

UNIVERSIDAD DE GRANADA

FACULTAD DE CIENCIAS



DEPARTAMENTO DE QUÍMICA FÍSICA

**RELACIÓN ENTRE ESTABILIDAD, ESTRUCTURA
Y EVOLUCIÓN DE PROTEÍNAS: ESTUDIOS
MUTACIONALES, TERMODINÁMICOS,
CINÉTICOS Y CRISTALOGRÁFICOS”**

Raquel Godoy Ruiz

**TESIS DOCTORAL
GRANADA
2006**

UNIVERSIDAD DE GRANADA
FACULTAD DE CIENCIAS
DEPARTAMENTO DE QUÍMICA FÍSICA

Memoria presentada para optar al Grado de Doctor en Ciencias Químicas

Fdo: Raquel Godoy Ruiz

Licenciada en Ciencias Químicas por la Universidad de Granada

Granda, 20 de Julio de 2006

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
Universidad de Granada

Fdo:
José Antonio Gavira Gallardo
Investigador
Instituto Andaluz de Ciencias de la
Tierra
Laboratorio de Estudios
Cristalográficos
CSIC

A mis padres

A mi hermano

AGRADECIMIENTOS

Quiero agradecer a José Manuel Sánchez Ruiz, en primer lugar, el haberme dado la oportunidad de llevar una tesis doctoral en su grupo de investigación. Pero mucho más importante, el haberme enseñado a enfrentarme a un proyecto científico con objetividad, perspectiva, creatividad e ingenio.

A Beatriz Ibarra Molero, debo agradecer su paciencia por el tiempo que ha pasado conmigo en el laboratorio. Todas las reuniones que hemos mantenido durante cuatro años, me han enseñado a razonar y me han transmitido su entusiasmo. Pero mucho más importante, es que siempre he podido tener su consejo y apoyo.

A José Antonio Gavira Gallardo, agradezco su paciencia y esfuerzo, pues he tenido la oportunidad de aprender técnicas de cristalización de la mano de un experto y de introducirme en el mundo de la cristalografía con la gracia gaditana.

A Victor Muñoz (University of Maryland), por haberme aceptado y dado la fabulosa acogida en su grupo de investigación durante las dos estancias que he realizado en Estados Unidos.

A Juan Ma. García Ruiz (Laboratorio de Estudios Cristalográficos), pues gracias a la colaboración mantenida con su grupo de investigación, he podido llevar a cabo la cristalización de las proteínas estudiadas en esta tesis doctoral.

A Javier Sancho (Universidad de Zaragoza) y su grupo de investigación, pues siempre me han acogido muy bien en Zaragoza y he tenido la oportunidad de colaborar con ellos en trabajos muy interesantes.

A mis compañeros del grupo de investigación: Fernando, David, Antonio y resto de compañeros del departamento: José, Irene, Javier, Manuel, Nico, Lorena, José Luis, Sisí...y como no a José Miguel, gracias a todos ellos. Un agradecimiento muy especial a M^a Mar, Eva, Isa y Ana, que me han ayudado tanto con las prácticas cuando menos tiempo tenía. No olvido Mercedes Guzman y Antonio Parody, pues con ellos di mis primeros pasos en el laboratorio. A mis compañeros de Maryland: Urmi, Athi, Adam por su acogida y el apoyo que me dieron en Maryland.

A mis amigos, con los que he compartido tantos almuerzos y buenos ratos: Adela, Julia, Alegría, Silvia, Ana, Andrés y Jose Manuel. Y a los que ya no están tan cerca, como Paco y Miriam. A Mourad, gracias por su apoyo incondicional. A Sandra, que a pesar de la distancia, siempre hemos tenido un momento para ponernos al día. A Bertrand, que siempre me ha animado y apoyado cuando más agobiada estaba.

Finalmente, a mis padres, a quien se lo debo todo, pues sin su cariño y apoyo, jamás hubiera conseguido mi sueño. Y a mi hermano, a quien le agradezco su amistad, los buenos momentos compartidos y el haberme sacado del laboratorio para contarme los nuevos descubrimientos arqueológicos.

ÍNDICE

1	INTRODUCCIÓN	1
2	MATERIALES Y MÉTODOS	9
2.1	Materiales	9
2.1.1	Productos	9
2.1.2	Instrumentación	11
2.2	MÉTODOS.....	13
2.2.1	OBTENCIÓN DE ADN DE PROTEÍNAS MUTANTES	13
2.2.1.1	Mutagénesis Dirigida.....	13
2.2.1.2	Diseño de los oligonucleótidos cebadores.....	14
2.2.1.3	Condiciones de la PCR.....	16
2.2.1.4	Transformación de células supercompetentes XL1-Blue.....	20
2.2.1.5	Preparación del ADN-dc mutado	21
2.2.2	EXPRESIÓN Y PURIFICACIÓN DE PROTEÍNAS.....	21
2.2.2.1	Preparación de células competentes JF521	21
2.2.2.2	Transformación de células competentes JF521	22
2.2.2.3	Transformación de células supercompetentes BL21-DE3	22
2.2.2.4	Pruebas de expresión de proteína	22
2.2.2.5	Expresión de proteína a gran escala	23
2.2.2.6	Purificación de la proteína.....	25
2.2.3	PREPARACIÓN DE LAS MUESTRAS	29
2.2.3.1	Preparación de los tampones y diálisis.....	29
2.2.3.2	Preparación de disoluciones concentradas de desnaturizante.....	30
2.2.4	EXPERIMENTO de CALORIMETRÍA diferencial de barrido ...	31
2.2.4.1	Generalidades	31
2.2.4.2	Experimento calorimétrico	34
2.2.4.3	Análisis de los datos de DSC según la termodinámica de equilibrio	39
2.2.5	DICROÍSMO CIRCULAR	51
2.2.5.1	Bases teóricas	51
2.2.5.2	Espectros de dicroísmo circular de proteínas	53
2.2.5.3	Dicroísmo circular y la estructura secundaria de las proteínas ..	54

2.2.5.4	Dicroísmo circular y la estructura terciaria de las proteínas	56
2.2.5.5	Procedimiento experimental	59
2.2.6	FLUORESCENCIA	60
2.2.6.1	Generalidades	60
2.2.6.2	Análisis de las cinéticas de repliegamiento-despliegamiento inducido por urea	62
2.2.6.3	Estados de transición y método de los valores de ϕ	71
2.2.6.4	Ensayos de despliegamiento	73
2.2.7	COMPARACIÓN DE SECUENCIAS DE PROTEÍNAS Y EVOLUCIÓN DE PROTEÍNAS	79
2.2.7.1	Conceptos generales sobre evolución	79
2.2.7.2	Alineamientos de secuencias realizados para tiorredoxina de <i>E. Coli</i>	81
2.2.8	LA CRISTALIZACIÓN DE PROTEÍNAS	85
2.2.8.1	Características de la solución de proteína	86
2.2.8.2	Metodología de cristalización utilizada: cristalización por contradifusión	87
2.2.8.3	Condiciones de cristalización elegidas	92
2.2.8.4	Protocolo de cristalización por contradifusión en capilares	95
2.2.8.5	Seguimiento del crecimiento cristalino	96
2.2.9	ESTRUCTURA DE PROTEÍNAS POR DIFRACCIÓN DE RAYOS X	98
2.2.9.1	Metodología de un experimento de difracción. Instrumentación utilizada	98
2.2.9.2	El problema de las fases	100
2.2.9.3	Método de resolución estructural: Reemplazamiento Molecular ..	105
2.2.9.4	Validación final de la estructura	109
	BIBLIOGRAFÍA	111

3	ESTUDIO DE LA RELACIÓN ENTRE ESTABILIDAD DE PROTEÍNAS, EVOLUCIÓN Y ESTRUCTURA BASADO EN UN ANÁLISIS MUTACIONAL DE ÁCIDOS CARBOXÍLICOS	119
---	---	-----

3.1	ESTUDIO DEL EFECTO DE LAS MUTACIONES E→D Y D→E SOBRE LA ESTABILIDAD DE TIORREDOXINA	120
3.1.1	Obtención de las proteínas mutadas del tipo E→D y D→E.....	120
3.1.2	Estudio de la estabilidad termodinámica de las formas mutantes de tiorredoxina mediante calorimetría diferencial de barrido.	121
3.2	CORRELACIÓN DE LOS EFECTOS DE MUTACIONES DE ÁCIDOS CARBOXÍLICOS SOBRE LA ESTABILIDAD CON SU FRECUENCIA DE APARICIÓN EN ALINEAMIENTOS DE SECUENCIAS	128
3.3	CONCLUSIONES.....	145
	BIBLIOGRAFÍA.....	147
4	PATRÓN DE ESTABILIDAD DE MUTACIONES HIDROFÓBICAS EN PROTEÍNAS QUE REFLEJA OPTIMIZACIÓN ESTRUCTURAL EVOLUTIVA.....	151
4.1	ESTUDIO DEL EFECTO DE MUTACIONES DE RESIDUOS HIDROFÓBICOS SOBRE LA ESTABILIDAD DE TIORREDOXINA.....	153
4.1.1	OBTENCIÓN DE LAS PROTEÍNAS MUTADAS DEL TIPO I-->V Y V-->I.....	153
4.1.2	ESTUDIO DE LA ESTABILIDAD TERMODINÁMICA DE LAS FORMAS MUTANTES DE TIORREDOXINA MEDIANTE CALORIMETRÍA DIFERENCIAL DE BARRIDO	154
4.2	CORRELACIÓN DE LOS EFECTOS DE MUTACIONES SOBRE LA ESTABILIDAD CON LA FRECUENCIA DE APARICIÓN DE RESIDUOS EN UN ALINEAMIENTO DE SECUENCIAS: TEST DE LA HIPÓTESIS DE PSEUDOEQUILIBRIO.	160
4.3	EL PATRÓN EXPERIMENTAL DE ESTABILIDAD-EMPAQUETAMIENTO PARA MUTACIONES HIDROFÓBICAS REFLEJA OPTIMIZACIÓN ESTRUCTURAL EVOLUTIVA.....	180
4.4	UN MODELO EVOLUTIVO SIMPLE PARA EL EFECTO DE MUTACIONES HIDROFÓBICAS SOBRE LA ESTABILIDAD DE PROTEÍNAS	181
4.5	APLICACIÓN DEL MODELO A LOS DATOS DE MUTACIONES DE TIORREDOXINA	190

4.5.1	OBTENCIÓN DE LAS PROTEÍNAS MUTADAS V16D, L42D, A67D y L99D	191
4.5.2	ESTUDIO DE LA ESTABILIDAD TERMODINÁMICA DE V16D, L42D, A67D y L99D MEDIANTE CALORIMETRÍA DIFERENCIAL DE BARRIDO.....	191
4.5.3	CARACTERIZACIÓN ESTRUCTURAL DE V16D, L42D, A67D y L99D MEDIANTE DICROÍSMO CIRCULAR	195
4.6	CONSERVACIÓN DEL AMBIENTE <i>VERSUS</i> CONSERVACIÓN DE RESIDUOS.....	197
4.7	RUGOSIDAD DEL PAISAJE DE ENERGÍA DE ESTABILIDAD Y SECUENCIA (ANALOGÍA CON UN JARDÍN JAPONÉS DE ROCAS)	199
4.8	CONCLUSIONES.....	200
	BIBLIOGRAFÍA.....	203
5	SELECCIÓN NATURAL DE LA ESTABILIDAD CINÉTICA: PROBABLE ORIGEN DE LAS CORRELACIONES ENTRE EFECTOS DE MUTACIÓN SOBRE LA ENERGÉTICA DE LAS PROTEÍNAS Y LAS FRECUENCIAS DE APARICIÓN DE AMINOÁCIDOS EN ALINEAMIENTOS DE SECUENCIAS	207
5.1	PLANTEAMIENTO INICIAL DEL PROBLEMA	207
5.2	DETERMINACIÓN DEL NIVEL DE ESTABILIDAD MÍNIMA PERMITIDA EVOLUTIVAMENTE PARA TIORREDOXINA.....	214
5.3	EFECTOS DE MUTACIONES <i>versus</i> ESTABILIDAD CINÉTICA	221
5.4	ESTUDIO EXPERIMENTAL DE LAS VELOCIDADES DE PLEGAMIENTO Y DESPLEGAMIENTO DE TIORREDOXINA DE <i>E. COLI</i> WT Y 27 MUTANTES SIMPLES	223
5.4.1	DESNATURALIZACIÓN EN EQUILIBRIO DE TIORREDOXINA ESTUDIADA POR ENSAYOS DE DESPLEGAMIENTO DE DOBLE SALTO	223
5.4.2	CINÉTICAS DE PLEGAMIENTO Y DESPLEGAMIENTO ESTUDIADAS MEDIANTE MEDIDAS DE FLUORESCENCIA DEL ESTADO ESTACIONARIO.....	252

5.4.3	ANÁLISIS GLOBALES DE LOS DATOS DE DESPLEGAMIENTO INDUCIDO POR UREA, EN EQUILIBRIO Y CINÉTICOS....	252
5.4.4	RESULTADOS	260
5.5	CONCLUSIONES.....	265
5.6	BIBLIOGRAFÍA.....	267
6	ESTUDIO CRISTALOGRAFICO DE TIORREDOXINA DE <i>E. COLI</i> Y LAS VARIANTES D2E, D9E, D10E, D13E, D43E, E44D, D47E, E48D Y E85D.....	273
6.1	DESCRIPCIÓN GENERAL DE LA ESTRUCTURA DE TIORREDOXINA	274
6.2	CRISTALIZACIÓN DE LA PROTEÍNA, RECOGIDA DE DATOS Y RESOLUCIÓN DE ESTRUCTURA	277
6.3	COMPARACIÓN DE LA ESTRUCTURA DE TIORREDOXINA SILVESTRE CON LA DE SUS VARIANTES	281
6.3.1	R.M.S.D. DE LA ESTRUCTURA DE 2TRX (C2) COMPARADA CON LA DE 2H74, 2H70, 2H76, 2H75, 2H73, 2H6Z, 2H71, 2H6Y y 2H72 (P6 ₁).....	281
6.3.2	R.M.S.D. DE LA ESTRUCTURA DE 2H6X (P6 ₁) COMPARADA CON LA DE 2H74, 2H70, 2H76, 2H75, 2H73, 2H6Z, 2H71, 2H6Y y 2H72 (P6 ₁).....	282
6.4	ESTUDIO LOCAL DEL EFECTO DE LA MUTACIÓN SOBRE LA ESTRUCTURA DE TIORREDOXINA	294
6.5	DETERMINACIÓN DE INTERACCIONES existentes en TRE AMINOÁCIDOS.....	310
6.5.1	ESTUDIO DE LOS ACOPLAMIENTOS ENTRE los aminoácidos CONSERVADOS EVOLUTIVAMENTE.....	310
6.5.2	INTERACCIONES ENERGÉTICAS locales PREDICHAS POR ANÁLISIS ESTRUCTURALES.....	323
6.6	CONCLUSIONES.....	334
6.7	CONCLUSIONES.....	335
6.8	BIBLIOGRAFÍA.....	337
7	RESUMEN Y CONCLUSIONES	343

ÍNDICE FIGURAS

Figura 2.1	15
Figura 2.2	19
Figura 2.3	24
Figura 2.4	30
Figura 2.5	33
Figura 2.6	37
Figura 2.7	38
Figura 2.8	44
Figura 2.9	55
Figura 2.10	57
Figura 2.11	58
Figura 2.12	61
Figura 2.13	65
Figura 2.14	68
Figura 2.15	69
Figura 2.16	72
Figura 2.17	75
Figura 2.18	82
Figura 2.19	84
Figura 2.20	88
Figura 2.21	91
Figura 2.22	97
Figura 2.23	100
Figura 2.24	103
Figura 2.25	106
Figura 2.26	107
Figura 3.1	124
Figura 3.2	125
Figura 3.3	126

Figura 3.4	131
Figura 3.5	135
Figura 3.6	136
Figura 3.7	137
Figura 3.8	138
Figura 3.9	139
Figura 3.10	140
Figura 3.11	141
Figura 3.12	142
Figura 3.13	144
Figura 4.1	156
Figura 4.2	157
Figura 4.3	158
Figura 4.4	164
Figura 4.5	165
Figura 4.6	166
Figura 4.7	167
Figura 4.8	171
Figura 4.9	173
Figura 4.10	174
Figura 4.11	175
Figura 4.12	176
Figura 4.13	177
Figura 4.14	179
Figura 4.15	182
Figura 4.16	186
Figura 4.17	192
Figura 4.18	194
Figura 4.19	196
Figura 4.20	201

Figura 5.1	215
Figura 5.2	216
Figura 5.3	218
Figura 5.4	220
Figura 5.5	222
Figura 5.6	225
Figura 5.7	226
Figura 5.8	227
Figura 5.9	228
Figura 5.10	229
Figura 5.11	230
Figura 5.12	231
Figura 5.13	232
Figura 5.14	233
Figura 5.15	234
Figura 5.16	235
Figura 5.17	236
Figura 5.18	237
Figura 5.19	238
Figura 5.20	239
Figura 5.21	240
Figura 5.22	241
Figura 5.23	242
Figura 5.24	243
Figura 5.25	244
Figura 5.26	245
Figura 5.27	246
Figura 5.28	247
Figura 5.29	248
Figura 5.30	249
Figura 5.31	250
Figura 5.32	251
Figura 5.33	254
Figura 5.34	255

Figura 5.35	256
Figura 5.36	257
Figura 5.37	262
Figura 5.38	263
Figura 5.39	264
Figura 6.1	275
Figura 6.2	276
Figura 6.3	285
Figura 6.4	286
Figura 6.5	289
Figura 6.6	290
Figura 6.7	292
Figura 6.8	293
Figura 6.9	295
Figura 6.10	297
Figura 6.11	298
Figura 6.12	300
Figura 6.13	302
Figura 6.14	303
Figura 6.15	305
Figura 6.16	306
Figura 6.17	308
Figura 6.18	314
Figura 6.19	315
Figura 6.20	317
Figura 6.21	318
Figura 6.22	319
Figura 6.23	320
Figura 6.24	321
Figura 6.25	322
Figura 6.26	323
Figura 6.27	326
Figura 6.28	327

Figura 6.29	328
Figura 6.30	329
Figura 6.31	330
Figura 6.32	331
Figura 6.33	332
Figura 6.33	333
Figura 6.34	334
Figura 6.35	335

ÍNDICE DE TABLAS

Tabla 2.1	17
Tabla 2.2	83
Tabla 2.3	93
Tabla 2.4	94
Tabla 2.5	95
Tabla 3.1	127
Tabla 3.2	132
Tabla 3.3	143
Tabla 4.1	159
Tabla 4.2	168
Tabla 4.3	170
Tabla 4.4	195
Tabla 5.1	212
Tabla 5.2	258
Tabla 5.3	259
Tabla 6.1	279
Tabla 6.2	283
Tabla 6.3	284
Tabla 6.4	287
Tabla 6.5	288
Tabla 6.6	336

ABREVIATURAS Y SÍMBOLOS

C_p^{ex}	Capacidad calorifica de exceso
C_p^D	Capacidad calorifica del estado desnaturalizado
C_p^N	Capacidad calorifica del estado nativo
ΔG_I	Energia de acoplamiento
‡ (\neq)	Estado de transicion
Abs	Absorbancia
APS	Persulfato amonico
ADN	Acido desoxirribonucleico
ASA	Area accesible al disolvente
C	Concentracion del desnaturalizante
$C_{1/2}$	Concentracion del desnaturalizante a la que $\Delta G=0$
CD	Dicroismo circular
D	Estado desnaturalizado de la proteina
dNTP's	desoxinucleotidos trifosfato
DSC	Calorimetria diferencial de barrido
Gdn HCl	Cloruro de guanidinio
GdnSCN	Tiocianato de guanidinio
I	Estado intermedio; Fuerza ionica, intensidad de fluorescencia
I_∞	Intensidad de fluorescencia en el equilibrio
IPTG	Isopropyl β -D-thiogalactopiranosido
K	Constante de equilibrio
k	Constante de velocidad aparente
k_D	Constante de velocidad de desplegamiento
k_F	Constante de velocidad de replegamiento
LB	Medio de cultivo bacteriano Laura Bertani
LEM	Metodo de extrapolacion lineal
N	Estado nativo de la proteina
PCR	Reaccion en cadena de la polimerasa
q_i	Carga del grupo I
R	Constante de los gases
r	relacion entalpia calorimetrica/entalpia van't Hoff

SDS	Lauril sulfato sodico
SDS-PAGE	Electroforesis en gel de poliacrilamida en presencia de lauril sulfato sodico
T	Temperatura (Kelvin, °C)
TB	Medio de cultivo terrorifico
TEMED	N, N, N', N'-tetrametiletilendiamina
T _H	Temperatura a la que $\Delta H=0$
T _m , T _m [*]	Temperaturas a las que $\Delta G=0$, temperatura a la que la mitad de la hebra de ADN se encuentra como doble helice (T _m)
T _r	Tiempo de respuesta del calorimetro
Trx WT	Tiorredoxina de <i>E.coli</i> forma silvestre
TS	Temperatura a la que $\Delta S=0$
UV	Ultravioleta
ΔC_p	Cambio de capacidad calorifica de desnaturalizacion a presion constante (tambien $\Delta_N^D C_p$)
$\Delta C_{p,m}$	Cambio de capacidad calorifica de desnaturalizacion a la T _m
ΔG	Cambio de la energia de Gibbs de desnaturalizacion (tambien $\Delta_N^D G$)
ΔG_w	Cambio de la energia de Gibbs de desnaturalizacion en disolucion acuosa
ΔH	Cambio de entalpia de desnaturalizacion (tambien $\Delta_N^D H$)
ΔH_m	Cambio de entalpia de desnaturalizacion a la T _m
ΔH^{vH}	Cambio de entalpia de van't Hoff
ΔS	Cambio de entropia de desnaturalizacion (tambien $\Delta_N^D S$)
ΔI	Amplitud de la cinetica
ΔS_m	Cambio de entropia de desnaturalizacion a la T _m
θ	Elipticidad
τ	Tiempo de relajacion
χ	Vector de protonacion
ω	Peso estadistico

ϵ_D Coeficiente de extincion molar para la luz polarizada a la derecha

ϵ_I Coeficiente de extincion molar para la luz polarizada a la izquierda

Ala	(A)	Alanine
Asn	(N)	Asparagine
Asp	(D)	Aspartate
Arg	(R)	Arginine
Cys	(C)	Cysteine
Gln	(Q)	Glutamine
Glu	(E)	Glutamate
Gly	(G)	Glycine
His	(H)	Histidine
Ile	(I)	Isoleucine
Leu	(L)	Leucine
Lys	(K)	Lysine
Met	(M)	Methionine
Phe	(F)	Phenylalanine
Pro	(P)	Proline
Ser	(S)	Serine
Thr	(T)	Threonine
Trp	(W)	Tryptophan
Tyr	(Y)	Tyrosine
Val	(V)	Valine

1. Introducción

INTRODUCCIÓN

La gran capacidad de las proteínas para llevar a cabo tareas moleculares con enorme eficacia y especificidad hace que tengan numerosas aplicaciones tanto en biotecnología y nanobiotecnología (enzimas como catalizadores en procesos industriales, biosensores, electrónica molecular, nanomáquinas, etc.) como en biomedicina (proteínas terapéuticas). La mayoría de estas aplicaciones requiere alguna modificación de la proteína de interés, para mejorarla con respecto a alguna característica fundamental. El ejemplo más obvio lo proporciona la estabilidad (Korkegian *et al.*, 2005); las proteínas naturales suelen presentar una estabilidad marginal, bien porque estén naturalmente seleccionadas en base a otras propiedades, o bien porque una estabilidad marginal sea beneficiosa para la función (garantice la flexibilidad necesaria), para el plegamiento correcto (evite que en el proceso de plegamiento la proteína quede atrapada en estados incorrectamente plegados) o para la unicidad estructural (Godoy-Ruiz *et al.*, 2004). En cualquier caso, la baja estabilidad limita o impide muchas aplicaciones prácticas y debe corregirse mediante las modificaciones (usualmente mutaciones) adecuadas.

Desde un punto de vista metodológico, tradicionalmente son dos los enfoques que se han empleado para optimizar la estabilidad de proteínas: el diseño racional (o computacional), para encontrar, conocida la estructura de la proteína, las mutaciones que, presumiblemente, producen el efecto deseado (Ibarra-Molero y Sanchez-Ruiz, 2002, Campos *et al.*, 2004) Este enfoque requiere de un conocimiento, más o menos detallado, de las fuerzas implicadas en el mantenimiento de la estructura nativa. Como ejemplo de aplicación de este enfoque, podemos citar el procedimiento de optimización de las interacciones electrostáticas en proteínas, desarrollado en nuestro grupo de investigación y que ha sido aplicado exitosamente (Sanchez-Ruiz y Makhatadze, 2001). En la práctica, nuestros conocimientos acerca de los principios básicos que determinan la estructura/estabilidad/función de las proteínas son aún incompletos, por lo que el enfoque racional/computacional presenta limitaciones. La alternativa obvia es la llamada evolución dirigida *in vitro* (segundo enfoque tradicional de abordar el diseño de estabilidad de proteínas) (Amstutz *et al.*, 2001 y Eijsink *et al.*, 2005),

con la que se crea aleatoriamente una biblioteca (“*library*”) de variantes, que se analiza en términos de la característica deseada (“*screening*”), sometiéndose las variantes seleccionadas a nuevos ciclos de aleatorización/selección que supuestamente mejoran las proteína con respecto a la propiedad deseada. Un factor limitante en la evolución dirigida es la eficacia del procedimiento de *screening* ya que, a no ser que se puedan analizar “*libraries*” muy grandes, pueden no llegar a detectarse variantes interesantes (ver Eijsink *et al.*, 2005, para una discusión sobre éste y otros problemas asociados al procedimiento de “*screening*” de Novozymes). La ventaja de este procedimiento es que no necesita del conocimiento de factores que limitan la estabilidad. A la vista de esto, se ha propuesto recientemente la posibilidad de combinar ambos enfoques, el diseño racional/computacional y la evolución dirigida (“*computationally-focused directed evolution*”, ver referencias Voigt *et al.*, 2001, Ibarra-Molero y Sanchez-Ruiz, 2002 y Wunderlich *et al.*, 2005). Así, cálculos computacionales basados en la estructura de la proteína pueden utilizarse para decidir un número comparativamente pequeño de posiciones prometedoras a aleatorizar para conseguir una “*library*” de variantes de pequeño tamaño (procedimiento puesto a punto por nuestro grupo de investigación en colaboración con la empresa Novozymes; ver Rodríguez-Larrea *et al.*, 2006).

Hasta ahora hemos planteado los dos enfoques que, tradicionalmente, se vienen empleado para diseñar la estabilidad de proteínas. No obstante existe un enfoque adicional, que hasta el momento no se ha usado mucho porque no se conoce bien la relación entre estabilidad y evolución de proteínas, que consiste en utilizar la información evolutiva que nos ofrecen los alineamientos de secuencias. Este enfoque hace uso del “*consensus approach*” para proteínas; se trata de la posibilidad de llevar a cabo evolución dirigida *in vitro* guiada (enfocada) en base a la información que se obtiene del análisis de alineamientos de secuencias, es decir, realizando mutaciones que sugieren los alineamientos de secuencias. Es una posibilidad novedosa, apoyada por trabajos recientes que indican que las correlaciones evolutivas detectadas en alineamientos de secuencias pueden reflejar, efectos de interés, tales como comunicación alostérica a larga distancia en el estado nativo, efectos en el estado desnaturalizado, e incluso pueden

proporcionar la información necesaria para especificar el “*fold*” (Godoy-Ruiz *et al.*, 2004 y Socolich *et al.*, 2005).

Así pues, en esta tesis doctoral nos hemos planteado estudiar esta relación entre estabilidad y evolución de proteínas, por su interés en aplicaciones biotecnológicas. Hemos llevado a cabo un análisis bioinformático de los alineamientos de secuencias utilizando tiorredoxina como proteína modelo (obtenidos de una búsqueda BLAST en la base de datos de secuencias UniProt/TrMBL, utilizando la secuencia de tiorredoxina de *E. coli* como “*query*”) Hemos estudiado el efecto de 27 mutaciones conservativas (13 mutaciones del tipo glutamato (E)↔aspartato (D) y 14 mutaciones del tipo valina (V)↔isoleucina(I) de tiorredoxina de *E. coli*, sugeridas por sus coeficientes en las matrices de PAM y BLOSUM, a todos los niveles. Hemos caracterizado la estabilidad termodinámica de tiorredoxina silvestre y 27 variantes de tiorredoxina, mutaciones del tipo E↔D y V↔I, mediante calorimetría diferencial de barrido de alta sensibilidad (Godoy-Ruiz *et al.*, 2004, 2005). Estos estudios se han completado con estudios cinéticos de plegamiento/desplegamiento, para la determinación de estabilidad cinética y “estructura” de estados de transición a partir del análisis de valores de ϕ . Para ello, hemos llevado a cabo medidas cinéticas de la fluorescencia del plegamiento/desplegamiento en disoluciones de desnaturante (urea) y ensayos de desplegamiento en equilibrio de doble salto, de tiorredoxina silvestre y las 27 variantes de tiorredoxina. Además, se han realizado análisis globales de las representaciones de *Chevron* y los perfiles del equilibrio resultantes (Godoy-Ruiz *et al.*, 2006).

Los estudios termodinámicos y cinéticos se han completado con un estudio cristalográfico de tiorredoxina silvestre y variantes, nueve mutaciones de ácidos carboxílicos del total de las obtenidas. Se ha puesto a punto la cristalización de tiorredoxina silvestre y sus variantes, utilizando técnicas de cristalización por contradifusión en capilares (García-Ruiz, 2003). Los cristales obtenidos se han difractado empleando rayos X y resuelto sus estructuras por el método de reemplazamiento molecular. Las estructuras determinadas resultantes se han depositado en el Protein Data Bank (códigos de acceso: 2H6X.pdb/WT, 2H6Y.pdb/E48D, 2H70.pdb/D9E, 2H71.pdb/D47E, 2H72.pdb/E85D, 2H73.pdb/D43E, 2H74.pdb/D2E, 2H75.pdb/D13E y 2H76.pdb/D10E). A partir de

los datos cristalográficos se han realizado análisis de flexibilidad de la proteína, influencia de las mutaciones en la estructura tridimensional teniendo en cuenta el empaquetamiento cristalino y se han buscado correlaciones estructurales de la mutación de la proteína variante y posiciones acopladas detectadas en alineamientos de secuencias, mediante el estudio de las desviaciones cuadráticas medias (R.M.S.D.), de los factores-B, de los contactos locales, de las diferencias en área superficial accesible y del desplazamiento atómico medio pesado por el error asociado a la posición atómica (Lockless y Ranganathan, 1999 y Jain y Ranganathan, 2004).

- Amstutz, P., P. Forrer, C. Zahnd and A. Pluckthun (2001). *Curr Opin Biotechnol* **12**(4): 400-5.
- Campos, L.A., M.M. Garcia-Mira, R. Godoy-Ruiz, J.M. Sanchez-Ruiz and J. Sancho (2004). *J Mol Biol* **344**(1): 223-37.
- Eijsink, V.G., S. Gaseidnes, T.V. Borchert and B. van den Burg (2005). *Biomol Eng* **22**(1-3): 21-30.
- García-Ruiz, J.M. (2003). *Methods in Enzymology* **368**: 130-154.
- Godoy-Ruiz, R., R. Perez-Jimenez, B. Ibarra-Molero and J.M. Sanchez-Ruiz (2004). *J Mol Biol* **336**(2): 313-8.
- Godoy-Ruiz, R., R. Perez-Jimenez, B. Ibarra-Molero and J.M. Sanchez-Ruiz (2005). *Biophys J* **89**(5): 3320-31.
- Godoy-Ruiz, R., F. Ariza, D. Rodriguez-Larrea, R. Perez-Jimenez, B. Ibarra-Molero and J.M. Sanchez-Ruiz (2006). *J Mol Biol* **In press**.
- Ibarra-Molero, B. and J.M. Sanchez-Ruiz (2002). *J. Phys. Chem. B* **106**(26): 6609-6613.
- Jain, R.K. and R. Ranganathan (2004). *Proc Natl Acad Sci U S A* **101**(1): 111-6.
- Korkegian, A., M.E. Black, D. Baker and B.L. Stoddard (2005). *Science* **308**(5723): 857-60.
- Lockless, S.W. and R. Ranganathan (1999). *Science* **286**(5438): 295-9.
- Rodriguez-Larrea, D., S. Minning, T.V. Borchert and J.M. Sanchez-Ruiz (2006). *J Mol Biol* **360**(3): 715-24.
- Sanchez-Ruiz, J.M. and G.I. Makhatadze (2001). *Trends Biotechnol* **19**(4): 132-5.
- Socolich, M., S.W. Lockless, W.P. Russ, H. Lee, K.H. Gardner and R. Ranganathan (2005). *Nature* **437**(7058): 512-8.
- Voigt, C.A., S.L. Mayo, F.H. Arnold and Z.G. Wang (2001). *Proc Natl Acad Sci U S A* **98**(7): 3778-83.
- Wunderlich, M., A. Martin, C.A. Staab and F.X. Schmid (2005). *J Mol Biol* **351**(5): 1160-8.

2. Materiales y Métodos

2 MATERIALES Y MÉTODOS

2.1 MATERIALES

2.1.1 PRODUCTOS

A continuación se detallan los reactivos utilizados para todos los experimentos realizados.

De DIFCO, se emplearon extracto de levadura (Ref. 212750), triptona (Ref. 211705) y agar (Ref. 214010).

De MERCK fueron

Cloruro sódico. PM=58.44 g/mol (Ref. 1.06404.1000). Calidad reactivo para análisis.

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

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

Cloruro cálcico. PM= 110.99 g/mol (Ref. 233-140-8).

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).

Los siguientes productos se obtuvieron de la casa comercial SIGMA:

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

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

Lauryl sulfato sódico. PM=288.4 g/mol (Ref.L-4509).

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

Sulfato de estreptomicina. PM=1457.4 g/mol (Ref. S-9137).

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

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

Urea. PM=60.05 g/mol (Ref. U-5378).

Acrilamida y bis-acrilamida. PM=71.08 g/mol (Refs. K-31503209/324)

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

Kanamicina A (C₁₈H₃₆N₄O₁₁·H₂SO₄). PM=582.6 g/mol (Ref. K-1377).

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

De la casa PANREAC:

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

Los reactivos necesarios para la reacción de mutagénesis dirigida fueron suministrados por STRATAGENE y son los siguientes:

PfuTurbo® ADN polimerasa (2.5 U/μl). 10×tampón de reacción (Tris-HCl 200 mM pH 8.8, KCl 100mM, (NH₄)₂SO₄ 100mM, MgSO₄ 20 mM, Triton® X-100 1%, BSA 1 mg/ml) (Ref. #600250).

Mezcla de nucleótidos (dNTP mix) 10Mm. (Ref. #200519-52).

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

Enzima de restricción Dpn I (10 U/μl) (Ref. #500402-51).

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

Estos componentes se pueden adquirir por separado o bien en un kit denominado comercialmente: *QuickChange™ Site-Directed Mutagenesis Kit*. (Ref. #200515-51).

Los tubos Falcon® 2059 de polipropileno fueron suministrados por BECTON-DICKINSON (Ref.#352059).

QIAGEN suministró el kit de purificación de ADN: *QIAprep® Spin Miniprep* (Ref.#27104).

Los reactivos para la cristalización de proteínas fueron suministrados por la casa HAMPTON RESEARCH. Los capilares para la cristalización fueron suministrados por HILGENBERG GMBH. Refs. 4007601 (0.1 mm diámetro), 4007602 (0.2 mm diámetro) y 4007603 (0.3 mm diámetro).

El agua utilizada en la preparación de las disoluciones fue siempre destilada, desionizada mediante una resina de intercambio iónico (1μS) y nuevamente desionizada y filtrada con un equipo Milli-Q (Millipore).

Para la diálisis se utilizaron membranas de SPECTRA/POR con un punto de corte entre 6000-8000 Daltons (diámetro: 23 mm, Ref.#132650; diámetro:50 mm, Ref.#132665), tamaño de poro que permite el paso de moléculas de peso molecular menor al de la proteína (11756 Da). Antes de su utilización es necesario hidratar las membranas al menos durante 30 minutos con agua Milli-Q.

Los filtros utilizados fueron de diferentes tamaños y con diámetros de poro de 0.22 y 0.45 μm. Se compraron a las casas PALL, MILLIPORE y WHATMAN.

Los concentradores fueron de Amicon (Millipore).

El vector pET30a (+) con el gen que codifica para tiorredoxina *WT* fue suministrado por TOP Gene Technologies (www.topgenetech.com).

Además, en la expresión y purificación de tiorredoxina de *E. coli* se necesitaron: cepas JF521 de *E. coli* y plásmido pTK100 que contiene el gen que codifica para tiorredoxina, amablemente proporcionadas por la Dra. María Luisa Tasayco y su grupo de Investigación (Biochemistry Division, Department of Chemistry The City College of New York).

2.1.2 INSTRUMENTACIÓN

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

Un autoclave “Autester-D”, un baño termostático con agitación orbital, Unitronic 320 OR y una estufa de cultivos “Incubat”, todos de la casa Selecta.

La amplificación del ADN mutado se ha realizado en un termociclador marca TECHNE PROGENE.

El lisado de los cultivos celulares se ha llevado a cabo en una prensa *French*, de la casa AMINCO.

Una microcentrífuga con cabezal angular, de Heraeus. Centrífugas KOKUSAN Modelo H-251, con rotor M1 y Hettich Modelo Roto Super 40, con rotor A.69. Y una ultracentrífuga Optima™ LE-80K de BECKMAN COULTIER, con rotor 45Ti.

Las medidas de pH se han realizado con dos tipos de electrodos de vidrio CRISON: 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). Estos electrodos estaban conectados a un pH-metro digital CRISON Digit-501, de sensibilidad 0.01 unidades. La calibración del pH-metro se realizó sistemáticamente con disoluciones estándar de pH 4.00 y 7.02 (a 25°C), también de la casa CRISON.

La concentración de disoluciones de proteína se ha realizado con las células de ultrafiltración de MILLIPORE (3 y 50 mL).

Espectrofotómetro PERKIN ELMER modelo Lambda 25 de doble haz para la obtención de los espectros ultravioleta-visible de las distintas muestras de proteína con las que se trabajó.

Las electroforesis de geles de poliacrilamida se han realizado con una cubeta vertical Mighty Small SE250/SE260.

Para llevar a cabo la purificación de la proteína se utilizaron una columna de exclusión molecular de 2L modelo XK-50 de *Sephacryl*TM S-100 *High Resolution*, un inyector, un colector de fracciones modelo Frac-100 y un detector LKB-Control Unit UV-1, todos ellos de la casa de AMERSHAM PHARMACIA BIOTECH AB. Una columna de intercambio iónico de 250 mL Fractogel® EMD DEAE(M), de MERCK. Un equipo de HPLC (*High Performance Liquid Chromatography*) KONIC-500-A Series que más tarde fue sustituido por una bomba peristáltica de la casa GILSON. Como sistema de registro un registrador LLOYD *Graphic 450*.

Para los experimentos de Calorimetría Diferencial de Barrido (DSC, *Differential Scanning Calorimetry*) se ha utilizado un calorímetro VP-DSC (MicroCal Incorporated).

Los experimentos de dicroísmo circular (CD, *Circular Dicroism*) se han realizado con un espectropolarímetro JASCO modelo J-715 equipado con un control de temperatura Jasco, PTC-348WI.

Los experimentos de fluorescencia se han realizado con un espectrofluorímetro PERKIN ELMER modelo LS55. La termostatación de las muestras y de la cubeta situada en el portacubetas se llevó a cabo recirculando agua desde un baño termostático a 25°C de la marca SELECTA modelo *Ultraterm* 6000383. La temperatura se comprobaba directamente en la cubeta mediante una sonda de platino tipo PT100 conectada a un termómetro digital LAUDA modelo R 42/2 que permite apreciar hasta 0.01°C. Como foco frío en los baños utilizados se dispuso de un dedo frío inmerso en el baño, también de la marca SELECTA.

Para las medidas de índice de refracción se ha utilizado un refractómetro de mano ATAGO modelo R5000, para medir la concentración de urea. La urea se ha purificado mediante una resina de intercambio iónico AG501-X8 y una columna dimensiones 1.5x20 cm, ambas de BIORAD.

El análisis de los resultados se ha realizado con los siguientes programas comerciales: SigmaPlot 2000 y SigmaPlot 8 para Windows de la casa SPSS Inc., MLAB de la compañía Civilized Software Inc. y MatLab® 6.5 de MathWorks Inc.

2.2 MÉTODOS

2.2.1 OBTENCIÓN DE ADN DE PROTEÍNAS MUTANTES

2.2.1.1 Mutagénesis Dirigida

Hemos utilizado el método de mutagénesis dirigida Quikchange™ de STRATAGENE, que permite realizar mutaciones (puntuales y múltiples) en un plásmido de cadena doble. La mutagénesis dirigida es una técnica muy útil para determinar la relación entre estabilidad y estructura de proteínas. El método usado presenta muchas ventajas respecto a otras técnicas, por ejemplo no necesita la subclonación en vectores del bacteriófago M13 o un ADN parental de cadena sencilla. En cuatro pasos genera ADN mutado con más de un 80% de eficiencia. Con este método se pueden hacer mutaciones puntuales, eliminar e insertar uno o varios aminoácidos. El procedimiento básico utiliza el vector de ADN de doble cadena (ADN-dc) que incluye el gen que expresa para la proteína y dos oligonucleótidos cebadores sintéticos que contienen la mutación deseada, complementarios y de cadena sencilla (ver Figura 2.1). La reacción en cadena de la polimerasa (PCR, Nobel de Química a K. Mullis en 1993) se lleva a cabo en un termociclador; comienza con la desnaturalización térmica del vector de doble cadena. Los oligonucleótidos cebadores, cada uno complementario a una cadena del vector, se extienden sobre el vector durante la etapa de anillamiento y sirven de cebadores a la PfuTurbo® ADN polimerasa que se encarga de la adición de nucleótidos, replicando ambas cadenas de plásmido parental con una alta fidelidad sin desplazar los oligonucleótidos cebadores (etapa de extensión). Ésto genera un plásmido mutado puntualmente (ADN-dc). Una vez finalizado el programa del termociclador la mezcla de reacción es tratada con la enzima endonucleasa Dpn I, específica para ADN metilado y hemimetilado (secuencia diana: 5'-Gm6ATC-3'). Se usa para digerir el ADN parental y así obtener el ADN sintetizado mediante la reacción de polimerización que contiene la mutación (ADN-dc). Ésto es posible gracias a que el ADN aislado de la mayoría de *E. coli* es dam metilado (dam+) y por lo tanto susceptible a la digestión por Dpn I. El vector que incorpora la mutación deseada, de estructura abierta, es seguidamente transformado en la células supercompetentes XL1-Blue (células comerciales preparadas para incorporar la molécula de ADN bicatenaria).

La poca cantidad de ADN parental requerida, la alta fidelidad de la PfuTurbo® ADN polimerasa y el bajo número de ciclos necesarios en el programa del termociclador dan como resultado una alta eficiencia de mutación y un descenso de las mutaciones aleatorias durante la reacción.

2.2.1.2 Diseño de los oligonucleótidos cebadores.

Cada pareja de oligonucleótidos cebadores (oligos) deben de ser diseñados individualmente de acuerdo con la mutación que se desee hacer. Los dos oligos son complementarios entre si y cada uno de ellos, a una de las cadenas del plásmido (ADN-dc). Para el diseño se pueden seguir las siguientes recomendaciones:

Si definimos la temperatura de fusión del oligo (T_m) como la temperatura a la cual el 50% del oligo no se encuentra enlazado al vector de ADN (ADN-dc) o a su complementario, es aconsejable diseñar oligos cuya T_m oscile entre los 73 y 78°C (longitud aproximada de 30 a 40 bases). La razón es que elevadas T_m , favorecen altas temperaturas de anillamiento (T_a) como se explica en sección 2.2.2.3. La T_m de los oligos se puede estimar con la siguiente fórmula:

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

donde N es el número total de bases del oligo y %G/C es tanto por ciento de Guanidinas/Citosinas.

Además, cuando se diseñen los oligos se intentará colocar la mutación en el centro de los mismos. Si ésto no es posible, se intenta minimizar el desplazamiento con respecto al centro. Por otro lado, el % GC debe ser superior al 40% y que el final de la secuencia del oligo contenga una o varias G/C, con objeto de que los extremos del mismo formen enlaces más estables.

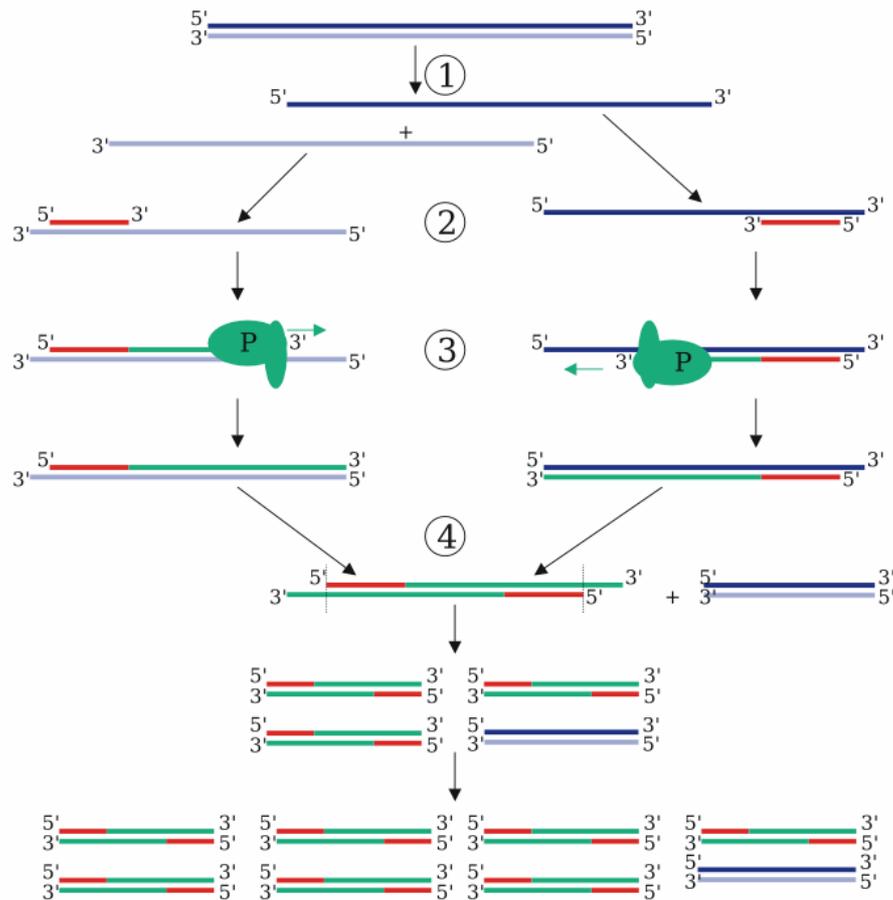


Figura 2.1: Síntesis de la cadena de ADN mutado mediante PCR. Se pueden distinguir principalmente 3 etapas: 1) desnaturalización del ADN-dc parental, 2) anillamiento de los oligos que contienen la mutación deseada y 3) extensión de los oligos con *Pfu*-ADN Polimerasa.

Y por último, el factor más importante parece encontrarse la estrategia seguida para realizar la mutación del codón. Parece influir mucho en la eficiencia el que se realice un único cambio de nucleótido o uno doble. Se ha observado que al realizar un doble cambio la eficiencia en la PCR mejora. Para diseñar dobles cambios se utilizan las siguientes estrategias:

- Mutar el primer y segundo nucleótido del codón. Es el más común y sencillo de hacer ya que sólo se necesita tener el tercer nucleótido del codón a mutar común.

- Mutar el segundo y tercer nucleótido del codón. Sólo podrá realizarse cuando el primer nucleótido del codón es común para el aminoácido de origen y el de destino. Si se puede, se elige el codón con mayor propensión en *E. coli*.

