495	Enterococcus	17229385	Nostoc
153181008	Listeria	16331440	Synechocystis
28377165	Lactobacillus	22299829	Thermosynechococcus
28379765	Lactobacillus	22297898	Thermosynechococcus
150393692	Staphylococcus	16329237	Synechocystis
138896249	Geobacillus	22299630	Thermosynechococcus
30264587	Bacillus	17229697	Nostoc
16079902	Bacillus	17229859	Nostoc
28378864	Lactobacillus	22298354	Thermosynechococcus
153179313	Listeria	17229358	Nostoc
29375972	Enterococcus	126696505	Prochlorococcus
15901605	Streptococcus	16331825	Synechocystis
110798962	Clostridium	17227548	Nostoc1
110800418	Clostridium	1351239	Pea Chloroplast
15894825	Clostridium	2507458	Spiol Chloroplast
15896334	Clostridium	11135474	Wheat Chloroplast
20807685	Thermoanaerobacter	15594012	Pisum Chloroplast
76789276	Chlamydia	11135407	Branan Chloroplast
15836191	Chlamydomonas	46199419	Thermus
119357517	Chlorobium	15807833	Deinococcus
119357012	Chlorobium	46199687	Thermus
29345629	Bacteroides	15805968	Deinococcus
150024368	Flavobacterium	147669275	Dehalococcoides
34539910	Porphyromonas	118047160	Chloroflexus
29347639	Bacteroides	118048687	Chloroflexus
29346087	Bacteroides	118046691	Chloroflexus
34540117	Porphyromonas	15606934	Aquifex
29346866	Bacteroides	42521808	Bdellovibrio
29345628	Bacteroides	39998535	Geobacter
32477354	Rhodopirellula	42523902	Bdellovibrio
32476401	Rhodopirellula	120602368	Desulfovibrio
15608608	Mycobacterium	39998370	Geobacter
57116870	Mycobacterium	116619824	Solibacter
62391823	Corynebacterium	116619449	Solibacter
72163169	Thermobifida	94970094	Acidobacteria
21219405	Streptomyces	34556879	Wolinella
72160576	Thermobifida	15645443	Helicobacter
15607956	Mycobacterium	57237155	Campylobacter
21219599	Streptomyces	15646067	Helicobacter
62391938	Corynebacterium	34557886	Wolinella

34556999	Wolinella	67005950	Escherichia
159184127	Agrobacterium	16130507	Escherichia
150398433	Sinorhizobium	30063983	Shigella
17988305	Brucella	16765969	Salmonella
15603883	Rickettsia	16123427	Yersinia
108935910	Bovin Mitochondrio	27366792	Vibrio
194226778	Equus		
21361403	Homo Mitochondrion		
16758038	Rattus Mitochondrio		
9903609	Mus Mitochondrion		
74318624	Thiobacillus		
121635072	Neisseria		
74316054	Thiobacillus		
126454139	Burkholderia		
33602206	Bordetella		
74318419	Thiobacillus		
74316241	Thiobacillus		
33602001	Bordetella		
66043570	Pseudomonas		
27364380	Vibrio		
16124003	Yersinia		
16767191	Salmonella		
30064924	Shigella		

References

1. Holmgren, A. Thioredoxin catalyzes the reduction of insulin disulfides by dithiothreitol and dihydrolipoamide. *J Biol Chem* **254**, 9627-32 (1979).
2. Holmgren, A. Reduction of disulfides by thioredoxin. Exceptional reactivity of insulin and suggested functions of thioredoxin in mechanism of hormone action. *J Biol Chem* **254**, 9113-9 (1979).
3. Ibarra-Molero, B., Loladze, V.V., Makhatadze, G.I. & Sanchez-Ruiz, J.M. Thermal versus guanidine-induced unfolding of ubiquitin. An analysis in terms of the contributions from charge-charge interactions to protein stability. *Biochemistry* **38**, 8138-49 (1999).
4. Perez-Jimenez, R. et al. Diversity of chemical mechanisms in thioredoxin catalysis revealed by single-molecule force spectroscopy. *Nat Struct Mol Biol* **16**, 890-6 (2009).
5. Wiita, A.P. et al. Probing the chemistry of thioredoxin catalysis with force. *Nature* **450**, 124-7 (2007).

**4. OBTENCIÓN DE LAS ESTRUCTURAS DE
TIORREDOXINAS PRECÁMBRICAS POR
CRISTALOGRAFÍA DE RAYOS-X Y ANÁLISIS
EVOLUTIVO A LO LARGO DE 4 MIL MILLONES
DE AÑOS**

— — — — —
OBTENCIÓN DE LAS ESTRUCTURAS DE TIORREDOXINAS PRECÁMBRICAS
POR CRISTALOGRAFÍA DE RAYOS-X Y ANÁLISIS EVOLUTIVO A LO LARGO DE — — — — —
4 MIL MILLONES DE AÑOS

4.1. Introducción.

Como hemos visto en el capítulo anterior, nuestros estudios sugieren que las tiorredoxinas ancestrales estaban adaptadas para operar en las condiciones de temperatura y acidez de los mares primitivos. Además, las tiorredoxinas ancestrales emplean mecanismos de reducción similares a los de las tiorredoxinas actuales. Esto implica que la divergencia catalítica mostrada por las tiorredoxinas actuales se estableció hace unos 4 mil millones de años antes del presente. Estos resultados tienen, claramente, implicaciones desde el punto de vista evolutivo. Indudablemente, la obtención y el análisis de las estructuras de las tiorredoxinas resucitadas puede expandir el alcance y las implicaciones evolutivas de los estudios de las proteínas ancestrales resucitadas. Por este motivo, hemos determinado las estructuras tridimensionales de las tiorredoxinas ancestrales mediante cristalografía de rayos-X. Cabe destacar que estas son las primeras proteínas precámbricas cuya estructura ha sido determinada experimentalmente. Por este motivo, nos encontramos en una posición única para entender a nivel molecular la evolución de las estructuras de estas enzimas.

Hasta el momento, el factor que más ha contribuido a nuestra comprensión de la evolución de las proteínas ha sido el gran número de secuencias proteicas descubiertas en la era genómica, ya que la identidad de secuencia a menudo se puede tomar como una indicación de homología (ascendencia común). Por otra parte, poco se sabe con certeza acerca de la evolución de las estructuras de las proteínas, a pesar del gran número de plegamientos (*folds*) diferentes revelados por las estructuras depositadas en el *Protein Data Bank*. Existen varios factores que contribuyen a esta desfavorable situación. Si bien, en general se admite que las estructuras de las proteínas cambian a un ritmo más lento que las secuencias, en los últimos años, se han acumulado evidencias que apoyan que las estructuras de proteínas no son inalterables y, por tanto, que pueden cambiar en el curso de la evolución (Grishin 2001; Taylor 2007; Murzin 2008; Valas, Yang et al. 2009). De hecho, el mapeado de la secuencia en la estructura posee propiedades de "*shape-covering*" (Caetano-Anolles, Wang et al. 2009), lo que significa que estructuras diferentes podrían estar a unas pocas mutaciones de distancia en el espacio de la secuencia, como se ha demostrado experimentalmente (Cordes, Walsh et al. 1999). Además, se acepta generalmente la posibilidad de la evolución convergente de los *folds* y, por consiguiente, que dos proteínas tengan un ancestro común, no implica necesariamente la similitud de sus estructuras (Orengo, Jones et al. 1994; Grishin

2001; Krishna y Grishin 2004; Taylor 2007; Valas, Yang et al. 2009; Schaeffer y Daggett 2011). De hecho, existen dificultades para aplicar algunos de los principios básicos del análisis evolutivo a las estructuras de proteínas, tales como la identificación de los caracteres ancestrales y derivados (Taylor 2007; Caetano-Anolles, Wang et al. 2009). Por lo tanto, muchas clasificaciones de *fold*s de proteínas actuales son fenéticas (basadas en la determinación de la semejanza estructural) y la posibilidad de las clasificaciones filéticas (basadas en las relaciones evolutivas) sigue siendo una cuestión sin resolver (Taylor 2007; Valas, Yang et al. 2009). Como resultado, las estimaciones de edad para los diferentes *fold*s de proteínas son inciertos y basados en métodos indirectos, tales como el censo de *fold*s (asignados) de los genomas (Winstanley, Abeln et al. 2005; Caetano-Anolles, Wang et al. 2009). Además, la utilidad del concepto de *fold* está en entredicho, ya que varios autores han discutido que el espacio del *fold* debe ser considerado como continuo en vez de discreto (Honig 2007; Xie y Bourne 2008; Sadreyev, Kim et al. 2009).

Nosotros proponemos que la resurrección de proteínas ancestrales puede aliviar estos problemas y puede emplearse para estudiar de la evolución de la estructura de proteínas. La reconstrucción de secuencias de proteínas ancestrales se basa en los análisis filogenéticos de secuencias actuales y, por lo tanto, no está sujeto a las incertidumbres asociadas con la clasificación de las estructuras. Además, como hemos visto en el capítulo anterior, es posible resucitar proteínas del período precámbrico (Gaucher, Govindarajan et al. 2008; Perez-Jimenez, Li et al. 2009), permitiéndonos estudiar una escala temporal en que podemos esperar que se produzcan cambios significativos en la estructura de las proteínas.

Desgraciadamente, el número de estructuras de proteínas ancestrales que figuran en la bibliografía es muy escaso y está principalmente representado por estructuras relativamente recientes. Por ejemplo, se ha estudiado la estructura de la hemoglobina de mamut lanudo (43000 años) usando técnicas de modelado molecular a partir de la estructura de la hemoglobina humana (Campbell, Roberts et al. 2010). En otro estudio, han sido capaces de resolver la estructura de una vancomicina ancestral obtenida a partir del ADN conservado en *permafrost* (30000 años) (D'Costa, King et al. 2011). Sin embargo, hasta la fecha solo se han publicado tres trabajos describiendo la estructura de proteínas resucitadas empleando métodos de reconstrucción filogenéticos. Uno de ellos es la variante $\beta 9$ de sulfotransferasa (450 millones de años) (Alcolombri, Elias et al. 2011); otro es el

precursor de glucocorticoides y mineralocorticoides de vertebrados (Ortlund, Bridgham et al. 2007; Carroll, Ortlund et al. 2011); y, por último, el precursor de galectinas (congerinas) de pez (Konno, Kitagawa et al. 2011).

En este capítulo analizamos el uso de la resurrección de proteínas para comprender la evolución de estructura de proteínas. Para llevar a cabo este análisis, en primer lugar hemos resuelto la estructura mediante cristalografía de rayos-X de todas las tiorredoxinas resucitadas pertenecientes a organismos del periodo precámbrico (Fig. 4.1). En segundo lugar, hemos realizado un detallado análisis comparativo de las estructuras de las tiorredoxinas ancestrales con las tiorredoxinas actuales de *Escherichia coli* y humana. Este estudio podría revelar la evolución del *fold* de sus estructuras, además de permitirnos comprender los mecanismos moleculares implicados en la gran estabilidad y el mantenimiento de la actividad a pH ácido.

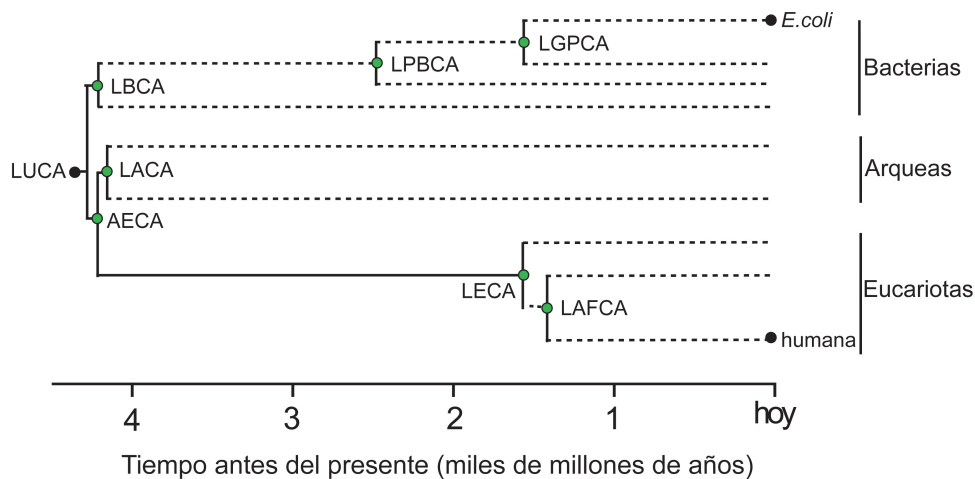


Figura 4.1. Árbol filogenético esquemático que muestra el tiempo geológico en el que vivieron diferentes organismos extintos. Las líneas discontinuas muestran bifurcaciones posteriores. Los tiempos de divergencia se han recopilado de múltiples fuentes y se resumen en el libro de Hedges, S. (Hedges 2009).

4.2. Determinación de las estructuras de tiorredoxinas precámbricas por cristalografía de rayos-X.

El requisito fundamental para poder abordar la determinación estructural de proteínas mediante la cristalografía de rayos-X es disponer de monocristales, homogéneos y de buena calidad, tal como detallamos en el apartado 2.2.7.1 de Materiales y Métodos. Para ello, la solución de proteína debe reunir una serie de características, como son: una concentración elevada, una pureza tan alta como sea posible, homogeneidad y estabilidad en las condiciones de cristalización (Ducruix y Giegé 2000).

Respecto a la **concentración** de la proteína, hemos usado diferentes valores dependiendo de si se trataba del muestreo primario o de la optimización de las condiciones. En la mayoría de los casos, la concentración de proteína se fijó en torno a 30-50 mg/mL. A partir de las observaciones obtenidas, decidimos aumentar la concentración entre los 50 y 100 mg/mL, condiciones óptimas para obtener un barrido más amplio de la sobresaturación en contradifusión.

La **pureza** de la solución de proteína debe ser tan alta como sea posible. Las impurezas pueden impedir la nucleación o limitar el tamaño de los cristales, ya que su interposición puede abortar el crecimiento y/o introducir irregularidades que disminuyan la calidad del cristal obtenido. Para asegurarnos de que la pureza es máxima, siempre hemos realizado una SDS-PAGE inmediatamente antes de comenzar el experimento de cristalización.

Además, es muy importante que la proteína sea **homogénea**. La heterogeneidad puede aparecer como resultado de la manipulación experimental (grados distintos de oxidación, proteólisis, saturación parcial con ligandos, agregación en grado variable, desnaturalización parcial, deamidación, etc.). Para obtener la máxima homogeneidad, hemos mantenido la solución de proteína a una temperatura de 4°C durante la purificación siempre que ha sido posible, evitando así el calentamiento de la muestra y la acción de las proteasas bacterianas. Tras la cromatografía de filtración en gel, hemos reunido solo las fracciones más puras de la proteína (aquellas que corresponden al máximo del pico del cromatograma). Durante el procesado previo al experimento de cristalización, hemos evitado tratamientos que

puedan introducir heterogeneidad, tales como tratamientos térmicos, la liofilización o la congelación de la solución de proteína.

El *buffer* en que está disuelta nuestra muestra debe ser aquel en que la **solubilidad** de la proteína sea máxima, ya que la selección del *buffer* puede afectar a la conservación y por tanto al proceso de cristalización. Con el objetivo de mantener la muestra lo mas homogénea posible, es importante mantener un control preciso del resto de reactivos que componen nuestro medio tamponante. Para ello, hemos realizado los experimentos de cristalización empleando *buffers* con muy baja fuerza iónica: 10 mM HEPES a pH 7.0, 10 mM HEPES pH 8.0 y 10 mM NaAc a pH 5.0. Para el muestreo primario de condiciones no hemos empleado ningún tipo de aditivo ni ligando en la solución de proteína. Aunque durante la optimización de las condiciones de cristalización, hemos añadido diferentes aditivos a la mezcla de precipitante. Estos aditivos han correspondido a sales de cationes divalentes, ya que un buen número de estructuras de tiorredoxina se han resuelto a partir de cristales crecidos en presencia de cobre (Holmgren y Soderberg 1970; Soderberg, Holmgren et al. 1974; Katti, LeMaster et al. 1990; Nikkola, Gleason et al. 1993).

4.2.1. Cristalización de tiorredoxinas precámbricas

En una primera aproximación, hemos realizado los primeros experimentos de cristalización con cuatro tiorredoxinas precámbricas: LBCA, LACA, AECA y LPBCA. Todas ellas tienen una cola de histidinas en el extremo N-terminal de la proteína. Hemos seleccionado en principio estas cuatro tiorredoxinas porque son las más estables y las más activas de todas las tiorredoxinas resucitadas (como ya hemos mencionado en el capítulo 3 del presente trabajo), por tanto el estudio de su estructura tridimensional resulta más interesante.

Como ya hemos mencionado, para cristalizar una proteína es necesario reducir su solubilidad hasta que alcanza el estado de sobresaturación. Uno de los métodos más ventajosos para alcanzar este objetivo es la contradifusión en capilares, debido a que se recorre una amplia zona del diagrama de fases en un único experimento. Esta técnica permite encontrar las condiciones de cristalización óptimas para un par proteína/agente precipitante minimizando el consumo de proteína (Otalora, Gavira et al. 2009).

Basándonos en estas premisas, la búsqueda inicial de condiciones de cristalización se llevó a cabo en capilares empleando la técnica de contradifusión. Para esta búsqueda inicial de condiciones empleamos el kit comercial de 24 condiciones (GCB-CSK 24) en las *Granada Crystallization Box* (GCB) comercializado por Triana S&T. La selección de estas 24 condiciones está basada en un amplio estudio sobre la efectividad de las 96 soluciones de precipitantes mas empleadas en la cristalización de macromoléculas (Jancarik y Kim 1991; Kimber, Vallee et al. 2003). Además, hemos empleado la contradifusión en la configuración de 3 capas con soluciones de precipitantes que han producido cristales de tiorredoxinas en el pasado (Holmgren y Soderberg 1970; Soderberg, Holmgren et al. 1974; Katti, LeMaster et al. 1990) expandida en el rango de pH 3.5 a 8.5 (estas técnicas, así como las composiciones de las condiciones empleadas, están explicadas detalladamente en el apartado 2.2.7.1 del capítulo Materiales y Métodos).

El diámetro interno de los **capilares** empleados en el muestreo primario en GCB ha sido de 0.1 ó 0.2 mm dependiendo de la cantidad disponible de proteína. En la configuración 3 capas solo hemos usado capilares de 0.1 mm de diámetro.

La concentración del gel de agarosa de la capa separadora en la configuración 3 capas ha sido de 0.5% de agarosa disuelta en NaAc 50mM pH 3.5 (p/v). En la configuración GCB, la agarosa, a una concentración de 0.5% (p/v), está equilibrada en la solución de precipitante según las especificaciones de la casa comercial. Las diferentes condiciones utilizadas en el muestreo primario de estas proteínas se resumen en la siguiente tabla:

Tabla 4.1 Parámetros del muestreo primario utilizadas para tiorredoxinas ancestrales con cola de Histidinas

Proteína	[Proteína] (mg/ml)	Buffer	Condiciones	Cap. Ø (mm)	T (°C)
LPBCA	46.5	NaAc 10mM pH 3.5	3 Capas GCB-CSK 24	0.1 0.1	4 4 y 20
LPBCA	37	Hepes 10mM pH 7.0	3 Capas GCB-CSK 24	0.1 0.1	4 4 y 20
LPBCA	37	Hepes 10mM pH 8.0	3 Capas GCB-CSK 24	0.1 0.1	4 4 y 20
LBCA	65	Hepes 10mM pH 7.0	GCB-CSK 24	0.2	4 y 20
LACA	90	Hepes 10mM pH 7.0	GCB-CSK 24	0.2	4 y 20
AECA	28	Hepes 10mM pH 8.0	GCB-CSK 24	0.2	4 y 20
AECA	32	NaAc 10mM pH 3.5	GCB-CSK 24	0.2	4 y 20

Durante el período de seguimiento del muestreo primario de las tiorredoxinas con cola de histidinas, no hemos observado ningún cristal, tan solo precipitados y separación de fases en algunas condiciones. Como podemos observar en la tabla 4.1., hemos probado numerosas condiciones en que variamos la solución de proteína (diferente concentración, composiciones de la solución tamponante, pHs, etc.), diámetro del capilar, kit de condiciones y temperatura. Todo ello, sumado a la amplia zona del diagrama de fases que barre el método de contradifusión, nos sugiere que la cristalización está impedida por la cola de histidinas de alguna manera. Por esta razón, comenzamos a emplear las tiorredoxinas sin la cola de histidinas para los sucesivos experimentos de cristalización.

Al eliminar la cola de histidinas de las proteínas por la posible interferencia en la cristalización, los ensayos iniciales realizados anteriormente no proveen ninguna información. Por tanto, comenzamos una nueva serie de ensayos de cristalización en la etapa de muestreo primario de condiciones. La purificación de las proteínas sin la cola de histidinas se detalla en el apartado 2.2.3.2. del capítulo Materiales y Métodos. Siguiendo el razonamiento del anterior, seleccionamos las tiorredoxinas LPBCA, LBCA, LACA y AECA para los ensayos iniciales. Para realizar el muestreo primario, hemos empleado solamente el método de contradifusión en capilares con la configuración de GCB. El kit empleado es el de 24 condiciones de Triana S&T con capilares de 0.1 mm de diámetro interno lo que reduce el consumo de proteína a unos 400 nL por capilar. En este caso, hemos tomado como variable adicional la concentración de tiorredoxina (Tabla 4.2). Hemos realizado los experimentos por duplicado, almacenando un kit a 4°C y otro a 20°C. Al emplear tiorredoxinas sin cola de histidinas, el resultado del muestreo primario tras el seguimiento de los experimentos es positivo, ya que obtenemos agregados cristalinos y cristales pequeños en algunas condiciones.

Debido a que las condiciones donde hemos obtenido cristales presentan como agente precipitante PEG, y que se han obtenido cristales de tiorredoxina empleando este agente precipitante anteriormente (Nikkola, Gleason et al. 1993; Schultz, Chivers et al. 1999), decidimos ampliar el muestreo primario con dos kits comerciales de Triana S&T: Mix-PEG y PEG1500 con un rango de pHs de 4 a 9 (su composición se detalla en el apartado 2.2.7.1. del capítulo de Materiales y Métodos). Empleando estos nuevos kits, hemos procedido a realizar el muestreo primario de la misma manera que hemos descrito en el párrafo anterior.

Como resultado del muestreo primario, obtuvimos cristales pequeños de LPBCA y LACA, así como agregados cristalinos de LPBCA, LBCA, LACA y AECA. Sin embargo, los cristales no eran de suficiente calidad (por inspección óptica) para ser difractados, aunque sí eran lo suficientemente grandes para realizar un test de difracción (apartado 2.2.7.2. del capítulo Materiales y Métodos). Realizamos el test de difracción con el equipo de rayos-X de BRUKER a un cristal de LPBCA sin extraerlo del capilar. El resultado del test fue un patrón de difracción típico de cristales de proteína.

Tras obtener resultados positivos con esta estrategia de cristalización, decidimos aplicar la misma secuencia al resto de tiorredoxinas precámbricas: LECA, LGPCA y LAFCA. Los parámetros de muestreo primario se resumen en la siguiente tabla:

Tabla 4.2. Parámetros del muestreo primario utilizadas para tiorredoxinas ancestrales sin cola de Histidinas con el método de contradifusión

Proteína	[Proteína] (mg/ml)	Buffer	Kit de condiciones	Cap. Ø (mm)	T (°C)
LPBCA	50	Hepes 10mM pH 7.0	GCB-CSK 24 Mix-PEG448 PEG1500	0.1	4 y 20
LBCA	59	Hepes 10mM pH 7.0	GCB-CSK 24 Mix-PEG448 PEG1500	0.1	4 y 20
LACA	85	Hepes 10mM pH 7.0	GCB-CSK 24 Mix-PEG448 PEG1500	0.1	4 y 20
AECA	50	Hepes 10mM pH 7.0	GCB-CSK 24 Mix-PEG448 PEG1500	0.1	4 y 20
LECA	40	NaAc 10mM pH 5.5	GCB-CSK 24 Mix-PEG448 PEG1500	0.1	4 y 20
LGPCA	30	NaAc 10mM pH 5.5	GCB-CSK 24 Mix-PEG448 PEG1500	0.1	4 y 20
LAFCA	104	NaAc 10mM pH 5.5	GCB-CSK 24 Mix-PEG448 PEG1500	0.1	4 y 20

Trascurrido el tiempo apropiado para el crecimiento de los cristales (diferente para cada tiorredoxina), obtenemos cristales de las tiorredoxinas precámbricas AECA, LACA, LPBCA, LAFCA y LGPCA en diferentes condiciones. Para las proteínas LPBCA y AECA han crecido cristales tanto a 20° como a 4°C es las condiciones señaladas (Tabla 4.4.) Sin embargo, para el resto de proteínas solo han crecido cristales a 4°C. Basándonos en los resultados de los experimentos con el kit Mix-PEGs y

PEG1500 se puede deducir que todas las proteínas deben de ser suficientemente solubles a pH 4.0 ya que todos los capilares están límpidos.

La tiorredoxina LBCA no cristalizó en ninguna condición, solo observamos agregados cristalinos o separaciones de fases que en ningún caso culminaron en el crecimiento de un cristal. En el caso de la tiorredoxina LECA, obtuvimos un único cristal tipo *rod* (cristal que ocupa todo el diámetro interno del capilar) en la condición 16 del kit de 24 condiciones, pero no se pudo reproducir el resultado en los capilares colocados posteriormente.

Para tratar de cristalizar estas dos proteínas hemos cambiado de estrategia de cristalización. En esta ocasión, hemos utilizado un robot de cristalización que emplea el método de difusión de vapor en modo gota sentada (apartado 2.2.7.1 de Materiales y Métodos). Este sistema nos permite testear un elevado número de condiciones diferentes al mismo tiempo, empleando muy poco volumen de muestra. Hemos usado dos kits comerciales de 48 condiciones para el muestreo primario: PEGRx-1 Screen™ y PEG/Ion screen™ (*Hampton Research*). La composición de las condiciones de estos kits se puede encontrar en www.hamptonresearch.com. Hemos optado por estos kits, basados en diferentes disoluciones de polietilenglicol, porque todas las demás tiorredoxinas ancestrales han cristalizado en soluciones que contienen este agente precipitante. Para formar la gota de 0.4 µL, hemos mezclado 0.2 µL de proteína con 0.2 µL de agente precipitante y hemos rellenado el reservorio con 50 µl de agente precipitante. El experimento lo realizamos por duplicado y las placas se colocan a 4°C y 20°C. La tabla 4.3. muestra un resumen de los experimentos de cristalización empleando este método.

Tabla 4.3. Parámetros del muestreo primario utilizadas para tiorredoxinas ancestrales sin cola de Histidinas con el método difusión de vapor.

Proteína	[Proteína] (mg/ml)	Buffer	Condiciones	Gota (µL)	T (°C)
LECA	30	NaAc 10mM pH 5.5	PEG Rx 1+ PEG/Ion	0.2/0.2	4 y 20
LECA	58	NaAc 10mM pH 5.5	PEG Rx 1+ PEG/Ion	0.2/0.2	4 y 20
LBCA	50	NaAc 10mM pH 5.5	PEG Rx 1+ PEG/Ion	0.2/0.2	4 y 20
LBCA	89	NaAc 10mM pH 5.5	PEG Rx 1+ PEG/Ion	0.2/0.2	4 y 20

Durante el seguimiento del crecimiento cristalino observamos cristales de ambas proteínas en varias condiciones. De la proteína LBCA solamente obtuvimos cristales en una condición del kit PEGRx a 20°C usando la solución más concentrada de proteína (89 mg/ml) (Tabla 4.4). Cabe destacar que esta condición no se ha probado en ningún kit anterior. Respecto a la proteína LECA obtuvimos cristales en dos condiciones del kit PEGRx y en una condición del kit PEG/ion (Tabla 4.4) a 20°C. Empleando la disolución más diluida observamos cristales en la condición 3 (0.2 M Fluoruro de amonio, 20% PEG 3500) del kit PEG/ion a 20°C que no aparecen cuando usamos la solución más concentrada de proteína. Como en el caso de la tiorredoxina LBCA, ninguna de estas condiciones ha sido probada anteriormente. La tabla 4.4 resume las todas las condiciones donde hemos obtenido cristales de tiorredoxinas ancestrales.

Tabla 4.4. Tabla resumen con las condiciones de cristalización donde hemos obtenido cristales de tiorredoxinas ancestrales en la etapa de muestreo primario.

Proteína	Condición	Kit	T (°C)
LPBCA	- 30% PEG 1500	GBC-CSK 24	20 y 4
	- NaAc 0.1M, PEG 1500 30%, pH 5.0	PEG 1500	20 y 4
	- Tris-HCl 0.1M, PEG 1500 30%, pH 6.0		
	- NaAc 0.1M, PEG 400 20%, PEG 4000 15%, PEG 8000 10%; pH 5.0	MixPEGs	20 y 4
- Tris-HCl 0.1M, PEG 400 20%, PEG 4000 15%, PEG 8000 10%; pH 6.0			
LBCA	- 0.1M Ácido Cítrico, PEG 1000 14%; pH 3.5	PEG-Rx	20
LACA	- Tris-HCl 0.1M, PEG 400 20%, PEG 4000 15%, PEG 8000 10%; pH 9.0	MixPEGs	4
AECA	- Tris-HCl 0.1M, PEG 400 20%, PEG 4000 15%, PEG 8000 10%; pH 9.0	MixPEGs	20 y 4
LECA	-30% PEG 400, 0.2M CaCl ₂ , 0.1M Hepes pH 7.50	GBC-CSK 24	20
	- 0.2M Fluoruro Sódico, PEG 3350 20%	PEG-Ion	20
	- 0.2M Fluoruro Amónico, PEG 3350 20%		
	- 0.1M Tris pH 8.0, <i>Jeffamine M600 30%; pH 7.0</i>	PEG-Rx	20
- 0.1M Tris pH 8.0, PEG Monometil Éter 2000 30%			
LGPCA	- 30% PEG 4000, 0.2M MgCl ₂ , 0.1M Tris-HCl pH 8.50	GBC-CSK 24	4
	- 30% PEG 400, 0.2M Na Citrato, 0.1M Tris-HCl pH 8.50		
	- 25% PEG 4000, 0.2M NH ₄ Sulfato, 0.1M AcNa pH 4.60		
LAFCA	- 30% PEG 4000, 0.2M MgCl ₂ , 0.1M Tris-HCl pH 8.50	GBC-CSK 24	4
	- 30% PEG 4000, 0.2M AcNH ₄ , 0.1M NaAc pH 4.60		

Los cristales de tiorredoxina obtenidos por contradifusión en capilares de 0.1 mm son, generalmente, inadecuados para un experimento de difracción al ser muy pequeños. Por esta razón, hemos llevado a cabo una etapa de optimización de las condiciones para mejorar la calidad y el tamaño de los cristales. Respecto a los

cristales obtenidos por difusión de vapor, los cristales de LECA eran de buena calidad y bastante grandes (por inspección óptica), así que los preservamos en nitrógeno líquido hasta la toma de datos. Por el contrario, los cristales obtenidos de LBCA eran pequeños y fue necesario un proceso de optimización de las condiciones.

4.2.2. Optimización de condiciones de cristalización

Para mejorar los cristales de las tiorredoxinas ancestrales, hemos usado el método de contradifusión en capilares con la configuración GBC. Para optimizar las condiciones con este método existen varios parámetros que podemos alterar.

Respecto a la composición de las condiciones de los kits, hemos empleado las mismas que en el muestreo primario: CSK 24, Mix-PEGs y PEG1500. El kit de 24 condiciones y el kit PEG1500 los hemos usado sin ninguna alteración, variando otros parámetros como la concentración de proteína o el diámetro interno del capilar. Sin embargo, en el kit Mix-PEGs hemos ampliado el rango de pH, de 4.5 a 10 en incrementos de pH de 0.5 unidades. Aparte de estos kits, hemos utilizado otros fabricados "a la carta". Uno de ellos es el Mix-PEG con 1mM de Cobre o 1mM de Zinc a un pH de 4 a 10, debido a que obtuvimos cristales de algunas proteínas en condiciones que contenían estos iones. Para optimizar los cristales de la proteína LBCA, hemos realizado un muestreo a distintos pHs fabricando un kit compuesto de 15% de PEG 1500, 10 mM Ácido cítrico de pH 3 a 7. Por otro lado, hemos sustituido los capilares de 0.1 o 0.2 de diámetro interno por los capilares de 0.3 mm con el objetivo de producir cristales de mayor tamaño. Además, en algunos casos, hemos aumentado la concentración de proteína para que haya más cantidad disponible para formar cristales. Para evitar el posible deterioro de la proteína, así como la heterogeneidad de la muestra, en la optimización de los cristales hemos partido de tiorredoxinas recién purificadas.

Tras el proceso de optimización de las condiciones, hemos obtenidos cristales de suficiente calidad y tamaño (por inspección óptica) para ser difractados. La tabla 4.5. resume las condiciones que han producido los cristales de tiorredoxinas precámbricas empleados en la resolución de las estructuras.

Tabla 4.5. Resumen de las condiciones donde han crecido los cristales de tiorredoxinas precámbricas empleados para la resolución de las estructuras. Además se muestra información relevante de cada tiorredoxina precámbrica.

Proteína	LPBCA	AECA	LACA	LAFCA	LGPCA	LECA	LBCA
Información:							
Nº aminoácidos	105	105	105	105	105	106	105
Peso molecular (Da)	11990.9	12019.0	12018.1	11894.2	11711.5	12034.0	12044.9
pI teórico	5.27	5.10	6.62	4.67	4.69	4.86	4.80
Cristalización							
[Proteína] (mg/mL)	56.2	71.0	84.7	30.0	104.0	58.0	60.0
Buffer proteína	10 mM HEPES pH 7.0	10 mM NaAc pH 5.5	10 mM HEPES pH 7.0	10 mM NaAc pH 5.5	10 mM NaAc pH 5.5	10 mM NaAc pH 5.5	10 mM NaAc pH 5.5
Método	Contradifusión GCB	Contradifusión GCB	Contradifusión GCB	Contradifusión GCB	Contradifusión GCB	Difusión Vapor Robot	Contradifusión GCB
Condición	30% PEG 1500	25% PEG1500 pH 9.0	Tris-HCl 0.1M, 20% PEG 400, 15% PEG 4000, 10% PEG 8000; pH 9.0	0.1M Tris-HCl, 30% PEG 4000, 0.2 M MgCl ₂ , pH 8.5	30% PEG 400, 0.1 M Tris-HCl, 0.2 M Citrato sódico, pH 8.5	Acetato Magnesio tetrahidratado, 20% PEG 3350	15% PEG 1500, 0.01 M Ácido Cítrico, pH 3.0
Temperatura (°C)	4	4	4	4	4	20	20

4.2.3. Recogida de datos de difracción

Los primeros cristales que obtuvimos en la condición 6 (30% PEG 1500) a 4°C del kit de 24 condiciones pertenecían a la tiorredoxina LPBCA, que es la tiorredoxina ancestral más estable (121°C de T_m). Hemos seleccionado tres cristales de buena calidad por inspección óptica, los hemos extraído del capilar y los hemos sumergido en solución crioprotectora, compuesta de 90% de solución madre (30% PEG 1500) y 10% de glicerol. A continuación, los hemos sobre-enfriado en nitrógeno líquido. Uno de los cristales lo difractamos en la fuente de rayos X BRUKER siguiendo el protocolo explicado en Materiales y Métodos. Obtuvimos un juego de datos de difracción hasta una resolución de 2.5 Å. Tras el análisis de datos, el cristal pertenece al grupo espacial monoclinico (P1 21 1) y los parámetros de la celdilla unidad son: $a=55\text{Å}$, $b=30\text{Å}$, $c=59\text{Å}$, $\alpha=90^\circ$, $\beta=116^\circ$, $\gamma=90^\circ$. Los otros dos cristales, de mejor calidad, los almacenamos en un vaso criogénico *Dewar* y los enviamos al ESRF. Los experimentos de difracción de dichos cristales los llevamos a cabo en la línea ID14-1 del ESRF. De uno de ellos, obtuvimos un juego de datos completo hasta una resolución de 1.6 Å. La tabla 4.6. muestra los datos estadísticos de la difracción. El cristal pertenece al grupo espacial monoclinico P1 21 1 y unos parámetros de celdilla unidad similares a los obtenidos en el fuente de rayos-X BRUKER. Cabe destacar que hemos obtenido datos a mejor resolución en el ESRF empleando cristales de la misma condición de cristalización y con el mismo tratamiento.

La tiorredoxina AECA produce cristales en la condición 25% PEG1500 pH 9.0 a 4°C del kit Mix-PEG de optimización. En esta ocasión, no extraemos los cristales y lo difractamos directamente dentro del capilar. Como tenemos varios capilares con cristales, seleccionamos uno de ellos y lo difractamos en la fuente de rayos-X BRUKER. Tras el experimento, resulta un juego de datos hasta 4 Å, lo que significa que tenemos que intentar recoger mejores datos. Por ello, guardamos varios capilares cortados y colocados en el soporte con pegamento de contacto; tras lo cual, lo sumergimos en nitrógenos líquido y lo almacenamos en un *Dewar*. El cristal en el interior del capilar lo difractamos en la línea ID23-2 del ESRF. Los cristales difractan a una resolución máxima de 2.6 Å. El cristal pertenece al grupo espacial monoclinico (P1 21 1), aunque tiene diferentes parámetros de celdilla unidad que LPBCA (Tabla 4.6.). Es relevante señalar que tiene 3 monómeros en la unidad asimétrica, algo que no ocurre con ninguna otra tiorredoxina ancestral.

La tiorredoxina LACA produce cristales en la condición Tris-HCl 0.1M, 20% PEG 400, 15% PEG 4000, 10% PEG 8000 a pH 9.0. En esta ocasión, extraemos los cristales del capilar pero no utilizamos crioprotector. Los cristales los congelamos en nitrógeno líquido y los almacenamos en un *Dewar* para difractarlos en el ESRF. La línea utilizada es la ID23-2, igual que en el caso de AECA. En esta ocasión, el cristal difracta hasta una resolución máxima de 1.75 Å y el grupo espacial es triclinico (P1).

En la condición 1 (30% PEG 4000, 0.2M Cloruro de Magnesio, 0.1M Tris-HCl, pH 8.50) del kit de 24 condiciones, obtenemos cristales de la tiorredoxina LAFCA. Extraemos los cristales del capilar, los congelamos directamente en nitrógeno líquido y los almacenamos en un *Dewar*. Hemos llevado a cabo el experimento de difracción en la línea BM16 de ESRF. Conseguimos un juego de datos hasta una resolución máxima de 2.2 Å. El grupo espacial al que pertenece el cristal es ortorrómbico (P212121) y existe un monómero en la unidad asimétrica.

Los cristales seleccionados de tiorredoxina LGPCA pertenecen a la condición 15 (30% PEG 400, 0.1 M Tris-HCl, 0.2 M Citrato sódico, pH 8.50) del kit de 24 condiciones. Hemos mantenido los cristales dentro del capilar durante el experimento de difracción en la línea ID29 del ESRF. El cristal pertenece al grupo espacial monoclinico (P1 21 1) con dos monómeros en la unidad asimétrica.

Respecto a la tiorredoxina LECA, recurrimos al método de difusión de vapor para cristalizarla. En la condición acetato magnesio tetrahidratado, 20% PEG 3350 del muestreo primario creció un cristal de buena calidad, identificado por inspección óptica. Tomamos el cristal con ayuda de un lazo, lo congelamos directamente en nitrógeno líquido y lo almacenamos en un *Dewar*. La difracción del cristal la realizamos en la línea ID23-1 del ESRF. El grupo espacial al que pertenece el cristal es monoclinico (C2) y tiene 2 monómeros en la unidad asimétrica.

Por último, los cristales seleccionados para la difracción de la tiorredoxina LBCA son de la condición optimizada: 15% PEG 1500, ácido cítrico 0.1M pH 3.0. El cristal lo extrajimos del capilar, lo congelamos en nitrógeno líquido y lo almacenamos en un *Dewar* para transportarlo al ESRF. La línea utilizada para los experimentos de difracción fue la ID23-1. El cristal pertenece al grupo espacial cúbico (F 4 3 2) y tiene dos monómeros en la unidad asimétrica.

Cabe destacar que los experimentos de difracción descritos anteriormente son aquellos que han generado los juegos de datos que se emplearon para resolver las estructuras de las diferentes proteínas ancestrales. Muchos de los cristales se han difractado en diferentes líneas y a partir de otras condiciones de difracción e incluso hemos empleado diferentes *stocks* de proteína. Sin embargo, algunos juegos de datos puede que no estuvieran suficientemente completos a una resolución aceptable, hemos obtenidos datos a baja resolución, el cristal decaía rápidamente o el tratamiento de crioprotección lo había destruido.

Una vez analizados los distintos juegos de datos y averiguado los parámetros de la celdilla unidad y el grupo espacial, los hemos integrado, indexado y escalado. La tabla 4.6. recoge únicamente las estadísticas del proceso de difracción de aquellos juegos de datos que se emplearon para resolver la estructura que presentaban la mejor resolución y calidad de los datos.

Es destacable que cada tiorredoxina ancestral ha cristalizado con un polimorfismo diferente, ya que aunque los cristales de las proteínas LPBCA, AECA, LGPCA y LECA pertenecen al grupo espacial monoclinico todas presentan celdillas unidades diferentes. El cristal de LACA es triclinico, el de LAFCA es ortorrómbico y el de LBCA es cúbico.

Las tiorredoxinas LECA, LACA, LPBCA, LGPCA y LBCA tienen 2 monómeros en la unidad asimétrica (ASU) según el cálculo del coeficiente de Matthews, estas se denominan cadena A y B. En el caso de LAFCA solo hay 1 monómero en la unidad asimétrica y AECA tiene 3, denominadas A, B y C. Los juegos de datos obtenidos a mejor resolución son de los cristales de LGPCA y LECA, ambas con 1.3 Å. La de peor resolución es AECA con 2.6 Å.

Todos los cristales, excepto el de la tiorredoxina LPBCA se han congelado sin utilizar crioprotector adicional y han difractado correctamente. Los juegos de datos de las proteínas AECA y LGPCA se obtuvieron por difracción del cristal en el interior del capilar.

Tabla 4.6. Tabla resumen con las estadísticas de la recogida de datos de difracción.

	LPBCA	AECA	LACA	LAFCA	LGPCA	LECA	LBCA
Difracción rayos-X							
Fuente radiación	Sincrotrón (ESRF)	Sincrotrón (ESRF)	Sincrotrón (ESRF)	Sincrotrón (ESRF)	Sincrotrón (ESRF)	Sincrotrón (ESRF)	Sincrotrón (ESRF)
Temperatura (K)	100	100	100	100	100	100	100
Línea	ID14-1	ID23-2	ID23-2	BM16	ID29	ID23-1	ID23-1
λ (Å)	0.930	0.873	0.873	0.980	0.976	0.973	0.973
Detector	CCD ADSC Quantum 210	CCD MAR225	CCD MAR225	CCD ADSC Quantum 210	Pilatus	ADSC Quantum 315r	ADSC Quantum 315r
Datos de difracción							
Grupo espacial	P 1 21 1	P1 21 1	P 1	P 21 21 21	P 1 21 1	C 1 2 1	F 4 3 2
Celdilla unidad							
a, b, c (Å)	55.1 30.2 59.0	37.5 48.7 91.0	32.2 36.3 48.1	37.5 42.8 55.9	36.1 62.9 42.9	58.4 47.8 73.8	193.0, 193.0, 193.0
β (°)	117.0	93.2	108.0	90.0	109.0	98.5	90.0
Resolución (Å)	48.45 - 1.65 (1.74 - 1.65)	45.51 - 2.65 (2.74 - 2.65)	45.32 - 1.75 (1.84 - 1.75)	20.00 - 2.20 (2.28 - 2.20)	34.09 - 1.30 (1.35 - 1.30)	36.60 - 1.30 (1.37 - 1.30)	58.10 - 2.45 (2.53 - 2.45)
R_{sym} (%)	4.5 (34.1)	6.5 (59.9)	10.5 (39.6)	10.3 (48.7)	5.6 (42.8)	4.7 (6.6)	13.0 (93.4)
I/σ_I	20.10 (3.80)	16.21 (3.24)	8.50 (2.70)	23.26 (4.95)	12.47 (2.68)	18.40 (7.90)	12.92 (4.18)
Complejidad (%)	99.8 (100.0)	97.9 (97.9)	97.6 (88.7)	99.0 (100.0)	97.2 (97.7)	94.4 (89.8)	99.9 (100.0)
Reflexiones únicas	21198	9592	18887	4904	43308	46574	11808
Multiplicidad	3.6 (3.7)	4.0 (4.0)	1.0 (2.9)	10.4 (10.7)	2.8 (2.8)	2.0 (1.8)	7.7 (8.0)

Entre paréntesis se muestran los datos de alta resolución

Estos resultados evidencian el elevado polimorfismo de los cristales de tiorredoxina, ya que no se observan coincidencias de las interacciones intermoleculares cuando se superponen los diferentes empaquetamientos cristalinos. El polimorfismo es un efecto habitualmente observado en la cristalización de proteínas. Se encuentra inducido por el cambio de las condiciones químico-físicas que se emplean durante el rastreo de condiciones de cristalización. Todas las tiorredoxinas precámbricas han cristalizado en un grupo espacial y parámetros de celdilla unidad diferentes al de otras estructuras de tiorredoxina previamente cristalizadas. Sin embargo, como veremos más adelante, aun obteniendo diversos polimorfismos, el plegamiento es idéntico es todas ellas y es el correspondiente a la familia de las tiorredoxinas.

4.2.4. Resolución del modelo estructural

Al igual que en los apartados anteriores, la estrategia ha sido analizar los datos obtenidos en el ESRF (a mayor resolución) de la proteína LPBCA. Para solucionar el problema de las fases, hemos utilizado el método de reemplazamiento molecular (explicado en detalle en el apartado 2.2.7.3. del capítulo de Materiales y Métodos). La búsqueda de la solución la hemos llevado a cabo empleando el programa *MolRep* que pertenece al paquete de programas CCP4 (Collaborative Computational Project 1994).

En el caso de las tiorredoxinas LECA, LGPCA y LACA hemos encontrado solución de reemplazamiento empleando la tiorredoxina de *E. coli* (PDB: 2TRX). Con la tiorredoxina LAFCA, hemos encontrado solución con la tiorredoxina humana (PDB: 1ERV), algo que parece razonable, ya que está próxima filogenéticamente a ella. Para las tiorredoxinas LBCA y AECA, ninguna de estas tiorredoxinas actuales generaba una solución válida, por lo que probamos la tiorredoxina de LACA como modelo de reemplazamiento, resultando la selección adecuada.