- Mutar el primer nucleótido del codón a mutar y el tercero del codón anterior. Es una estrategia con más posibilidades, pero es necesario que el codón a mutar presente el segundo y tercer nucleótido común al aminoácido de origen y al de destino.

- Menos posibilidades tiene la mutación del tercer nucleótido del codón a mutar y el primero del codón posterior. Para ello es necesario que el aminoácido posterior tenga una alta degeneración (Leu y Arg).

- Por último, una mutación del primer y tercer nucleótido del codón también parece funcionar mejor que un cambio simple. Por ello, aunque como última opción, se puede usar este cambio.

Existen ciber páginas muy útiles para ayudarnos con el diseño de oligos. Se pueden destacar:

<http://www.idtdna.com/analyzer/Applications/OligoAnalyzer/Default.aspx>

http://www.genebee.msu.su/services/rna2_reduced.html

<http://mbcf.dfc.harvard.edu/docs/oligocalc.html>

<http://www.basic.northwestern.edu/biotools/oligocalc.html>

<http://www.pitt.edu/~rsup/OligoCalc.html>

http://www.ch.embnet.org/software/LALIGN_form.html

http://bioinformatics.org/sms/rev_comp.html

2.2.1.3 Condiciones de la PCR.

En este punto se van a comentar las cantidades de cada reactivo utilizadas generalmente para realizar la mutagénesis dirigida, debido a que es muy importante la proporción en la que se encuentre los oligos y el ADN parental. Es importante que la cantidad de oligos se encuentren en exceso respecto a la cantidad de ADN parental, de esta forma nos aseguramos que el número de moléculas de oligos es superior al de ADN parental y que no existirán problemas en la etapa de anillamiento.

La mezcla de reacción consta de los siguientes reactivos:

40 μ l de agua Milli-Q destilada y desionizada

5 μ l de tampón de reacción 10 \times

1 μ l de mezcla de nucleótidos dNTP 10 mM

1 μ l de ADN parental (100 ng/ μ l)

1 μ l de primer directo 5' \rightarrow 3' (200 ng/ μ l)

1 μ l de primer reverso 3' \rightarrow 5' (200 ng/ μ l)

1 μ l de PfuTurbo ADN Polimerasa

El programa que se utiliza en el termociclador para realizar la PCR es el siguiente:

La temperatura de anillamiento (T_a) es una elección muy importante en el experimento de mutagénesis dirigida. Aunque se recomienda una temperatura 5°C inferior a la T_m de los oligos, T_a puede oscilar desde 55°C hasta 72°C. Tenemos pues cierta libertad de elección. El problema es que hay que encontrar un compromiso entre T_a y especificidad del producto de PCR. Esto es debido a que cuanto más baja es T_a , más fácilmente van a unirse los oligos a la doble cadena desnaturalizada de ADN parental pero con una menor especificidad, es decir, se producen inserciones o uniones incorrectas de los oligos a la hebra de ADN parental. Por el contrario, una T_a elevada nos garantiza una hibridación con una probabilidad baja de uniones incorrectas entre el oligo y la hebra de ADN parental.

En la PCR se pueden distinguir 3 etapas (ver Figura 2.1 y Tabla 2.1):

- Desnaturalización del ADN-cd parental: durante la desnaturalización a 95°C, la doble cadena de ADN se abre para originar dos cadenas simples.

- Anillamiento de los oligos: Los oligos se unen a las cadenas simples de ADN-cs parental. Esto ocurre a temperaturas entre 50-72°C

- Extensión: La polimerasa lleva a cabo la reacción de polimerización creando así una nueva cadena, complementaria a la hebra parental a la temperatura de 72°C.

Tabla 2.1: Programa de la PCR seleccionado en el termociclador.

Programa	Número de ciclos	Temperatura	Duración
1	1	95°C	2 min.
2	30	95°C Tm-5°C 72°C	30 seg. 30seg. 2 min./kb plásmido
3	1	72°C	10 min.

Una vez finalizado el programa del termociclador se le añade a cada tubo de PCR 1 µl de enzima de restricción Dpn I. La enzima Dpn I es específica para ADN metilado y hemimetilado (secuencia diana: 5'-Gm6ATC-3'). Se agita suavemente y se incuba 1 hora a 37°C. De esta forma se digiere el ADN-dc parental, que es el no mutado. Y por tanto, queda libre el ADN-dc sintetizado por PCR, de estructura lineal, que contiene la mutación deseada.

En este trabajo de investigación se han utilizado dos plásmidos diferentes: pTK100 y pET30a (+) (Figura 2.2), cada uno de ellos con el gen que codifica para tiorredoxina insertado.

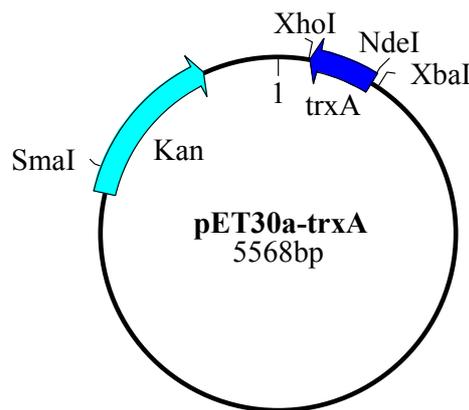
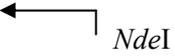


Figura 2.2: Mapa del plásmido pET30-trxA. El gen que codifica para tiorredoxina silvestre (azul oscuro), ha sido clonado entre *XhoI* y *NdeI*. La longitud total del plásmido es de 5568 pares de bases.

A continuación se detalla tanto la secuencia de bases nitrogenadas del mencionado gen para tiorredoxina forma silvestre (*WT*, *Wild Type*), así como la secuencia de aminoácidos a la que da lugar:

pET30a(+) ←  NdeI

GTAATTCCTCTANAAATACCTTGTTTTTTTAAGAAGGAGATACATATGAGCGATA
AAATTATTCACCTGACTGACGACAGTTTTGACACGGATGTA CTCAAAGCGGACGGGG
CGATCCTCGTCGATTTCTGGGCAGAGTGGTGCGGTCCGTGCAAATGATCGCCCCG
ATTCTGGATGAAATCGCTGACGAATATCAGGGCAA ACTGACCGTTGCAA ACTGAAC
ATCGATCAAACCCTGGCACTGCGCCGAAATATGGCATCCGTGGTATCCCGACTCT
GCTGCTGTTCAAAAACGGTGAAGTGGCGGCAACCAAAGTGGGTGCACTGTCTAAAG
GTCAGTTGAAAGAGTTCCTCGACGCTAACCTGGCGTAACTCGAGCACCACCACCAC
CACCACTGAGATCCGGCTGCTAACAAAGCCCGAAAGGAAGCTGAGTTGGCTGCTGC
CACCGCTGAGCAATAACTAGCATAACCCCTTGGGGCCTCTAAACGGGTCTTGAGGG
GTTTTTTGCTGAAAGGAGGAACTATATCCGATTGGCGAATGGGACGCGCCCTGTA
GCGGCGCATTAAAGCGCGGGGGTGTGGTGGTTACGCGCAGCGTGACCGCTACACT
TGCCAGCGCCCTAGCGCCCGTCCCTTTGCTTTCTTCCCTTCCCTTCTCGCCACGTT
CGCCGGCTTTCCCGTCAAGCTCTAAATCGGGGGCTCCCTTAGGGTTCCGATTTA
GTGCTTTACGGCACCTCGACCCCAAAAACTTGATTAGGGTGATGGTTCACGTAGTG
GGCCATCGCCCTGATANACGGTTTTTCGCCCTTGACGTTGGAGT

XhoI → pET30a(+)

Secuencia de aminoácidos para tiorredoxina *WT*:

SDKIIHLTDDSFDTDVLKADGAILVDFWAEWCGPCKMIAPILDEIADEYQG
KLTVAKLNIDQNP GTAPKYGIRGIPTLLL FKNGEVAATKVGALS KGLKE
FLDANLA

2.2.1.4 Transformación de células supercompetentes XL1-Blue.

Las células XL1-Blue se encuentran congeladas a -80°C en alícuotas de 200 μl . Una correcta manipulación evitará pérdidas de eficiencia en la transformación, por lo que si se saca una alícuota del congelador hay que mantenerla en hielo todo el tiempo. Una vez descongelada, hacemos alícuotas de 50 μl de células en los tubos Falcon® 2059 de polipropileno (una por cada ADN-dc mutado producto de PCR que se desee transformar en XL1-Blue), previamente enfriados en hielo. El resto de las células que no se vayan a utilizar hay que congelarlas rápidamente de nuevo a -80°C .

Es importante utilizar los tubos Falcon® 2059, ya que otro tipo de tubos podrían ser degradados por el β -mercaptoetanol, utilizado para aumentar la eficiencia de la transformación. Además, el grosor del tubo es el adecuado para la etapa del pulso de calor según se explica a continuación.

El protocolo a seguir para transformar las células supercompetentes XL1-Blue con el ADN-dc mutado producto de PCR es el siguiente:

Las alícuotas de células supercompetentes XL1-Blue se mantienen en hielo durante la transformación. En primer lugar, se añaden 0.8 μl de β -mercaptoetanol 1.42 M a cada alícuota de 50 μl , diluyéndolo así hasta a una concentración final 25 mM. Ésto se agita suavemente y se incuban los tubos 10 minutos en hielo, agitando cada 2 minutos. A continuación se añaden de 0.1-50 ng de ADN-dc (mínimo 1 μl de ADN-dc) producto de PCR (previamente tratado con Dpn I) a cada alícuota de células y se agita suavemente. Es importante en este paso realizar un control con una muestra de ADN de concentración conocida, para comprobar que la transformación ha ocurrido sin problemas, ya que desconocemos la concentración del ADN-dc producto de PCR. Y se incuban 30 minutos en hielo.

El pulso de calor es la etapa más crítica en la transformación. Consiste en introducir los tubos 45 segundos en un baño termostatzado a 42°C . La duración de este pulso de calor puede variar, nunca menos de 45 segundos ni más de 60 segundos. Seguidamente, se incuban los tubos en hielo durante 2 minutos. Se añaden 0.5 ml de medio SOC precalentado a 42°C y de nuevo se incuban los tubos en un baño orbital termostatzado a 37°C durante 1 hora. Con el pulso de calor se favorece la incorporación de las moléculas de ADN-dc mutado de estructura lineal

en las células. Una vez dentro las células mediante enzimas ligasas convertirán el ADN-dc lineal en ADN-dc circular.

Finalmente, se plaquean 100-200 μ l de mezcla de transformación en placas con LB-Agar a una concentración de 40 μ g/ml de kanamicina y se incuban en una estufa termostaticada a 37°C toda la noche durante al menos 16 horas. Al cabo de este periodo de tiempo crecerán sólo las colonias de células que contengan el plásmido con el gen que ofrece resistencia a kanamicina.

2.2.1.5 Preparación del ADN-dc mutado

De las placas crecidas durante toda la noche, según se ha descrito en el apartado anterior se eligen un número determinado de colonias y se crecen en un preinóculo de medio líquido según se describe en el apartado 2.2.2.1. El ADN-dc es extraído de dichas células utilizando el kit QIAprep®, en cuyo protocolo se explican detalladamente todos los pasos a seguir. Por último, se necesita secuenciar el ADN preparado anteriormente. Todas las muestras preparadas en este trabajo de investigación se han enviado al servicio de secuenciación *Sistemas Genómicos* en Valencia, (www.sistemasgenomicos.com).

2.2.2 EXPRESIÓN Y PURIFICACIÓN DE PROTEÍNAS

En este apartado se describen las etapas necesarias para expresar y purificar tiorredoxina, formas silvestre y mutante.

Se han utilizado dos tipos de células para expresar proteína: células JF521 y células supercompetentes BL21-DE3. Las células BL21-DE3 se compran listas para usar, mientras que las JF521 es necesario hacerlas competentes.

2.2.2.1 Preparación de células competentes JF521

Los genes de tiorredoxina silvestre de *E. coli* y los mutantes ($E \rightarrow D$ y $D \rightarrow E$) se encuentran clonados en pTK100. Para su expresión se usó las cepas JF521 de *E. coli*. Estas células se almacenan en el congelador a -80°C. Para poder utilizarlas, lo primero que hay que hacer es plaquear durante varios días consecutivos en placas Petri con medio sólido LB-Agar, con objeto de obtener una colonia diferenciada. Seguidamente es necesario hacerlas competentes, es decir, prepararlas para que puedan incorporar la molécula de ADN bicatenaria. El procedimiento es el siguiente: se elige una colonia de las placas crecidas y se inocula un tubo con 5 mL de medio LB que se crece durante toda la noche a 37°C.

Al día siguiente, se transfiere 1 mL de cultivo crecido durante la noche a un matraz de 250 ml con 50 ml de medio LB y se incuba a 37°C hasta que alcanza una absorbancia de 0.6-0.7 a 660 nm. A continuación se centrifugan durante 10 minutos a 8000 r.p.m. (4500 g con rotor M1). Se tira el sobrenadante y las células obtenidas se resuspenden, suavemente, en 5 ml de CaCl₂ 50 mM estéril, mantenido a 4°C. Se incuban en hielo durante 30 minutos y se vuelven a centrifugar 10 minutos a 3500 r.p.m. después de tirar el sobrenadante, se resuspenden nuevamente en 5 ml de CaCl₂ 50 mM y se mantienen en hielo durante toda la noche. Estas células guardadas en el frigorífico a 4°C conservan su competencia durante 4 ó 5 días, aunque ésta vaya disminuyendo progresivamente.

2.2.2.2 Transformación de células competentes JF521

Se toma una alícuota de 200µl de células competentes JF521 a la que se añade 2 µl de ADN. A partir de este punto, los tiempos de incubación y el choque térmico son los mismos que los descritos en el apartado 2.2.1.4 para las células supercompetentes XLI-Blue. La mezcla resultante de la transformación se crece, realizando diferentes diluciones, durante toda la noche en placas con medio sólido LB-Agar más kanamicina (40 µg/ml) a 37°C.

2.2.2.3 Transformación de células supercompetentes BL21-DE3

Los genes de tiorredoxina silvestre y los mutantes ($I \rightarrow V$ y $V \rightarrow I$) se encuentran clonados en el vector pET30a (+). Su expresión se realizó en células BL21-DE3. Este vector es más pequeño que pTK100 (pET30a-trxA consta de 5568 pb frente a 9600 pb de pTK100-trx), con lo cual se acorta la duración de los experimentos de mutagénesis dirigida.

El protocolo a seguir para la transformación de las células supercompetentes BL21-DE3 es similar al usado para las células supercompetentes XLI-Blue (sección 2.2.1.4) con excepción de los tiempos de incubación y condiciones del choque térmico. En este caso, se realiza una incubación en hielo de 5 minutos seguida del choque térmico (30 segundos a 42°C), más 2 minutos de incubación en hielo.

2.2.2.4 Pruebas de expresión de proteína

La cantidad de proteína producida tanto por las células competentes JF521 como por las células supercompetentes BL21-DE3 transformadas ha sido evaluada

mediante la realización de electroforesis en condiciones desnaturalizantes (*PAGE-SDS*). La composición de todos los geles hechos en este trabajo es del 16.5%T y 3%C en el gel separador y del 10%T y 3%C en el gel apilador.

La metodología a seguir varía dependiendo del vector con el que estemos trabajando: pTK100 o pET30a (+). Se pretende determinar qué colonia produce más tiorredoxina, para ello se procede como se explica a continuación: de las placas crecidas durante toda la noche se eligen unas 10 colonias y se inoculan 10 tubos que contienen 5 ml de medio líquido LB más kanamicina (40 µg/ml) como se describe en el apartado 2.2.2.1. Se dejan crecer en un baño con agitación a 37°C durante toda la noche. Sólo para el caso en el que estemos trabajando con pET30a (+), hay que inducir los cultivos con IPTG (ver detalles en sección 2.2.2.5) y después se dejan creciendo toda la noche a 37°C. Al día siguiente se toman 100 µL de cada uno de esos tubos y se centrifugan. Se elimina el sobrenadante resultante, se resuspende el pellet en 15 µL de agua Milli-Q y se agita la mezcla aproximadamente 1 minuto en vortex. Finalmente, se añaden 15 µL de tampón de carga de la muestra, se da un pulso de centrifuga y se cargan las muestras en el gel. Es aconsejable cargar una calle del gel con una muestra de tiorredoxina o un indicador de pesos moleculares, con objeto de identificar más fácilmente la banda correspondiente a tiorredoxina. Terminada la electroforesis se deja el gel en tinción durante 1 hora aproximadamente y después se pasa a la disolución de destinción y finalmente a la de fijación (ver Figura 2.3-B).

2.2.2.5 Expresión de proteína a gran escala

El siguiente paso es expresar tiorredoxina a gran escala, una vez elegida la colonia que más expresa tiorredoxina (ésto nos lo indica la calle del gel de electroforesis que contenga la banda más intensa correspondiente a tiorredoxina. Se han seguido dos protocolos dependiendo del vector con el que se ha trabajado. Ambos protocolos se describen a continuación.

A) Expresión de proteína utilizando pTK100:

Se inocula la colonia elegida en un tubo con 5 mL de LB más kanamicina a 40 µg/mL (preinóculo). Se crece en un baño orbital a 37°C hasta que su absorbancia a 660 nm es de 1.0, pues a este valor nos encontramos en la fase exponencial de crecimiento. La absorbancia no debe pasar de 1.0. Alcanzada la

absorbancia deseada se transfieren los 5 mL de preinoculo a un matraz de 250 mL con 50 mL de medio LB más kanamicina a 40 $\mu\text{g}/\text{mL}$ (inóculo) y se crece durante 2-3 horas en baño con agitación hasta que alcance una absorbancia de 0.9.

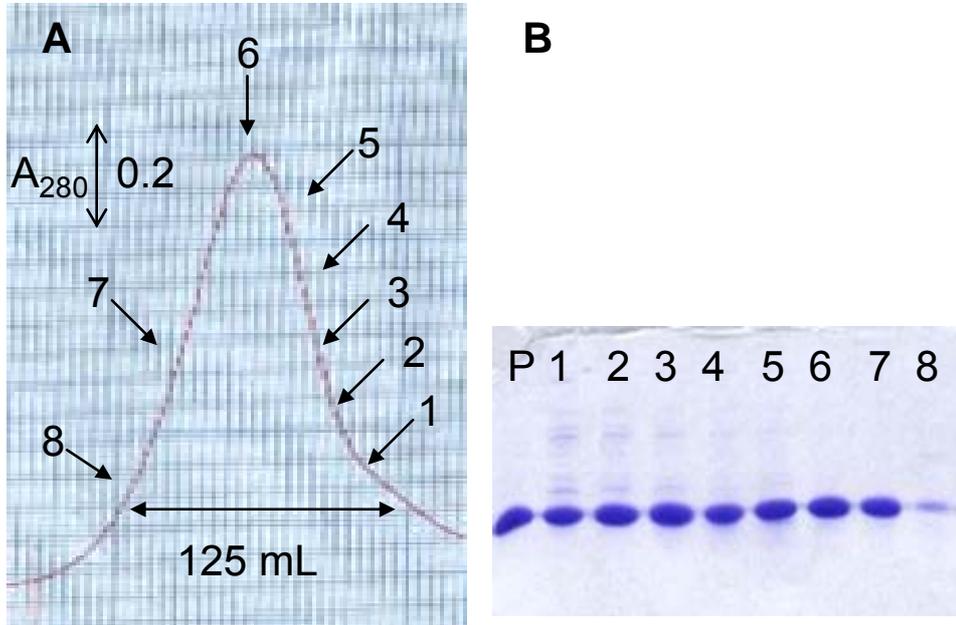


Figura 2.3: A) Pico del mutante de tiorredoxina E44D en la elución de la columna de exclusión molecular. Se corresponde con el segundo pico de la elución. B) Gel de poliacrilamida en presencia de SDS que muestra la electroforesis discontinua correspondiente a la elución de las fracciones (1-8) de la cromatografía A). Como patrón de peso molecular (P) se ha utilizado una muestra de tiorredoxina silvestre.

Hay que evitar que la absorbancia sea mayor que 1.0. Cuando el inóculo ha alcanzado la absorbancia deseada se transfieren 5 mL a un matraz de 2 L con 750 mL de medio TB más kanamicina a 40 $\mu\text{g}/\text{mL}$ y se crece en baño con agitación durante toda la noche. Al día siguiente, comienza la etapa de separación de tiorredoxina.

B) Expresión de proteína utilizando pET30a (+):

Se inocula la colonia elegida en un tubo con 5 mL de LB más kanamicina a 40 $\mu\text{g}/\text{mL}$ (preinóculo). Se crece en un baño orbital a 37°C hasta que su absorbancia a 660 nm es de 0.8, entonces se transfiere 1 mL de preinóculo a un matraz de 250 mL con 50 mL de medio LB más kanamicina a 40 $\mu\text{g}/\text{mL}$ (inóculo)

y se crece durante 1-2 horas en incubador con agitación hasta que alcance una absorbancia de 0.8. Cuando el inóculo ha alcanzado dicha absorbancia se transfieren 5 mL a un matraz de 2 L con 500 mL de medio LB más kanamicina a 40 µg/mL y se crece en el incubador con agitación hasta que alcanza una absorbancia de 0.6. En ese momento, es necesario inducir el cultivo con 0.5 mL de IPTG 400 mM y se deja crecer toda la noche. Al día siguiente, comienza la etapa de separación de tiorredoxina.

2.2.2.6 Purificación de la proteína

Son necesarios 4 pasos para separar la proteína. El siguiente protocolo describe con detalle estos cuatro pasos.

A) Lisis celular y eliminación de ácidos nucleicos

Los cultivos crecidos durante la noche se centrifugan a 8000 r.p.m. durante 10 minutos (centrífuga Kokusan o Hettich). El sobrenadante se elimina y el precipitado celular o *pellet* se conserva a 4 °C. Se añaden 3 mL de tampón TE por 1 gr de *pellet* y se resuspende dicho *pellet* en el tampón Tris-EDTA 30:1 mM (TE). La disolución resultante la pasamos por la prensa French dos veces. La prensa somete a las células en su interior a una elevada presión (1200 p.s.i.) que las paredes celulares resisten. Al salir, por diferencia de presión, tiene lugar la ruptura de las paredes celulares, obteniéndose un homogeneizado formado por restos celulares más ácidos nucleicos, proteínas en disolución. Para eliminar los restos de paredes celulares, se centrifuga ese homogeneizado a 8000 r.p.m. durante 10 minutos. Ahora se aprovecha la disolución sobrenadante y se elimina el precipitado. El siguiente paso será la precipitación del ADN y ARN presentes en el sobrenadante con una disolución de sulfato de estreptomicina (SE) al 10 % preparada en TE. La cantidad de disolución de sulfato de estreptomicina que hay que preparar se calcula con la siguiente regla:

$$X = 0.25 \cdot S \quad (2.2)$$

donde: X representa el volumen (mL) de disolución de sulfato de estreptomicina y S representa el volumen (mL) de disolución sobrenadante.

La disolución de SE al 10% preparada se transfiere a un embudo de decantación y se sitúa encima del vaso que contiene la disolución sobrenadante, permitiendo un goteo lento y constante durante toda la noche a una temperatura de 4°C. Además, en el vaso debe existir una agitación suave y permanente durante toda la precipitación. De esta forma, aparecerá un precipitado blanco y poco consistente que al día siguiente se centrifuga en la ultracentrífuga a 20000 r.p.m., durante 30 minutos a 4°C. La disolución sobrenadante se filtra con filtros de 0.45 µm para eliminar pequeñas partículas que no hayan precipitado mediante la centrifugación. Llegados a este punto, la columna de exclusión molecular tiene que estar equilibrada con un volumen de TE doble al volumen del gel que contiene, es decir, unos 4 L de TE. El volumen máximo de disolución de proteína que se puede inyectar en la columna es de 150 mL.

B) Cromatografía de filtración en gel

En este punto partimos de una disolución de varias proteínas y otros componentes (sales, lípidos o polisacáridos). Se ha elegido la cromatografía de filtración en gel o exclusión molecular como primer paso para separar tiorredoxina de otras proteínas de mayor peso molecular, lípidos, polisacáridos u otras impurezas. Como soporte se ha utilizado Sephacryl S-100 de AMERSHAM PHARMACIA BIOTECH AB. Es una matriz de polímeros (poliacrilamida/dextrano) fuertemente hidrofílicos, que al ser puestos en contacto con el agua, adquieren el aspecto de un gel. Las micropartículas en que están divididos estos soportes son esferas formadas por enrejados de fibras del polímero correspondiente. Los huecos o poros de estos enrejados tendrán un tamaño determinado por la concentración de fibras y su grado de entrecruzamiento. Las micropartículas de gel al ponerse en contacto con el agua se hinchan considerablemente, de tal forma que se puede estimar que un 80 % del lecho cromatográfico empaquetado es líquido. Cuando el eluyente (TE) comienza a fluir arrastra el líquido de los intersticios de estas micropartículas pero no el líquido atrapado por éstas ya que se encuentra protegido por el entramado de fibras.

En el HPLC o en la bomba peristáltica, se fija una velocidad de flujo de 5 mL/min, una vez introducida toda la disolución de proteína en la columna, se sigue circulando tampón TE a una velocidad de 5mL/min. La parte superior de la columna se encuentra conectada a una lámpara de ultravioleta y ésta a su vez a un

registro gráfico, que detectará la salida de cualquier componente, ya sea impureza o proteína (ver Figura 2.3A). El colector de fracciones, se programa para recoger fracciones de 10 mL. De esta forma, cualquier componente registrado por su absorbancia con la lámpara de ultravioleta, está fácilmente identificado.

La pureza de las fracciones correspondientes al pico de tiorredoxina se ha evaluado mediante la realización de *PAGE-SDS*. El protocolo a seguir se explica a continuación: se toman 15 μL de disolución de proteína eluida y se le añaden otros 15 μL de tampón de carga. La mezcla resultante la cargo en gel de poliacrilamida (ver Figura 2.3 A).

Una vez secado el gel, el siguiente paso es hacer una densitometría. De esta forma se evalúa la pureza de las fracciones de disolución de tiorredoxina eluidas en la cromatografía de exclusión molecular. Llegados a este punto la columna de intercambio iónico tiene que estar equilibrada para seguidamente pasar la disolución de tiorredoxina como se describe en la siguiente sección.

C) Cromatografía de intercambio iónico

En este tipo de cromatografía la fase estacionaria tiene grupos cargados, anclados a la resina, que interaccionan electrostáticamente con iones de signo contrario de la fase móvil. Las resinas se caracterizan por tener muy buenas propiedades de flujo y una elevada capacidad de intercambio iónico. Por eso son muy útiles para la eliminación de impurezas iónicas de reactivos, o la eliminación de cationes metálicos de tampones y reactivos.

El protocolo a seguir se describe a continuación: las fracciones resultantes de la cromatografía de filtración en gel de mayor pureza, se unen para pasarlas por la columna de intercambio iónico. Ésta tiene un diámetro de 2.6 cm y una longitud de 21cm. Se ha empaquetado utilizando como soporte una resina intercambiadora de iones de tipo débil, denominada Fractogel EMD DEAE-650 (el grupo cargado es DiEtilAmino Etilo). La disolución de proteína es succionada por una bomba peristáltica a una velocidad de flujo de 2 mL/min.

Seguidamente la columna se equilibra con 200 mL de TE para posteriormente comenzar con el gradiente de elución de la proteína. Al pH de trabajo la disolución de proteína tendrá una carga neta positiva. Al viajar a través de la resina, ésta intercambiará el contraión Na^+ . Se puede escribir la siguiente ecuación de intercambio iónico:



El equilibrio anterior está gobernado por la constante:

$$K = \frac{[\text{Resina} - \text{P}][\text{Na}^+]}{[\text{Resina} - \text{Na}^+][\text{P}^+]} \quad (2.4)$$

Cuyo valor concreto está dictado por las afinidades de la fase estacionaria por cada uno de los iones en juego, P^+ y Na^+ . Las variaciones del pH del medio que alteran la carga neta de la proteína y la concentración de Na^+ en el medio, podrán desplazar el equilibrio en un sentido o en otro.

El gradiente de concentración se consigue con un sistema de vasos comunicantes, uno de ellos contiene una disolución de NaCl 0.5 M en TE pH 8.3 y el otro, que está directamente conectado a la columna, sólo tampón TE. La concentración de NaCl en éste último aumenta progresivamente. El gradiente de sal se introduce en la columna a una velocidad de flujo de 2 mL/min. Si establecido el anterior equilibrio (ecuación 2.3), se pasa por la columna una fase móvil con una concentración muy elevada de Na^+ (fuerza iónica elevada), el equilibrio 2.3 se desplazará a la izquierda, con resultado de la liberación de la proteína antes retenida. Para que la liberación se produzca es necesario pasar una elevada concentración de ion (Na^+ u otro ion). En definitiva, el equilibrio de intercambio iónico (ecuación 2.3) puede considerarse como una competencia entre los iones que se desplazan por la fase móvil por unirse a los grupos cargados del intercambiador. Cuanto mayor sea la concentración de un ion dado en la fase móvil, más grupos cargados del intercambiador ocuparán, desplazando al ion que se encuentra en minoría.

Como resultado se recogen fracciones de proteína a una concentración mucho mayor. Para determinar la pureza de dichas fracciones, nuevamente se realiza una electroforesis y se evalúa mediante una densitometría.

Después de la cromatografía de intercambio iónico, la proteína está casi lista para poder ser utilizada. El último paso es realizar una diálisis exhaustiva en el tampón que se van a hacer los experimentos.

2.2.3 PREPARACIÓN DE LAS MUESTRAS

2.2.3.1 Preparación de los tampones y diálisis

El primer paso de la diálisis consiste en la preparación del tampón en el que va a ir disuelta la proteína. El procedimiento general que se ha seguido en este trabajo consiste en prepararlo a partir de la mezcla de dos disoluciones de igual concentración de tampón y de pHs superior e inferior al deseado, respectivamente. El pH más básico lo da el propio tampón y el ácido se consigue añadiendo un volumen de la forma ácida del tampón o en su defecto HCl.

Los detalles acerca de los diferentes tampones utilizados, pHs y condiciones de trabajo en general, se encuentran en las secciones 2.2.8.3, 3.1.2, 4.1.2 y 5.4.2.

Como ya se ha indicado en la sección 2.2.2.6, las proteínas purificadas utilizadas en el presente trabajo (tioredoxina *WT* y sus distintas formas mutantes) se encuentran disueltas en tampón de elución correspondiente a la purificación de intercambio iónico. Por lo tanto, el primer paso en la preparación de la diálisis es tomar con ayuda de una micropipeta el volumen apropiado de disolución de proteína dependiendo de cuál sea su concentración. Lo introducimos en la membrana de diálisis preparada previamente como se explica en el apartado 2.1.1. La bolsa se cierra mediante pinzas especiales (de Spectrum Medical Industries) dejando una pequeña cámara de aire. Finalmente, se introduce en un recipiente que contiene un volumen de tampón 500 a 600 veces mayor al de la muestra y se pone a agitar suavemente durante unas 8 horas a 4°C. Transcurrido este tiempo se repone el tampón que baña la bolsa de diálisis por otro volumen fresco y de nuevo dejamos en agitación otras 8 horas aproximadamente. El proceso se repite una tercera vez de manera que transcurrido el tiempo indicado extraemos la disolución dializada de proteína de la bolsa. Con objeto de eliminar cualquier residuo insoluble centrifugamos dicha disolución en la microcentrífuga (un pulso a máxima velocidad) o la filtramos utilizando filtros de tamaño de poro de 0.45 µm.

La concentración de proteína en disolución se ha determinado espectrofotométricamente registrando el espectro ultravioleta-visible. El coeficiente de extinción utilizado para tioredoxina silvestre y sus formas mutantes ha sido 14000 M⁻¹cm⁻¹ a 280 nm (Langstemo *et al.*, 1989). Para registrar el espectro se toma una alícuota de la disolución de proteína y se diluye de manera

que su absorbancia a 280 nm sea próxima a la unidad. De esta forma nos aseguramos estar en el rango lineal de la dependencia de la absorbancia con la concentración. Sistemáticamente se realizaron tres medidas con diferentes diluciones para tomar el valor medio de absorbancia a 280 nm. En la Figura 2.4 se muestra un espectro ultravioleta-visible típico para tiorredoxina silvestre.

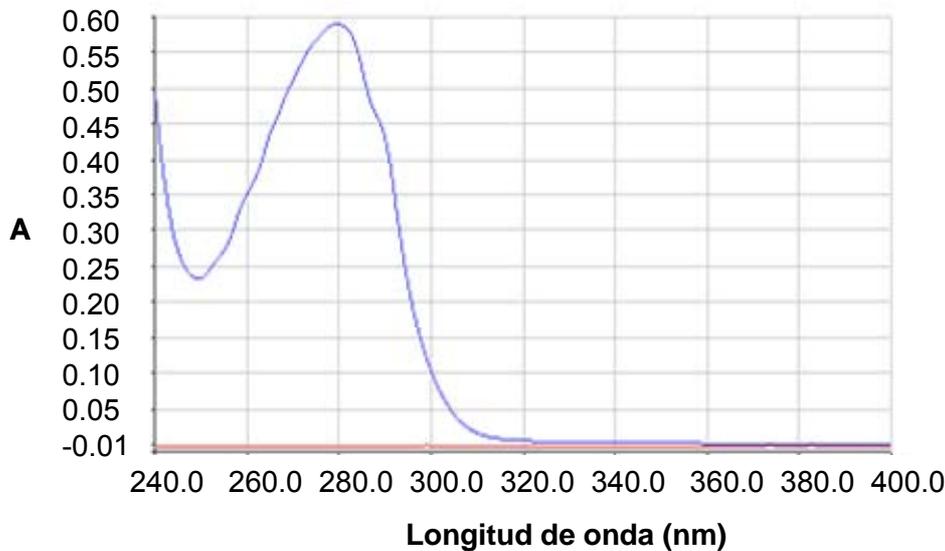


Figura 2.4: Espectro ultravioleta-visible de tiorredoxina silvestre de *E. coli* en tampón Hepes 5mM pH 7.0. La concentración de proteína es de 0.50 mg/mL y para la obtención del espectro se utilizó una cubeta de cuarzo de 1 cm de paso de luz.

2.2.3.2 Preparación de disoluciones concentradas de desnaturizante

Las disoluciones de urea descomponen lentamente originando iones amonio y cianato (Hagel *et al.*, 1971). Estos últimos pueden interactuar con los grupos amino de la proteína (Stark, 1965) de manera que es aconsejable usar una disolución fresca de urea, preparada el mismo día que se van a hacer los experimentos. A continuación se detallan los pasos seguidos en la preparación de una disolución madre de urea, aproximadamente 10 M, en tampón HEPES 5 mM, pH 7:

En primer lugar se pesa la cantidad de necesaria de urea para disolverla después en un volumen de agua Milli-Q algo menor al final. Dado que estamos muy cerca de la saturación (la solubilidad de urea a 25°C es 10.49 M) y con objeto de acelerar el proceso de solubilización, calentamos ligeramente sin llegar a superar los 30°C ya que se favorecería la descomposición. A continuación pasamos la disolución a través de la resina de intercambio iónico como se mencionó en la sección 2.1.2, recogiendo diferentes fracciones de disolución de urea. Éste proceso se repite dos o tres veces. Una vez purificada la urea, preparamos tampón HEPES 200 mM a pH 7 según se ha descrito en la sección anterior 2.2.3.1. A cada fracción de urea vamos se añade el volumen necesario de HEPES 200 mM para tener una concentración final de HEPES 5 mM. Ayudándonos de un matraz aforado enrasamos el volumen final de cada fracción con agua o más disolución de urea. Después se mide el pH para confirmar que no hay cambios significativos ya que la presencia de desnaturalizantes influye en las medidas de pH (Acevedo *et al.*, 2002). Para finalizar, se filtra la disolución resultante con un filtro de 0.45 µm.

Para determinar la concentración exacta de las disoluciones de urea se utiliza la medida del índice de refracción, n . La diferencia entre el índice de refracción de la disolución de urea y el tampón acuoso en el que se disuelve, $\Delta n = n_{urea} - n_{tampón}$, nos dará un Δn el cual está relacionado con la concentración de urea mediante la expresión de Warren y Gordon, 1966:

$$[\text{urea}] = 117.66 \cdot \Delta n + 29.753 \cdot (\Delta n)^2 + 185.568 \cdot (\Delta n)^3 \quad (2.5)$$

2.2.4 EXPERIMENTO DE CALORIMETRÍA DIFERENCIAL DE BARRIDO

2.2.4.1 Generalidades

La calorimetría diferencial de barrido (DSC, *Differential Scanning Calorimetry*) es la técnica más poderosa para la caracterización energética de los cambios conformacionales inducidos por temperatura en macromoléculas biológicas tales como proteínas, ácidos nucleicos y biomembranas. Existen diversas revisiones sobre los aspectos más interesantes de esta técnica (Privalov, 1979, Privalov, 1980, 1982, 1989, Sanchez-Ruiz y Mateo, 1987, Chowdry y Cole,

1989, Freire *et al.*, 1990, Cooper y Johnson, 1994, Sanchez-Ruiz, 1995, Ibarra-Molero y Sanchez-Ruiz, 2006, 2006).

Mediante un experimento de calorimetría diferencial de barrido es posible obtener la capacidad calorífica molar aparente de una proteína o de cualquier macromolécula como función de la temperatura. El posterior análisis de esta dependencia nos dará una caracterización termodinámica completa de la transición que ha tenido lugar por efecto de la temperatura. En un termograma típico de DSC ésta se manifiesta por un pico, correspondiente a un proceso endotérmico de acuerdo con el Segundo Principio de la Termodinámica (supuesto el proceso de equilibrio), tanto más agudo cuanto más cooperativa sea la transición.

En general, de los experimentos de DSC analizados según la termodinámica de equilibrio podemos obtener información acerca de:

1. La capacidad calorífica parcial absoluta de una molécula.
2. La función de partición y consecuentemente, la población de estados intermedios y los parámetros termodinámicos [cambio de entalpía (ΔH), cambio de entropía (ΔS) y cambio de la capacidad calorífica (ΔC_p)] asociados a la transición inducida por la temperatura.

El calorímetro que se ha usado ha sido el VP-DSC. Este calorímetro tiene una elevada sensibilidad y reproducibilidad, de manera que es posible obtener resultados de calidad con unos 50 μg de proteína en la célula de muestra. Las células tienen formas de moneda y están hechas de tántalo, material resistente a la corrosión de ácidos pero que puede deteriorarse por la acción de bases fuertes. El volumen efectivo de las células es de 0.5 mL aproximadamente y se encuentran dispuestas en el interior de una coraza cilíndrica que asegura la adiabaticidad, accediéndose a ellas a través de unos capilares de 1.5 mm de diámetro interno. El VP-DSC opera en un rango de temperaturas que va desde -10 a 130 $^{\circ}\text{C}$. Es capaz de registrar barridos tanto calentando como enfriando (en este caso en modo no adiabático) usando un mecanismo interior Peltier, de manera que no se requieren dispositivos externos de calentamiento o enfriamiento. Las velocidades de barrido son seleccionadas por el usuario dentro del rango de 0 a 1.5 K/min. Se dispone de un sistema mecánico de presurización que alcanza hasta 2 atmósferas.

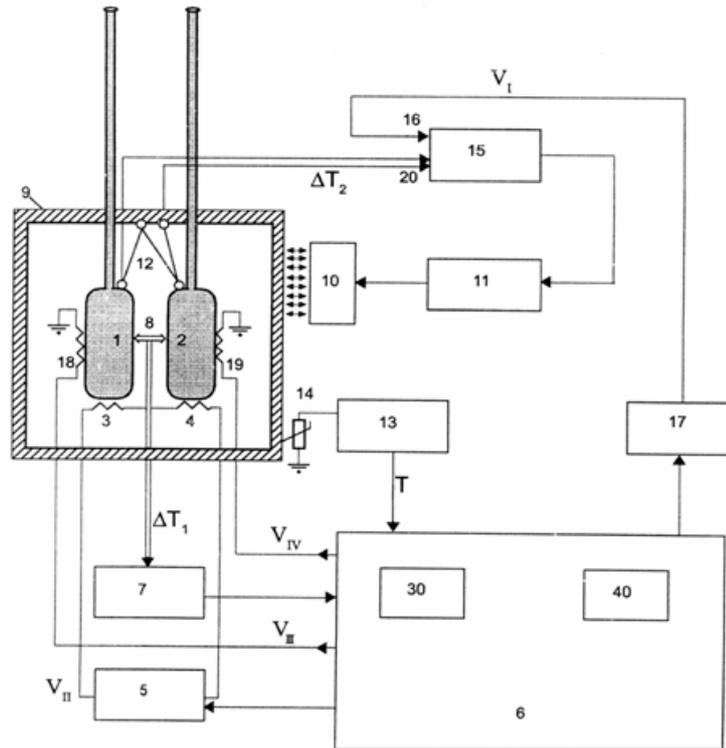


Figura 2.5: Esquema del calorímetro diferencial de barrido VP-DSC: (1,2) Células de muestra y referencia con los tubos capilares de entrada para el llenado y limpieza; (3,4) calentadores principales situados sobre las células, controlados por el voltaje V_{II} de la fuente de alimentación (5). Ésta a su vez está conectada al computador (6) a través de una tarjeta convertora analógico/digital (A/D) (30), (18,19) elementos auxiliares de calentamiento, activados a través de la tarjeta A/D (30), utilizados en la calibración y en el modo de retroalimentación; dispositivo de medida del efecto térmico (7) y sensor de cristal (8) para medir la diferencia de temperatura ΔT_1 entre las dos células. Rodeando a las dos células se encuentra una coraza (9) con dispositivos de calentamiento/enfriamiento (10) dirigidos por un controlador (11), que responde a una señal que procede de un sistema de amplificación (15). El amplificador recibe, a su vez, una señal entrante en un terminal (20) procedente del sensor (12) que mide la diferencia de temperatura ΔT_2 entre las células y la coraza. El otro terminal (16) recibe un voltaje V_{BIB} de la fuente de alimentación (17), controlada por el computador a través de la tarjeta A/D (30). La coraza tiene además un dispositivo (13) para la medida de temperaturas absolutas activado mediante un sensor (14) dispuesto en la coraza térmica, de manera que la señal que sale de él pasa por el computador a través de la tarjeta A/D (30). Las señales calibradas de entrada al computador (que corresponden a ΔT_1 y ΔT_2), así como la temperatura absoluta T son seguidas continuamente durante el experimento y almacenadas en la memoria del computador (40) a intervalos señalados por el operador (Plotnikov *et al.*, 1997).

En la Figura 2.5 se muestra un esquema del instrumento (ver pie de figura para más detalles). La repetitividad de líneas base sucesivas sin rellenar es 2 $\mu\text{cal}/\text{min}$ y rellenando es 5 $\mu\text{cal}/\text{min}$ (ver Figura 2.6). Además, es posible trabajar en modo isoterma durante largos periodos de tiempo, lo que resulta muy conveniente si se quiere seguir cinéticas a diferentes temperaturas o estudiar la estabilidad de drogas u otros compuestos químicos. En el VP-DSC el tiempo de respuesta es seleccionado por el operador. Así para el caso de transiciones anchas, como sucede con la mayoría de las proteínas, en las que se requiere una alta sensibilidad para resolver el pico, es aconsejable usar tiempos de respuesta mayores. Por el contrario para transiciones agudas del tipo gel-cristal líquido en lípidos se aconseja un tiempo de respuesta corto para evitar distorsionar la forma del termograma. No obstante el aumento en el ruido, inherente al uso de tiempos de respuesta menores, va a reducir la calidad de los datos. Para detalles adicionales acerca del diseño y características del VP-DSC ver Plotnikov *et al.*, 1997.

El operador accede al control del VP-DSC a través del programa VPViewer. Aquí es donde podemos modificar los parámetros experimentales (temperatura de inicio y final, velocidad de barrido...) así como parar o iniciar un barrido, adquirir y almacenar los datos y presentarlos a tiempo real.

2.2.4.2 Experimento calorimétrico

A) Preparación de la experiencia:

La muestra de proteína se dializa frente al tampón correspondiente como se indica en el apartado 2.2.3.1. Debemos ser rigurosos con la preparación de la muestra de manera que podamos garantizar su pureza e integridad. La concentración de ésta debe determinarse de una manera precisa, ya que un error en la concentración se reflejará directamente en la medida de los distintos parámetros termodinámicos que se derivan del termograma. Por ejemplo, un error del 5% en la concentración introducirá un 5% de error en la determinación del cambio de entalpía. El tampón del diálisis se utiliza como referencia. Es importante, antes de comenzar el experimento calorimétrico que las disoluciones de partida hayan sido previamente filtradas y se les haya comprobado su pH.

Antes de introducir en el calorímetro las disoluciones, tanto la de muestra como la de referencia, es necesario desgasificarlas en vacío. El tiempo puede variar dependiendo del volumen pero normalmente oscila en torno a 5 minutos en el caso del tampón y 3 para la disolución de proteína. Con ésta hay que ser especialmente cuidadoso ya que, si el proceso de desgasificación se alarga demasiado, podríamos estar modificando la concentración de proteína en disolución. El volumen mínimo requerido para rellenar el calorímetro es 1 mL aproximadamente.

B) La línea base instrumental:

Idealmente si las dos células del calorímetro fuesen idénticas, bastaría un solo barrido para determinar la diferencia en capacidad calorífica entre las soluciones de muestra y referencia. Pero en la práctica, las dos células no son exactamente iguales y por tanto es necesario registrar por lo menos dos barridos para sustraer y eliminar la contribución a la señal debida a aspectos instrumentales.

Así pues, en primer lugar registramos la línea base instrumental que coincide con el perfil obtenido cuando las dos células están llenas con el tampón en el que se ha dializado y diluido la muestra. Las primeras líneas base no son reproducibles y para obtener resultados óptimos es preciso hacer varios barridos hasta alcanzar la completa reproducibilidad. Dicha línea base tampón-tampón se le restará al perfil proteína-tampón con objeto, como ya se ha indicado, de eliminar las distorsiones instrumentales (Figura 2.6). Esta sustracción da lugar, dentro de la mayor parte del rango de temperaturas registrado, a valores negativos de capacidad calorífica de la proteína indicando que la capacidad calorífica de la disolución de referencia es mayor que la de la muestra. Esto es lógico si tenemos en cuenta que el contenido en agua de esta última es menor (la proteína ha desplazado al agua) y que la capacidad calorífica del agua líquida es mayor que la de la proteína.

El VP-DSC actúa cíclicamente, por esta razón es necesario registrar varias líneas base de tampón y es importante no romper la historia térmica del aparato a la hora del llenado o del lavado de las células, de manera que se recomienda hacerlo en la fase de enfriamiento y por debajo de los 30°C (próximo a la temperatura ambiente). El mayor grado de solapamiento entre sucesivos barridos,

con o sin rellenar las células, se obtiene cuando el instrumento permanece en el mismo ciclo térmico (igual velocidad de barrido, igual temperatura de inicio y final de barrido, el mismo tiempo de equilibración entre el final de un barrido y el comienzo del siguiente) durante todos los experimentos que serán comparados.

C) Barrido de la muestra:

Un aspecto importantísimo antes de comenzar un nuevo experimento es asegurarnos de que la historia térmica del instrumento se ha establecido. El mayor grado de solapamiento entre sucesivos barridos, con o sin rellenar las células, se obtiene cuando el instrumento permanece en el mismo ciclo térmico (igual velocidad de barrido, igual temperatura de inicio y final de barrido, el mismo tiempo de equilibración entre el final de un barrido y el comienzo del siguiente) durante todos los experimentos que serán comparados.

Así pues, antes de pasar al barrido de la muestra debemos registrar al menos dos líneas base tampón-tampón sucesivas (sin rellenar la células) de manera que la diferencia entre ambas se encuentre dentro de las especificaciones del instrumento ($\pm 2 \mu\text{cal/grado}$). Ésto será indicativo de que el aparato se encuentra estable y que su historia térmica no se ha interrumpido.

Para registrar el termograma correspondiente a la proteína en estudio basta rellenar la célula de muestra con nuestra disolución, dejando en la célula de referencia el tampón con el que se realizó la línea base instrumental (que es el mismo en el que va disuelta la muestra; ver Figura 2.6).