El refinamiento lo llevamos a cabo con el programa *Phenix* como se ha descrito en el apartado 2.2.7.4 del capítulo Materiales y Métodos. Entre cada ciclo de refinamiento, la reconstrucción manual la realizamos con el programa gráfico *COOT*. Para posicionar los aminoácidos y las cadenas laterales hemos utilizado los mapas de densidad electrónica $|2F_0 - F_c|$ contorneados a 1σ y los mapas $|F_0 - F_c|$ contorneados a 3σ . El resumen de los datos y estadísticas del refinamiento se recoge en la tabla 4.7.

Los mapas de densidad electrónica han sido siempre continuos para la cadena principal y la mayoría de las cadenas laterales. En caso de ausencia de densidad electrónica de las cadenas laterales, hemos cortado átomo a átomo la cadena y hemos comprobado su desaparición tras un nuevo ciclo de refinamiento.

La conformación del puente disulfuro, que conecta los residuos CysXXCys siempre ha quedado claramente definida en mapa $|2F_0 - F_c|$ y $|F_0 - F_c|$. En algunos casos, y debido a un efecto comprobado de la radiación, el puente disulfuro se ha roto parcialmente y esta composición se ha modelado al 30 o 50% de ocupación. Las moléculas de agua las hemos añadido automáticamente en el *COOT* con los parámetros por defecto y las hemos refinado con *Phenix*, tras lo cual hemos realizado una comprobación visual.

Tabla 4.7. Tabla resumen con las estadísticas del proceso de construcción y el refinamiento del modelo estructural.

Proteína	LPBCA	AECA	LACA	LAFCA	LGPCA	LECA	LBCA
Código PDB	2YJ7		2YNX	2YPM	2YN1	2Y01	4BA7
Modelo reemplazamiento	2TRX	LACA	2TRX	1ERV	2TRX	2TRX	LACA
Refinamiento							
Resolución (Å)	48.45 - 1.65	45.51 - 2.65	45.32 - 1.75	20.00 - 2.20	34.09 - 1.30	36.60 - 1.30	58.10 - 2.45
R / R _{free} (%)	18.19/21.98	21.83/28.08	15.76/21.52	16.65/23.92	16.76/20.42	15.43/18.47	15.83/21.08
No. átomos	2117	2494	3934	918	4121	4371	3764
Proteína	1852	2469	1817	866	1825	1954	1819
Aguas	264	25	184	52	357	433	69
Ligandos	1	0	2	0	2	7	8
Media factores B(Å ²)	21.80	66.60	19.80	31.00	18.30	16.20	44.20
R.M.S.D							
Longitud enlaces (Å)	0.007	315.000	0.013	0.009	0.011	0.011	0.015
Ángulo enlaces (°)	0.970	0.005	1.390	1.090	1.350	1.290	1.480
Ramachandran (%)							
Valores permitidos	100.0	98.0	98.0	99.0	100.0	98.0	100.0
Valores atípicos	0.0	0.0	0.0	0.0	0.0	0.0	0.0

Los modelos estructurales de tiorredoxinas ancestrales siguen la topología canónica del plegamiento de tiorredoxina (Katti, LeMaster et al. 1990) (un ejemplo se muestra en la Fig. 4.2). Cada una de las cadenas está formada por un dominio N-terminal tipo $\beta_1\alpha_1\beta_2\alpha_2\beta_3\alpha_3$ y un dominio C-terminal $\beta_4\beta_5\alpha_4$ distribuidos en un núcleo central de tres láminas paralelas y dos antiparalelas, rodeadas por cuatro hélices alfa. Los extremos N-terminal y C-terminal se mantienen unidas mediante enlaces de hidrógeno entre las láminas β_4 y β_2 , quedando alineadas de manera antiparalela en el conjunto de la molécula. Las hélices α_2 y α_4 se encuentran a un lado de las láminas β y las hélices α_1 y α_3 al otro lado (Fig. 4.3).

Las tiorredoxinas ancestrales conservan el centro activo de reconocimiento del sustrato, formado por la secuencia Cys-Gly-Pro-Cys (Katti, LeMaster et al. 1990), en el cual las dos cisteínas forman el puente disulfuro responsable de la actividad oxido-reductasa. Al igual que en otras tiorredoxinas, el centro activo está rodeado mayormente por residuos hidrofóbicos que están conservados en la mayoría de tiorredoxinas ancestrales, aunque en las posiciones 29 y 35 (numeración de la proteínas ancestrales) encontramos algunas variaciones. El residuo Asp 25, que es importante en la actividad oxido-reductasa (Chivers, Prehoda et al. 1997; Chivers y Raines 1997), también está conservado en todas las estructuras.

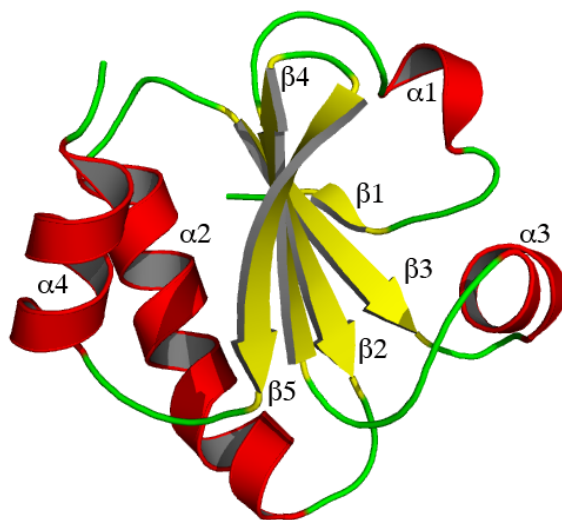


Figura 4.2. Esquema de la estructura de tiorredoxina ancestral LPBCA (cadena A de PDB: 2YJ7). Se muestran en rojo las hélices alfa, en amarillo las hojas beta. La asignación de estructura secundarias se ha realizado con DSSP.

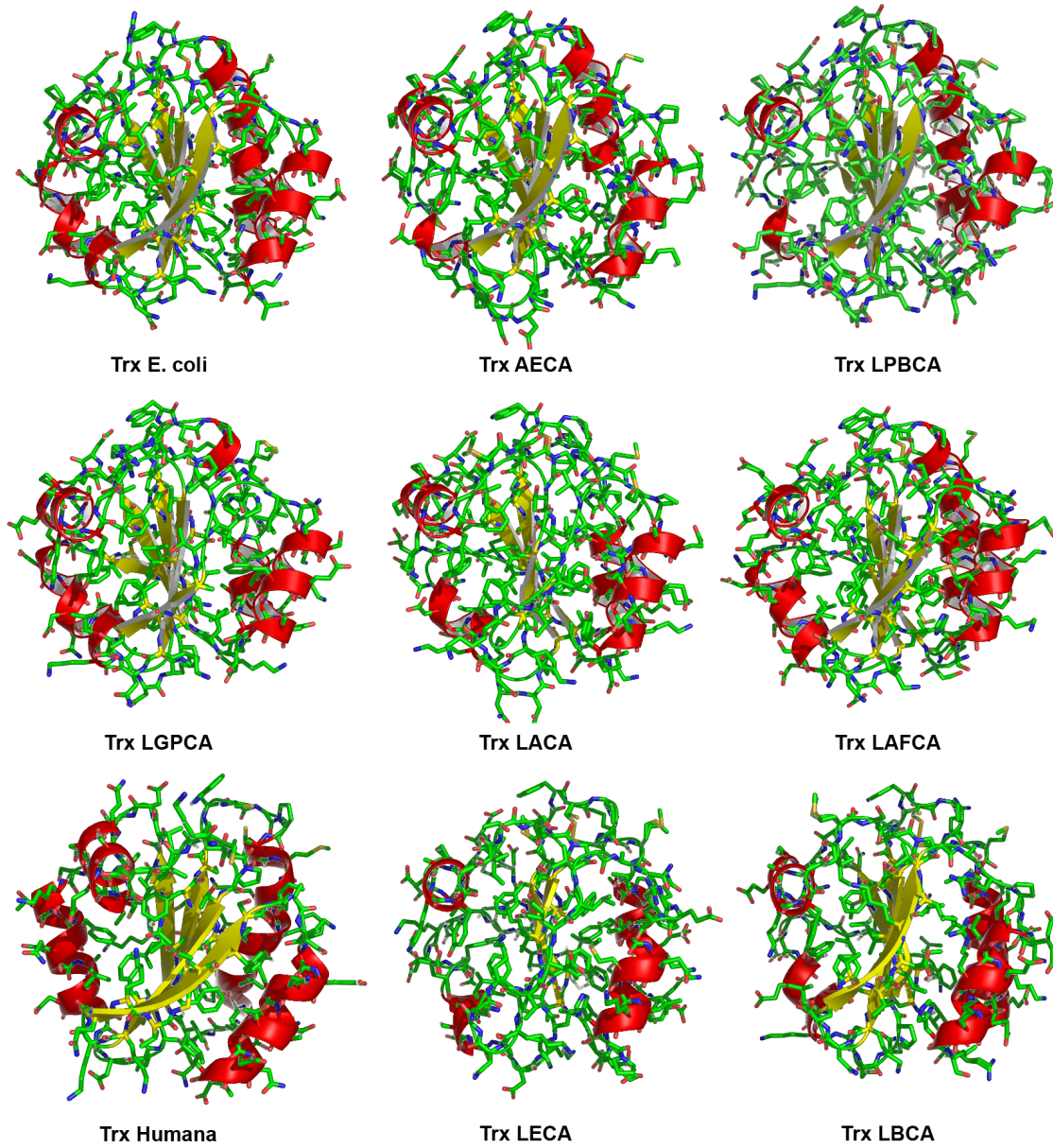


Figura 4.3. Modelos estructurales de las tiorredoxinas precámbricas resucitadas obtenidas en este trabajo y las tiorredoxinas actuales de *E. coli* (PBD:2TRX) y Humana (PDB:1ERU). En rojo se muestra las hélices alfa y en amarillos las hojas beta, ambos elementos asignados mediante el programa DSSP.

4.3. Análisis de la evolución de la estructura de tiorredoxinas a lo largo de 4 mil millones de años

Una vez obtenidos los modelos estructurales de las tiorredoxina ancestrales (Fig. 4.3) nos hemos planteado extender el uso de la resurrección de proteínas precámbricas para tratar de comprender la evolución de las estructuras de las proteínas. Con este objetivo, hemos realizado un análisis comparativo de las estructuras de las tiorredoxinas precámbricas resueltas por nuestro grupo (datadas en hasta ~4 mil millones de años) junto a las tiorredoxinas actuales de *E. coli* y humana.

Al realizar la superposición de las diferentes estructuras de tiorredoxinas (Fig. 4.4) podemos apreciar que todas las tiorredoxinas ancestrales tienen la topología típica de plegamiento de las tiorredoxinas (Fig. 4.2). Es más, al comparar las diferentes proteínas ancestrales entre sí o al compararlas con las tiorredoxinas actuales de *E.coli* (PDB:2TRX) y humana (PDB:1ERU) no hemos observado diferencias relevantes en términos de organización espacial global.

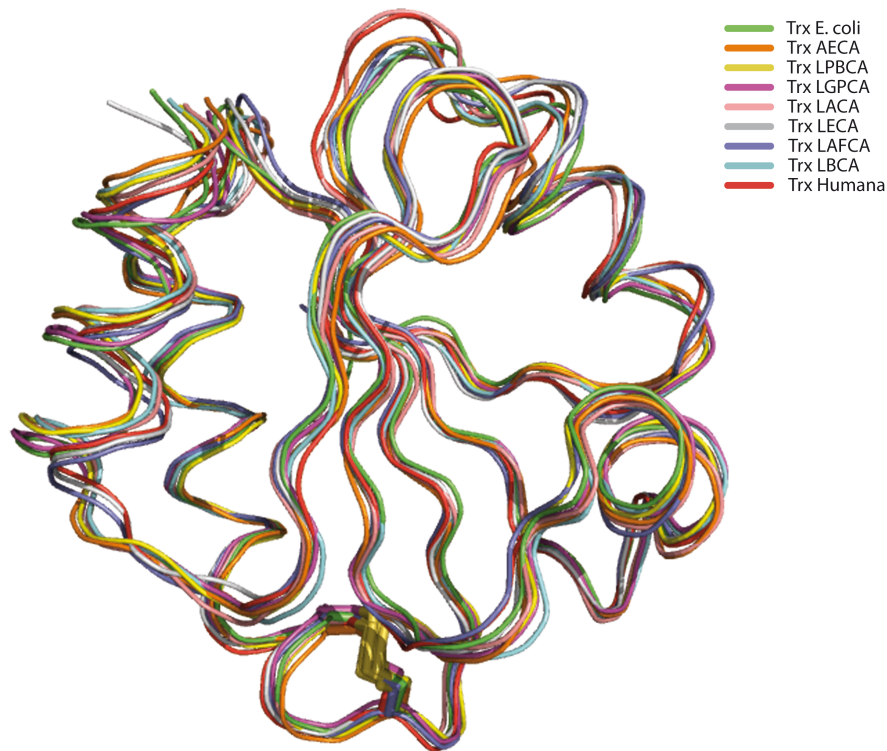


Figura 4.4. Superposición de las cadenas polipeptídicas principales (cadena A) de estructuras de tiorredoxinas ancestrales y la tiorredoxina humana y de *E. coli*. En dorado (parte inferior) se representa el puente disulfuro presente en todas las tiorredoxinas.

Para comparar las estructuras de las tiorredoxinas más detalladamente, hemos calculado la desviación cuadrática media (R.M.S.D) entre los carbonos alfa del esqueleto polipeptídico (Tabla 4.8) a partir de la superposición de las cadenas A de las estructuras. Estos cálculos los hemos realizado con el programa *DALI Pairwise Comparison* (apartado 2.2.8.2 del capítulo de Materiales y Métodos).

Tabla 4.8. Matriz de valores de R.M.S.D (en Å) de la cadena polipeptídica principal (cadena A) de tiorredoxinas ancestrales, humana y de *E. coli*.

Trx	LBCA A	LPBCA A	LGPCA A	<i>E.coli</i> A	AECA A	LACA A	LECA A	LAFCA	Humana
LBCA A		0.6	0.9	0.8	0.9	1.0	1.1	1.2	1.2
LPBCA A	0.6		0.8	0.7	0.9	1.0	1.1	1.2	1.2
LGPCA A	0.9	0.8		0.7	1.2	1.3	1.2	1.2	1.2
<i>E.coli</i> A	0.8	0.7	0.7		1.1	1.3	1.2	1.2	1.3
AECA A	0.9	0.9	1.2	1.1		0.8	1.2	1.3	1.2
LACA A	1.0	1.0	1.3	1.3	0.8		1.1	1.3	1.0
LECA A	1.1	1.1	1.2	1.2	1.2	1.1		0.5	0.7
LAFCA A	1.2	1.2	1.2	1.2	1.3	1.3	0.5		0.8
Humana	1.2	1.2	1.2	1.3	1.2	1.0	0.7	0.8	

Las similitudes de la estructura se han confirmado por los bajos valores de desviación cuadrática media. De hecho, el mayor valor lo encontramos entre las tiorredoxinas humana y de *E. coli* (1.3 Å), que es similar al valor que obtenemos al comparar la tiorredoxina humana y cualquier tiorredoxina de la rama procariota del árbol filogenético (Fig. 4.1) o las tiorredoxinas más antiguas (LBCA, LACA, AECA, LPBCA). Cuando comparamos las tiorredoxinas de la rama procariota entre sí, los valores de R.M.S.D. son similares al valor obtenido al comparar la cadena A y B de la estructura de tiorredoxina de *E. coli* (0.7 Å) y, por tanto, pueden ser consideradas idénticas.

Estos resultados están en concordancia con la identidad de secuencia al comparar las diferentes tiorredoxinas entre sí (Tabla 4.9). Es remarcable mencionar que al representar el porcentaje de identidad de secuencia frente a el R.M.S.D. comparando las tiorredoxinas actuales de *E. coli* y humana con las tiorredoxinas ancestrales, hemos observado una clara correlación (Fig. 4.5).

Tabla 4.9. Matriz de valores de identidad de secuencia (en %) entre las diferentes tiorredoxinas ancestrales y actuales.

TRX	LBCA A	LPBCA A	LGPCA A	<i>E.coli</i> A	AECA A	LACA A	LECA A	LAFCA	Humana
LBCA A		87	68	56	85	77	57	51	36
LPBCA A	87		69	57	76	73	53	48	34
LGPCA A	68	69		83	60	57	45	42	31
<i>E.coli</i> A	56	57	83		52	52	38	36	26
AECA A	85	76	60	53		92	61	56	41
LACA A	77	73	57	52	92		56	51	35
LECA A	57	53	45	38	59	56		86	55
LAFCA	51	48	42	36	54	51	86		58
Humana	36	34	31	26	38	35	55	58	

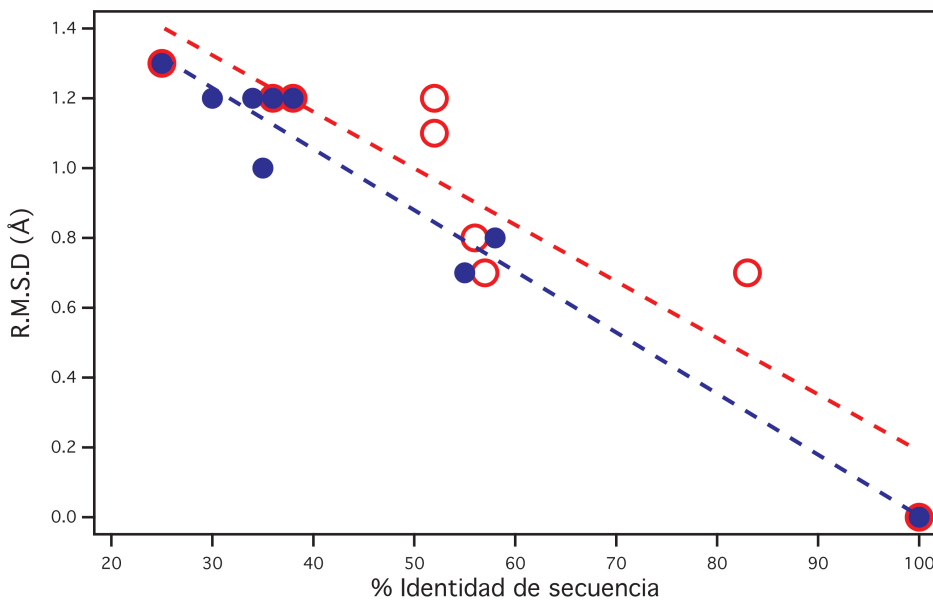


Figura 4.5. Representación del R.M.S.D. en función de la identidad de secuencia. En círculos rojos vacíos se representan los valores obtenidos al comparar las tiorredoxinas ancestrales y la humana con la trx de *E. coli*. La línea discontinua roja corresponde a un ajuste lineal a los datos de trx de *E.coli*. En círculos azules rellenos se representan los valores obtenidos al comparar las tiorredoxinas ancestrales y la de *E. coli* con la trx humana. La línea discontinua azul corresponde a un ajuste lineal a los datos de trx humana.

Otro factor que hemos analizado para corroborar la similitud de las estructuras de tiorredoxinas precámbricas es la distribución de regiones flexibles (obtenida a partir de los factores B) en las estructuras. Para llevar a cabo este análisis, hemos calculamos la media normalizada de factores B para cada residuo del esqueleto polipeptídico (Fig. 4.6) a partir de la superposición de las cadenas A de las estructuras de tiorredoxinas.

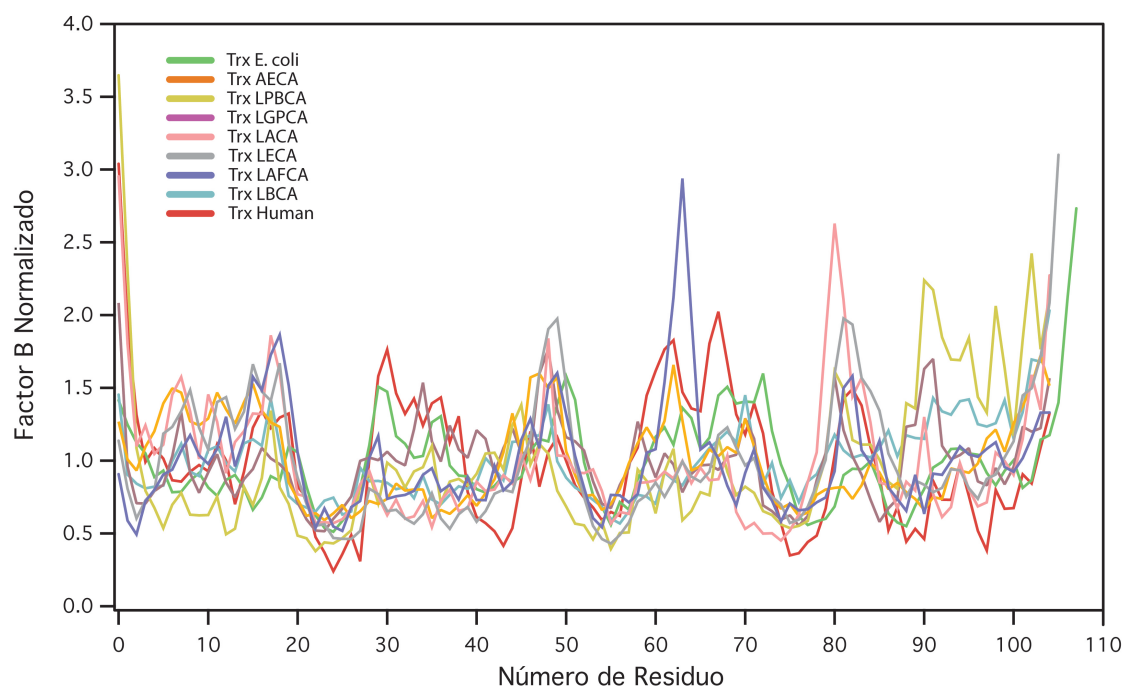


Figura 4.6. Representación de los valores de factores beta normalizados (valores medios de la cadena principal de cada residuo/valor medio total de la proteína) de la cadena principal para cada residuo.

Los valores de los factores térmicos B asociado a cada átomo de la proteína dan idea de la movilidad/flexibilidad de ese grupo de átomos ó región de la cadena polipeptídica. En la figura 4.6, podemos observar una regionalización de altos y bajos valores de factores B normalizados, que corresponden a regiones de la estructura de la tiorredoxinas más o menos flexibles, respectivamente. La región del centro activo (residuos 25-35), tiene una baja flexibilidad en la mayoría de las proteínas, excepto en la trx humana en la que se observa un cambio brusco desde una región de bajos factores B (los menores de todas las tiorredoxinas) a altos valores de B. Las otras regiones con menor flexibilidad (menor valor de factor B normalizado) corresponden a las regiones que forman hojas beta en la estructura (ver Fig. 4.7.), mientras que las regiones que presentan mayor flexibilidad corresponden a lazos o extremos de alfa hélices (Fig. 4.6.). En el caso de la trx Humana y LAFCA hay una región de alta flexibilidad en torno al residuo 65 que se aleja de la media y que corresponde a la hélice $\alpha 3$ (Fig. 4.7.). Pese a las leves diferencias observadas en la distribución de regiones flexibles, este estudio nos permite afirmar que las estructuras son muy similares, ya que hemos identificado un patrón común de regiones flexibles en las tiorredoxinas estudiadas.

Las estructuras de tiorredoxinas ancestrales presentan, por tanto, el plegamiento típico de tiorredoxina, constituyendo unos fósiles moleculares de 4 mil millones de años. Este resultado confirma la visión más aceptada que mantiene que las estructuras de las proteínas evolucionan lentamente. Sin embargo, hemos podido apreciar diferencias significativas en la longitud de la hélice $\alpha 1$ en las estructuras de las tiorredoxinas ancestrales y, consecuentemente, en los lazos que flanquean esta hélice (Fig. 4.7). Este resultado es relevante y coincidente con los resultados anteriores que demuestran que existe una diferencia sustancial de la primera hélice cuando se comparan un gran número de proteínas que muestran el *fold* de tiorredoxina (Qi and Grishin 2005). Así se observa que las estructuras de la tiorredoxina de *E. coli* y humana actuales difieren fundamentalmente en la longitud de esta hélice 4 aminoácidos en el caso de *trx* de *E. coli* y 10 aminoácidos en el caso de la tiorredoxina humana (Fig. 4.7.). Este hecho plantea una cuestión obvia: ¿cuál de las dos características estructurales (hélice larga en *trx* humana o hélice corta en *trx* de *E. coli*) es ancestral para ambas tiorredoxinas y cuál es derivada?

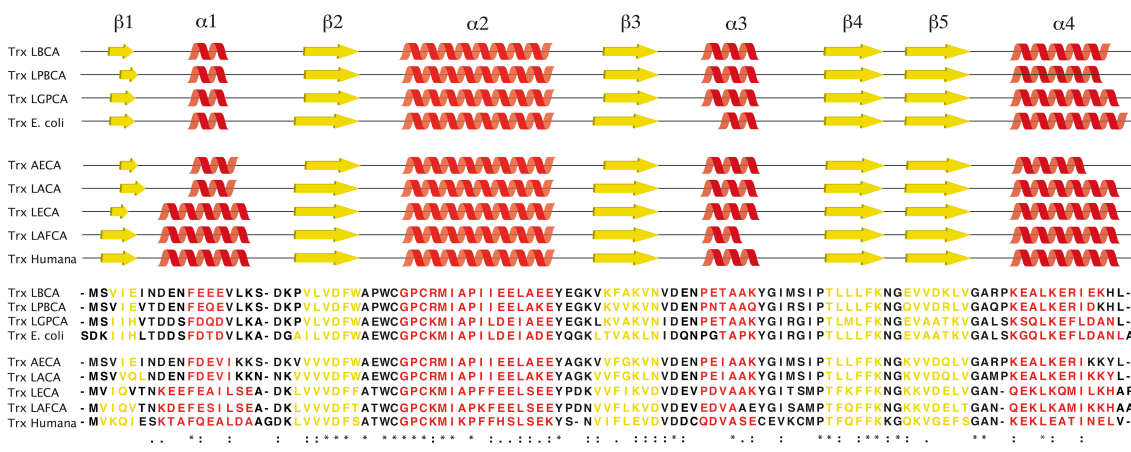


Figura 4.7. Diagrama y alineamiento de secuencias representando la asignación de estructura secundaria de las tiorredoxina ancestrales y actuales de *E. coli* y humana empleando el programa DSSP (apartado 2.2.8.3 de Materiales y Métodos). En rojo se representan las alfa hélices y en amarillo se representan las hojas beta. Las cuatro primeras secuencias corresponden a la rama procariota del árbol filogenético empleado para la reconstrucción. Las últimas cinco secuencias corresponden a la rama arquea/eucariota del árbol. En la parte inferior se muestra información relativa al alineamiento de secuencias. Con "*" aparecen marcados los residuos idénticos, con ":" y "." se marcan las posiciones conservadas y semi-conservadas, respectivamente.

Dado que este tipo de cuestiones evolutivas cruciales no pueden ser contestadas mediante el análisis comparativo de las estructuras actuales, la resurrección de proteínas ancestrales, junto con la determinación de sus estructuras, puede proporcionar una respuesta inmediata. La figura 4.8.A muestra la representación de la longitud de la hélice $\alpha 1$ en función del tiempo geológico, con datos de

tiorredoxinas humanas y de *E. coli*, así como los datos para las tiorredoxinas precámbricas. Las tres estructuras de tiorredoxinas más antiguas (LBCA, AECA, y LACA) presentan una hélice $\alpha 1$ corta. Además, la longitud de los lazos que la flanquean corroboran este resultado. Podemos apreciar como en la rama bacteriana no ha cambiado la longitud de la hélice $\alpha 1$ a lo largo de la evolución, mientras que en la rama eucariota, la longitud de la hélice ha aumentado sustancialmente (Fig. 4.8.A).

Para corroborar los resultados obtenidos, hemos estudiado la hélice $\alpha 1$ de las tiorredoxinas precámbricas y actuales (y los lazos que la flanquean) de forma aislada. Para ello, hemos identificado los enlaces de hidrógeno canónicos de la hélice $\alpha 1$, es decir, aquellos que se establecen entre el grupo amino de la cadena principal de un aminoácido (i) con el grupo carboxilo de la cadena principal del aminoácido (i+4) (Fig. 4.8.B). De nuevo, podemos observar, claramente, como el número de enlaces de hidrógeno de la hélice aumenta en la rama eucariota de 4 enlaces en trx de AECA a 8 en LECA, LAFCA y humana; mientras que en la rama procariota el número de enlaces de hidrógeno permanece constante a lo largo de 4 mil millones de años (Fig. 4.8.B).

Por tanto, la respuesta a la cuestión fundamental que nos hemos formulado es que la hélice corta que presenta la tiorredoxina de *E. coli* es muy probable que sea la característica estructural ancestral (podríamos especular que es muy probablemente la que presente la tiorredoxina de LUCA), mientras que la hélice larga de la tiorredoxina humana es una característica derivada a lo largo de la evolución.

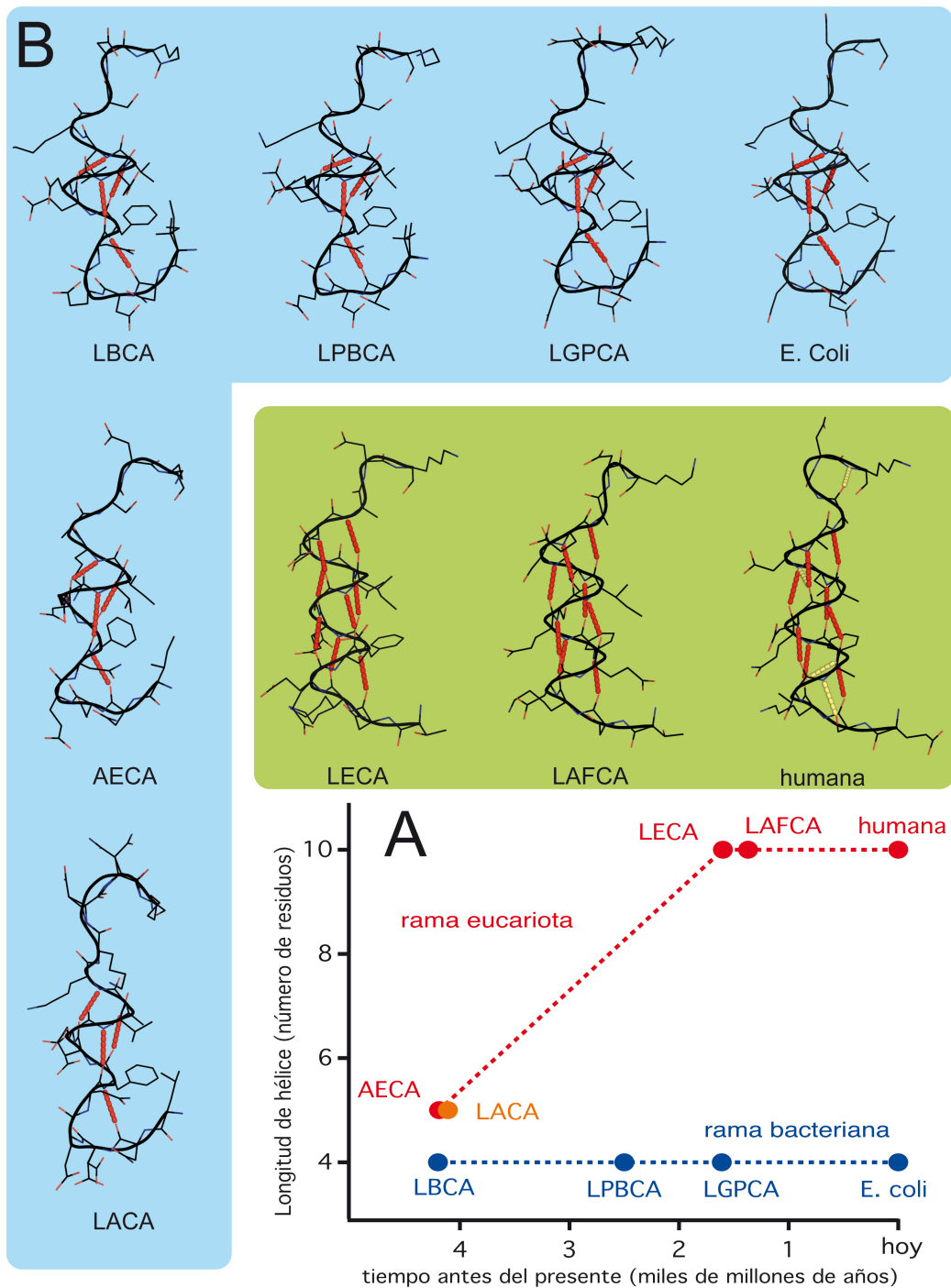


Figura 4.8. A: Representación de la longitud de la hélice $\alpha 1$ en función del tiempo geológico estimado para las tiorredoxinas estudiadas. En rojo se representa las tiorredoxinas pertenecientes a la rama eucariota del árbol filogenético. En naranja se muestra el último ancestro común de las arqueas. En rojo se representan las tiorredoxinas pertenecientes a la rama procariota del árbol filogenético. B: representación del fragmento de estructura que corresponde a la hélice $\alpha 1$ y los lazos que la flaquean en cada tiorredoxina. En rojo se representan los enlaces de hidrógeno canónicos de una hélice alfa ($i, i+4$). Con fondo azul se muestran las tiorredoxinas con 4 enlaces de hidrógeno canónicos y con fondo verde se muestran las tiorredoxinas con 8 enlaces de hidrógeno canónicos.

En resumen, hemos proporcionado evidencias que conducen a que la estructura de tiorredoxina ha cambiado poco a lo largo de 4 mil millones de años. Este resultado es realmente destacable, dado el elevado número de diferencias en la secuencia (más del 50% en algunos casos) entre las tiorredoxinas ancestrales y actuales. También constituye un resultado relevante porque apoya que, a pesar de la naturaleza geoméricamente continua del espacio estructural de las proteínas (Sadreyev, Kim et al. 2009), los análisis evolutivos basados en la resurrección de proteínas ancestrales podrían conducir a la determinación de prototipos estructurales bien definidos. Por supuesto, nuestros resultados experimentales indican algunos cambios en las estructuras de tiorredoxinas a lo largo de 4 mil millones de años y, aunque estos cambios son pequeños, hemos mostrado que están relacionados claramente con la señal filogenética y proporcionan una interpretación evolutiva de las diferencias estructurales entre las tiorredoxinas humana y de *E. coli*. En general, hemos demostrado que la resurrección ancestral de proteínas precámbricas es una poderosa herramienta que nos permite explorar y probar las hipótesis sobre la evolución de las estructuras de proteínas.

4.4. Bibliografía

- Alcolombri, U., M. Elias and D. S. Tawfik (2011). "Directed Evolution of Sulfotransferases and Paraoxonases by Ancestral Libraries." Journal of molecular biology **411**(4): 837-853.
- Caetano-Anolles, G., M. Wang, D. Caetano-Anolles and J. E. Mittenthal (2009). "The Origin, Evolution and Structure of the Protein World." The Biochemical journal **417**(3): 621-637.
- Campbell, K. L., J. E. Roberts, L. N. Watson, J. Stetefeld, A. M. Sloan, A. V. Signore, J. W. Howatt, J. R. Tame, N. Rohland, T. J. Shen, J. J. Austin, M. Hofreiter, C. Ho, R. E. Weber and A. Cooper (2010). "Substitutions in Woolly Mammoth Hemoglobin Confer Biochemical Properties Adaptive for Cold Tolerance." Nat Genet **42**(6): 536-540.
- Carroll, S. M., E. A. Ortlund and J. W. Thornton (2011). "Mechanisms for the Evolution of a Derived Function in the Ancestral Glucocorticoid Receptor." PLoS genetics **7**(6): e1002117.
- Chivers, P. T., K. E. Prehoda, B. F. Volkman, B. M. Kim, J. L. Markley and R. T. Raines (1997). "Microscopic Pka Values of Escherichia Coli Thioredoxin." Biochemistry **36**(48): 14985-14991.
- Chivers, P. T. and R. T. Raines (1997). "General Acid/Base Catalysis in the Active Site of Escherichia Coli Thioredoxin." Biochemistry **36**(50): 15810-15816.
- Collaborative Computational Project, N. (1994). "The Ccp4 Suite: Programs for Protein Crystallography." Acta crystallographica. Section D, Biological crystallography **50**(Pt 5): 760-763.
- Cordes, M. H., N. P. Walsh, C. J. McKnight and R. T. Sauer (1999). "Evolution of a Protein Fold in Vitro." Science **284**(5412): 325-328.
- D'Costa, V. M., C. E. King, L. Kalan, M. Morar, W. W. Sung, C. Schwarz, D. Froese, G. Zazula, F. Calmels, R. Debruyne, G. B. Golding, H. N. Poinar and G. D. Wright (2011). "Antibiotic Resistance Is Ancient." Nature **477**(7365): 457-461.
- Ducruix, A. and R. Giegé (2000). Crystallization of Nucleic Acids and Proteins: A Practical Approach Oxford, Oxford University Press.
- Gaucher, E. A., S. Govindarajan and O. K. Ganesh (2008). "Palaeotemperature Trend for Precambrian Life Inferred from Resurrected Proteins." Nature **451**(7179): 704-707.
- Grishin, N. V. (2001). "Fold Change in Evolution of Protein Structures." Journal of structural biology **134**(2-3): 167-185.
- Hedges, S. B. K., S. (2009). The Timetree of Life Oxford Univ. Press, Oxford.

- Holmgren, A. and B. Soderberg (1970). "Crystallization and Preliminary Crystallographic Data for Thioredoxin from Escherichia Coli B." Journal of molecular biology **54**(2): 387-390.
- Honig, B. (2007). "Protein Structure Space Is Much More Than the Sum of Its Folds." Nature structural & molecular biology **14**(6): 458.
- Jancarik, J. and S.-H. Kim (1991). "Sparse Matrix Sampling: A Screening Method for Crystallization of Proteins." Journal of Applied Crystallography **24**(4): 409-411.
- Katti, S. K., D. M. LeMaster and H. Eklund (1990). "Crystal Structure of Thioredoxin from Escherichia Coli at 1.68 a Resolution." Journal of molecular biology **212**(1): 167-184.
- Kimber, M. S., F. Vallee, S. Houston, A. Necakov, T. Skarina, E. Evdokimova, S. Beasley, D. Christendat, A. Savchenko, C. H. Arrowsmith, M. Vedadi, M. Gerstein and A. M. Edwards (2003). "Data Mining Crystallization Databases: Knowledge-Based Approaches to Optimize Protein Crystal Screens." Proteins **51**(4): 562-568.
- Konno, A., A. Kitagawa, M. Watanabe, T. Ogawa and T. Shirai (2011). "Tracing Protein Evolution through Ancestral Structures of Fish Galectin." Structure **19**(5): 711-721.
- Krishna, S. S. and N. V. Grishin (2004). "Structurally Analogous Proteins Do Exist!" Structure **12**(7): 1125-1127.
- Murzin, A. G. (2008). "Metamorphic Proteins." Science **320**(5884): 1725-1726.
- Nikkola, M., F. K. Gleason, J. A. Fuchs and H. Eklund (1993). "Crystal Structure Analysis of a Mutant Escherichia Coli Thioredoxin in Which Lysine 36 Is Replaced by Glutamic Acid." Biochemistry **32**(19): 5093-5098.
- Orengo, C. A., D. T. Jones and J. M. Thornton (1994). "Protein Superfamilies and Domain Superfolds." Nature **372**(6507): 631-634.
- Ortlund, E. A., J. T. Bridgham, M. R. Redinbo and J. W. Thornton (2007). "Crystal Structure of an Ancient Protein: Evolution by Conformational Epistasis." Science **317**(5844): 1544-1548.
- Otalora, F., J. A. Gavira, J. D. Ng and J. M. Garcia-Ruiz (2009). "Counterdiffusion Methods Applied to Protein Crystallization." Prog Biophys Mol Biol **101**(1-3): 26-37.
- Perez-Jimenez, R., J. Li, P. Kosuri, I. Sanchez-Romero, A. P. Wiita, D. Rodriguez-Larrea, A. Chueca, A. Holmgren, A. Miranda-Vizueté, K. Becker, S. H. Cho, J. Beckwith, E. Gelhaye, J. P. Jacquot, E. A. Gaucher, J. M. Sanchez-Ruiz, B. J. Berne and J. M. Fernandez (2009). "Diversity of Chemical Mechanisms in Thioredoxin Catalysis Revealed by Single-Molecule Force Spectroscopy." Nat Struct Mol Biol **16**(8): 890-896.
- Qi, Y. and N. V. Grishin (2005). "Structural Classification of Thioredoxin-Like Fold Proteins." Proteins **58**(2): 376-388.

- Sadreyev, R. I., B. H. Kim and N. V. Grishin (2009). "Discrete-Continuous Duality of Protein Structure Space." Current Opinion in Structural Biology **19**(3): 321-328.
- Schaeffer, R. D. and V. Daggett (2011). "Protein Folds and Protein Folding." Protein engineering, design & selection : PEDS **24**(1-2): 11-19.
- Schultz, L. W., P. T. Chivers and R. T. Raines (1999). "The Cxxc Motif: Crystal Structure of an Active-Site Variant of Escherichia Coli Thioredoxin." Acta crystallographica. Section D, Biological crystallography **55**(Pt 9): 1533-1538.
- Soderberg, B. O., A. Holmgren and C. I. Branden (1974). "Structure of Oxidized Thioredoxin to 4 with 5 a Resolution." Journal of molecular biology **90**(1): 143-152.
- Taylor, W. R. (2007). "Evolutionary Transitions in Protein Fold Space." Current Opinion in Structural Biology **17**(3): 354-361.
- Valas, R. E., S. Yang and P. E. Bourne (2009). "Nothing About Protein Structure Classification Makes Sense except in the Light of Evolution." Current Opinion in Structural Biology **19**(3): 329-334.
- Winstanley, H. F., S. Abeln and C. M. Deane (2005). "How Old Is Your Fold?" Bioinformatics **21 Suppl 1**: i449-458.
- Xie, L. and P. E. Bourne (2008). "Detecting Evolutionary Relationships across Existing Fold Space, Using Sequence Order-Independent Profile-Profile Alignments." Proceedings of the National Academy of Sciences of the United States of America **105**(14): 5441-5446.

5. ANÁLISIS ENERGÉTICO-ESTRUCTURAL DE TIORREDOXINAS PRECÁMBRICAS

5.1. Introducción

En el capítulo 3 de la presente memoria, hemos abordado la reconstrucción de las secuencias de varias tiorredoxinas precámbricas (~ 4 mil millones de años antes del presente) a partir de un análisis filogenético de más de 200 secuencias distintas de tiorredoxinas actuales, la resurrección de estas proteínas en el laboratorio y la determinación de su estabilidad (por calorimetría diferencial de barrido) y su actividad catalítica (por microscopía de fuerza atómica) (Perez-Jimenez, Ingles-Prieto et al. 2011). La gran estabilidad y la catálisis a pH ácido observadas en las tiorredoxinas más antiguas han proporcionado las pruebas de adaptación de los organismos a las condiciones calientes y ácidas estimadas en los océanos primitivos (Perez-Jimenez, Ingles-Prieto et al. 2011). Este resultado tiene gran relevancia en términos evolutivos, así como desde el punto de vista de la ingeniería de proteínas, ya que hemos conseguido proteínas con propiedades modificadas.

Tanto la alteración de la estabilidad como de la actividad catalítica de las enzimas tiene un gran interés en la industria biotecnológica y biomédica. En esta tesis doctoral nos hemos centrado en el estudio de la estabilidad térmica de las proteínas, ya que algunas de las tiorredoxinas ancestrales resucitadas son hiperestables, tal como refleja su temperatura de desnaturalización (T_m) por encima de los 110°C (Perez-Jimenez, Ingles-Prieto et al. 2011). En el caso de la tiorredoxina de LPBCA, la T_m llega a alcanzar los 121°C, 32°C más alta que la de tiorredoxina de *E. coli*.

Se sabe que la estabilidad térmica de las proteínas es marginal (Schellman 1987; Taverna y Goldstein 2002), tal como indican los bajos valores experimentales del cambio de energía libre Gibbs de desplegamiento (ΔG) (del orden de decenas de kJ/mol) a temperatura fisiológica. Esto es debido a que los componentes entálpico y entrópico (del orden de cientos de KJ/mol) de ΔG (ΔH y $-T\cdot\Delta S$) se cancelan casi totalmente. Desde el punto de vista molecular, el bajo valor de ΔG para el desplegamiento de proteínas es el resultado de un balance muy delicado entre diferentes interacciones atómicas en la estructura de las proteínas, las que se oponen al plegamiento y las que lo dirigen. Sin embargo, el conocimiento que tenemos sobre la relación estructura/energética de las proteínas es muy escaso. En la bibliografía se pueden encontrar multitud de estudios que abordan el papel de diferentes interacciones en la estabilidad de proteínas, tales como los enlaces de hidrógeno (Shirley, Stanssens et al. 1992; Myers y Pace 1996; Gao, Bosco et al. 2009; Pace 2009), los puentes salinos (Sun, Sauer et al. 1991; Kumar y Nussinov

1999; Kumar, Tsai et al. 2000), las interacciones hidrofóbicas (Kauzmann 1954; Kauzmann 1959; Eriksson, Baase et al. 1992; Pace, Fu et al. 2011) y las interacciones de carga superficiales (Ibarra-Molero, Makhatadze et al. 1999; Sanchez-Ruiz y Makhatadze 2001; Sundd, Iverson et al. 2002). Además, se ha evaluado la contribución a la estabilidad debida a la propensión los aminoácidos de aparecer en determinados elementos de estructura secundaria (Zhang, Baase et al. 1992; Munoz, Blanco et al. 1995; Malakauskas y Mayo 1998; Trevino, Schaefer et al. 2007; Fu, Grimsley et al. 2009) o al número de prolinas y glicinas que poseen las proteínas (Matthews, Nicholson et al. 1987; Suzuki, Hatagaki et al. 1991). No obstante, la cuantificación de la contribución de estos factores a la estabilidad de proteínas no está aun clara, e incluso los resultados y conclusiones de estos trabajos son a veces opuestos.

En este capítulo, nos hemos propuesto averiguar las posibles causas de la hiperestabilidad de las tiorredoxinas precámbricas resucitadas, ya que la resurrección de tiorredoxinas ancestrales y la posterior obtención de sus estructuras tridimensionales nos coloca en una posición única para entender los mecanismos moleculares responsables de la hiperestabilidad. Con este objetivo, hemos realizado un estudio comparativo de las estructuras de las tiorredoxinas precámbricas y las tiorredoxinas actuales de *Escherichia coli* y humana dirigido a evaluar las diferentes contribuciones energético/estructurales a la estabilidad térmica de estas proteínas.