La operación de vaciado de la célula de muestra y posterior llenado con la disolución de proteína conviene hacerla lo más rápido posible con objeto de tener el bloque calorimétrico destapado poco tiempo y así evitar irreproducibilidades. Para eliminar cualquier burbuja que hayamos podido introducir durante el llenado, empujamos el émbolo de la jeringa hacia abajo en movimientos cortos pero intensos, de manera que el flujo del líquido a presión que entra desplace a cualquier burbuja hacia fuera de la célula. Para conocer exactamente la concentración de proteína en el interior del calorímetro y teniendo en cuenta que se produce una dilución del 1 o 2% de la muestra en la célula, al no estar ésta completamente seca, lo más conveniente es registrar el espectro ultravioleta-visible de la solución sobrante de llenar la célula, cuya composición es igual a la del interior de ésta.

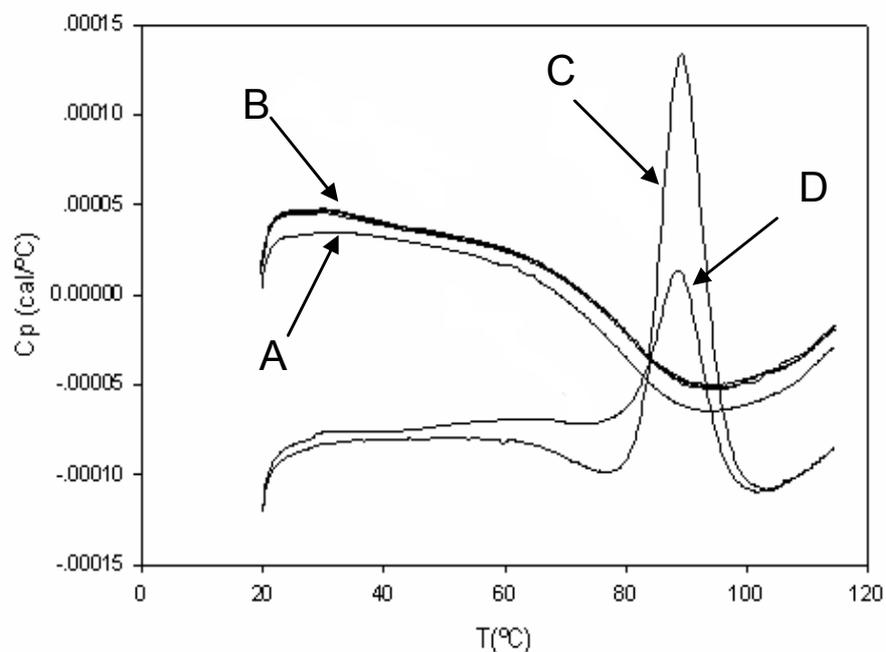


Figura 2.6: A) Primer barrido tampón-tampón; B) Cinco primeros barridos tampón-tampón Hepes 5 mM pH7; C) Barrido con célula de muestra llena con disolución de tiorredoxina silvestre concentración 0.5 mg/mL; D) Segundo barrido con célula de muestra llena con disolución de tiorredoxina.

Para nuestra proteína, un perfil típico de DSC muestra un pico (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. En la Figura 2.7 se indican algunos de los parámetros más significativos en una transición calorimétrica.

D) Segundo barrido de la muestra:

Generalmente la reversibilidad de una muestra se expresa en términos del porcentaje de área bajo la curva que se recupera en el segundo barrido de la misma (Figura 2.6-D). Éste es un aspecto muy importante ya que no se puede aplicar un análisis basado en la termodinámica de equilibrio si la muestra presenta una transición irreversible y cinéticamente controlada.

Una vez acabados los experimentos se deben lavar las células inmediatamente con agua Milli-Q en abundancia para evitar que se adhieran a las paredes de éstas posibles agregados que se hayan formado. Es conveniente limpiar con cierta periodicidad con ácido fórmico 10% o con ácido nítrico 63% en los casos que pueda existir agregación.

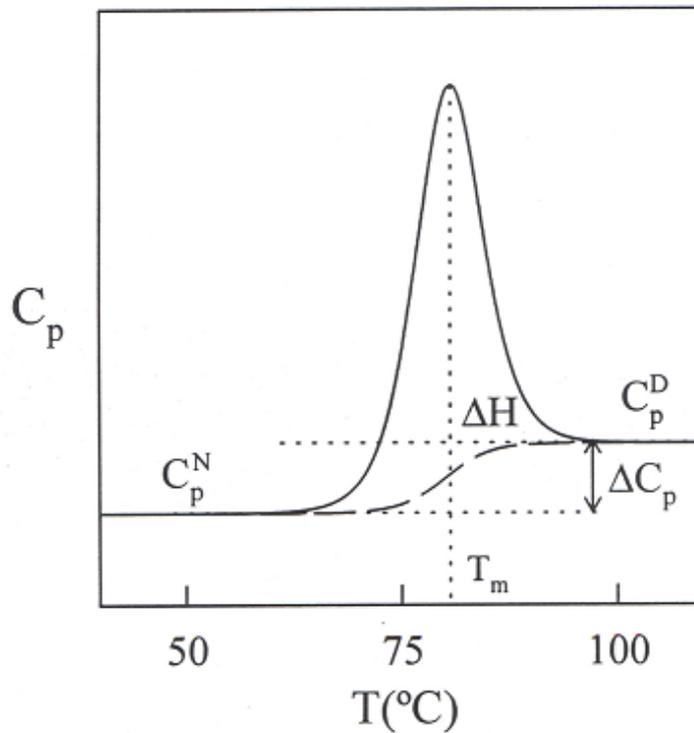


Figura 2.7: Curva de capacidad calorífica (trazo continuo) en función de la temperatura calculada para una proteína hipotética que despliega según un modelo de equilibrio de dos estados. C_p^N representa la capacidad calorífica del estado nativo y C_p^D la del estado desplegado. En trazo discontinuo se representa la línea base química (línea que conecta de un modo suave los valores de capacidad calorífica pre- y postransicionales, y se puede usar como referencia para calcular la capacidad calorífica de exceso). El área encerrada entre las dos curvas equivale a la entalpía de desnaturalización, ΔH . Los parámetros utilizados en la simulación son: $\Delta C_p = C_p^D - C_p^N = 7 \text{ kJ}\cdot\text{K}^{-1} \text{ mol}^{-1}$; $\Delta H = 400 \text{ kJ}\cdot\text{mol}^{-1}$ y $T_m = 80^\circ\text{C}$.

2.2.4.3 Análisis de los datos de DSC según la termodinámica de equilibrio

Los ficheros generados por el programa de adquisición de datos contienen el valor de una señal proporcional a la variación de la capacidad calorífica de la disolución de proteína (en mV) con la temperatura.

Los ficheros se importan al programa MicroCal Origin para DSC v 4.1 y una vez aquí se resta la línea base instrumental y se normalizan los datos según la velocidad de barrido y la concentración de proteína presente en el experimento. Las trazas obtenidas se analizan con el programa comercial MLAB introduciendo la subrutina de ajuste que se describirá en el Apéndice I (Programa 1).

Bajo ciertas condiciones experimentales, la desnaturalización de muchas proteínas globulares simples y algunas complejas es reversible (Privalov, 1979, Privalov, 1982), recuperándose en gran medida la estructura nativa tras enfriar la disolución. Es en estos casos (tiorredoxina WT y todas las formas mutantes estudiadas en este trabajo de investigación tienen una reversibilidad mayor al 90%) en los que es posible aplicar la formulación basada en la termodinámica de equilibrio, suponiendo que en todo momento existe equilibrio entre los diferentes estados poblados de la proteína. Normalmente, se acepta que un proceso es de equilibrio cuando se obtiene reversibilidad calorimétrica (para una discusión más detallada ver Sanchez-Ruiz, 1995, Plaza del Pino et al., 2000, Ibarra-Molero y Sanchez-Ruiz, 2006, 2006).

Una vez comprobado que un sistema es caloriméricamente reversible se intenta ajustar el modelo termodinámico más sencillo posible a los perfiles obtenidos. El modelo seguido para el análisis de datos en este trabajo de investigación se describe a continuación.

A) Modelo de equilibrio de dos estados

Es el mecanismo de desnaturalización más simple, y es aquél en el que sólo se encuentran poblados de forma significativa los estados nativo (N) y desplegado (D) de la proteína.



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

$$K = \frac{[D]}{[N]} \quad (2.7)$$

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:

$$\Delta_N^D J = J(D) - J(N) \quad (2.8)$$

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:

$$\langle \Delta H \rangle = x_D \Delta_N^D H = \frac{K \cdot \Delta_N^D H}{1 + K} \quad (2.9)$$

$$C_p^{ex} = \frac{\partial \langle \Delta H \rangle}{\partial T} = \frac{(\Delta_N^D H)^2}{RT^2} \cdot \frac{K}{(1 + K)^2} + x_D \cdot \Delta_N^D C_p \quad (2.10)$$

donde $x_D = \frac{K}{1 + K}$ es la fracción de proteína que se encuentra en el estado desnaturalizado 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:

$$\frac{\partial \ln K}{\partial T} = \frac{\Delta_N^D H}{RT^2} \quad (2.11)$$

$$\frac{\partial \Delta_N^D H}{\partial T} = \Delta_N^D C_P \quad (2.12)$$

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 2.10 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 sigmoideal). 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*, $\Delta_N^D H$. Ésta depende de la temperatura según la ecuación de Kirchoff. 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 esta entalpía a la T_m , y se denomina ΔH_m . Como a esta temperatura el cambio de energía de Gibbs es cero el cambio de entropía a la T_m se puede calcular con:

$$\Delta S_m = \frac{\Delta H_m}{T_m} \quad (2.13)$$

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 estándar:

$$\Delta_N^D H(T) = \Delta_N^D H(T_m) + \int_{T_m}^T \Delta_N^D C_P \cdot dT \quad (2.14)$$

$$\Delta_N^D S(T) = \frac{\Delta_N^D H(T_m)}{T_m} + \int_{T_m}^T \frac{\Delta_N^D C_P}{T} \cdot dT \quad (2.15)$$

$$\Delta_N^D G(T) = \Delta_N^D H(T) - T \cdot \Delta_N^D S(T) \quad (2.16)$$

Si se considera el cambio de capacidad calorífica constante con la temperatura las ecuaciones (2.14), (2.15) y (2.16) pasarían a ser:

$$\Delta_N^D H(T) = \Delta H_m + \Delta_N^D C_P \cdot (T - T_m) \quad (2.17)$$

$$\Delta_N^D S(T) = \frac{\Delta H_m}{T_m} + \Delta_N^D C_P \cdot \ln\left(\frac{T}{T_m}\right) \quad (2.18)$$

$$\Delta_N^D G(T) = \Delta H_m \cdot \left[1 - \frac{T}{T_m}\right] + \Delta_N^D C_P \cdot \left[T - T_m - T \cdot \ln\left(\frac{T}{T_m}\right)\right] \quad (2.19)$$

La dependencia de ΔG con la temperatura se conoce como curva de estabilidad y un ejemplo está representado en la Figura 2.8 A (Sanchez-Ruiz, 1995). Las propiedades fundamentales de la curva de estabilidad se encuentran recogidas en Becktel y Shellman, 1987; Schellman, 1987. A continuación comentamos las más significativas:

- La curvatura viene dada por $\frac{\partial^2 \Delta_N^D G}{\partial T^2} = -\frac{\Delta_N^D C_P}{T}$, que tiene siempre valores negativos porque el cambio de capacidad calorífica de desnaturalización es siempre positivo.

- La pendiente de la curva es $-\frac{\partial \Delta_N^D G}{\partial T} = \Delta_N^D S$ y presenta un único extremo (un máximo) a la temperatura, T_S , a la que $\Delta_N^D S = 0$ (Figura 2.8-B, tomada de Sanchez-Ruiz, 1995). La temperatura T_H , a la que la entalpía de desnaturalización es cero, es ligeramente inferior a T_S . La proteína nativa es estable en el rango de temperatura en que $\Delta_N^D G > 0$. Según se puede ver en la Figura 2.8-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 hace unos 30 años (Brandts, 1964) y demostrada experimentalmente con posterioridad en estudios (véase por ejemplo Privalov, 1986, Ibarra-Molero *et al.*, 1999) que sugieren que la desnaturalización fría es una propiedad común a las proteínas globulares, confirmando así las características ya mencionadas de la curva de estabilidad de proteínas.

Así pues, si tenemos los valores de T_m , ΔH_m y ΔC_p determinados a partir del termograma de DSC, entonces 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 2.9, veremos que ésta 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 ésta a partir del término debido al desplazamiento del equilibrio de desnaturalización por efecto de la temperatura en la ecuación 2.10 y considerando $T=T_m$ (y por tanto $K=1$):

$$\Delta H^{vH} = 4RT^2 \cdot \left[\frac{\Delta C_{p,m}}{\Delta_N^D H_m} \right] \quad (2.20)$$

donde $\Delta C_{p,m}$ es la capacidad calorífica de exceso a la temperatura T_m medida sobre la línea base química. El cociente $\Delta C_{p,m} / \Delta_N^D H_m$ no depende de la cantidad de proteína presente en la célula del calorímetro, pero sí es sensible a la forma (anchura) de la transición. Obviamente, no puede haber dos cambios de entalpía para el mismo proceso. A la ecuación (2.20) se llega 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

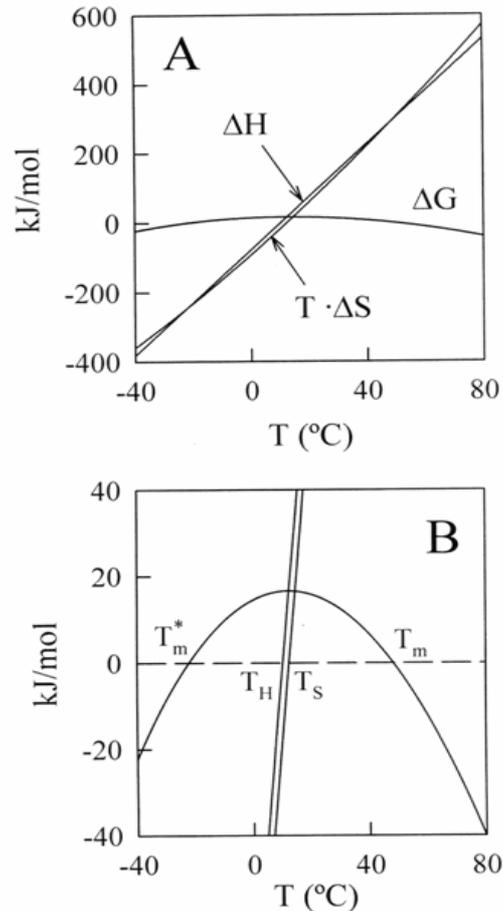


Figura 2.8: (A) Efecto de la temperatura sobre los cambios de entalpía, entropía y energía libre de Gibbs para la desnaturalización de una proteína hipotética. (B) Ampliación que muestra las características principales de la curva de estabilidad de proteínas (la representación de $\Delta_N^D G$ frente a T); se han indicado las temperaturas de desnaturalización por calor (T_m) y por frío (T_m^*), así como las temperaturas en las que los cambios de entalpía y de entropía de desnaturalización son cero (T_H y T_S). Los valores de $\Delta_N^D H$, $\Delta_N^D S$ y $\Delta_N^D G$ se calcularon usando las ecuaciones (2.17) a (2.19) con los siguientes parámetros: $T_m = 48.3$ °C, $\Delta_N^D H(T_m) = 290$ kJ·mol⁻¹ y $\Delta_N^D C_p = 7.6$ kJ·K⁻¹·mol⁻¹. Figura tomada de Sanchez-Ruiz, 1995.

Hoff y la calorimétrica son iguales dentro del error experimental. Si por el contrario existen diferencias significativas entre los valores se concluye que el proceso de desnaturalización no ocurre siguiendo un modelo de dos estados. Los resultados del test del modelo de dos estados se suelen dar como la relación entre las entalpías calorimétrica y de van't Hoff: $r = \Delta_N^D H_m / \Delta H^{vH}$. 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, ya que en este caso la entalpía de van't Hoff y la calorimétrica podrían no estar bien 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.

B) Modelo de equilibrio multiestado

En ocasiones existen transiciones calorimétricas reversibles que no pueden ser explicadas con un modelo tan sencillo como el de dos estados. Son los casos en los que aparecen varias transiciones separadas, o bien, aquellos en los que aún existiendo un sólo pico, las entalpías calorimétrica y de van't Hoff no coinciden. Es aquí donde el parámetro r resulta de gran utilidad. Si $r < 1$ indica que existe cooperatividad intermolecular, es decir, la unidad cooperativa no es un monómero sino un dímero u otro estado de agregación superior al que se suponía inicialmente. Si $r > 1$ debemos considerar la presencia de estados intermedios en la desnaturalización térmica de la proteína en estudio. En este apartado nos centraremos en este último caso.

Un mecanismo general para la desnaturalización térmica de proteínas en equilibrio fue propuesto por Freire y Biltonen, 1978. Estos autores sugieren que el estado nativo (I_0) pasa al estado desnaturalizado (I_n) a través de $n-1$ estados intermedios:

$$I_0 \Leftrightarrow I_1 \Leftrightarrow I_2 \Leftrightarrow \dots \Leftrightarrow I_{n-1} \Leftrightarrow I_n \quad (2.21)$$

El análisis se simplifica si usamos el formalismo de la función de partición. Tomamos como referencia el estado nativo y definimos los pesos estadísticos (ω_i), para cada uno de los estados, como el cociente entre su concentración y la del estado de referencia:

$$\omega_i = \frac{[I_i]}{[I_0]} = K_i = \exp(-\Delta_N^i \Delta G / RT) \quad (2.22)$$

Toda la información relevante sobre el sistema está contenida en la función de partición (Z) que se define como la suma de los pesos estadísticos:

$$Z = \sum_{i=0}^n \omega_i = \sum_{i=0}^n \exp(-\Delta_N^i G / RT) = 1 + \sum_{i=1}^n \exp(-\Delta_N^i G / RT) \quad (2.23)$$

La fracción de proteína (x_i) presente en un determinado estado I_i , viene dada por el cociente entre el peso estadístico de ese estado y la función de partición:

$$x_i = \omega_i / Z = (1/Z) \cdot \exp(-\Delta_N^i G / RT) \quad (2.24)$$

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

$$\langle \Delta H \rangle = \sum_{i=0}^n x_i \cdot \Delta_N^i H = (1/Z) \cdot \sum_{i=0}^n \Delta_N^i H \cdot \exp(-\Delta_N^i G / RT) \quad (2.25)$$

y la capacidad calorífica de exceso:

$$C_p^{ex} = \delta \langle \Delta H \rangle / \delta T = (\langle \Delta H^2 \rangle - \langle \Delta H \rangle^2) / (RT^2) + \langle \Delta C_p \rangle \quad (2.26)$$

donde $\langle \Delta H^2 \rangle$ es el valor medio de $\Delta H^2 \left[= \sum x_i \cdot (\Delta_N^i H)^2 \right]$. El primer término de la ecuación (2.26) se debe al desplazamiento del equilibrio de desnaturalización

inducido por la temperatura, mientras que el segundo término es la capacidad calorífica media de la proteína $\left[\langle \Delta C_p \rangle = \sum \Delta_N^i C_p \cdot x_i \right]$ y corresponde a la línea base química.

A partir de la ecuación (2.25) y mediante una doble integración es posible obtener la función de partición como función de la temperatura:

$$\ln[Z(T)] = \int_{T_0}^T (1/RT^2) \cdot \langle \Delta H \rangle \cdot dT = \int_{T_0}^T (1/RT^2) \cdot \left(\int_{T_0}^T C_p^{ex} \cdot dT \right) \cdot dT \quad (2.27)$$

donde T_0 es una temperatura en la que toda la proteína está presente en estado nativo.

Vemos cómo un perfil de DSC contiene toda la información requerida para caracterizar completamente el proceso de desnaturalización en equilibrio, como ya señalaron Freire y Biltonen. Estos autores desarrollaron un procedimiento recursivo por pasos para determinar el número de estados intermedios significativamente poblados y caracterizados termodinámicamente. También es posible encontrar en la bibliografía modos alternativos de deconvolucionar un perfil de DSC (Gill *et al.*, 1985; Filimonov y Rogov, 1996).

Es posible generalizar el formalismo de la función de partición descrito para incluir efectos de unión de ligando y oligomerización (Thorolfsson *et al.*, 2002).

C) Determinación de las capacidades caloríficas absolutas de los estados nativo y desnaturalizado a partir de los termogramas experimentales

Siguiendo el proceso descrito por Kholodenko y Freire, 1999, es posible determinar valores de capacidad calorífica absoluta a partir de la dependencia con la concentración de proteína de la función de capacidad calorífica aparente ΔC_{app} , que es la capacidad calorífica de la proteína tomando como referencia la línea base tampón-tampón. ΔC_{app} es un número negativo ya que en la célula de muestra la cantidad de agua es menor (y ésta tiene una capacidad calorífica mayor que la proteína). ΔC_{app} de una proteína puede expresarse como:

$$\Delta C_{app} = C \cdot V_0 \cdot 10^{-6} \cdot \left(C_{p,P} - \frac{V_P}{V_W} C_{p,W} \right) \quad (2.28)$$

dónde C es la concentración de proteína expresada en mM, V_0 es el volumen de la célula del calorímetro en mL, $C_{p,P}$ y $C_{p,W}$ son las capacidades caloríficas de la proteína y el agua (tampón), respectivamente y, V_P y V_W son los volúmenes molares de la proteína y el agua (Georgescu *et al.*, 2001). El volumen específico de la proteína se puede calcular mediante procedimientos descritos (Makhatadze y Privalov, 1990). A partir de la ecuación anterior se puede calcular la capacidad calorífica absoluta mediante la representación de ΔC_{app} frente a la concentración de proteína. Para una temperatura dada, dicha representación debe ajustarse a una línea recta y pasar por el origen; de su pendiente, $V_0 \cdot 10^{-6} \cdot C_{p,P}$, podremos calcular la capacidad calorífica absoluta. Este procedimiento requiere varios termogramas correspondientes a varias concentraciones de proteína; sin embargo, muestra muchas ventajas en comparación a los métodos tradicionales:

- a) Cada valor de capacidad calorífica absoluta (a una temperatura dada) se calcula de varios experimentos correspondientes a distintas concentraciones de proteína. Ésto decrece el error asociado al valor de capacidad calorífica y nos permite una estimación de dicho error. Además, como la representación gráfica de ΔC_{app} frente a C se hace con distintos experimentos (cada valor de ΔC_{app} se ha calculado con una línea base tampón-tampón diferente), el error asociado a la línea base instrumental se aleatoriza y se tiene en cuenta en el análisis.
- b) Los datos de ΔC_{app} derivados de experimentos llevados a cabo a diferentes valores de pH (o distintos valores de velocidad de barrido), se pueden incluir en la misma representación gráfica de ΔC_{app} frente a C . Ésto permite determinar el efecto del pH sobre ΔC_{app} y en caso de que no exista ningún efecto, permite usar los datos a diferentes valores de pH para calcular valores de capacidad calorífica absoluta, decreciendo de este modo el error asociado.
- c)

D) Estimación teórica de las capacidades caloríficas de los estados nativo y desnaturalizado

A partir de la composición de aminoácidos es posible determinar la capacidad calorífica teórica de los estados nativo y desnaturalizado. Para el caso

del estado desnaturalizado podemos emplear el enfoque de Privalov y Makhatadze, 1990, el cual supone un estado desnaturalizado completamente solvatado:

$$C_p^D = \sum_{i=1}^{20} n_i \cdot C_{p,i} + (N-1) \cdot C_{p,NH-CO} + C_{p,NH_2} + C_{p,COOH} \quad (2.29)$$

donde N es el número de aminoácidos de la proteína, el subíndice i se refiere a los 20 aminoácidos naturales, n_i es el número de aminoácidos del tipo i , y $C_{p,NH-CO}$, $C_{p,COOH}$ y C_{p,NH_2} son las capacidades caloríficas de la unidad peptídica y los grupos carboxilo y amino terminal para los que se emplean los valores de representaciones polinómicas de los datos de Privalov y Makhatadze, 1990.

Los valores teóricos de capacidad calorífica de proteínas en estado nativo se calculan como una función de la temperatura a partir de la relación empírica dada por Freire, 1995:

$$C_p^N = (1.323 + 0.0067 \cdot T) \cdot M_r \quad (2.30)$$

donde M_r es la masa molecular de la proteína y el resultado se obtiene en $J \cdot K^{-1} \cdot mol^{-1}$.

Diferencias entre los valores de capacidad calorífica absoluta determinados experimentalmente (representación gráfica de ΔC_{app} frente a C) y los valores teóricos calculados según la ecuación 2.30 pueden explicarse por cambios en el área superficial accesible al disolvente (ASA) (Gomez *et al.*, 1995), según:

$$\Delta C_p = 1.88 \cdot \Delta ASA_{ap} - 1.09 \cdot \Delta ASA_{pol} \quad (2.31)$$

donde los cambios en el área superficial son en Å^2 y el cambio en capacidad calorífica es en $J \cdot K^{-1} \cdot mol^{-1}$. (Los valores de los coeficientes ΔASA_{ap} y ΔASA_{pol} corresponden a los calculados a 25°C; dichos coeficientes cambian algo con la temperatura, pero en un análisis cualitativo como el que se hace en el apartado C, no se tiene en cuenta dicha dependencia). De acuerdo a la

ecuación 2.28 un valor de capacidad calorífica (determinado experimentalmente) menor que el predicho para una proteína completamente solvatada (ecuación 2.29), se puede explicar en términos de enterramiento de superficie apolar, ya que un enterramiento de superficie polar daría lugar a un aumento de la capacidad calorífica (obsérvense los coeficientes de ΔASA_{ap} y ΔASA_{pol} en ecuación 2.31). Si tenemos en cuenta la relación existente en proteínas para ambos tipos de áreas ($\Delta ASA_{ap} / \Delta ASA_{pol} = 0.59$, Spolar y Record, 1994), valores de capacidades caloríficas determinadas experimentalmente menores que los predichos se suelen explicar en términos de enterramiento de superficie apolar junto con el porcentaje correspondiente de superficie polar. Esta conclusión parece ser consistente con la existencia de estudios experimentales (a menudo basados en RMN) que sugieren la existencia de interacciones entre residuos apolares en proteínas desplegadas o fragmentos de proteínas desplegadas, más concretamente estudios experimentales sobre la formación de agrupaciones hidrofóbicas en el estado desplegado (Garvey *et al.*, 1989, Evans *et al.*, 1991, Neri *et al.*, 1992, Lumb y Kim, 1994, Demarest *et al.*, 1998, Robic *et al.*, 2003 y Guzman-Casado *et al.*, 2003).

2.2.5 DICROÍSMO CIRCULAR

2.2.5.1 Bases teóricas

El dicroísmo circular (DC) se basa en que las moléculas quirales, entre las que están las proteínas, interaccionan de distinta manera con la luz polarizada circularmente hacia la derecha (*lpcd*) y la izquierda (*lpci*) en ausencia de campos magnéticos (Woody, 1996). Concretamente, el DC se define como la diferencia entre la absorción de la luz polarizada circularmente hacia la derecha y hacia la izquierda.

La luz polarizada circularmente se consigue con dos haces de luz ortogonales que están polarizados en el plano y que se encuentran fuera de fase en 90°. El vector eléctrico de la luz polarizada circularmente se puede expresar matemáticamente (Michel y Thulstrup, 1986; Kliger *et al.*, 1990) como:

$$\mathbf{E}_{\pm} = E_0(\mathbf{i} \pm i\mathbf{j}) \exp[2\pi i(vt - z/\lambda)] \quad (2.32)$$

donde + indica luz circular polarizada a la derecha y – indica luz circular polarizada a la izquierda, E_0 es la amplitud de la onda, \mathbf{i} y \mathbf{j} son los vectores unidad en las coordenadas x e y respectivamente en un sistema de coordenadas cartesiano hacia la derecha en el que + z es la dirección de propagación, i es $(-1)^{1/2}$ y v y λ son la frecuencia y la longitud de onda de la luz.

La absorbancia de luz polarizada hacia la izquierda se puede expresar como:

$$A_i = \log_{10} \left(\frac{I_i^0}{I_i} \right) = \varepsilon_i cl \quad (2.33)$$

Y hacia la derecha:

$$A_d = \log_{10} \left(\frac{I_d^0}{I_d} \right) = \varepsilon_d cl \quad (2.34)$$

donde I^0 e I son las intensidades de luz polarizada que inciden y que salen de la muestra respectivamente, l es la distancia que recorre la luz a través de la muestra,

C es la concentración de soluto quiral y ε es el coeficiente de extinción molar del soluto para la luz polarizada hacia el lado correspondiente. En consecuencia, según la definición de DC:

$$\Delta A = A_i - A_d = \varepsilon_i \cdot cl - \varepsilon_d \cdot cl = (\varepsilon_i - \varepsilon_d) \cdot cl = \Delta\varepsilon \cdot cl \quad (2.35)$$

donde las unidades de $\Delta\varepsilon$ son decimal molar.

Además de expresar el DC en forma de $\Delta\varepsilon$, en bioquímica también se utiliza el método de medida original propuesto por Lowry (Lowry, 1935). Éste se basa en que cuando una luz polarizada circularmente pasa a través de un medio circularmente dicroico, la absorción diferencial de las dos componentes circulares hace que a la salida de la muestra cada componente siga estando circularmente polarizada, pero los radios de los círculos trazados por los vectores eléctricos son ahora diferentes, y cuando estas dos ondas luminosas opuestas, polarizadas circularmente, se combinan, el resultado sea luz polarizada elípticamente. Cuando los vectores eléctricos de las dos componentes circulares van en el mismo sentido, la suma de sus magnitudes es el eje mayor de la elipse, y cuando van en sentidos opuestos, la resta de sus magnitudes nos da el valor del eje menor de la elipse. Ahora el DC se puede caracterizar por la relación entre los ejes mayor y menor, que es la tangente de un ángulo θ , al que se denomina elipticidad. Como θ es normalmente muy pequeño, puede expresarse como:

$$\theta(rad) = \tan \theta = \frac{|\mathbf{E}_i| - |\mathbf{E}_d|}{|\mathbf{E}_i| + |\mathbf{E}_d|} = \frac{\exp\left(\frac{-A_i}{2}\right) - \exp\left(\frac{-A_d}{2}\right)}{\exp\left(\frac{-A_i}{2}\right) + \exp\left(\frac{-A_d}{2}\right)} \quad (2.36)$$

Expandiendo las exponenciales, teniendo en cuenta el valor despreciable de ΔA frente a la unidad y convirtiendo a grados:

$$\theta(grad) = 180 \cdot \ln 10 \cdot \Delta A / 4\pi = 32.98 \Delta A \quad (2.37)$$

obtenemos que la elipticidad es proporcional al dicroísmo circular.

La elipticidad molar se define como:

$$[\theta] = \frac{\theta}{10 \cdot c \cdot l} \quad (2.38)$$

o bien la elipticidad molar normalizada:

$$[\theta]_{MRW} = \frac{\theta \cdot M_r}{10 \cdot c \cdot l \cdot N_A} \quad (2.39)$$

donde θ se introduce en mrad (datos crudos salida del espectropolarímetro), M_r es la masa molecular y N_A el número de aminoácidos; $[\theta]$ y $[\theta]_{MRW}$ se expresan en $\text{grados} \cdot \text{dmol}^{-1} \cdot \text{cm}^2$. La concentración se introduce en dmol/cm^3 .

2.2.5.2 Espectros de dicroísmo circular de proteínas

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

- Determinar si una proteína se encuentra plegada y en tal caso, podemos caracterizar su estructura secundaria (ultravioleta lejano), estructura terciaria (ultravioleta cercano) y clasificarla dentro de una familia estructural.
- Comparar las estructuras de proteínas obtenidas mediante otras técnicas (rayos X o RMN) o las estructuras de diferentes mutantes de una misma proteína.
- Estudiar la estabilidad conformacional de una proteína (estabilidad térmica, frente a cambios de pH o en presencia de agentes desnaturizantes) y cómo esta estabilidad se altera dependiendo de las condiciones utilizadas.
- Determinar los cambios conformacionales asociados a interacciones entre moléculas asimétricas (interacciones proteína-proteína, proteína-DNA, proteína-ligando y DNA-ligando).

2.2.5.3 Dicroísmo circular y la estructura secundaria de las proteínas

La estructura secundaria de las proteínas se puede determinar mediante DC en el ultravioleta lejano (190-250 nm). Esta región es muy sensible a los cambios estructurales, fundamentalmente a los de estructura secundaria, ya que la contribución de la estructura terciaria a la señal de DC en estas longitudes de onda es despreciable. A estas longitudes de onda el cromóforo es el enlace peptídico. Aunque el enlace peptídico *per se* no es una estructura quiral (con capacidad de absorber de modo distinto luz de diferente polarización), la interacción entre grupos peptídicos adyacentes dispuestos en determinadas orientaciones determina que las proteínas presenten actividad óptica en la región del UV lejano, donde absorbe el enlace peptídico. Cada elemento principal de estructura secundaria de las proteínas (hélices α , láminas β y estructura no regular) presenta un espectro de dicroísmo circular característico (Figuras 2.11 y 2.12).

En concreto las hélices α presentan un espectro de DC muy característico en la región del UV lejano. El espectro de DC de una proteína con estructura de hélice α presenta un doble mínimo a 222 y a 208-210 nm y un máximo más intenso a 191-193 nm (ver Figura 2.9). La intensidad de las tres bandas (dependiente de los ángulos diedros ϕ y Ψ) da una idea de la cantidad de contenido helicoidal que tiene la proteína.

Las proteínas con estructura β suelen presentar un único pico negativo y un único pico positivo cuyas intensidades son menores que las de las hélices. El mínimo se encuentra normalmente entre 210 y 225 nm (ver Figura 2.10) aunque se han descrito algunos ejemplos (Manavalan y Johson, 1983; Wu *et al.*, 1992) donde proteínas que por RMN se ha determinado que son todo β tienen una banda negativa intensa alrededor de 200 nm, como los péptidos con estructura no regular. La explicación que se ha dado a estos casos es que, según los datos de difracción de rayos X, estas proteínas tienen hojas β cortas y bastante irregulares. El pico positivo es de mayor intensidad que el negativo y aparece entre 190 y 200 nm (Figura 2.10).

Los péptidos y proteínas con estructura no regular (los bucles de las conformaciones nativas y, en general, las conformaciones desplegadas) presentan habitualmente una banda negativa fuerte cerca de 200 nm (véase Figura 2.11) y

algunas bandas débiles entre 220 y 230 nm que pueden ser tanto positivas como negativas.

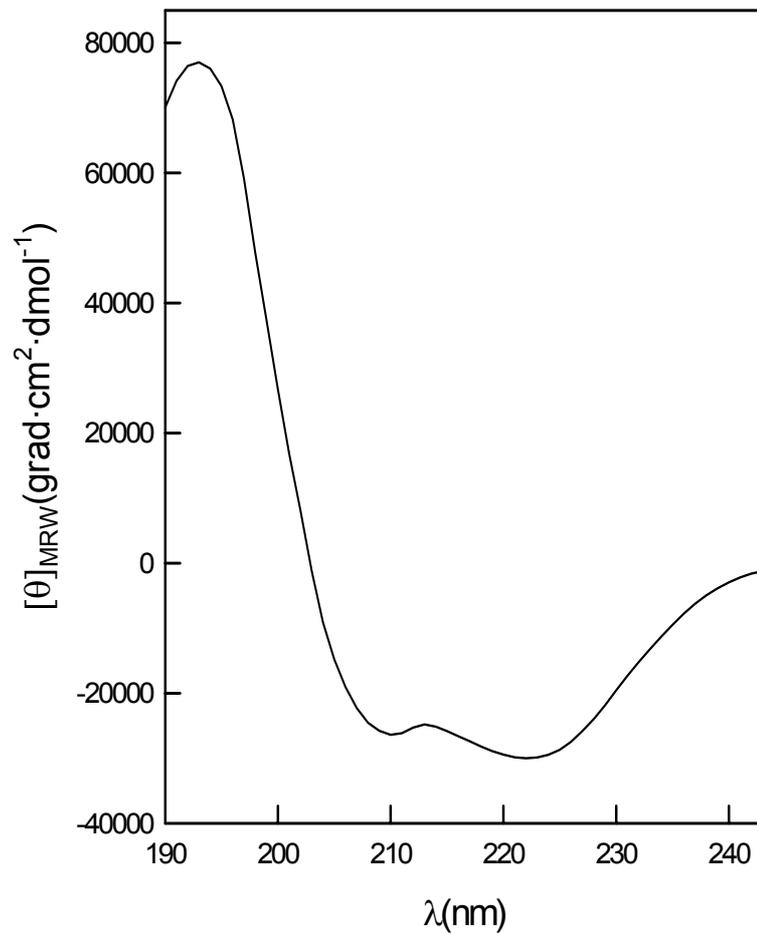


Figura 2.9. Espectro ultravioleta lejano de DC de una proteína con estructura de hélice α (Chen et al., 1972).

Hay que recordar, sin embargo, que en el UV lejano también absorben los aminoácidos aromáticos y que, aunque sus cadenas laterales no son quirales, se

comportan como tales cuando así lo propicia su entorno proteico. Por esta razón el espectro de algunas proteínas aparece dominado por contribuciones peculiares debidas a estos residuos. Con esta salvedad, la actividad óptica de los enlaces peptídicos es de gran utilidad para conocer el contenido de estructura secundaria de las proteínas. Así, cada proteína tiene un espectro de dicroísmo circular en el ultravioleta lejano que es, aproximadamente, una combinación lineal de los espectros característicos de las estructuras secundarias y no periódicas que contiene.

2.2.5.4 Dicroísmo circular y la estructura terciaria de las proteínas

El espectro ultravioleta cercano (250-350 nm) de una proteína es sensible a cambios en la estructura terciaria. A estas longitudes de onda los cromóforos son los aminoácidos aromáticos y los enlaces disulfuro cuando se encuentran inmovilizados en un entorno estructural asimétrico, típicamente el estado nativo de algunas proteínas.

La contribución de la fenilalanina al espectro de CD es pequeña, debido a su bajo coeficiente de extinción; además, es poco sensible a las alteraciones de su ambiente. Presenta una banda de absorción en la región 250-270 nm. La tirosina tiene una banda de absorción más intensa a 276 nm y un hombro a 283 nm (banda que puede desplazarse hacia el rojo debido a la formación de puentes de hidrógeno por el grupo fenólico). El triptófano presenta la banda de absorción más intensa centrada a 282 nm (también puede desplazarse hacia el rojo debido a la formación de un puente de hidrógeno por el grupo indol). A diferencia de lo que ocurría en el lejano, las absorciones de las cadenas laterales de estos residuos aromáticos no se mezclan con las del grupo peptídico (que no absorbe en esta región). Así, el espectro de dicroísmo circular de una proteína en el ultravioleta cercano es una pequeña huella dactilar de la conformación nativa cuya pérdida nos indicará que han ocurrido cambios conformacionales apreciables.

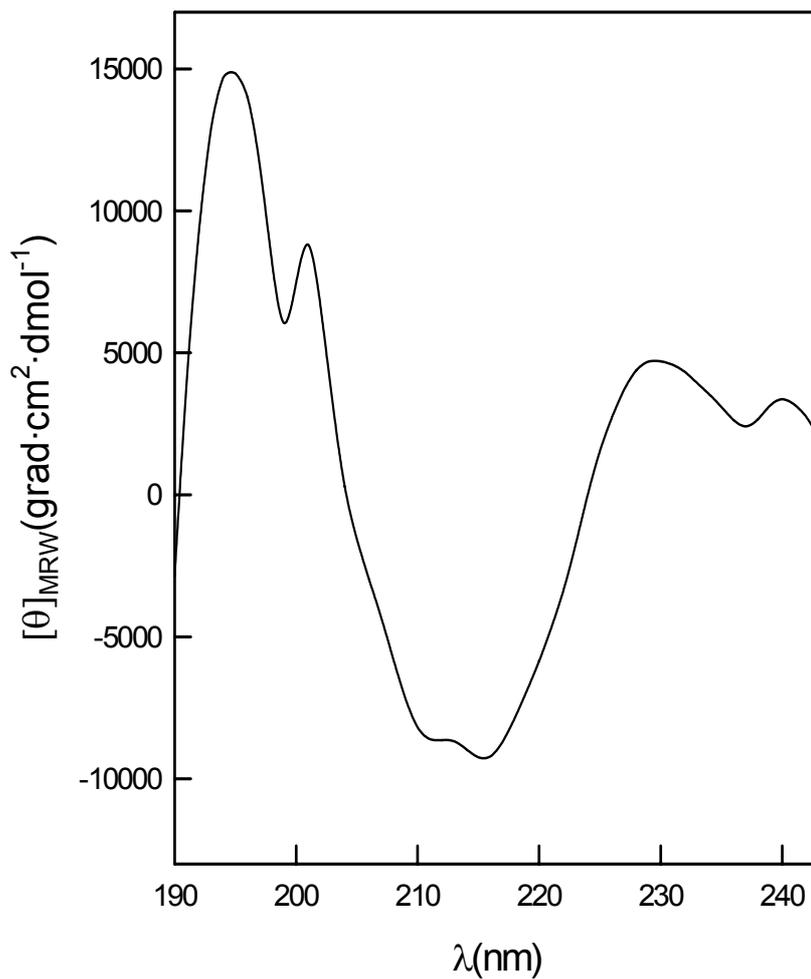


Figura 2.10. Espectro ultravioleta lejano de DC de una proteína con estructura de hoja β (Chen et al., 1972).

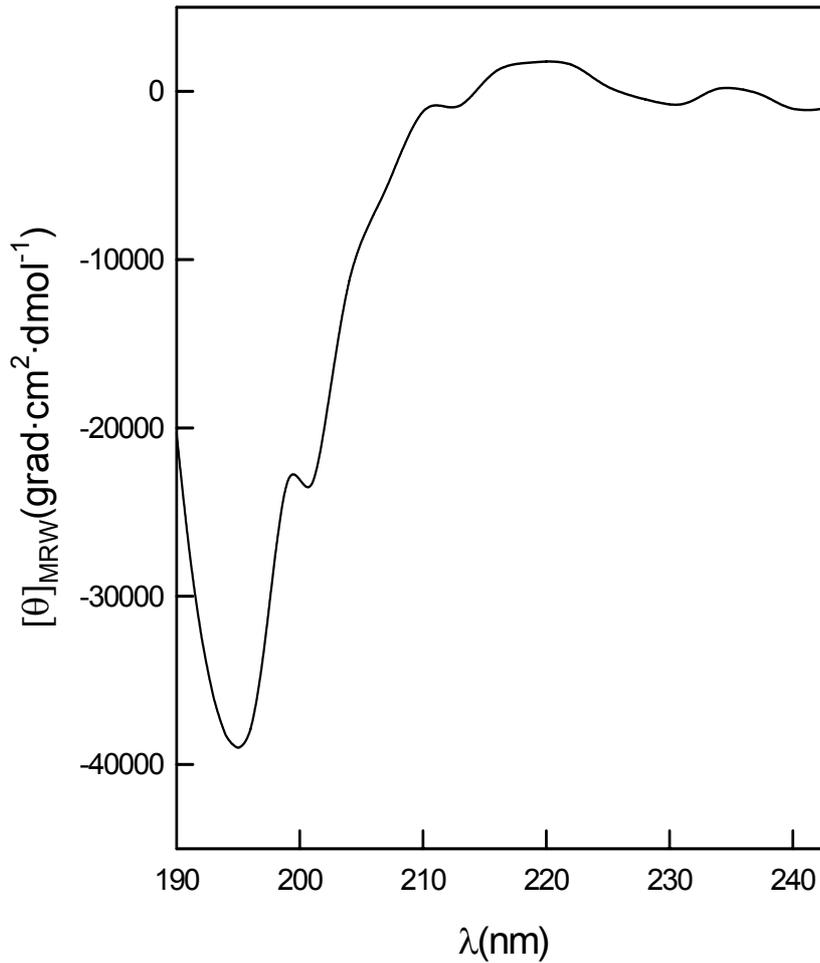


Figura 2.11. Espectro de ultravioleta lejano de DC de una proteína con estructura no regular (Chen et al., 1972).

2.2.5.5 Procedimiento experimental

Los experimentos se han llevado a cabo en un espectropolarímetro JASCO modelo J-715 (sección 2.1.2). La proteína se dializó en tampón HEPES 5 mM, pH 7 preparado como se indica en el apartado 2.2.3.1.

Se han tomado los espectros de DC en el lejano y cercano de tiorredoxina silvestre y formas mutantes a diferentes temperaturas, con objeto de observar la conservación/pérdida de la estructura correspondiente al estado nativo. Los experimentos en el ultravioleta lejano se registraron desde 260 hasta 200 nm; se utilizaron cubetas de 1 mm de paso de luz y concentraciones de muestra de 0.18 mg/mL. Para el ultravioleta cercano (325-260 nm), el coeficiente de extinción de las bandas de absorción de los grupos aromáticos es menor, por lo que fue posible utilizar cubetas de 1 cm de paso de luz y concentraciones de proteína de 0.3-1mg/mL.

Debido a limitaciones técnicas no se pudieron realizar espectros a temperaturas superiores a 90°C. Inicialmente, se debe de registrar un espectro con la cubeta llena del tampón de diálisis. Posteriormente, el blanco se sustraerá al espectro obtenido con la disolución de proteína con el fin de eliminar las diferentes contribuciones instrumentales y del disolvente al espectro de DC. Antes de registrar cada espectro, se deja la muestra equilibrar a la temperatura deseada al menos 1 min.

El programa de adquisición del instrumento permite promediar N espectros de DC consecutivos de forma que el nivel de ruido del espectro resultante es $N^{1/2}$ veces menor que el de cada espectro individual, ganándose de esta manera sensibilidad. En todos los casos se realizó una acumulación de 8 espectros.

2.2.6 FLUORESCENCIA

2.2.6.1 Generalidades

La fluorescencia intrínseca de las proteínas se usa comúnmente para detectar y seguir cambios conformacionales y más concretamente cambios en el ambiente de los fluoróforos. Hemos elegido esta técnica para nuestros estudios de desnaturalización por codisolventes debido a su gran sensibilidad, bajo requerimientos de muestra y fácil disponibilidad.

A continuación detallamos algunos de los aspectos experimentales relativos a la proteína estudiada. Tiorredoxina silvestre (WT) (y todas sus formas mutantes obtenidas) posee dos triptófanos en las posiciones 28 y 31. Los espectros de excitación de soluciones de tiorredoxina nativa y desplegada presentan un máximo centrado a 276 nm y un hombro alrededor de 290 nm, espectro similar a los espectros de absorción del ultravioleta cercano. Los espectros de emisión de fluorescencia de tiorredoxina, nativa y desplegada, se han registrado siguiendo la excitación a 276 nm. Éstos presentan un máximo muy ancho centrado a 344 nm (ver Figura 2.12) característico de los triptófanos; sin embargo, la intensidad del perfil de emisión de fluorescencia en presencia de altas concentraciones de urea es mucho mayor, desplazándose el máximo de emisión de 344 a 351 nm. Estos cambios reflejan en gran medida la separación estérica del triptófano 28 del puente disulfuro existente entre las cisteínas 32 y 35, razón por la cual disminuye el efecto de *quenching* que existe entre ambos y por lo tanto, aumenta la intensidad de fluorescencia (Reutimann *et al.*, 1981 y Kelley y Stellwagen, 1984). Este análisis nos ha permitido observar que la máxima diferencia entre el espectro de emisión de tiorredoxina nativa y desplegada ocurre a 350 nm, por lo que la hemos fijado como la longitud de onda a la que registramos nuestras medidas de emisión de fluorescencia.

A continuación, hay que establecer el rango de concentración de proteína en disolución en el que la señal de fluorescencia cambia linealmente. Este aspecto es muy importante a la hora de poner a punto cualquier método fluorimétrico ya que interesa obtener la máxima relación señal/ruido, en un rango lineal para evitar efectos no deseados como el de filtro interno. Para ello hemos medido la intensidad de fluorescencia de disoluciones de proteína de concentración comprendida en el rango de 0.01 a 0.06 mg/mL encontrando una dependencia

lineal como en este rango de concentraciones. Esto se ha comprobado tanto para la proteína nativa como desplegada en alta concentración de urea.