5.2. Resultados y Discusión

5.2.1. El efecto hidrofóbico

El efecto hidrófobo se considera como una de las principales fuerzas del plegamiento desde el trabajo pionero de Kauzmann en los años cincuenta (Kauzmann 1954; Kauzmann 1959). Dentro de las contribuciones hidrofóbicas a la estabilidad de proteínas podemos distinguir entre (Pace, Fu et al. 2011): i) Las relativas a la cantidad de superficie apolar que se expone al disolvente acuoso debido al proceso de desplegamiento de la proteína y ii) La eficiencia en el empaquetamiento del núcleo hidrofóbico de la proteína (relacionada con la contribución de las interacciones de van der Waals entre los residuos apolares).

La cantidad de superficie apolar que se expone al disolvente acuoso debido al desplegamiento de la proteína puede ser evaluado a partir de los cálculos de área superficial de la proteína accesible al disolvente (ASA). Podemos suponer que las proteínas con una área apolar menos expuesta podrían ser más estables, ya de los residuos hidrofóbicos normalmente se encuentran enterrados en el núcleo de la proteína. Como, *a priori*, no sabemos el ASA del estado desplegado de la proteína, hemos empleado varias aproximaciones para modelar la exposición del estado desplegado: el modelo de los tripéptidos y el modelo de los límites superior e inferior propuestos por Rose (Creamer, Srinivasan et al. 1995; Creamer, Srinivasan et al. 1997)(para más detalles sobre los cálculos de ASA consultar el apartado 2.2.9.1 del capítulo Materiales y Métodos). La figura 5.1 muestra el cambio en el ASA durante el desplegamiento de las tiorredoxinas ancestrales y actuales.

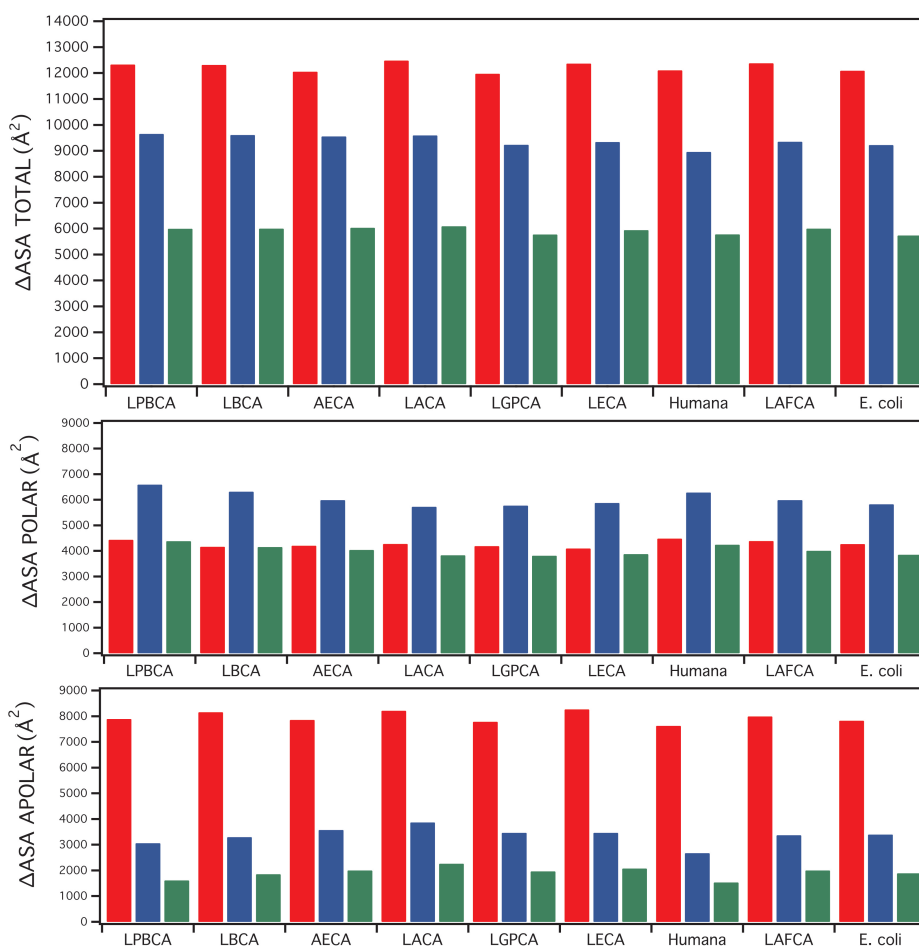


Figura 5.1. Valores de cambio de ASA durante el desplegamiento de las tiorredoxinas, colocadas de mayor a menor T_m . Se muestra el Δ ASA de la proteína total, de los residuos polares y de los residuos apolares. Para los cálculos hemos empleado la estructura tridimensional de las tiorredoxinas ancestrales actuales de *E. coli* (PDB: 2TRX) y humana (PDB: 1ERU). Para modelar la exposición del estado desplegado se han empleado los modelos de los tripéptidos (rojo), límite superior (azul) y límite inferior (verde).

Las diferencias del cambio de ASA de las tiorredoxinas actuales respecto a las tiorredoxinas ancestrales son relativamente pequeñas (alrededor del 0.1% al 20%), independientemente del modelo utilizado para modelar la exposición del estado desplegado. Concretamente, en el caso del cambio de ASA apolar no hemos apreciado diferencias significativas entre las diferentes tiorredoxina (Fig. 5.2).

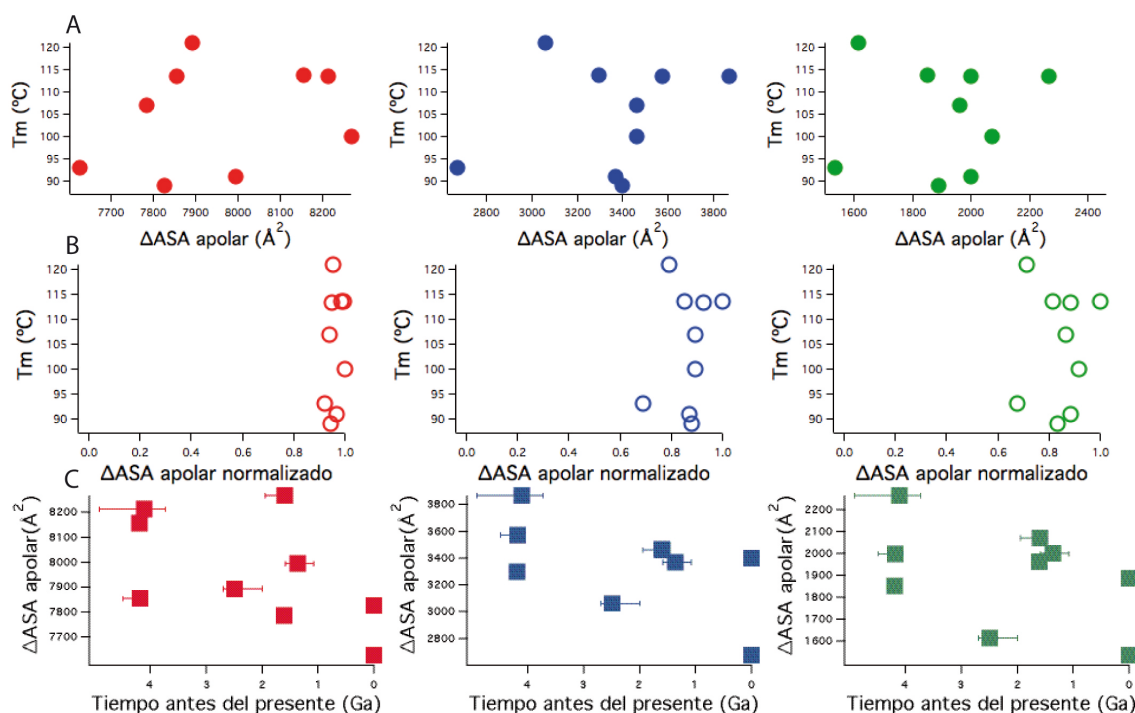


Figura 5.2. Representación de la T_m de las tiorredoxinas precámbricas y actuales de *E.coli* y humana en función de ΔASA apolar (A) y ΔASA apolar normalizada por el valor máximo (B). C: Representación de ΔASA apolar en función de tiempo de divergencia estimado para las tiorredoxinas analizadas. Los tiempos de divergencia están recogidos en (Hedges y Kumar 2009). Para modelar la exposición del estado desplegado de las se han empleado los modelos de los tripéptidos (rojo), límite superior (azul) y límite inferior (verde).

Por otra parte, la eficiencia en el empaquetamiento del núcleo hidrofóbico de la proteína está relacionada con la presencia de cavidades internas (Eriksson, Baase et al. 1992). De manera que un empaquetamiento ineficiente sería reflejo de la existencia de cavidades en el interior de la proteína. La tabla 5.1 muestra la estimación del número, el volumen y el área de las cavidades internas en las estructuras nativas de las tiorredoxinas actuales y ancestrales. Esta estimación la hemos realizado con el programa *MSroll* (en el apartado 2.2.9.2. de Materiales y Métodos se explican estos cálculos con detalle).

Tabla 5.1. Estimación de las cavidades internas totales en las tiorredoxina. El método utilizado es el propuesto por Connolly (Connolly 1983) empleando una sonda de radio 1.2 Å. El volumen de la cavidad es el volumen total que puede ser ocupado por la sonda, mientras que el área de la cavidad es el área barrida por la sonda en el interior de la cavidad.

Proteína (PDB)	T _m (°C)	Número cavidades	Volumen total cavidades (Å ³)	Área total cavidades (Å ²)
Trx LPBCA (2YJ7)	121.0	7	251.4	417.7
Trx LBCA (4BA7)	113.7	6	201.0	343.9
Trx AECA	113.5	9	237.9	399.3
Trx LACA (2YNX)	113.4	6	168.6	272.4
Trx LGPCA (2YN1)	107.0	7	192.0	331.9
Trx LECA (2Y01)	100.0	6	153.9	245.2
Trx humana (1ERU)	93.0	4	56.8	95.6
Trx LAFCA (2YPM)	91.0	4	62.3	102.4
Trx <i>E. coli</i> (2TRX)	89.0	4	135.3	222.3

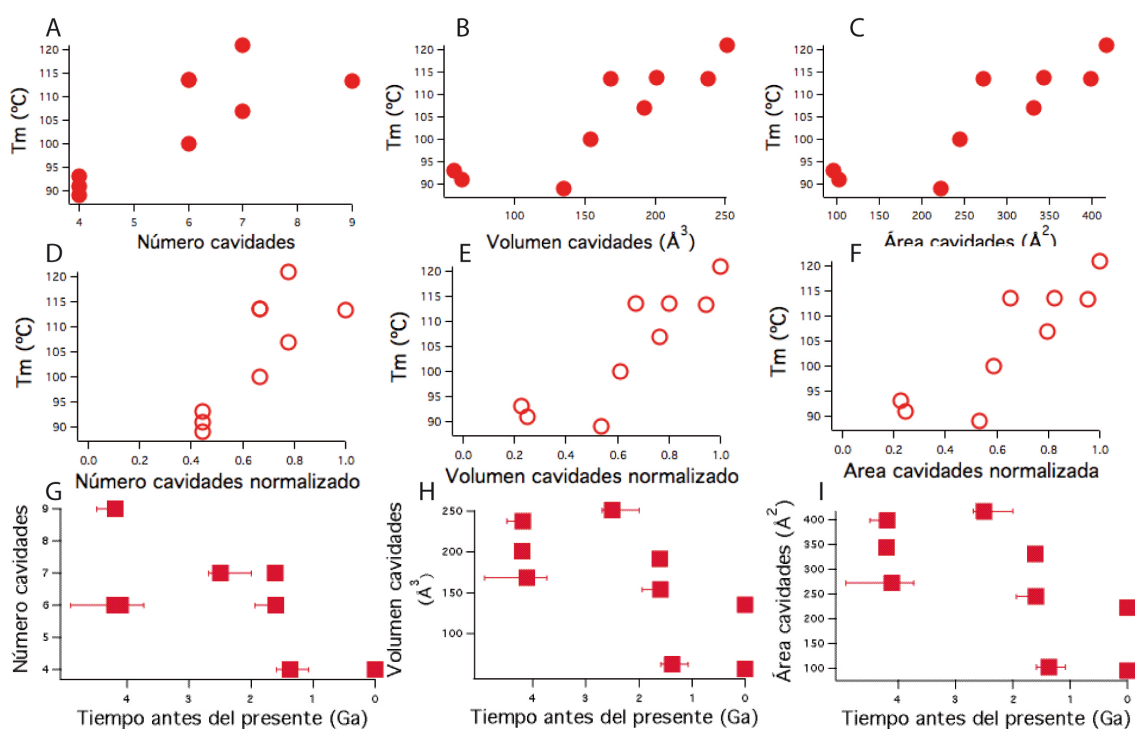


Figura 5.3. A-C: representación de la temperatura de desnaturalización de las tiorredoxinas precámbricas y actuales frente al número de cavidades (A), el volumen total de cavidades (B) y el área total de las cavidades (C). Además, hemos representado la t_m frente al número de cavidades normalizado por el valor máximo (D), el volumen total de cavidades normalizado por el valor máximo (E) y el área total de las cavidades normalizada por el valor máximo (F). G-I: Representación del número de cavidades (G), el volumen total de cavidades (H) y el área total de cavidades (I) en función del tiempo de divergencia estimado para las tiorredoxinas analizadas. Los tiempos de divergencia están recogidos en (Hedges y Kumar 2009).

El número de cavidades internas y el volumen total son similares en todas las tiorredoxinas excepto en la tiorredoxina humana y LAFCA, que son las menos estables (Tabla 5.1). De hecho, hemos observado cierta correlación de la estabilidad térmica de las proteínas con el volumen de cavidades (Fig. 5.3), aunque, sorprendentemente, el resultado es lo contrario de que cabría esperar, ya que por ejemplo, la tiorredoxina ancestral más estable (LPBCA) presenta un número y volumen total de las cavidades de 251.5 \AA^3 , mientras que la tiorredoxina menos estable (trx de *E. coli*), presenta un menor número y volumen total de cavidades de 135.3 \AA^3 . Por ello, es necesario localizar estas cavidades y realizar un estudio más detallado de las mismas.

En general, no hemos apreciado diferencias significativas en el cambio de ASA apolar (Fig. 5.2.) entre las diferentes tiorredoxinas y en el volumen y número de cavidades hemos encontrado una tendencia contraria a la esperada (Fig. 5.3). Estos resultados, por tanto, rebaten que el efecto hidrófobo sea el responsable principal de la diferencia en la estabilidad entre las tiorredoxinas actuales y ancestrales. Por supuesto, esta conclusión debemos tomarla con precaución, debido a que la contribución total del efecto hidrofóbico a la energética de plegamiento es del orden de cientos de kJ/mol y pequeñas diferencias (por debajo del margen de precisión de los cálculos) podrían tener una contribución significativa a la estabilidad de la proteína.

5.2.2. El número de enlaces de hidrógeno

Antes de los años 90, la opinión generalizada era que los enlaces de hidrógeno contribuían relativamente poco a la estabilidad de las proteínas, ya que, previsiblemente, los enlaces de hidrógeno intracatenarios podrían sustituirse por enlaces de hidrógeno igual de fuertes con moléculas de agua. Este enfoque ha sido cuestionado principalmente por el Dr. Pace y sus colaboradores en varios estudios experimentales (Pace 2009; Pace, Fu et al. 2011). Estos estudios sugieren que un enlace de hidrógeno podría contribuir aproximadamente 1 kcal/mol a la estabilidad, a pesar de la gran incertidumbre (en el orden del valor mismo) de esta estimación.

Para estimar el número de enlaces de hidrógeno de las tiorredoxinas ancestrales y actuales, hemos calculado el número total de enlaces de hidrógeno intramoleculares potenciales (Tabla 5.2). Además, hemos estimado el número de enlaces de hidrógeno que forman una red óptima, es decir, los enlaces que podrían

contribuir en mayor medida a la estabilidad de la proteína (apartado 2.2.9.3. de Materiales y Métodos).

Hay que tener en cuenta que los resultados de estas dos estimaciones no se pueden comparar directamente. El motivo principal es que el cálculo de enlaces de hidrógeno potenciales emplea distancias y ángulos arbitrarios, asignando a todos los enlaces de hidrógeno el mismo valor. Por otro lado, el cálculo de enlaces de hidrógeno que forman la red óptima utiliza un cálculo de campo de fuerza, que asigna una puntuación a cada enlace según su contribución a la estabilidad de la proteína. Por esta razón, el resultado del cálculo de enlaces de hidrógeno potenciales puede dar más de un enlace de hidrógeno por átomo, mientras que el cálculo de enlaces de hidrógeno que forman la red óptima solo tiene en cuenta el mejor enlace de hidrógeno para cada átomo.

Tabla 5.2. Número de enlaces de hidrógeno en las estructuras de tiorredoxinas (ordenadas de mayor a menor T_m). El número total de enlaces de hidrógenos potenciales se ha calculado con el programa *PFIS*. El número de enlaces que forman la red óptima se han calculado con el servidor *Whatif optimal hydrogen bonds network*.

Proteína (PDB)	Enlaces entre cadenas principales	Enlaces entre cadenas principales y laterales	Enlaces entre cadenas laterales	Nº de enlaces potenciales	Nº de enlaces de hidrógeno en red óptima
Trx LPBCA (2YJ7)	137	20	16	173	57.0
Trx LBCA (4BA7)	111	17	7	135	51.8
Trx AECA	52	10	2	64	41.7
Trx LACA (2YNX)	93	11	11	115	53.8
Trx LGPCA (2YN1)	119	21	11	151	56.2
Trx LECA (2Y01)	147	35	6	188	50.8
Trx humana (1ERU)	69	11	8	88	57.2
Trx LAFCA (2YPM)	86	20	10	116	50.6
Trx <i>E. coli</i> (2TRX)	82	16	12	110	54.5

Los resultados muestran que no existen diferencias significativas respecto al número de enlaces de hidrógeno (Fig. 5.4). La proteína más estable, LPBCA con 121°C de T_m , tiene 173 enlaces de hidrógeno potenciales, mientras que una de las

proteínas menos estables, LECA con 100°C de T_m , forma 188 enlaces. En el caso de los enlaces de hidrógeno que estarían formando la red óptima, la tiorredoxina LPBCA tiene prácticamente el mismo número de enlaces de hidrógeno que la tiorredoxina humana (93°C de T_m), que es una de las proteínas menos estables. Por tanto, no existen indicios que apunten a la posible contribución de los enlaces de hidrógeno a la hiperestabilidad de algunas tiorredoxinas ancestrales.

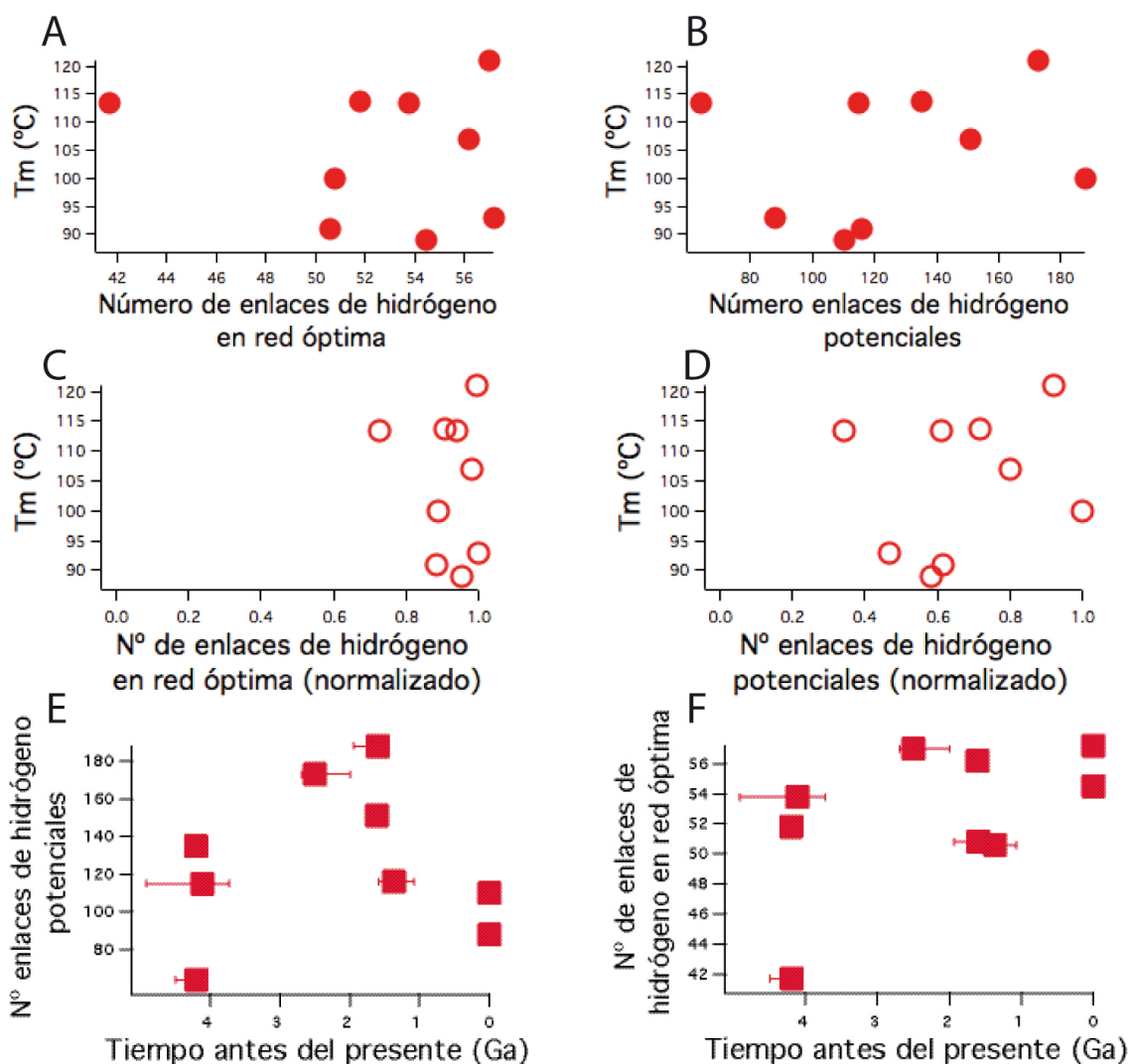


Figura 5.4. A: Representación de la T_m de las tiorredoxinas precámbricas y actuales de *E. coli* y humana frente al número de enlaces de hidrógeno formando una red óptima (A) y número de enlaces de hidrógenos potenciales (B). Representación de la T_m frente al número de enlaces de hidrógeno formando una red óptima (C) y número de enlaces de hidrógenos potenciales normalizados por el valor máximo (D). Representación del número de enlaces de hidrógeno formando una red óptima (E) y número de enlaces de hidrógeno potenciales (F) en función del tiempo estimado de divergencia de las tiorredoxinas estudiadas. Los tiempos de divergencia están recogidos en (Hedges y Kumar 2009).

Cabe destacar, que los datos recogidos en la tabla 5.2 los hemos obtenido manteniendo las conformaciones alternativas de los residuos en las estructuras de

tioredoxinas, ya que no sabemos con seguridad qué cadena podría estar formando enlaces de hidrógeno. No obstante, hemos realizado este cálculo eliminando las conformaciones alternativas con menor ocupación y número total de enlaces varía ligeramente, pero el número de enlaces que forman la red óptima no varía significativamente.

En general, estos cálculos rebaten la hipótesis de que haya una elevada contribución de los enlaces de hidrógeno a la estabilidad de las tioredoxinas analizadas. De nuevo, esta conclusión debemos tomarla con la debida cautela, dada la gran incertidumbre asociada a las estimaciones actuales de la contribución de enlaces de hidrógeno y, en particular, al hecho de que diferentes tipos de enlaces de hidrógeno puede contribuir de forma diferente y de una manera que es difícil determinar (Gao, Bosco et al. 2009).

5.2.3. La distribución de carga superficial

Hace unos 30 años, Wada y Nakamura analizaron 44000 interacciones de carga en la superficie de 14 proteínas y concluyeron que las cargas en las proteínas están, en promedio, rodeadas de cargas de signo opuesto (Wada y Nakamura 1981). Estudios experimentales más recientes han apoyado el carácter estabilizante global de las interacciones carga-carga en la superficie. Además, el rediseño de la distribución de carga superficial ha demostrado ser un método eficiente para la estabilización de las proteínas (Loladze, Ibarra-Molero et al. 1999). Más aún, la optimización de las interacciones entre cargas de superficie ha demostrado ser un factor determinante de la elevada estabilidad de algunas proteínas termófilas e hipertermófilas (Sanchez-Ruiz y Makhatadze 2001).

Hemos evaluado las interacciones de carga superficiales de las tioredoxinas actuales y ancestrales empleando un sencillo cálculo de Tanford-Kirkwood (apartado 2.2.9.4 de Materiales y Métodos) que ha sido validado en varios estudios experimentales de nuestro grupo (Ibarra-Molero, Loladze et al. 1999; Perez-Jimenez, Godoy-Ruiz et al. 2005). Para cada residuo cargado, el modelo Tanford-Kirkwood proporciona un valor medio de la energía de interacción con todos los demás residuos cargados en la proteína (Fig. 5.5). Los valores negativos indican interacciones favorables entre cargas de signo opuesto, mientras que los valores positivos indican las interacciones desestabilizantes entre las cargas del mismo signo.

ANÁLISIS ENERGÉTICO-ESTRUCTURAL DE TIORREDOXINAS
PRECÁMBRICAS

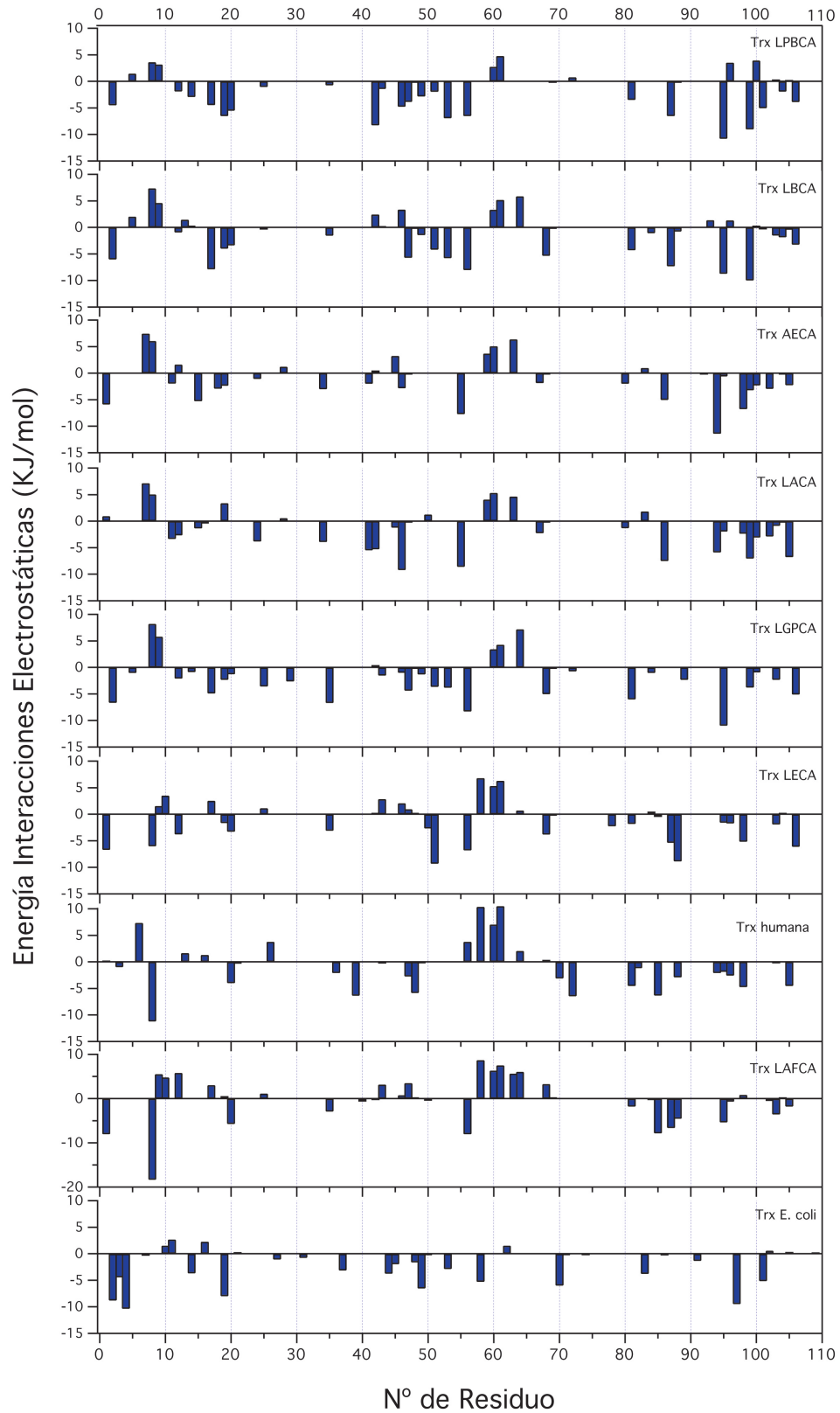


Figura 5.5. La figura muestra la energía de las interacciones carga-carga para cada residuo ionizable de las tiorredoxinas ancestrales y actuales (ordenadas de mayor a menor T_m). Los valores negativos se consideran interacciones favorables para la estabilidad y los valores positivos son interacciones desfavorables para la estabilidad.

La mayor parte de las energías de interacción electrostáticas mostradas en la figura 5.5. son interacciones favorables (valor negativo). Sin embargo, el efecto estabilizante prácticamente se contrarresta con las interacciones electrostáticas desestabilizantes. Sorprendentemente, la tiorredoxina de *E. coli* (89°C de T_m), es la que tiene menor número de interacciones desestabilizantes, siendo la proteína menos estable. Además, destaca la presencia de dos regiones desestabilizantes en todas las tiorredoxinas (excepto en la tiorredoxina de *E. coli*) una de ellas en torno al residuo 60 y otra en torno al residuo 10. Otros factores que hemos tenido en cuenta es el número de residuos ionizables, la energía total debida a las interacciones carga-carga de las tiorredoxinas y su carga neta a pH 7.0 (Fig. 5.6). Al comparar estos parámetros tampoco hemos observado diferencias significativas. Por ejemplo, el valor energía determinado para la tiorredoxina más estable (LPBCA) y la menos estable (*E. coli*) vemos que es prácticamente idéntico (-40.19 KJ/mol y -39.69 KJ/mol, respectivamente) (Fig. 5.6).

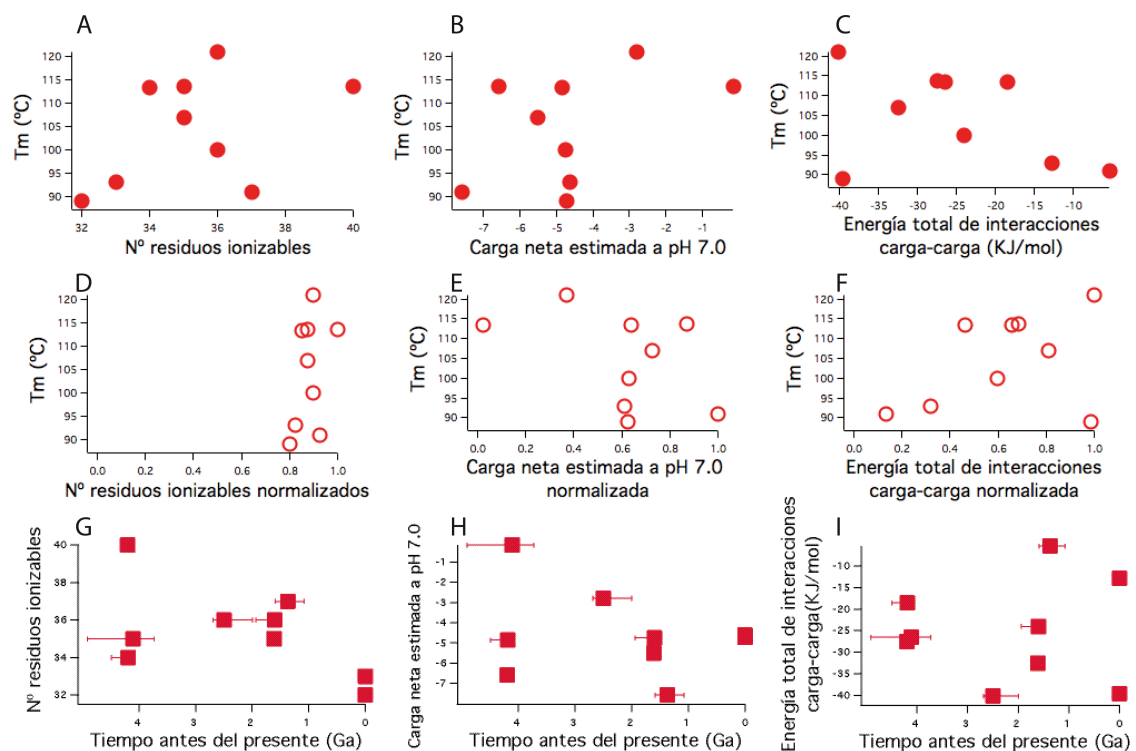


Figura 5.6. A-C: Representación de la T_m de las tiorredoxinas precámbricas y actuales de *E. coli* y humana frente al número residuos ionizables (A), energía total de interacciones carga-carga (B) y carga neta de la proteína a pH 7.0 (C). D-E: Representación de la T_m de las de las tiorredoxinas analizadas frente al número residuos ionizables normalizado (D), energía total de interacciones carga-carga normalizada (E) y carga neta normalizada de la proteína a pH 7.0 (F). Los valores han sido normalizadas por el valor máximo. Representación del número de residuos ionizables (G), energía total de interacciones carga-carga (H) y carga neta estimada a pH 7.0 (I), en función del tiempo estimado de divergencia de las tiorredoxinas estudiadas. Los tiempos de divergencia están recogidos en (Hedges y Kumar 2009).

Estos resultados sugieren que la distribución de cargas superficiales no parece que sea un buen indicador de la variación de la estabilidad en las tiorredoxinas precámbricas y actuales.

5.2.4. El número de puentes salinos

En las proteínas plegadas, los residuos cercanos con cargas opuestas pueden interactuar y formar un puente salino si se encuentran a una distancia menor de 4 Å. A su vez, estos puentes pueden formar redes a lo largo de la superficie de la proteína. Diversos estudios, tanto experimentales como teóricos, postulan que la contribución de los puentes salinos a la estabilidad pueden ser desde estabilizantes (Xu, Lin et al. 1997; Kumar y Nussinov 1999), insignificantes (Serrano, Horovitz et al. 1990) o incluso desestabilizantes (Sun, Sauer et al. 1991). Por otra parte, la disposición espacial de los residuos que forman el puente salino, parece ser un factor importante en su contribución a la estabilidad del puente salino (Kumar y Nussinov 1999). Aunque no está claro que los puentes salinos y las redes que forman contribuyan significativamente a la estabilidad de proteínas en organismos mesófilos, sí podrían ser responsables en gran medida de la mayor termoestabilidad de las proteínas de organismos termófilos e hipertermófilos (Vieille y Zeikus 2001).

Hemos determinado el número total de puentes salinos presentes en los distintos modelos estructurales y la energía total de interacción asociadas a ellos (Tabla 5.3.). Con los programas *ESBRI* y *Whatif* hemos estimado el número de puentes salinos y con el programa *SATK-ENG* hemos calculado la energía total de interacción (estos cálculos se describe de forma detallada en el apartado 2.2.9.5 de materiales y métodos).

Tabla 5.3. Estimación del número de puentes salinos en las tiorredoxinas precámbricas y actuales (dispuesta de mayor a menor T_m). Además se ha calculado la energía de interacción total de los puentes salinos. El valor negativo de la energía indica que la interacción es estabilizante.

Proteína (PDB)	T_m (°C)	Nº Puentes Salinos	Energía de interacción Total (KJ/mol)
Trx LPBCA (2YJ7)	121.0	8	-34.517
Trx LBCA (4BA7)	113.7	8	-30.521
Trx AECA	113.5	6	-27.735
Trx LACA (2YNX)	113.4	6	-28.012

Trx LGPCA (2YN1)	107.0	7	-34.303
Trx LECA (2Y01)	100.0	5	-14.167
Trx humana (1ERU)	93.0	6	-24.184
Trx LAFCA (2YPM)	91.0	4	-20.036
Trx <i>E. coli</i> (2TRX)	89.0	8	-42.388

No hemos apreciado diferencias significativas en la energía total de interacción debida a puentes salinos. Por ejemplo, la tiorredoxina de *E. coli*, con la T_m más baja, presenta mayor energía total de interacción debida a puentes salinos que la tiorredoxina más estable, teniendo el mismo número de puentes salinos. Aunque sí parece existir cierta correlación con el número de puentes salinos que tendremos que estudiar con más detalle en la estructura. (Fig. 5.7.).

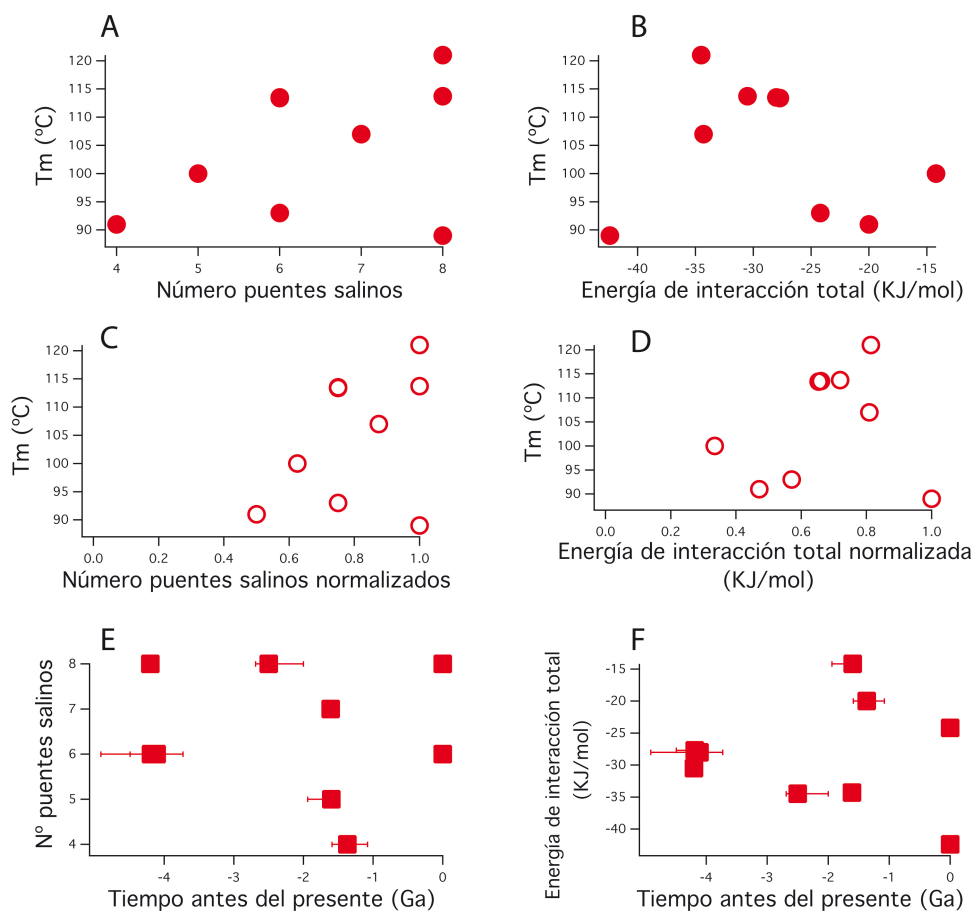


Figura 5.7. Representación de la T_m de las tiorredoxinas precámbricas y actuales de *E. coli* y humana frente al número puentes salinos (A) y la energía total asociada a los puentes salinos de cada proteína (B). Representación de la T_m de las tiorredoxinas analizadas frente al número normalizado por el valor máximo de puentes salinos (C) y la energía total normalizada por el valor máximo asociada a los puentes salinos de cada proteína (D). Representación del número de puentes salinos (E) y la energía de interacción total asociada a los mismos (F) en función del tiempo estimado de divergencia de las tiorredoxinas estudiadas. Los tiempos de divergencia están recogidos en (Hedges y Kumar 2009).

Hay que destacar que las tiorredoxinas de la rama procariota presentan una red conservada de puentes salinos, que implica los residuos 44-96-48-100. Esta red no está en las tiorredoxinas de la rama eucariota. En todos los casos, la energía de interacción era negativa, lo que quiere decir que todas las interacciones eran estabilizantes. Como podemos deducir de estos resultados (Fig. 5.7), el número y la energía de los puentes salinos tampoco puede explicar la variación de la termoestabilidad de las tiorredoxinas ancestrales.

5.2.5. La propensión de aminoácidos en elementos de estructura secundaria

Cuando empezaron a emerger las primeras estructuras cristalinas de proteínas en los años 60, se observó que ciertos aminoácidos aparecían con más frecuencia en los elementos de estructura secundaria (α -hélices, hojas- β y giros- β) (Davies 1964). Basándose en esta información, Chou y Fasman, en los años 70 (Chou y Fasman 1974), elaboraron una escala de propensión de elementos de estructura secundaria, y propusieron un método para predecir, a partir de la secuencia de aminoácidos, dónde se formarán dichos elementos en la estructura nativa.

Desde entonces se ha estudiado ampliamente y con diferentes enfoques la propensión que tienen los distintos aminoácidos de encontrarse en los elementos de estructura secundaria, tanto de alfa hélices (Munoz y Serrano 1995; Luque, Mayorga et al. 1996; Rohl, Chakrabartty et al. 1996; Pace y Scholtz 1998), hojas beta (Minor y Kim 1994), y giros beta (Guruprasad y Rajkumar 2000). Gracias a las diferentes escalas de propensión que se han propuesto en estos trabajos, podemos estimar los residuos más propensos a pertenecer a cada elemento de estructura secundaria e incluso su posición preferida dentro de los mismos. En los últimos años, empleando estas escalas, varios investigadores han conseguido estabilizar proteínas reemplazando un residuo con baja propensión por un elemento de estructura secundaria por otro residuo con mayor propensión para dicho elemento (Zhang, Baase et al. 1992; Munoz, Blanco et al. 1995; Malakauskas y Mayo 1998; Trevino, Schaefer et al. 2007; Fu, Grimsley et al. 2009).

A partir de la asignación de estructura secundaria (apartado 2.2.8.3 del capítulo de Materiales y Métodos), hemos calculado la propensión de cada aminoácido de las

tioredoxinas estudiadas de pertenecer a cada elemento de estructura secundaria (apartado 2.2.6.9 de Materiales y Métodos). Hemos empleado el servidor AGADIR para calcular, en términos de helicidad, la propensión de aminoácidos de situarse en hélices alfa, (Fig. 5.8.A). Para estimar la propensión relativa de aminoácidos de situarse en láminas beta (Fig. 5.8.B) hemos usado la escala termodinámica del Dr. Kim y para estimar la propensión de aminoácidos de situarse en giros beta, en términos de potencial de posición para cada tipo de giro beta, (Fig. 5.8.C) hemos usado la escala estadística del Dr. S. Rajkumar.

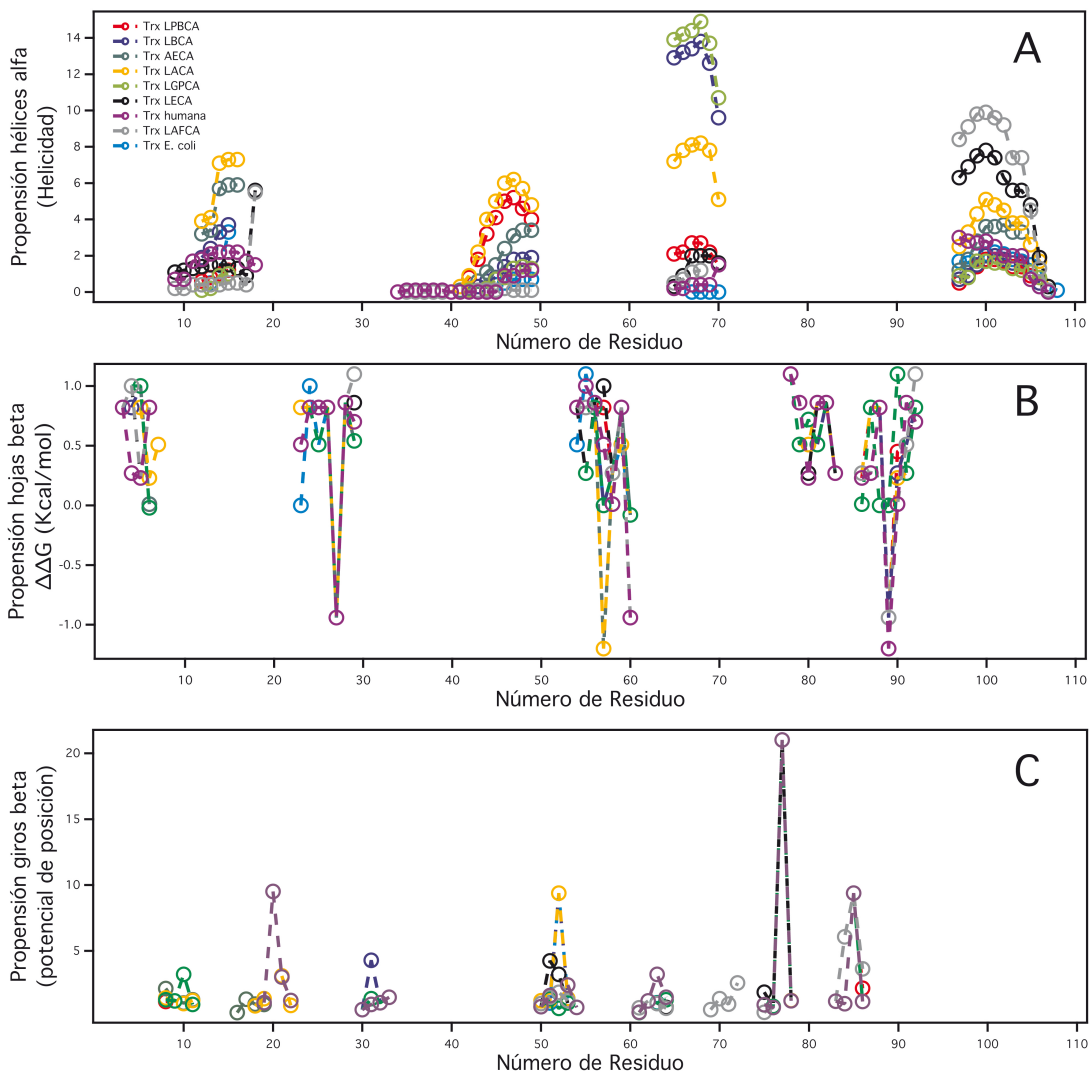


Figura 5.8. Propensión de los aminoácidos de situarse en elementos de estructura secundaria calculada para las tioredoxinas actuales y precámbricas. En la leyenda se han ordenado las tioredoxinas de mayor a menor T_m . A: propensión de los aminoácidos a situarse en alfa hélices. B: propensión de los aminoácidos a situarse en hojas beta. C: propensión de los aminoácidos a situarse en giros beta. En todas tioredoxinas se han asignado los diferentes elementos de estructura secundaria y se ha asignado la propensión para cada aminoácido dependiendo del elemento al que pertenece.

Como se puede apreciar en la figura 5.8, en el caso de las hojas beta y giros beta la propensión de los aminoácidos son similares para todas las tiorredoxinas. En el caso de la propensión de aminoácidos de alfa hélice (helicidad, según AGADIR) observamos diferencias entre unas tiorredoxinas y otras. El ejemplo más claro lo encontramos en las en las hélices $\alpha 2$ y $\alpha 3$ de las tiorredoxinas LACA, AECA, LBCA y LPBCA, que tienen más helicidad que las demás tiorredoxinas en estas hélices. Este resultado es interesante porque estas cuatro tiorredoxinas son las más estables. Además, las tiorredoxinas LACA, AECA y LBCA también tienen mayor propensión en la hélice $\alpha 1$. Por otro lado, las tiorredoxinas ancestrales menos estables (LAFCA y LECA) tiene mayor propensión en la hélice 4, por lo que esta hélice podría no tener mucha repercusión en la estabilidad. Los residuos de las cuatro hélices alfa de las tiorredoxinas actuales presentan escasa o nula propensión de situarse en alfa hélices. Estos resultados sugieren que este parámetro podría estar contribuyendo a aumentar la estabilidad de las tiorredoxinas ancestrales. De hecho, hemos observado cierta correlación al representar la temperatura de desnaturalización en función de la helicidad media de cada tiorredoxina estudiada (Fig. 5.9A.), que no se observa en el caso de las hojas beta y los giros beta (Fig. 5.9).

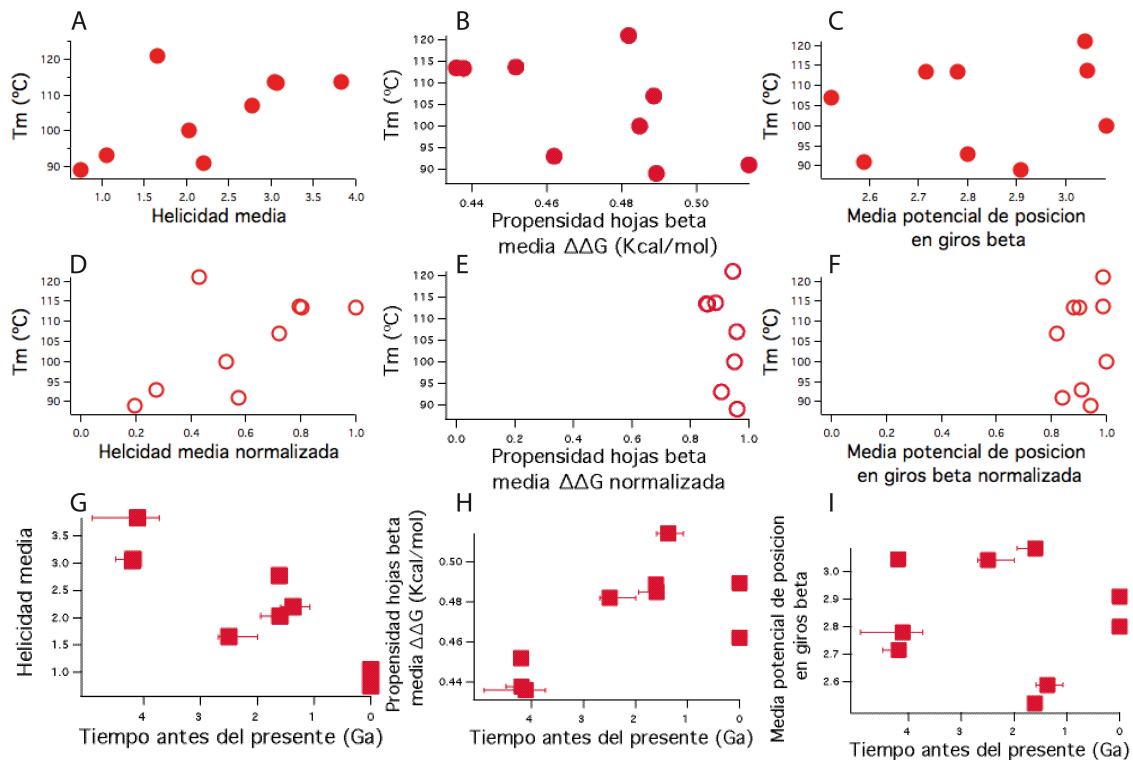


Figura 5.9. A-C: Representación de la T_m de las tiorredoxinas precámbricas y actuales de *E. coli* y humana frente a la media de helicidad obtenida del programa AGADIR (A), a la media propensión relativa de hojas beta obtenida de la escala termodinámica (Minor y Kim 1994) (B) y media del potencial de posición de los residuos en diferentes tipos de giros beta (C). D-E: Representación de la T_m frente a helicidad normalizada (D), a propensión de hoja beta normalizada (E) y potencial de posición normalizado de giro beta (F). Los valores han sido normalizadas por el valor máximo. Representación de la helicidad media (G), propensión de hoja beta (H) y carga neta estimada a pH 7.0, en función del tiempo estimado de divergencia de las tiorredoxinas estudiadas. Los tiempos de divergencia están recogidos en (Hedges y Kumar 2009).

5.2.6. El número de prolinas y glicinas

Está aceptado que las prolinas y glicinas tienen efectos contrarios sobre la entropía conformacional del estado desplegado de las proteínas (y, por tanto sobre la estabilidad de las proteínas). El laboratorio del Dr. B. Matthews propuso que se puede aumentar la estabilidad de proteínas (de estructura tridimensional conocida) disminuyendo la entropía conformacional por adición de prolinas o eliminación de glicinas en la secuencia de aminoácidos (Matthews, Nicholson et al. 1987). En el caso de la prolina, el aumento de estabilidad puede ser consecuencia la formación de estructuras residuales en el estado desplegado, lo cual disminuye la entropía conformacional. Esto es posible porque la prolina tiene su cadena lateral formado un anillo pirrolidina, lo cual restringe el número de conformaciones que puede adoptar, restringiendo a su vez las conformaciones del residuo que le precede (Schimmel y Flory 1968). Además, se ha demostrado que un incremento en el

número de prolinas, especialmente si están colocadas en posiciones específicas de la estructura, contribuye a aumentar la estabilidad de las proteínas (Watanabe, Masuda et al. 1994). Por el contrario, la glicina tiene el mayor grado de entropía conformacional de todos los aminoácidos (Ramachandran, Ramakrishnan et al. 1963). Por este motivo, este aminoácido está relacionado con el aumento de la entropía conformacional del estado desplegado de la proteína, lo cual puede disminuir su estabilidad térmica (Matthews, Nicholson et al. 1987).

Para estimar el posible efecto de las glicinas y prolinas en la estabilidad de las tiorredoxinas ancestrales, hemos determinado el número de las prolinas y glicinas en las secuencias de tiorredoxinas ancestrales y tiorredoxinas actuales de *E. coli* y humana (Fig. 5.10.). Para ver si existe alguna correlación entre la estabilidad térmica (T_m) y el número de estos residuos, hemos representado la temperatura de desplegamiento frente al número de Pro (Fig. 5.11.A) y frente a número de Gly (Fig. 5.11.B) para cada una de las proteínas ancestrales y actuales de *E. Coli* y humana.

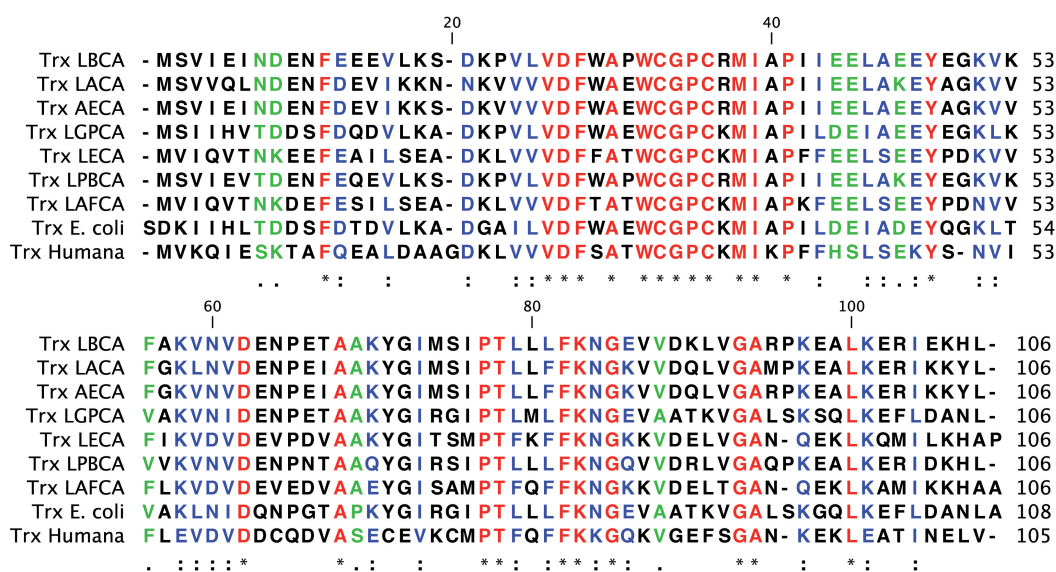


Figura 5.10. Alineamiento de secuencias de tiorredoxinas ancestrales y actuales (*E. coli* y humana). En rojo y "*" se muestran los residuos idénticos, en azul y ":" los residuos conservados y en verde y "." los residuos semiconservados.

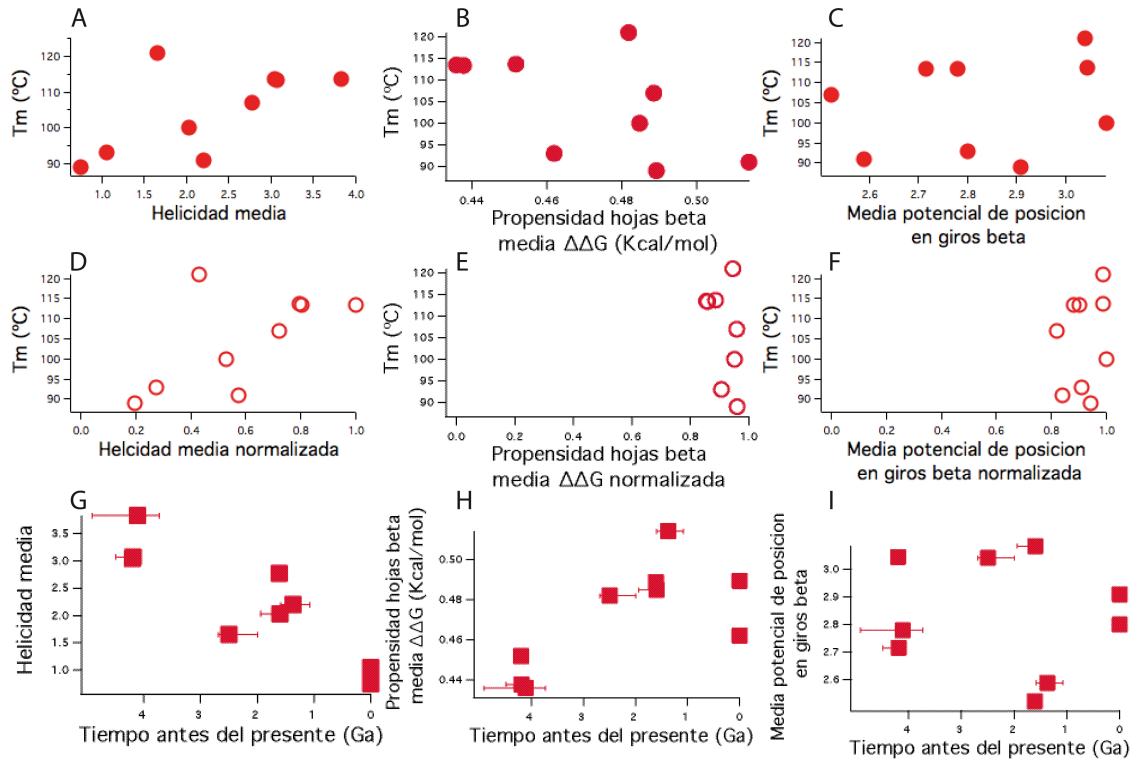


Figura 5.11. A: Representación de la temperatura de desnaturalización (T_m) frente al número de Pro de las tiorredoxinas estudiadas. B: Representación de la T_m frente al número de Gly de las tiorredoxinas estudiadas. C: Representación de la T_m frente al número normalizado de Pro. D: Representación de la T_m frente al número normalizado de Gly. Representación del nº de Pro (E) y Gly (F) en función del tiempo estimado de divergencia para las tiorredoxinas estudiadas. Los tiempos de divergencia están recogidos en (Hedges y Kumar 2009).

Los datos sugieren que el número de prolinas podrían contribuir en mayor medida a la estabilidad que el número de glicinas (figura 5.11. A-D). Aunque para ambos parámetros observamos una buena correlación. Sin embargo, para comprobar esta hipótesis es necesario realizar un estudio mutacional experimental más detallado.

En resumen, no hemos apreciado diferencias evidentes en términos de área polar y apolar accesible al disolvente, número de enlaces de hidrógenos y distribución de cargas superficiales ni propensión de aminoácidos de situarse en hojas y giros beta cuando hemos comparado las estructuras de las tiorredoxinas ancestrales entre sí o con las tiorredoxinas actuales de *E. coli* y humana. Por supuesto, estas conclusiones debemos tomarlas con mucha precaución, ya que la contribución total de algunos de estos parámetros a la energética de plegamiento de la proteína es del orden de cientos de KJ/mol, y pequeñas diferencias (por debajo del margen de precisión de estos cálculos) podrían tener una contribución significativa a la estabilidad de las tiorredoxinas.

Por otro lado, parece que existe cierta correlación de la estabilidad de las tiorredoxinas con el número de puentes salinos, la propensión de residuos de situarse en hélices alfa y el número de glicinas y prolinas. Además, hemos observado una correlación de la T_m con el volumen de las cavidades detectadas en las distintas tiorredoxinas que no podemos explicar, pues hemos percibido que las proteínas más estables presentan cavidades con mayor volumen.

Para estimar matemáticamente el posible efecto de la contribución estos factores a la estabilidad tendríamos que realizar un ajuste a los datos de un modelo de regresión lineal múltiple, minimizando la suma del error cuadrático de los mismos. No obstante, para comprobar cualquier resultado que pudiésemos obtener, son necesarios estudios experimentales con el fin de evaluar de manera empírica la contribución de estos factores a la estabilidad de las tiorredoxinas.

5.3. Bibliografía

- Chou, P. Y. and G. D. Fasman (1974). "Conformational Parameters for Amino Acids in Helical, Beta-Sheet, and Random Coil Regions Calculated from Proteins." Biochemistry **13**(2): 211-222.
- Connolly, M. L. (1983). "Solvent-Accessible Surfaces of Proteins and Nucleic Acids." Science **221**(4612): 709-713.
- Creamer, T. P., R. Srinivasan and G. D. Rose (1995). "Modeling Unfolded States of Peptides and Proteins." Biochemistry **34**(50): 16245-16250.
- Creamer, T. P., R. Srinivasan and G. D. Rose (1997). "Modeling Unfolded States of Proteins and Peptides. Ii. Backbone Solvent Accessibility." Biochemistry **36**(10): 2832-2835.
- Davies, D. R. (1964). "A Correlation between Amino Acid Composition and Protein Structure." Journal of molecular biology **9**: 605-609.
- Eriksson, A. E., W. A. Baase, X. J. Zhang, D. W. Heinz, M. Blaber, E. P. Baldwin and B. W. Matthews (1992). "Response of a Protein Structure to Cavity-Creating Mutations and Its Relation to the Hydrophobic Effect." Science **255**(5041): 178-183.
- Fu, H., G. R. Grimsley, A. Razvi, J. M. Scholtz and C. N. Pace (2009). "Increasing Protein Stability by Improving Beta-Turns." Proteins **77**(3): 491-498.
- Gao, J., D. A. Bosco, E. T. Powers and J. W. Kelly (2009). "Localized Thermodynamic Coupling between Hydrogen Bonding and Microenvironment Polarity Substantially Stabilizes Proteins." Nat Struct Mol Biol **16**(7): 684-690.
- Guruprasad, K. and S. Rajkumar (2000). "Beta-and Gamma-Turns in Proteins Revisited: A New Set of Amino Acid Turn-Type Dependent Positional Preferences and Potentials." Journal of biosciences **25**(2): 143-156.
- Hedges, S. B. and S. Kumar (2009). The Timetree of Life. Oxford, Oxford Univ. Press.
- Ibarra-Molero, B., V. V. Loladze, G. I. Makhatadze and J. M. Sanchez-Ruiz (1999). "Thermal Versus Guanidine-Induced Unfolding of Ubiquitin. An Analysis in Terms of the Contributions from Charge-Charge Interactions to Protein Stability." Biochemistry **38**(25): 8138-8149.
- Ibarra-Molero, B., G. I. Makhatadze and J. M. Sanchez-Ruiz (1999). "Cold Denaturation of Ubiquitin." Biochim Biophys Acta **1429**(2): 384-390.
- Kauzmann, W. (1954). Symposium on the Mechanism of Enzyme Action. Denaturation of proteins and enzymes. W. McElroy y B. Glass. Baltimore, McCollum-Pratt Institute, Johns Hopkins University Press: 70-110.
- Kauzmann, W. (1959). "Some Factors in the Interpretation of Protein Denaturation." Advances in protein chemistry **14**: 1-63.

- Kumar, S. and R. Nussinov (1999). "Salt Bridge Stability in Monomeric Proteins." Journal of molecular biology **293**(5): 1241-1255.
- Kumar, S., C. J. Tsai and R. Nussinov (2000). "Factors Enhancing Protein Thermostability." Protein Eng **13**(3): 179-191.
- Loladze, V. V., B. Ibarra-Molero, J. M. Sanchez-Ruiz and G. I. Makhatadze (1999). "Engineering a Thermostable Protein Via Optimization of Charge-Charge Interactions on the Protein Surface." Biochemistry **38**(50): 16419-16423.
- Luque, I., O. L. Mayorga and E. Freire (1996). "Structure-Based Thermodynamic Scale of Alpha-Helix Propensities in Amino Acids." Biochemistry **35**(42): 13681-13688.
- Malakauskas, S. M. and S. L. Mayo (1998). "Design, Structure and Stability of a Hyperthermophilic Protein Variant." Nature structural biology **5**(6): 470-475.
- Matthews, B. W., H. Nicholson and W. J. Becktel (1987). "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 **84**(19): 6663-6667.
- Minor, D. L., Jr. and P. S. Kim (1994). "Measurement of the Beta-Sheet-Forming Propensities of Amino Acids." Nature **367**(6464): 660-663.
- Munoz, V., F. J. Blanco and L. Serrano (1995). "The Hydrophobic-Staple Motif and a Role for Loop-Residues in Alpha-Helix Stability and Protein Folding." Nat Struct Biol **2**(5): 380-385.
- Munoz, V. and L. Serrano (1995). "Elucidating the Folding Problem of Helical Peptides Using Empirical Parameters. Ii. Helix Macrodipole Effects and Rational Modification of the Helical Content of Natural Peptides." Journal of molecular biology **245**(3): 275-296.
- Myers, J. K. and C. N. Pace (1996). "Hydrogen Bonding Stabilizes Globular Proteins." Biophys J **71**(4): 2033-2039.
- Pace, C. N. (2009). "Energetics of Protein Hydrogen Bonds." Nat Struct Mol Biol **16**(7): 681-682.
- Pace, C. N., H. Fu, K. L. Fryar, J. Landua, S. R. Trevino, B. A. Shirley, M. M. Hendricks, S. Iimura, K. Gajiwala, J. M. Scholtz and G. R. Grimsley (2011). "Contribution of Hydrophobic Interactions to Protein Stability." J Mol Biol.
- Pace, C. N. and J. M. Scholtz (1998). "A Helix Propensity Scale Based on Experimental Studies of Peptides and Proteins." Biophysical journal **75**(1): 422-427.
- Perez-Jimenez, R., R. Godoy-Ruiz, B. Ibarra-Molero and J. M. Sanchez-Ruiz (2005). "The Effect of Charge-Introduction Mutations on E. Coli Thioredoxin Stability." Biophys Chem **115**(2-3): 105-107.
- Perez-Jimenez, R., A. Ingles-Prieto, Z. M. Zhao, I. Sanchez-Romero, J. Alegre-Cebollada, P. Kosuri, S. Garcia-Manyes, T. J. Kappock, M. Tanokura, A. Holmgren, J. M. Sanchez-Ruiz, E. A. Gaucher and J. M. Fernandez (2011).

- "Single-Molecule Paleoenzymology Probes the Chemistry of Resurrected Enzymes." Nature structural & molecular biology **18**(5): 592-596.
- Ramachandran, G. N., C. Ramakrishnan and V. Sasisekharan (1963). "Stereochemistry of Polypeptide Chain Configurations." Journal of molecular biology **7**: 95-99.
- Rohl, C. A., A. Chakrabartty and R. L. Baldwin (1996). "Helix Propagation and N-Cap Propensities of the Amino Acids Measured in Alanine-Based Peptides in 40 Volume Percent Trifluoroethanol." Protein science : a publication of the Protein Society **5**(12): 2623-2637.
- Sanchez-Ruiz, J. M. and G. I. Makhatadze (2001). "To Charge or Not to Charge?" Trends Biotechnol **19**(4): 132-135.
- Schellman, J. A. (1987). "The Thermodynamic Stability of Proteins." Annual review of biophysics and biophysical chemistry **16**: 115-137.
- Schimmel, P. R. and P. J. Flory (1968). "Conformational Energies and Configurational Statistics of Copolypeptides Containing L-Proline." Journal of molecular biology **34**(1): 105-120.
- Serrano, L., A. Horovitz, B. Avron, M. Bycroft and A. R. Fersht (1990). "Estimating the Contribution of Engineered Surface Electrostatic Interactions to Protein Stability by Using Double-Mutant Cycles." Biochemistry **29**(40): 9343-9352.
- Shirley, B. A., P. Stanssens, U. Hahn and C. N. Pace (1992). "Contribution of Hydrogen Bonding to the Conformational Stability of Ribonuclease T1." Biochemistry **31**(3): 725-732.
- Sun, D. P., U. Sauer, H. Nicholson and B. W. Matthews (1991). "Contributions of Engineered Surface Salt Bridges to the Stability of T4 Lysozyme Determined by Directed Mutagenesis." Biochemistry **30**(29): 7142-7153.
- Sundd, M., N. Iverson, B. Ibarra-Molero, J. M. Sanchez-Ruiz and A. D. Robertson (2002). "Electrostatic Interactions in Ubiquitin: Stabilization of Carboxylates by Lysine Amino Groups." Biochemistry **41**(24): 7586-7596.
- Suzuki, Y., K. Hatagaki and H. Oda (1991). "A Hyperthermostable Pullulanase Produced by an Extreme Thermophile, *Bacillus Flavocaldarius* Kp 1228, and Evidence for the Proline Theory of Increasing Protein Thermostability." Appl Microbiol Biotechnol **34**(6): 707-714.
- Taverna, D. M. and R. A. Goldstein (2002). "Why Are Proteins Marginally Stable?" Proteins **46**(1): 105-109.
- Trevino, S. R., S. Schaefer, J. M. Scholtz and C. N. Pace (2007). "Increasing Protein Conformational Stability by Optimizing Beta-Turn Sequence." J Mol Biol **373**(1): 211-218.
- Vieille, C. and G. J. Zeikus (2001). "Hyperthermophilic Enzymes: Sources, Uses, and Molecular Mechanisms for Thermostability." Microbiol Mol Biol Rev **65**(1): 1-43.

Wada, A. and H. Nakamura (1981). "Nature of the Charge Distribution in Proteins." Nature **293**(5835): 757-758.

Watanabe, K., T. Masuda, H. Ohashi, H. Mihara and Y. Suzuki (1994). "Multiple Proline Substitutions Cumulatively Thermostabilize Bacillus Cereus Atcc7064 Oligo-1,6-Glucosidase. Irrefragable Proof Supporting the Proline Rule." Eur J Biochem **226**(2): 277-283.

Xu, D., S. L. Lin and R. Nussinov (1997). "Protein Binding Versus Protein Folding: The Role of Hydrophilic Bridges in Protein Associations." Journal of molecular biology **265**(1): 68-84.

Zhang, X. J., W. A. Baase and B. W. Matthews (1992). "Multiple Alanine Replacements within Alpha-Helix 126-134 of T4 Lysozyme Have Independent, Additive Effects on Both Structure and Stability." Protein science : a publication of the Protein Society **1**(6): 761-776.

6. ESTUDIO EXPERIMENTAL DE LA ESTABILIDAD TÉRMICA DE LA TIORREDOXINA ANCESTRAL LPBCA

6.1. Introducción.

Como hemos visto en el capítulo anterior, al comparar las estructuras de tiorredoxinas precámbricas con las tiorredoxinas actuales (de *E. coli* y la humana), no hemos apreciado diferencias significativas en términos de área superficial accesible al disolvente, empaquetamiento hidrofóbico, número de enlaces de hidrógeno y puentes salinos, distribución de carga superficial y propensión en hojas beta y giros beta. No obstante, parece haber diferencias en el número de prolina y glicinas, así como en la propensión de residuos de situarse en hélices alfa. Por motivos metodológicos, hemos optado por estudiar experimentalmente la posible contribución de las prolina y glicinas en la estabilidad de la tiorredoxina de LPBCA y *E. coli*.

Por otro lado, al comparar las secuencias de estas dos tiorredoxinas, nos hemos percatado de la presencia de una mutación muy poco frecuente a lo largo de la evolución. Esta mutación ha sido de una leucina (en LPBCA) a una lisina (en *E. coli*) en la posición 90. Algunos trabajos recientes han subrayado que, en general, este tipo de mutaciones no conservativas modifican la estabilidad de las proteínas (Bershtein, Segal et al. 2006; Zeldovich, Chen et al. 2007; Tokuriki y Tawfik 2009; Tokuriki y Tawfik 2009). Además, los estudios experimentales muestran que la mayoría de estas mutaciones son desestabilizantes y, en consecuencia, su acumulación pueden comprometer el denominado "fitness" de la proteína. Consecuentemente, las mutaciones no conservativas podrían causar mutaciones de residuos cercanos en la estructura (por ejemplo, para acomodar al nuevo residuo), las cuales podrían afectar también a la estabilidad de la proteína.

En este capítulo, hemos tratado de identificar experimentalmente los mecanismos moleculares responsables de la hiperestabilidad de la tiorredoxina precámbrica LPBCA. Para llevar a cabo este estudio, hemos realizado un análisis mutacional de la hiperestabilidad de la tiorredoxina LPBCA (la más estable) comparada con la tiorredoxina de *E. coli* (la menos estable), con el objetivo de identificar los aminoácidos responsables de la diferencia de estabilidad (33°C) entre ambas tiorredoxinas. Concretamente, nos hemos centrado en estudiar el número de glicinas y prolina, así como en la mutación no conservativa K90L y mutaciones próximas a esta posición en las estructuras. Hemos abordado este análisis con un enfoque basado en la construcción y caracterización de bibliotecas combinatoriales de variantes de tiorredoxinas, seguida de la reconstrucción de la población

mutacional de la biblioteca con el método de regresión de mínimos cuadrados parciales.

6.2. Resultados y Discusión

6.2.1. Evaluación del papel de las prolinas y glicinas en la estabilidad de tiorredoxina de *E. coli* y LPBCA

Como primer paso en el análisis del posible efecto de las prolinas y glicinas en la diferencia de estabilidad entre la tiorredoxina de LPBCA (121 °C) y la trx de *E. coli* (~87°C), hemos determinado el número total de glicinas y prolinas de estas dos tiorredoxinas (Tabla 6.1).

Tabla 6.1. Número total de prolinas y glicinas presentes en la secuencia de las tiorredoxinas de *E. coli* y LPBCA.

Proteína	Nº Prolinas	Nº Glicinas
Trx LPBCA	7	5
Trx <i>E. coli</i>	5	9

Podemos observar claramente como el número total de Pro y Gly es muy diferente en ambas tiorredoxinas. Esta diferencia es mucho más marcada en el caso del número de glicinas, y podría ser la responsable de la menor estabilidad de la tiorredoxina de *E. coli*. Para estudiar la posición y conservación de estos residuos en ambas tiorredoxinas, hemos realizado un alineamiento de secuencias de ambas proteínas (Fig. 6.1.).

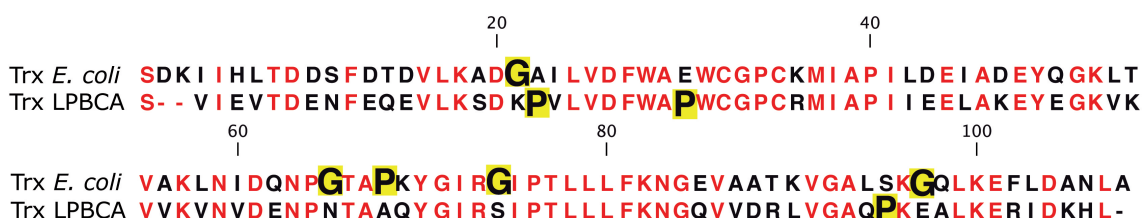


Figura 6.1. Alineamiento de secuencias de trx de *E. coli* y trx de LPBCA. El alineamiento se ha realizado con el programa *CLC Sequence viewer*. En rojo se muestran las posiciones idénticas en ambas secuencias. Destacado en amarillo están las posiciones no conservadas donde aparecen residuos de Pro o Gly, que marcan las posiciones susceptibles de ser mutadas.

Para llevar a cabo el análisis mutacional de la hiperestabilidad de la tiorredoxina LPBCA, a partir del alineamiento de secuencias (Fig. 6.1.), hemos seleccionado un total de 8 posiciones susceptibles de ser mutadas, tomando a la trx de *E. coli* como *background*. Con estas 8 posiciones, hemos construido una biblioteca combinatorial de variantes de tiorredoxina de *Escherichia coli* (apartado 2.2.4. de Materiales y Métodos), de manera que hemos identificado las posiciones en que aparezca una Pro o Gly no conservada en ambas secuencias (resaltadas en amarillo en figura 6.1) y las hemos sustituido por los residuos que aparecen en trx de LPBCA en dichas posiciones. Esta biblioteca contendrá, por tanto, 256 variantes de tiorredoxina con todas las posibles combinaciones de mutación.

Posteriormente, hemos purificado alrededor de un 10% de las variantes de tiorredoxina de esta biblioteca (apartado 2.2.3.1 de Materiales y métodos) y hemos caracterizado su estabilidad térmica (T_m) mediante calorimetría diferencial de barrido (apartado 2.2.6. de Materiales y Métodos). Para caracterizar una muestra representativa de la biblioteca, hemos seleccionado 45 colonias de bacterias aleatoriamente. De estas 45 colonias, solo 31 han expresado tiorredoxina, como se pudo apreciar en los controles de sobreexpresión. Tras la secuenciación de estas variantes de tiorredoxina, solo obtuvimos la secuencia completa de 24. Una vez identificadas las mutaciones que presenta cada variante de tiorredoxina, hemos realizado la determinación de su T_m mediante DSC (Fig. 6.2).

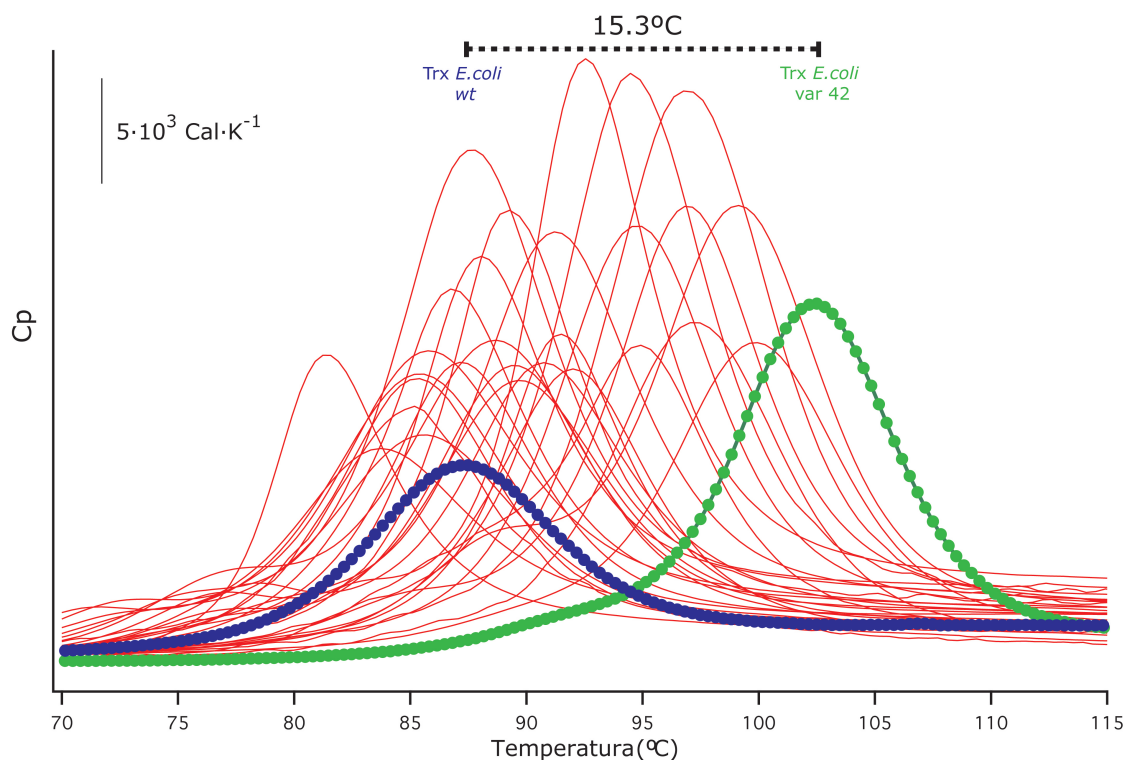


Figura 6.2. Termogramas de DSC de las variantes analizadas experimentalmente de la biblioteca combinatorial de tiorredoxina de *E. coli*. En azul se representa el termograma correspondiente a la tiorredoxina de *E. coli*. En rojo se representan los termogramas de las variantes de tiorredoxina de la biblioteca combinatorial. En verde se representa la variante de la biblioteca más estable. Los experimentos se realizaron a una velocidad de barrido de 150 K/hora y a diversas concentraciones de proteína (0.3-0.7 mg/ml). Los termogramas de las diferentes proteínas no están normalizadas por la concentración, de ahí la diferencia en la altura de los picos.

Los datos de DSC muestran que hay algunas variantes de tiorredoxina con una T_m superior a la tiorredoxina de *E. coli* (Fig. 6.2). De todas ellas, la variante 42 presenta la mayor T_m (102.5°C). Esta variante posee solamente 4 mutaciones (A22P, P68A, G74S, G97E) que aumentan su estabilidad 15.3°C respecto a *trx* de *E. coli* (Tabla 6.2).

Tabla 6.2. Variantes de tiorredoxina examinadas experimentalmente y su temperatura de desnaturalización (T_m). Para cada mutante, se muestra si la posición ha sido mutada en la secuencia (gris) o no (blanco). También se muestra la T_m de la *trx* de *E. coli*, que hemos empleado como *background*. Resaltado en amarillo está la variante de tiorredoxina con la T_m experimental más elevada.

# Variante	G21K	A22P	E30P	G65N	P68A	G74S	S95P	G97E	T_m (°C)
<i>E.coli trx</i>									87.3
1									94.9
2									97.3
5									89.3
7									94.5
8									93.1

12									89.7
13									91.2
14									85.9
15									88.0
16									99.1
17									96.9
19									88.7
22									89.5
23									87.6
24									91.9
25									83.8
27									94.7
32									85.4
36									85.3
39									90.8
40									99.8
41									91.5
42									102.5
43									81.3

Nuestro objetivo tras el análisis experimental de una pequeña población de la biblioteca combinatorial de variantes de tiorredoxina de *E. coli* es tratar de reconstruir la población completa de mutantes de la biblioteca; es decir, predecir la T_m de las 256 posibles variantes de tiorredoxinas. Para ello, hemos empleado el método estadístico de regresión por mínimos cuadrados parciales (PLS-R, de sus siglas en inglés). Este método está explicado en detalle en el apartado 2.2.10. de Materiales y Métodos. Cabe destacar que antes del análisis PLS hemos normalizado los datos de T_m (variable dependiente). Las variables independientes las obtenemos a partir de información de las secuencias, de manera que tomará valores de 1 si hay presencia de mutación y 0 si no hay mutación en una posición determinada.

Tras obtener la matriz de datos para las variantes analizadas experimentalmente, hemos utilizado el método de *bootstrapping* (apartado 2.2.10.2 de Materiales y Métodos) para minimizar el posible error experimental. Usando este método, hemos generado 20 réplicas de los datos aleatorias, permitiendo repeticiones y sin variar el número de muestra.

Una vez que ajustamos el modelo a cada una de las réplicas generadas (empleando la ecuación 2.18 de Materiales y Métodos), es relativamente sencillo predecir los datos de la temperatura de desnaturalización de todas las variantes de la biblioteca (Fig. 6.3). Para esta predicción hemos utilizado el número óptimo de variables latentes calculadas por el programa. Aunque este número depende de la réplica

usada en la regresión, en todos los casos ha fluctuado entre 3 y 9, valores considerablemente menores que el número de variables independientes.

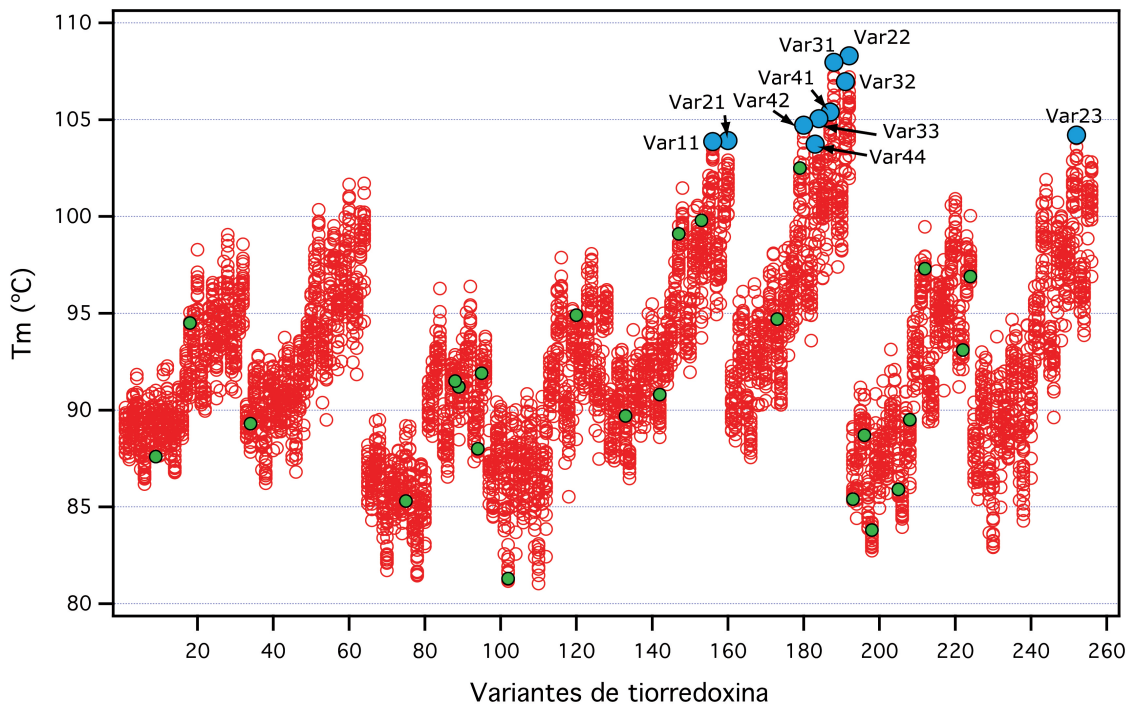


Figura 6.3. Gráfica de reconstrucción del espacio mutacional completo de la biblioteca de prolina y glicina. En verde se representa el valor de T_m experimental para las 24 variantes de tiorredoxina analizadas experimentalmente. En rojo se representan los valores predichos de T_m para cada variante de tiorredoxina. Los círculos azules rellenos corresponden a las 10 variantes de tiorredoxina con mayor T_m predicha.

A partir de las predicciones, es simple identificar aquellas variantes de la biblioteca con la mayor T_m predicha. En esta ocasión, hemos seleccionado las 10 variantes con mayor T_m predicha (Fig. 6.3) y hemos visto qué mutaciones presentan (tabla 6.3). Estas nuevas variantes (más la que tiene las 8 mutaciones) se generaron por la compañía TOP gene a partir de algunas de las variantes estudiadas experimentalmente. Luego, las hemos purificado y hemos determinado su T_m mediante DSC (Fig. 6.4).

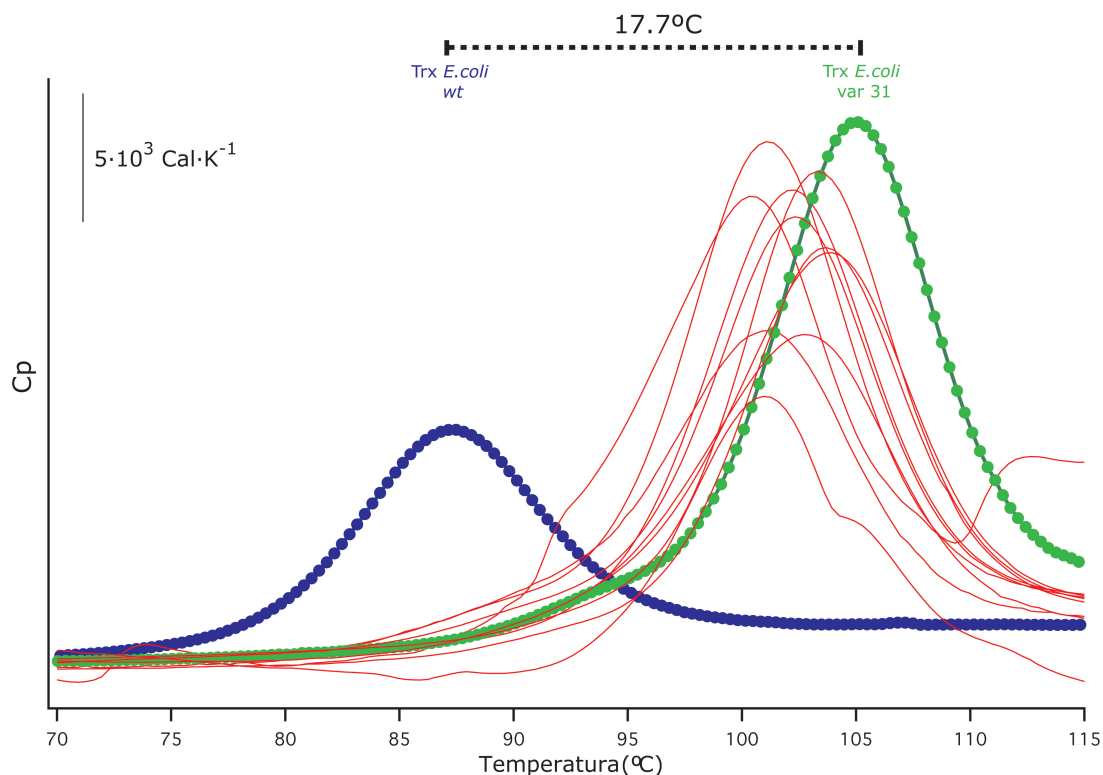


Figura 6.4. Termogramas de DSC de las variantes con mayor T_m predicha de las variantes de la biblioteca combinatorial de tiorredoxina de *E. coli*. En azul se representa el termograma correspondiente a la tiorredoxina de *E. coli*. En rojo se representan los termogramas de las variantes de tiorredoxina que hemos analizado experimentalmente. En verde se representa la variante de la biblioteca más estable. Los experimentos se realizaron a una velocidad de barrido de 150 K/hora y a diversas concentraciones de proteína (0.3-0.7 mg/ml). Los termogramas de las diferentes proteínas no están normalizadas por la concentración, de ahí la diferencia en la altura de los picos.

Cuando analizamos la T_m experimental, observamos que el mutante más estable es la variante 31 con 105°C (y no la variante 22 como predijo el modelo). Esta variante de tiorredoxina tiene 6 de las 8 mutaciones posibles (Tabla 6.3) y ha aumentado la estabilidad 17.7°C.

Los datos de T_m de las 256 posibles variantes de la biblioteca predichos a partir de las 24 variantes originales son muy parecidos a los obtenidos experimentalmente (Tabla 6.3.). Podemos afirmar, por tanto, que la predicción del modelo ha sido buena, ya que la variante más estable experimentalmente es la segunda variante más estable predicha.

Tabla 6.3. Variantes de tiorredoxina examinadas experimentalmente y su temperatura de desnaturalización (T_m). Para cada mutante, se muestra si la posición ha sido mutada en la secuencia (gris) o no (blanco). Resaltado en amarillo está la variante de tiorredoxina con la T_m experimental más elevada.

Variante	G21K	A22P	E30P	G65N	P68A	G74S	S95P	G97E	T_m predicha (°C)	T_m experimental (°C)
Var22									108.3	104.0
Var31									108.0	105.0
Var32									107.0	102.3
Var41									105.4	103.4
Var33									105.0	102.7
Var42									104.7	103.7
Var23									104.2	100.9
Var11									103.9	101.1
Var21									103.9	102.2
Var44									103.7	101.0
Var12									102.8	99.9

En resumen, la variante de tiorredoxina más estable (Var31) presenta las mutaciones: G21K, A22P, G65N, P68A, G74S, G97E (Fig. 6.5). De todas las posibles mutaciones, esta variante tiene todos los reemplazamientos de glicina. Este resultado aparentemente apoya la hipótesis de que las glicinas aumentan la entropía conformacional del estado desplegado de la proteínas, pudiendo contribuir a disminuir la estabilidad de las mismas (Matthews, Nicholson et al. 1987). El caso de las mutaciones en que intervienen prolina las conclusiones no son tan claras. Solo una mutación de adición de prolina parece repercutir en un aumento de la estabilidad (A22P). En el caso de la mutación P68A está más claro, ya que la prolina 68 se localiza en medio de una alfa hélice en la estructura de la tiorredoxina de *E. coli*. La incapacidad de prolina para formar los puentes de hidrogeno necesarios hace que la presencia de este residuo impida (o dificulte) la formación de la hélice. Además, su cadena lateral puede presentar impedimentos estéricos con el esqueleto de la hélice, lo que provocaría una flexión del eje de la hélice de $\sim 30^\circ\text{C}$ (Richardson 1981). Por ello, la mutación a otro residuo (en este caso a alanina), puede contribuir a relajar los impedimentos estéricos, lo que se puede traducir en un aumento de la estabilidad.