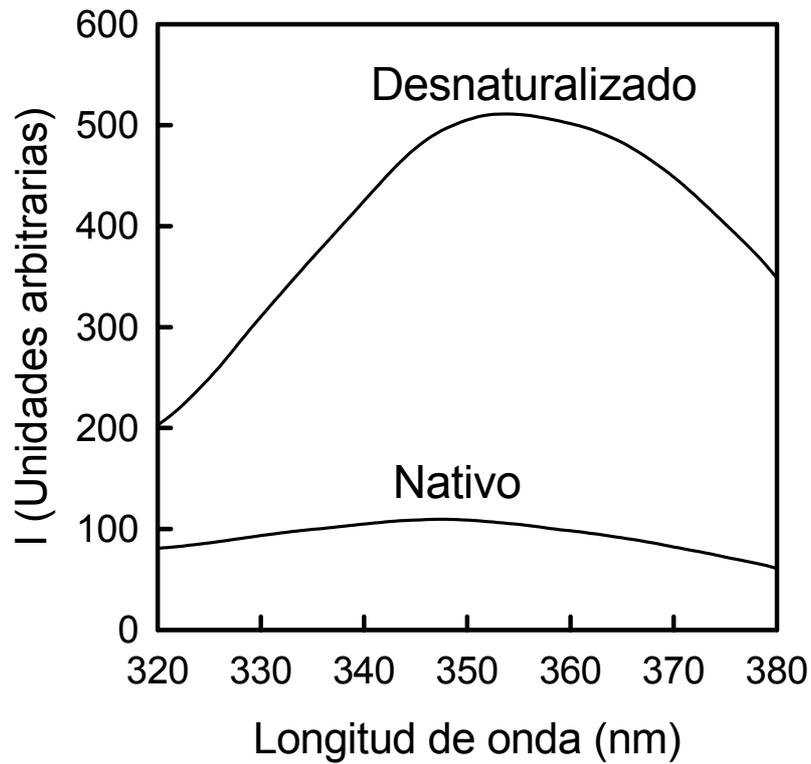


Figura 2.12: Espectros de emisión de fluorescencia a 25°C de tiorredoxina en estado nativo y desnaturalizado. La proteína nativa está disuelta en tampón HEPES 5 mM, pH7 mientras que la desplegada se encuentra en urea 8 M, HEPES 5 mM pH 7. La concentración de proteína es 0.05 mg/mL en ambos casos. La velocidad de barrido seleccionada es de 500 nm/min y las rendijas de excitación y emisión son de 10 nm, respectivamente. Se ha utilizado una cubeta de cuarzo de 1 cm de paso de luz. La longitud de onda de excitación seleccionada ha sido 276 nm; se puede observar como la máxima diferencia entre los dos espectros correspondientes a la proteína nativa y desplegada se produce a 350 nm.

Por último, se deben fijar los valores óptimos de anchura de las rendijas de excitación y emisión. Si la calidad de la señal es suficiente se suele usar una rendija de excitación estrecha para así reducir la intensidad de la radiación que llega a la muestra y evitar la potencial descomposición fotoquímica de la misma. La apertura de la rendija de emisión dependerá de la señal que se consiga. Debemos jugar con diferentes valores hasta conseguir una buena relación señal/ruido. Para todos los experimentos de tiorredoxina y cualquiera de sus formas mutantes se han utilizado rendijas de emisión y excitación de 10 nm, respectivamente. Además, parte de los experimentos se han realizado en un PerkinElmer LS45 que posee rendijas de emisión y excitación de 10 nm, respectivamente, sin posibilidad de variarlas.

2.2.6.2 Análisis de las cinéticas de replegamiento-desplegamiento inducido por urea

En este trabajo de investigación se han caracterizado las cinéticas en la dirección del replegamiento-desplegamiento inducido por urea, de tiorredoxina (*WT* y todas sus formas mutantes) a una temperatura de 25°C, excitando a 276 nm y siguiendo la emisión de fluorescencia a 350 nm.

La proteína se ha dializado en tampón HEPES 5 mM, pH 7, preparado según se ha indicado en el apartado 2.2.3.1. La disolución concentrada de urea se ha preparado siguiendo el procedimiento descrito en el apartado 2.2.3.2. Las condiciones experimentales relativas al fluorímetro son las detalladas en 2.2.6.1.

A) Cinéticas en la dirección del desplegamiento:

Se parte de una disolución de tiorredoxina nativa (*WT* o cualquier variante mutante de tiorredoxina) de concentración en torno a 10 mg/mL en tampón acuoso a 25°C y transferimos una alícuota directamente a la cubeta del fluorímetro en donde tenemos una disolución de concentración conocida de urea, termostaticada; se diluye el volumen necesario de disolución concentrada de proteína de manera que la concentración final sea de 0.04 mg/mL. Una vez añadida la proteína a la cubeta, con la cánula de la pipeta automática se homogeniza bien la disolución y rápidamente hay que comenzar a medir el aumento de intensidad de fluorescencia en función del tiempo (manteniendo la agitación en el interior de la cubeta, siempre que sea posible en el fluorímetro utilizado). Para evitar posibles efectos de fotodescomposición de la muestra en el

caso de cinéticas lentas, los fluorímetros utilizados constan de una lámpara pulsante, de manera que la muestra sólo está siendo irradiada intermitentemente. Una vez finalizada la cinética, medimos el índice de refracción de la disolución para conocer exactamente la concentración de urea, ya que se ha producido una dilución a causa de la adición de la disolución de proteína (ecuación 2.5). De este modo, se han estudiado las cinéticas de desplegamiento en presencia de concentraciones de desnaturalizante desde la $C_{1/2}$ hasta 9 M aproximadamente (véase Figura 2.14-B). Este límite superior viene dado por la limitación instrumental que supone no poder caracterizar cinéticas cuyos tiempos de relajación sean comparables al tiempo de respuesta del fluorímetro y al tiempo muerto debido a la mezcla manual de la muestra que se realiza.

Las cinéticas en la dirección del desplegamiento observadas fueron cinéticas simples que se ajustaron bien a perfiles monoexponenciales. Para cinéticas de primer orden, la dependencia de la intensidad de fluorescencia con el tiempo viene dada por la ecuación de velocidad siguiente:

$$I = I_{\infty} - \Delta I \cdot e^{(-t/\tau)} \quad (2.40)$$

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

$$\tau = \frac{1}{k} \quad (2.41)$$

B) Cinéticas en la dirección del replegamiento.

El procedimiento experimental es muy similar al descrito para las cinéticas de desplegamiento. En este caso, se parte de una disolución de proteína que se ha incubado en urea 9 M, tampón HEPES 5 mM, pH7 (concentración a la cual la proteína se encuentra totalmente desplegada) el tiempo necesario para que alcance el equilibrio (3 horas aproximadamente). Las cinéticas de replegamiento son seguidas mediante el descenso de la intensidad de fluorescencia con el tiempo,

que ocurre al transferir una alícuota de proteína en urea 9 M, tampón HEPES 5 mM, pH 7, a la cubeta donde será diluida con tampón HEPES 5 mM, pH 7. De este modo, se pueden llegar a obtener disoluciones de proteína en diferente concentración de urea, que comprenden desde la $C_{1/2}$ hasta alrededor de 4 M (véase Figura 2.14-A). Este límite inferior de concentración de desnaturizante se ha impuesto porque en el estado desnaturizado, en ausencia de restricciones, los enlaces X-Pro presentan las dos conformaciones, *trans* y *cis*. Ésto implica la existencia de procesos paralelos de plegamiento: uno rápido para aquellas moléculas desplegadas que presentan para cada prolina el isómero “correcto” (el presente en la conformación nativa), y otros más lentos que incluyen una o varias fases adicionales debidas a la lenta isomerización de una o más prolinas (Wilson *et al.*, 1986; Georgescu *et al.*, 1998). Para saber la concentración exacta de desnaturizante a la que se hizo el experimento medimos el índice de refracción de la disolución que contiene la cubeta al finalizar éste (ecuación 2.5).

Las cinéticas en la dirección de replegamiento observadas fueron cinéticas simples que se ajustaron bien a perfiles monoexponenciales. En este caso, la dependencia de la intensidad de fluorescencia con el tiempo viene dada por la ecuación de velocidad siguiente:

$$I = I_{\infty} + \Delta I \cdot e^{(-t/\tau)} \quad (2.42)$$

donde I_{∞} , al igual que en la ecuación 2.40 descrita para los procesos de desplegamiento, es la intensidad de fluorescencia a tiempo infinito, ΔI la amplitud de la exponencial y τ el tiempo de relajación.

Los perfiles de intensidad de fluorescencia versus tiempo obtenidos para los procesos de desplegamiento-replegamiento a distintas concentraciones de urea y 25°C se han analizado según las ecuaciones 2.40 y 2.42, respectivamente. Ambos casos corresponden a cinéticas de primer orden. Los ajustes se han realizado con el programa comercial MatLab (Programas 2 y 3 de Apéndice I); en todos los casos los ajustes resultantes han sido excelentes como se muestra en la Figura 2.14-B. La desviación estándar típicamente observada ha sido el 0.03% del total de la amplitud de las cinéticas exponenciales.

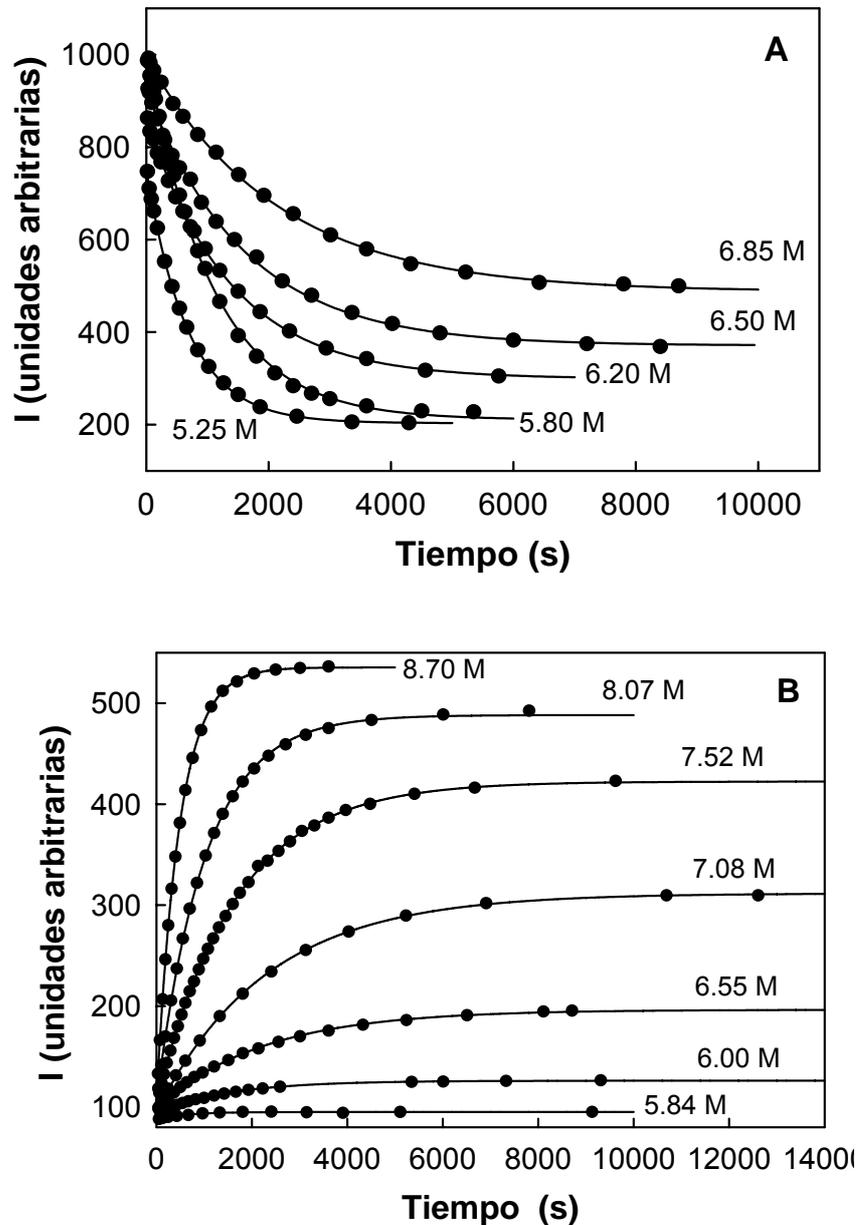


Figura 2.13: (A) Perfiles cinéticos en la dirección del replegamiento de tiorredoxina *WT*, registrados a 25°C después de transferir una alícuota de tiorredoxina desplegada en urea 9 M (tampón HEPES 5 mM, pH 7), hasta disoluciones de concentración de urea igual a: 5.25 M, 5.80 M, 6.20 M, 6.50 M y 6.85 M. (B) Perfiles cinéticos en la dirección del desplegamiento de tiorredoxina *WT*, registrados a 25°C después de transferir una alícuota de proteína nativa (en tampón HEPES 5 mM, pH 7) hasta disoluciones de concentración de urea igual a: 5.84 M, 6.00 M, 6.55 M, 7.08 M, 7.52 M, 8.07 M y 8.70 M. Las líneas continuas representan el ajuste a los datos experimentales según la

C) Análisis según el modelo cinético de dos estados

El análisis de los datos cinéticos lo hemos llevado a cabo aplicando el modelo cinético de dos estados (Ibarra-Molero y Sanchez-Ruiz, 1996, 1997). Según el modelo cinético de dos estados tenemos que:



$$K = \frac{[D]}{[N]} = \frac{k_U}{k_F} \quad (2.44)$$

donde k_F y k_D son las constantes de velocidad correspondientes a los procesos de plegamiento y desplegamiento, respectivamente. La constante de velocidad aparente, k , viene dada por la suma de éstas de manera que:

$$\ln k = \ln(k_F + k_U) \quad (2.45)$$

Las ecuaciones 2.46 y 2.47 relacionan las constantes k_F y k_D con la concentración de urea y pueden deducirse fácilmente de la teoría del estado de transición (Chen *et al.*, 1992):

$$\ln k_F = \ln k_{1/2} + \frac{m_{U-\ddagger}}{RT} \cdot (C - C_{1/2}) \quad (2.46)$$

$$\ln k_D = \ln k_{1/2} + \frac{m_{N-\ddagger}}{RT} \cdot (C - C_{1/2}) \quad (2.47)$$

donde $m_{D-\ddagger}$ y $m_{N-\ddagger}$ describen el efecto de concentración de desnaturalizante sobre las energías libres de Gibbs de activación para el plegamiento y desplegamiento, respectivamente y se expresan:

$$m_{D-\ddagger} = -\left(\frac{\delta\Delta G_{U-\ddagger}}{\delta C}\right) \quad (2.48)$$

$$m_{N-\ddagger} = -\left(\frac{\delta\Delta G_{N-\ddagger}}{\delta C}\right) \quad (2.49)$$

donde se ha supuesto una dependencia lineal del cambio en las energías de Gibbs de activación con la concentración de desnaturalizante; $k_{1/2}$ es el valor de k_F y k_U cuando $C=C_{1/2}$; $C_{1/2}$ es el valor de concentración de desnaturalizante en el que la mitad de las moléculas de proteína están en su estado nativo y la otra mitad en el estado desnaturalizado; $m_{U-\ddagger}$ y $m_{N-\ddagger}$ están relacionados con el cambio de área superficial accesible expuesta al disolvente durante el plegamiento y desplegamiento, respectivamente, es decir, con el tamaño de la proteína. Más concretamente, $m_{U-\ddagger}$ se relaciona con el cambio de área expuesta al disolvente, producido en el paso desde el estado desplegado al estado de transición y $m_{N-\ddagger}$ se relaciona con el cambio de área expuesta en el proceso de desplegamiento, desde el estado nativo al estado de transición.

La representación gráfica del logaritmo neperiano de las constantes de velocidad aparentes obtenidas del análisis de cada cinética en función de la concentración de desnaturalizante (ecuaciones 2.45 a 2.47), tiene forma de “V” y se denomina *Chevron* (véase Figura 2.14); su mínimo coincide de manera aproximada con $C_{1/2}$, donde $k_U=k_F$. La rama de la derecha (concentraciones de desnaturalizante altas) corresponde al proceso de desplegamiento de la proteína inducido por saltos a concentraciones de urea mayores que $C_{1/2}$ (ecuación 2.47); en esta rama se cumple que $k_U>k_F$ y $m_{U-\ddagger}$ se considera la pendiente de la recta. La rama de la izquierda (concentraciones de desnaturalizante bajas) corresponde al proceso de replegamiento de la proteína inducido por saltos a concentraciones de urea menores que $C_{1/2}$ (ecuación 2.46); en esta rama se cumple que $k_F>k_U$ y $m_{F-\ddagger}$ se considera la pendiente de la recta. En la Figura 2.15 se muestra un esquema de los procesos predominantes en cada una de las regiones que se distinguen en el *Chevron*.

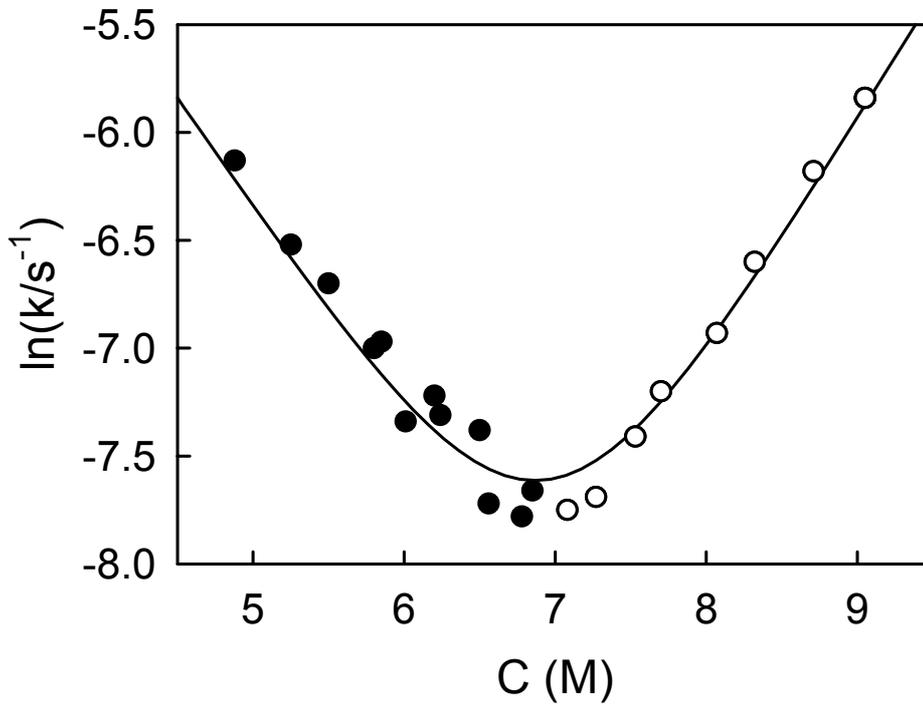


Figura 2.14: Dependencia de las constantes de velocidad de replegamiento (círculos negros)-desplegamiento (círculos blancos) con la concentración de urea (representación denominada Chevron) para tiorredoxina silvestre. Los círculos negros y blancos corresponden a los experimentos cinéticos realizados en las direcciones del replegamiento (Figura 2.14-B) y desplegado (Figura 2.14-A), respectivamente. La línea continua representa el mejor ajuste no lineal por mínimos cuadrados de la dependencia predicha por las ecuaciones 2.45 a 2.47 a los datos experimentales.

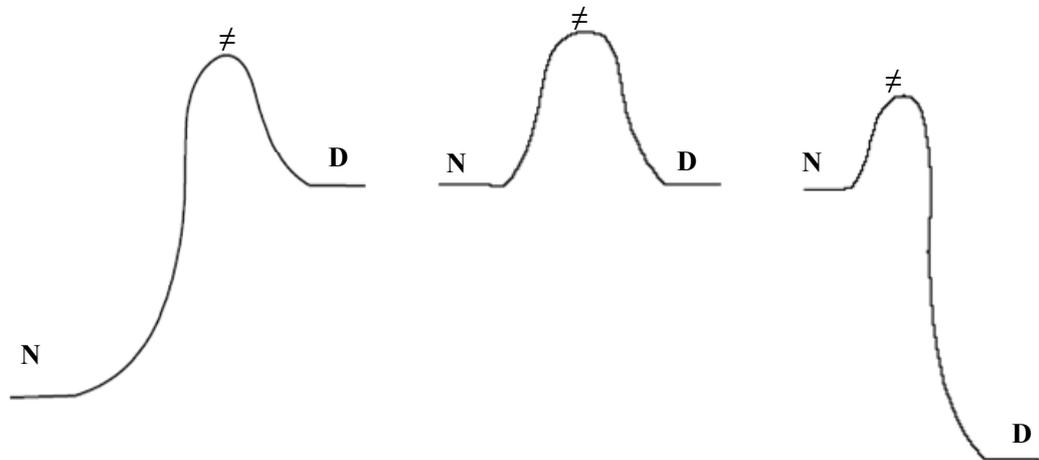


Figura 2.15: Diagramas de energía simplificados relativos a los estados nativo, desplegado y de transición, correspondientes a una proteína hipotética que despliega según el modelo de dos estados. A baja concentración de desnaturante (rama izquierda del *Chevron*) el estado más favorecido energéticamente es el nativo; en estas concreciones la constante de equilibrio, K , es mucho menor a la unidad, $k_F \gg k_U$ y por tanto la constante de velocidad aparente que medimos en los experimentos cinéticos se puede aproximar a k_F . A alta concentración de urea (rama de la derecha del *Chevron*) el estado energéticamente más favorecido es el desplegado; K es mucho mayor a la unidad, $k_U \gg k_F$ y por tanto la constante de velocidad aparente que medimos en los experimentos cinéticos equivale de manera aproximada a k_U .

La dependencia lineal de $\ln k$ con C que se observa en las dos ramas del Chevron implica una dependencia lineal del cambio de energía de Gibbs de activación con C .

El valor de $m_{1/2}$ correspondiente al equilibrio se puede calcular a partir de los valores cinéticos $m_{U-\ddagger}$ y $m_{N-\ddagger}$, dado que:

$$\Delta_N^\ddagger G = \Delta G + \Delta G_U^\ddagger \quad (2.50)$$

resulta:

$$m_{1/2} = m_{N-\ddagger} - m_{U-\ddagger} \quad (2.51)$$

Esta última expresión (ecuación 2.51) constituye un *test de dos estados* en cinética. Este análisis nos será útil en el estudio estructural del estado de transición y del estado desnaturalizado.

El análisis de este tipo de representación ha sido ampliamente descrito en diferentes trabajos (Beasty *et al.*, 1986, Hurle *et al.*, 1987, Matthews y Hurle, 1987, Perry *et al.*, 1989, Chen *et al.*, 1989a y Chen y Schellman, 1989b, Chen *et al.*, 1992, Ibarra-Molero y Sanchez-Ruiz, 1996).

Para las distintas proteínas estudiadas en este trabajo de investigación, los *Chevrons* obtenidos se han interpretado de acuerdo al modelo cinético de dos estados según el cual los procesos de replegamiento y desplegamiento ocurren sin que ningún intermedio cinético llegue a poblarse, al menos significativamente. Ésto es así ya que, en nuestras condiciones de trabajo el desplegamiento y replegamiento de tiorredoxina y sus formas mutantes se describen cinéticamente mediante una única exponencial.

El ajuste no lineal por mínimos cuadrados de las ecuaciones 2.45, 2.46 y 2.47 a los perfiles de $\ln k$ versus C obtenidos a 25°C ha sido siempre excelente, como se observa en la Figura 2.14, lo que apoya la validez del modelo cinético de dos estados. El análisis de los datos experimentales se ha hecho utilizando el programa comercial MLAB y MatLab (Apéndice I, programas 5 y 6). Como ya se ha explicado anteriormente, debemos introducir los datos experimentales en parejas (C , $\ln k$) dispuestas en una matriz $M_{N \times 2}$ (Apéndice I, programa 4).

2.2.6.3 Estados de transición y método de los valores de ϕ

El análisis de valores de ϕ es el único método con el cual se puede determinar la estructura del estado de transición a la resolución de aminoácidos individuales. El conocimiento de las estructuras del estado de transición es de elevada importancia para entender el problema del plegamiento, ya que el estado de transición representa, por definición, el punto de mayor energía del plegamiento de una proteína. El conocimiento de las estructuras del estado de transición nos puede dar información sobre los procesos que determinan la velocidad de plegamiento y puede usarse para rediseñar la velocidad de plegamiento de un modo racional.

El ajuste de nuestros datos experimentales a un modelo de dos estados (linealidad de las dos ramas del *Chevron*, véase Figura 2.14) nos indica que las únicas especies pobladas durante los procesos de replegamiento y desplegado son el estado D y el estado N. El estado de transición presenta una elevada energía; esta característica del estado de transición es probablemente debida a las moléculas de agua que están asociadas a los residuos que formarán el núcleo de la proteína. Durante las últimas etapas del plegamiento, se produce la expulsión de las moléculas de agua y la sustitución de sus interacciones con la proteína, por puentes de hidrógeno intramoleculares. Debido a sus características termodinámicas, el estado de transición resulta difícil de caracterizar estructural y energéticamente. Con el método de los valores de ϕ se obtiene información acerca de la importancia en el estado de transición de las interacciones que establece la cadena lateral de cada residuo, lo cual puede ser utilizado para inferir información sobre la estructura de este estado.

El principio del método aplicado a una proteína sencilla que se pliega en ausencia de intermediario, es el siguiente: i) mediante mutagénesis dirigida se elimina la interacción de una cadena lateral con otros residuos; ii) mediante análisis termodinámico se calcula la constante de equilibrio y de ahí, la energía libre del proceso de plegamiento para la proteína no mutada, ΔG_{N-D}^{WT} y para la variante mutante, ΔG_{N-D}^{mut} , y la variación de energía libre de plegamiento causada por la mutación, $\Delta\Delta G_{N-D}$; iii) mediante análisis cinético (apartado 2.2.6.2) se calcula la constante de velocidad y de ahí la energía libre de activación de

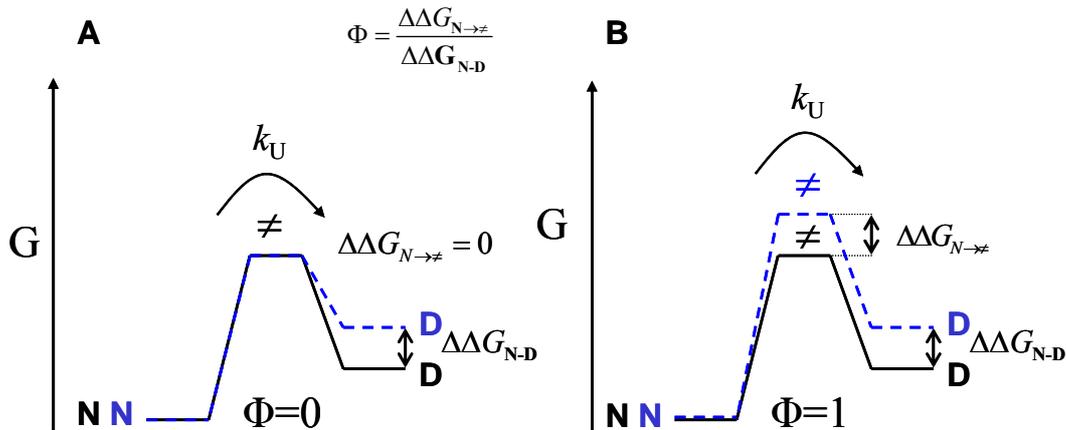


Figura 2.16: Representación del análisis de valores de ϕ realizado. Para el caso de una proteína que despliega por un mecanismo dos estados, se ilustra: A) el caso de una mutación que origina una variación de energía libre de plegamiento, $\Delta\Delta G_{N-D}$; para este caso se calcula un valor de $\phi=0$; ésto implica que la interacción está igualmente formada en el estado de transición y en el estado N. B) Si la mutación origina una variación de energía libre de activación causada, $\Delta\Delta G_{\ddagger-D}$ y además, una variación de energía libre de plegamiento, $\Delta\Delta G_{N-D}$, se calcula un valor $\phi=1$; ésto implica que la interacción no está formada en el estado de transición, es decir, la región de la mutación se encuentra desnaturalizada en el estado de transición. El valor de ϕ , depende de la cantidad de estructura que se ha formado en el estado de transición en la posición de la mutación. De este modo, estudiando el efecto sobre la estabilidad de un conjunto de mutaciones estratégicamente distribuidas sobre la proteína, se puede mapear la estructura del estado de transición.

la reacción para la proteína no mutada, ΔG_{N-D}^{WT} , para la mutante, ΔG_{N-D}^{mut} , y la variación de energía libre de activación causada por la mutación, $\Delta\Delta G_{\ddagger-D}$, y iv) se calcula el valor $\phi = \Delta\Delta G_{\ddagger}^{\ddagger} / \Delta\Delta G_{N-D}$. Un valor de $\phi=0$ implica que la interacción está igualmente formada en el estado de transición y en el estado N y un valor $\phi=1$ implica que la interacción no está formada en el estado de transición, es decir, la

región de la mutación se encuentra desnaturalizada en el estado de transición. Un valor fraccionado de ϕ implica que la interacción no está completamente formada en el estado de transición. De este modo, realizando un número suficiente de mutaciones, puede definirse qué regiones del estado de transición presentan una estructura nativa y cuáles no presentan estructura nativa formada.

2.2.6.4 Ensayos de desplegamiento

Ya se ha visto en la sección 2.2.4.1 cómo la no coincidencia de las entalpías calorimétrica y de van't Hoff puede ser un indicativo de la presencia de intermedios significativamente poblados en el desplegamiento térmico de la proteína. En el caso de transiciones inducidas por cambios en el disolvente, no existe un criterio similar para confirmar la validez del modelo de dos estados. A menudo la coincidencia de varios perfiles obtenidos siguiendo propiedades espectroscópicas diferentes, tales como absorbancia, fluorescencia, elipticidad, etc. se ha tomado como una buena evidencia a pesar de que intermedios que mostrasen propiedades similares al estado nativo no serían detectados. Por tanto, para determinar la naturaleza del desplegamiento inducido por desnaturalizantes, no es suficiente con seguir cambios en las propiedades físicas ya que éstas contienen contribuciones de todas las especies presentes, que en ocasiones pueden no ser fácilmente separables. En cambio medir directa e independientemente la concentración de moléculas en estado nativo, desplegado y posibles intermedios, en función de la concentración de desnaturalizante constituye una prueba más determinante.

En este trabajo de investigación hemos aplicado un método cinético, propuesto por Mücke y Schmid, 1994, para determinar la concentración de moléculas de proteína en el estado nativo. En efecto, es posible conocer la fracción molar de proteína presente en el estado nativo, dadas unas condiciones determinadas de concentración de desnaturalizante y temperatura, a partir de las cinéticas de desnaturalización observadas al transferir la proteína a condiciones fuertemente desnaturalizantes. En estas condiciones, los estados parcialmente desplegados normalmente desnaturalizan más rápidamente que el estado nativo (Schmid, 1992), de manera que es posible seguir independientemente el desplegamiento de éste y calcular la cantidad presente en la muestra original a partir de la amplitud de las cinéticas observadas.

Hemos aplicado los ensayos de desplegamiento para determinar la fracción de estado nativo en muestras de proteína incubadas en presencia de diferentes concentraciones de desnaturizante en condiciones de equilibrio.

Para cinéticas de primer orden la dependencia de la intensidad de fluorescencia con el tiempo viene dada por la ecuación 2.40. Es necesario realizar un experimento control en el que se siga la cinética de desplegamiento de la proteína, transferida desde tampón acuoso hasta exactamente las mismas condiciones de los ensayos de desplegamiento de las muestras. Así la amplitud de esta cinética, ΔI^0 , corresponderá al 100% nativo ($x_N=1$). La comparación de este control con el resto de las muestras nos dará una doble información:

1.- Si los valores de τ para cada cinética de desplegamiento permanecen constantes, indicaría que las moléculas de N que existen en el rango de concentración de desnaturizante en donde tiene lugar la transición, son indistinguibles de aquéllas que existen en condiciones nativas.

2.- La amplitud relativa de la cinética de desplegamiento con respecto al control es una medida directa de la fracción de moléculas de proteína que están en el estado nativo, en las condiciones en las se encontraba la muestra original:

$$x_N = \frac{\Delta I}{\Delta I^0} \quad (2.52)$$

En la Figura 2.18-A hemos representado algunos perfiles cinéticos correspondientes a ensayos de desplegamiento realizados a muestras incubadas a diferentes concentraciones de urea. Claramente la amplitud de estas cinéticas va disminuyendo al aumentar la concentración de desnaturizante en la muestra original.

Si disponemos del perfil x_N concentración de desnaturizante (Figura 2.18-B) podemos compararlo con los perfiles derivados del cambio de cualquier propiedad espectroscópica por efecto del disolvente. Si las curvas coinciden, indicaría que la presencia de intermedios de características diferentes a D es poco probable. Por el contrario si la concentración de moléculas en estado nativo decrece más rápidamente que la propiedad física medida, entonces ésto sugeriría que existe un intermedio significativamente poblado que es diferente a D en esa propiedad y que se despliega más rápidamente en el ensayo de desplegamiento.

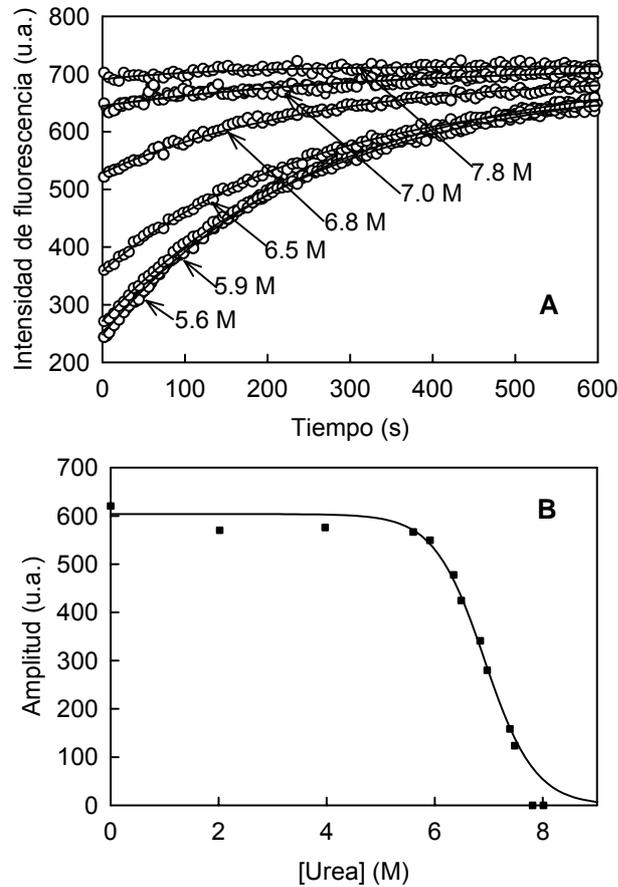


Figura 2.17: Determinación del perfil de desplegamiento inducido por urea en equilibrio, mediante ensayos de desplegamiento de doble salto, para tiorredoxina *WT* de *E. coli*. A) Para registrar los perfiles de intensidad de fluorescencia *versus* tiempo se ha transferido la disolución de proteína (previamente incubada a 25°C y a las concentraciones de urea mostradas) hasta condiciones fuertemente desnaturizantes (8.5 M). La línea continua representa el mejor ajuste por mínimos cuadrados de la ecuación 2.40 a los datos experimentales (círculos). B) Amplitudes de desplegamiento calculadas del ajuste de la ecuación 2.40 a las cinéticas de desplegamiento mostradas en A. Estas amplitudes son proporcionales a la cantidad de proteína en estado nativo en la solución inicial y se representan frente a la concentración de urea en esas soluciones. La línea continua se ha predicho con la ecuación 2.53 con los parámetros

Las ventajas que presenta este método cinético a la hora de caracterizar una transición inducida por codisolvente frente al procedimiento “clásico” de seguimiento de una propiedad espectral son:

- a) No es necesario presuponer ningún tipo de mecanismo de desplegamiento a priori ya que medimos directamente x_N .
- b) No se requiere ninguna extrapolación de líneas base; obviamente éstas serán igual a 1 y 0 en la pre- y post-transición.

Por otra parte, también es importante señalar que el ensayo de desplegamiento cinético descrito sería capaz de discriminar entre moléculas nativas y parcialmente desplegadas, pero no distinguiría entre éstas y las desplegadas. Este test tampoco detectaría tampoco intermedios que desplegaran tan lentamente como lo hace la proteína nativa, aunque tales especies son poco frecuentes debido a la gran barrera energética que separa ambos estados.

A) Procedimiento experimental y análisis de los perfiles cinéticos:

La proteína se ha dializado en tampón HEPES 5 mM, pH7 preparado como se indica en la sección 2.2.3.1. La disolución de urea se preparó como se indicó en el apartado 2.2.3.2. Los parámetros instrumentales relativos al fluorímetro se indican en el apartado 2.2.6.1.

La concentración de proteína incubada en presencia de desnaturizante (tampón HEPES 5 mM, pH 7), hasta alcanzar el equilibrio fue de 10 mg/mL. En todos los casos las cinéticas de desplegamiento se han iniciado diluyendo una alícuota de la muestra en condiciones fuertemente desnaturizantes, siendo la concentración final de proteína en la célula del fluorímetro de aproximadamente 0.04 mg/mL. La dilución de la disolución de proteína se hace con una disolución de compensación de urea, que en cada caso contiene, la concentración de urea necesaria para compensar la disolución de proteína, en urea (tampón HEPES 5 mM, pH 7), de partida. De esta manera, se consigue siempre la misma concentración final de urea en la cubeta.

Interesa que la concentración de desnaturizante final sea alta, para asegurarnos que la proteína se va a desplegar completamente, pero debemos tener en cuenta que al aumentar la concentración de desnaturizante el tiempo de relajación disminuirá. Por tanto, hay que llegar a una situación de compromiso para conseguir tiempos de relajación mayores a varios minutos, claramente

superiores al tiempo de respuesta del instrumento, pero a concentraciones de urea superiores al rango donde ocurre la transición. Así pues, las condiciones elegidas para realizar los ensayos de desplegamiento fueron 25°C y una concentración en torno a 8.5 M de urea en la cubeta del fluorímetro. En estas condiciones el tiempo de relajación del estado nativo es aproximadamente 180 .segundos.

En todos los casos se ha medido el índice de refracción de la disolución final de proteína, para conocer la concentración exacta de urea (ecuación 2.5). Los perfiles de intensidad de fluorescencia en función del tiempo, para cada concentración de urea a la que hemos realizado los ensayos de desplegamiento, se han analizado haciendo uso de la ecuación de velocidad de primer orden (ecuación 2.40) de manera que son tres los parámetros ajustables: I_{∞} , ΔI y τ . Como ya se ha explicado anteriormente, la amplitud de la exponencial en cada caso, ΔI , va a ser proporcional a la fracción de proteína en estado nativo en la muestra original (ecuación 2.52).

En la Figura 2.17 se muestran algunas de las cinéticas correspondientes a los ensayos de desplegamiento para diferentes concentraciones de urea. El ajuste de la ecuación 2.40 a los datos experimentales ha sido excelente en todos los casos. Como ya comentado en esta sección, el hecho de que el valor de τ permanezca constante, dentro del error experimental, supone un control en tanto en cuanto nos asegura que, en primer lugar, las moléculas nativas presentes en la transición son indistinguibles de aquéllas que existen en condiciones nativas y, en segundo lugar, que las condiciones finales de concentración de urea y temperatura en la célula de medida son las mismas en todos los casos. No obstante, se ha podido observar que a alta concentración de desnaturizante el ruido en el tiempo de relajación aumenta. Esto es lógico si pensamos que en estas condiciones la población de estado nativo es prácticamente nula, con lo que la amplitud de las cinéticas observadas, ΔI (que es proporcional a x_N), tiende a cero de manera que los algoritmos usados en el programa de análisis tienen dificultad a la hora de encontrar los parámetros óptimos del ajuste. Por esta razón, se ha fijado a cero la amplitud de las cinéticas que ocurren a alta concentración de urea

La subrutina de ajuste en el entorno MatLab de los perfiles de intensidad de fluorescencia en función del tiempo es la que se ha detallado en el apartado 2.2.6.2 (Apéndice I, Programa 2).

B) Análisis de los perfiles de ΔI versus C según el modelo de dos estados:

Podemos representar los valores de ΔI obtenidos del análisis de las cinéticas correspondientes a los ensayos de desplegamiento realizados a 25°C versus a concentración de urea (Figura 2.17-B).

Si se supone un modelo de dos estados para el proceso de desplegamiento, entonces a partir de la ecuación 2.53 se llega a la expresión que liga ΔI con C (Ibarra-Molero y Sanchez-Ruiz, 1996, 1997):

$$\Delta I = \frac{\Delta I^0}{1 + e^{m_{1/2}(C-C_{1/2})/RT}} \quad (2.53)$$

donde ΔI^0 , $m_{1/2}$ y $C_{1/2}$ son los parámetros ajustables. C es la concentración de desnaturalizante (urea), $C_{1/2}$ es la concentración de urea a la cual la energía libre de desplegamiento es cero y $m_{1/2}$ da la dependencia de la concentración de urea de la energía libre de desplegamiento a la $C_{1/2}$:

$$m_{1/2} = -\left(\frac{\partial \Delta G}{\partial C}\right)_{C_{1/2}} \quad (2.54)$$

En este tipo de análisis las líneas base de los estados nativo y desplegado están perfectamente definidas (líneas planas y a una altura igual a ΔI^0 y 0, respectivamente), con lo que se evita la incertidumbre asociada a la determinación de líneas base.

Los ajustes no lineales por mínimos cuadrados de la ecuación 2.40 a los perfiles experimentales fueron siempre excelentes. La subrutina de ajuste utilizada se describe en Apéndice I, Programas 7 y 8. Por último, es importante destacar que así como en la ecuación 2.53 se supone que la desnaturalización sigue el modelo de dos estados, en el cálculo de x_N como es igual a $\Delta I/\Delta I_0$ (ecuación 2.53) no se hace ninguna suposición a priori en este sentido, de manera que constituye un excelente método para comprobar la validez del modelo de dos estados (Ibarra-Molero y Sanchez-Ruiz, 1996, 1997).

Tanto los ajustes de los *Chevron* (programas 4, 5 y 6 de Apéndice I) como los de los perfiles de equilibrio de amplitudes frente a la concentración de urea (programas 7 y 8), se han realizado con programas escritos en entorno MataLab y MLAB, con objeto de comparar los errores obtenidos en ambos entornos. En el Capítulo 5, se explicará cómo llevan ajustes globales de los *Chevron* y los perfiles de equilibrio de amplitudes frente a concentración de urea, con el fin de acotar el error de los parámetros ajustables.

2.2.7 COMPARACIÓN DE SECUENCIAS DE PROTEÍNAS Y EVOLUCIÓN DE PROTEÍNAS

2.2.7.1 Conceptos generales sobre evolución

La comparación de secuencias de proteínas es una herramienta muy poderosa para caracterizar secuencias de proteínas y extraer información acerca de su historia evolutiva, pertenecientes a diferentes organismos, a lo largo de billones de años. En la Figura 2.18 se muestra un árbol de la vida (figura extraída de Dayhoff *et al.*, 1978), en el que se puede seguir la evolución, desde hace 4 billones de años en la historia de la tierra, de los diferentes grupos de organismos.

La similaridad existente entre secuencias de proteínas es una medida de la extensión en la que las secuencias de proteínas están relacionadas. Mediante una comparación de secuencias se pueden detectar secuencias de proteínas *homólogas*. Es decir, si dos secuencias de proteínas presentan una similaridad significativa, muy probablemente esas proteínas comparten un ancestro común en el pasado. Para que dos proteínas sean *homólogas* deben de poseer al menos un 0.25-0.30 de similaridad en su secuencia y compartir la misma estructura dimensional. Las secuencias *homólogas* se pueden dividir en dos grupos: (1) secuencias *ortólogas*: secuencias que se diferencian porque se encuentran en diferentes especies. Es decir, secuencias que a pesar de tener la misma estructura tridimensional y función biológica presentan una secuencia aminoacídica diferente debido probablemente a la divergencia evolutiva de especies; (2) secuencias *parálogas*: secuencias que se diferencian debido a una duplicación de los genes que codifican para la secuencias de las proteínas y por lo cual divergen de modo distinto en la evolución.

A lo largo de la evolución se pueden distinguir selección natural y patrones de sustituciones de aminoácidos. La sustitución de un aminoácido (mutación no-sinónima) es el resultado de dos procesos: en primer lugar, la frecuencia de aparición de la mutación en el gen que codifica la proteína; en segundo lugar, es la fijación de dicha mutación por la población (especies), como la nueva forma predominante. Las razones por las que se pueden dar las sustituciones pueden ser principalmente dos: la primera, debida a la degeneración del código genético (es más probable que ocurran mutaciones que exijan el cambio de una sola base); la segunda, de tipo funcional (el nuevo aminoácido debe tener una función similar al que existía previamente). De esta forma, en la evolución se aceptan las denominadas mutaciones conservativas (entre aminoácidos física y químicamente similares) con efectos casi-neutros (Lázló, 1999).

Las frecuencias con las que han ocurrido las mutaciones a lo largo de la evolución quedan recogidas en las matrices de mutaciones: PAM (desarrolladas por Dayhoff y su grupo, Dayhoff *et al.*, 1978, ver Tabla 2.2-a) y BLOSUM Henikoff y Henikoff, 1992, ver Tabla 2.2-b). Estas matrices se basan en una base de datos que recoge 1572 cambios dentro de 71 grupos de proteínas altamente relacionadas entre sí. Las matrices de PAM muestran el porcentaje de mutaciones aceptadas y éste ha sido calculado de los diferentes tipos de cambios observados en esas secuencias de proteínas. Cada cambio se ha tabulado y se ha introducido en la matriz, de manera que ésta nos muestra todos los posibles cambios entre aminoácidos. Si se observa la Tabla 2.2, se puede ver como cada matriz posee coeficientes distintos. Ésto es debido a que en las matrices de Dayhoff el porcentaje de mutaciones conservadas se calcula entre secuencias de proteínas muy relacionadas (secuencias con al menos un 85% de similaridad). Las matrices de BLOSUM, sin embargo realizan predicciones de mutaciones sobre bloques de secuencias que pueden tener una similaridad menor (por ejemplo, las matrices derivadas de una base de datos de segmentos de secuencias alineados con una similaridad del 45% y 80% se denominan BLOSUM 45 y BLOSUM 80, respectivamente). Además, consideran distintos porcentajes de conservación de residuos polares/apolares (0.49 en matrices de Dayhoff frente a 0.93 en matrices BLOSUM). Ésto puede ser debido al hecho de que las regiones de la superficie de las proteínas (lazos, β -giros), que son más variables, se encuentran

subrepresentadas, y por el contrario, las regiones más conservadas (elementos de estructura secundaria) que forman el núcleo hidrofóbico de las proteínas, se encuentran sobrerrepresentadas. Es decir, el bajo porcentaje de conservación de residuos polares en las matrices de Dayhoff se debe principalmente al hecho de que éstas consideran la conservación de la hidrofiliidad más que la del propio residuo en si (Lázló, 1999).

2.2.7.2 Alineamientos de secuencias realizados para tiorredoxina de *E. Coli*

Los alineamientos de secuencias se obtuvieron realizando una búsqueda con BLAST 2 (1996-2003, W. Gish, <http://blast.wustl.edu>) en la base de datos UniProt/TrEMBL (<http://www.ebi.ac.uk/trembl/>), utilizando la secuencia de tiorredoxina de *E. coli* como secuencia a comparar (secuencia *query* en la búsqueda) con el resto de secuencias de la base de datos. Para hacer el alineamiento se dejaron las variables de BLAST 2 fijadas por defecto. Las secuencias resultantes de la búsqueda se alinearon a la secuencia *query* usando el algoritmo de Smith-Waterman (Smith y Waterman, 1981), método muy riguroso que calcula, de forma local, la similaridad entre dos secuencias.

De todas las secuencias alineadas se han filtrado aquellas que tenían una similaridad con la secuencia *query* mayor del 0.25 (las similaridades se calcularon como el número de coincidencias entre cualquier secuencia alineada con la secuencia *query* y el número de residuos de ésta última). Es decir, todas se han rechazado todas aquellas secuencias con una similaridad con la secuencia *query* menor de 0.25% (ver fragmento del alineamiento de secuencias que se muestra en la Figura 2.19). El corte de 0.25 de similaridad se ha elegido porque se considera que secuencias pertenecientes a distintas especies que tienen una similaridad con la secuencia *query*, en secuencia de al menos 0.25 tienen una estructura tridimensional similar (ver secciones 3.2 y 4.2; Godoy-Ruiz *et al.*, 2004, 2005, Dokholyan y Shakhnovich, 2001 y referencias incluídas en este artículo). Por lo tanto, en la búsqueda realizada para tiorredoxina de *E. coli* (secuencia *query*) en este trabajo de investigación se obtuvo un alineamiento de un total de 491 secuencias, para el que se asumió que la mayoría de las secuencias alineadas compartían la misma estructura tridimensional que la secuencia *query*. Además, hay que resaltar que las 491 secuencias alineadas, resultaron con un parámetro *p*

con la secuencia *query* $<10^{-6}$ (el parámetro *p* es la significancia estadística, es decir, es la probabilidad de que el alineamiento de una secuencia dada con la secuencia *query* esté ocurriendo al azar. Se considera que la búsqueda es significativa si $p > 10^{-2}$). Más concretamente, se pudo comprobar que del total de las 491 secuencias alineadas, 428 secuencias se alinearon con un valor de $p < 10^{-6}$, 479 secuencias con $p < 10^{-4}$ y 489 secuencias con $p < 10^{-2}$.

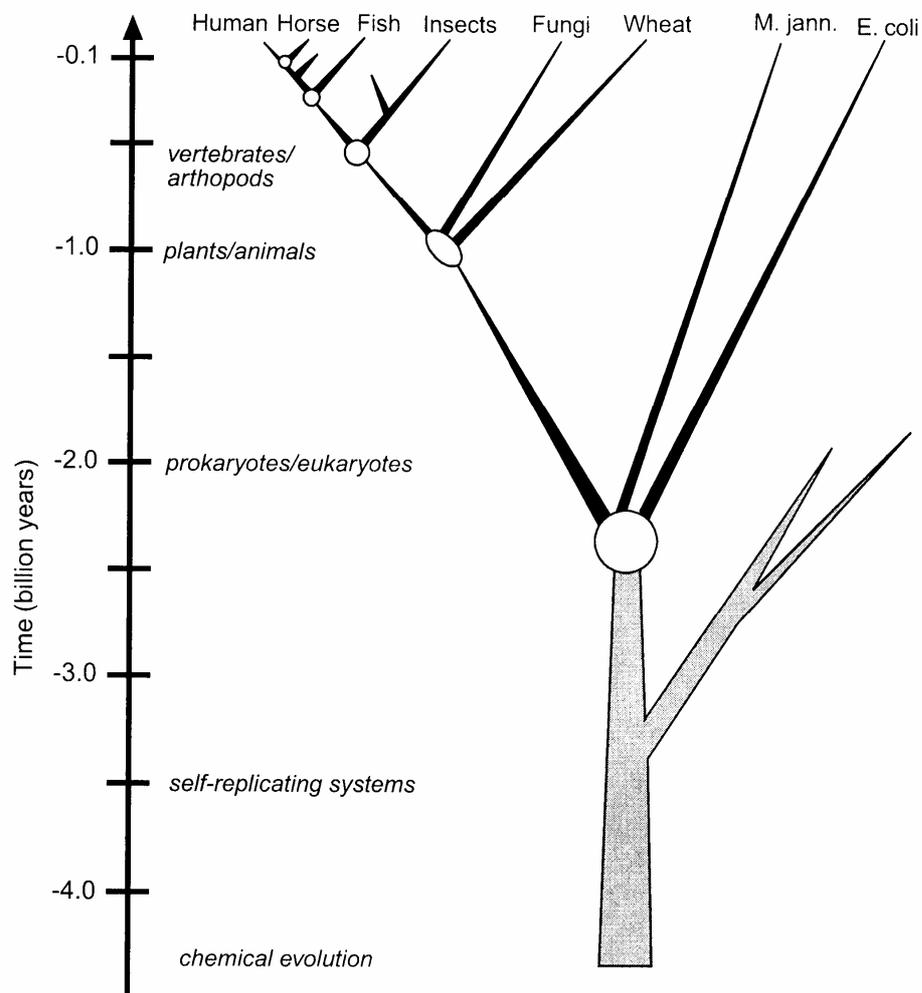


Figura 2.18: Árbol de la vida. Figura extraída de Dayhoff *et al.*, 1978.

Mediante un estudio mucho más exhaustivo (sección 4.2; Godoy-Ruiz *et al.*, 2005), se ha determinado la clasificación biológica del alineamiento de secuencias obtenido para tiorredoxina de *E.coli*. Del total de las 491 secuencias alineadas, 219 secuencias pertenecían a la taxa de bacteria, 249 a la de eucariota y 18 a la de archaea (y para 5 secuencias, no fue posible encontrar su clasificación taxonómica en UniProt/TrEMBL). Dentro de la taxa de eucariota, 123 secuencias pertenecían al grupo de metazoa, 81 a viridiplantae y en menor cantidad también se pudieron observar secuencias pertenecientes al grupo de fungi, alveolata, etc. Dentro del grupo de secuencias clasificadas como bacteria, 101 secuencias resultaron ser proteobacterias, 51 secuencias firmicutes y en menor cantidad también se recontaron cianobacteria, actinobacteria, etc. Tan sólo 62 secuencias dentro del grupo de proteobacterias correspondían a γ -proteobacterias (clase a la que pertenece *E. coli*).



Figura 2.19: Fragmento del alineamiento de secuencias (archivo TRX25.aln) obtenido al utilizar la secuencia de tiorredoxina de *E. coli* como secuencia a comparar (secuencia query), en una búsqueda realizada con BLAST2 (1996-2003, W. Gish, <http://blast.wust1.edu>) en la base de datos UniProt (<http://www.ebi.ac.uk/trembl>) y dejando las opciones de búsqueda de BLAST 2 fijadas por defecto.

2.2.8 LA CRISTALIZACIÓN DE PROTEÍNAS

La cristalización de moléculas en solución es un fenómeno de equilibrio químico reversible cuya cinética y termodinámica están controladas por las características físicas (temperatura, gravedad, superficie, viscosidad, constante dieléctrica o vibraciones) y químicas (pH, fuerza iónica, iones metálicos o detergentes) del solvente y del soluto. Bajo ciertas condiciones el sistema segrega de la solución un precipitado que disminuye la energía libre del sistema (núcleo crítico o un agregado no específico) (Weber, 1991). Esta disminución de energía libre es tanto mayor cuanto más ordenada sea la fase segregada.

El fenómeno de crecimiento cristalino, típicamente se considera dividido en tres fases o etapas; nucleación, crecimiento y cese del crecimiento. La formación de un núcleo, proporciona al sistema superficie para seguir creciendo. La nucleación es el proceso por el cual una fase sólida se segrega de la solución para formar agregados cristalinos termodinámicamente estables. Cuando la nucleación ocurre en una solución libre de partículas y superficies extrañas se denomina nucleación homogénea. La presencia de partículas extrañas puede provocar la aparición de los núcleos a unos valores de sobresaturación inferiores a los requeridos para que se produzca la nucleación homogénea (nucleación heterogénea). El crecimiento de un cristal implica una serie de pasos que describen la migración de las moléculas desde el seno de la solución a la superficie del cristal y su posterior incorporación desplazando las moléculas de solvente. Hay tres causas que provocan que un cristal deje de crecer y sea de baja calidad: el consumo de las unidades de crecimiento (se ha alcanzado una situación de equilibrio termodinámico), el incremento de la energía de superficie por acumulo de imperfecciones o la incorporación de impurezas sobre las caras del cristal (Kam *et al.*, 1978; Feher y Kam, 1985). Las impurezas (moléculas de proteína con cierta desnaturalización, fragmentos proteícos u otras proteínas copurificadas, partículas de polvo) no se incorporan perfectamente, ya que el número de contactos efectivos es menor que para una molécula en estado nativo. Sin embargo, conforme el cristal crece la concentración de impurezas aumenta.

Todos estos elementos se oponen al principio de uniformidad, y por tanto, a la cristalización. A continuación se explican los requerimientos mínimos de las disoluciones de proteína que se pretenden cristalizar, así como los pasos de

búsqueda y refinamiento de las condiciones de cristalización, y la metodología de cristalización empleada.

2.2.8.1 Características de la solución de proteína

Las disoluciones de proteína que se desean cristalizar deben reunir principalmente las siguientes características (McPherson, 1999, McPherson, 1982 y Ducruix y Giegé, 1999):

1. Una concentración elevada.
2. Una pureza tan alta como sea posible.
3. Homogeneidad y estabilidad en las condiciones de cristalización.

Más concretamente, las condiciones que se han utilizado en este trabajo han sido las siguientes:

Para alcanzar la sobresaturación, las soluciones de proteína se concentraron siempre hasta una concentración de 30 mg/mL. Concentraciones tan elevadas las conseguimos utilizando células de ultrafiltración de MILLIPORE (apartado 2.1.2).

La pureza de la solución de proteína debe de ser tan alta como sea posible, aceptándose comúnmente que al menos ha de ser del 95% (determinada mediante electroforesis de geles de poliacrilamida, *PAGE-SDS*; véase sección 2.2.2.6). Las impurezas pueden impedir la nucleación o limitar el tamaño de los cristales, ya que su interposición puede abortar el crecimiento del cristal, y/o introducir irregularidades que disminuyan la calidad del cristal obtenido. Por eso, cuando los cristales que se obtienen con una proteína siempre difractan mal o son pequeños, es buena estrategia revisar el grado de pureza e intentar aumentarla.

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.). Hay que evitar durante la preparación de la proteína someterla a tratamientos que puedan introducir heterogeneidad, tales como los tratamientos térmicos, que pueden producir pequeños cambios aun en proteínas que los resisten (manteniendo su actividad). Igualmente, si una proteína se eluye de una determinada purificación en dos picos distintos debe intentarse la cristalización por separado de la proteína presente en cada pico, sin mezclarlos. No es recomendable usar la liofilización como modo de

preservar las proteínas para cristalizar, ya que muchas veces, al redissolver la proteína liofilizada, una parte de ella agrega, lo que generalmente entorpece fuertemente el proceso de cristalización. La congelación y descongelación sucesivas tampoco son recomendables, ya que pueden también causar que una fracción de la proteína agregue o se desnaturalice, al menos parcialmente.

Las condiciones de cristalización (pH, temperatura, adición de ligandos y de protectores de grupos –SH) deben buscar preservar al máximo la estabilidad de la proteína. Ello afecta a la conservación de la proteína tras su purificación así como a la solución en la que se disuelve para intentar la cristalización.

2.2.8.2 Metodología de cristalización utilizada: cristalización por contradifusión

La estrategia general para la cristalización de proteínas es reducir la solubilidad de la macromolécula hasta que ésta alcanza el estado de sobresaturación. Existen diversas formas de llevar una solución de proteína a un estado sobresaturado (McPherson, 1999): por apantallamiento de la carga, por eliminación de moléculas de agua, por disminución de la constante dieléctrica del medio o por descenso de la temperatura. Se han desarrollado diferentes técnicas para llevar una solución subsaturada de proteína a la región sobresaturada donde la nucleación y el crecimiento son posibles. Las técnicas más empleadas son la de mezcla directa (“*batch*”) o la de difusión en fase vapor. La cristalización por el método de *batch* se consigue mediante la mezcla directa de la solución de proteína con un agente precipitante. Como resultado de la adición del agente precipitante, se consigue la sobresaturación en la solución de proteína (Figura 2.20-A). En la técnica de difusión en fase vapor, la solución de proteína se mezcla con la solución de agente precipitante (en una gota sentada o colgante de una superficie) para que la mezcla se sitúe en la región sobresaturada del diagrama de solubilidad. La gota se enfrenta, en un sistema cerrado, a una solución que contiene el agente precipitante (reservorio) a mayor concentración. La migración del agua de la gota al reservorio (ley de Raoult) aumenta la concentración de agente precipitante y de proteína en la gota. El paulatino aumento de concentración, representado como una trayectoria en el diagrama de solubilidad, consigue que la solución entre en la región sobresaturada en la que pueden tener lugar la nucleación y el crecimiento de los cristales (Figura 2.20-B).

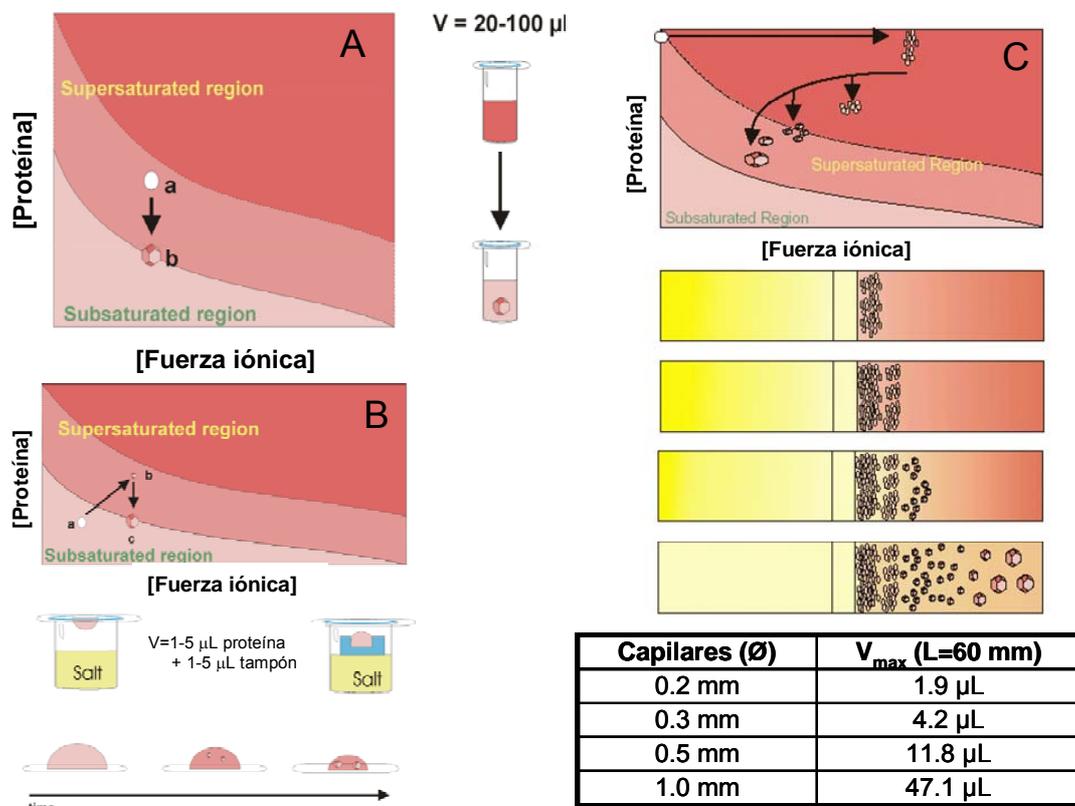


Figura 2.20: Comparación entre las técnicas de cristalización convencionales de difusión en fase vapor y cristalización por contradifusión en capilares. Los diagramas de fases muestran las zonas sobresaturadas en un rango de concentraciones de proteína crecientes como una función de la fuerza iónica en la disolución de proteína. Las regiones lábiles y metaestables se representan en rojo oscuro y rojo claro, respectivamente (A-C). La región subsaturada se representa en amarillo. Durante el proceso de cristalización, la disolución de proteína viaja hacia la región sobresaturada disminuyendo la solubilidad de la proteína. En la región lábil ocurre la nucleación, acto seguido la solubilidad de la disolución de proteína disminuye hasta la región metaestable donde el crecimiento del cristal es más favorable (a→b→c). En la tabla, se muestran los distintos diámetros de capilar que se pueden usar y el volumen de disolución de proteína necesario en cada caso. Figura extraída de Ng *et al.*, 2003.

Una nueva técnica que ha venido tomando auge en los últimos años es la contradifusión, basada en la difusión líquido-líquido. Para desarrollarla la solución de agente precipitante y la de proteína se ponen en contacto en un medio en el que el transporte de masa está controlado por la difusión. Este requerimiento se satisface en condiciones de gravedad reducida, en medios gelificados o en capilares, o en combinación de éstos (García-Ruiz, 1991). Otros dos requerimientos necesarios para que esta técnica desarrolle todo su potencial son unas condiciones iniciales muy alejadas del equilibrio (elevada concentración de agente precipitante o de proteína) y una dimensión de desarrollo. En un capilar, para amortiguar los elevados valores de sobresaturación que se alcanzarían al poner ambas cámaras (la de agente precipitante y la de proteína) en contacto se añade una tercera cámara o tampón físico, entre éstas (Figuras 2.20-C y Figura 2.21). La difusión del agente precipitante, con un mayor coeficiente de difusión que la proteína, en la cámara de proteína originará un gradiente continuo de sobresaturación hasta alcanzar condiciones de equilibrio que será dependiente del espacio y del tiempo a lo largo de todo el capilar (García-Ruiz, 2003; Ng *et al.*, 2003). De esta forma, el transporte de masa y la precipitación son consecutivos, y a diferencia de las otras técnicas, la precipitación, nucleación y crecimiento del cristal pueden ocurrir simultáneamente en el capilar. La cinética del proceso completo estará controlada por el proceso que tenga una cinética más lenta. Para reducir la velocidad de sobresaturación, la mejor estrategia es minimizar la cinética del proceso de transporte de masa. Y esto a su vez se consigue minimizando los procesos de convección y permitiendo que el transporte de masa ocurra según las leyes de la difusión. Además, se ha demostrado que los geles son un medio muy bueno para crecer los cristales de alta calidad (para el caso de agarosa, una concentración de 0.12% p/v es suficiente para eliminar la convección). Además, cuando los cristales crecen en geles, permanecen todo el tiempo suspendidos, con lo cual se evita la sedimentación (Ng *et al.*, 2003).

En resumen, en un experimento de cristalización por contradifusión se recorre una amplia zona en el diagrama de fases, lo que permite buscar las condiciones de cristalización que aseguran cristales de elevada calidad (García-Ruiz, 2003). La formación del cristal ocurre como resultado de la progresión del frente de nucleación resultante de los tres procesos que tienen lugar: transporte de masa, nucleación y crecimiento del cristal. De esta manera, los cristales obtenidos

son el resultado de un crecimiento selectivo en diferentes ambientes de sobresaturación (García-Ruiz, 2003 y Carotenuto *et al.*, 2002).

Existen diferentes formas de implementar la contradifusión (García-Ruiz, 2003). El empleo de capilares de rayos-X de 0.1 mm de diámetro asegura la minimización de los fenómenos convectivos y reduce significativamente la cantidad de proteína necesaria para llevar a cabo los experimentos. Además, el crecimiento de los cristales en el interior de un capilar evita el estrés hidrodinámico y mecánico durante su crecimiento y elimina la manipulación de los cristales cuando estos se preparan para la difracción. A diferencia de los otros métodos de cristalización en los que la difracción se realiza utilizando un lazo como soporte del cristal, la cristalización en capilares permite difractar el cristal en el capilar, sin necesidad de extraerlo del capilar; de este modo, al no existir manipulación se evita la degradación del cristal (Gavira *et al.*, 2002).

Una vez que se han establecido el conjunto de condiciones de cristalización, el siguiente paso es su optimización, con objeto de producir cristales de elevada calidad para la difracción de rayos-X. Ésto se consigue variando los parámetros físico o químicos, mencionados anteriormente, que afectan a la solubilidad de la proteína.

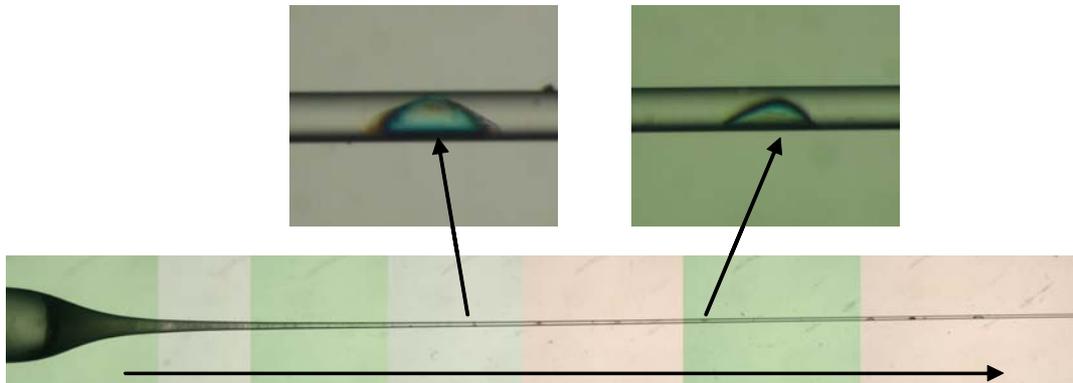


Figura 2.21: Ejemplo ilustrativo de crecimiento cristalino utilizando la técnicas de contradifusión. Se trata del mutante I23V de tiorredoxina; en la imagen se muestra el gradiente en el crecimiento del cristal resultado del acoplamiento entre difusión y precipitación. La precipitación de la proteína tiene lugar cuando su solubilidad varía con la concentración de agente precipitante. Cuando las dos soluciones entran en contacto, la sobresaturación es muy elevada y se forman los precipitados (comienzo del capilar) suelen ser amorfos o de una baja calidad cristalina. Conforme el agente precipitante difunde en la cámara de proteína (dirección indicada por la flecha negra), tienen lugar nuevas precipitaciones de proteína, esta vez bajo condiciones de sobresaturación menores. Ésto favorece la formación de cristales de mayor tamaño y una calidad mayores. De esta forma la cristalización en contradifusión explora un amplio número de condiciones de cristalización en un sólo experimento.

2.2.8.3 Condiciones de cristalización elegidas.

Las distintas condiciones bajo las que se ha cristalizado tiorredoxina silvestre (*WT*) y mutantes previamente se muestran en la Tabla 2.3:

Tabla 2.3: Condiciones de cristalización según la bibliografía.

Proteína	Condiciones de cristalización	Referencias
	Método usado para la cristalización: microdiálisis	
1 Tiorredoxina de <i>E. coli</i> (2TRX.pdb)	Agente precipitante: 2-metil-2,4-pentanodiol (MPD) 40% (v,v), 1 mM Ac ₂ Cu en NaAc 10 mM pH 3.8 a 4°C Concentración de proteína: 5mg/mL Límite de resolución: 1.68 Å	Holmgren y Söderberg, 1970; Katti <i>et al.</i> , 1990
2 Mutante K36E de tiorredoxina de <i>E. coli</i> (2TIR.pdb)	Método usado para la cristalización: difusión en fase vapor. Agente precipitante: PEG 6000 17%, 1 mM Ac ₂ Cu, NaAc 15 µM pH 4.8 a 4°C Límite de resolución: 2 Å	Nikkola <i>et al.</i> , 1993
	Método usado para la cristalización: microdiálisis	
3 Tiorredoxina de <i>E. coli</i>	Agente precipitante: NaAc 10 mM pH 3.7, MPD 48-60% (v/v), 1 mM Ac ₂ Cu Concentración de proteína: 10 mg/mL Límite de resolución: 2.8 Å	Söderberg <i>et al.</i> , 1974
	Método usado para la cristalización: difusión en fase vapor	
4 Mutante de tiorredoxina P40S de <i>E. coli</i>	Agente precipitante: NaAc 100 mM pH 3.8, Etanol 25%, 10 mM Ac ₂ Cu a 4°C Concentración de proteína: 5mg/mL Límite de resolución: 1.8 Å	Rudresh <i>et al.</i> , 2002, Holmgren y Söderberg, 1970
5 Motivo	Método usado para la cristalización:	Schultz <i>et al.</i> ,

CXXC de tioredoxina de <i>E. coli</i>	difusión en fase vapor. Agente precipitante: metil éter PEG 2000 (10%), 1 mM Ac ₂ Cu en tampón succinato 50 mM pH 4.2 a 4°C Concentración de proteína: 10 mg/mL Límite de resolución: 1.8 Å	1999
---	--	------

En la bibliografía se han encontrado distintas condiciones de cristalización bajo las cuales se ha cristalizado tioredoxina (Tabla 2.3). Hemos simulado todas ellas utilizando la técnica de contradifusión en capilares; los resultados obtenidos nos permitieron realizar sucesivas optimizaciones de las condiciones de cristalización hasta obtener cristales de calidad adecuada para difractar en un tiempo de equilibración mínimo de dos semanas (Tabla 2.4).

Tabla 2.4: Primera optimización de las condiciones de cristalización.

Condiciones de cristalización	
Técnica: Cristalización por contradifusión en capilares.	
Diámetro de capilar: 0.1 cm o 0.2 cm.	
Concentración de proteína: 30 mg/mL en NaAc 15 mM pH 3.5	
Tampón físico: Agarosa 0.5% en NaAc 15 mM pH 3.5	
Temperatura: 4°C	
1	Agente precipitante: MPD 60% (v,v) en NaAc 10 mM pH 3.8, Ac ₂ Cu 1 mM
2	Agente precipitante: PEG 6000 30% en NaAc 15 mM pH 4.8, Ac ₂ Cu 1 mM
3	Agente precipitante: PEG 6000 30% en NaAc 15 mM pH 5.2, Ac ₂ Cu 1 mM
4	Agente precipitante: Etanol 25% en NaAc 10 mM pH 3.8, Ac ₂ Cu 10 mM

De estas cuatro condiciones sólo se obtuvieron cristales de calidad elevada de las condiciones 1. Por lo tanto, una vez seleccionada la condición 1 como la mejor condición de cristalización para nuestras proteínas, se realizó una segunda optimización de la condición. Se llevaron a cabo dos experimentos paralelamente, las condiciones de cristalización de cada uno de ellos se recogen en la Tabla 2.5.

Tabla 2.5: Optimización de las condiciones de cristalización ampliando el rango de pH (3.6-7.8).

Condiciones de cristalización	
Técnica: Cristalización por contradifusión en capilares.	
Diámetro de capilar: 0.2 cm	
Concentración de proteína: 30 mg/mL en NaAc 15 mM pH 3.5	
Tampón físico: Agarosa 0.5% en NaAc 15 mM pH 3.5	
Temperatura: 4°C	
1	Agente precipitante: MPD 60% (v,v) en NaAc 15 mM, HEPES 15 mM pH 7, Ac ₂ Cu 1 mM
Técnica: Cristalización por contradifusión en capilares.	
Diámetro de capilar: 1mm	
Concentración de proteína: 30 mg/mL en NaAc 15 mM pH 3.5	
Tampón físico: Agarosa 0.5% en NaAc 15mM pH 3.5	
Temperatura: 4°C	
2 Agente precipitante:	
1	MPD 60% (v,v) en NaAc 15 mM pH 3.6, Ac ₂ Cu 1 mM
2	MPD 60% (v,v) en NaAc 15 mM pH 3.8, Ac ₂ Cu 1 mM
3	MPD 60% (v,v) en NaAc 15 mM pH 4.1, Ac ₂ Cu 1 mM
4	MPD 60% (v,v) en NaAc 15 mM pH 4.5, Ac ₂ Cu 1 mM
5	MPD 60% (v,v) en NaAc 15 mM pH 4.9, Ac ₂ Cu 1 mM
6	MPD 60% (v,v) en HEPES 15 mM pH 5.4, Ac ₂ Cu 1 mM
7	MPD 60% (v,v) en HEPES 15 mM pH 5.7, Ac ₂ Cu 1 mM
8	MPD 60% (v,v) en HEPES 15 mM pH 6.3, Ac ₂ Cu 1 mM
9	MPD 60% (v,v) en HEPES 15 mM pH 6.9, Ac ₂ Cu 1 mM
10	MPD 60% (v,v) en HEPES 15 mM pH 7.5, Ac ₂ Cu 1 mM
11	MPD 60% (v,v) en HEPES 15 mM pH 8.0, Ac ₂ Cu 1 mM

2.2.8.4 Protocolo de cristalización por contradifusión en capilares

Ya hemos visto como una forma de obtener un medio difusivo para realizar experiencias en contradifusión es el empleo de capilares y medios gelificados (sección 2.2.8.2). Su preparación la hemos llevado a cabo en capilares de borosilicato transparentes a los rayos X. Hemos utilizado la técnica de contradifusión con una configuración de tres cámaras (Figura 2.21), o mejor dicho, dos cámaras separadas por una tercera cámara amortiguadora (tampón físico). El diámetro de los capilares depende de la disponibilidad de proteína, de forma que podemos emplear capilares de 0.1 o 0.2 de diámetro interno.

El protocolo de preparación de una experiencia en contradifusión bajo la condición de cristalización 1.1 (Tabla 2.5) se explica a continuación:

1. La concentración de proteína debe ser aproximadamente de 30 mg/mL.
2. Se utilizan capilares de borosilicato de 0.2 mm de diámetro interno. Se cortan para que su longitud sea de 6 cm aproximadamente.
3. Se prepara la disolución de agarosa 0.5% (*p/v*) en NaAc 15 mM pH 3.5. Para facilitar su disolución, se calienta en el microondas durante algunos minutos sin permitir que llegue a hervir. Una vez disuelta, la solución debe incubarse a 45-50 °C en un baño o baño termostático durante toda la experiencia para evitar que gelifique.
4. **Cámara de proteína.** Se pipetea 6 µL de disolución de proteína y se mezclan, sobre un trozo de parafilm, con 1 µL de tampón de azarosa (0.5% *p/v*). Esta mezcla, se absorbe por capilaridad utilizando el capilar de 0.2 mm. Una vez conseguido, se sella el extremo del capilar (menor diámetro) con cera líquida.
5. **Cámara amortiguadora.** Por el otro extremo del capilar (mayor diámetro) se añaden 3 µL de tampón agarosa 0.5% (*p/v*) en NaAc 15 mM pH 3.5 (tampón físico), con cuidado de no introducir burbujas. El tampón de agarosa no interacciona químicamente en el proceso de cristalización, su función es disminuir la velocidad de mezcla de las disoluciones que separa (disolución de proteína y agente precipitante).
6. **Cámara salina.** Para terminar, se adicionan 30 µL de disolución de agente precipitante: MPD 60% en Hepes 15mM pH7, NaAc 15 mM, Ac₂Cu 1 mM. La concentración de la disolución de agente precipitante

es crítica, ya que establecerá la velocidad a la cual se mueve el frente de sobresaturación y la duración total del experimento.

7. El capilar se termina sellando con cera líquida, se deposita en una bandeja y se almacena a 4°C dónde permanecerá durante el tiempo necesario hasta que se observen cristales (entre 2 y 3 semanas en la mayoría de los casos).

El protocolo de preparación de una experiencia en contradifusión bajo las condiciones de cristalización 2.1-11 (Tabla 2.5) se explica a continuación:

Los puntos 1., 2. y 3. son exactamente iguales que en el caso anterior.

4. **Cámara de proteína.** Se introduce el capilar de 0.1 mm de diámetro interno directamente en la disolución de proteína, y observamos como la disolución de proteína asciende por capilaridad hasta la zona cónica del capilar. Una vez conseguido, se sella el extremo del capilar (menor diámetro) con cera líquida.
5. **Cámara amortiguadora.** Por el otro extremo del capilar (mayor diámetro) se añaden 3 μ L de agarosa 0.5% (*p/v*) en NaAc 15 mM pH 3.5 (tampón físico), con cuidado de no introducir burbujas.
6. **Cámara salina.** Para terminar, se adicionan 30 μ L de disolución de agente precipitante. Como para esta condición se realiza un barrido de pH, se montan 11 capilares, uno para cada condición salina (11 pHs diferentes, barriendo el intervalo desde pH 3.6 hasta pH 8.0).
7. El capilar se termina sellando con cera líquida, se deposita en una bandeja y se almacena a 4°C dónde permanecerá durante el tiempo necesario hasta que se observen cristales (entre 2 y 3 semanas en la mayoría de los casos).

Para poder identificar los capilares, se numerarán en el mismo orden en el que se han preparado. Y la bandeja se etiquetará con el nombre del experimento y la fecha en la que se ha realizado.

2.2.8.5 Seguimiento del crecimiento cristalino.

El seguimiento de la evolución espacial y temporal de las experiencias de crecimiento cristalino se ha realizado con un equipo de microscopía óptica de transmisión o una lupa binocular equipados con una cámara digital conectada a un

ordenador personal (PC). De esta forma, podemos conseguir imágenes de cualquier etapa del crecimiento cristalino. En la Figura 2.22 se muestran algunas imágenes de los cristales que se han difractado de las diferentes proteínas cristalizadas.

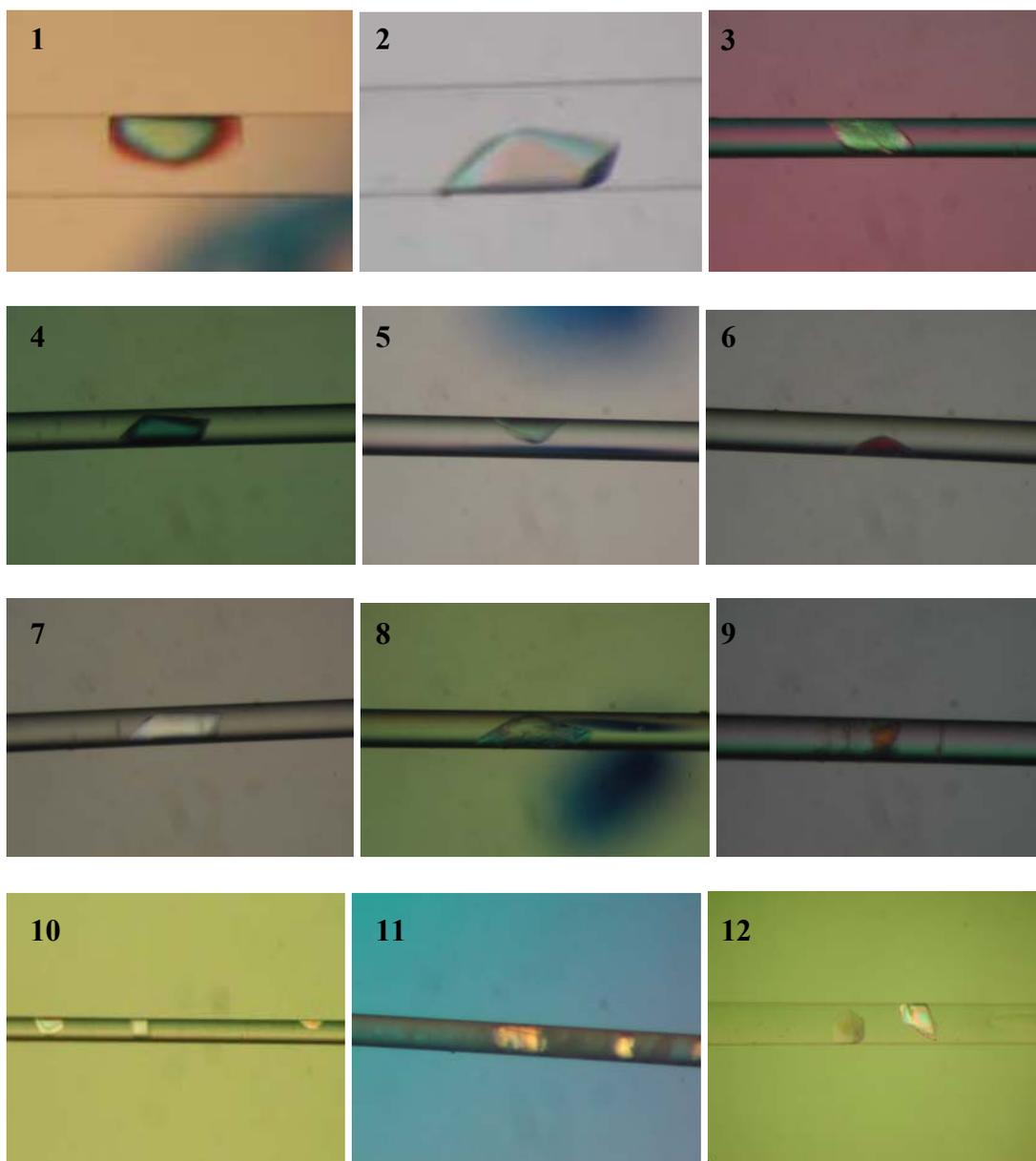


Figura 2.22: Algunas imágenes de cristales de algunas proteínas diseñadas utilizando métodos de contradifusión bajo las condiciones 1 y 2 (Tabla 2.5). Mutantes de tiorredoxina: D2E (1), D9E (2), D10E (3), D13E (4), D43E (5), E44D (6), D47E (7), E48D (8), E85D (9) y forma silvestre (WT) de tiorredoxina de *E.coli* (10, 11, 12).

2.2.9 ESTRUCTURA DE PROTEÍNAS POR DIFRACCIÓN DE RAYOS X

Para determinar la estructura de macromoléculas a nivel atómico de resolución disponemos de dos aproximaciones: la difracción (de rayos X y neutrones) de monocristales y la resonancia magnética nuclear (RMN). A diferencia de la RMN, la difracción no presenta ninguna limitación en cuanto al peso molecular de la proteína, pero necesita un cristal sobre el que realizar los estudios de difracción. Otras técnicas como la microscopía electrónica o la microscopía de fuerza atómica suministran una imagen a baja resolución del relieve general de la superficie de la proteína. Las técnicas bioinformáticas permiten modelar la estructura de una proteína a partir de las estructuras de sus homólogos secuenciales, pero ni la precisión ni la fidelidad de los modelos son comparables a lo obtenible mediante difracción de rayos X o RMN. En resumen, la ventaja de la difracción de rayos X es, que realizando un estudio estructural para proteínas se genera un modelo con la distribución espacial de los átomos con una precisión que depende de la calidad del cristal (el modelo nos proporciona el plegamiento y detalles de los cristales difractados, como lo son las distancias y ángulos de enlace, estado térmico vibracional de los átomos, empaquetamiento cristalino, etc). Toda esta información es muy útil para entender la función biológica de las proteínas.

Un estudio cristalográfico se divide en las siguientes etapas:

- Clonaje, expresión y purificación de la proteína.
- Cristalización.
- Difracción.
- Asignación de Fases.
- Refinamiento de la estructura.
- Validación de la estructura.

2.2.9.1 Metodología de un experimento de difracción. Instrumentación utilizada

El primer paso de una experiencia de difracción es preparar el cristal que se desea difractar. En nuestro caso, como hemos utilizado capilares, ésto es relativamente sencillo. Una vez elegido el cristal, se corta el capilar que lo contiene (1 cm aproximadamente) y se sella por ambos extremos con un

pegamento de secado rápido y se monta sobre una cabeza goniométrica que posee varios grados de libertad (Figura 2.23). Seguidamente el cristal se centra con una cámara digital que se encuentra alineada en el plano del haz de rayos X. Se utiliza el programa Proteum 2Server.

Antes de recoger los datos se hace un test de difracción en el que se expone el cristal durante 15-20 segundos y se gira con un grado de libertad 0.5°. La imagen recogida se indexa y se define la celdilla unidad y el grupo de Laue. El cálculo de la estrategia del experimento de difracción se hace para recoger el máximo de reflexiones posibles en el menor tiempo de exposición. Una vez calculada la estrategia, se importa y se comienza la recogida de reflexiones.

En el experimento de difracción, el haz incidente produce calor y radicales libres en las moléculas de proteínas y de solvente, provocando con el tiempo daños irreversibles en el cristal. Luego, es necesario subenfriar el cristal a 100 K con un flujo laminar de nitrógeno gas (Krioflex). Por lo que el cristal debe encontrarse saturado previamente en una solución crioprotectora (hemos utilizado MPD que actúa a la vez como agente precipitante) que reemplace a las moléculas de agua no estructural, evitando así la formación de hielo y la consiguiente ruptura del cristal.

La mayoría de los cristales obtenidos en este trabajo de investigación se han difractado utilizando un generador de rayos-X de ánodo rotatorio Microstar de BRUKER Microfocus (Montel Optics) junto con un detector CCD (*Charged Coupled Device*) Proteum X8. 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 ($\lambda=1.54 \text{ \AA}$). El generador de alimentación del tubo opera típicamente a 45 kV y 60 mA (Figura 2.23).

En ocasiones los datos se han adquirido empleando radiación sincrotrón con longitudes de onda por debajo de 1 \AA generadas en el acelerador de partículas ESRF (*European Synchrotron Radiation Facilities* en Grenoble, línea BM16).

Se han recogido un número imágenes para obtener un 99% de completitud en el rango de resolución seleccionado, mientras se hacía girar al cristal, al menos con un grado de libertad, en intervalos del orden de 0.5°, con un tiempo de exposición y distancia cristal-detector determinados del análisis preliminar de la calidad de difracción. Del análisis de estas imágenes (posición e intensidad de los

puntos) se puede derivar, la geometría de la celdilla elemental cristalina y su simetría (grupo espacial) que junto a la intensidad asociada a cada punto recíproco, constituye toda la información que se puede obtener directamente de un experimento de difracción en condiciones normales. El integrado y escalado de la intensidad de cada reflexión se ha hecho con el programa *SAINTE* y la corrección de absorción con el programa *SADABS* (ambos pertenecientes al paquete de programas *PROTEUM*). La reducción de los datos se realizó con *XPREP*.

Cuando los cristales se difractaron en el *ESRF*, los datos de difracción se indexaron e integraron utilizando el programa *DENZO*, se escalaron usando el programa *SCALEPACK* (programas pertenecientes al paquete de programas *HKL2000*).

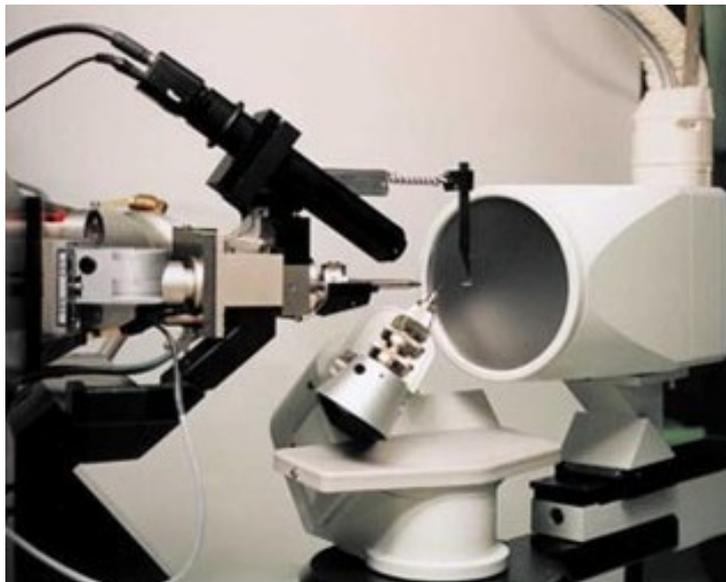


Figura 2.23: Foto del difractor de rayos X de ánodo rotatorio Microstar de *BRUKER* Microfocus (Montel Optics) junto con un detector *CCD (Charged Coupled Device) Proteum X8*

2.2.9.2 El problema de las fases

En un experimento de difracción, los rayos-X son dispersados por los electrones de los átomos existentes en un cristal. La amplitud intrínseca del rayo-X dispersado por cada tipo de átomo, es dependiente de la densidad electrónica de

cada átomo. Cuanto mayor sea la densidad electrónica en cualquier punto del espacio, mayor será la amplitud del haz dispersado. Si se tiene en cuenta esto en la ecuación de propagación de una onda, hay que incluir un factor de dispersión atómico f para cada átomo. En cristalografía, un factor de estructura F (ecuación 2.55) se define como la amplitud de cada haz dispersado en valores específicos de los índices de Miller ($h k l$) e incluye todos los átomos en la celdilla unidad:

$$F(hkl) = \sum_{j=1}^N f_j e^{2\pi i(hx_j + ky_j + lz_j)} \quad (2.55)$$

donde N es el número de átomos en la celdilla unidad y

$$\mathbf{S} \cdot \mathbf{r}_j = (ha^*, kb^*, lc^*)(x_j a, y_j b, z_j c) = hx_j + ky_j + lz_j \quad (2.56)$$

Esta expresión del factor de estructura (ecuación 2.55) considera los elementos en los que se produce la difracción como puntos fijos y discretos en el espacio (definidos por el vector \mathbf{r}_j). Se compone de dos partes: el factor de dispersión atómico f_j que define el máximo de amplitud del rayo-X dispersado si ese átomo se encuentra en el origen de coordenadas de la celdilla unidad ($\phi=0$) y sólo depende del tipo de átomo que actúa como nudo de dispersión; es un número real. El factor exponencial contiene la información relacionada a la fase relativa de la onda dispersada; es un número complejo.

Los electrones no son cargas puntuales localizadas en un determinado punto del espacio, es más adecuado considerarlos como una densidad electrónica continua $\rho(\mathbf{r})$. Luego, el objetivo de un estudio estructural es la determinación de la *densidad electrónica*, $\rho(x,y,z)=\rho(\mathbf{r})$ para cada punto de la celdilla unidad cristalina (\mathbf{r}). El factor de estructura puede reescribirse en términos de densidad electrónica, para cualquier punto del espacio en el que consideremos un elemento diferencial de volumen: $\delta x \delta y \delta z = \delta V$, de la siguiente forma:

$$F(\mathbf{S}) = \int \rho(\mathbf{r}) e^{2\pi i \mathbf{S} \cdot \mathbf{r}} \delta V \quad (2.57)$$

Si consideramos que en el espacio de un cristal existe una probabilidad continua de encontrar electrones, entonces el factor de estructura se puede escribir como una integral extendida a todo el volumen de la muestra:

$$F(\mathbf{S}) = \int V \rho(\mathbf{r}) e^{2\pi i \mathbf{S} \cdot \mathbf{r}} dV \quad (2.58)$$

Ésta ecuación es una serie de Fourier y es similar a la dada para átomos discretos (ecuación 2.55), ya que en las posiciones atómicas es donde existe la mayor probabilidad de encontrar los electrones. El factor exponencial ($2\pi i \mathbf{S} \cdot \mathbf{r}$) depende de la fase de los elementos dispersores y de su posición en el espacio. Las ecuaciones 2.55 y 2.58 nos dicen, que si conocemos el tipo y posición de cada elemento dispersor relativo al origen del cristal (es decir, si conocemos la estructura de la molécula), en el espacio, podremos calcular los factores de estructura $F(\mathbf{S})$. Y por tanto, si conocemos $F(\mathbf{S})$ conoceremos las distribuciones de densidades electrónicas en el espacio real.

La transformación del espacio recíproco (patrón de difracción) al espacio cristalino (distribución espacial de la densidad electrónica) implica la resolución la *transformada de Fourier*:

$$\rho(\mathbf{r}) = \frac{1}{V} \int dV^* e^{-2\pi i \mathbf{S} \cdot \mathbf{r}} F(\mathbf{S}) \quad (2.59)$$

donde V^* es el elemento de volumen en el espacio recíproco y V es el volumen en la celdilla unidad. La transformada de Fourier dada por la ecuación 2.59, nos proporciona una función que describe la densidad electrónica en cualquier punto del espacio real en términos del vector de dispersión en el espacio recíproco.

Como la difracción ocurre siguiendo un patrón regular y, de acuerdo a las condiciones de von Laue, en puntos discretos, las densidades electrónicas se pueden calcular de la suma de los factores de estructura $F(\mathbf{S})$ para los índices de Miller ($h k l$). Ésto es una serie de Fourier para reflexiones discretas (ecuación 2.60), en oposición a la transformada de Fourier continua (ecuación 2.59):

$$\rho(\mathbf{r}) = \frac{1}{NV} \sum_{h=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} \sum_{l=-\infty}^{\infty} F(hkl) e^{-2\pi i \mathbf{S} \cdot \mathbf{r}} \quad (2.60)$$

Cuando las densidades electrónicas calculadas con la ecuación 2.60 se visualizan para las posiciones (x, y z) de la celdilla unidad cristalina, se genera un *mapa de densidad electrónica* para la celdilla unidad del cristal (Figura 2.24). Este mapa muestra las posiciones y los tipos de átomos (número de electrones para un volumen dado) en la unidad asimétrica del cristal.