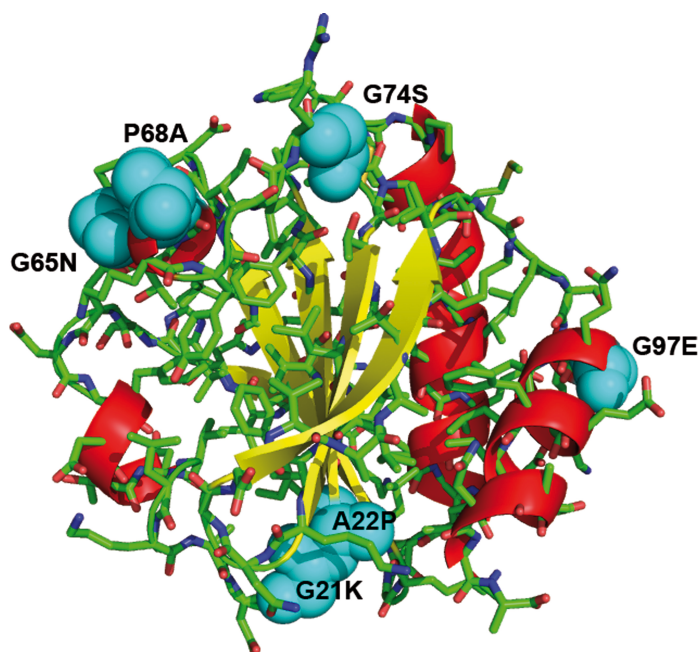


Figura 6.5. Estructura tridimensional de la tiorredoxina de *E.coli* (PDB: 2TRX). En la estructura se han representado los elementos de estructura secundaria asignados con el programa DSSP. Se destacan (esferas-azules) los residuos de la tiorredoxina de *E. coli* que han sido mutados en la variante más estable (var31).

Los datos de T_m experimental de las variantes de tiorredoxina más estables, los hemos incorporado al juego de datos de las otras 24 variantes para generar un nuevo juego de datos. A continuación, hemos generado de nuevo 20 réplicas por *bootstrapping* y hemos ajustado el modelo anterior a este nuevo set de datos para predecir los valores de T_m de toda la biblioteca. Las predicciones resultantes de este segundo ajuste se ajustan más a la realidad. En esta ocasión, se identifica la variante 31 como la más estable y la variante 22 como la segunda más estable, tal como indican los datos experimentales. Además, solamente aparece una nueva variante entre las 10 más estables, cuyo valor de T_m predicho es el más bajo. Por este motivo, no hemos considerado oportuno determinar experimentalmente su termoestabilidad. Este procedimiento supone una validación experimental de todo nuestro enfoque.

6.2.2. Evaluación del papel de la mutación no conservativa K90L en la estabilidad de tiorredoxina de *E. coli* y LPBCA

Hasta el momento, hemos obtenido una variante de tiorredoxina $\sim 18^{\circ}\text{C}$ más estable que la tiorredoxina de *E. coli*. Este aumento en la estabilidad se ha conseguido con tan solo 6 mutaciones (G21K, A22P, G65N, P68A, G74S, G97E). Para tratar de explicar la diferencia de estabilidad restante entre la tiorredoxina de *E. coli* y la de LPBCA ($\sim 16^{\circ}\text{C}$), hemos vuelto a fijarnos en el alineamiento de secuencias, ya que ha llamado nuestra atención la presencia de mutaciones poco frecuentes (no conservativas) entre ambas secuencias. El ejemplo más evidente es la sustitución de una leucina (en *trx* de LPBCA) por una lisina (en *trx* de *E. coli*) en la posición 90. Como hemos señalado en la introducción de este capítulo, este tipo de cambios podría tener repercusiones en la estabilidad de la tiorredoxina.

Para estimar el número de mutaciones no conservativas, a partir del alineamiento de secuencias de ambas tiorredoxinas (Fig. 6.1), hemos determinado qué mutaciones son conservativas y cuales no lo son, empleando la matriz PAM 250. En esta matriz quedan recogidas las frecuencias con las que han ocurrido las mutaciones a lo largo de la evolución. Se basa en una base de datos que recoge 1572 cambios dentro de 71 grupos de proteínas altamente relacionadas entre sí (Dayhoff, Schwartz et al. 1978). Las mutaciones neutras tienen un coeficiente cero, mientras que mutaciones conservativas tienen coeficientes positivos y las no conservativas tienen un coeficiente negativo (Fig. 6.6).

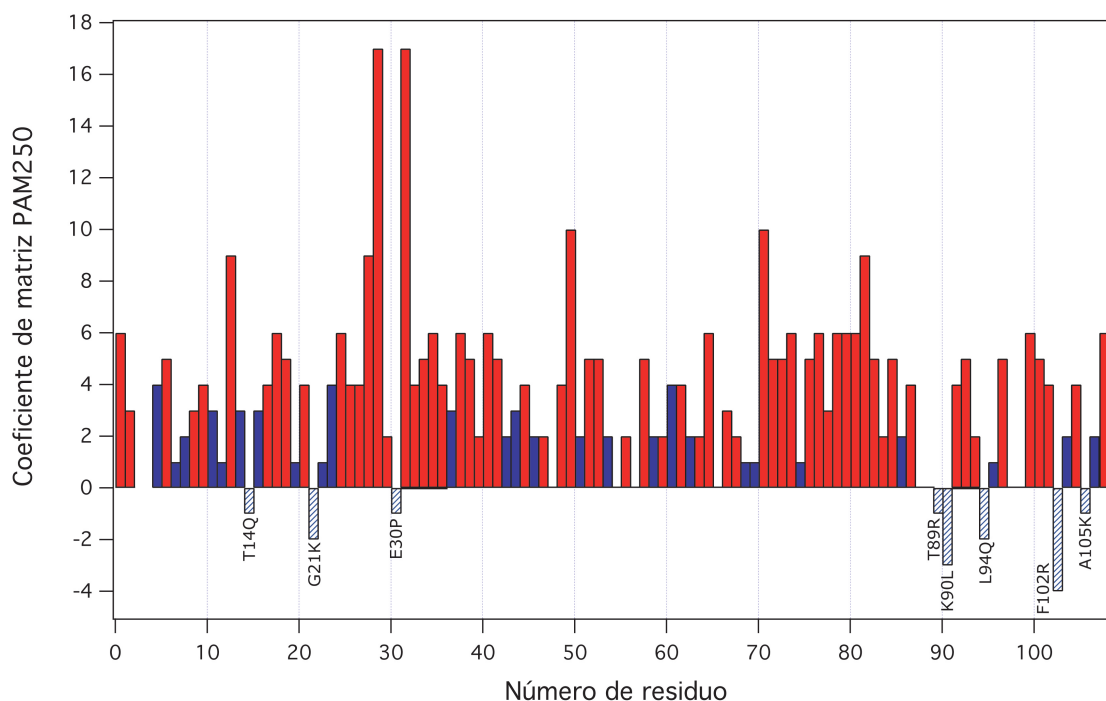


Figura 6.6. Valores de la matriz PAM250. Se han asignado los coeficientes asociados a los cambios observados en el alineamiento de secuencias de tiorredoxina LPBCA y de *E. coli* (Fig. 6.1.). En rojo se representan las posiciones donde no hay mutación. En azul se representan los distintos valores para cada mutación. Los valores positivos corresponden a mutaciones conservativas (azul sólido) y los valores negativos corresponden a mutaciones no conservativas (azul a rayas).

Al cuantificar los cambios en la secuencia mediante la matriz PAM250, observamos que hay 8 mutaciones no conservativas (Fig. 6.6.). Para tratar de identificar cuál de ellas podría tener mayor repercusión en la estabilidad, nos hemos fijado en las que tienen mayor coeficiente, estas son K90L y F102R. Al observar estas dos posiciones en la superposición de las estructuras de estas tiorredoxinas, nos dimos cuenta de que hay un cambio brusco en la accesibilidad de los residuos K90 (trx *E. coli*) y 90L (trx LPBCA) (Fig. 6.7.), algo que luego confirmaron los cálculos de ASA para estos residuos. En el primer caso, la leucina aparece completamente enterrada en la estructura de la tiorredoxina LPBCA (0% de ASA), mientras que la lisina está parcialmente expuesta al disolvente (42% de ASA) (Fig. 6.7). Estos valores de ASA son previsible, ya que se ha sustituido un residuo hidrofóbico (que prefiere estar enterrado en la molécula) por un residuo cargado (que suele situarse en la superficie o estableciendo una interacción electrostática). Por otro lado, la posición 90 está situado más cerca del núcleo de la estructura, mientras que el residuo 102 se sitúa en el extremo C-terminal de la proteína. Por estos motivos, hemos seleccionado la posición 90 para un análisis experimental, ya que posiblemente el impacto en la estabilidad de la proteína podría ser mayor.

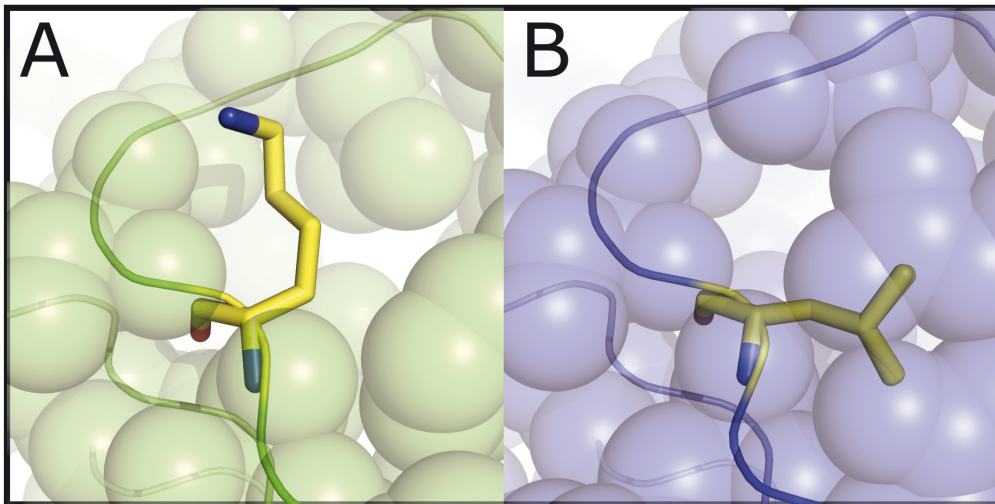


Figura 6.7. A: Posición 90 en la estructura de *trx* de *E. coli* (PDB: 2TRX) (verde). B: Posición 90 en la estructura de LPBCA (PDB: 2YJ7) (azul). Se muestra en detalle los residuos de la posición 90 (amarillo). Se puede observar como la leucina (B) está completamente enterrada en la molécula (0% ASA) y la lisina está parcialmente expuesta al disolvente (A) (42% ASA).

Para comprobar experimentalmente el efecto de la mutación K90L en la estabilidad térmica de tiorredoxina, hemos obtenido los mutantes de *trx* *E. coli* K90L, Variante 31 K90L y *trx* LPBCA L90K por mutagénesis dirigida (apartado 2.2.1 de Materiales y Métodos). Una vez expresados y purificados estos mutantes, hemos llevado a cabo la determinación de su T_m mediante calorimetría diferencial de barrido (Fig. 6.8).

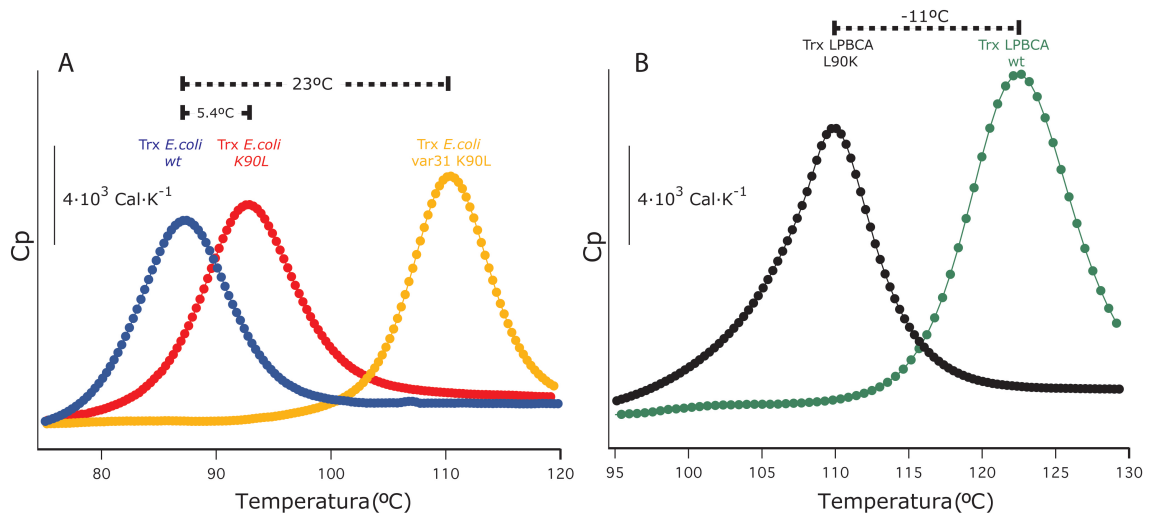


Figura 6.8. Termogramas de DSC de los mutantes de tiorredoxinas. A: Termogramas de DSC de la tiorredoxina de *E. coli* (azul) y los mutantes de *trx* de *E. coli* K90L (rojo) y variante de la biblioteca de tiorredoxina 31 K90L (naranja). B: Termogramas de DSC de la tiorredoxina de LPBCA (verde) y el mutante de *trx* de LPBCA L90K (negro). Los experimentos se realizaron a una velocidad de barrido de 150 K/hora y a diversas concentraciones de proteína (0.7-1.0 mg/ml). Los termogramas de las diferentes proteínas no están normalizadas por la concentración, de ahí la diferencia en la altura de los picos.

Tabla 6.4. Mutantes generados mutando la posición 90 y su respectiva T_m experimental.

Mutante	T_m (°C)
Trx <i>E. coli</i> K90L	92.7
Trx Variante 31 K90L	110.3
Trx LPBCA L90K	109.9

Sorprendentemente, hemos aumentado la T_m de la tiorredoxina de *E. coli* 5.4°C únicamente realizando esta mutación (Tabla 6.4), mientras que la mutación L90K en trx de LPBCA ha disminuido la T_m 11°C (Fig. 6.8). La Variante 31 de la biblioteca con la mutación K90L ha incrementado 5°C su T_m (Fig. 6.8). La posición de estos residuos en la estructura de trx de *E. coli* se puede observar en la figura 6.9.

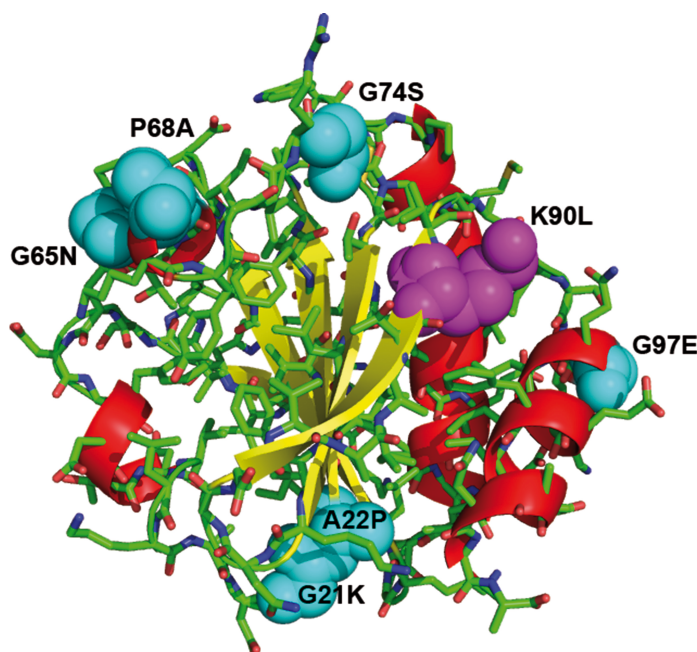


Figura 6.9. Estructura tridimensional de la tiorredoxina de *E. coli* (PDB: 2TRX). En la estructura se han representado los elementos de estructura secundaria asignados con el programa DSSP. Se destacan los residuos de la tiorredoxina de *E. coli* que han sido mutados en la variante más estable de la biblioteca P/G (esferas-cian) y la posición 90 (esferas-magenta).

En vista de los resultados obtenidos, hemos comprobado como la mutación no conservativa K90L tiene un fuerte impacto sobre la estabilidad de la tiorredoxina. Esta conclusión es muy relevante desde el punto de vista de la ingeniería de proteínas, pues podríamos identificar y revertir estas mutaciones para aumentar la estabilidad de las proteínas de interés. Por supuesto, para comprobar esta hipótesis,

es necesario ampliar este estudio experimental a todas las tiorredoxinas ancestrales.

Hasta el momento, hemos conseguido explicar 23°C de la diferencia de estabilidad con tan 7 mutaciones sobre la secuencia de la tiorredoxina de *E. coli* (G21K, A22P, G65N, P68A, G74S, K90L, G97E). Como ya hemos mencionado anteriormente, el cambio K90L es muy poco frecuente en la evolución, esto nos lleva a pensar que este tipo de mutaciones podría conllevar a su vez cambios en residuos próximos en la estructura (con el fin de acomodar al nuevo residuo). Así mismo, estos nuevos cambios podrían afectar a la estabilidad de la proteína. Por esta razón, para explicar la diferencia de estabilidad restante (10°C) entre la tiorredoxina de *E. coli* y la de LPBCA, hemos identificado los residuos que están próximos a la posición 90 en ambas estructuras. Hemos empleado el programa Esfera2 (apartado 2.2.9.7. de Materiales y Métodos) para identificar los residuos que estén en contacto (a una distancia de corte de 8 Å) con los residuos de la posición 90 en las estructuras de tiorredoxinas de *E. coli* y LPBCA. A continuación, hemos examinado cuáles de estos residuos son diferentes en el alineamiento estructural (apartado 2.2.8.2. de Materiales y Métodos) de ambas estructuras. Siguiendo este procedimiento, hemos seleccionado un total de 8 posiciones susceptibles de ser mutadas (A87V, A88D, T89R, L94Q, Q98A, F102R, L103I y N106H). Para estudiar experimentalmente el efecto que pueden tener estas mutaciones sobre la estabilidad hemos construido una biblioteca combinatorial (apartado 2.2.4. de Materiales y Métodos) utilizando como *background* la variante de tiorredoxina 31(K90L). Para construir esta biblioteca, hemos sustituido el aminoácido presente en una posición determinada en tiorredoxina de *E. coli* por el residuo que presenta la tiorredoxina LPBCA en dicha posición. La biblioteca, al igual que la anterior, contendrá 256 posibles variantes con todas las posibles combinaciones de mutación.

Para caracterizar la biblioteca en esta ocasión, hemos seleccionado aleatoriamente 40 colonias de las placas de LB-Agar para llevar a cabo el test de expresión. De las 40 colonias, 35 expresaban la tiorredoxina. Una vez determinada la secuencia de estas 35 variantes, 34 de ellas tenían la secuencia de aminoácidos completa (Tabla 6.5.). A continuación, hemos purificamos las variantes de tiorredoxina (apartado 2.2.2.1 de Materiales y Métodos) y hemos determinado su estabilidad térmica mediante calorimetría diferencial de barrido (Fig. 6.10.).

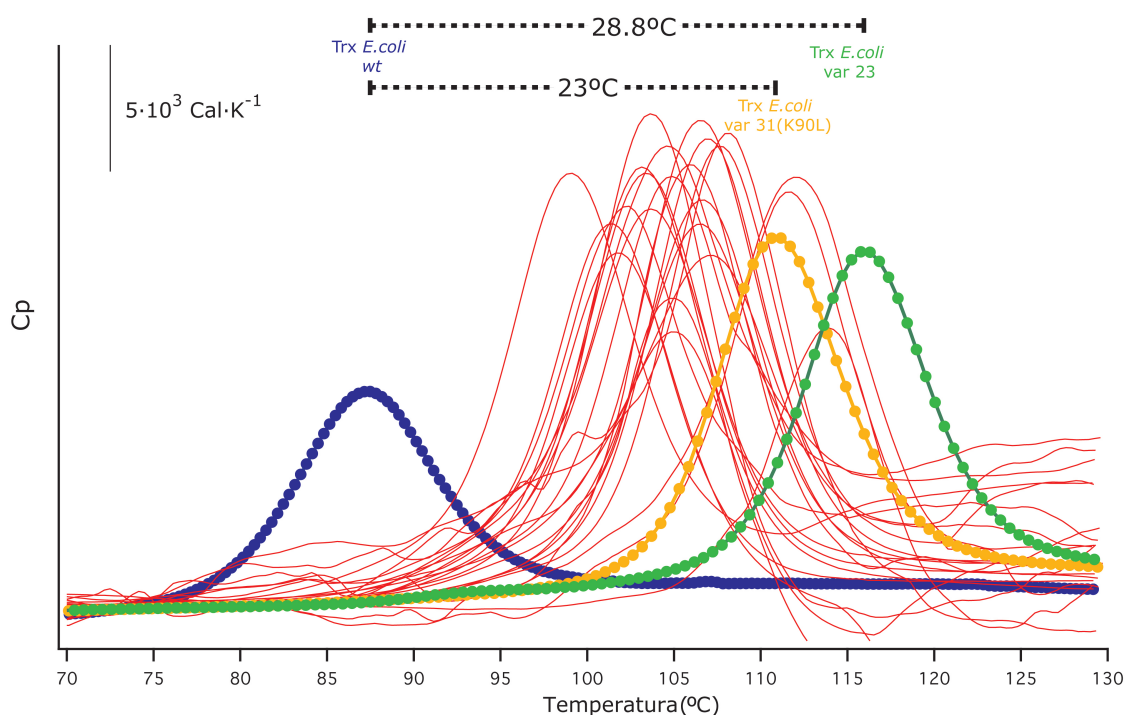


Figura 6.10. Termogramas de DSC de algunas variantes de la biblioteca combinatorial de la variante de *trx 31(K90L)*. En azul se representa el termograma correspondiente a la tiorredoxina de *E. coli*. En naranja se representa el termograma correspondiente a la variante 31 (K90L). En rojo se representan los termogramas de las variantes de la biblioteca que hemos analizado experimentalmente y en verde se representa la variante de la biblioteca más estable. Los experimentos se realizaron a una velocidad de barrido de 150 K/hora y a diversas concentraciones de proteína (0.3-0.7 mg/ml). Los termogramas de las diferentes proteínas no están normalizados por la concentración, de ahí la diferencia en la altura de los picos.

En esta ocasión, la mayoría de las variantes analizadas presentan una T_m menor que la variante de tiorredoxina 31 (K90L); no obstante, hay algunas variantes con mayor T_m . La variante de tiorredoxina 23, con 4 nuevas mutaciones (A87V, T89R, Q98A, N106H) muestra la mayor T_m (116.1°C). Con un total de 11 mutaciones, hemos conseguido un aumento de estabilidad de la tiorredoxina silvestre de *E. coli* de ~29°C.

Tabla 6.5: Variantes de tiorredoxina examinadas experimentalmente y su temperatura de desnaturalización (T_m). Para cada mutante, se muestra si la posición ha sido mutada en la secuencia (gris) o no (blanco). También se muestra la T_m de la variante de tiorredoxina 31 (K90L), que hemos empleado como *background* de la biblioteca. Resaltado en amarillo está la variante de tiorredoxina con la T_m experimental más elevada.

MUTANTE	A87V	A88D	T89R	L94Q	Q98A	F102R	L103I	N106H	T_m (°C)
<i>Var31K90L</i>									110.3
1									106.4
2									99.1
3									104.8

ESTUDIO EXPERIMENTAL DE ESTABILIDAD TÉRMICA DE LA
TIORREDOXINA ANCESTRAL LPBCA

5									106.5
6									107.2
8									111.7
9									101.4
10									103.5
11									106.6
12									104.9
13									104.6
14									102.3
15									103.2
16									110.5
18									103.7
19									112.2
20									105.1
21									106.0
22									108.7
23									116.1
25									101.4
26									104.8
27									106.7
28									112.7
29									107.8
30									108.1
31									99.84
32									106.4
34									111.6
35									113.9
36									103.7
37									103.2
38									101.9
39									104.2

Tras la caracterización de una pequeña población de variantes de la biblioteca, el objetivo es reconstruir el espacio mutacional completo de la misma, es decir, predecir la T_m de las 256 variantes que componen la biblioteca. Para ello, hemos seguido el mismo procedimiento que con la biblioteca anterior. Con las variantes de la biblioteca analizadas experimentalmente, hemos formado un juego de datos, a partir del cual, hemos generado 20 réplicas con el método de *bootstrapping*. A cada réplica, hemos ajustado el modelo de la ecuación 2.18 de Materiales y Métodos. Como en el caso anterior, hemos empleado el método PLS-R para predecir la T_m de todas las variantes de la biblioteca, utilizando el número óptimo de variables latentes calculadas por el programa (entre 2 y 8, según la réplica empleada). Los datos de la predicción de la T_m de todas las variantes de tiorredoxina se encuentran recogidos en la figura 6.11.

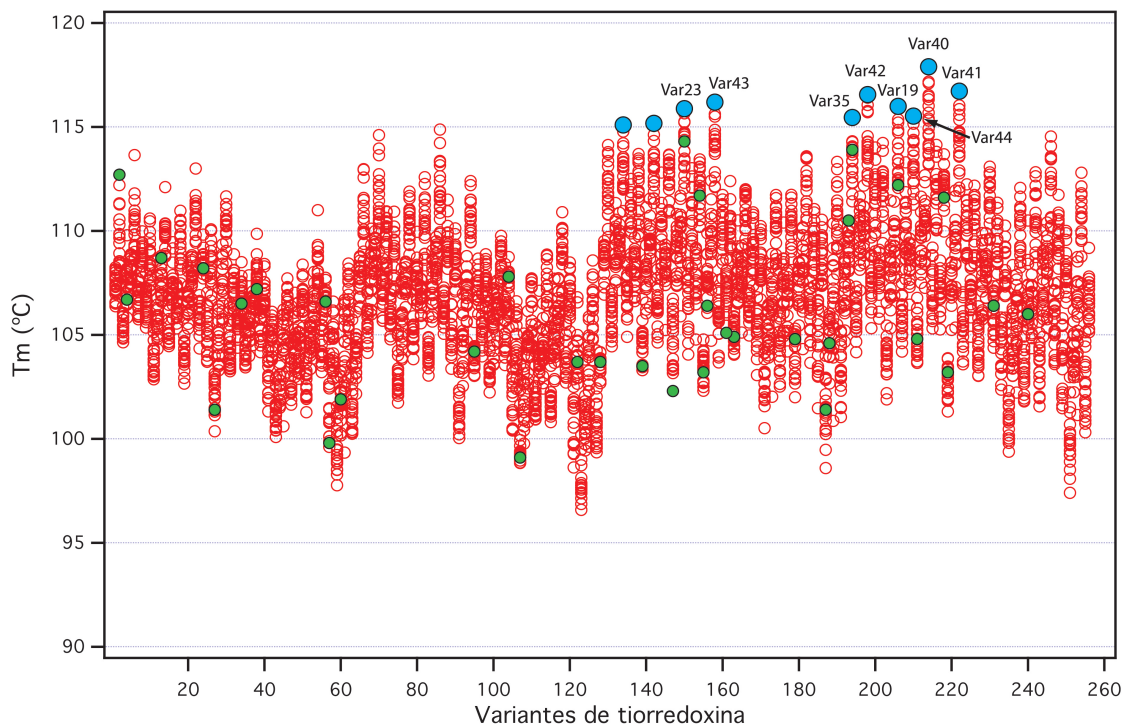


Figura 6.11. Gráfica de reconstrucción del espacio mutacional completo de la biblioteca de residuos próximos a la posición 90. En verde se representa el valor de T_m experimental para las 24 variantes de tioredoxina analizadas experimentalmente. En rojo se representan los valores predichos de T_m para cada variante de tioredoxina. Los círculos azules rellenos corresponden a las 10 variantes de tioredoxina con mayor T_m predicha, aunque solo hemos seleccionado 8 para su caracterización experimental.

Una vez obtenida la predicción, hemos seleccionado los 8 mutantes con mayor T_m predicha para su caracterización experimental (Fig. 6.11.). Estos mutantes los hemos generado en nuestro laboratorio por mutagénesis dirigida a partir de las variantes de tioredoxina analizadas previamente (apartado 2.2.1. de Materiales y Métodos). Además, hemos generado la variante con las 8 mutaciones. Una vez comprobada que la secuencia de las variantes es correcta (Tabla 6.6.), hemos purificado las diferentes tioredoxinas y hemos determinado la T_m mediante DSC (Fig. 6.12.).

ESTUDIO EXPERIMENTAL DE ESTABILIDAD TÉRMICA DE LA
TIORREDOXINA ANCESTRAL LPBCA

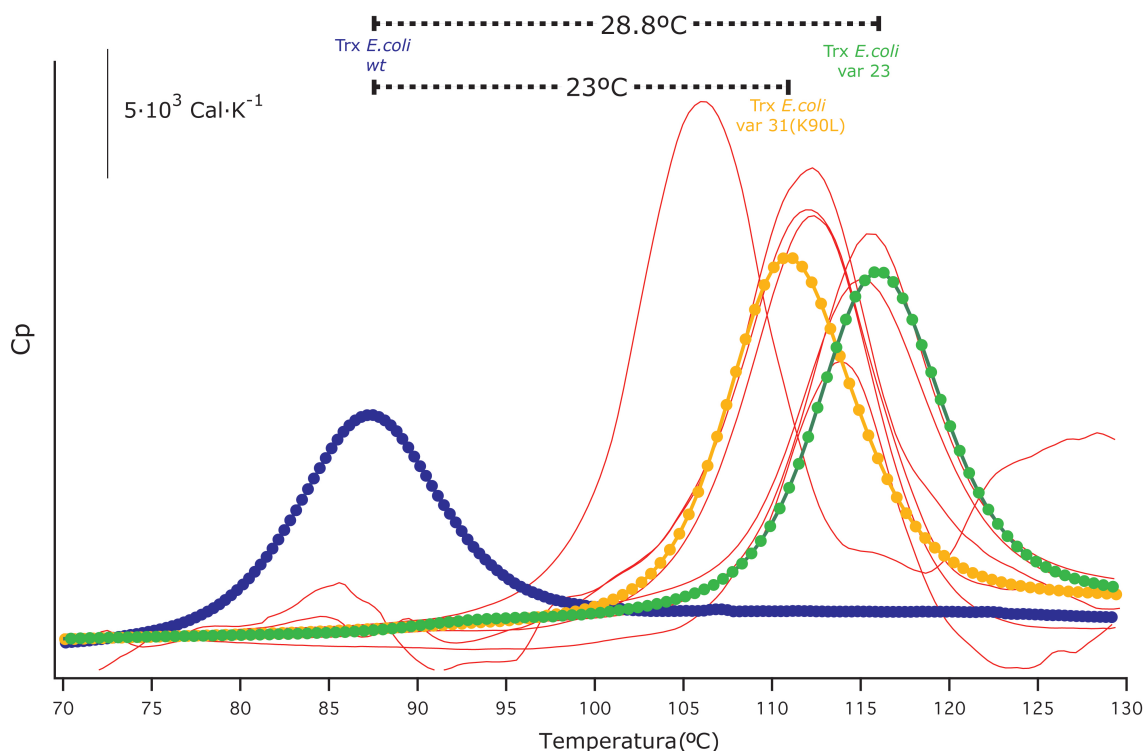


Figura 6.12. Termogramas de DSC de las variantes con mayor T_m predicha de la biblioteca combinatorial de variantes de tiorredoxina 31(K90L). En azul se representa el termograma correspondiente a la tiorredoxina de *E. coli*. En naranja se representa el termograma correspondiente a la variante 31 (K90L). En rojo se representan los termogramas de las variantes de tiorredoxina que hemos analizado experimentalmente. En verde se representa la variante de la biblioteca más estable. Los experimentos se realizaron a una velocidad de barrido de 150 K/hora y a diversas concentraciones de proteína (0.3-0.7 mg/ml). Los termogramas de las diferentes proteínas no están normalizadas por la concentración, de ahí la diferencia en la altura de los picos.

Tabla 6.6. Variantes de tiorredoxina examinadas experimentalmente y su temperatura de desnaturalización (T_m). Para cada mutante, se muestra si la posición ha sido mutada en la secuencia (gris) o no (blanco). Resaltado en amarillo está la variante de tiorredoxina con la T_m experimental más elevada.

Variante	A87V	A88D	T89R	L94Q	Q98A	F102R	L103I	N106H	T_m predicha (°C)	T_m exp. (°C)
Var40									117.9	115.6
Var41									116.7	112.3
Var42									116.5	115.0
Var43									116.2	112.3
19									116.0	112.2
23									115.9	116.1
Var44									115.5	114.2
35									115.4	113.9
Var45									109.1	106.1

En este caso, entre las variantes más estables de la biblioteca hay varios que ya hemos analizado en el primer muestreo de la biblioteca (mutantes 23, 35 y 19) y aparecen 4 nuevas variantes con una T_m predicha superior a la variante más estable medida experimentalmente. Sin embargo, al comprobar la T_m experimentalmente la variante más estable continúa siendo la variante 23 con 11 mutaciones (Fig. 6.12 y Tabla 6.6.).

Hemos añadido los datos de las nuevas variantes más estables al juego de datos anterior para construir un nuevo juego de datos y hemos repetido el proceso de *bootstrapping* y regresión PLS con el objetivo de encontrar alguna variante más estable. Los datos observados tras la segunda predicción mostraron que no había ninguna nueva variante más estable, aunque sí había una predicción que se ajustaba mejor a los valores de T_m de las variantes de tiorredoxina medidos experimentalmente.

Los resultados del estudio experimental de la diferencia de estabilidad de la tiorredoxina de *E. coli* y la LPBCA muestran que hemos conseguido aumentar la estabilidad de la tiorredoxina de *E. coli* $\sim 29^\circ\text{C}$ realizando 11 mutaciones (G21K, A22P, G65N, P68A, G74S, G97E, K90L, A87V, T89R, Q98A, N106H) (Fig. 6.13).

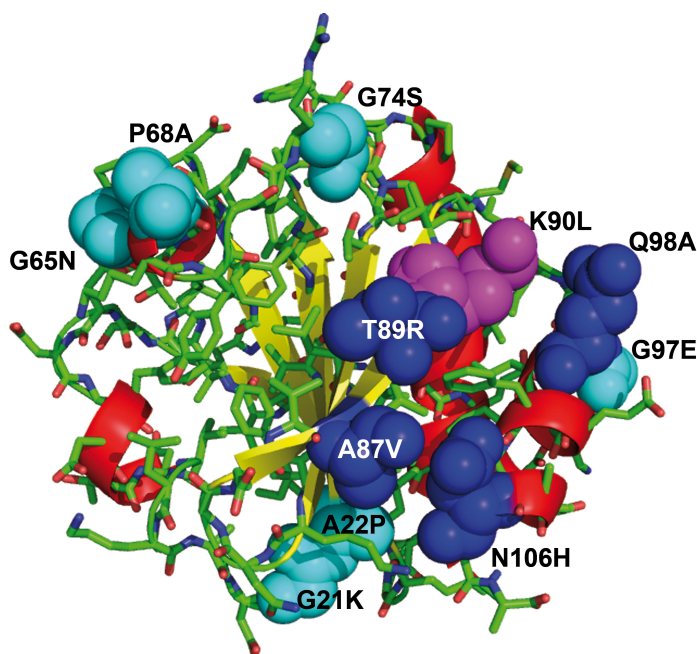


Figura 6.13. Estructura tridimensional de la tiorredoxina de *E. coli*. En la estructura se han representado los elementos de estructura secundaria asignados con el programa DSSP. Se destacan los residuos de la Tiorredoxina de *E. coli* que han sido mutados en la variante más estable de la biblioteca de prolina y glicinas (esferas-cian), la posición 90 (esferas-azul) y las posiciones mutadas en la variante más estable de la última biblioteca (esferas-magenta).

En resumen, para tratar de explicar la diferencia de estabilidad de 33°C entre la tiorredoxina precámbrica LPBCA y la tiorredoxina actual de *Escherichia coli*, hemos abordado el estudio mutacional experimental del número de prolina y glicinas en ambas tiorredoxinas, así como el estudio de la mutación no conservativa K90L y los residuos próximos a esta posición. Utilizando un enfoque de bibliotecas combinatoriales de mutantes de tiorredoxinas, junto con el método de regresión de mínimos cuadrados parciales, hemos logrado obtener una variante de tiorredoxina con una temperatura de desnaturalización de 116.1°C. Esto implica que hemos conseguido explicar casi 29°C de la diferencia de estabilidad de la tiorredoxina precámbrica LPBCA y la tiorredoxina actual de *E. coli*, así como identificar los aminoácidos implicados en la pérdida de la estabilidad de la tiorredoxina de *E. coli* a lo largo de la evolución.

Es importante destacar que la tiorredoxina de LPBCA tiene 46 residuos diferentes a la tiorredoxina de *E. coli*, y solo realizando 11 mutaciones hemos podido explicar, en gran medida, la diferencia de estabilidad entre ambas proteínas. Solo nos quedarían por explicar 5°C de diferencia, que podrían ser debidos a una o varias contribuciones que hemos analizado en el capítulo 5 de la presente tesis doctoral. En este contexto, aunque no hayamos percibido diferencias significativas en ninguno de los factores analizados en el capítulo 5, ligeras diferencias (por debajo de límite de precisión de los cálculos) podrían tener un efecto sobre la estabilidad, dado que la estabilidad de las proteínas es marginal (del orden de decenas de KJ/mol). No obstante, sí hemos apreciado diferencias al comparar la propensión de alfa hélices, el número de puentes salinos y el volumen de las cavidades en las tiorredoxinas precámbricas y actuales de *E.coli* y humana, existiendo cierta correlación con la estabilidad de las proteínas. Por tanto, el aumento de la estabilidad podría ser debido también a estos factores. Por ejemplo, el caso más claro es la mutación P68A, donde la prolina tiene una bajísima propensión de situarse en alfa hélices, mientras que la alanina tiene una gran propensión. En cualquier caso, para estudiar este efecto sería necesario disponer de la estructura tridimensional de este mutante y validar los resultados teóricos mediante un análisis experimental.

En general, los resultados obtenidos en este trabajo sugieren que la eliminación de residuos de glicina en tiorredoxina podría tener como consecuencia un aumento de la estabilidad de esta proteína, ya que el mutante más estable tiene 4 glicinas menos. Además, la ocurrencia de la mutación no conservativa K90L y la mutación de

algunos residuos próximos a esta posición, implica una pérdida de estabilidad de la tiorredoxina de *E. coli*. Este resultado podría tener implicaciones generales, ya que el enfoque de caracterización de algunas mutaciones poco conservativas y su posterior análisis experimental podría ser útil para incrementar la estabilidad térmica de las proteínas. Por supuesto, esta conclusión debemos tomarla con precaución, ya que es necesario comprobar experimentalmente que se sigue esta tendencia en el resto de tiorredoxinas precámbricas.

6.3. Bibliografía

- Bershtein, S., M. Segal, R. Bekerman, N. Tokuriki and D. S. Tawfik (2006). "Robustness-Epistasis Link Shapes the Fitness Landscape of a Randomly Drifting Protein." Nature **444**(7121): 929-932.
- Dayhoff, M. O., R. M. Schwartz and B. C. Orcutt (1978). "{a Model of Evolutionary Change in Proteins}." Atlas of protein sequence and structure **5**(suppl 3): 345-351.
- Matthews, B. W., H. Nicholson and W. J. Becktel (1987). "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 **84**(19): 6663-6667.
- Richardson, J. S. (1981). "The Anatomy and Taxonomy of Protein Structure." Advances in protein chemistry **34**: 167-339.
- Tokuriki, N. and D. S. Tawfik (2009). "Chaperonin Overexpression Promotes Genetic Variation and Enzyme Evolution." Nature **459**(7247): 668-673.
- Tokuriki, N. and D. S. Tawfik (2009). "Stability Effects of Mutations and Protein Evolvability." Current Opinion in Structural Biology **19**(5): 596-604.
- Zeldovich, K. B., P. Chen and E. I. Shakhnovich (2007). "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 **104**(41): 16152-16157.

7. CONCLUSIONES

La presente tesis doctoral expone varias aportaciones al estudio de proteínas ancestrales resucitadas. Las conclusiones más relevantes obtenidas en este trabajo son las siguientes:

1. Las tiorredoxinas ancestrales emplean mecanismos de reducción similares a los de las tiorredoxinas actuales. Esto implica que la divergencia catalítica mostrada por las tiorredoxinas actuales se estableció muy temprano en la evolución, hace unos 4 mil millones de años antes del presente.
2. Las tiorredoxinas ancestrales resucitadas se adaptaron a lo largo de la evolución a los cambios del entorno, ya que poseen una elevada estabilidad térmica y las más antiguas presentan actividad a pH ácido, condiciones inherentes a los océanos primitivos
3. La estructura de las tiorredoxinas se estableció muy temprano en la evolución, hace unos 4 mil millones de años antes del presente, y ha sufrido pocas variaciones desde entonces, por lo que puede considerarse como un fósil molecular.
4. Las pequeñas diferencias apreciadas en las estructuras de las tiorredoxinas precámbricas están limitadas a los ancestros de las tiorredoxinas eucariotas y proporcionan una interpretación evolutiva de las diferencias estructurales observadas entre las tiorredoxinas actuales de *Escherichia coli* y humana.
5. Las 11 mutaciones responsables de la mayor parte de la hiperestabilidad de la tiorredoxina de LPBCA están relacionadas con el número de prolinas y glicinas, la mutación no conservativa K90L y algunos aminoácidos próximos en la estructura a esta posición.

Los resultados del estudio de tiorredoxinas resucitadas obtenidos en esta tesis doctoral podrían tener las siguientes implicaciones generales:

- I. La resurrección de proteínas ancestrales puede convertirse en una de las principales fuentes de la próxima generación de proteínas con aplicaciones biotecnológicas y biomédicas.
- II. La resurrección de proteínas ancestrales puede convertirse en una herramienta muy útil para estudiar la evolución de las estructuras de proteínas.
- III. El estudio de algunas mutaciones no conservativas (y los residuos próximos en la estructura) podría utilizarse en ingeniería de proteínas para incrementar la estabilidad térmica de las proteínas.

8. RESUMEN

En la presente memoria para la obtención del grado de Doctor por la Universidad de Granada abordamos la resurrección y caracterización energética, biofísica y estructural de tiorredoxinas ancestrales pertenecientes a organismos extintos del periodo Precámbrico (4.5 a 0.5 mil millones de años antes del presente, aproximadamente). Este estudio tiene implicaciones evolutivas, ya que contribuye a ampliar la información sobre el entorno físico que rodeaba a los organismos extintos, así como implicaciones en ingeniería de proteínas, ya que las proteínas ancestrales estudiadas presentan unas propiedades alteradas, inusuales o extremas en términos de estabilidad o actividad catalítica.

La presente memoria comienza con una introducción general en la que repasamos brevemente la historia de la resurrección de proteínas ancestrales y discutimos las implicaciones evolutivas y en ingeniería de proteínas que tiene este enfoque. A continuación detallamos las diferentes etapas en las que hemos llevado a cabo el trabajo de esta tesis doctoral y enumeramos los objetivos concretos que nos hemos propuesto. Seguidamente, explicamos detalladamente cada uno de los procedimientos experimentales utilizados para la consecución de los resultados expuestos en la presente memoria. Cabe destacar que, para la caracterización de las tiorredoxinas ancestrales ha sido necesario utilizar gran diversidad de técnicas experimentales (ensayos bioquímicos, cristalografía de rayos-X, calorimetría diferencial de barrido, procedimientos de biología molecular, etc.)

Tras los detalles técnicos, mostramos los resultados de la primera etapa del trabajo, que ha consistido en la expresión, purificación y caracterización biofísica de siete tiorredoxinas ancestrales pertenecientes a organismos del período Precámbrico. Las tiorredoxinas resucitadas son extremadamente estables, presentando temperaturas de desnaturalización hasta unos 30-35 grados superiores a las de las tiorredoxinas actuales. Además, las tiorredoxinas resucitadas más antiguas (~ 4 mil millones de años antes del presente) muestran una actividad, tanto a pH 7.0 como a pH 5.0, mucho más elevada que las tiorredoxinas actuales. A partir de los resultados obtenidos, podemos concluir que las tiorredoxinas ancestrales emplean mecanismos de reducción similares a los de las tiorredoxinas actuales. Aunque han mantenido su mecanismo de reducción inalterado, sí han sido capaces de adaptarse a lo largo de la evolución a los cambios del entorno en términos de temperatura y acidez.

Para tratar de incrementar el alcance y las implicaciones de la resurrección de proteínas ancestrales, hemos llevado a cabo la resolución de las estructuras tridimensionales de las siete tiorredoxinas resucitadas mediante cristalografía de rayos-X. Además, hemos realizado un detallado análisis comparativo de las estructuras de las tiorredoxinas ancestrales con las tiorredoxinas actuales de *Escherichia coli* y humana. Los resultados de este estudio muestran que la estructura de las tiorredoxinas han experimentado pocos cambios a lo largo de 4 mil millones de años, por lo que puede considerarse como un fósil molecular. Los pequeños cambios observados en las estructuras, relacionados principalmente con los ancestros de las tiorredoxinas eucariotas, proporcionan una interpretación evolutiva a las diferencias estructurales de las tiorredoxinas de *E. coli* y humana. Estos resultados sugieren que la resurrección ancestral de proteínas puede constituir una herramienta muy útil para estudiar la evolución de la estructura de las proteínas.

Dado que disponemos de las estructuras de tiorredoxinas precámbricas estamos en un posición única para estudiar los mecanismos moleculares responsables de la alta estabilidad observada en las tiorredoxinas resucitadas. Con este objetivo, hemos realizado un estudio computacional comparativo de las estructuras de las tiorredoxinas ancestrales y las tiorredoxinas actuales de *E. coli* y humana, dirigido a evaluar las diferentes contribuciones energético/estructurales (interacciones hidrofóbicas, enlaces de hidrógeno, entropía conformacional, etc.) a la estabilidad de estas proteínas. Los resultados de este estudio muestran contribuciones significativas en lo que respecta a la propensión de los residuos por alfa hélices y a parámetros relacionados con la entropía conformacional del estado desplegado (número de prolina y glicinas).