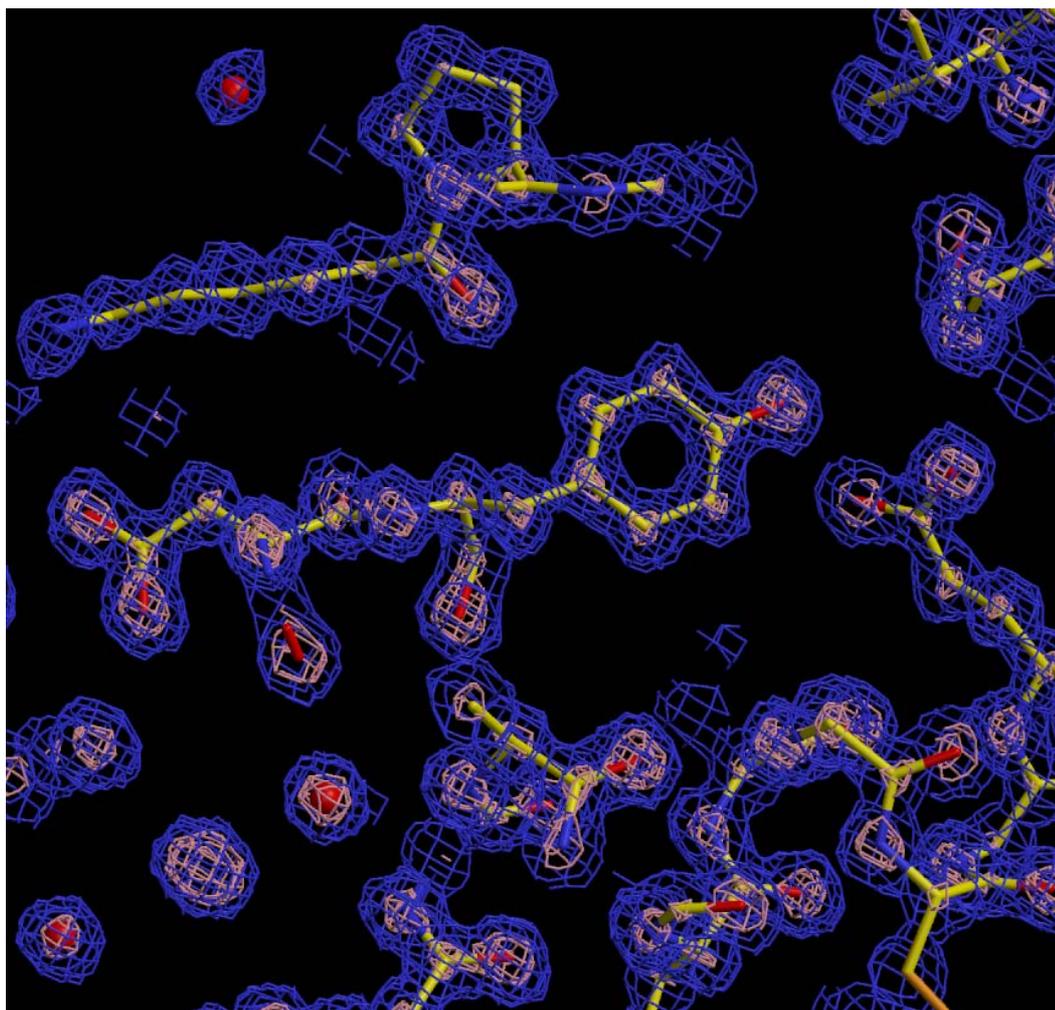


Figura 2.24: Mapa de densidad electrónica. En contorno azul, se representa $2F_0 - F_C$; en contorno violeta se representa $F_0 - F_C$; en rojo, zonas de densidad electrónica negativa. El esqueleto del modelo se ajusta frente a $2F_0 - F_C$.

No obstante, el proceso aparentemente sencillo está complicado por el llamado “problema de la fase”. En el experimento de difracción, la energía total recogida es proporcional a la intensidad ($I(\mathbf{S})$) y por lo tanto, a su amplitud ($|F(\mathbf{S})|^2$). Luego, el problema es que podemos calcular la amplitud (magnitud) del factor de estructura, pero no sus fases (dirección). En términos de notación compleja, se puede expresar mediante la siguiente expresión:

$$I(\mathbf{S}) = |F(hkl)|^2 = |F(\mathbf{S})|^2 = F(\mathbf{S})F^*(\mathbf{S}) \quad (2.61)$$

donde $F^*(\mathbf{S})$ es el conjugado complejo de $F(\mathbf{S})$, y se define como:

$$F^*(S) = \sum_{j=1}^N f_j e^{-2\pi i \mathbf{S} \cdot \mathbf{r}_j} = F_{re} - iF_{im} \quad (2.62)$$

En resumen, a partir de las medidas del espectro de difracción se obtienen las intensidades $I(hkl)$. Estas intensidades nos proporcionan fácilmente las amplitudes $|F(hkl)|$ de los factores de estructura, pero, y aquí viene el problema, no sus fases $\phi(hkl)$. Es decir, poseemos la parte de información que nos proporciona $F(\mathbf{S})$, que está relacionada con el número y tipo de átomos dispersores respecto al origen de coordenadas de la celdilla unidad; sin embargo, no podemos medir la información correspondiente a las fases, relacionada con la posición absoluta de los átomos en la celdilla unidad y por lo tanto no podemos calcular el mapa de densidad electrónica.

El problema de las fases puede resolverse mediante tres métodos:

a) Introduciendo átomos altamente dispersores, o método de Reemplazamiento Isomorfo Múltiple (*MIR, Multiple Isomorphous Replacement*).

b) Introduciendo átomos dispersores anómalos, o método de *Dispersión Anómala Múltiple (MAD, Multi-wavelength Anomalous Diffraction)*.

c) Mediante el *Reemplazamiento Molecular (MR, Molecular Replacement)*, haciendo uso de un modelo estructural de una proteína homóloga, previamente determinada.

Con cualquiera de estos tres métodos podemos calcular unas fases iniciales, $\Phi_c(hkl)$, que junto con las amplitudes experimentales, $|F(hkl)|$, nos permitirán calcular una función de densidad electrónica (ecuación 2.59) aproximada, y sobre la cual podremos construir nuestro modelo estructural. En el siguiente apartado, se describe brevemente el método de reemplazamiento molecular utilizado para resolver el problema de las fases, pues ha sido el utilizado para resolver la estructura de las proteínas estudiadas en este trabajo de investigación (proceso resumido en Figura 2.25).

2.2.9.3 Método de resolución estructural: Reemplazamiento Molecular

Cuando se dispone de un modelo estructural de una proteína con una secuencia de aminoácidos homóloga, el problema de la fase se puede resolver mediante la técnica del reemplazo molecular (*RM*). La estructura de esta proteína homóloga se considera como si fuera la proteína que va a determinarse y sirve como un primer modelo que posteriormente será refinado. Este procedimiento está basado en la observación de que proteínas homólogas en su secuencia peptídica muestran un plegamiento muy similar. El problema, en este caso, consiste en transferir la estructura molecular de la proteína conocida, desde su propio empaquetamiento cristalino hasta el cristal de la proteína con estructura desconocida. El posicionamiento de la molécula conocida en la celdilla unidad de la proteína desconocida requiere determinar su correcta orientación y su posición precisa. Ambas operaciones, rotación y traslación, se calculan mediante las funciones denominadas de rotación y de traslación (**R** y **T** en Figura 2.26, respectivamente).

La calidad de las soluciones de estas funciones se expresa mediante los coeficientes de correlación entre las funciones de Patterson experimental y la calculada con la proteína conocida. Un coeficiente de correlación alto entre dichas funciones equivale a un buen acuerdo entre el espectro de difracción experimental y el calculado con la proteína conocida. Una vez orientada y trasladada convenientemente la molécula conocida, se calcula un mapa de densidad electrónica usando los factores de estructura experimentales.

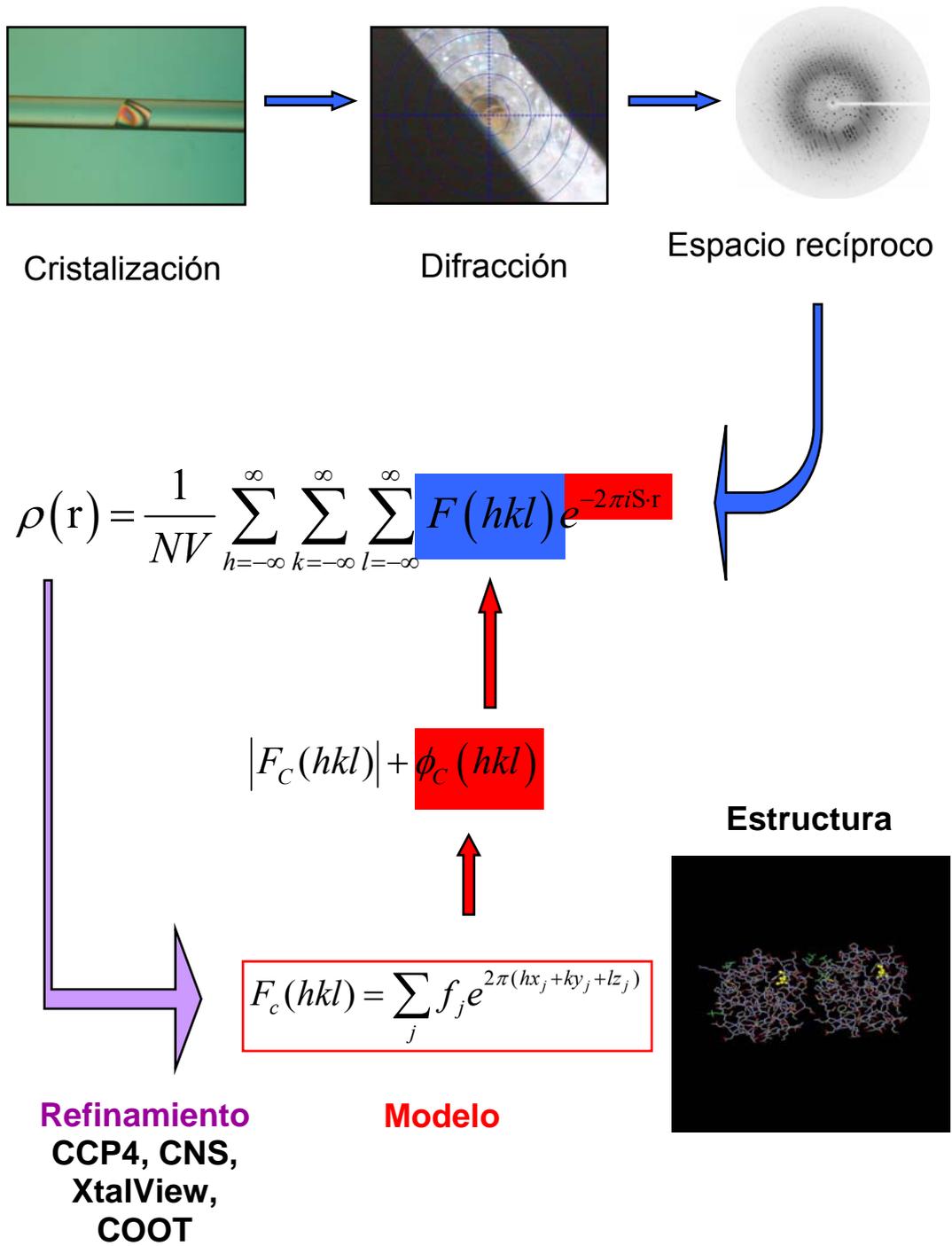


Figura 2.25: Ilustración del proceso de determinación de la estructura tridimensional de una macromolécula por difracción de rayos-X.

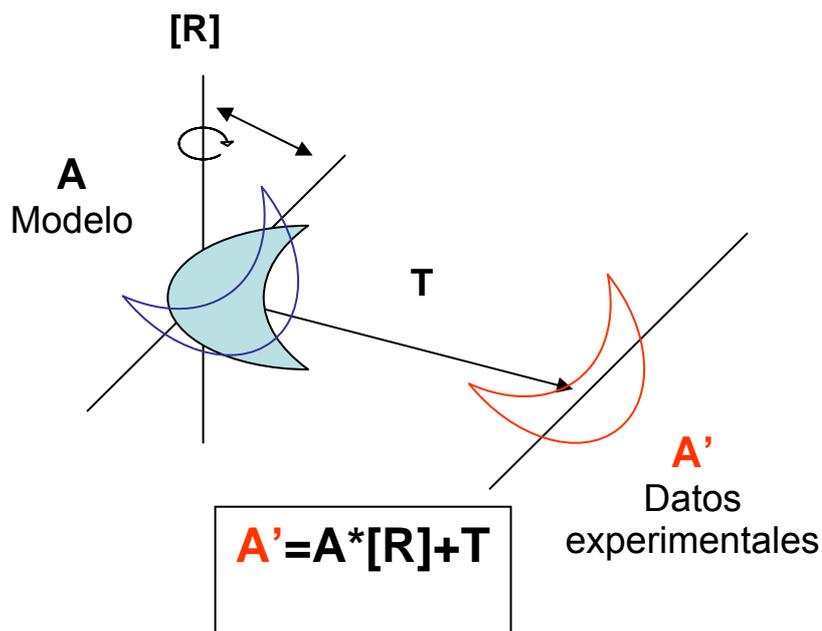


Figura 2.26: De forma gráfica se puede explicar el reemplazamiento molecular como las operaciones que permiten superponer nuestro modelo, molécula **A**, con nuestros datos experimentales, **A'**. Las dos operaciones necesarias son las de rotación, representada por la matriz de rotación **[R]** y traslación, representada por el vector **T**.

Una vez solucionado el problema de las fases, se puede calcular de forma directa la función de densidad electrónica (ecuación 2.59) que está definida en cada punto de la celdilla unidad. La representación gráfica de los valores de esta función da lugar a lo que se denomina *mapa de densidad electrónica*. Si la distribución de fases estimadas es correcta para el conjunto de módulos de los factores de estructura experimentales, el mapa mostrará zonas con densidad electrónica positiva sobre las cuales deberá acoplarse el esqueleto polipeptídico (Figura 2.24). La calidad de los mapas de densidad electrónica depende de la exactitud de los datos de difracción y en especial de la resolución de los mismos. Para obtener una resolución alta en los mapas de densidad electrónica es muy importante haber realizado los experimentos de difracción en condiciones de baja temperatura (aunque no siempre es necesario este requisito), lo que permite un mayor tiempo de vida del cristal bajo el haz de rayos X y estabiliza las regiones

con una gran flexibilidad conformacional. Por otro lado, cuanto mayor sea la resolución de los datos de difracción, mayor detalle observaremos en el mapa y se podrán derivar, con mayor precisión, parámetros tales como distancias o ángulos interatómicos.

Con la información que suministra el mapa de densidad electrónica, se corrige manualmente la cadena polipeptídica del modelo estructural utilizado y es posible ampliarlo, es decir, se pueden construir otras zonas de la cadena polipeptídica no incluidas en nuestro modelo previo. Es en este momento cuando se utilizan los algoritmos de cálculo iterativo, que agilizan el proceso de convergencia entre el modelo estructural y las observaciones experimentales (espectro de difracción). En esta etapa, se ajustan los parámetros estructurales (x, y, z, B) para optimizar el acuerdo entre las observaciones y la predicción. El factor térmico, B, representa la amplitud cuadrática media de vibración térmica del átomo; cuanto menor sea este parámetro, más localizado está el átomo. Para el caso de proteínas, hay que introducir constricciones en la estereoquímica del modelo.

Entre los diversos tipos de algoritmos que se utilizan en el refinamiento, encontramos un *refinamiento de cuerpo rígido*, en el cual se mantiene la estructura de la proteína como un todo que se modifica sólo en términos de traslaciones y giros; un *annealing simulado*, basado en las técnicas de Monte Carlo, equivale al proceso físico de calentar la proteína para enfriarla lentamente y obtener la estructura cristalina.

La bondad del refinamiento se evalúa con el denominado *factor R*, que se define como:

$$R = \frac{\sum |F(hkl)| - |F_{calc}|}{\sum |F(hkl)|} \quad (2.63)$$

Teniendo en cuenta que el refinamiento es un proceso iterativo de ajuste manual y minimización de R empleando los programas de refinamiento, el factor R mide las diferencias entre el espectro observado experimentalmente ($|F(hkl)|$) y el calculado con el modelo estructural obtenido ($|F_{calc}|$), por lo que cuanto menor sea este factor, mejor será el ajuste del modelo al experimento; un valor de

$R > 0.59$ nos indica que el modelo utilizado es erróneo. Además, durante el proceso de refinamiento también hay que controlar el porcentaje de residuos que se encuentran en zonas no permitidas del diagrama de ramachandran.

El proceso de refinamiento se llevó a cabo con los programas *CNS* (Brünger *et al.*, 1998), *REFMAC* (perteneciente al paquete de programas CCP4 (Collaborative Computacional Project, Number 4, <http://www.ccp4.ac.uk/main.html>, Murshudov *et al.*, 1997) y el programa gráfico XtalView (McRee, 1993).

2.2.9.4 Validación final de la estructura

Al final del proceso de refinamiento, toda la información sobre el modelo geométrico de la proteína queda reflejada en un fichero que se deposita en la base de datos: *Protein Data Bank* (<http://www.rcsb.org/pdb/Welcome.do>, Bernstein *et al.*, 1977) en donde se revisa, valida y recibe un código que lo identifica. Este fichero contiene toda la información concerniente a las condiciones de cristalización de la proteína, datos de difracción, datos de refinamiento (coordenadas atómicas, factores térmicos, factores de ocupación, grupo espacial, dimensiones de la celdilla unidad, etc). Existen páginas donde se puede comprobar la validez de la estructura que se piensa depositar en el *Protein Data Bank*: <http://molprobitry.biochem.duke.edu/>.

BIBLIOGRAFIA

- Acevedo, O., M. Guzman-Casado, M.M. Garcia-Mira, B. Ibarra-Molero and J.M. Sanchez-Ruiz (2002). *Anal Biochem* **306**(1): 158-61.
- Beasty, A.M., M.R. Hurle, J.T. Manz, T. Stackhouse, J.J. Onuffer and C.R. Matthews (1986). *Biochemistry* **25**: 2965-2974.
- Becktel, W.J. and J.A. Shellman (1987). *Biopolimers* **26**: 1859-1877.
- Bernstein, F.C., T.F. Koetzle, G.J.B. Williams, E.F.J. Meyer, M.D. Brice, J.R. Rodgers, O. Kennard, T. Shimanouchi and M. Tasumi (1977). *J. Mol. Biol.* **112**: 535-542.
- Brandts, J.F. (1964). *J. Am. Chem. Soc.* **86**: 4291-4301.
- Brünger, A.T., P.D. Adams, G.M. Clore, P. Gros, R.W. Grosse-Kunstleve, J.-S. Jiang, J. Kuszewski, M. Nilges, N.S. Pannu, R.J. Read, L.M. Rice, T. Simonson and G.L. Warren (1998). *Acta Cryst.* **D54**(905-921).
- Carotenuto, L., C. Piccolo, D. Castagnolo, M. Lappa, A. Tortora and J.M. García-Ruiz (2002). *Acta Cryst.* **D 58**: 1628-1632.
- Chen, B.L. and J.A. Schellman (1989b). *Biochemistry* **28**: 685-691.
- Chen, B.L., W.A. Bease and J.A. Schellman (1989a). *Biochemistry* **28**: 691-699.
- Chen, B.L., W.A. Bease, H. Nicholson and J.A. Schellman (1992). *Biochemistry* **31**: 1464-1476.
- Chowdry and Cole (1989). *Trends Biotechnol* **7**: 11-18.
- Cooper, A. and C.M. Johnson (1994). *Methods in Molecular Biology* **22**(a): 109-124.
- Dayhoff, M., R.M. Schwartz and B.C. Orcutt (1978). Atlas of Protein Sequence and Structure. M.e. Dayhoff. Silver Spring, MD, National Biomedical Research Foundation. **vol. 5 (3)**: 345-352.
- Demarest, S.J., R. Fairman and D.P. Raleigh (1998). *J Mol Biol* **283**: 279-291.
- Dokholyan, N.V. and E.I. Shakhnovich (2001). *J Mol Biol* **312**(1): 289-307.
- Ducruix, A. and R. Giegé (1999). Oxford, Oxford University Press.
- Evans, P.A., K.D. Topping, D.N. Woolfson and C.M. Dobson (1991). *Proteins* **9**: 248-266.

- Feher, G. and Z. Kam (1985). *Methods in Enzymol.* **114**: 77.
- Filimonov, V.V. and V.V. Rogov (1996). *J. Mol. Biol.* **255**: 767-777.
- Freire, E. (1995). Protein Stability and Folding. B.A. Shirley. Totowa, New Jersey, Humana Press: 191-218.
- Freire, E. and R.L. Biltonen (1978). *Biopolimers* **17**: 463-479.
- Freire, E., W.W. van Osdol, O.L. Mayorga and J.M. Sanchez-Ruiz (1990). *Annu Rev Biophys Biophys Chem* **19**: 159-88.
- García-Ruiz, J.M. (1991). *Key Eng. Mater.* **88**: 87-106.
- García-Ruiz, J.M. (2003). *Methods in Enzymology* **368**: 130-154.
- Garvey, E.P., J. Swan and C.R. Matthews (1989). *Proteins* **6**: 259-266.
- Gavira, J.A., D. Toh, J. Lopez-Jaramillo, J.M. García-Ruiz and J.D. Ng (2002). *Acta Cryst.* **D58**(1147-1154).
- George, D.G., L.T. Hunt and W.C. Barker (1988). New York, A.R. Liss.
- Georgescu, R.E., M.M. Garcia-Mira, M.L. Tasayco and J.M. Sanchez-Ruiz (2001). *Eur J Biochem* **268**(5): 1477-85.
- Georgescu, R.E., J.-H. Li, M.E. Goldberg, M.L. Tasayco and A.F. Chaffotte (1998). *Biochemistry* **1998**: 10286-10297.
- Gill, S.C., B. Richey, B. Bishop and J. Wyman (1985). *Biophys Chem* **21**: 1-14.
- Godoy-Ruiz, R., R. Perez-Jimenez, B. Ibarra-Molero and J.M. Sanchez-Ruiz (2004). *J Mol Biol* **336**(2): 313-8.
- Godoy-Ruiz, R., R. Perez-Jimenez, B. Ibarra-Molero and J.M. Sanchez-Ruiz (2005). *Biophys J* **89**(5): 3320-31.
- Gomez, J., V.J. Hilser, D. Xie and E. Freire (1995). *Proteins* **22**: 404-412.
- Guzman-Casado, M., A. Parody-Morreale, S. Robic, S. Marqusee and J.M. Sanchez-Ruiz (2003). *J Mol Biol* **329**(4): 731-43.
- Hagel, P., J.J.T. Gerding, W. Fieggen and H. Bloemendal (1971). *Biochem Biophys. Acta* **243**: 366.
- Henikoff, S. and T.G. Henikoff (1992). *Proc Natl Acad Sci U S A* **89**: 10915-10919.
- Holmgren, A. and B.O. Söderberg (1970). *J. Mol. Biol.* **54**: 387-390.

- Hurle, M.R., G.A. Michelotti, M.M. Crisanti and C.R. Matthews (1987). *Proteins: Struct., Funct., Genet.* **2**(54-63).
- Ibarra-Molero, B. and J.M. Sanchez-Ruiz (1996). *Biochemistry* **35**(47): 14689-702.
- Ibarra-Molero, B. and J.M. Sanchez-Ruiz (1997). *Biochemistry* **36**(31): 9616-24.
- Ibarra-Molero, B. and J.M. Sanchez-Ruiz (2006). Protein folding and misfolding. M. V, Royal Society of Chemistry: In press.
- Ibarra-Molero, B. and J.M. Sanchez-Ruiz (2006). Advanced techniques in biophysics. J.L.a.A. Arrondo, A., Elsevier: 27-48.
- Ibarra-Molero, B., V.V. Loladze, G.I. Makhatadze and J.M. Sanchez-Ruiz (1999). *Biochemistry* **38**(25): 8138-49.
- Kam, Z., H.B. Shore and G. Feher (1978). *J. Mol. Biol.* **123**: 539.
- Katti, S.K., D.M. LeMaster and H. Eklund (1990). *J. Mol. Biol.* **212**: 167-184.
- Kelley, R.F. and E. Stellwagen (1984). *Biochemistry* **23**: 5095-5102.
- Kholodenko, V. and E. Freire (1999). *Anal Biochem* **270**(2): 336-8.
- Kliger, D.S., J.W. Lewis and C.E. Randall (1990). New York, Academy Press.
- Langstemo, K., J. Fuchs and C. Woodward (1989). *Biochemistry* **28**: 3211-3220.
- Lázló, P. (1999). Budapest (Hungary), Hungarian Academy of Sciences.
- Lowry, T.M. (1935). London, Green.
- Lumb, K.J. and P.S. Kim (1994). *J Mol Biol* **236**: 412-420.
- Makhatadze, G.I. and P.L. Privalov (1990). *J Mol Biol* **213**(2): 375-84.
- Manavalan, P. and W.C.J. Johnson (1983). *Nature* **305**: 831-832.
- Matthews, C.R. and M.R. Hurle (1987). *BioEssays* **6**: 254-257.
- McPherson, A. (1982). New York.
- McPherson, A. (1999). Cold Spring Harbor, New York, Cold Spring Harbor Laboratory Press.
- McRee, D.E. (1993). New York.
- Michel, J. and E.W. Thulstrup (1986). New York.

- Mücke, M. and F.X. Schmid (1994). *Biochemistry* **33**: 12930-12935.
- Murshudov, G.N., A.A. Vagin and E.J. Dodson (1997). *Acta Cryst.* **D53**(240-255).
- Neri, D., M. Billeter, G. Wider and K. Wüthrich (1992). *Science* **257**: 1559-1563.
- Ng, J.D., J.A. Gavira and J.M. García-Ruiz (2003). *Jour. of Struct. Biolog.* **142**: 218-231.
- Nikkola, M., F.K. Gleason, J.A. Fuchs and H. Eklund (1993). *Biochemistry* **32**: 5093-5098.
- Perry, K.M., J.J. Onuffer, M.S. Gittelman, L. Barmat and C.R. Matthews (1989). *Biochemistry* **34**: 8621-8630.
- Plotnikov, V.V., J.M. Brandts, L.N. Lin and J.F. Brandts (1997). *Analytical Biochemistry* **250**: 237-244.
- Privalov, P.L. (1979). *Adv Prote Chem* **4**: 41-50.
- Privalov, P.L. (1980). *Pure Appl Chem* **52**: 479-497.
- Privalov, P.L. (1982). *Adv Prote Chem* **35**: 1-104.
- Privalov, P.L. (1986).
- Privalov, P.L. (1989). *Annu Rev Biophys Biophys Chem* **18**: 47-69.
- Privalov, P.L. and G.I. Makhatadze (1990). *J Mol Biol* **213**(2): 385-91.
- Reutimann, H., B. Straub, P.L. Luisi and A. Holmgren (1981). *J Biol Chem* **256**: 6796-6803.
- Robic, S., M. Guzman-Casado, J.M. Sanchez-Ruiz and S. Marqusee (2003). *Proc Natl Acad Sci U S A* **100**(20): 11345-9.
- Rudresh, J., V. Dani, A. Mitra, S. Srivastava, S.P. Sarma, R. Varadarajan and S. Ramakumar (2002). *Protein Eng* **15**: 627.
- Sanchez-Ruiz, J.M. (1995). New York, Plenum Press.
- Sanchez-Ruiz, J.M. and P.L. Mateo (1987). *Revis Biol Celular* **11**: 15-45.
- Schellman, J.A. (1987). *Annu Rev Biophys Chem* **16**: 115-137.
- Schmid, F.X. (1992). Protein Folding. T.E. Creighton. New York, Freeman.

Schultz, L.W., P.T. Chivers and R.T. Raines (1999). *Acta Cryst.* **D55**: 1533-1538.

Smith, T.F. and M.S. Waterman (1981). *J Mol Biol* **147**: 195-197.

Söderberg, B.O., A. Holmgren and C.I. Brändén (1974). *J Mol Biol* **90**(143-52).

Spolar, R.S. and M.T. Record (1994). *Science* **263**: 777-784.

Stark, G.R. (1965). *Biochemistry*.

Thorolfsson, M., B. Ibarra-Molero, P. Fojan, S.B. Petersen, J.M. Sanchez-Ruiz and A. Martinez (2002). *Biochemistry* **41**(24): 7573-85.

Warren, J.E. and R.G. Gordon (1966). *J Phys Chem* **67**: 1524.

Weber, P.C. (1991). *Adv. Prot. Chem.* **41**.

Wilson, J., R.F. Kelley, W. Shalongo, D. Lowery and E. Stellwagen (1986). *Biochemistry* **25**(23): 7560-6.

Woody, R. (1996). Circular Dichroism and the Conformational Analysis of Biomolecules. New York, Fashman Plenum: 25-67.

Wu, J., J.T. Yang and C.S.C. Wu (1992). *Anal Biochem* **200**: 359-364.

**3. Estudio de la relación entre estabilidad
de proteínas, evolución y estructura
basado en un análisis mutacional de
ácidos carboxílicos**

3 ESTUDIO DE LA RELACIÓN ENTRE ESTABILIDAD DE PROTEÍNAS, EVOLUCIÓN Y ESTRUCTURA BASADO EN UN ANÁLISIS MUTACIONAL DE ÁCIDOS CARBOXÍLICOS

La estabilidad termodinámica de las proteínas es marginal (Schellman, 1987; Dill, 1990; Makhatadze y Privalov, 1995; Sanchez-Ruiz, 1995), según indican los bajos valores experimentales del cambio de energía libre de desnaturalización (del orden de algunas decenas de kJ/mol) a temperatura fisiológica. Esta baja estabilidad puede haberse seleccionado durante la evolución por diferentes razones beneficiosas, por ejemplo, una baja estabilidad puede asociarse con una alta flexibilidad de las proteínas (como la que se requiere para la función), puede haberse seleccionado para facilitar la degradación de las proteínas o para ayudar a que la proteína no quede atrapada en estados plegados incorrectos durante el plegamiento (Sanchez-Ruiz, 1995). Además, las interacciones desestabilizantes, como por ejemplo las que implican el enterramiento de grupos polares (Lumb y Kim, 1995; Knappenberger, 2002) pueden ser necesarias para conseguir una única estructura. Por otro lado, el hecho de que, probablemente, muchas mutaciones (mutaciones neutras) puedan ser ligeramente desestabilizantes, puede estar contribuyendo al hecho de que la estabilidad de las proteínas sea marginal. Estas mutaciones neutras, pero desestabilizantes pueden haberse acumulado durante algunos estadios tempranos de la evolución hasta el punto de llegar a una estabilidad marginal, de manera que una desestabilización mayor hubiera comprometido el estado nativo de las proteínas.

En cualquier caso, una vez que se ha alcanzado la estabilidad marginal, pueden ocurrir mutaciones en posiciones no restringidas, con probabilidades relacionadas con sus efectos sobre la estabilidad (es decir, la probabilidad de que ocurra una mutación desestabilizante es mucho menor). En los últimos años se ha sugerido, mayormente en trabajos teóricos (Dokholyan y Shakhnovich, 2001; Shortle, 2003), que las secuencias de proteínas se pueden considerar como sistemas de mecánica estadística que representan estados discretos de energía. De tal forma, las secuencias, en tiempo evolutivo, se visitan de acuerdo a su energía (evitando las proteínas con energías altas en el estado nativo), de manera análoga a la distribución de Boltzmann de energías para un sistema en equilibrio. Nos

referimos a esta hipótesis como la “hipótesis de pseudo-equilibrio” (por carecer de un nombre mejor).

El aceptable nivel de éxito del llamado “concepto de consenso” para la estabilización de proteínas apoya experimentalmente la hipótesis de pseudo-equilibrio (Steipe *et al.*, 1994; Lehmann *et al.*, 2000) y emplea un análisis estadístico de los alineamientos de secuencias para predecir mutaciones estabilizantes. En este capítulo, realizamos una evaluación cuantitativa y sistemática de la hipótesis de pseudo-equilibrio haciendo un análisis de las mutaciones de residuos glutamato y aspartato existentes en tiorredoxina de *E. coli*.

3.1 ESTUDIO DEL EFECTO DE LAS MUTACIONES E→D Y D→E SOBRE LA ESTABILIDAD DE TIORREDOXINA

Los residuos glutamato (E) y aspartato (D) son física-químicamente similares y pueden sustituirse fácilmente uno por el otro en el curso de la evolución de proteínas, como muestran sus correspondientes coeficientes en las matrices de PAM (George *et al.*, 1990) y BLOSUM (Henikoff y Henikoff, 1992). La observación de dichas matrices nos indica que los residuos E y D son muy conservativos (véase sección 2.2.7.1 en *Materiales y métodos*). Por otro lado, los valores de pK de agrupaciones de residuos glutamato determinados por RMN se encuentran alrededor de 4, mientras que los valores de pK de los residuos aspartato varían mucho más. Ésto sugiere que los residuos aspartato son más sensibles que los residuos glutamato a las interacciones con el ambiente que los rodea (Forsyth *et al.*, 2002). Por lo tanto, el hecho de que las mutaciones E→D y D→E tengan un efecto pequeño, pero significativo sobre la estabilidad de la proteína, junto al hecho de que E y D son residuos que se encuentran frecuentemente sobre la superficie (lo que significa que no se encuentran restringidos estructural o funcionalmente), hace que las mutaciones del tipo E→D y D→E sean particularmente útiles para demostrar la relación entre estabilidad y evolución.

3.1.1 OBTENCIÓN DE LAS PROTEÍNAS MUTADAS DEL TIPO E→D Y D→E.

En este trabajo de investigación, se han obtenido, mediante técnicas de ingeniería genética un total de 13 variantes de tiorredoxina silvestre (WT) de *E. coli*: 5 proteínas con una mutación simple del tipo E→D (E30D, E44D, E48D,

E85D y E101D) y 8 proteínas con una mutación simple del tipo D→E (D2E, D9E, D10E, D13E, D43E, D47E, D61E y D104E). Las mutaciones deseadas se han diseñado como se explicó en la sección 2.2.1.2 y se obtuvieron mediante mutagénesis dirigida (véase para más detalles la sección 2.2.1.3). Para este tipo de mutaciones se trabajó con el gen de tiorredoxina *WT* clonado en el vector pTK100 (sección 2.2.2.1) y las células JF521 (secciones 2.2.2.1 y 2.2.2.2). La expresión y purificación de proteína se explicó detalladamente en las secciones 2.2.2.5 y 2.2.2.6, respectivamente. La pureza de las proteínas purificadas se ha evaluado mediante una electroforesis *PAGE-SDS* (sección 2.2.2.6).

3.1.2 ESTUDIO DE LA ESTABILIDAD TERMODINÁMICA DE LAS FORMAS MUTANTES DE TIORREDOXINA MEDIANTE CALORIMETRÍA DIFERENCIAL DE BARRIDO.

La estabilidad de tiorredoxina *WT* y de todas las proteínas mutadas obtenidas (5 proteínas con una mutación del tipo E→D y 8 proteínas con una mutación del tipo D→E) se ha determinado por Calorimetría Diferencial de Barrido (*DSC*). Los experimentos se llevaron a cabo a una velocidad de barrido de 1.5 K/min en el intervalo de temperatura de 10-115 °C (véanse detalles de la técnica y análisis de los datos en las secciones 2.2.4.2 y 2.2.4.3, respectivamente). Las disoluciones de proteínas utilizadas en los experimentos de calorimetría se dializaron en HEPES 5 mM, pH 7 y su concentración se determinó espectrofotométricamente según se explicó en la sección 2.2.3.1. Para todas las variantes de tiorredoxina obtenidas, se han realizado varios experimentos de calorimetría a concentraciones de proteína de 0.5 mg/mL y menores, ya que tiorredoxina de *E. coli* presenta efecto de la concentración sobre la temperatura de desnaturalización cuando se trabaja a altas concentraciones de proteína (Ladbury *et al.*, 1993); en este rango de concentraciones, no se han observado efectos de concentración sobre la desnaturalización. Además, la desnaturalización de todas las variantes de tiorredoxina estudiadas ha sido altamente reversible (véase un termograma ilustrativo en la Figura 2.6, sección 2.2.4.2 de Materiales y métodos) y los ajustes de los perfiles de capacidad calorífica al modelo de dos estados en equilibrio, han sido excelentes (véanse en las Figuras 3.1-2 los ajustes realizados y la sección 2.2.4.3 para detalles sobre el modelo de equilibrio de dos estado) y similares a los obtenidos previamente para tiorredoxina *WT* (Georgescu *et al.*,

3. Relación entre estabilidad de proteínas, evolución y estructura...

2001). Merece la pena destacar que el análisis de barrera variable de tiorredoxina, propuesto por Munoz y Sanchez-Ruiz (2004), muestra la desnaturalización térmica de tiorredoxina como un escenario con dos macro-estados. Hemos comprobado que los efectos de las mutaciones sobre la energía libre de desnaturalización (valores de $\Delta\Delta G$ calculados según la ecuación 3.1; ver Tabla 3.1) fueron pequeños, por lo que pudimos usar la ecuación de Schellman (Schellman, 1987) para calcularlos sin introducir errores significativos. La ecuación de Schellman se puede escribir como:

$$\Delta\Delta G = \Delta H_m^0 \left(\frac{\Delta T_m}{T_m^0} \right) \quad (3.1)$$

donde $\Delta\Delta G$ es el efecto sobre el cambio de energía libre de Gibbs (ecuación 3.2) debido a la mutación; T_m^0 y ΔH_m^0 son la temperatura de desnaturalización y el cambio en entalpía de desnaturalización (a la temperatura de desnaturalización), respectivamente, para la forma *WT* de tiorredoxina.

$$\Delta\Delta G = \Delta G(\text{mutante}) - \Delta G(WT) \quad (3.2)$$

ΔT_m es el cambio en la temperatura de desnaturalización debido a la mutación:

$$\Delta T_m = T_m(\text{mutante}) - T_m^0 \quad (3.3)$$

De este modo, hemos caracterizado el efecto de cinco mutaciones E→D sobre la estabilidad de tiorredoxina de *E. coli*. Todas estas mutaciones han sido claramente desestabilizantes (Figuras 3.1 y 3.3), lo cual sugiere que los ambientes de las posiciones mutadas (posiciones con residuos glutamato en tiorredoxina *WT* de *E. coli*) no se encuentran optimizados para mantener interacciones estabilizantes con residuos aspartato (o al menos no lo están tanto como para el caso de residuos E). Acto seguido, hemos estudiado el efecto de 8 variantes de tiorredoxina en las que se ha mutado un residuo aspartato a un residuo glutamato (8 mutantes individuales del tipo D→E. De todos los aspartatos existentes en

3. Relación entre estabilidad de proteínas, evolución y estructura...

tiorredoxina, no se mutado el que originaría la variante D26→E, ya que el D26 se encuentra formando parte del sitio activo de tiorredoxina. Podría anticiparse que, si todas las mutaciones E→D han resultado ser desestabilizantes, todas o la mayoría de las mutaciones de D→E serían estabilizantes. Sin embargo, experimentalmente hemos comprobado que la mayoría de las mutaciones D→E son desestabilizantes o tienen un efecto pequeño sobre la estabilidad y solamente algunas de ellas son ligeramente estabilizantes (Figuras 3.2 y 3.3). Es decir, parece que la mayoría de las posiciones con residuos D en la proteína *WT* tienen ambientes optimizados para la interacción con estos residuos. En términos evolutivos y como sugiere la hipótesis de pseudo-equilibrio, ésto se podría entender como que la evolución ha situado la mayoría de los residuos de ácidos carboxílicos en posiciones de tiorredoxina *WT* en las que puedan mantener interacciones estabilizantes.

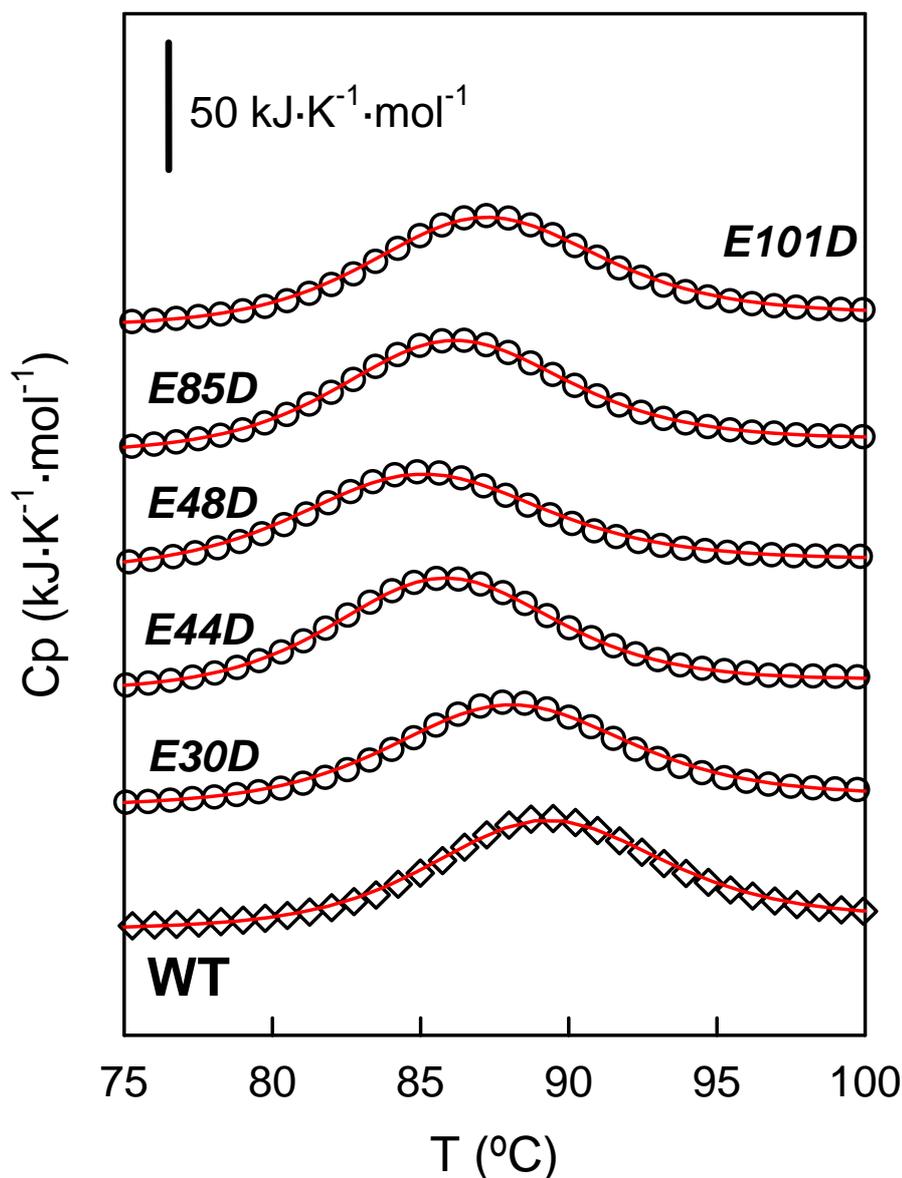


Figura 3.1: Perfiles de capacidad calorífica frente a temperatura de tiorredoxina silvestre (WT) (rombos) y los mutantes *E30D*, *E44D*, *E48D*, *E85D* y *E101D* (círculos), obtenidos mediante Calorimetría Diferencial de Barrido (DSC). Los experimentos se han llevado a cabo a 1.5 K/min. Las disoluciones de proteína se dializaron en HEPES 5 mM, pH7. Para cada proteína se han realizado un mínimo de tres experimentos, a concentraciones de proteína de 0.5 mg/mL y menores, comprobando de este modo que no existe efecto de la concentración de proteína sobre la temperatura de desnaturalización. La línea roja representa el ajuste de los datos experimentales al modelo de equilibrio de dos estados. Los parámetros del ajuste se recogen en la Tabla 3.1. Con objeto de obtener una mayor claridad de la figura, los perfiles de las variantes de tiorredoxina han sido desplazados en el eje de capacidades caloríficas.

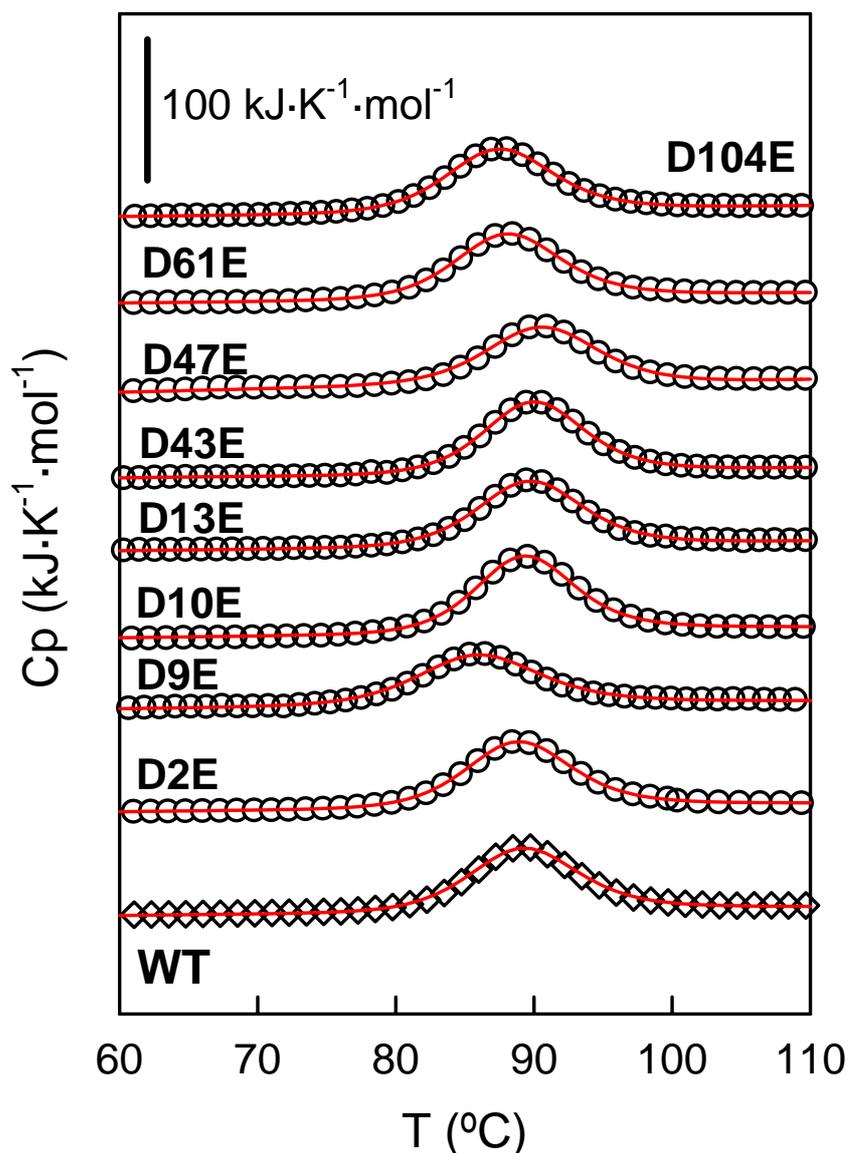


Figura 3.2: Perfiles de capacidad calorífica frente a temperatura de tiorredoxina silvestre (*WT*) (rombos) y los mutantes D2E, D9E, D10E, D13E, D43E, D47E, D61E y D104E (círculos), obtenidos mediante Calorimetría Diferencial de Barrido (*DSC*). Los experimentos se han llevado a cabo a 1.5 K/min. Las disoluciones de proteína se dializaron en HEPES 5 mM, pH7. Para cada proteína se han realizado un mínimo de tres experimentos, a concentraciones de proteína de 0.5 mg/mL y menores, comprobando de este modo que no existe efecto de la concentración de proteína sobre la temperatura de desnaturalización. La línea roja representa el ajuste de los datos experimentales al modelo de equilibrio de dos estados. Los parámetros del ajuste se recogen en la Tabla 3.1. Con objeto de obtener una mayor claridad de la figura, los perfiles de las variantes de tiorredoxina han sido desplazados en el eje de capacidades caloríficas.

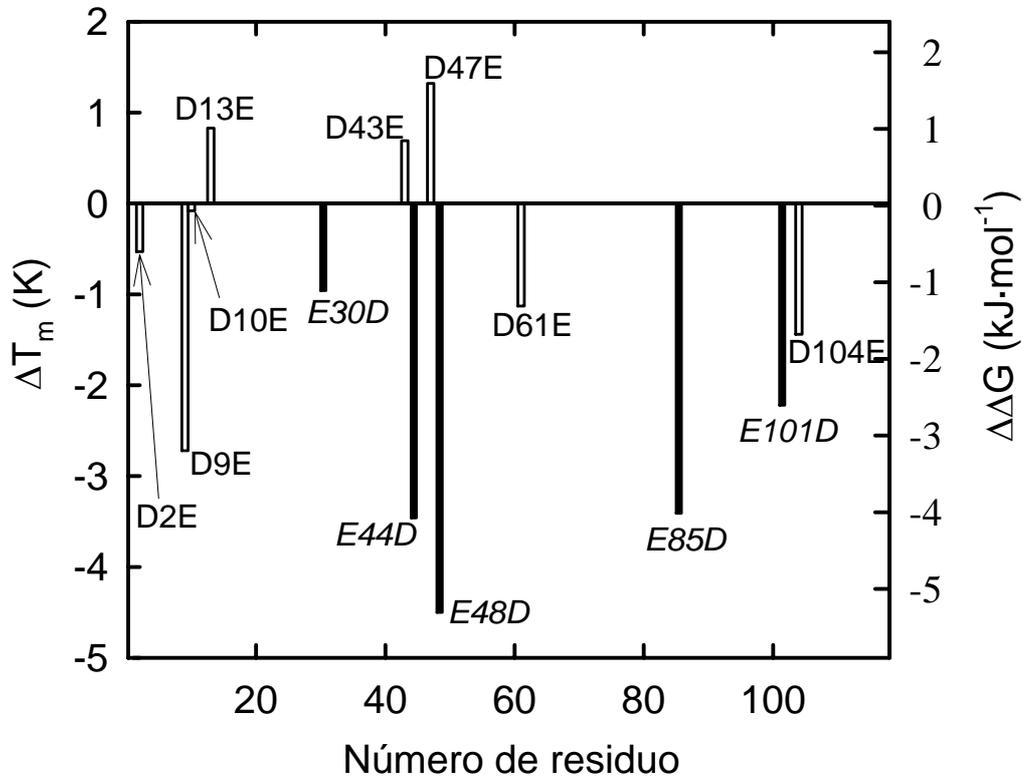


Figura 3.3: Efecto de las mutaciones glutamato (E)→aspartato(D) y aspartato(D)→glutamato(E) sobre la estabilidad de tiorredoxina de *E. coli* (Tabla 3.1). Los cambios en estabilidad se han determinado por el efecto de la mutación sobre la temperatura de desnaturalización (eje de la izquierda) y sobre el cambio en energía libre (eje de la derecha). Las mutaciones D→E y E→D se muestran con barras blancas y negras, respectivamente.

Tabla 3.1: Temperatura de desnaturalización (T_m), cambio en la temperatura de desnaturalización respecto a tiorredoxina *WT* (ecuación 3.3), cambio en la entalpía de desnaturalización (ΔH_m) y el efecto de la mutación en el cambio de la energía libre de Gibbs (ecuación 3.2) derivados del ajuste al modelo de equilibrio de dos estados de cada una de las proteínas estudiadas. Los valores dados tienen en cuenta el sentido de la mutación $D \rightarrow E$.

Proteína	T_m (°C)	ΔT_m (°C)	ΔH (kJ/mol)	$\Delta\Delta G$ (kJ/mol)
WT	89.05	-	430.0	-
D2E	88.52	-0.53	443.9	-0.6
D9E	86.33	-2.72	355.4	-3.2
D10E	88.97	-0.08	456.7	-0.1
D13E	89.88	0.83	435.6	1.0
E30D	88.09	0.96	419.0	1.1
D43E	89.74	0.69	452.3	0.8
E44D	85.59	3.46	438.3	4.1
D47E	90.37	1.32	416.1	1.6
E48D	84.55	4.5	395.8	5.3
D61E	87.92	-1.13	431.9	-1.3
E85D	85.74	3.31	426.9	3.93
E101D	86.83	3.22	429.3	2.6
D104E	87.61	-1.44	427.90	-1.7

3.2 CORRELACIÓN DE LOS EFECTOS DE MUTACIONES DE ÁCIDOS CARBOXÍLICOS SOBRE LA ESTABILIDAD CON SU FRECUENCIA DE APARICIÓN EN ALINEAMIENTOS DE SECUENCIAS

Con objeto de profundizar en la relación existente entre la estabilización D *versus* E y evolución hemos hecho el análisis de un alineamiento de 500 secuencias resultado de una búsqueda realizada con el programa BLAST (Altschul *et al.*, 1997), utilizando tiorredoxina de *E. coli* como secuencia a comparar (secuencia *query*). En dicho alineamiento cuantificamos la frecuencia de aparición de un residuo determinado ($X=D$ o E) en una posición dada (p) usando la siguiente expresión:

$$Z_x(p) = \sum_s w_x(p,s) \quad (3.4)$$

donde $w_x(p,s)$ es un peso asociado al tipo de residuo X en la posición p de una secuencia dada s . Este peso tiene un valor de cero si el residuo presente en la posición p en la secuencia s no es del tipo X . Si el residuo presente en p y s es del tipo X , consideramos dos posibilidades: 1) El residuo presente en la posición p en la secuencia *query* no es del tipo X . Entonces, el peso se fija igual al valor de la similaridad de la secuencia s con la secuencia *query* (es decir, igual al número de coincidencias de la secuencia s en el alineamiento, con la secuencia *query*, dividido por el número de residuos de esta última); 2) El residuo en p en la secuencia *query* es del tipo X . Entonces el peso se fija igual a 1 menos la similaridad de la secuencia. Es decir, se recuentan el número de apariciones de un residuo en una determinada posición en el alineamiento de secuencias y dichas apariciones se pesan por la similaridad existente entre la secuencia y la secuencia *query*, dando un peso cero a cualquier secuencia que es idéntica a la secuencia *query*. De este modo, se tiene en cuenta que coincidencias entre secuencias con elevada similaridad reflejan cómo las proteínas correspondientes divergieron recientemente en la evolución.

3. Relación entre estabilidad de proteínas, evolución y estructura...

La hipótesis de pseudo-equilibrio sugiere que el valor de la siguiente función de energía libre:

$$\Gamma_{D \rightarrow E(p)} = RT \ln \left[\frac{Z_E(p)}{Z_D(p)} \right] \quad (3.5)$$

debería reflejar el efecto de las mutaciones en el sentido $D \rightarrow E$ en la posición p sobre la estabilidad. En la ecuación 3.5, R es la constante de los gases (8.314 kJ/Kmol) y T es una temperatura que puede ser interpretada como una medida de la presión evolutiva (Dokholyan y Shakhnovich, 2001). En nuestros cálculos, consideramos siempre $T=300$ K. Además, con objeto de poder comparar los cambios de mutaciones $D \rightarrow E$ y $E \rightarrow D$ sobre la estabilidad de tiorredoxina, con la frecuencia de aparición de estos residuos en un alineamiento de secuencias realizado para tiorredoxina, consideramos todas las mutaciones en el mismo sentido; es decir, para el caso de las mutaciones del tipo $E \rightarrow D$, los cambios en estabilidad se han calculado, invirtiendo el signo a los valores obtenidos con las ecuaciones 3.3 y 3.4, o bien, utilizando la siguiente expresión:

$$\Delta T_m = (T_m(WT, E) - T_m(mutante, D)) \quad (3.6)$$

Hay que destacar que en el alineamiento obtenido para tiorredoxina de *E.coli*, la mayoría de las secuencias poseen una similaridad con la secuencia *query* de 0.25 o mayor, lo que nos indica que la mayoría de las correspondientes proteínas comparten el mismo plegamiento que tiorredoxina. Teniendo en cuenta la hipótesis de pseudo-equilibrio, los efectos de las mutaciones en el sentido $D \rightarrow E$ sobre la estabilidad de tiorredoxina (un total de 13 valores experimentales de $\Delta\Delta G_{D \rightarrow E}$ correspondientes a los distintos residuos carboxílicos existentes en la proteína) deben de correlacionar con la frecuencia de aparición $\Gamma_{D \rightarrow E}$ de dichos residuos en un alineamiento de secuencias hecho para tiorredoxina de *E. coli*. Para comprobar dicha correlación, se han calculado las frecuencias de aparición $\Gamma_{D \rightarrow E}$ de los residuos carboxílicos estudiados considerando distintos rangos de similaridades con la secuencia *query*, desde 0.25 hasta 0.6. En cada caso

3. Relación entre estabilidad de proteínas, evolución y estructura...

particular, la sumatoria en la ecuación 3.5 que define Z se extiende a todas las secuencias comprendidas en el rango de similaridad especificado (Figura 3.4).

Los resultados han sido sorprendentes. En todos los casos estudiados, las correlaciones (de las 13 variantes de tiorredoxina estudiadas) han sido muy buenas y robustas es decir, se ha obtenido una correlación buena cuando se ha usado un rango amplio de similaridad entre secuencias para el cálculo de Z (un rango de similaridad de 1-0.3 comprende 309 secuencias), pero también cuando se ha usado un rango más estrecho de similaridad entre secuencias (un rango de similaridad de 1-0.6 comprende sólo 23 secuencias). Además, hay que señalar que el valor de la pendiente de la correlación de $\Delta\Delta G_{D \rightarrow E}$ versus $\Gamma_{D \rightarrow E}$ es razonablemente cercano a la unidad (aproximadamente 0.35) y que el valor de la ordenada en el origen es próximo a cero (Figura 3.4).

En la Tabla 3.2 se pueden observar los valores de la pendiente (RT) y la ordenada en el origen (A) del mejor ajuste lineal por mínimos cuadrados realizado para cada una de las correlaciones obtenidas.

Además, se ha calculado para cada correlación su significancia estadística (p), es decir, la probabilidad de que una correlación se dé al azar. En todos los casos se han obtenido valores de p del orden de 10^{-4} ; si tenemos en cuenta que una correlación se considera significativa si tiene $p < 10^{-2}$, los valores de p obtenidos son una comprobación más de la robustez de las correlaciones de $\Delta\Delta G_{D \rightarrow E}$ versus $\Gamma_{D \rightarrow E}$.

Para el cálculo de las frecuencias de aparición de un residuo en un alineamiento de secuencias (Z) se han utilizado los programas FRC-2.exe y Sec.exe (desarrollados en el grupo de investigación de J.M. Sánchez-Ruiz); estos programas utilizan el fichero de alineamientos de secuencias llamado Trx.aln (parámetros de entrada: 500 secuencias; 2 tandas; corte 30). El cálculo de los valores de p se ha realizado con el programa *scriptp.m* creado en Matlab (Apéndice II, programa 9).

3. Relación entre estabilidad de proteínas, evolución y estructura...

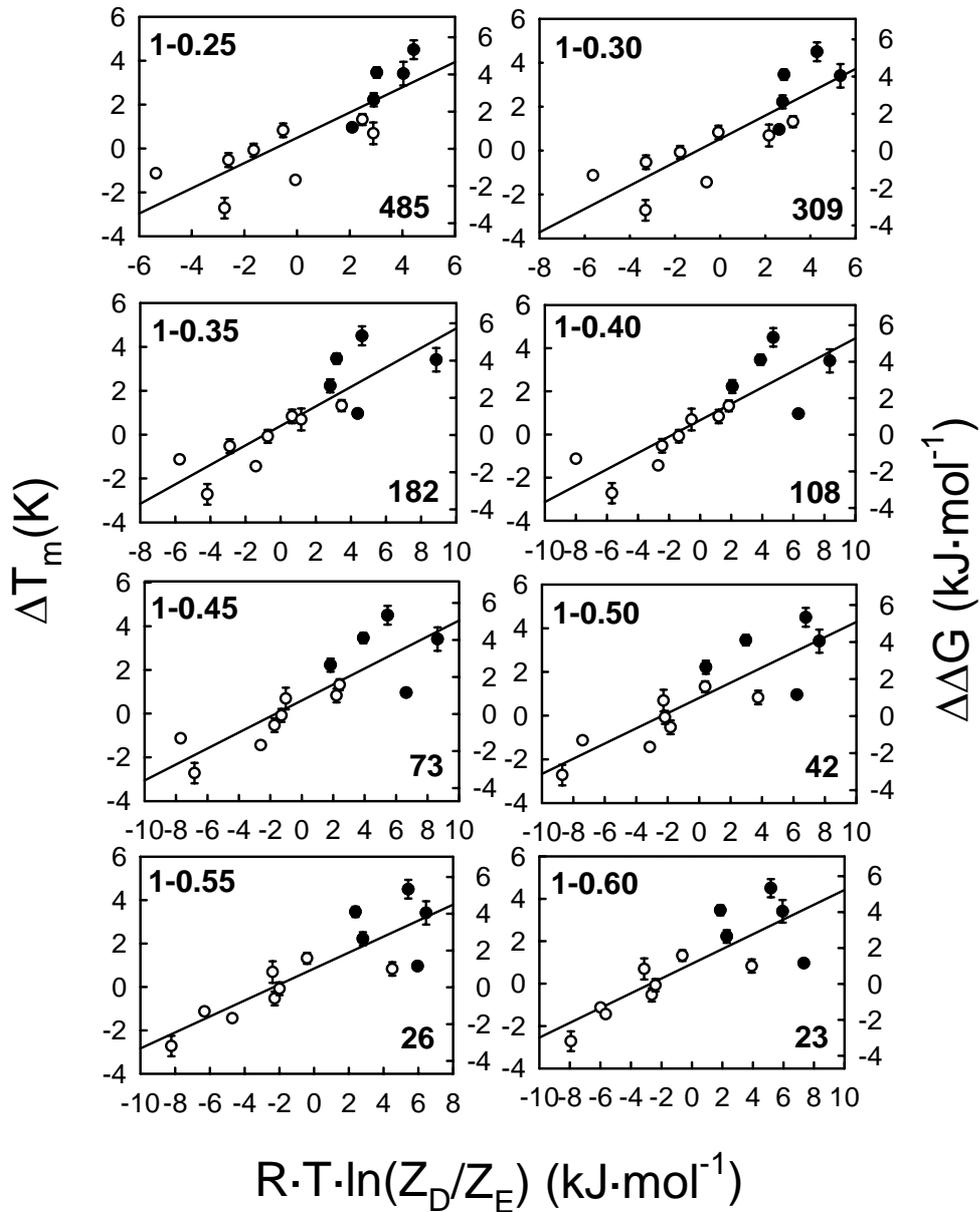


Figura 3.4: Correlación entre los efectos de mutaciones en el sentido $D \rightarrow E$ sobre la estabilidad de tiorredoxina y la frecuencia de aparición de los residuos D y E , Z_D y Z_E , respectivamente (ecuación 3.4) en un alineamiento de secuencias realizado para tiorredoxina de *E. coli*. Los símbolos negros representan las mutaciones originales del tipo $E \rightarrow D$, a las que se les ha cambiado el signo de su efecto sobre la estabilidad (ecuación 3.6). Todos los símbolos representan un promedio de, al menos, tres determinaciones de estabilidad realizadas para cada mutante y su desviación estándar. En cada panel se representa: en la esquina superior izquierda el rango de similaridad utilizado y en la esquina inferior derecha el número de secuencias existente en el rango de similaridad considerado.

3. Relación entre estabilidad de proteínas, evolución y estructura...

Los resultados mostrados hasta el momento, indican cómo la evolución molecular introduce los residuos D y E en las posiciones estudiadas de manera aleatoria, pero con frecuencias que reflejan aproximadamente los efectos correspondientes sobre la estabilidad de las proteínas. Los valores diferentes de $\Delta\Delta G_{D \rightarrow E}$ (y $\Gamma_{D \rightarrow E}$) obtenidos para cada residuo carboxílico que se muestran en la Figura 3.4, demuestran que la contribución que realiza un residuo carboxílico a la estabilidad de la proteína es dependiente de la posición. Esta dependencia refleja muy probablemente las interacciones de los residuos D/E con el ambiente que los rodea en la estructura nativa de la proteína (aunque no podemos excluir efectos del estado desnaturalizado (Klein-Seetharaman *et al.*, 2002; Guzman-Casado *et al.*, 2003; Robic *et al.*, 2003); véanse las conclusiones en la sección 3.3). Sin embargo, el ambiente de un grupo no es “estático”, sino que cambia durante el curso de la evolución molecular. Así, considerando la hipótesis de pseudo-equilibrio, se puede explicar cómo la presencia simultánea de grupos que mantienen interacciones estabilizantes se favorece a lo largo de la evolución gracias al carácter estabilizante de las interacciones en las que participan.

Tabla 3.2: Valores resultantes del ajuste lineal por mínimos cuadrados realizado a las correlaciones de $\Delta\Delta G_{D \rightarrow E}$ versus $\Gamma_{D \rightarrow E}$ que se muestran en la Figura 3.4.

Similaridad	A (kJ/Kmol)	RT (kJ/Kmol)	r^2	θ (K)	p
0.25	0.457±0.350	0.580±0.116	0.696	69.76 ± 13.95	4·10 ⁻⁴
0.30	0.504±0.338	0.535±0.103	0.711	64.35 ± 12.39	4·10 ⁻⁴
0.35	0.368±0.344	0.455±0.085	0.713	54.73 ± 10.24	4·10 ⁻⁴
0.40	0.633±0.340	0.381±0.049	0.702	45.83 ± 9.01	4·10 ⁻⁴
0.45	0.576±0.334	0.367±0.070	0.715	44.14 ± 8.42	3·10 ⁻⁴
0.50	0.683±0.323	0.350±0.066	0.718	42.10 ± 7.95	3·10 ⁻⁴
0.55	0.822±0.329	0.371±0.070	0.717	44.62 ± 8.46	3·10 ⁻⁴
0.60	0.906±0.358	0.351±0.075	0.664	42.22 ± 9.07	7·10 ⁻⁴

3. Relación entre estabilidad de proteínas, evolución y estructura...

En un intento por determinar interacciones entre grupos de residuos que se encuentran conservados en la evolución, se ha realizado un test de asociación de χ^2 . Los cálculos se basan en la descripción existente en el capítulo 3 de Ewens y G.R. (2001). Se consideran dos posiciones p_A y p_B , y el tipo de aminoácido presente en la secuencia de tiorredoxina en estas posiciones: X_A y X_B , respectivamente. Además, si se considera el alineamiento de secuencias obtenido para tiorredoxina de *E. coli* (secuencia *query*), se denomina Y_{11} al número de secuencias en el cual podemos encontrar el residuo X_A en la posición p_A y el residuo X_B en la posición p_B . Del mismo modo, se denomina Y_{12} al número de secuencias en el cual el tipo de aminoácido en la posición p_A es del tipo X_A pero el aminoácido en la posición p_B no es del tipo X_B . Con este mismo criterio se definen Y_{21} y Y_{22} . Se definen también E_{11} , E_{12} , E_{21} y E_{22} , como el número de secuencias en las cuales se cumple la hipótesis nula, es decir, son las secuencias en las cuales se considera que no existe correlación entre los residuos X_A y X_B en las posiciones p_A y p_B . El test de asociación de χ^2 para dos tipos de aminoácidos en dos posiciones específicas se basa en la siguiente estadística:

$$\chi^2 = \frac{\sum (Y_{JK} - E_{JK})^2}{E_{JK}} \quad (3.7)$$

donde la sumatoria tiene cuatro términos correspondientes a las cuatro posibilidades existentes para el subíndice JK (11, 12, 21, 22). Esta estadística se ajusta a una distribución de χ^2 con un grado de libertad y se ha obtenido un valor de χ^2 de 3.84 correspondiente al 5% de significancia. Además, y como se explica más adelante, muchas de las asociaciones detectadas (véase panel inferior de la Figuras 3.5-12) son estadísticamente significativas, muy probablemente consecuencia de la similaridad existente entre las secuencias que participan.

Como ejemplo ilustrativo, se muestra en las Figuras 3.5-12 los valores de χ^2 resultantes de la asociación entre las posiciones 2, 9, 15, 61, 104 (D en tiorredoxina *WT* de *E. coli*), 30, 44, 48 (E en tiorredoxina *WT* de *E. coli*) y el resto de posiciones en la secuencia de tiorredoxina. En todos los casos, se ha utilizado el conjunto de secuencias comprendido en el rango de similaridad 1-0.3 con la

3. Relación entre estabilidad de proteínas, evolución y estructura...

secuencia *query* (tioredoxina de *E. coli*). Se puede observar como muchos de los valores de χ^2 obtenidos son mayores al valor correspondiente al 5% de significancia (3.84). Además, hemos considerado que la mayoría de estas asociaciones significativas no nos están ofreciendo una información estructural, sino que solamente reflejan la similaridad entre secuencias (es decir, dos residuos que se encuentran en la secuencia de tioredoxina de *E. coli*, es muy probable que se encuentren también en secuencias que tienen una similaridad significativa a la de tioredoxina). El cálculo de χ^2 lo hemos realizado para todos los residuos ácidos carboxílicos en tioredoxina *WT*, excluyendo el residuo D26 que forma parte del centro activo de tioredoxina. Las interacciones detectadas se recogen en la Tabla 3.3. Hemos encontrado 10 interacciones residuo-residuo con un valor de $\chi^2 > 50$ (un límite sugerido heurísticamente del análisis de las interacciones residuo-residuo encontradas en el test de D9 que se muestra en la Figura 3.7).

En la Figura 3.7 los dos valores de χ^2 más elevados (61 y 133) corresponden a los residuos N63 y T66, que están muy próximos a D9 (2.58 Å) en la estructura nativa de tioredoxina de *E. coli* (Katti *et al.*, 1990) y claramente se encuentran interaccionando (véase Tabla 3.3). Tres de las 10 interacciones residuo-residuo detectadas son “falsos positivos” (véanse Figuras 3.6, 3.11 y 3.12), es decir, corresponden a residuos que se encuentran a una distancia de más de 8 Å en la estructura nativa de tioredoxina *WT*. Sin embargo, seis de ellos corresponden a contactos residuo-residuo en la estructura nativa de tioredoxina (Katti *et al.*, 1990) con distancias entre los átomos más cercanos aproximadamente de 3 Å o menores (véase Tabla 3.3). Luego podemos concluir diciendo que cada una de estas interacciones contribuye algunos kJ/mol a la estabilidad termodinámica, ya que cada una de las mutaciones de ácidos carboxílicos que se han realizado originan cambios en la energía libre de desnaturalización de esta magnitud (Figuras 3.1, 3.2 y 3.3).

Para la obtención de las distancias entre residuos se han utilizado los programas: DIS-ALL.exe y ASA-ENG.exe (parámetros de entrada: radios de Chotia; tamaño sonda 1.4 Å; 2000 puntos); para el cálculo de los valores de χ^2 se han utilizado los siguientes programas: ASSOC.exe, ASSOC2 y CHI2.exe (ambos necesitan el fichero Trx.aln; parámetros de entrada: 500 secuencias; 2 tandas;

corte 30). Todos estos programas han sido desarrollados en el grupo de investigación de J.M. Sanchez-Ruiz.

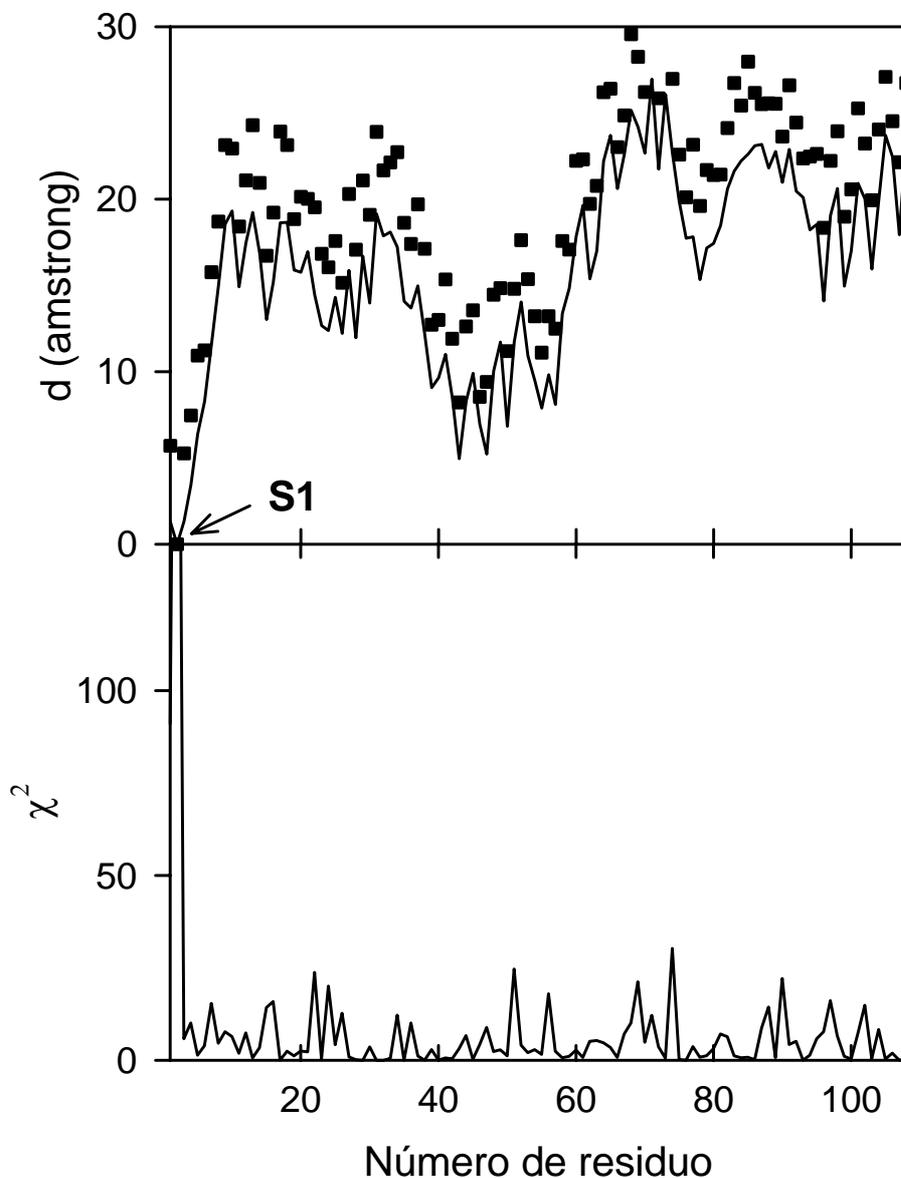


Figura 3.5: Interacciones residuo-residuo en las que participa el aspartato **D2** de tioredoxina sugeridas por un test de asociación de χ^2 del alineamiento de secuencias obtenido para tioredoxina de *E. coli*. El panel superior muestra la distancia de todos los residuos de tioredoxina a D2. Se muestran dos tipos de distancias: las distancias entre los carbonos β (cuadrados negros) y las distancias entre los átomos más cercanos (línea continua). El panel inferior da los valores de χ^2 correspondientes al test de asociación entre cada residuo en tioredoxina y D2.

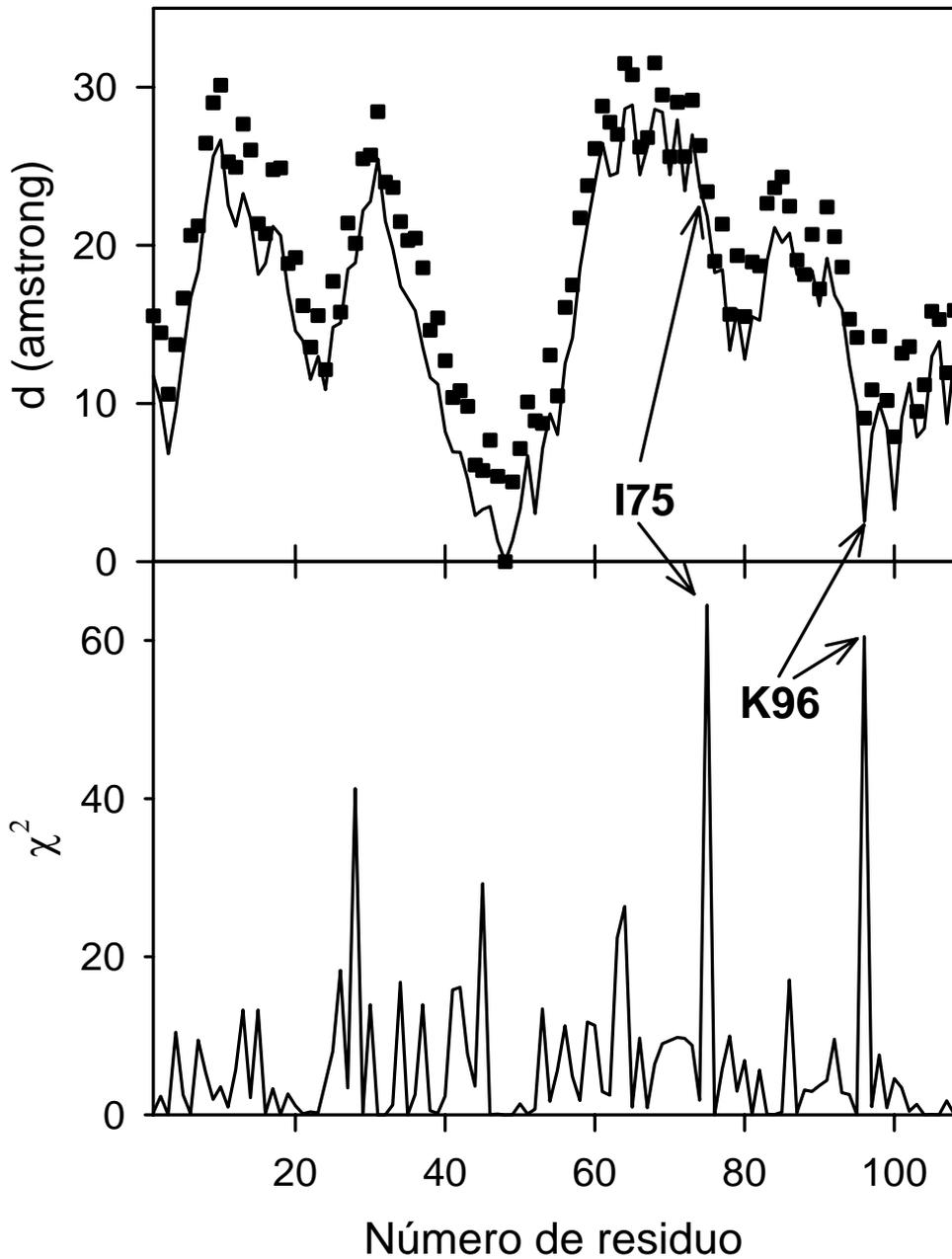


Figura 3.6: Interacciones residuo-residuo en las que participa el aspartato **E48** de tioredoxina sugeridas por un test de asociación de χ^2 del alineamiento de secuencias obtenido para tioredoxina de *E. coli*. El panel superior muestra la distancia de todos los residuos de tioredoxina a E48. Se muestran dos tipos de distancias: las distancias entre los carbonos β (cuadrados negros) y las distancias entre los átomos más cercanos (línea continua). El panel inferior da los valores de χ^2 correspondientes al test de asociación entre cada residuo en tioredoxina y E48.

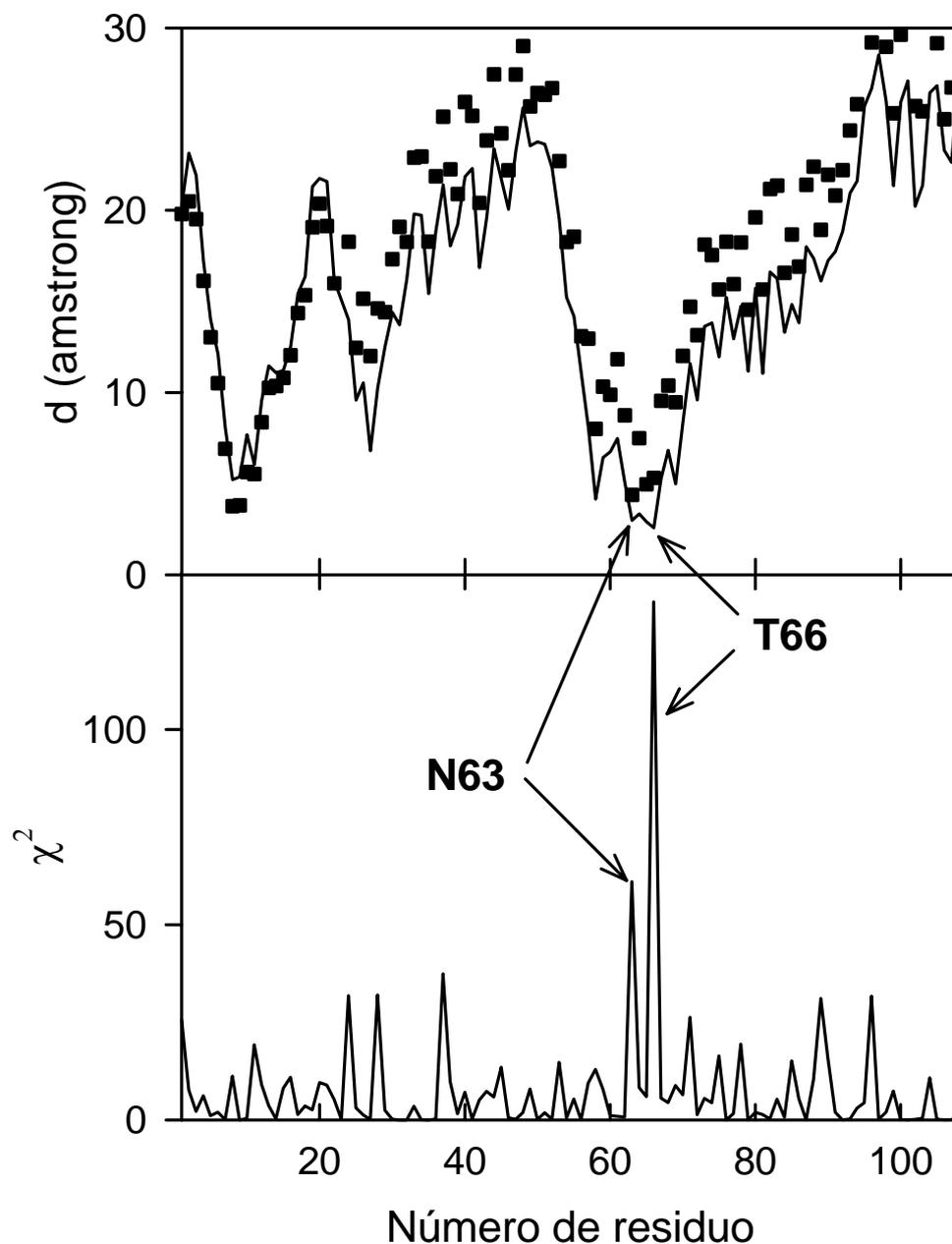


Figura 3.7: Interacciones residuo-residuo en las que participa el aspartato **D9** de tiorredoxina sugeridas por un test de asociación de χ^2 del alineamiento de secuencias obtenido para tiorredoxina de *E. coli*. El panel superior muestra la distancia de todos los residuos de tiorredoxina a D9. Se muestran dos tipos de distancias: las distancias entre los carbonos β (cuadrados negros) y las distancias entre los átomos más cercanos (línea continua). El panel inferior da los valores de χ^2 correspondientes al test de asociación entre cada residuo en tiorredoxina y D9.

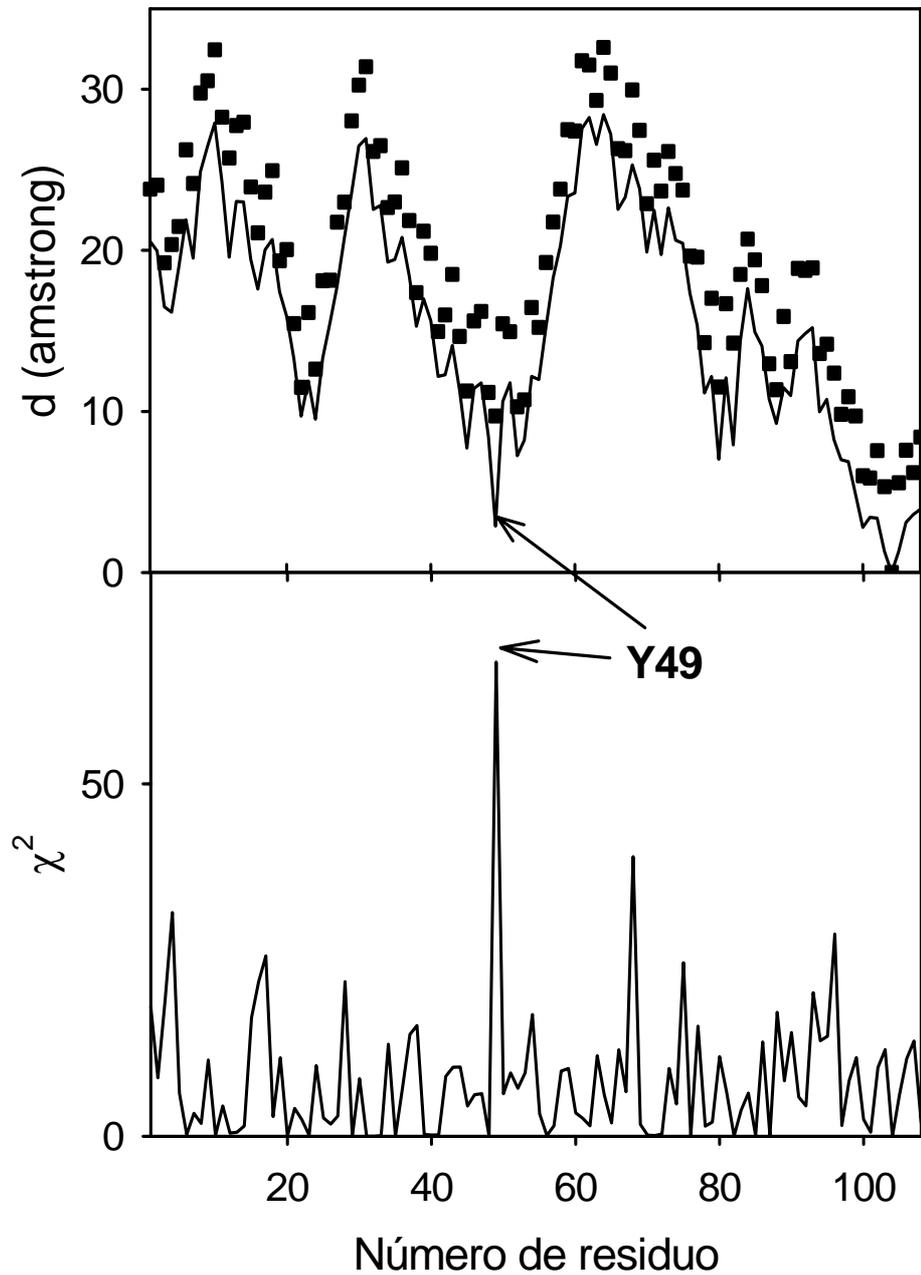


Figura 3.8: Interacciones residuo-residuo en las que participa el aspartato **D104** de tiorredoxina sugeridas por un test de asociación de χ^2 del alineamiento de secuencias obtenido para tiorredoxina de *E. coli*. El panel superior muestra la distancia de todos los residuos de tiorredoxina a D104. Se muestran dos tipos de distancias: las distancias entre los carbonos β (cuadrados negros) y las distancias entre los átomos más cercanos (línea continua). El panel inferior da los valores de χ^2 correspondientes al test de asociación entre cada residuo en tiorredoxina y D104.

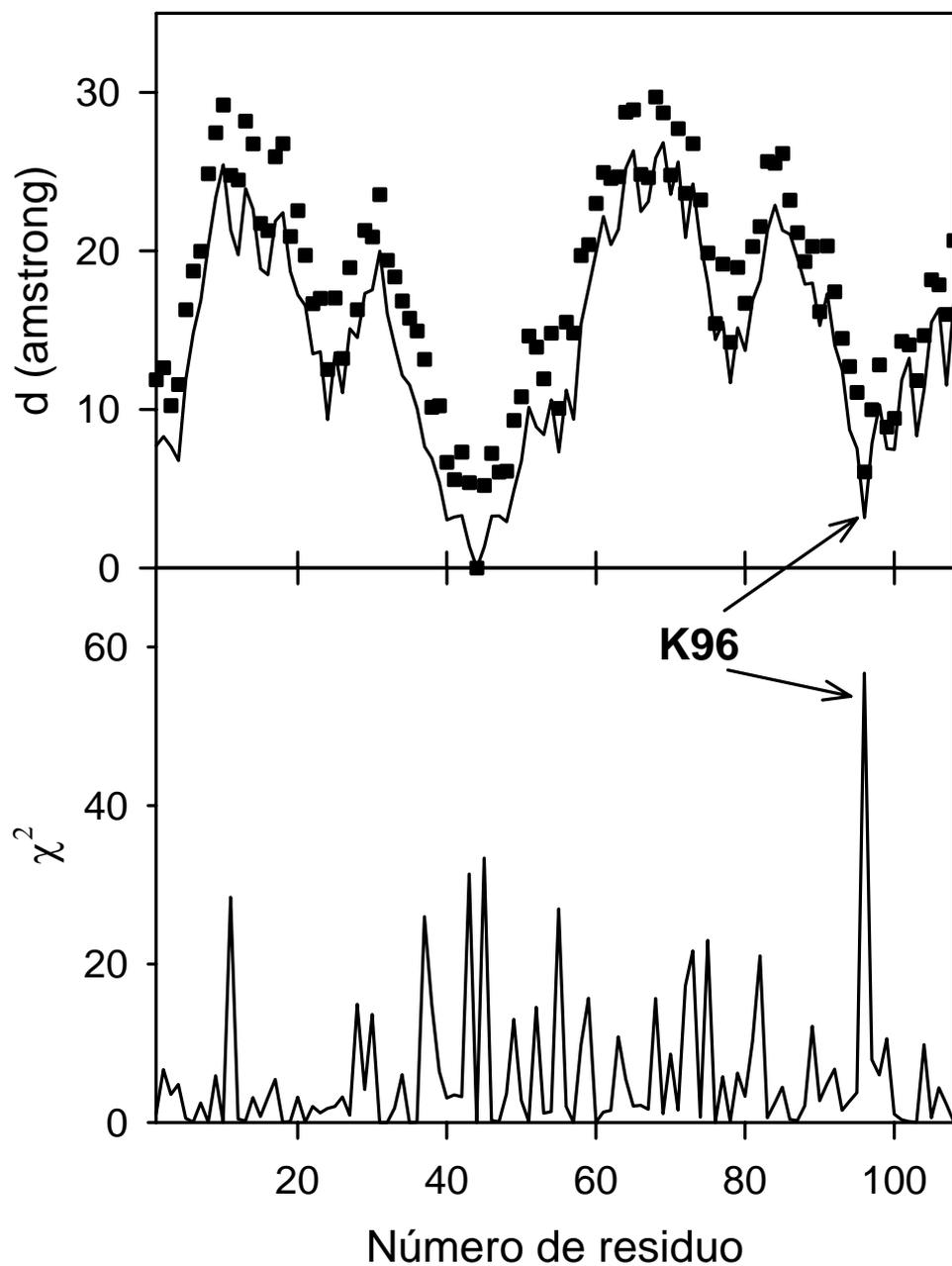


Figura 3.9: Interacciones residuo-residuo en las que participa el aspartato **E44** de tiorredoxina sugeridas por un test de asociación de χ^2 del alineamiento de secuencias obtenido para tiorredoxina de *E. coli*. El panel superior muestra la distancia de todos los residuos de tiorredoxina a E44. Se muestran dos tipos de distancias: las distancias entre los carbonos β (cuadrados negros) y las distancias entre los átomos más cercanos (línea continua). El panel inferior da los valores de χ^2 correspondientes al test de asociación entre cada residuo en tiorredoxina y E44.

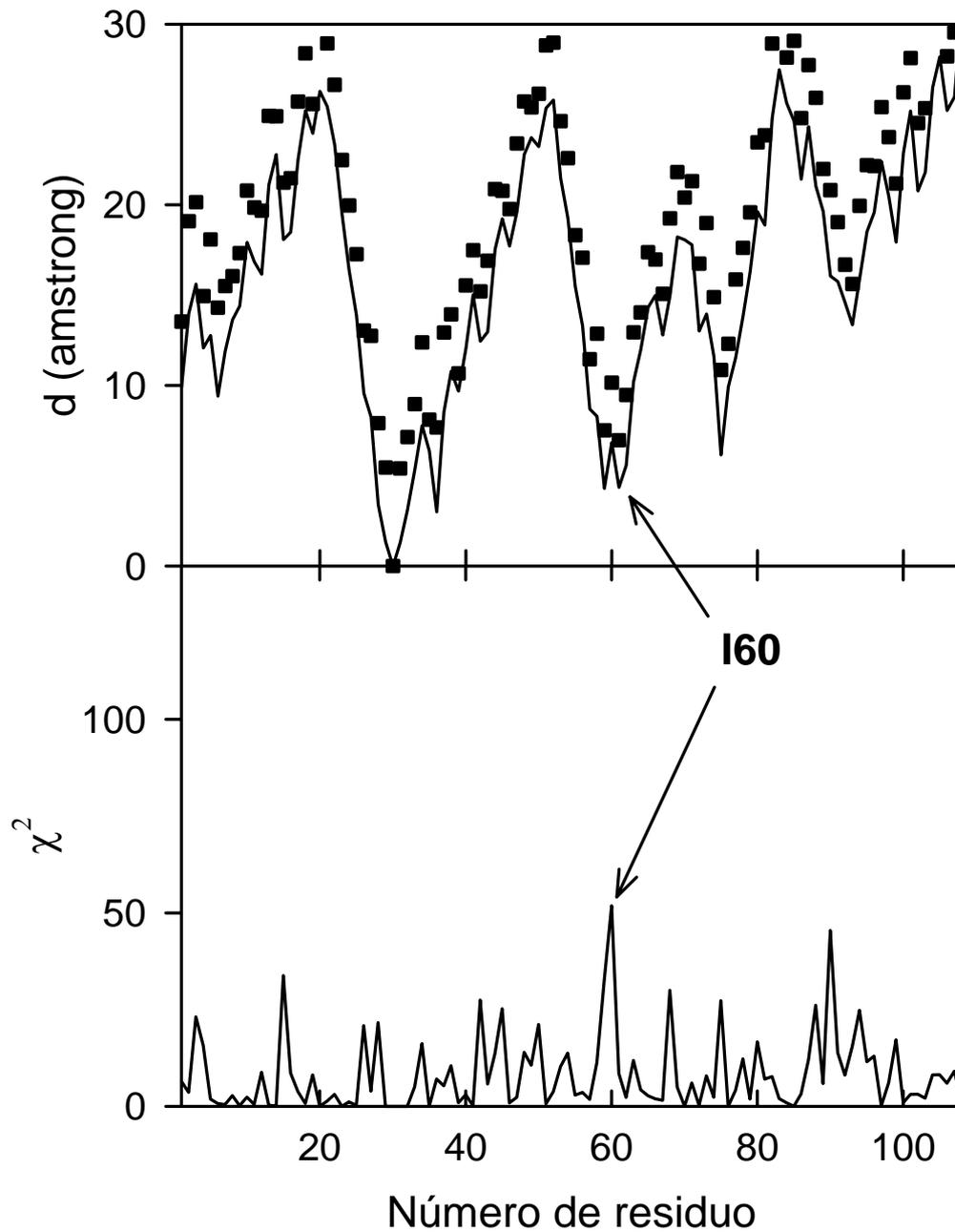


Figura 3.10: Interacciones residuo-residuo en las que participa el aspartato E30 de tioredoxina sugeridas por un test de asociación de χ^2 del alineamiento de secuencias obtenido para tioredoxina de *E. coli*. El panel superior muestra la distancia de todos los residuos de tioredoxina a E30. Se muestran dos tipos de distancias: las distancias entre los carbonos β (cuadrados negros) y las distancias entre los átomos más cercanos (línea continua). El panel inferior da los valores de χ^2 correspondientes al test de asociación entre cada residuo en tioredoxina y E30.