Finalmente, hemos llevado a cabo un análisis mutacional de la hiperestabilidad de la tiorredoxina LPBCA (121°C de T_m), en comparación con la tiorredoxina actual de *E. coli* (87.3°C de T_m). En este estudio nos hemos centrado en el número de prolina y de glicinas, así como en la mutación no conservativa K90L y los residuos próximos en la estructura a dicha posición. Para llevar a cabo este análisis experimental, hemos utilizado un enfoque de bibliotecas combinatoriales de variantes de tiorredoxinas, en combinación con la regresión de mínimos cuadrados parciales. Los resultados del análisis experimental muestran que con tan solo 11 mutaciones pueden explicarse alrededor de 29°C de la diferencia de estabilidad (de 33°C) entre las tiorredoxinas analizadas. Estos resultados podrían tener

implicaciones generales, pues sugieren que el estudio de algunas mutaciones no conservativas (y los residuos próximos en estructura) podría ser utilizado en ingeniería de proteínas para modificar la estabilidad de las proteínas.

9. ABREVIATURAS Y SÍMBOLOS

A	Actividad
Abs	Absorbancia
ADN	Ácido desoxirribonucleico
AECA	Último ancestro común de arqueas y eucariotas
AFM	Microscopía de Fuerza Atómica
Amp	Ampicilina
ASA	Área superficial accesible al disolvente
c	Concentración
CD	Dicroísmo Circular
C_p^{ex}	Capacidad calorífica de exceso
C_p^D	Capacidad calorífica del estado desnaturalizado
C_p^N	Capacidad calorífica del estado nativo
D	Estado desplegado
dNTPs	Mezcla de desoxinucleótidos trifosfato
D.O.	Densidad Óptica
DSC	Calorimetría Diferencial de Barrido
DTNB	Ácido 5,5'-ditiobis-2-nitrobenzoico
DTT	Ditiotreitol
EDTA	Ácido etilendiaminotetraacético
E. coli	<i>Escherichia coli</i>
ESRF	European Synchrotron Radiation Facility
F	Fuerza
Ga	Giga-año
IPTG	Isopropil- β -D-1-tiogalactopiranosido
K	Constante de equilibrio
Kan	Kanamicina
LACA	Último ancestro común de arquea
LAFCA	Último ancestro común de los animales y hongos
LB	(Medio de cultivo) Luria Bertani
LBCA	Último ancestro común de las bacterias
LECA	Último ancestro común de eucariotas
LPBCA	Último ancestro común grupos deinococcus y thermus
LGPCA	Último ancestro común de las γ -proteobacterias
MPD	2-Metil-2,4-pentanodiol
Mw	Masa Molecular
n	Orden del proceso
N	Estado nativo
NIPALS	Nonlinear Iterative Partial Least Squares
Pb	Pares de bases
PCA	Análisis de Componentes Principales
PCR	Reacción en Cadena de la Polimerasa
PEG	Polietilenglicol
PLS-R	Regresión por Mínimos Cuadrados Parciales
Pm	Peso molecular
R	Constante de los gases

RMN	Resonancia Magnética Nuclear
R.M.S.D.	Desviación cuadrática media
r.p.m.	Revoluciones por minuto
SDS	Dodecilsulfato sódico
SDS-PAGE	Electroforesis en gel de poliacrilamida con SDS
SOC	<i>Super optimal broth</i> con represión catabólica
T	Temperatura
TE	Tris-EDTA
T _m	Temperatura de desnaturalización por calor, temperatura a la cual la mitad de la hebra de ADN se encuentra como doble hélice
Trx	Tiorredoxina
U	Estado desnaturalizado
U.A.	Unidades de Absorbancia
UV	Ultravioleta
β	Barrera de energía libre de plegamiento
ΔC _p	Cambio de capacidad calorífica de desnaturalización (también Δ _N ^D C _p)
ΔG	Cambio de la energía de Gibbs de desnaturalización (también Δ _N ^D G)
ΔH	Cambio de entalpía de desnaturalización (también Δ _N ^D H)
ΔH ^{VH}	Cambio de entalpía de van't Hoff
ΔS	Cambio de entropía de desnaturalización (también Δ _N ^D S)

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	Isoleucina
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

10. APÉNDICE

Protein Folding Drives Disulfide Formation

Pallav Kosuri,^{1,2,*} Jorge Alegre-Cebollada,² Jason Feng,² Anna Kaplan,² Alvaro Inglés-Prieto,⁵ Carmen L. Badilla,² Brent R. Stockwell,^{2,3,4} Jose M. Sanchez-Ruiz,⁵ Arne Holmgren,⁶ and Julio M. Fernández^{2,*}

¹Graduate Program in Biochemistry & Molecular Biophysics

²Department of Biological Sciences

³Department of Chemistry

⁴Howard Hughes Medical Institute

Columbia University, New York, NY 10027, USA

⁵Facultad de Ciencias, Departamento de Química-Física, Universidad de Granada, 18071 Granada, Spain

⁶Medical Nobel Institute for Biochemistry, Department of Medical Biochemistry and Biophysics, Karolinska Institutet, 17177 Stockholm, Sweden

*Correspondence: pallav@caa.columbia.edu (P.K.), jfernandez@columbia.edu (J.M.F.)

<http://dx.doi.org/10.1016/j.cell.2012.09.036>

SUMMARY

PDI catalyzes the oxidative folding of disulfide-containing proteins. However, the sequence of reactions leading to a natively folded and oxidized protein remains unknown. Here we demonstrate a technique that enables independent measurements of disulfide formation and protein folding. We find that non-native disulfides are formed early in the folding pathway and can trigger misfolding. In contrast, a PDI domain favors native disulfides by catalyzing oxidation at a late stage of folding. We propose a model for cotranslational oxidative folding wherein PDI acts as a placeholder that is relieved by the pairing of cysteines caused by substrate folding. This general mechanism can explain how PDI catalyzes oxidative folding in a variety of structurally unrelated substrates.

INTRODUCTION

Protein folding is a sensitive reaction that can be easily affected by a multitude of factors such as mutations and changes in the physical or chemical environment. Failure to correctly fold is the basis of numerous disorders of central importance to modern medicine (Dobson, 2003). In particular, the third of human proteins that traverse the secretory pathway and that possess disulfide bonds pose unresolved challenges to our understanding of protein folding and disease (Gething and Sambrook, 1992; Ron and Walter, 2007; Schröder and Kaufman, 2005). Protein disulfide isomerase (PDI) introduces disulfide bonds into folding proteins and is the main catalyst of oxidative folding in humans (Wilkinson and Gilbert, 2004). Recent studies have revealed a link between disulfide chemistry and the pathogenesis of misfolding diseases and specifically implicated PDI as a novel target for treatment of several neurodegenerative disorders including Alzheimer's disease (Hoffstrom et al., 2010; Uehara et al., 2006). These studies stress the importance of understanding how PDI catalyzes oxidative folding.

Human PDI catalyzes the formation of disulfides (oxidase activity) as well as the rearrangement of incorrectly formed disulfide bonds (isomerase activity) (Wilkinson and Gilbert, 2004). The enzyme consists of two catalytically active A domains and two redox-inactive B domains. Isolated A domains have been shown to effectively catalyze the introduction of disulfides into protein substrates; meanwhile the full-length protein is generally thought to be required for efficient isomerase activity (Darby and Creighton, 1995b). PDI belongs to a ubiquitous family of enzymes that catalyze thiol-disulfide exchange (Wilkinson and Gilbert, 2004). In addition to PDI, this family includes other oxidoreductases such as thioredoxin (TRX), glutaredoxin, and the bacterial Dsb enzymes (Martin, 1995). All of these enzymes share a characteristic structural fold and a highly conserved Cys-X-X-Cys motif in their active sites (Martin, 1995). Their mechanism of action has been revealed through numerous studies over the past 40 years. In all cases, the reaction mechanism involves the formation of a mixed disulfide between a cysteine in the substrate and the N-terminal cysteine in the active site of the enzyme (Holmgren, 1985; Walker et al., 1996) (Figure S1 available online). The C-terminal cysteine can attack and cleave the mixed disulfide, thereby spontaneously releasing the enzyme (Walker and Gilbert, 1997; Wilkinson and Gilbert, 2004). Whereas spontaneous release is necessary during reduction of substrate disulfide bonds, it is unknown how this activity affects catalysis of oxidative folding.

Secretory proteins are synthesized as linear polypeptides and emerge from the ribosomal channel via the translocon into the endoplasmic reticulum (ER) (Rapoport et al., 1996; Van den Berg et al., 2004; Walter et al., 1984). Emerging sequentially into the ER, the nascent polypeptide encounters PDI, which catalyzes cotranslational oxidative folding (Bulleid and Freedman, 1988; Molinari and Helenius, 1999). This reaction is mediated by the formation of a mixed disulfide bond between the PDI enzyme and a cysteine in the nascent polypeptide (Figure 1A) (Frang and Kaiser, 1999; Gilbert, 1995; Sevier and Kaiser, 2002). The mixed disulfide is then transferred to the folding polypeptide. Given the crucial roles of mixed disulfides in oxidoreductase catalysis, many studies have been focused on these ephemeral intermediates. The molecular structures of mixed disulfide complexes have been reported for several enzymes

(Dong et al., 2009; Paxman et al., 2009; Qin et al., 1995). In addition, mixed disulfide complexes in the process of oxidative folding have been characterized in living cells (Di Jeso et al., 2005; Kadokura and Beckwith, 2009; Kadokura et al., 2004; Molinari and Helenius, 1999). Although these studies have provided us with snapshots of mixed disulfide complexes, their dynamics during oxidative protein folding remain unknown. In order to study the intersection of covalent chemistry and protein folding, a method is needed that can independently measure these two concurrent processes.

Here we demonstrate single-molecule techniques that measure the kinetics of enzyme attachment and release and the kinetics of protein oxidation and folding. We used an atomic force microscope (AFM) to mimic the initial stages of oxidative folding by extending an individual polypeptide to a state wherein the substrate cysteines were spatially separated, and a mixed disulfide complex was formed with the catalytic domain A1 from human PDI (hereafter referred to as PDIA, Figure 1B). We then allowed the polypeptide to refold in the absence of external force and probed the progress of oxidative folding by applying additional force pulses. Our data are consistent with a model in which PDIA acts as a placeholder that allows the substrate to guide the pairing of cysteines into native disulfide bonds. This mechanism is general in its nature and therefore does not require substrate specificity.

RESULTS

Creating a Mixed Disulfide Complex between PDIA and a Single Unfolded Protein

Formation of a PDI-substrate mixed disulfide is thought to enable oxidative folding of newly synthesized proteins. We sought to create such a mixed disulfide complex and, starting from this state, study the progression of oxidative folding. A mixed disulfide complex can be formed through two pathways (Figure S1). In the first pathway, an oxidized PDI enzyme reacts with a free substrate cysteine (Fränd and Kaiser, 1999; Tu et al., 2000). In the second pathway, the mixed disulfide is formed when reduced PDI reacts with a substrate disulfide (Lundström and Holmgren, 1990). Regardless of the pathway, the resulting mixed disulfide complex is identical. Although we could use both pathways (see Figures 6 and S7), the second pathway proved more advantageous in our experiments because we could then directly detect the formation of the mixed disulfide, as described in detail below.

We used in our experiments human PDI A1 (PDIA). Although essentially inactive in disulfide isomerization as compared to the full-length enzyme, PDIA is sufficient for catalysis of oxidative folding (Darby and Creighton, 1995b). The substrate we chose was the 27th Ig domain from human cardiac titin containing a buried disulfide between residues 32 and 75 (hereafter referred to as I27³²⁻⁷⁵), which has been well characterized in the past (Ainavarapu et al., 2007; Wiita et al., 2006, 2007).

Our strategy to establish mixed disulfide complexes is illustrated in Figure 1C, and an experimental recording is displayed in Figure 1D. We used a custom-built AFM to apply a constant calibrated force to the termini of a single I27³²⁻⁷⁵ protein, while measuring its extension (Fernandez and Li, 2004). A force of

130–150 pN enables protein unfolding but does not break any of its covalent bonds (Grandbois et al., 1999; Wiita et al., 2006). The polypeptide chain can thus unravel only up to the disulfide bond, resulting in the 11 nm extension step seen in the recording (Figure 1D). Meanwhile, the 11 nm extension exposes the 32–75 disulfide to the solvent and enables reactions with enzymes present in the surrounding media (Wiita et al., 2006). In Figure 1D, a reduced PDIA enzyme present in solution reacts with the now solvent-exposed disulfide, creating a mixed disulfide. The enzymatic reaction is captured as a 14 nm step that results from mechanical unraveling of the rest of the substrate as soon as the original disulfide is broken. A 14 nm step in our experiments thus indicates the formation of a mixed disulfide complex. Once a mixed disulfide was acquired in a fully extended polypeptide, we could proceed to study its effect on protein folding.

Mixed Disulfide Complexes with PDIA Enable Oxidative Folding

We used a three-part experimental protocol consisting of two mechanically equivalent force pulses separated by a refolding interval [*denature* – *folding* – *probe*] to study the effect of PDIA mixed disulfide complexes on the folding of an (I27³²⁻⁷⁵)₈ polypeptide substrate (Figure 1E). A polyprotein was used because it yields multiple events within a single recording, thereby providing a stronger fingerprint of the reactions. Figure 1F shows how each substrate domain was first completely extended and linked to a PDIA enzyme during an initial *denature* pulse, as described in the previous section. The force was subsequently switched off, and the substrate was allowed to fold for a set time $\Delta t_{\text{folding}}$ (Δt_F). In order to detect the formation of disulfide bonds and protein folding during Δt_F , we once again applied force, thereby halting the folding reaction. This *probe* pulse was identical to the initial *denature* pulse and in the same manner allowed us to detect folded domains and disulfide bonds. A 25 nm step appearing during the *probe* pulse reports that a domain had folded but not formed a disulfide during Δt_F . An 11 nm step, on the other hand, indicates that a domain had folded during Δt_F and also acquired a disulfide. Given that PDIA was present in the surrounding media throughout the experiment, 14 nm steps would be seen in the *probe* pulse if newly formed disulfide bonds were again broken. In summary, a 25 nm step provided a signature of a natively folded domain with no disulfide, whereas an 11 nm step (followed by a 14 nm step) indicated a natively folded domain with a disulfide (Figure 1F).

Figure 1G shows an experimental recording of the [*denature* – *folding* – *probe*] experiment in a solution containing reduced PDIA, and Figure 1H shows histograms of the step sizes detected in several such recordings. During the *denature* pulse, 11 and 14 nm steps were seen, indicating formation of mixed disulfide complexes between PDIA and the substrate. When the force was removed ($\Delta t_F = 5$ s), the substrate rapidly collapsed as seen in the experimental trace (Figure 1G). The 11 and 14 nm steps seen in the *probe* pulse unambiguously showed that disulfides had been formed in the substrate during folding. The reaction path for these domains is readily interpreted. When the force was removed, the substrate collapsed and allowed the free cysteine in each domain to attack the mixed

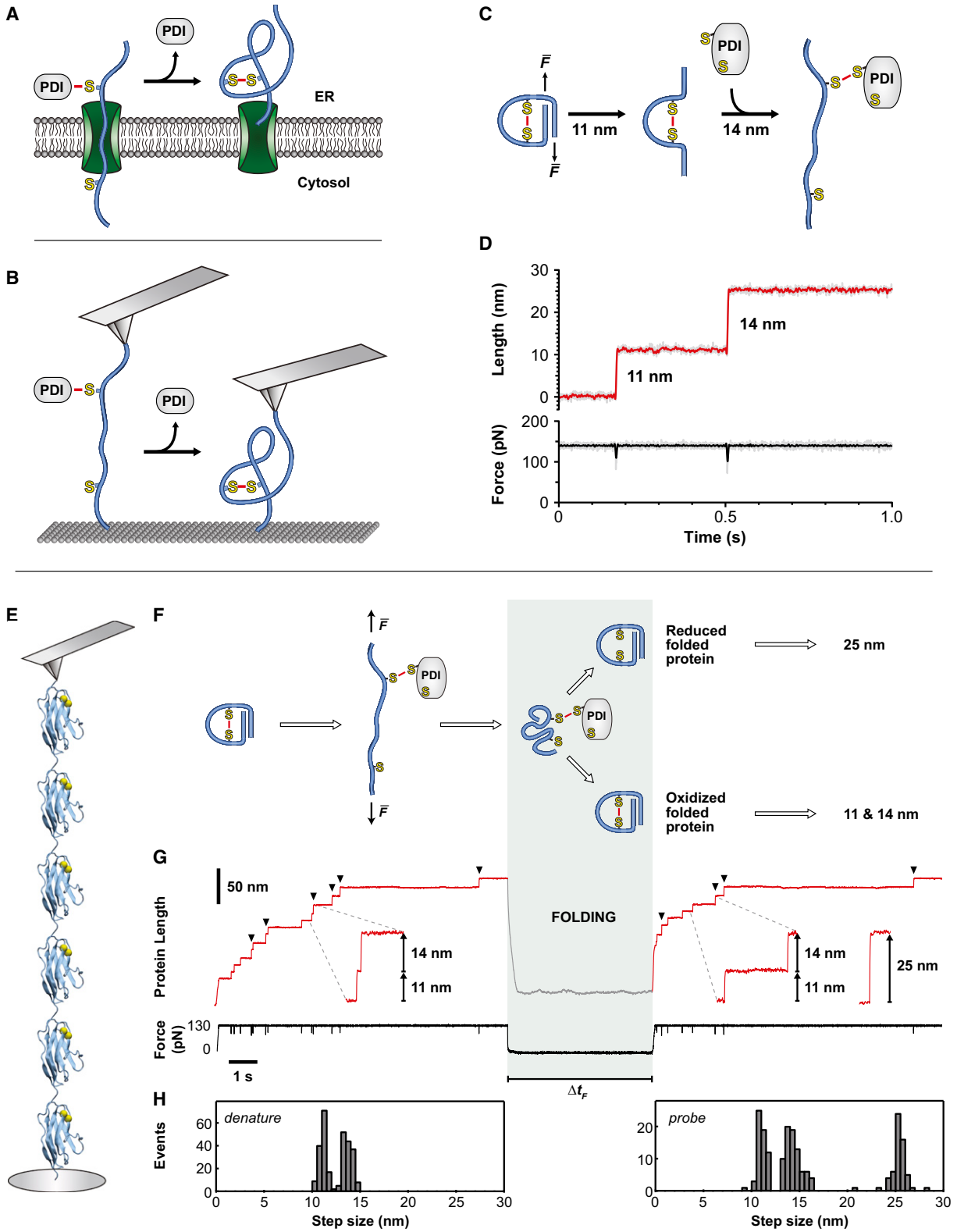


Figure 1. A Single-Molecule Approach to Study Oxidative Folding

(A) As part of the secretory pathway, protein disulfide isomerase (PDI) forms mixed disulfide complexes with nascent polypeptides (blue) undergoing ER translocation. These complexes are thought to enable oxidative folding.

disulfide bond, thus releasing the PDla enzyme and establishing the intramolecular 32–75 disulfide bond. The appearance of 11 nm unfolding steps is consistent with these domains also acquiring their native fold. Additional experiments performed with full-length bovine PDI showed similar results, validating our use of PDla (see Figure S2C for a representative recording).

Detection of Spontaneous Enzyme Release

In addition to 11 and 14 nm steps, 25 nm steps were also seen in the *probe* pulse (Figures 1G, 1H, and S2B), revealing that some substrate domains had folded without forming disulfide bonds. Two scenarios can account for these 25 nm steps: either (1) PDla was still attached to the folded substrate but had failed to introduce the 32–75 disulfide bond, or (2) PDla had spontaneously released before substrate oxidation could be realized. We could thus distinguish between these scenarios by using two oxidoreductase enzymes with vastly different release rates: wild-type (WT) and C35S TRX.

TRX is capable of catalyzing the same reactions as PDI (Carvalho et al., 2006; Debarbieux and Beckwith, 2000; Lundström et al., 1992) but releases spontaneously from the mixed disulfide on a much faster (sub-millisecond) timescale, as inferred from reactions with nonprotein substrates (Chivers and Raines, 1997; Lappi and Ruddock, 2011). Consequently, in a [*denature* – *folding* – *probe*] experiment with WT TRX, all enzymes would have released from the substrate before the end of the *denature* pulse. We would thus expect only 25 nm steps in the *probe* pulse as the substrate folds in the absence of the enzyme. Experiments with TRX confirmed that only 25 nm steps were detected in the *probe* pulse (Figure 2A).

Spontaneous release from the mixed disulfide is mediated by the C-terminal cysteine in the active site of oxidoreductases (Cys35 in TRX). We mutated this cysteine to abolish spontaneous release (TRX C35S, Figure 2B; see also Figure S3). This resulted in the complete absence of 25 nm steps and full oxidation of the substrate (as indicated by 11 and 14 nm steps in the *probe* pulse). Thus, every domain that folded in the presence of a mixed disulfide complex successfully completed oxidative folding. These observations indicate that 25 nm steps are only caused by spontaneous release of the enzyme from the mixed disulfide complex prior to folding, in agreement with scenario (2) described above (see Figure S4C for further support of this

conclusion). We could therefore use 25 nm steps as an indicator of spontaneous enzyme release.

Measurement of the PDla Spontaneous Release Rate

In our experiments, we could detect the formation of individual mixed disulfide complexes (14 nm steps in the *denature* pulse), their presence during protein folding (14 nm steps in the *probe* pulse), and their spontaneous dissociation prior to folding (25 nm steps in the *probe* pulse). By varying the time before folding (the duration of the *denature* pulse, Δt_D ; see Figures 3 and S4A) and measuring the resulting proportion of 25 nm steps in the *probe* pulse, we could thus measure the rate of PDla spontaneous release from a protein substrate. In these experiments, we used a force protocol optimized for long experiments with split *denature* and *probe* pulses (see Experimental Procedures).

For a *denature* pulse duration $\Delta t_D = 5$ s, ~50% of the refolded domains contained disulfide bonds (11 nm, 14 nm steps in *probe* pulse of Figure 3A). In contrast, for $\Delta t_D = 30$ s, the enzymes had more time to release from the mixed disulfide complex, which led to 80% of folded domains not having acquired disulfide bonds (25 nm steps in *probe* pulse of Figure 3B).

Figure 3D shows a simple model of the PDla dissociation process. This model contains three rates, representing the rate of spontaneous enzyme release (k_{off}), the rate of oxidative folding from the mixed disulfide complex (k_{ox}), and the rate of folding of the reduced substrate (k_{fold}). The model also assumes that PDla can dissociate during both Δt_D and Δt_F . Notably, the total amount of refolding remained constant for all values of Δt_D (Figure S4B), leading to the conclusion that k_{ox} and k_{fold} were approximately equal (this is further verified by the kinetic data in Figure 4C). Given that the folding and oxidation rates were similar, it follows that 14 versus 25 nm steps report on the PDla enzyme being “on” and “off,” respectively. We measured the dissociation of PDla as the ratio between the number of 25 nm steps and the total number of refolded domains in the *probe* pulse. In Figure 3C, this dissociated fraction is displayed as a function of total time before probing ($\Delta t_D + \Delta t_F$). We obtained the PDla release rate by fitting these data with a single exponential curve constrained to zero at time zero. The result of the fit was consistent with a spontaneous PDla release rate $k_{off} = 0.10 \pm 0.03$ s⁻¹.

(B) In this study, we used an atomic force microscope (AFM) to create mixed disulfide complexes between PDla enzymes and a single extended protein. Starting from this state, we investigated how PDla catalyzes oxidative folding.

(C) Mixed disulfide complexes were formed by applying a constant stretching force to a folded protein containing a buried disulfide, thus unfolding the protein and exposing the disulfide. A reduced PDla enzyme could then form a mixed disulfide with one of the cysteines in the substrate through S_N2-attack by the enzyme thiolate on the substrate disulfide (charges have been omitted for clarity; see also Figure S1).

(D) In an experimental recording of the end-to-end length of the I27^{32–75} substrate under force, unfolding of the substrate was detected as an 11 nm extension step. Formation of a mixed disulfide was detected as a 14 nm extension step arising from the cleavage of the 32–75 disulfide in the substrate (unfiltered data shown in light grey).

(E) A mechanical force was applied to a polyprotein consisting of repeated I27^{32–75} domains, in a solution containing reduced PDla.

(F) The substrate was extended, and mixed disulfides were formed between PDla enzymes and a cysteine in each domain, as described in (C). The force was then removed, and the resulting folding and oxidation of the substrate were probed after a preset folding interval Δt_F .

(G) Representative recording showing extension and force measurements for the [*denature* – *folding* – *probe*] force protocol. Arrowheads indicate formation of mixed disulfide complexes with PDla (14 nm steps). After refolding for 5 s, four domains had folded and acquired disulfides, as revealed by the subsequent 11 and 14 nm steps in the *probe* pulse. Other traces revealed refolding without disulfide formation (25 nm step, inset; see also Figure S2B).

(H) Step-size histograms compiled from several recordings confirm that PDla catalyzed oxidative folding in some domains, whereas other domains refolded in a reduced state ($n = 137$).

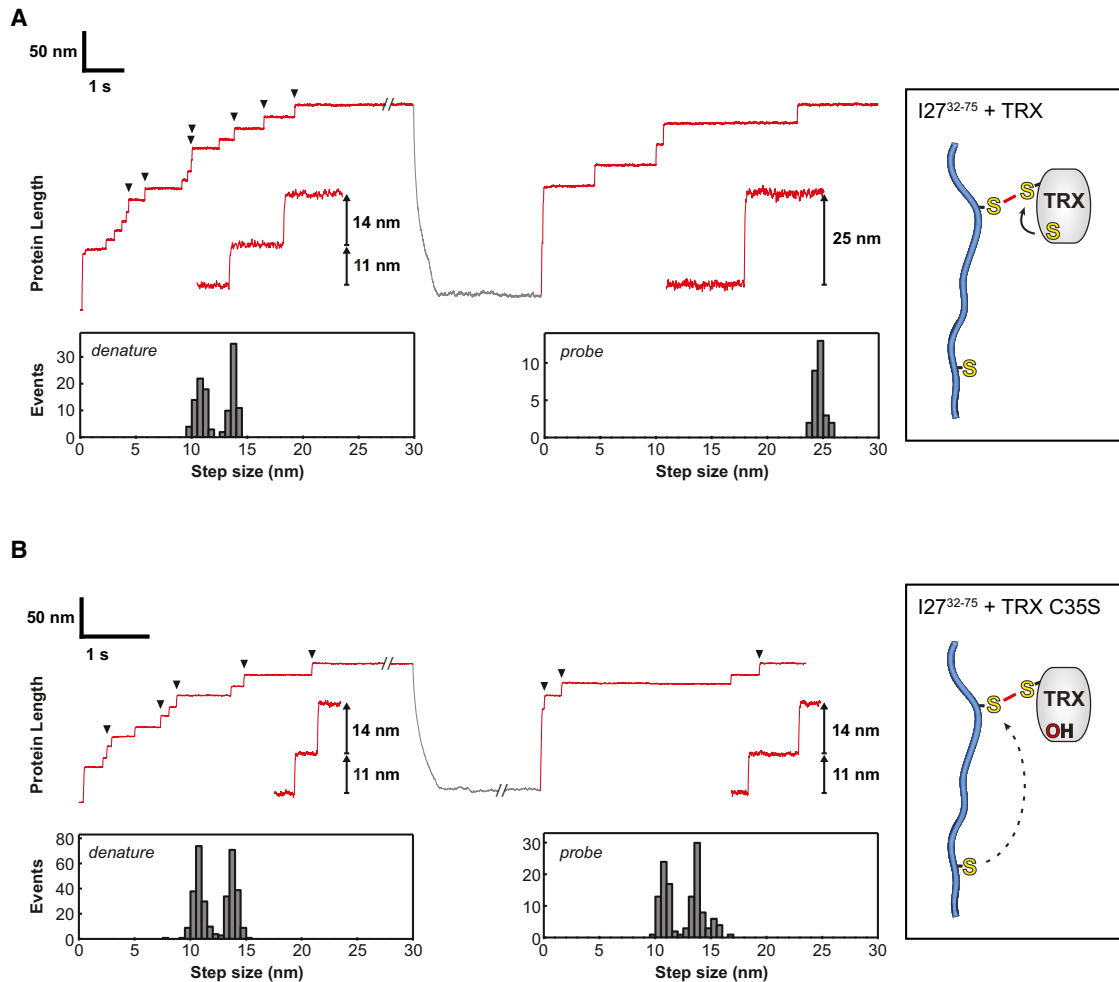


Figure 2. Enzyme Release Determines Outcome of Oxidative Folding

(A) Mixed disulfide complexes with thioredoxin (TRX) do not catalyze disulfide formation in the folding I27^{32–75} polypeptide. In this recording, seven domains were completely extended during the *denature* pulse. Four of these then refolded, albeit without disulfides, as seen from the 25 nm steps in the *probe* pulse. The rapid release mechanism of TRX accounts for its inability to catalyze disulfide formation (right panel) ($n = 61$).

(B) The C35S mutation in TRX replaces a reactive sulfur atom with oxygen and thus prevents spontaneous release of the enzyme (see Figure S3). TRX C35S catalyzed disulfide formation in the folding polypeptide by remaining in the mixed disulfide complex upon substrate folding ($\Delta t_F = 3$ s). Step-size histograms show that disulfides had been formed in all refolded domains ($n = 169$). Arrowheads indicate mixed disulfide complex formation.

Kinetics of Oxidative Protein Folding from an Extended State

A mixed disulfide complex with PDla enables the catalysis of oxidative folding, yet it is unknown how protein folding is affected by the covalent attachment of this enzyme. Our approach allowed us to directly measure the effect of a covalently bound PDla on protein folding.

We first set out to establish the refolding properties of I27^{32–75} in isolation, with and without the 32–75 disulfide bond. Using the [*denature* – *folding* – *probe*] protocol, we first measured folding kinetics of the reduced substrate. Figure 4A shows a representative trace for a refolding time $\Delta t_F = 3$ s. Steps of 25 nm are seen in both the *denature* and *probe* pulses, confirming that none of the domains contained disulfide bonds (in the absence of enzymes, we never observed disulfide formation during folding; Figure S5 shows the equivalent experiment with

oxidized I27^{32–75}). A longer refolding time of $\Delta t_F = 10$ s allowed for more domains to refold (Figure 4B). By varying the refolding time Δt_F , we could measure the kinetics of refolding. We found that the reduced protein folded at a rate of 0.27 s^{-1} , whereas presence of the 32–75 disulfide increased the folding rate by nearly 30 times to 7.74 s^{-1} (Figure 4C).

Strikingly, the time course of PDla-catalyzed oxidative folding was similar to the time course of folding of the reduced substrate (Figure 4C). Despite the steric hindrance caused by covalent attachment, PDla apparently did not interfere with the folding protein. Single exponential kinetics accounted well for the PDla-catalyzed oxidative folding data (solid red line in Figure 4C), suggesting a single rate-limiting step late in the oxidative folding pathway. Statistical analysis showed that the rate of oxidation was marginally yet significantly faster than folding ($p < 0.01$), indicating that this rate-limiting step occurred before folding had

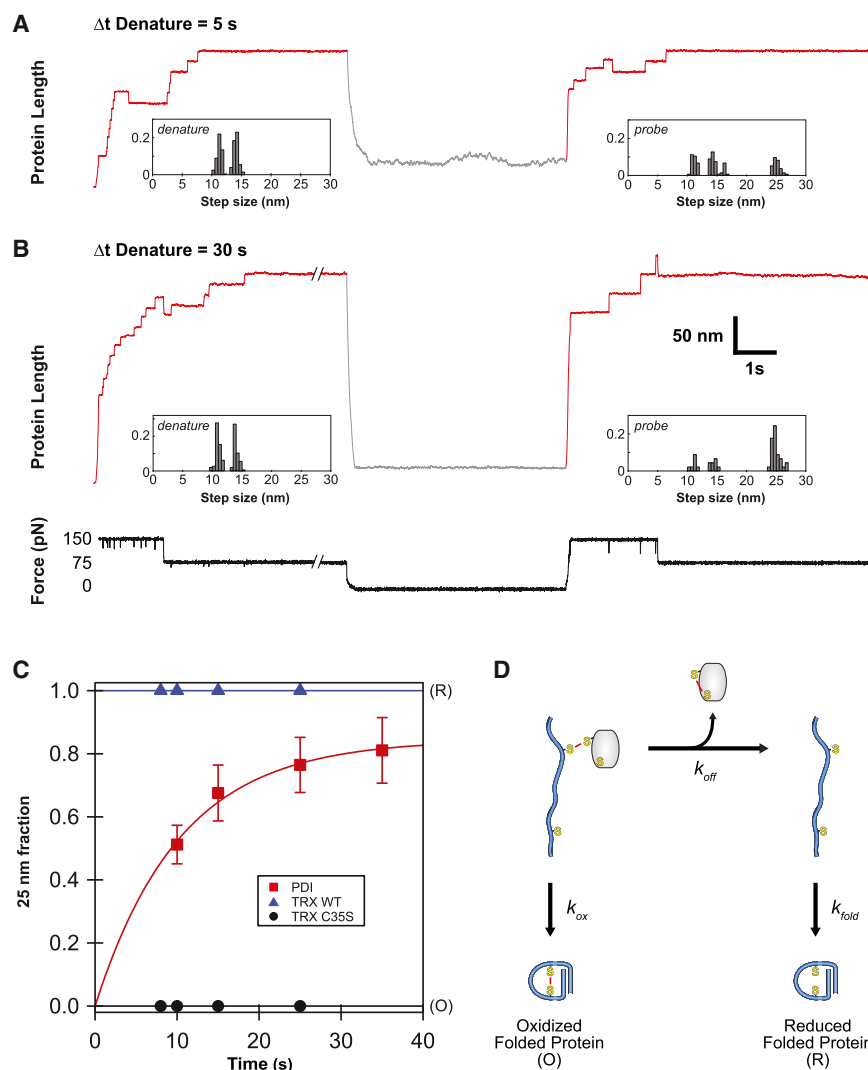


Figure 3. Measurement of PDIa Release Kinetics

(A) In an experiment with a 5 s denature pulse ($\Delta t_D = 5 \text{ s}$), $\sim 50\%$ of PDIa enzymes remained in the mixed disulfide complexes during the subsequent protein-folding interval. As a result, oxidative folding mostly completed successfully, as seen from the predominance of 11 and 14 nm steps in the probe pulse ($n = 105$).

(B) After a 30 s denature pulse ($\Delta t_D = 30 \text{ s}$), $\sim 80\%$ of mixed disulfides had been cleaved through spontaneous release of PDIa. Folding was still observed, but only few of the folded domains contained disulfides, as seen from the predominance of 25 nm steps in the probe pulse ($n = 66$). See also Figure S4A.

(C) Spontaneous enzyme release, measured as the fraction of folded domains that were reduced by the start of the probe pulse (25 nm steps, fraction \pm SEM). The fraction is displayed as a function of total time ($\Delta t_D + \Delta t_F$). The data for PDIa fall between the rapidly releasing TRX and the release-deficient TRX C35S. Assuming the model in (D), the rate of spontaneous PDIa release was calculated from an exponential fit (solid line) ($n > 100$).

(D) Kinetic model including the rate of spontaneous enzyme release (k_{off}), the rate of oxidative folding from the mixed disulfide complex (k_{ox}), and the rate of folding of the reduced substrate (k_{fold}).

completed. The native state was acquired rapidly after disulfide formation, as inferred from the fast refolding kinetics of the oxidized substrate (Figure 4C). Although PDIa-mediated oxidation could in theory be more accurately described as a multi-exponential process, fitting a double exponential to the PDIa data yielded only marginal improvement of the fit ($\chi^2 = 0.373$ versus $\chi^2 = 0.377$). Curiously, all experiments with enzymes (PDIa and TRX) showed lower absolute values of folding than the experiments without enzymes. However, we observed this effect also in experiments where both the substrate and the enzyme were reduced, wherein no mixed disulfides could have been formed (data not shown). This effect is therefore not caused by mixed disulfide complexes but rather by other, likely noncovalent, enzyme-substrate interactions.

Taken together, our data give an idea of how PDIa catalyzes disulfide formation in the folding protein. PDIa first becomes covalently attached to the substrate by forming a mixed disulfide complex. From there on, the enzyme acts as a passive placeholder, allowing the substrate to collapse and fold in an unin-

dered manner. At a late stage of folding, close in time to the acquisition of the native fold, PDIa catalyzes the formation of an intramolecular disulfide in the substrate. This reaction releases PDIa and allows the substrate to rapidly complete oxidative folding. An earlier study showed exceptionally high reactivity between PDI-substrate mixed disulfides and freely diffusing cysteine-containing peptides

(Darby and Creighton, 1995a). In light of these and our results, it appears that PDI-mediated oxidative folding is rate-limited not by the covalent chemical reaction but rather by the encounter rate of the reacting groups, which in turn is limited by protein folding.

Formation of Interdomain Disulfides Triggers Misfolding

The oxidative folding recordings with I27³²⁻⁷⁵ revealed in the probe pulse sporadic 15 nm steps, which were due to cleavage of interdomain disulfide bonds. Several lines of evidence support this conclusion. The most direct indication came from recordings where the events were clearly separated in time and thus allowed for higher precision when determining the step size. For instance, in the recording shown in Figure 5A, we achieved a spatial resolution better than 0.1 nm by localizing the centroids of Gaussians fitted to 1,000 unfiltered data points before and after each step. This analysis revealed that whereas the step size of the 32-75 reduction events was 13.8 nm (hereafter referred to as 14 nm steps for simplicity), the new steps had

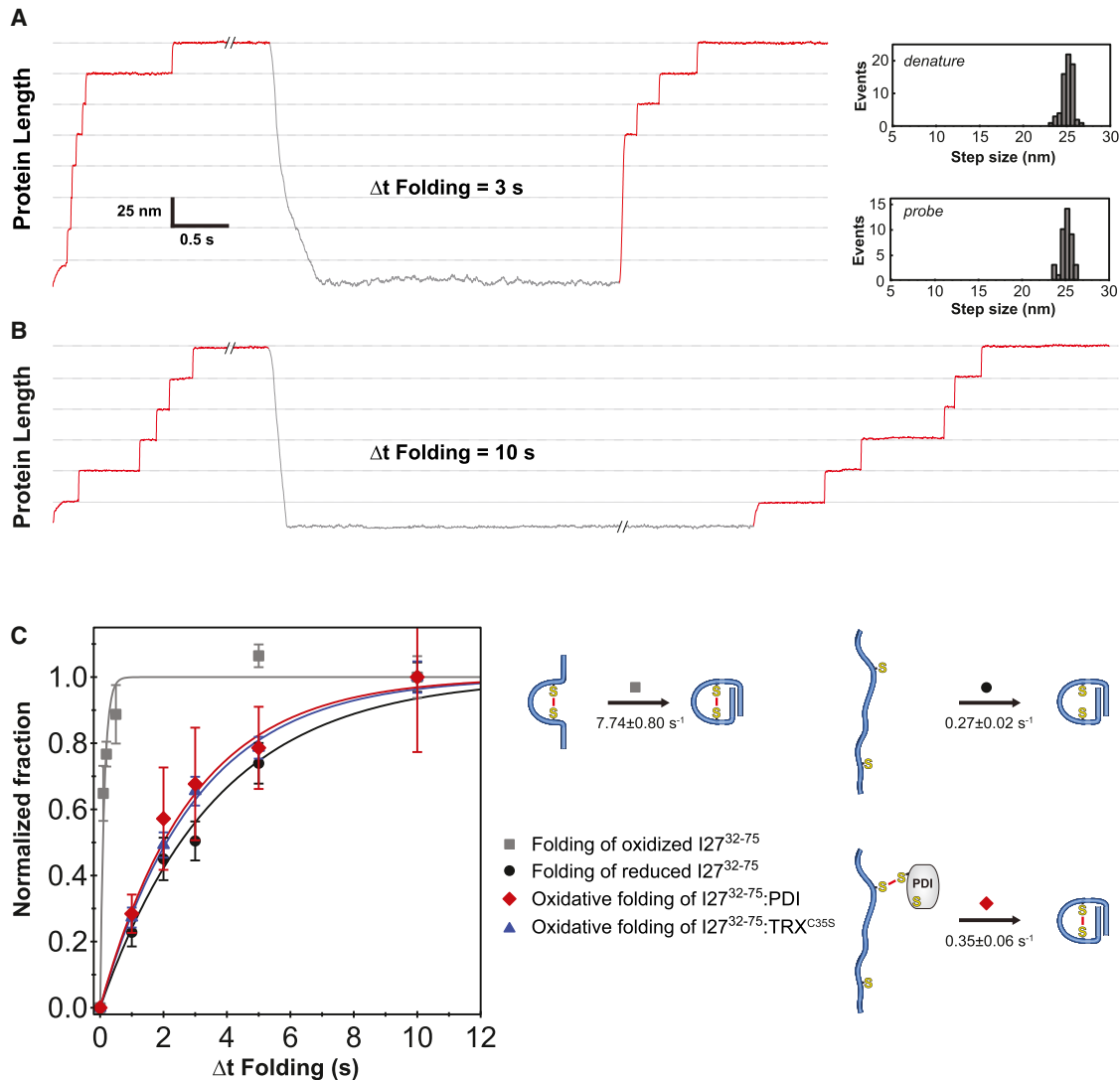


Figure 4. Measurement of Protein Folding from an Extended State

(A) Refolding experiments with prerduced I27³²⁻⁷⁵ polyproteins show no evidence of disulfide formation in the absence of enzymes ($n = 68$).

(B) Longer folding interval (Δt_f) increases refolding probability. Folding kinetics were measured by varying Δt_f .

(C) Time course of folding (gray squares and black circles) and oxidative folding (red diamonds and blue triangles) from an extended state (fraction \pm SEM; $n > 50$). I27 folding is accelerated nearly 30-fold by the presence of the 32–75 disulfide (Figure S5). The rates of disulfide formation catalyzed by PDI and TRX C35S are only marginally faster than the folding rate of the reduced substrate, indicating that disulfide formation occurred late in the oxidative folding process. Errors represent SEM.

a magnitude of 15.4 nm (hereafter referred to as 15 nm steps). The occurrence of 15 nm steps correlated well with the number of initially unfolded domains (Figures S6A and S6B); however, these steps did not represent unfolding events, as they were not induced by a high stretching force. Instead, their step size corresponded exactly to the expected elongation upon cleavage of a disulfide between Cys75 in one I27 domain and Cys32 in the next domain (15.4 ± 0.1 nm measured versus 15.5 nm predicted from Equation S1).

To investigate the prevalence of interdomain disulfides, we compiled an extended histogram from the oxidative folding data with PDI (Figure 5B). We then fit the data with multiple

fixed-width Gaussians and identified three clearly distinct peaks in the *probe* pulse that were not present in the *denature* pulse. One peak was centered at ~ 25 nm and corresponded to the unfolding of reduced domains. The two other peaks were centered at ~ 15 nm and ~ 30 nm and could be assigned to interdomain disulfides (Figure 5E). In total, the interdomain disulfides accounted for $\sim 20\%$ of the data.

Sporadic formation of interdomain disulfides was observed for different substrates (Figure 5D), including I27²⁴⁻⁵⁵ (Figure S7C) and I27 WT (Figure S7B). The high accuracy of our step assignment can be seen in Figure 5E, where the horizontal axis indicates the cysteine separation as a number of residues, and the

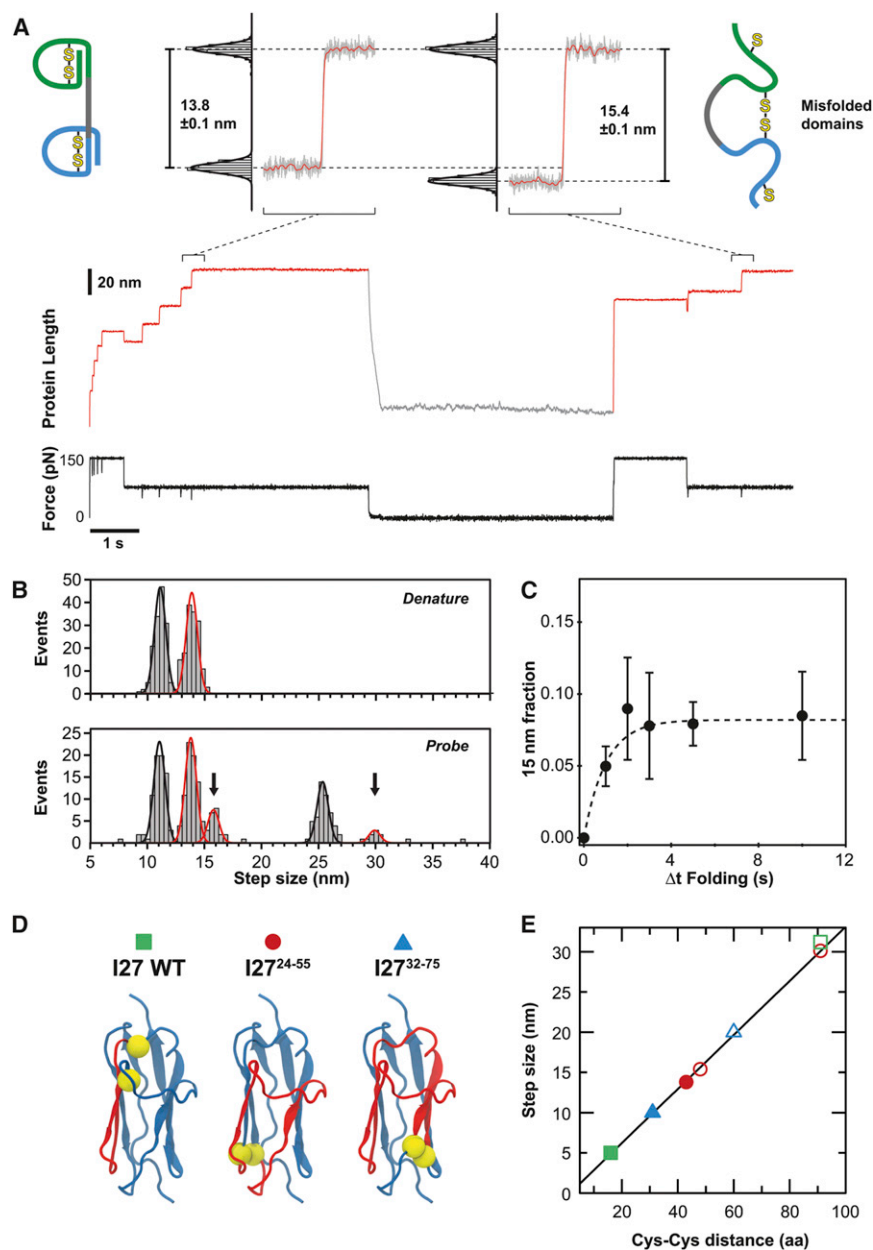


Figure 5. Non-native Interdomain Disulfides Prevent Native Folding

(A) Cleavage of a 32–75 intradomain disulfide resulted in a 13.8 nm step. Interdomain disulfides between Cys75 in one domain and Cys32 in an adjacent domain were detected as 15.4 nm steps. The recording shows unfolding and PDla-mediated disulfide cleavage of four domains in the *denature* pulse. Upon refolding, two interdomain disulfides were formed, but no refolding took place. Cleavage of the interdomain disulfides can be seen in the *probe* pulse as two 15.4 nm steps (values represent mean \pm SEM).

(B) Histograms reveal in the *probe* pulse populations of 15.4 nm and 30 nm steps (black arrows) that appear only after refolding in a mixed disulfide complex with PDla. These anomalous steps were not preceded by corresponding unfolding steps ($n = 302$; solid lines: Gaussian fits with $\sigma = 0.7$ nm).

(C) Kinetics of interdomain disulfide formation by PDla (fraction \pm SEM) ($n > 100$). See also Figure S6.

(D) The three substrates used in this study. Yellow spheres indicate locations of cysteine residues. Red segments are inextensible when an intradomain disulfide is present. Cleavage of a disulfide under force triggers extension of the corresponding red segment (Protein Data Bank [PDB]: 1TIT).

(E) Comparison of observed step sizes (symbols) with predicted values (Equation S1, solid line). Solid symbols represent cleavage of the predominant intradomain disulfides; open symbols represent cleavage of sporadically formed interdomain disulfides.

vertical axis indicates the measured step size. In all our experiments, interdomain disulfides were formed sporadically but did not appear to be favored during PDla-catalyzed oxidative folding.

The I27^{32–75} data showed that formation of interdomain disulfides precluded proper folding. Whereas nearly every 14 nm step in the *denature* and *probe* pulses was accompanied by a corresponding 11 nm step ($N_{11}/N_{14} = 97\% \pm 1\%$ and $91\% \pm 3\%$, respectively), 15 nm steps did not appear to be preceded by corresponding unfolding steps (Figure 5A). An 11 nm step sometimes appeared before a 15 nm step in our recordings. In all of these cases, however, the 11 nm step was also succeeded by a 14 nm step, showing that the 11+14 nm pair corresponded

correctly folded. Formation of non-native interdomain disulfides was thus sufficient to induce protein misfolding.

We explored the possibility that domain swapping had caused the formation of interdomain disulfides. Domain swapping has been observed in I27 in previous studies (Borgia et al., 2011; Oberhauser et al., 1999). However, these studies showed that domain-swapped structures retain mechanical stability. We found no evidence of such structures in our data and therefore conclude that the interdomain disulfides were likely not caused by domain swapping.

Remarkably, the interdomain disulfides appeared at a faster rate than native disulfides (1.0 s^{-1} , Figure 5C). The proportion of interdomain disulfides reached a plateau after ~ 2 s, indicating

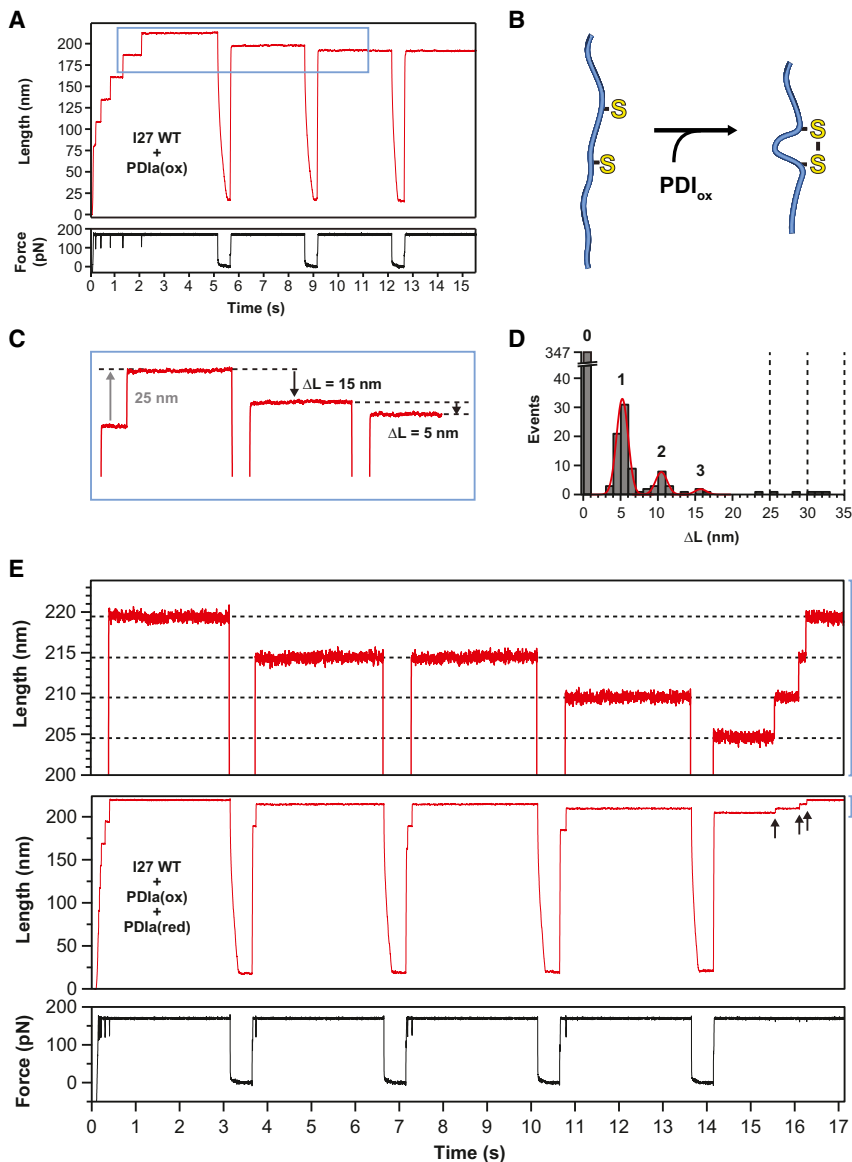


Figure 6. Disulfide Formation Catalyzed by Oxidized PDIA

(A) I27 WT was refolded in solution containing oxidized PDIA. After initial unfolding, the substrate was allowed to collapse for periods of 0.5 s between repeated probe pulses. Formation of a disulfide bond between Cys47 and Cys63 in a domain was detected as a 5 nm shortening of the stretched substrate. Occasionally, multiple oxidation events were detected after a single refolding interval, resulting in 10 or 15 nm steps. (B) PDIA-mediated disulfide formation was detected from the resulting substrate shortening. (C) Magnified view of the recording shown in (A). (D) Histogram of substrate shortenings ΔL in subsequent pulses, compiled from multiple recordings. Numbers above the peaks indicate the corresponding number of intradomain disulfides. Vertical dotted lines indicate predicted shortenings upon formation of interdomain disulfides ($n = 439$). (E) Refolding of I27 WT in solution containing 1:1 molar ratio of reduced PDIA and oxidized PDIA. Similar to the result in (A), disulfide formation was detected as shortening of the substrate. These disulfides could then be cleaved by reduced PDIA enzymes, as seen from the 5 nm extension steps (arrows). See also Figure S7A.

experiments, non-native disulfides could potentially be resolved by reduced PDIA or through intramolecular rearrangement involving the attack of a free cysteine in the substrate, as shown in a previous study (Alegre-Cebollada et al., 2011). However, we did not detect a decrease in interdomain disulfides over time (Figure 5C), indicating that they are relatively stable and not prone to isomerization on the timescale of our experiments.

Formation of Disulfides by Oxidized PDIA

In the ER, PDI-substrate mixed disulfides are thought to form mainly as a result of

that they were formed at an early stage of folding. This is consistent with the substrate being less structured early on during the folding process. If the interdomain disulfides were formed through random collision between cysteines in different domains, then they would be more likely to form at an early stage, as compared to later stages of folding. In this scenario, the folding pathway determines the fidelity of disulfide formation. As domains segregate and tertiary structure starts to form, interdomain contacts lessen, which in turn decreases the probability of forming interdomain disulfides. The accuracy of disulfide formation thus reaches a maximum as the protein approaches the native state.

In cells, non-native interdomain disulfides have been observed during the oxidative folding of low-density lipoprotein receptor (LDL-R) in the ER (Jansens et al., 2002). These disulfides were isomerized into a native configuration on a timescale of hours. In our

oxidized PDI reacting with free cysteines in a substrate (Frاند and Kaiser, 1999). To test whether our results were valid also under these conditions, we investigated the reaction of oxidized PDIA with a reduced substrate. Initially, we studied PDIA-mediated oxidation of reduced I27³²⁻⁷⁵. However, we found that folded I27³²⁻⁷⁵ became oxidized at a rate too rapid to enable observation of oxidative folding in a single molecule (Figures S6C and S6D). We therefore chose a substrate with two cysteines that are known to be buried, WT I27 (see Figure 5D). The two cysteines in I27 WT are close in the native structure yet have not been found to form a disulfide (Improta et al., 1996). We speculated that folding in the presence of oxidized PDIA could induce disulfide formation. Formation of an intradomain disulfide in I27 WT shortens its stretched length by 5 nm (Figure 6B). We used a polyprotein consisting of eight repeats of I27 WT to investigate disulfide formation in the presence of

12.5 μM oxidized PDla. The initial staircase in Figure 6A shows only 25 nm steps, verifying that all domains were natively folded and reduced at the beginning of the experiment. When a domain becomes unfolded, oxidized PDla can form mixed disulfide complexes with the newly exposed cysteines. After exposure, followed by a brief refolding interval in the absence of force ($\Delta t_F = 0.5$ s), the substrate was again stretched but did not reach its fully extended length. This shortening, ΔL , may be due to intramolecular disulfide formation (Figures 6B and 6C). A histogram of all observed shortenings for a large number of recordings ($n = 439$) is shown in Figure 6D. Although the majority of refolding intervals did not yield shortenings, three peaks can be seen centered at shortenings of 5, 10, and 15 nm. We reasoned that these three peaks corresponded to the formation of 1, 2, and 3 intradomain disulfide bonds, in separate substrate domains. To prove this hypothesis, we repeated the experiment in the presence of a 1:1 mixture of reduced and oxidized PDla and now observed 5 nm steps as the newly formed disulfides were cleaved by reduced PDla (Figures 6E and S7A). Although present (see Figure S7B), interdomain disulfide formation was not frequently observed in our experiments. We did not observe mechanically stable folded structures in oxidized domains; nevertheless, the data indicate a strong preference for intradomain over interdomain disulfides during PDla-mediated oxidation of I27 WT. As the native fold of the substrate appeared to determine the disulfide formation pattern, this confirms that protein folding drives disulfide formation also in the context of oxidized PDla.

DISCUSSION

For a protein, the path to a native fold is lined with potential traps (Dobson, 2003). Protein folding has become a field of intense research due to its fundamental importance as well as its relevance in numerous diseases. The formation of disulfide bonds adds an additional layer of complexity to the folding pathway of many proteins. For these proteins, cells have evolved dedicated pathways to ensure efficient oxidative folding (Chakravarthi et al., 2009; Sevier and Kaiser, 2002).

In eukaryotes, nascent polypeptides targeted to the ER are exported from the cytosol cotranslationally (Rapoport et al., 1996). Emerging in the ER lumen, polypeptides that contain cysteines form mixed disulfide complexes with oxidase enzymes such as PDI (Bulleid and Freedman, 1988; Molinari and Helenius, 1999). However, it has remained unknown how a mixed disulfide complex affects protein folding and how oxidation and folding are coupled. To investigate this, we used single-molecule AFM to reconstitute mixed disulfide complexes between PDla and an unfolded model protein. Our approach enabled direct and independent measurements of protein folding and disulfide formation.

Does disulfide formation drive protein folding, or conversely, does protein folding provide a driving force for native disulfide formation? Both ideas have been proposed in previous studies (Camacho and Thirumalai, 1995; Wedemeyer et al., 2000; Welker et al., 2001; Wilkinson and Gilbert, 2004), without any consensus being reached in the field. In our experiments, the presence of a mixed disulfide with PDla enabled disulfide formation in a folding protein. Whereas protein folding appeared to rate-limit oxidative

folding, attachment of PDla did not impede the protein-folding rate. Furthermore, native disulfide formation was catalyzed by PDla at a late stage of protein folding. The enzyme remained in the mixed disulfide complex until the substrate had attained a near-native state. At this point, the substrate disulfide was formed, and PDla was released. By enabling oxidation while interfering minimally with protein folding, we conclude that PDla functions more like a passive placeholder than as an active folding catalyst. Our results thus indicate that protein folding provides the driving force during PDla-mediated oxidative folding.

Nascent polypeptides are constrained in an extended state as they traverse the ribosomal exit tunnel and Sec translocase (Becker et al., 2009). Before folding can take place, extended polypeptides first have to undergo entropic and hydrophobic collapse (Berkovich et al., 2010; Fernandez and Li, 2004; Garcia-Manyes et al., 2009). In the absence of force, extended full-length proteins collapse in a fraction of a second and then sample an ensemble of compact conformations that over a time-scale of several seconds acquire their native contacts. Such “molten globule” precursor states have been studied extensively, and they are today recognized as a significant deviation from the classical two-state model of protein folding (Baldwin et al., 2010; Garcia-Manyes et al., 2009; Kuwajima, 1989; Ptit-syn, 1995). We found that the kinetics of disulfide formation corresponded closely to the kinetics of formation of the mechanically stable native state. A molten globule state was thus generally not sufficient for native disulfide formation. Instead, PDla catalyzes disulfide formation with high efficiency only after a large fraction of the native contacts have been acquired (Figure 7).

Cotranslational folding in eukaryotes is limited by the ribosomal translation rate of ~ 5 amino acids per second (Ingolia et al., 2011) (Figure 7). Nuclear magnetic resonance (NMR) studies of cotranslational folding have shown that, although proteins can start acquiring some structural elements while still attached to the ribosome, they acquire their native fold in a domain-wise fashion (Cabrita et al., 2009; Eichmann et al., 2010). For a typical 100 amino acid domain, this leaves around 20 s during which its cysteines are relatively accessible and therefore able to form mixed disulfides with PDI. Based on the estimated concentration of PDI in the ER (up to 1 mM) (Lyles and Gilbert, 1991; Zapun et al., 1992) and the rate of mixed disulfide formation ($>0.1 \text{ mM}^{-1}\text{s}^{-1}$; see Figures S6C and S6D), there is a high probability that a mixed disulfide is present before folding occurs.

When an entire protein domain has been translocated into the ER, the mixed disulfide has to remain in place as the protein folds into its native state. What happens if PDI dissociates before its substrate is completely folded? An exposed substrate cysteine can form a new mixed disulfide. However, protein folding can cause the burial of cysteine residues, which in turn can prevent thiol-disulfide exchange. In an earlier study, prematurely folded domains could acquire disulfides only after spontaneous unfolding, which could take up to an hour (Walker and Gilbert, 1995). To avoid unproductive waiting times and/or consumption of energy through active unfolding of such substrates, PDI needs to introduce disulfides when the nascent protein folds for the first time. This in turn requires the spontaneous off-rate of PDI to be slower than the rate of protein folding. We measured the off-rate of human PDla to be 0.1 s^{-1} . Human PDI is thus an efficient

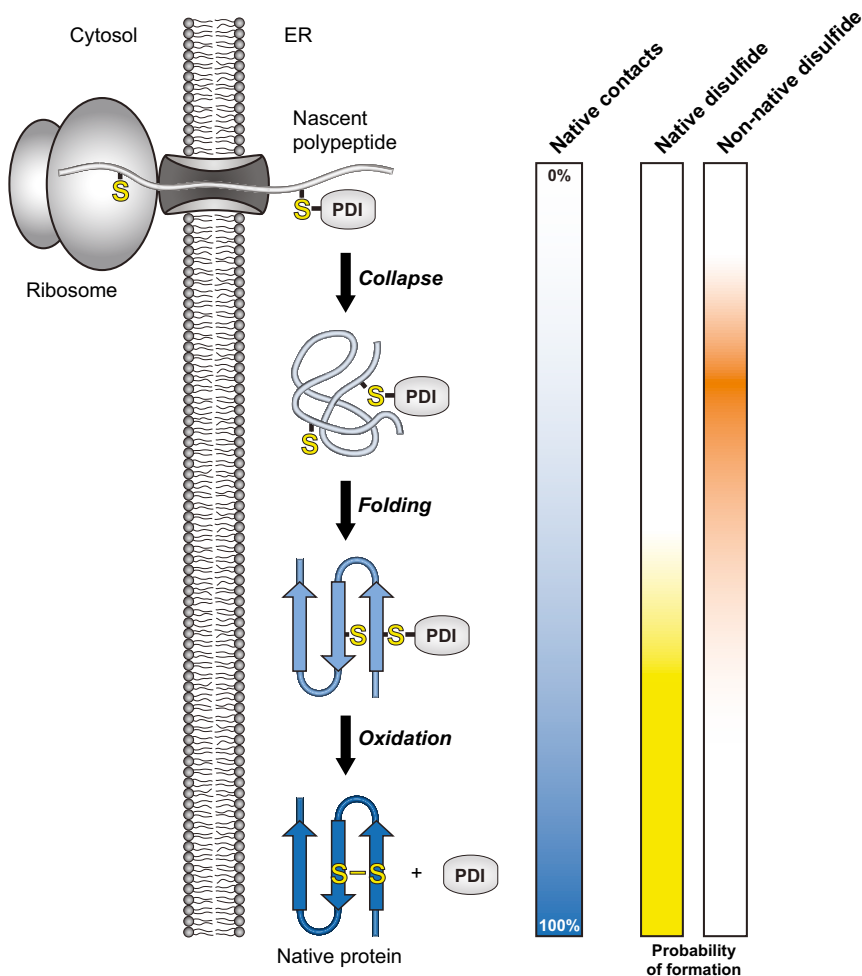


Figure 7. PDI-Mediated Oxidative Folding in the ER

A nascent polypeptide is kept physically extended while emerging from the ribosome and is unable to fold or form disulfide bonds before it enters the ER. PDI enables oxidative folding in the ER lumen by forming a mixed disulfide with a cysteine in the nascent polypeptide. The polypeptide collapses before folding takes place, whereas PDI remains attached throughout this process. Non-native disulfides are sporadically formed during the early stages of folding. PDI favors the formation of native disulfides by allowing the polypeptide to fold into a near-native state before it catalyzes disulfide formation. By allowing protein folding to guide the pairing of cysteines, PDI can catalyze oxidative folding without the need for substrate-specific interactions. Graphs on right side: color intensity indicates percentage of native contacts (blue bar) and probability of disulfide formation (yellow and orange bars).

In view of this, Wilkinson and Gilbert proposed that it is the substrate, rather than PDI, that determines the pairing of cysteines (Wilkinson and Gilbert, 2004). Our data lend strong support to this theory by showing that covalent attachment of PDIa did not significantly interfere with the folding protein. PDIa could thereby favor native disulfides by allowing its substrate to decide, through folding, which cysteines to join. Because this mechanism does not rely on any substrate-specific interactions besides the mixed disulfide, it can explain how a single enzyme can catalyze the oxidative

oxidative folding catalyst for substrates that fold in less than 10 s after they collapse. This time limit is likely sufficient for most secreted proteins and certainly sufficient for small proteins with simple globular structure. However, protein folding can take place on a wide range of timescales. It is therefore tempting to suggest that the off-rate of PDI is adapted to the folding rate of its substrates. The human PDI family has at least 19 members. Perhaps this diversity is required to accommodate the range of protein-folding rates.

Although spontaneous release by PDIa precluded successful oxidative folding in our experiments, this mechanism can serve an important purpose as a release timer in situations where the mixed disulfide complex is unproductive (Walker and Gilbert, 1997).

How does PDI ensure correct pairing of cysteines in proteins with more than one disulfide? For a 4-disulfide protein, there are more than 700 possible disulfide-bonded configurations. Despite the vast number of possibilities, the majority of bonds formed by PDIa in our experiments were the correct intradomain disulfides. In order to catalyze oxidative folding in a wide range of substrates, PDI must rely on a catalytic mechanism that is general yet ensures correct bonding for each specific protein.

folding of a wide variety of proteins. Nevertheless, it is conceivable that other oxidases affect protein folding in diverse ways; for instance, we have found that mixed disulfides with the small molecule glutathione inhibit protein folding (unpublished data).

Oxidative folding in the cell is a highly complex process that involves many components, including oxidoreductases, chaperones, proline *cis-trans* isomerases, and small redox molecules such as glutathione. All of these components can interact with a protein during its folding process, and the resulting complexity has severely hindered detailed mechanistic studies of oxidative folding (Chakravarthi et al., 2009). We have here presented an approach that enables precise control and measurement of both folding and disulfide formation in single protein molecules. These methods can effectively be used to determine differences between oxidase enzymes, as well as the function of other factors involved in oxidative folding in the cell.

EXPERIMENTAL PROCEDURES

Single-Molecule AFM

The details of our custom-made atomic force microscope have been described previously (Fernandez and Li, 2004). We used silicon nitride cantilevers (Bruker MLCT) with a typical spring constant of 15 pN nm⁻¹, calibrated using

the equipartition theorem (Florin et al., 1995). The force-clamp mode provided a feedback time constant of 5 ms. The buffer used in all the experiments was 10 mM HEPES, 150 mM NaCl, 1 mM EDTA, degassed and at pH 7.2. The concentrations used were 10 μ M reduced TRX, 40 μ M reduced TRX C35S, 120 μ M reduced PDI A1, and 12.5 μ M oxidized PDI A1. These enzyme concentrations were chosen so as to yield similar rates of mixed-disulfide formation. The polyprotein substrate was added in a droplet and allowed to bind to a freshly evaporated gold coverslip before the experiments. Every experiment consisted of repeated trials where the tip was pressed against the surface at 1 nN for 0.5 s and subsequently retracted. If attachment was achieved, the pulse protocol was applied until detachment occurred. The oxidative folding force-clamp experiments used a triple-pulse [denature – folding – probe] force protocol. The first pulse was maintained at 130–150 pN for a time long enough to ensure complete unfolding and reduction of the substrate (at least 5 s). The second pulse was set at 0 pN and maintained for the desired amount of refolding time. The third pulse was set at a force identical to the first and maintained until complete unfolding and reduction could be ensured (at least 5 s). For measurement of the data in Figures 3, 4, and 5, split denature and probe pulses were used. This method allowed for a clearer separation of unfolding and reduction steps and has been described earlier (Wiita et al., 2006).

Data Analysis

We used custom-written software in IGOR Pro (Wavemetrics) to collect and analyze data. Recordings were low-pass filtered at 1 kHz for display purposes. Traces were selected based on the fingerprint consisting of at least two unfolding events in the denature pulse. Initially oxidized I27^{32–75} traces exhibiting step sizes other than 11 nm or 14 nm in the denature pulse were excluded from the analysis to ensure homogeneity. For initially reduced substrate samples, 25 nm steps were used as the selection criterion. More than 50% of all traces with repeated steps met these criteria. Only traces showing equal extension at the end of the probe pulse and the end of the denature pulse were included, to ensure that the same protein was stretched in the two pulses. Step-size histograms were generated using all steps >5 nm detected after the initial elastic extension in each force pulse. A 2.5 nm minimum cutoff was applied to the histogram in Figure 6D. Histograms in the main-text figures were compiled from representative experiments. The number of independent observations *n* was counted as the total number of protein domains (as observed in the denature pulse). Standard error of the mean (SEM) for fractions was estimated through the bootstrap method, where each recording was treated as an independent data point. SEM for fit parameters was determined as the standard error for the coefficient in the fit, given the measurement errors of the individual data points.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Extended Experimental Procedures and seven figures and can be found with this article online at <http://dx.doi.org/10.1016/j.cell.2012.09.036>.

ACKNOWLEDGMENTS

This work was supported by grants from the NIH (HL066030 and HL061228 to J.M.F.; CA097061, CA161061, and GM085081 to B.R.S.), HHMI (B.R.S.), the Arnold and Mabel Beckman Foundation (B.R.S.), the Spanish Ministry of Education (A.I.-P.), the Spanish Ministry of Science and Innovation (BIO2009-09562 and CSD2009-00088 to J.M.S.-R.), FEDER (J.M.S.-R.), Fundación Ibercaja (J.A.-C.), and the Training Program in Molecular Biophysics (T32GM008281 to A.K.). P.K. was supported by a Fulbright scholarship. P.K., J.A.-C., and J.M.F. designed the experiments. P.K., J.A.-C., and J.F. performed the experiments and analyzed the data. A.K. and B.R.S. provided PDI A1. A.H. provided TRX. A.I.-P. and J.M.S.-R. provided TRX C35S. P.K., J.A.-C., and J.M.F. wrote the manuscript. We thank the members of the Fernández lab for their illuminating feedback.

Received: January 11, 2012

Revised: July 3, 2012

Accepted: September 5, 2012

Published: November 8, 2012

REFERENCES

- Ainavarapu, S.R., Brujic, J., Huang, H.H., Wiita, A.P., Lu, H., Li, L., Walther, K.A., Carrion-Vazquez, M., Li, H., and Fernandez, J.M. (2007). Contour length and refolding rate of a small protein controlled by engineered disulfide bonds. *Biophys. J.* 92, 225–233.
- Alegre-Cebollada, J., Kosuri, P., Rivas-Pardo, J.A., and Fernández, J.M. (2011). Direct observation of disulfide isomerization in a single protein. *Nat. Chem.* 3, 882–887.
- Baldwin, R.L., Frieden, C., and Rose, G.D. (2010). Dry molten globule intermediates and the mechanism of protein unfolding. *Proteins* 78, 2725–2737.
- Becker, T., Bhushan, S., Jarasch, A., Armache, J.P., Funes, S., Jossinet, F., Gumbart, J., Mielke, T., Berninghausen, O., Schulten, K., et al. (2009). Structure of monomeric yeast and mammalian Sec61 complexes interacting with the translating ribosome. *Science* 326, 1369–1373.
- Berkovich, R., Garcia-Manyes, S., Urbakh, M., Klafter, J., and Fernandez, J.M. (2010). Collapse dynamics of single proteins extended by force. *Biophys. J.* 98, 2692–2701.
- Borgia, M.B., Borgia, A., Best, R.B., Steward, A., Nettels, D., Wunderlich, B., Schuler, B., and Clarke, J. (2011). Single-molecule fluorescence reveals sequence-specific misfolding in multidomain proteins. *Nature* 474, 662–665.
- Bulleid, N.J., and Freedman, R.B. (1988). Defective co-translational formation of disulphide bonds in protein disulphide-isomerase-deficient microsomes. *Nature* 335, 649–651.
- Cabrera, L.D., Hsu, S.T., Launay, H., Dobson, C.M., and Christodoulou, J. (2009). Probing ribosome-nascent chain complexes produced in vivo by NMR spectroscopy. *Proc. Natl. Acad. Sci. USA* 106, 22239–22244.
- Camacho, C.J., and Thirumalai, D. (1995). Modeling the role of disulfide bonds in protein folding: entropic barriers and pathways. *Proteins* 22, 27–40.
- Carvalho, A.P., Fernandes, P.A., and Ramos, M.J. (2006). Similarities and differences in the thioredoxin superfamily. *Prog. Biophys. Mol. Biol.* 91, 229–248.
- Chakravarthi, S., Jessop, C.E., and Bulleid, N.J. (2009). Oxidative folding in the endoplasmic reticulum. In *Oxidative Folding of Peptides and Proteins, Chapter 1.5* (Cambridge, UK: The Royal Society of Chemistry), pp. 81–104.
- Chivers, P.T., and Raines, R.T. (1997). General acid/base catalysis in the active site of *Escherichia coli* thioredoxin. *Biochemistry* 36, 15810–15816.
- Darby, N.J., and Creighton, T.E. (1995a). Characterization of the active site cysteine residues of the thioredoxin-like domains of protein disulfide isomerase. *Biochemistry* 34, 16770–16780.
- Darby, N.J., and Creighton, T.E. (1995b). Functional properties of the individual thioredoxin-like domains of protein disulfide isomerase. *Biochemistry* 34, 11725–11735.
- Debarbieux, L., and Beckwith, J. (2000). On the functional interchangeability, oxidant versus reductant, of members of the thioredoxin superfamily. *J. Bacteriol.* 182, 723–727.
- Di Jeso, B., Park, Y.N., Ulianich, L., Treglia, A.S., Urbanas, M.L., High, S., and Arvan, P. (2005). Mixed-disulfide folding intermediates between thyroglobulin and endoplasmic reticulum resident oxidoreductases ERp57 and protein disulfide isomerase. *Mol. Cell. Biol.* 25, 9793–9805.
- Dobson, C.M. (2003). Protein folding and misfolding. *Nature* 426, 884–890.
- Dong, G., Wearsch, P.A., Peaper, D.R., Cresswell, P., and Reinisch, K.M. (2009). Insights into MHC class I peptide loading from the structure of the tapasin-ERp57 thiol oxidoreductase heterodimer. *Immunity* 30, 21–32.
- Eichmann, C., Preissler, S., Riek, R., and Deuerling, E. (2010). Cotranslational structure acquisition of nascent polypeptides monitored by NMR spectroscopy. *Proc. Natl. Acad. Sci. USA* 107, 9111–9116.
- Fernandez, J.M., and Li, H. (2004). Force-clamp spectroscopy monitors the folding trajectory of a single protein. *Science* 303, 1674–1678.
- Florin, E.L., Rief, M., Lehmann, H., Ludwig, M., Dornmair, C., Moy, V.T., and Gaub, H.E. (1995). Sensing specific molecular-interactions with the atomic-force microscope. *Biosens. Bioelectron.* 10, 895–901.

- Frand, A.R., and Kaiser, C.A. (1999). Ero1p oxidizes protein disulfide isomerase in a pathway for disulfide bond formation in the endoplasmic reticulum. *Mol. Cell* 4, 469–477.
- García-Manyes, S., Dougan, L., Badilla, C.L., Brujic, J., and Fernández, J.M. (2009). Direct observation of an ensemble of stable collapsed states in the mechanical folding of ubiquitin. *Proc. Natl. Acad. Sci. USA* 106, 10534–10539.
- Gething, M.J., and Sambrook, J. (1992). Protein folding in the cell. *Nature* 355, 33–45.
- Gilbert, H.F. (1995). Thiol/disulfide exchange equilibria and disulfide bond stability. *Methods Enzymol.* 251, 8–28.
- Grandbois, M., Beyer, M., Rief, M., Clausen-Schaumann, H., and Gaub, H.E. (1999). How strong is a covalent bond? *Science* 283, 1727–1730.
- Hoffstrom, B.G., Kaplan, A., Letso, R., Schmid, R.S., Turmel, G.J., Lo, D.C., and Stockwell, B.R. (2010). Inhibitors of protein disulfide isomerase suppress apoptosis induced by misfolded proteins. *Nat. Chem. Biol.* 6, 900–906.
- Holmgren, A. (1985). Thioredoxin. *Annu. Rev. Biochem.* 54, 237–271.
- Improta, S., Politou, A.S., and Pastore, A. (1996). Immunoglobulin-like modules from titin I-band: extensible components of muscle elasticity. *Structure* 4, 323–337.
- Ingolia, N.T., Lareau, L.F., and Weissman, J.S. (2011). Ribosome profiling of mouse embryonic stem cells reveals the complexity and dynamics of mammalian proteomes. *Cell* 147, 789–802.
- Jansens, A., van Duijn, E., and Braakman, I. (2002). Coordinated nonvectorial folding in a newly synthesized multidomain protein. *Science* 298, 2401–2403.
- Kadokura, H., and Beckwith, J. (2009). Detecting folding intermediates of a protein as it passes through the bacterial translocation channel. *Cell* 138, 1164–1173.
- Kadokura, H., Tian, H., Zander, T., Bardwell, J.C., and Beckwith, J. (2004). Snapshots of DsbA in action: detection of proteins in the process of oxidative folding. *Science* 303, 534–537.
- Kuwajima, K. (1989). The molten globule state as a clue for understanding the folding and cooperativity of globular-protein structure. *Proteins* 6, 87–103.
- Lappi, A.K., and Ruddock, L.W. (2011). Reexamination of the role of interplay between glutathione and protein disulfide isomerase. *J. Mol. Biol.* 409, 238–249.
- Lundström, J., and Holmgren, A. (1990). Protein disulfide-isomerase is a substrate for thioredoxin reductase and has thioredoxin-like activity. *J. Biol. Chem.* 265, 9114–9120.
- Lundström, J., Krause, G., and Holmgren, A. (1992). A Pro to His mutation in active site of thioredoxin increases its disulfide-isomerase activity 10-fold. New refolding systems for reduced or randomly oxidized ribonuclease. *J. Biol. Chem.* 267, 9047–9052.
- Lyles, M.M., and Gilbert, H.F. (1991). Catalysis of the oxidative folding of ribonuclease A by protein disulfide isomerase: dependence of the rate on the composition of the redox buffer. *Biochemistry* 30, 613–619.
- Martin, J.L. (1995). Thioredoxin—a fold for all reasons. *Structure* 3, 245–250.
- Molinari, M., and Helenius, A. (1999). Glycoproteins form mixed disulphides with oxidoreductases during folding in living cells. *Nature* 402, 90–93.
- Oberhauser, A.F., Marszalek, P.E., Carrion-Vazquez, M., and Fernandez, J.M. (1999). Single protein misfolding events captured by atomic force microscopy. *Nat. Struct. Biol.* 6, 1025–1028.
- Paxman, J.J., Borg, N.A., Horne, J., Thompson, P.E., Chin, Y., Sharma, P., Simpson, J.S., Wielens, J., Piek, S., Kahler, C.M., et al. (2009). The structure of the bacterial oxidoreductase enzyme DsbA in complex with a peptide reveals a basis for substrate specificity in the catalytic cycle of DsbA enzymes. *J. Biol. Chem.* 284, 17835–17845.
- Ptitsyn, O.B. (1995). Molten globule and protein folding. *Adv. Protein Chem.* 47, 83–229.
- Qin, J., Clore, G.M., Kennedy, W.M., Huth, J.R., and Gronenborn, A.M. (1995). Solution structure of human thioredoxin in a mixed disulfide intermediate complex with its target peptide from the transcription factor NF kappa B. *Structure* 3, 289–297.
- Rapoport, T.A., Jungnickel, B., and Kutay, U. (1996). Protein transport across the eukaryotic endoplasmic reticulum and bacterial inner membranes. *Annu. Rev. Biochem.* 65, 271–303.
- Ron, D., and Walter, P. (2007). Signal integration in the endoplasmic reticulum unfolded protein response. *Nat. Rev. Mol. Cell Biol.* 8, 519–529.
- Schröder, M., and Kaufman, R.J. (2005). The mammalian unfolded protein response. *Annu. Rev. Biochem.* 74, 739–789.
- Sevier, C.S., and Kaiser, C.A. (2002). Formation and transfer of disulphide bonds in living cells. *Nat. Rev. Mol. Cell Biol.* 3, 836–847.
- Tu, B.P., Ho-Schleyer, S.C., Travers, K.J., and Weissman, J.S. (2000). Biochemical basis of oxidative protein folding in the endoplasmic reticulum. *Science* 290, 1571–1574.
- Uehara, T., Nakamura, T., Yao, D., Shi, Z.Q., Gu, Z., Ma, Y., Masliah, E., Nomura, Y., and Lipton, S.A. (2006). S-nitrosylated protein-disulphide isomerase links protein misfolding to neurodegeneration. *Nature* 441, 513–517.
- Van den Berg, B., Clemons, W.M., Jr., Collinson, I., Modis, Y., Hartmann, E., Harrison, S.C., and Rapoport, T.A. (2004). X-ray structure of a protein-conducting channel. *Nature* 427, 36–44.
- Walker, K.W., and Gilbert, H.F. (1995). Oxidation of kinetically trapped thiols by protein disulfide isomerase. *Biochemistry* 34, 13642–13650.
- Walker, K.W., and Gilbert, H.F. (1997). Scanning and escape during protein-disulfide isomerase-assisted protein folding. *J. Biol. Chem.* 272, 8845–8848.
- Walker, K.W., Lyles, M.M., and Gilbert, H.F. (1996). Catalysis of oxidative protein folding by mutants of protein disulfide isomerase with a single active-site cysteine. *Biochemistry* 35, 1972–1980.
- Walter, P., Gilmore, R., and Blobel, G. (1984). Protein translocation across the endoplasmic reticulum. *Cell* 38, 5–8.
- Wedemeyer, W.J., Welker, E., Narayan, M., and Scheraga, H.A. (2000). Disulfide bonds and protein folding. *Biochemistry* 39, 4207–4216.
- Welker, E., Wedemeyer, W.J., Narayan, M., and Scheraga, H.A. (2001). Coupling of conformational folding and disulfide-bond reactions in oxidative folding of proteins. *Biochemistry* 40, 9059–9064.
- Wiita, A.P., Ainavarapu, S.R., Huang, H.H., and Fernandez, J.M. (2006). Force-dependent chemical kinetics of disulfide bond reduction observed with single-molecule techniques. *Proc. Natl. Acad. Sci. USA* 103, 7222–7227.
- Wiita, A.P., Perez-Jimenez, R., Walther, K.A., Gräter, F., Berne, B.J., Holmgren, A., Sanchez-Ruiz, J.M., and Fernandez, J.M. (2007). Probing the chemistry of thioredoxin catalysis with force. *Nature* 450, 124–127.
- Wilkinson, B., and Gilbert, H.F. (2004). Protein disulfide isomerase. *Biochim. Biophys. Acta* 1699, 35–44.
- Zapun, A., Creighton, T.E., Rowling, P.J., and Freedman, R.B. (1992). Folding in vitro of bovine pancreatic trypsin inhibitor in the presence of proteins of the endoplasmic reticulum. *Proteins* 14, 10–15.

EXTENDED EXPERIMENTAL PROCEDURES

Protein Expression and Purification

The (I27^{32C-75C})₈, (I27^{24C-55C})₈, and (I27 WT)₈ polyprotein substrates were prepared as previously described (Ainavarapu et al., 2007; Carrion-Vazquez et al., 1999; Wiita et al., 2007). Human TRX and TRX C35S were expressed and purified as described elsewhere (Perez-Jimenez et al., 2008; Ren et al., 1993). We used a version of TRX C35S that included a C-terminal RSCC sequence and an N-terminal His tag. Human PDI A1 was generated by PCR from Ultimate ORF Clone IOH9865 (Invitrogen) as an NdeI-BamHI fragment. The insert (amino acids 18–134 in the full-length PDI sequence) was cloned into NdeI-BamHI sites of pET-15b vector (Novagen) containing the N-terminal His6 tag and expressed in *Escherichia coli* BL21-Gold (DE3) (Stratagene). Cells with the construct were grown in LB media supplemented with 100 μg/ml ampicillin at 37°C for 10 hr (OD_{600nm} ~0.6). Protein expression was induced with 0.5 mM isopropyl β-D-thiogalactoside overnight. Cells were lysed by sonication, and recombinant PDI A1 was first purified from cell lysate using Ni Sepharose 6 Fast Flow (GE Life Sciences) and then using gel filtration Superdex 300 column. The fractions containing PDI A1 were pooled together and dialyzed into 20 mM sodium phosphate buffer (pH 8) and verified by SDS-PAGE. Protein concentration was determined from absorbance measurements at 280 nm ($\epsilon = 20,000 \text{ M}^{-1} \text{ cm}^{-1}$). Full-length bovine PDI was obtained from Sigma (catalog no. P3818).

Preparation of Reduced and Oxidized Enzymes

Before beginning an experiment, the enzyme sample was incubated with reduced DTT for 20 min at room temperature, to ensure that the enzyme was fully reduced. 200 μM DTT was used with TRX (10 μM) and TRX C35S (40 μM). 500 μM DTT was used with PDI A1 (120 μM). The low concentrations of DTT used in our experiments did not cause any measurable redox activity on the substrate (see below, and also Figures S2A and S4D). For the PDI experiments shown in Figures 3, 4, 5, and 6, a solution was used containing reduced PDI A1 and no DTT.

Reduced PDI A1 was prepared by first reducing the enzyme in 10 mM reduced DTT for 10 min, and then removing the DTT in a Superdex 200 FPLC size exclusion column. The purified reduced enzyme was snap-frozen and stored under argon at –80°C until just before the experiment. Oxidized PDI A1 was prepared in a similar manner but was incubated with 10 mM GSSG (instead of DTT) prior to purification.

The presence of DTT did not affect our measurements, for the following reasons. The reaction rate of DTT with the exposed I27³²⁻⁷⁵ disulfide has been measured previously (Wiita et al., 2006), and at a concentration of 0.5 mM and a stretching force of 75 pN the reduction rate by DTT is estimated to be 0.008 s⁻¹. The trace in Figure S2A is consistent with this result, as no reduction steps are seen during the first 10 s in the presence of 0.5 mM DTT. By comparison, the rate of PDI-mediated reduction under the same conditions was measured to be 0.35 ± 0.04 s⁻¹. Consequently, in our experiments less than 2% of reduction steps can be attributed to DTT. The experiments in Figures 1 and 2 were performed in the presence of DTT. All PDI data in Figures 3, 4, 5, and 6 were obtained in the absence of DTT. Within each of these experiments, selected data points were tested using DTT, however no difference was found between the results with and without DTT (Figure S4D). The presence of DTT therefore did not interfere significantly with our enzymatic measurements.

We noticed that the proportion of 25 nm steps in the *probe* pulse started decreasing a few hours after adding reduced PDIs to the experimental setup (the same was not true for TRX, where only 25 nm steps were seen in the *probe* pulse for up to 24 hr after the experiment was started). Although the nature of this change remains unknown, it does not appear to involve the N-terminal cysteine in the PDI active site, as the rate of initial attack (mixed disulfide formation) did not change significantly throughout the same experiments. The change may therefore represent some conformational change of the enzyme, or a chemical modification affecting the C-terminal active site cysteine. In order to avoid any bias arising from this effect, we limited our recordings to the first ~4 hr after experimental setup, and made sure that all data points contained recordings that were acquired over this entire time span.

Single-Molecule AFM

Double-pulse experiments were used when acquiring the data shown in Figures 3, 4, and 5 (except for control experiments without enzymes), as well as in Figures S4B–S4D, S6A, and S6B. In these experiments, each *denature* and *probe* pulse consisted of two parts: a short pulse at a high force (150–170 pN for ~1 s) that ensured unfolding and exposure of all substrate disulfide bonds, followed by a longer pulse at low force (75 pN for a variable time). The double-pulse experiments gave results consistent with the single-pulse experiments, but were advantageous for two reasons. First, the rate of attack by PDI on the reduced disulfide bond was higher at lower forces, in a similar manner to TRX (Wiita et al., 2007). Therefore, mixed disulfide complexes could be established in a shorter amount of time by using a rapid unfolding pulse followed by a reduction pulse at lower force. This double-pulse protocol enabled a more accurate determination of the PDI off-rate, and made the effects of the PDI off-rate less significant when acquiring the data for, e.g., Figure 3C. Furthermore, the double-pulse protocol proved more practical as it lowered the frequency of early detachment. Since the protein is not covalently bound to the AFM tip, the rate of spontaneous detachment is highly dependent on the applied force. By maintaining the force at 75 pN for most of each recording, we were therefore able to acquire more useful traces in a given amount of time.

Mechanical Unfolding and Reduction of I27 Domains

A natively folded I27 domain can resist piconewton-range stretching forces applied to its termini for a stochastic amount of time. The rate of thermally driven unfolding increases exponentially with the applied force (Schlierf et al., 2004). The extension upon unfolding occurs in a single step because once the native fold is broken, the protein domain unravels in a downhill manner. If a disulfide is present, the domain unravels up until the disulfide in a single step. Importantly, the step size reveals how many residues have unraveled. For instance, an 11 nm step provides a signature of a natively folded and oxidized I27³²⁻⁷⁵ protein (Alegre-Cebollada et al., 2011). Similarly, once a disulfide is broken under force, the sequestered part of the protein unravels rapidly under the applied force. The expected step size upon cleavage of a disulfide was calculated as:

$$\Delta L = k_{WLC} \times (L_{AA} \times N_{AA} - L_{SS}) \quad (S1)$$

where k_{WLC} is the worm-like chain fractional extension, $L_{AA} = 0.4$ nm is the contour length of an amino acid, N_{AA} is the number of amino acids between the two cysteines and $L_{SS} = 0.8$ nm is the contour length of a cystine (Alegre-Cebollada et al., 2011).

Data Analysis

We relied on a unique fingerprint to verify the correct attachment of our substrate molecule. In the case of experiments with PDI and I27³²⁻⁷⁵, the fingerprint consisted of repeated 11 and 14 nm steps in the *denature* pulse. These steps are a signature of a folded and oxidized I27³²⁻⁷⁵ polyprotein. Since the focus of our study was on reactions occurring in the substrate subsequent to the initial unfolding, we were not biased by selecting for a particular initial state (although there may be many different initial states of our molecule of interest, we were only interested in probing the time evolution of a particular state, i.e., the natively folded and oxidized state). To ensure that the same protein was stretched in the two pulses, only traces showing equal extension at the end of the *probe* pulse and the end of the *denature* pulse were included in the analysis. An equal extension at the end of the two pulses is a strong indication of the same molecule being probed for the following reason. Since the attachment point on the substrate molecule varies between recordings, the standard deviation in full extension between different molecules was 44 nm. Considering our spatial resolution of < 1 nm, the probability of two randomly picked up molecules displaying the same length is therefore at most 0.9%. Moreover, the successful pick-up rate is ~1%. Therefore, the probability of picking up a new molecule with the same length on a given attempt is approximately 1 in 10,000. By comparison, nearly every trace that lasted more than 5 s after the onset of the *probe* pulse reached the same level as at the end of the *denature* pulse, thus verifying attachment of the same molecule.

Our accuracy in measuring a step size depends on multiple factors, including the signal stability during the experiment and the dwell time between subsequent steps. For instance, although the last step in the trace shown in Figure 5A is clearly separated from neighboring steps, and thus allows for ample averaging that yielded the 0.1 nm precision reported in the figure, the previous step in the same trace occurred immediately after a downward step induced by a shift in the applied force. Such close proximity of events precludes accurate determination of the extension level before the step, and thus severely limits our resolution. From the isolated cases where the resolution was adequate, we could clearly distinguish between populations of different step sizes (as shown in Figure 5A). On the other hand, when compiling histograms like those shown in 6B, in order to be consistent we could not use extensive averaging. We therefore only averaged 20 data points before and after each step when compiling histograms. An advantage of this approach is that the measurement error, at 0.7 nm, although relatively large, is consistent across data sets. Each peak can thus be fitted accurately with an individual Gaussian with a fixed width of 0.7 nm. We used the Multipeak algorithm in Igor Pro (Wavemetrics) for optimization of the fits.

When determining the PDIA on-rate (Figures S6C and S6D), only traces containing at least two steps with the same magnitude (>5 nm) were selected. Step detection was automated using a custom-made algorithm. The data were then binned in both dimensions, and the oxidized fraction of the substrate was measured as relative proportion of ~11 nm steps to the total number of steps within a time-bin. This proportion was then fit with a single-exponential curve, from which the oxidation rate was determined. The on-rate of PDIA is equal to the oxidation rate of the substrate under the assumption that the interconversion of mixed disulfides into substrate disulfides is much faster than the competing PDI off-rate ($k_{ox} \gg k_{off}$). This assumption appears well motivated since disulfide formation was faster than the off-rate even when substrate folding rate-limited disulfide formation (0.35 s^{-1} versus 0.1 s^{-1} , see Figures 3 and 4). In Figures S6C and S6D, the substrate was already folded, and we therefore expect the disulfide formation rate to be even higher, thus justifying our assumption. The on-rate for an unstructured substrate may be higher than our measured value; however, this would require further investigation to ascertain.

SUPPLEMENTAL REFERENCES

Ainavarapu, S.R., Brujic, J., Huang, H.H., Wiita, A.P., Lu, H., Li, L., Walther, K.A., Carrion-Vazquez, M., Li, H., and Fernandez, J.M. (2007). Contour length and refolding rate of a small protein controlled by engineered disulfide bonds. *Biophys. J.* 92, 225–233.

Alegre-Cebollada, J., Kosuri, P., Rivas-Pardo, J.A., and Fernández, J.M. (2011). Direct observation of disulfide isomerization in a single protein. *Nat. Chem.* 3, 882–887.

Carrion-Vazquez, M., Oberhauser, A.F., Fowler, S.B., Marszalek, P.E., Broedel, S.E., Clarke, J., and Fernandez, J.M. (1999). Mechanical and chemical unfolding of a single protein: a comparison. *Proc. Natl. Acad. Sci. USA* 96, 3694–3699.

- Darby, N.J., and Creighton, T.E. (1995a). Characterization of the active site cysteine residues of the thioredoxin-like domains of protein disulfide isomerase. *Biochemistry* 34, 16770–16780.
- Perez-Jimenez, R., Wiita, A.P., Rodriguez-Larrea, D., Kosuri, P., Gavira, J.A., Sanchez-Ruiz, J.M., and Fernandez, J.M. (2008). Force-clamp spectroscopy detects residue co-evolution in enzyme catalysis. *J. Biol. Chem.* 283, 27121–27129.
- Ren, X., Björnstedt, M., Shen, B., Ericson, M.L., and Holmgren, A. (1993). Mutagenesis of structural half-cystine residues in human thioredoxin and effects on the regulation of activity by selenodiglutathione. *Biochemistry* 32, 9701–9708.
- Schlierf, M., Li, H., and Fernandez, J.M. (2004). The unfolding kinetics of ubiquitin captured with single-molecule force-clamp techniques. *Proc. Natl. Acad. Sci. USA* 101, 7299–7304.
- Walker, K.W., and Gilbert, H.F. (1995). Oxidation of kinetically trapped thiols by protein disulfide isomerase. *Biochemistry* 34, 13642–13650.
- Walker, K.W., Lyles, M.M., and Gilbert, H.F. (1996). Catalysis of oxidative protein folding by mutants of protein disulfide isomerase with a single active-site cysteine. *Biochemistry* 35, 1972–1980.
- Wiita, A.P., Ainarapu, S.R., Huang, H.H., and Fernandez, J.M. (2006). Force-dependent chemical kinetics of disulfide bond reduction observed with single-molecule techniques. *Proc. Natl. Acad. Sci. USA* 103, 7222–7227.
- Wiita, A.P., Perez-Jimenez, R., Walther, K.A., Gräter, F., Berne, B.J., Holmgren, A., Sanchez-Ruiz, J.M., and Fernandez, J.M. (2007). Probing the chemistry of thioredoxin catalysis with force. *Nature* 450, 124–127.



Figure S1. Related to Figure 1

(A) Thiol-disulfide oxidoreductases in the TRX superfamily, including PDI, catalyze disulfide exchange by forming a mixed disulfide complex with the substrate (blue). Of the two cysteines in the active site, only the N-terminal cysteine engages in a mixed disulfide with substrates during normal activity (Walker et al., 1996). The mixed disulfide can be formed in one of two ways. In the first pathway, an oxidized enzyme reacts with a free substrate cysteine (this is the case in Figures 6, S6C, S6D, S7A, and S7B). This pathway is thought to be predominant during oxidative folding in vivo. In the second pathway, the mixed disulfide is formed through reaction of reduced PDI with a disulfide in the substrate (this is the case in Figures 1, 2, 3, 4, 5, S2, S3, S4, S5, and S7C). This reaction cleaves the substrate disulfide and is therefore necessary for reduction or isomerization activity. Regardless of the pathway, the resulting mixed disulfide complex is identical. (B) Pathway for PDI-mediated disulfide oxidation.

(C) Pathway for PDI-mediated disulfide reduction. All depicted cysteines exist in equilibrium between the protonated (SH) and deprotonated, reactive form (S⁻). For simplicity, only the sulfur is shown.

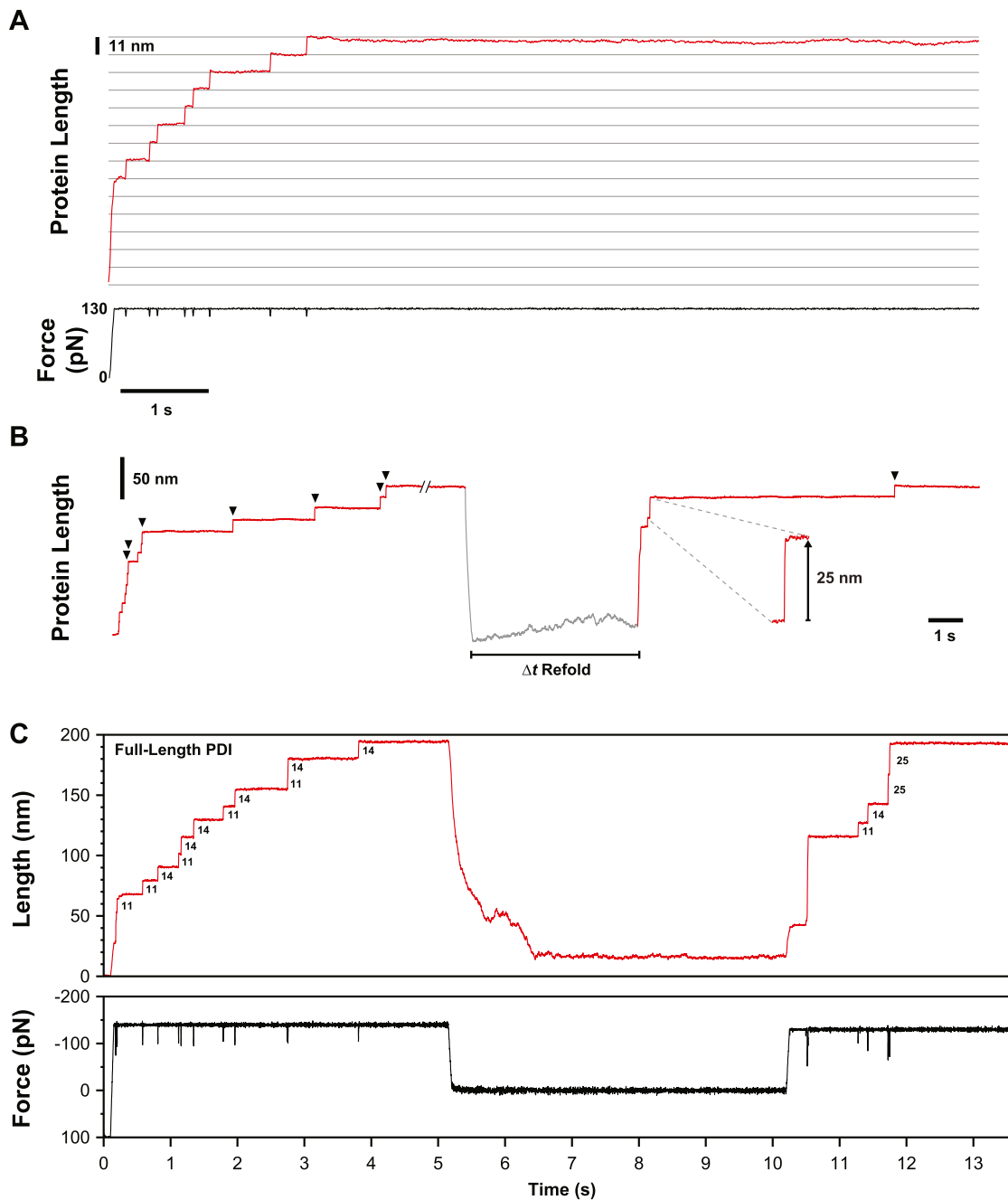


Figure S2. Related to Figure 1

(A) The $(I27^{32-75})_8$ polyprotein unfolded in the presence of only 500 μM DTT displayed 11 nm unfolding steps but did not show any 14 nm reduction steps.

(B) Oxidative folding with PDIa shows 25 nm steps in the *probe* pulse, evidence of incomplete oxidative folding. Arrowheads indicate 14 nm steps each resulting from the formation of a mixed disulfide complex.

(C) Oxidative folding of $(I27^{32-75})_8$ catalyzed by full-length PDI.

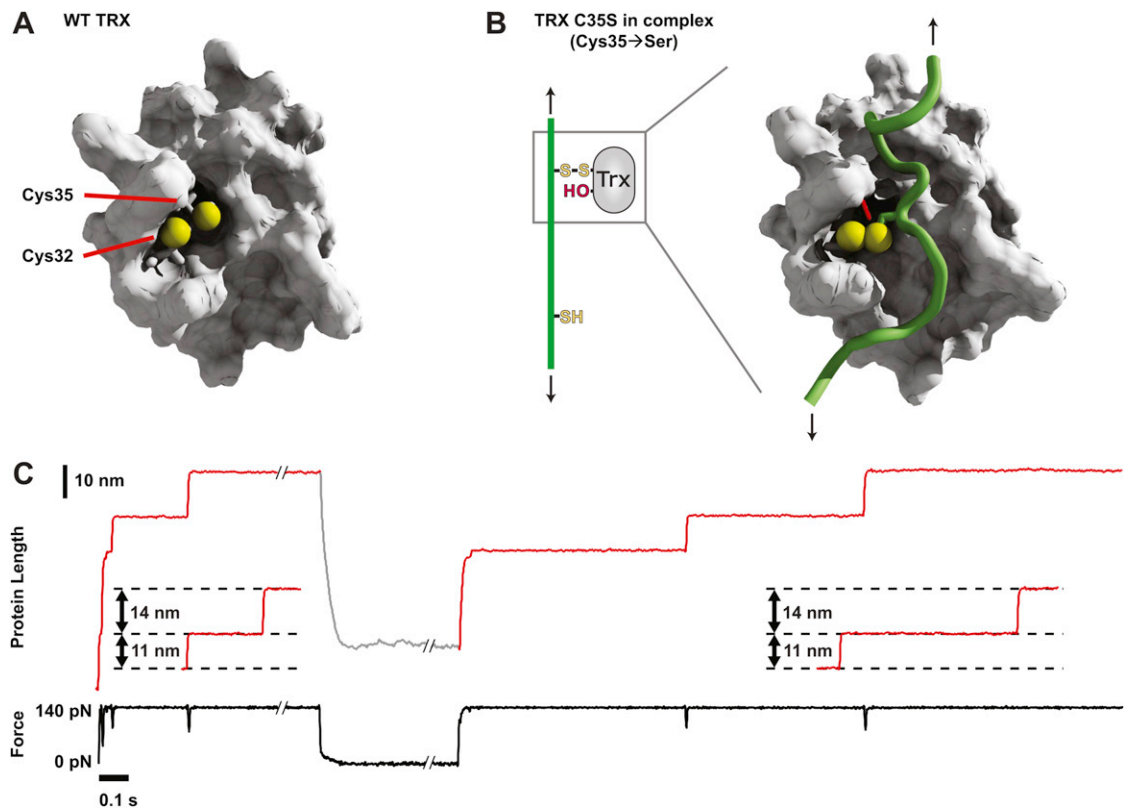


Figure S3. Related to Figure 2

(A) Structure of reduced human TRX, showing the two redox active sulfur atoms in the active site. The sulfur in Cys32 carries out the initial nucleophilic attack that cleaves the substrate disulfide and forms a mixed disulfide (PDB: 1TRS).

(B) Single cysteine mutant of human TRX in a mixed disulfide complex with a substrate fragment (Cys35Ala; PDB: 1MDI). In our experiments, an equivalent Cys35Ser mutation removed the mechanism for spontaneous release from the mixed disulfide complex.

(C) Experimental trace showing a single I27³²⁻⁷⁵ domain being unfolded (11 nm step) and forming a mixed disulfide with TRX C35S (14 nm step). After refolding, the 32-75 disulfide has been reformed, showing that the distal cysteine in the substrate was able to attack the mixed disulfide, despite the steric hindrance imposed by the enzyme. TRX C35S exhibited disulfide formation kinetics similar to that of PDIa (Figure 4C), suggesting that the catalytic mechanisms of TRX and PDI are highly similar. It thus appears that WT TRX is intrinsically capable of catalyzing disulfide formation but releases rapidly from the mixed disulfide complex and therefore prevents disulfide formation in folding proteins.

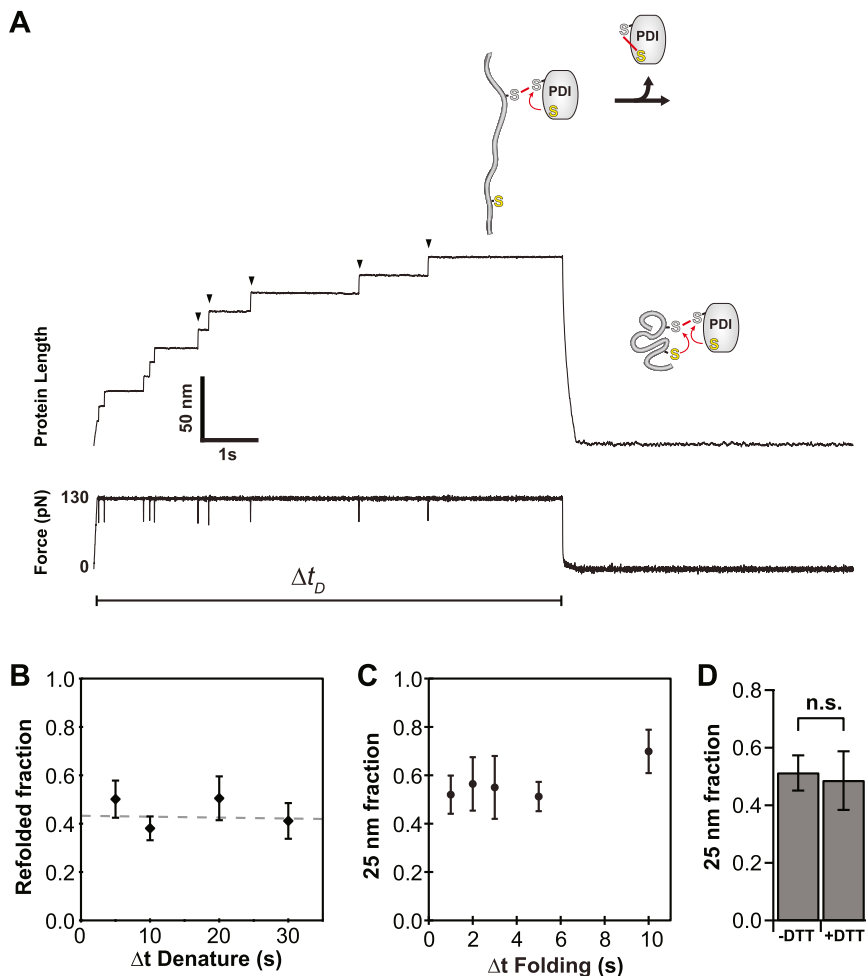


Figure S4. Related to Figure 3

(A) The kinetic competition during refolding of I27-PDIa mixed disulfide complexes. When the substrate is extended (during ongoing translocation, or during application of mechanical force in Δt_D), no substrate cysteine is available to cleave the mixed disulfide bond. However, the enzyme can release spontaneously through nucleophilic attack by the C-terminal cysteine in its active site. As soon as the substrate collapses (upon completion of translocation or removal of the mechanical force), its free cysteines become available. The outcome of oxidative folding now depends on the competing reaction kinetics of the cysteines in the enzyme and those in the substrate.

(B) Refolding of (I27³²⁻⁷⁵)₈ from mixed disulfide complexes with PDIa (measured as the sum of 11 and 25 nm steps in the *probe* pulse divided by the number of 11 nm steps in the *denature* pulse), for different durations of the *denature* pulse Δt_D (fraction \pm SEM). The folding time Δt_F was set to 5 s. There is no apparent dependence of the amount of folding on Δt_D , showing that folding is not noticeably altered by the presence of the mixed disulfide complex. The rates k_{ox} and k_{fold} shown in Figure 3D are therefore approximately equal. ($n > 50$).

(C) Fraction of 25 nm steps in *probe* pulse of PDIa experiments, as a function of refolding time (fraction \pm SEM). ($n > 50$). Length of the *denature* pulse $\Delta t_D = 5$ s. If some mixed disulfides were still present after the substrate had acquired a native fold, then the prevalence of 25 nm steps would be expected to decrease at long refolding times (as the domains became oxidized). The figure shows that the fraction of 25 nm steps does not decrease as a function of refolding time, indicating that the enzyme had released from all of these domains prior to folding.

(D) The presence of 500 μ M DTT did not affect the 25 nm fraction in the *probe* pulse (fraction \pm SEM) $\Delta t_D = 5$ s, $\Delta t_F = 5$ s ($n > 50$; $p < 0.01$).

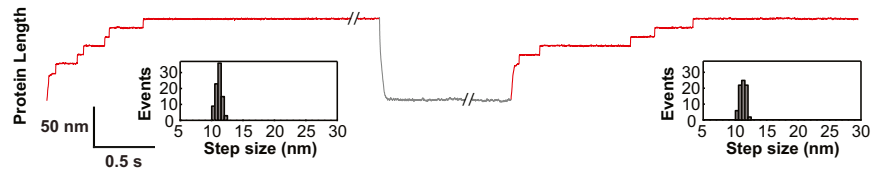


Figure S5. Related to Figure 4

Refolding of oxidized I27³²⁻⁷⁵ in the absence of enzymes. No cleavage of disulfide bonds was observed in these experiments (n = 86).

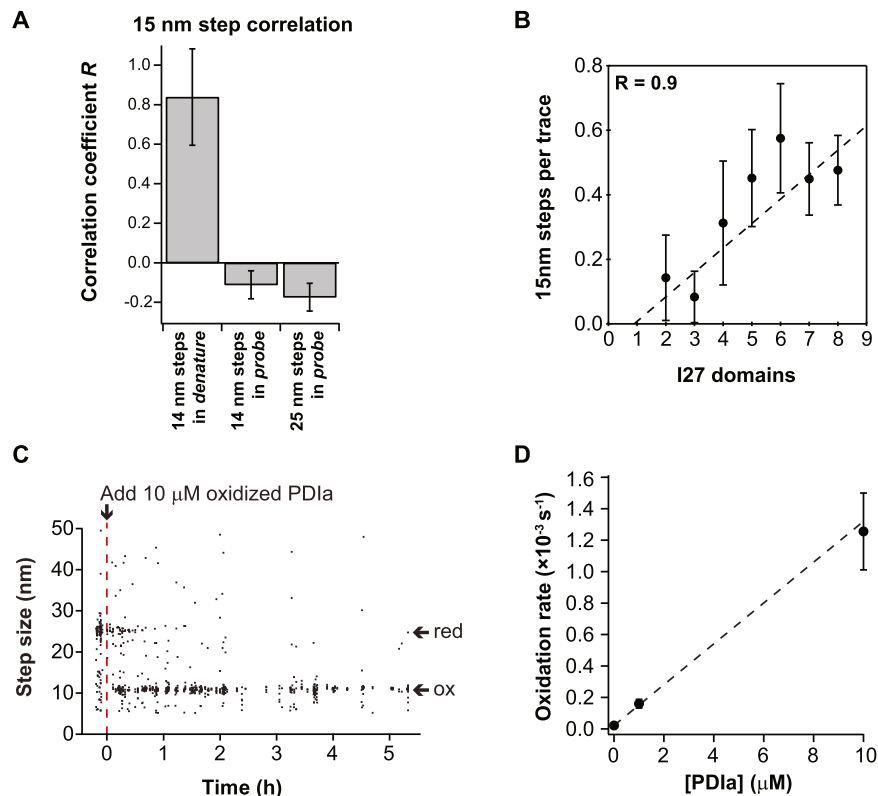


Figure S6. Related to Figure 5

(A) Formation of interdomain disulfide bonds prevents native folding of $(I27^{32-75})_8$. The appearance of 15 nm steps in the *probe* pulse correlated strongly with the number of mixed disulfide complexes that were formed in the *denature* pulse, and therefore with the number of domains in the substrate. The appearance of these steps was negatively correlated both with native disulfide formation and native protein folding. Folding is thus hindered by the formation of interdomain disulfides. (Data represent correlation coefficient \pm SEM.)

(B) Formation of inter-domain disulfide bonds (fraction \pm SEM) was linearly dependent on the number of unfolded $I27^{32-75}$ domains (as measured from the number of 11 and 14 nm steps in the *denature* pulse) ($n > 10$). The x-intercept of the fit was close to 1, thus verifying that at least two domains are required to form an interdomain disulfide. All of these results support the interpretation of anomalous steps in the *probe* pulse as arising from the cleavage of interdomain disulfides.

(C) PD1a-mediated oxidation of folded, pre-reduced $I27^{32-75}$. Substrate molecules were randomly picked up and unfolded in the AFM. The observed step sizes in the *denature* pulse are displayed as points as a function of the time elapsed since the addition of PD1a. 25 nm steps indicate reduced domains while 11 nm steps indicate oxidized domains ($n = 759$).

(D) The experiment in (C) was repeated with different concentrations of oxidized PDI. The data were binned and fit to single exponentials yielding the rate of PD1a-catalyzed oxidation of a folded $I27^{32-75}$ protein, $0.13 \text{ mM}^{-1} \text{ s}^{-1}$ (mean \pm SEM; $n > 500$). This value is equal to the PD1a on-rate under the assumption that $k_{ox} \gg k_{off}$. Our measurements are consistent with previous studies of the PDI on-rate, where it was found to be in the range of $0.1\text{--}0.6 \text{ mM}^{-1} \text{ s}^{-1}$ (Darby and Creighton, 1995a; Walker and Gilbert, 1995).

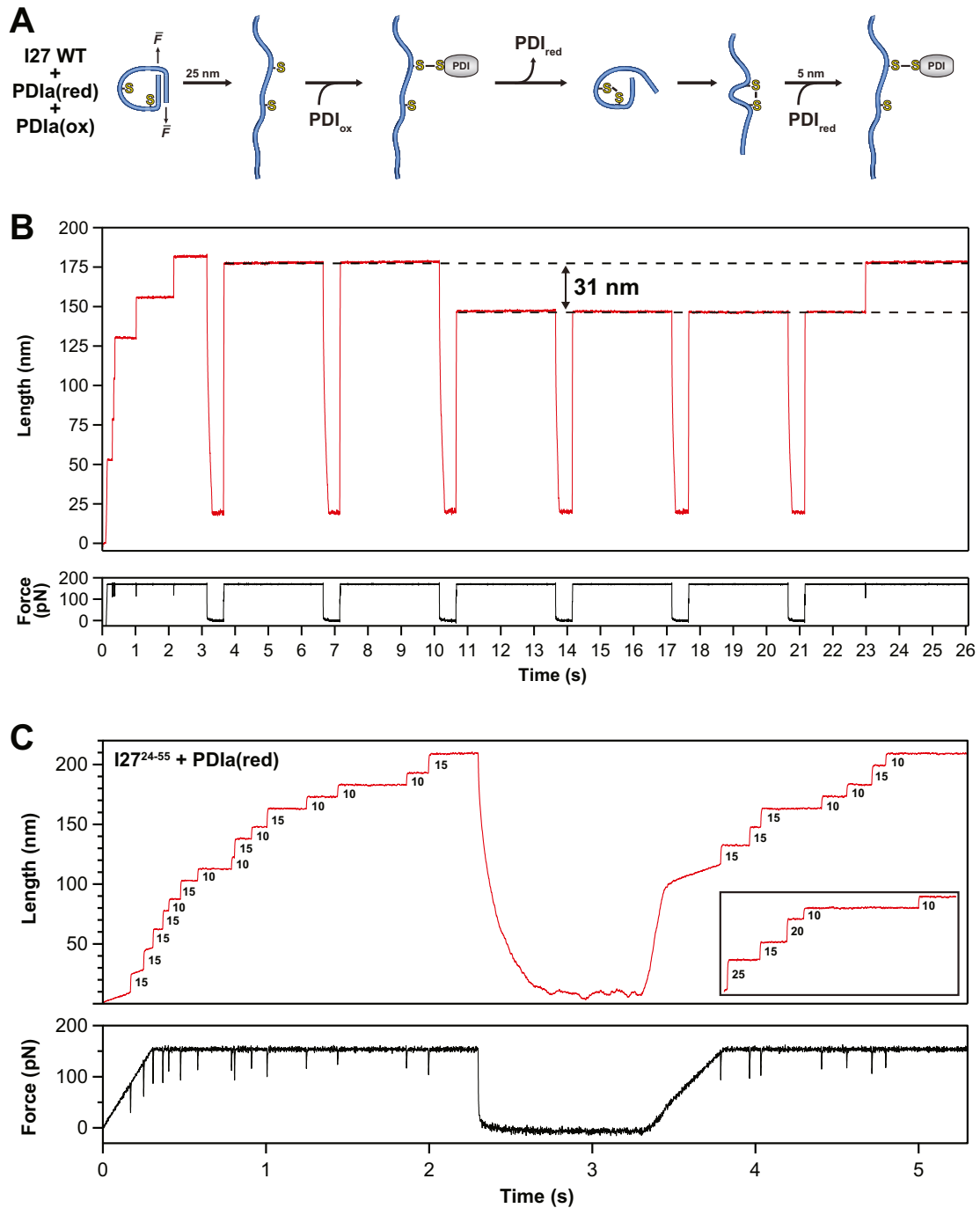


Figure S7. Related to Figure 6

(A) Diagram illustrating the refolding of I27 WT in solution containing reduced and oxidized PDla (as in Figure 6E). Shortenings corresponding to the formation of intradomain disulfides by oxidized PDla were occasionally seen. These disulfides could then be cleaved by reduced PDla, as seen from the 5 nm extension steps indicated by arrows in the experimental recording in Figure 6E.

(B) Refolding of I27 WT in solution containing 60 μ M reduced PDla and 60 μ M oxidized PDla, showing a 31 nm shortening followed by a 31 nm step, indicating the formation and cleavage of an interdomain disulfide. A 5 nm shortening (corresponding to formation of an intradomain disulfide) can be seen after the first refolding interval.

(C) Oxidative folding of (I27²⁴⁻⁵⁵)₂ in the presence of reduced PDla, under conditions equivalent to those in Figure 1G. Unfolding of an oxidized I27²⁴⁻⁵⁵ domain leads to a 15 nm step, whereas reduction of an intradomain disulfide leads to a 10 nm step. Inset: 25 nm steps were observed in the probe pulse, corresponding to reduced folded domains resulting from instances of spontaneous PDla dissociation. Sporadic 20 nm steps coincided with the predicted length for cleavage of an interdomain disulfide; see Figure 5E.

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

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 (ΔH) 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 , ΔH is positive, the unfolding process is endothermic and, hence, it is driven by a temperature increase (heat denaturation). Below T_H , ΔH 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-

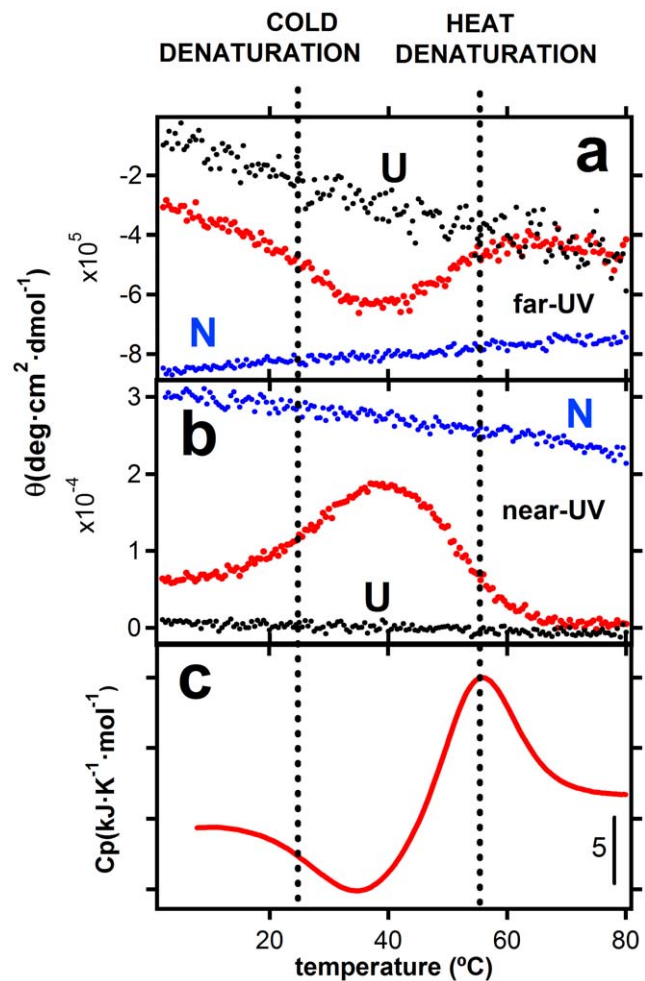


Figure 1. Heat and cold denaturation of thioredoxin at pH 7 in the presence of 2M guanidine, as followed by, a) far-UV circular dichroism (222 nm), b) near-UV circular dichroism (280 nm) and, c) differential scanning calorimetry. The three experiments (far-UV, near-UV and DSC) were carried out with the 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

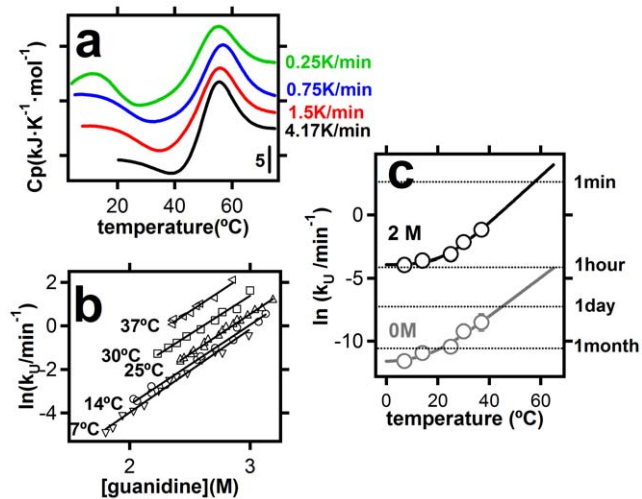


Figure 2. Cold denaturation and folding-unfolding kinetics. a) Scan-rate effect on the DSC profiles for thioredoxin heat and cold denaturation at pH 7 and in the presence of 2M guanidine. All the scanning experiments started at 2°C. The differences observed in the lowest temperature are related to the 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

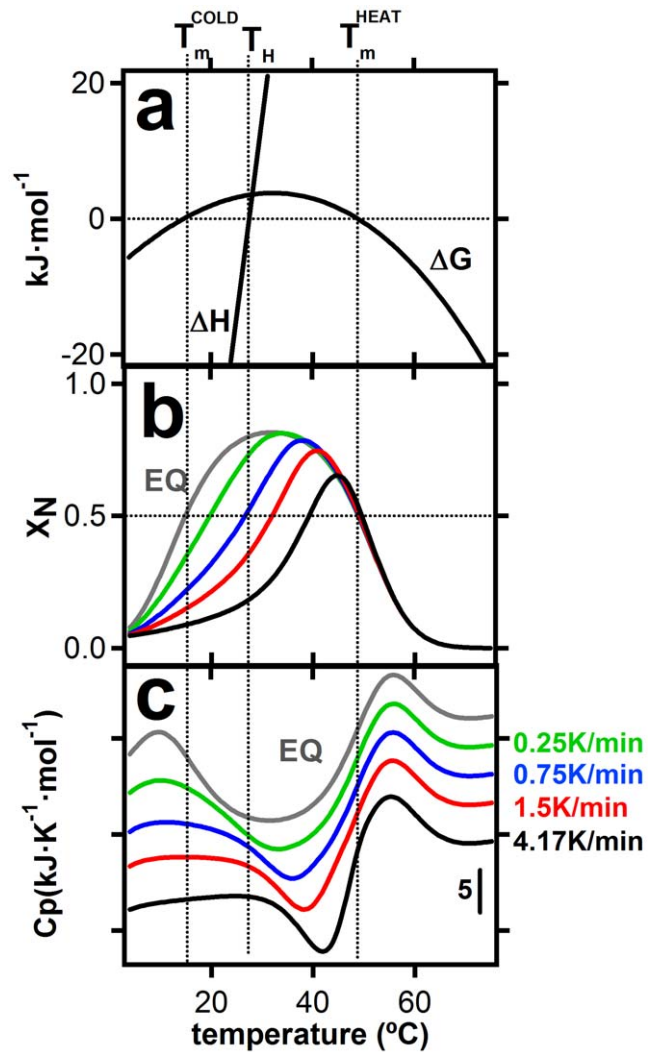


Figure 3. Theoretical simulation of temperature-scanning profiles for protein heat-cold denaturation including slow folding-unfolding kinetics. a) Folding-unfolding energetics used in the simulation (see Materials and Methods for details). Profiles of unfolding free energy and unfolding enthalpy versus 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 thiothyretin 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 [θ] 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

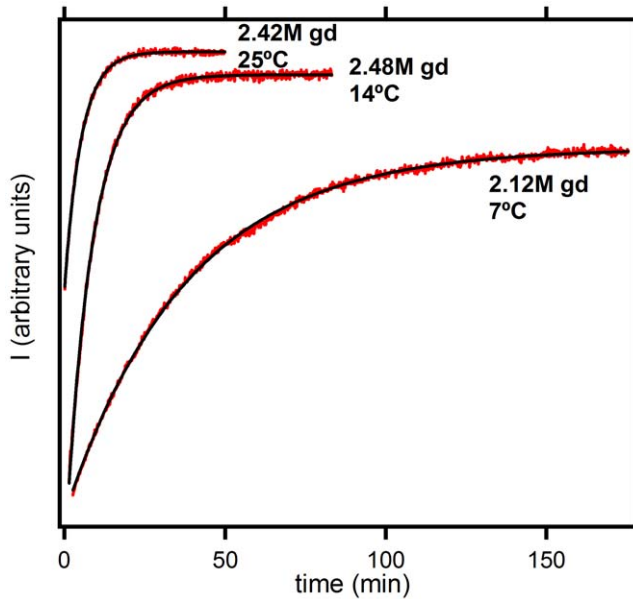


Figure 4. Representative examples of fluorescence intensity versus time profiles for thioredoxin unfolding. The numbers alongside the profiles stand for the temperature and the guanidine concentration used. Experimental data are shown in red and the black line represents the best fit of a single-exponential.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

linear dependence can be written as,

$$\ln k_U = \ln k_U^0 - \frac{m_U}{RT} C \quad (1)$$

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:

$$\ln k_u = \ln \{ \alpha + \beta \cdot e^{\gamma T} \} \quad (2)$$

where α , β and γ 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,

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

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 τ is the folding-unfolding relaxation time,

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

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 ,

$$X_N^{EQ} = \frac{1}{1 + K} \quad (5)$$

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,

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

where C_P^{EX} is an excess heat capacity defined with respect to the native heat capacity level, and ΔC_P and ΔH 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 (ΔH) and heat capacity (ΔC_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 , ΔH and ΔC_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 ΔS and ΔG 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), ΔH_R , ΔS_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 (ΔT 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⁻¹·mol⁻¹,
 $d\Delta C_P/dT = 0.1$ kJ·K⁻²·mol⁻¹.

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 ΔC_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 2H₂O stabilization of a staphylococcal nuclease mutant. *Proc Natl Acad Sci U S A* 88: 7715–7718.
- Ibarra-Molero B, Makhatazde GI, Sanchez-Ruiz JM (1999) Cold denaturation of ubiquitin. *Biochim Biophys Acta* 1429: 384–390.
- Yang WY, Gruebele M (2005) Kinetic equivalence of the heat and cold structural transitions of lambda_{da6-85}. *Philos 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 fluorine 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 YV, Venyamov SY, 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 YV, 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, Makhatazde 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, Hilser 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.
- Whitten ST, Kurtz AJ, Pometum MS, Wand AJ, Hilser 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. Sadqi 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 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, Bystroff 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 Ø, 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. Connelly 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. Rodríguez-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 Maria 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, Makhatadze 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.

ÁLVARO INGLÉS PRIETO

c/ Amberes 19 3B 10005 Cáceres Spain · +34-669538166 · ingles.prieto@gmail.com

CURRICULUM VITAE

PERSONAL INFORMATION:

Name: Álvaro Inglés Prieto
Date of birth: 11/02/1984
Address: C/ Amberes 19 3B. Cáceres. 10005. Spain
Telephone: +34 669538166
E-mail: ingles.prieto@gmail.com



EDUCATION:

PhD Degree (Chemistry)

Place: University of Granada.
Starting date: January 2008.
Estimated Finishing date: December 2012
Academic record marks: -

Advanced Studies Diploma (Master's Degree)

Place: University of Seville.
Starting date: January 2008.
Finishing date: May 2009
Academic record marks: **Excellent**

Bachelor's Degree in Biochemistry

Place: University of Salamanca.
Starting date: October 2002.
Finishing date: June 2007
Academic record marks: Excellent with honors=4, Excellent=17, Outstanding=19, Pass=4.
Academic average grade: **2.54 (High Outstanding)**

WORKING EXPERIENCE. PARTICIPATION IN RESEARCH PROJECTS

1. Project name: Protein engineering and evolution
Place: Department of Physical Chemistry
University of Granada
Duration: October 2008 – December 2012
Director: Dr. José Manuel Sánchez Ruiz (sanchezr@ugr.es).

ÁLVARO INGLÉS PRIETO

c/ Amberes 19 3B 10005 Cáceres Spain · +34-669538166 · ingles.prieto@gmail.com

2. Project name: From protein structure and dynamics to tailored enzymes (CSD2009-00088)
Place: Department of Physical Chemistry
University of Granada
Duration: October 2008 – December 2012
Director: Dr. José Manuel Sánchez Ruiz (sanchezr@ugr.es), Victor Muñoz, José Luis Lopez Lacomba, Eva de Alba, Antonio Romero, Rafael Giraldo-Suárez.
3. Project name: Protein engineering (BIO2006-07332)
Place: Department of Physical Chemistry
University of Granada
Duration: October 2008 – December 2012
Director: Dr. José Manuel Sánchez Ruiz (sanchezr@ugr.es).
4. Project name: Initiation, elongation, blocking, termination and topology of replication (BFU2008-00408).
Place: Centro de Investigaciones Biológicas
University Complutense - CSIC
Duration: September 2007 – September 2008
Director: Dr. Jorge Bernardo Schwartzman Blinder (schwartzman@cib.csic.es)
5. Project name: Purification and Characterization of the Calx-beta domain of integrin Beta 4 (BFU2006-01929)
Place: Cancer Research Centre
University of Salamanca - CSIC
Duration: October 2005 – July 2007
Director: Dr. José María de Pereda Vega (pereda@usal.es)
6. Project name: Study of the molecular mechanisms involved in the regulation of the activation of STAT6 by IL4.
Place: Research Unit
San Pedro de Alcántara Hospital (Cáceres)
Duration: 9 non-consecutive months (summers of 2003, 2004, 2005)
Director: Dr. José Zamorano Quirantes (uinvcc@ses.juntaextremadura.net)

LABORATORY AND COMPUTER SKILLS

- **Molecular biology techniques** mostly working with recombinant DNA (cloning, site-directed mutagenesis, plasmid isolation, polyprotein synthesis for AFM, etc.).
- **Expression and purification of recombinant proteins** using different chromatographic techniques such as size-exclusion chromatography, affinity chromatography, ionic exchange, etc.
- **Physico-chemical characterization** of proteins including Differential Scanning Calorimetry, Fluorescence and circular dichroism spectroscopies.
- **Protein crystallization** using sitting drop and counter-diffusion techniques. Basic knowledge of X-ray data collection, analysis and structural determination using molecular replacement method.
- **Setting up activity assays** for diverse enzymes.

ÁLVARO INGLÉS PRIETO

c/ Amberes 19 3B 10005 Cáceres Spain · +34-669538166 · ingles.prieto@gmail.com

- **Single molecule studies** of enzyme catalysis using Atomic Force Microscopy.
- **Immunoprecipitation/western blots assays.**
- Working knowledge of **scientific software**, such as Igor Pro, Origin, Pymol, CCP4, Phenix, EndNote X4, Unscrambler X10, etc.
- **Complex analysis** of experimental data, including Partial Least Square Regression, Bootstrapping, etc.

PUBLICATIONS

1.- Kosuri, Pallav; Alegre-Cebollada, Jorge; Feng, Jason; Kaplan, Anna; **Inglés-Prieto, Álvaro**; Badilla, Carmen L.; Stockwell, Brent R.; Sánchez- Ruiz, José M.; Holmgren, Arne; Fernández, Julio M. *Protein folding drives disulfide formation*. **Cell**. 151 - 4, pp. 794 - 806. 2012. ISSN 1097-4172

Type of production: Article

2.- Perez-Jimenez, Raul; **Inglés-Prieto, Álvaro**; Zhao, Zi-ming; Sánchez-Romero, Inmaculada; Alegre-cebollada, Jorge; Kosuri, Pallav; Garcia-manyes, Sergi; Kappock, T Joseph; Tanokura, Masaru; Holmgren, Arne; Sanchez-Ruiz, Jose Manuel; Gaucher, Eric A.; Fernandez, Julio M. *Single-molecule paleoenzymology probes the chemistry of resurrected enzymes*. **Nature Structural and Molecular Biology**. 18 - 5, pp. 592 - 596. 2011. ISSN 1545-9993

Type of production: Article **Index measuring impact:** 13.685

3.- Romero-Romero, Maria Luisa; **Inglés-Prieto, Álvaro**; Ibarra-Molero, Beatriz; Sanchez-Ruiz, Jose Manuel. *Highly Anomalous Energetics of Protein Cold Denaturation Linked to Folding-Unfolding Kinetics*. **PLoS One**. 6 - 7, 2011. ISSN 1932-6203

Type of production: Article **Index measuring impact:** 4.411

4.- Alonso-Garcia, Noelia; **Inglés-Prieto, Álvaro**; Sonnenberg, Arnaud; De Pereda, Jose Maria. *Structure of the calx-b domain of the integrin b4: insights into function and cation-independent stability*. **Acta crystallogr. D Biol. crystallogr.** 65 - 8, pp. 858 - 871. 2009. ISSN 0907-4449

Type of production: Article **Index measuring impact:** 2.257

PRIZES AND GRANTS:

Grant name: Fellowship for undergraduate students
Donor: Spanish Ministry of Education and Sciences
Duration: December 2006 – June 2007
Place: Cancer Research Centre
University of Salamanca - CSIC

Grant name: JAE predoctoral research training grant
Donor: Consejo Superior de Investigaciones Científicas.
Duration: January 2008 - September 2008.
Place: Centro de Investigaciones Biológicas (CSIC).

ÁLVARO INGLÉS PRIETO

c/ Amberes 19 3B 10005 Cáceres Spain · +34-669538166 · ingles.prieto@gmail.com

Grant name: Plan propio short-term fellowship.
Donor: University of Granada.
Duration: August 2009 - October 2009.
Place: Dr. George Makhatadze Lab. (RPI Biology Faculty)

Grant name: Estancia de excelencia short-term fellowship.
Donor: Junta de Andalucía.
Duration: August 2009 - October 2009.
Place: Dr. George Makhatadze Lab. (RPI Biology Faculty, Troy, USA)

Grant name: FPU predoctoral research training grant.
Donor: Spanish Ministry of Education.
Duration: July 2009 - November 2012.
Place: Department of Physical Chemistry
University of Granada.

Grant name: FPU short-term fellowship.
Donor: Spanish Ministry of Education.
Duration: July 2010 - September 2010.
Place: Ana Fernández-Sesma lab. (Mount Sinai School of Medicine, New York, USA).

Grant name: Boehringer Ingelheim travel grant.
Donor: Boehringer Ingelheim Company.
Duration: August 2011 - October 2011.
Place: Julio Fernandez lab (Columbia University, New York, USA).

Grant name: EMBO short-term fellowship.
Donor: European Molecular Biology Organization.
Duration: August 2011 - October 2011.
Place: Julio Fernández lab (Columbia University, New York, USA).

TEACHING EXPERIENCE

Type of teaching: Official teaching for undergraduates.
Place: University of Granada
Subject: Physical Chemistry (Introducción a la Experimentación en Química-Física).
Course: 3rd year Bachelor's Degree in Chemistry (2011-2012)
Duration: 70 hours

Type of teaching: Official teaching for undergraduates.
Place: University of Granada
Subject: Physical Chemistry (Introducción a la Experimentación en Química-Física).
Course: 3rd year Bachelor's Degree in Chemistry (2010-2011)
Duration: 60 hours

ÁLVARO INGLÉS PRIETO

c/ Amberes 19 3B 10005 Cáceres Spain · +34-669538166 · ingles.prieto@gmail.com

SHORT-TERM STAYS

1. Place: Center For Biotechnology And Interdisciplinary Studies (**Rensselaer Polytechnic Institute**. Troy, NY, USA.)
Duration: 3 months
Director: Dr. George Makhatadze
2. Place: Department of Microbiology (**Mount Sinai School Of Medicine**. New York, NY, USA.)
Duration: 3 months
Director: Dr. Ana Fernández-Sesma
3. Place: Department of Biological Sciences (**Columbia University**, New York, NY, USA.)
Duration: 3 months
Director: Dr. Julio Fernández

KNOWLEDGE OF LANGUAGES

- **Native level of Spanish**
- **Fluent level of English.**
 - 2010. English course in New York (Teacher's College. Columbia University). 5 weeks. Level: Advanced
 - 2007. English course in Hastings (UK). 3 weeks. Level: Intermediate/advanced
- **Basic level of Italian.**
 - 2008. Italian course in Veglie (Italy). 2 weeks. Level B1

ANOTHER DATA

- University Debate League 2005: University of Salamanca Team.
- University Debate League 2006: University of Salamanca Team.
- InterCEU 2006 (Debate Competition): University of Salamanca Team.
- Person in charge of the computers room in the Fray Luis de León Hall of Residence (2004/2005).
- Person in charge of the library in the Fray Luis de León Hall of Residence (2004/2005).