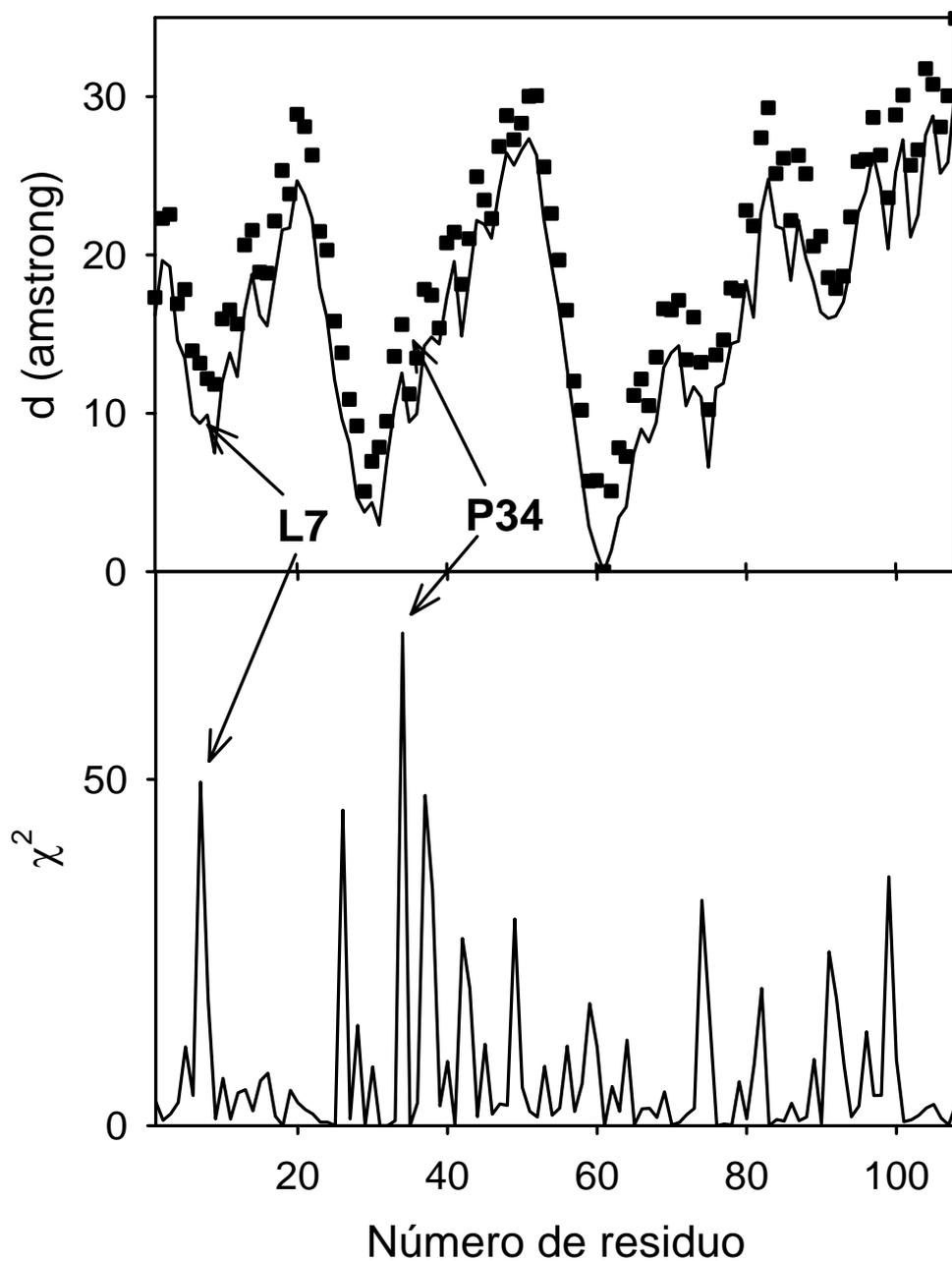


Figura 3.11: Interacciones residuo-residuo en las que participa el aspartato **D61** de tiorredoxina sugeridas por un test de asociación de χ^2 del alineamiento de secuencias obtenido para tiorredoxina de *E. coli*. El panel superior muestra la distancia de todos los residuos de tiorredoxina a D61. Se muestran dos tipos de distancias: las distancias entre los carbonos β (cuadrados negros) y las distancias entre los átomos más cercanos (línea continua). El panel inferior da los valores de χ^2 correspondientes al test de asociación entre cada residuo en tiorredoxina y D61.

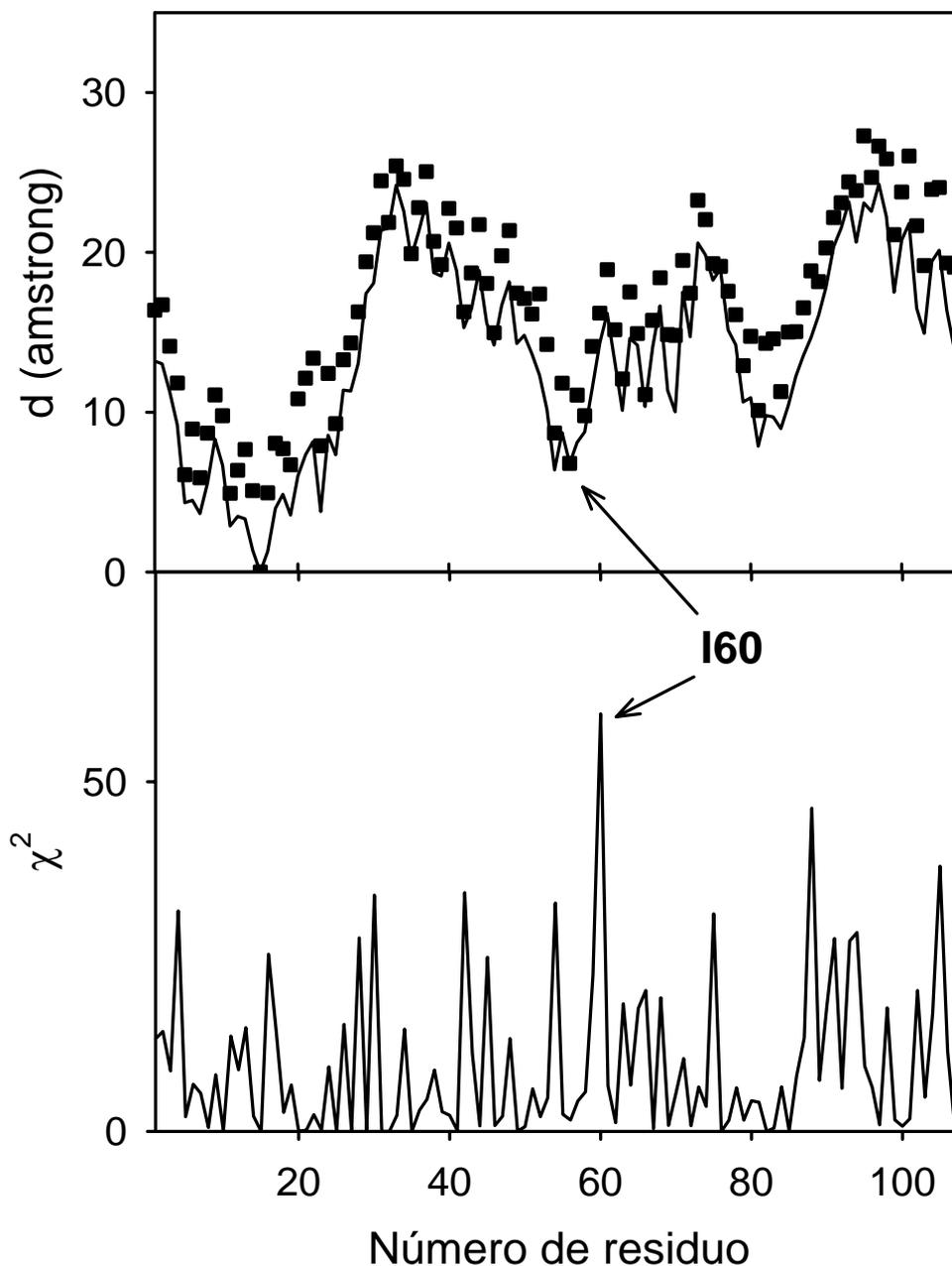


Figura 3.12: Interacciones residuo-residuo en las que participa el aspartato D15 de tiorredoxina sugeridas por un test de asociación de χ^2 del alineamiento de secuencias obtenido para tiorredoxina de *E. coli*. El panel superior muestra la distancia de todos los residuos de tiorredoxina a D15. Se muestran dos tipos de distancias: las distancias entre los carbonos β (cuadrados negros) y las distancias entre los átomos más cercanos (línea continua). El panel inferior da los valores de χ^2 correspondientes al test de asociación entre cada residuo en tiorredoxina y D15.

Tabla 3.3: Interacciones residuo-residuo en las que participan los residuos glutamato y aspartato de tiorredoxina de *E. coli* sugeridas por un análisis de valores de asociación de χ^2 del alineamiento de secuencias hecho para tiorredoxina. En la tabla se muestran sólo las interacciones con un valor de χ^2 mayor de 50 y se han ordenado en función de la distancia entre los átomos más cercanos en la estructura nativa de tiorredoxina.

Residuos	Distancia entre los átomos más cercanos (Å)	Átomos más cercanos	χ^2
D2-S1	1.30	N C	90.9
E48-K96	2.54	OE1...NZ	60.5
D9-T66	2.58	OD2 OG1	132.6
D104-Y49	2.87	OD1...OH	67.3
D9-N63	2.98	N...OD1	61.0
E44-K96	3.17	OE1 NZ	56.7
E30-I60	6.84	N...CG1	51.8
D61-P34	12.53	OD2 CD	71.1
D15-I60	14.34	OD1...N	59.7
E48-I75	21.82	OE2 C	64.5

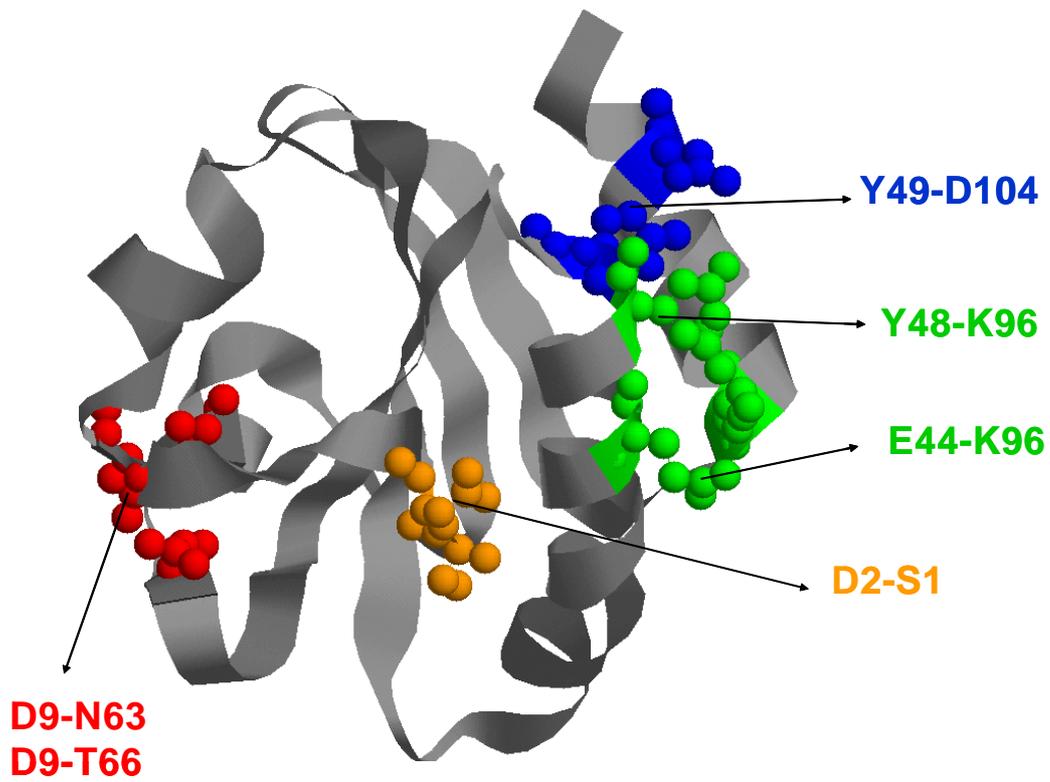


Figura 3.13: Ilustración de la molécula de tiorredoxina (2TRX.pdb) en la que se representan las interacciones residuo-residuo, en las que participan los residuos glutamato y aspartato de tiorredoxina de *E. coli*, sugeridas por el análisis de asociación de χ^2 en el alineamiento de secuencias realizado para tiorredoxina de *E. coli*. Las interacciones que se muestran tienen un valor de χ^2 mayores de 50 (Tabla 3.3).

3.3 CONCLUSIONES

Los resultados que hemos mostrado en este capítulo (Figuras 3.3 y 3.5-12) nos muestran claramente que, para posiciones que no se encuentran limitadas estructural o funcionalmente, las mutaciones ocurren a lo largo de la evolución con probabilidades relacionadas con sus efectos sobre la estabilidad (Godoy-Ruiz *et al.*, 2004). Son una aportación más que ayuda a clarificar el mecanismo (muy poco conocido) de fijación de mutaciones durante la evolución. Estos resultados sugieren que los ambientes de muchos grupos en proteínas pueden encontrarse optimizados por interacciones estabilizantes con un alto grado de especificidad (por ejemplo, aspartato *versus* glutamato). Hemos mostrado (Figuras 3.4, 3.5-3.12, y Tabla 3.3) que dichas interacciones estabilizantes se pueden detectar en un alineamiento de secuencias, resultado que, como explicamos a continuación, puede ser muy útil en la derivación de información estructural sólo realizando análisis comparativos de secuencias.

Aproximadamente, sólo la mitad de las secuencias de genomas conocidos son homólogas a proteínas de estructura conocida (Chothia *et al.*, 2003); ésto ha hecho que uno de los objetivos principales en biología estructural sea la predicción de estructura conociendo la secuencia de proteínas y la búsqueda de información estructural (contactos entre residuos) mediante el análisis de posiciones correlacionadas en un alineamiento de secuencias. Sin embargo, los resultados obtenidos hasta el momento no parecen ser muy alentadores. Por ejemplo, en un estudio reciente (Fariselli *et al.*, 2001), utilizando redes neuronales y un amplio conjunto de datos, se ha conseguido sólo un 20% de precisión en la predicción de contactos en la proteína; mientras que se piensa que es necesario por lo menos un 50% de precisión para determinar la estructura tridimensional.

Los resultados expuestos en este apartado (Figura 3.6-3.9 y Tabla 3.3) demuestran cómo se pueden detectar interacciones estabilizantes (del orden de decenas de kJ/mol) en alineamientos de secuencias con un 60-70% de precisión en la predicción. Además, existe la posibilidad de que dichas interacciones puedan estar formándose en el estado desnaturalizado y ser detectadas mediante un análisis de un alineamiento de secuencias. Estudios recientes (Shortle y Ackerman, 2001 y Klein-Seetharaman *et al.*, 2002) sugieren la existencia de interacciones de largo alcance y procesos cooperativos en estados

3. Relación entre estabilidad de proteínas, evolución y estructura...

desnaturalizados. En Guzman-Casado *et al.*, 2003 se muestran evidencias de este comportamiento complejo para el caso de ribonucleasa H termofílica; experimentalmente se observó un cambio abrupto de la pendiente en una representación de la temperatura de desnaturalización *versus* al pH. Un estudio mutacional realizado sobre ribonucleasa H termofílica mostró que se puede destruir la estructura residual formada en el estado desnaturalizado haciendo que las características termodinámicas de la proteína termofílica reviertan a las de la mesofílica. Este comportamiento se encontraba relacionado con una toma extremadamente cooperativa de protones en el estado desnaturalizado (formación de agrupaciones de residuos hidrofóbicos) en un rango estrecho de pH. La formación de estructura residual en el estado desnaturalizado puede modular la estabilidad de ribonucleasa H termofílica y como tal, puede estar diseñada por la evolución (Robic *et al.*, 2003), ya que garantiza una temperatura de desnaturalización alta sin que su energía libre de desnaturalización llegue a ser excesivamente alta (lo cual haría a la proteína muy rígida).

BIBLIOGRAFÍA

- Altschul, S.F., T.L. Madden, A.A. Schäffer, H. Zhang, J. Zhang, Z. Z Zhang, W. Miller and D.J. Lipman (1997). *Nucleic Acids Res* **25** **3389-3402**.
- Chothia, C., J. Gough, C. Vogel and S.A. Teichmann (2003). *Science* **300**: 1701-1703.
- Dill, K.A. (1990). *Biochemistry* **29**(31): 7133-55.
- Dokholyan, N.V. and E.I. Shakhnovich (2001). *J Mol Biol* **312**(1): 289-307.
- Ewens, W.J. and G. G.R. (2001). New York, Springer.
- Fariselli, P., O. Olmea, A. Valencia and R. Casadio (2001). *Proteins* **5**: 157-162.
- Forsyth, W.R., J.M. Antosiewicz and A.D. Robertson (2002). *Struct Funct Genet* **48**(388-403).
- George, D.G., W.C. Baker and L.T. Hunt (1990). *Methods Enzymol* **183**: 333-351.
- Georgescu, R.E., M.M. Garcia-Mira, M.L. Tasayco and J.M. Sanchez-Ruiz (2001). *Eur J Biochem* **268**(5): 1477-85.
- Godoy-Ruiz, R., R. Perez-Jimenez, B. Ibarra-Molero and J.M. Sanchez-Ruiz (2004). *J Mol Biol* **336**(2): 313-8.
- Guzman-Casado, M., A. Parody-Morreale, S. Robic, S. Marqusee and J.M. Sanchez-Ruiz (2003). *J Mol Biol* **329**(4): 731-43.
- Henikoff, S. and T.G. Henikoff (1992). *Proc Natl Acad Sci U S A* **89**: 10915-10919.
- Katti, S.K., D.M. LeMaster and H. Eklund (1990). *J. Mol. Biol.* **212**: 167-184.
- Klein-Seetharaman, J., M. Oikawa, S.B. Gimshaw, J. Wirmer, E. Duchardt and T. Ueda (2002). *Science* **295**: 1719-1722.
- Knappenberger, J.E., Smith, J.E., Thorpe, S.H., Zitzewitz, J.Z. & Matthews, C.R. (2002). *J Mol Biol* **321**: 1-6.
- Ladbury, J.E., R. Wynn, H.W. Hellinga and J.M. Sturtevant (1993). *Biochemistry* **32**(29): 7526-30.
- Lehmann, M., L. Pasamontes, S.F. Lassen and M. Wyss (2000). *Biochim Biophys Acta* **1543**(2): 408-415.

3. Relación entre estabilidad de proteínas, evolución y estructura...

Lumb, K.J. and P.S. Kim (1995). *Biochemistry* **34**(27): 8642-8.

Makhatadze, G.I. and P.L. Privalov (1995). *Adv Prote Chem* **47**: 307-425.

Munoz, V. and J.M. Sanchez-Ruiz (2004). *Proc Natl Acad Sci U S A* **101**(51): 17646-51.

Robic, S., M. Guzman-Casado, J.M. Sanchez-Ruiz and S. Marqusee (2003). *Proc Natl Acad Sci U S A* **100**(20): 11345-9.

Sanchez-Ruiz, J.M. (1995). New York, Plenum Press.

Schellman, J.A. (1987). *Annu Rev Biophys Chem* **16**: 115-137.

Shortle, D. (2003). *Protein Sci* **12**(6): 1298-1302.

Shortle, D. and Ackerman (2001). *Science* **293**: 487-489.

Steipe, B., B. Schiller, A. Plückthun and S. Steinbacher (1994). *J Mol Biol* **240**: 182-192.

**4. Patrón de estabilidad de mutaciones
hidrofóbicas en proteínas que refleja
optimización estructural evolutiva**

4 PATRÓN DE ESTABILIDAD DE MUTACIONES HIDROFÓBICAS EN PROTEÍNAS QUE REFLEJA OPTIMIZACIÓN ESTRUCTURAL EVOLUTIVA

Con el desarrollo de las técnicas de ingeniería genética, se han llevado a cabo numerosos estudios físico-químicos sobre el efecto de mutaciones en proteínas. La interpretación de dichos estudios en términos de las fuerzas físicas responsables de la estructura de una proteína nativa ha proporcionado una base para el diseño de proteínas. Sin embargo, muchas características de proteínas naturales se espera que sean el resultado de la selección natural, lo que podría además incidir en la estabilidad. De hecho, algunas veces se ha sugerido que la estabilidad termodinámica marginal de las proteínas existentes en la naturaleza (valores experimentales de energía libre de desnaturalización del orden de algunas decenas de kJ/mol a temperatura fisiológica) se ha seleccionado durante la evolución debido a ventajas potenciales (Godoy-Ruiz *et al.*, 2004) asociadas a el grado de flexibilidad quizás requerido por función, para facilitar la degradación y ayudar a evitar que la proteína quede atrapada en estados plegados incorrectos durante el plegamiento (Sanchez-Ruiz, 1995). Además, determinadas interacciones desestabilizantes pueden ser requeridas por razones funcionales o para garantizar una estructura única (Lumb y Kim, 1995; Knappenberger, 2002).

Más sorprendentemente, trabajos recientes han propuesto la existencia de una relación entre evolución de proteínas y estabilidad de proteínas, incluso a nivel de los ambientes de residuos individuales. De hecho, en el Capítulo 3 de esta memoria (publicado en Godoy-Ruiz *et al.*, 2004) se describe un estudio exhaustivo de los efectos que tienen las mutaciones individuales de ácidos carboxílicos sobre la estabilidad de *Escherichia coli*. Hemos demostrado que todas las mutaciones de glutamato (E) → aspartato (D) realizadas sobre la forma silvestre (WT) de tiorredoxina de *E. coli* fueron desestabilizantes, así como la mayoría de las mutaciones de D → E. Como ya explicamos, los estudios sobre la estabilidad de todas las mutaciones (sentido de la mutación D → E) correlacionan, de manera robusta, con la frecuencia de aparición de dichos residuos en un alineamiento de secuencias (aproximadamente 500 secuencias) realizado con el programa BLAST. Estos resultados indican que los ambientes de residuos cargados en las superficies de proteínas pueden estar optimizados para

4. Un patrón de estabilidad de mutaciones hidrofóbicas...

crear interacciones estabilizantes con un grado remarcable de especificidad, hasta el punto de discriminar entre glutamato y aspartato.

En este capítulo, se describe el efecto de mutaciones de isoleucina (I) y valina (V) sobre la estabilidad de tiorredoxina de *E. coli* y se comparan los efectos experimentales de la mutación sobre la estabilidad de tiorredoxina, medidos con la frecuencia de aparición de I y V en un alineamiento de secuencias. Estas mutaciones $I \rightarrow V$ y $V \rightarrow I$ se espera que sean conservativas y que tengan efectos muy pequeños sobre la estructura de proteínas (en una región bien empaquetada, una mutación $I \rightarrow V$ debería crear una cavidad muy pequeña y una mutación $V \rightarrow I$ debería originar algo de tensión). De hecho, los residuos I y V pueden intercambiarse fácilmente en el curso de la evolución de proteínas como muestran los coeficientes de I/V en las matrices de PAM (George *et al.*, 1990) y BLOSUM (Hirsch *et al.*, 1971) (véase sección 2.2.7.1 en *Materiales y métodos*).

Además, queremos comprobar si el tipo de optimización evolutiva que hemos encontrado previamente para residuos cargados expuestos ocurre, además, en el interior hidrofóbico de proteínas. Más importante es el hecho de que el análisis de mutaciones hidrofóbicas permitirá de forma natural realizar una interpretación de la optimización evolutiva. La razón es que el ambiente de residuos hidrofóbicos enterrados se puede describir, por lo menos en una primera aproximación, en términos de un único parámetro: la densidad de empaquetamiento hidrofóbico. Ésto contrasta con la descripción del ambiente de residuos cargados, para la cual hay que tener en cuenta necesariamente algunos parámetros, relacionados con las interacciones electrostáticas de larga distancia, la solvatación de átomos cargados, la formación de pares iónicos, puentes de hidrógeno, etc.

De hecho, hemos encontrado que la representación gráfica del efecto de mutaciones hidrofóbicas sobre la estabilidad *versus* la densidad de empaquetamiento hidrofóbico muestra un patrón sorprendentemente simple que claramente refleja la optimización estructural evolutiva. La existencia de dicho patrón sugiere que los efectos de mutaciones sobre la estabilidad se puede interpretar (y predecir potencialmente) en base a “modelos evolutivos”. Por modelo evolutivo consideramos específicamente en este contexto un modelo en el cual ciertas características físicas del ambiente de los residuos se evalúan en función a cómo dichos ambientes se han seleccionado durante la evolución

(concepto opuesto al de “modelo físico” en el cual estas características se derivan de algún tipo de análisis energético de las características estructurales de la proteína).

Para ilustrar este novedoso acercamiento, hemos desarrollado un modelo simple (Godoy-Ruiz *et al.*, 2005) que explica el efecto de mutaciones hidrofóbicas en la estabilidad teniendo en cuenta la densidad de empaquetamiento hidrofóbico alrededor del sitio de la mutación en la estructura nativa, así como el tipo y el grado de optimización evolutiva del ambiente del residuo mutado. Hemos encontrado que dicho modelo tiene exitosamente en cuenta el efecto de mutaciones en las que se origina una cavidad $I \rightarrow V$ y mutaciones que introducen tensión $V \rightarrow I$, sobre la estabilidad de tiorredoxina. Desde un punto de vista más general, nuestro análisis sugiere directrices de cómo se pueden investigar sistemáticamente el papel de la optimización evolutiva en el plegamiento de proteínas y la estabilidad.

4.1 ESTUDIO DEL EFECTO DE MUTACIONES DE RESIDUOS HIDROFÓBICOS SOBRE LA ESTABILIDAD DE TIORREDOXINA.

4.1.1 OBTENCIÓN DE LAS PROTEÍNAS MUTADAS DEL TIPO $I \rightarrow V$ Y $V \rightarrow I$

Se han obtenido, mediante técnicas de ingeniería genética, un total de 14 variantes de tiorredoxina de *E. coli*: nueve proteínas con una mutación simple del tipo $I \rightarrow V$ (I4V, I5V, I23V, I38V, I41V, I45V, I60V, I72V e I75V) y cinco proteínas con una mutación simple del tipo $V \rightarrow I$ (V16I, V25I, V55I, V86I y V91I). Las mutaciones deseadas se diseñaron como se explicado en la sección 2.2.1.2 y se han obtenido mediante mutagénesis dirigida (véase para más detalles la sección 2.2.1.3). Para este tipo de mutaciones se trabajó con el gen de tiorredoxina *WT* insertado en el vector pET30a(+) (sección 2.2.1.1) y las células supercompetentes BL21-DE3 (sección 2.2.2.3). La expresión y purificación de proteína se ha explicado detalladamente en las secciones 2.2.2.5 y 2.2.2.6, respectivamente. La pureza de las proteínas purificadas se evaluó por electroforesis (PAGE-SDS; véase sección 2.2.2.6).

4.1.2 ESTUDIO DE LA ESTABILIDAD TERMODINÁMICA DE LAS FORMAS MUTANTES DE TIORREDOXINA MEDIANTE CALORIMETRÍA DIFERENCIAL DE BARRIDO

Hemos caracterizado el efecto sobre la estabilidad de tiorredoxina silvestre (*WT*) de todas las posibles mutaciones $I \rightarrow V$ (residuo *WT* =I) y $V \rightarrow I$ (residuo *WT* =V). Los experimentos se han llevado a cabo a una velocidad de barrido de 1.5 K/min en el intervalo de temperatura de 10-115 °C (véanse detalles de la técnica y análisis de los datos en las secciones 2.2.4.2 y 2.2.4.3, respectivamente). Las disoluciones de proteínas utilizadas en los experimentos de calorimetría se han dializado en HEPES 5 mM, pH 7 y su concentración se ha determinado espectrofotométricamente según se explicó en la sección 2.2.3.1. Para todas las variantes de tiorredoxina obtenidas, se han realizado varios experimentos de calorimetría a concentraciones de proteína de 0.5 mg/mL y menores, ya que tiorredoxina de *E. coli* presenta efecto de la concentración sobre la temperatura de desnaturalización cuando se trabaja a altas concentraciones de proteína (Ladbury *et al.*, 1993); en este rango de concentraciones, no se han observado efectos de concentración sobre la desnaturalización. Además, la desnaturalización de todas las variantes de tiorredoxina estudiadas fue altamente reversible (véase un termograma ilustrativo en la Figura 2.6, sección 2.2.4.2 de Materiales y métodos) y los ajustes de los perfiles de capacidad calorífica al modelo de dos estados en equilibrio, han sido excelentes (véanse en las Figuras 4.1-2 los ajustes realizados y la sección 2.2.4.3 para detalles sobre el modelo de equilibrio de dos estado) y similares a los obtenidos previamente para tiorredoxina *WT* (Georgescu *et al.*, 2001).

Los efectos de las mutaciones sobre la estabilidad de tiorredoxina (valores de $\Delta\Delta G$ calculados según la ecuación 3.1; ver Tabla 4.1) se han calculado a partir de los efectos de la mutación sobre la temperatura de desnaturalización (valores de ΔT_m), usando la aproximación de la ecuación de Shellman (ecuación 3.1). Un cálculo mucho más riguroso, basado en la extrapolación de Gibbs-Helmholtz (ecuación 2.22 en *Materiales y métodos*), ha dado como resultado valores de $\Delta\Delta G$ que difieren en 0.1 kJ/mol de aquellos calculados con la ecuación de Schellman (Tabla 4.1). La mayoría de los valores de $\Delta\Delta G$ y ΔT_m dados en este trabajo son el promedio de tres determinaciones independientes, con desviaciones típicas del orden de 0.1 K para ΔT_m .

4. Un patrón de estabilidad de mutaciones hidrofóbicas...

Un estudio de la accesibilidad de las posiciones mutadas al disolvente nos ha mostrado que todas ellas presentaban una baja accesibilidad al disolvente en la estructura nativa, excepto I5 y V91 que están significativamente expuestas (véase panel inferior de Figura 4.3). Experimentalmente hemos comprobado que todas las mutaciones $V \rightarrow I$ son desestabilizantes, así como la mayoría de las mutaciones $I \rightarrow V$ (Figura 4.1-3). Por lo tanto, a pesar de la similitud en la estructura química (V e I difieren sólo en un grupo metilo), los ambientes de la mayoría de las posiciones I/V en tiorredoxina se encuentran altamente optimizadas para interacciones estabilizantes con cada tipo de residuo específico (I o V) presente en la forma *WT*. Este es el mismo tipo de resultado que el descrito para residuos de ácidos carboxílicos (Godoy-Ruiz *et al.*, 2004 y sección 3.1).

Los cálculos de área superficial accesible (ASA) se han hecho usando una modificación del algoritmo de Shrake-Rupley (Shrake y Rupley, 1973). Este algoritmo sitúa aleatoriamente 2000 puntos alrededor de la esfera de Van der Waals que representa cada átomo. Se han usado un radio de 1.4 Å para la sonda del disolvente y los radios de Chotia (Chothia, 1975) para los átomos de la proteína. La accesibilidad de los residuos se calcula como el cociente entre el ASA de la cadena lateral en la estructura nativa y la del tripéptido Gly-X-Gly. Los cálculos se realizaron con el programa ASA-ENG.exe desarrollado en el grupo de investigación de J.M. Sánchez-Ruiz.

Hay que señalar que, en el análisis realizado, se han considerado todas las mutaciones en la dirección $I \rightarrow V$, lo cual quiere decir que se ha invertido el signo asociado a los efectos sobre la estabilidad para las mutaciones $V \rightarrow I$ (con residuo V en tiorredoxina *WT*), correspondiendo así los valores a efectos $I \rightarrow V$. Para clarificar la información se indica en las figuras de mayor relevancia el tipo de residuo (I o V) presente en la forma *WT* en esa posición (Figura 4.1-3). Y como se explicará más adelante, el propósito de considerar todos los efectos de mutaciones en una dirección dada ($I \rightarrow V$) es mostrar la optimización del ambiente. Como ejemplo, un valor elevado y positivo para el efecto de la mutación $\Delta\Delta G_{I \rightarrow V}$ en una posición determinada significará que el ambiente en esa posición está optimizado para V (sobre I), a pesar del residuo específico (I o V) que existe en esa posición en la forma *WT* (un ejemplo específico se da en la sección 4.6).

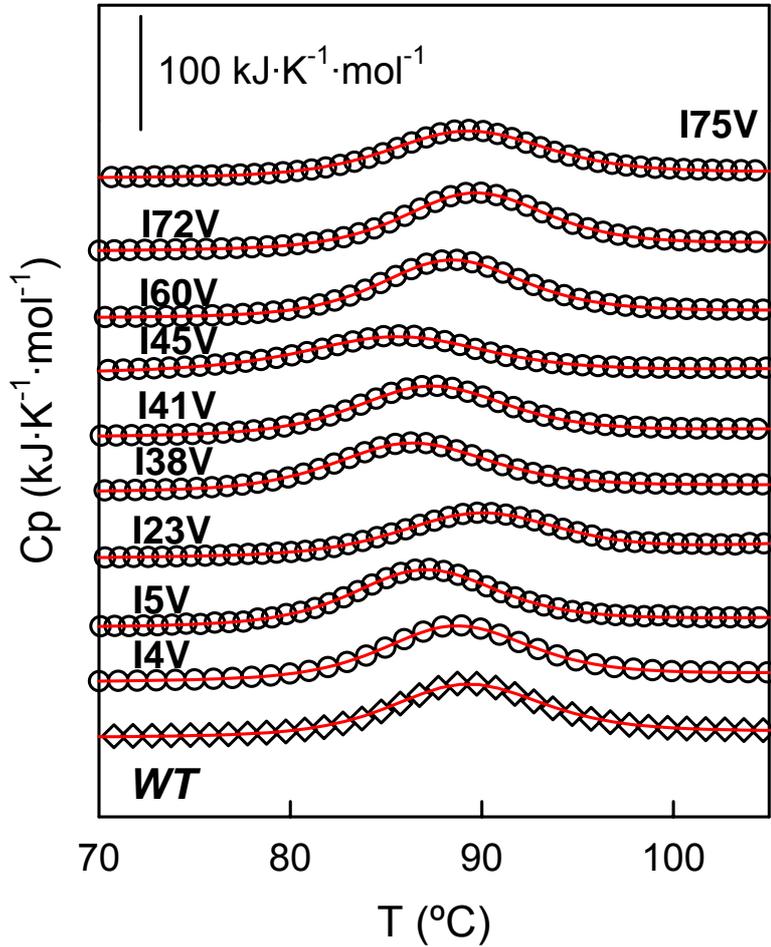


Figura 4.1: Perfiles de capacidad calorífica frente a temperatura de tiorredoxina silvestre (*WT*) (rombos) y los mutantes I4V, I5V, I23V, I38V, I41V, I45V, I60V, I72V y I75V (círculos), obtenidos mediante Calorimetría Diferencial de Barrido (*DSC*). Los experimentos se llevaron a cabo a 1.5 K/min. Las disoluciones de proteína se dializaron en HEPES 5 mM, pH 7. Para cada proteína se han realizado un mínimo de tres experimentos, a concentraciones de proteína de 0.5 mg/ml y menores, comprobando de este modo que no existe efecto de la concentración de proteína sobre la temperatura de desnaturalización. La línea roja representa el ajuste de los datos experimentales al modelo de equilibrio de dos estados. Los parámetros del ajuste se recogen en la Tabla 4.1. Con objeto de obtener una mayor claridad de la figura, los perfiles de las variantes de tiorredoxina han sido desplazados en el eje de capacidades caloríficas.

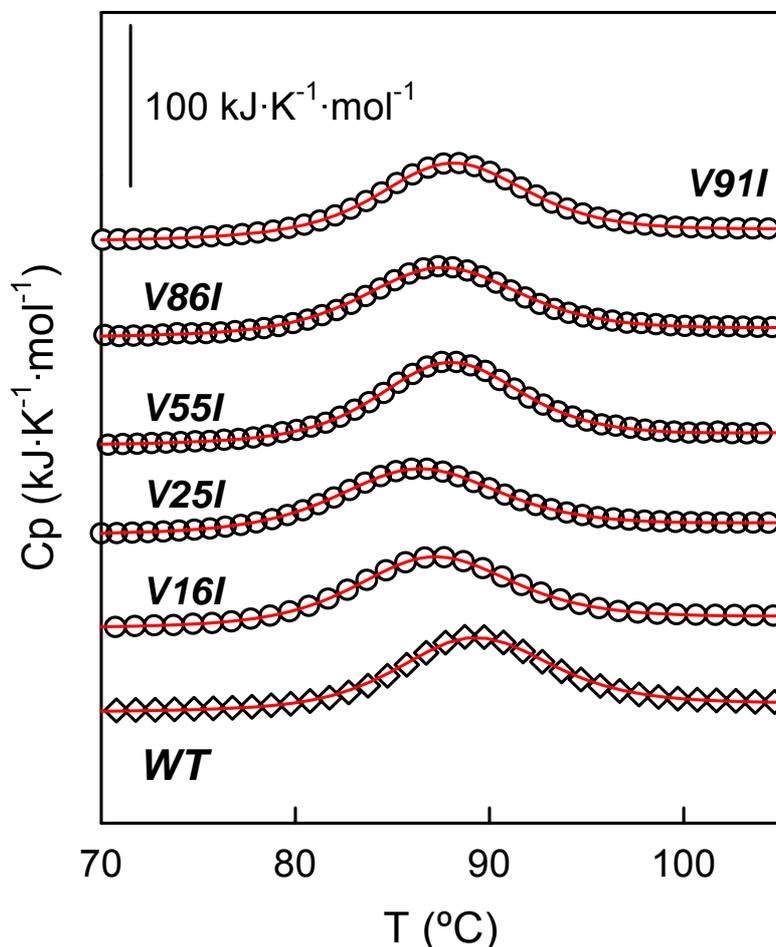


Figura 4.2: Perfiles de capacidad calorífica frente a temperatura de tiorredoxina silvestre (*WT*) (rombos) y los mutantes *V16I*, *V25I*, *V55I*, *V86I* y *V91I* (círculos), obtenidos mediante Calorimetría Diferencial de Barrido (*DSC*). Los experimentos se han llevado a cabo a 1.5 K/min. Las disoluciones de proteína se dializaron en HEPES 5 mM, pH7. Para cada proteína se han realizado un mínimo de tres experimentos, a concentraciones de proteína de 0.5 mg/mL y menores, comprobando de este modo que no existe efecto de la concentración de proteína sobre la temperatura de desnaturalización. La línea roja representa el ajuste de los datos experimentales al modelo de equilibrio de dos estados. Los parámetros del ajuste se recogen en la Tabla 4.1. Con objeto de obtener una mayor claridad de la figura, los perfiles de las variantes de tiorredoxina han sido desplazados en el eje de capacidades caloríficas.

4. Un patrón de estabilidad de mutaciones hidrofóbicas...

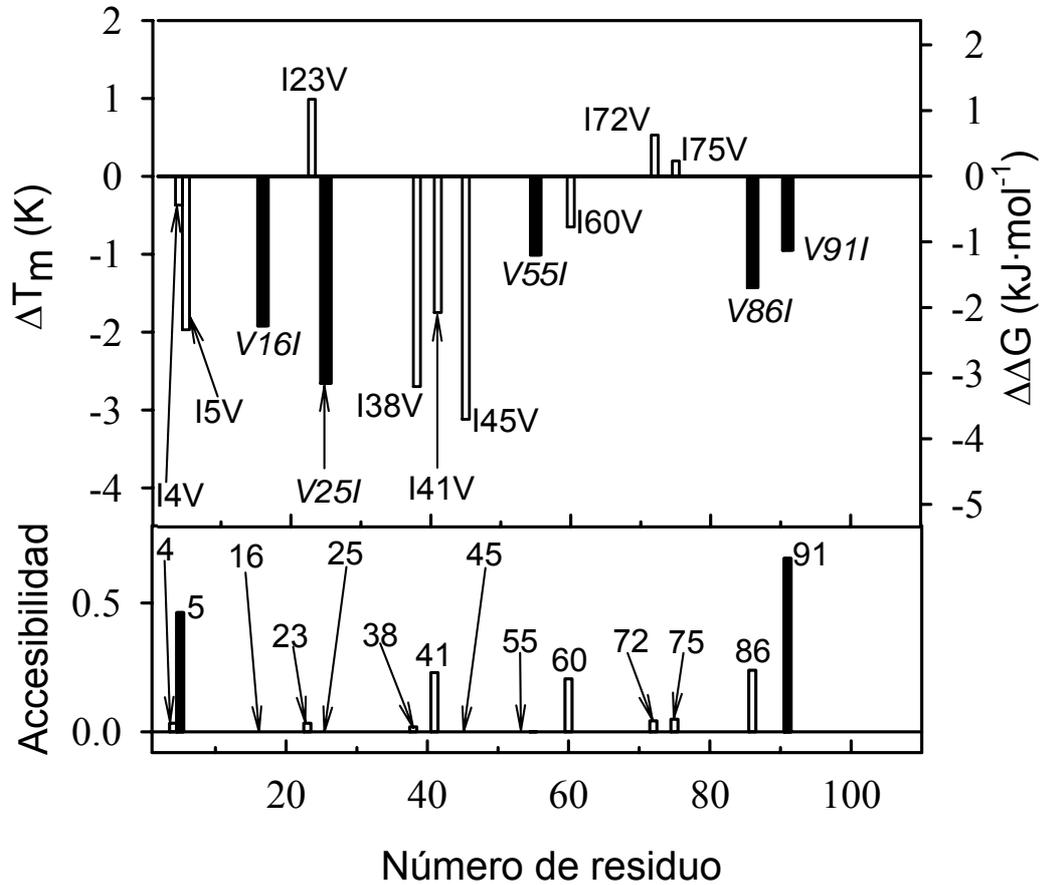


Figura 4.3: (Panel superior) Efecto de las mutaciones isoleucina → valina y valina → isoleucina sobre la estabilidad de tiorredoxina de *E. coli*. Los cambios en estabilidad se miden por el efecto de la mutación sobre la temperatura de desnaturalización (eje de la izquierda) y la energía libre de desnaturalización (eje de la derecha). Las mutaciones I → V y V → I se muestran con barras negras y blancas, respectivamente. Los números sobre las barras indican el residuo mutado en tiorredoxina WT. Panel inferior) Accesibilidad de la cadena lateral al disolvente en la estructura nativa del residuo mutado. Las ASAs se han calculado como se explica en el apartado 4.1.2. Las barras negras indican residuos con una accesibilidad alta (5 y 91).

Tabla 4.1: Temperatura de desnaturalización (T_m), cambio en la temperatura de desnaturalización respecto a tiorredoxina *WT* (ecuación 3.3), cambio en la entalpía de desnaturalización (ΔH_m), cambio en la energía libre de Gibbs (ecuación 3.2) y el efecto de la mutación en el cambio de la energía de Gibbs-Helmholtz, derivados del ajuste al modelo de equilibrio de dos estados de cada una de las proteínas estudiadas. Los valores dados tienen en cuenta el sentido de la mutación $I \rightarrow V$.

Proteína	T_m (°C)	ΔT_m (°C)	ΔH (kJ/mol)	$\Delta\Delta G$ (kJ/mol)	$\Delta\Delta G_{\text{Gibbs-Helmholtz}}$ (kJ/mol)
WT	89.05	-	430.0	-	-
I4V	88.68	-0.37	435.0	-0.4	-0.4
I5V	87.07	-1.98	443.7	-2.4	-2.4
V16I	87.13	1.92	416.7	2.3	2.2
I23V	90.04	0.99	411.2	1.2	1.0
V25I	86.39	-2.69	397.8	3.2	3.1
I38V	86.35	-2.70	411.1	-3.2	-3.1
I41V	87.30	-1.75	421.1	-2.1	-2.1
I45V	85.93	-3.12	394.2	-3.7	-3.6
V55I	88.04	1.01	435.1	1.2	1.2
I60V	88.40	-0.65	448.1	-0.8	-0.7
I72V	89.58	0.53	453.8	0.6	0.7
I75V	89.25	0.20	429.9	0.2	0.2
V86I	87.12	1.43	393.0	1.7	1.7
V91I	88.10	0.95	437.0	1.1	1.1

4.2 CORRELACIÓN DE LOS EFECTOS DE MUTACIONES SOBRE LA ESTABILIDAD CON LA FRECUENCIA DE APARICIÓN DE RESIDUOS EN UN ALINEAMIENTO DE SECUENCIAS: TEST DE LA HIPÓTESIS DE PSEUDOEQUILIBRIO.

Previamente se ha sugerido en la bibliografía (Dokholyan y Shakhnovich, 2001; Shortle, 2003; Godoy-Ruiz *et al.*, 2004) y fundamentalmente en base a trabajos teóricos, que en una escala de tiempo evolutiva, las secuencias se visitan de acuerdo a la energía libre, de una manera análoga a la distribución de energías de Boltzman para un sistema termodinámico en equilibrio. Nos referimos a esta proposición como la “hipótesis de pseudo-equilibrio” o hipótesis de Boltzman. Cuando la aplicamos a mutaciones puntuales neutras (o cuasi-neutras), la hipótesis de pseudo-equilibrio es equivalente a asumir que las mutaciones se fijan en el curso de la evolución con frecuencias que reflejan aproximadamente el efecto de la mutación sobre la estabilidad. Como hemos indicado previamente (Capítulo 3, publicado en Godoy-Ruiz *et al.*, 2004), la hipótesis de pseudo-equilibrio es apoyada por el aceptable nivel de éxito del llamado “concepto de consenso” para estabilización de proteínas (Lehmann *et al.*, 2000) que emplea un análisis estadístico de los alineamientos de secuencias para predecir mutaciones estabilizantes.

En el Capítulo 3 (Godoy-Ruiz *et al.*, 2004) se ha mostrado que el efecto de mutaciones del tipo glutamato → aspartato y aspartato → glutamato sobre la estabilidad de *E. coli*, muestra una correlación robusta con la frecuencia de aparición de estos residuos en un alineamiento de secuencias derivado de una búsqueda realizada con BLAST, usando la secuencia de tiorredoxina de *E. coli* (secuencia *query*) como secuencia a comparar, resultado que es consistente con la hipótesis de pseudo-equilibrio. En este capítulo, mostramos que la misma correlación se mantiene para mutaciones hidrofóbicas del tipo isoleucina → valina y valina → isoleucina. Además, realizamos (como se explica a continuación) una caracterización más detallada que vuelve a demostrar que se cumple la hipótesis de pseudo-equilibrio.

4. Un patrón de estabilidad de mutaciones hidrofóbicas...

Nuestro análisis de la correlación entre efectos de mutaciones sobre la estabilidad y los alineamientos de secuencias se basa en la siguiente ecuación:

$$\Delta\Delta G_{I(E)\rightarrow V(D)} = A + R\theta \cdot \ln\left(\frac{N_{V(D)}}{N_{I(E)}}\right) \quad (4.1)$$

En la ecuación. 4.1, el miembro de la izquierda $\Delta\Delta G_{I\rightarrow V}$ es el efecto sobre la estabilidad de una mutación $I \rightarrow V$ en una posición dada. En el miembro de la derecha, A es una constante (cuyo valor se espera que sea muy próximo a cero) y N_I (N_V) es el número de secuencias con I (V) en la posición dada en los alineamientos de secuencias realizados con BLAST usando la secuencia de tiorredoxina de *E. coli* como la secuencia a comparar (véase el apartado 2.2.7.2 en *Materiales y métodos* para una descripción de las características del alineamiento usado; véase también Tabla 1 de Apéndice III para conocer los números de las frecuencias de todos los aminoácidos en las posiciones estudiadas). El término $R\theta \cdot \ln(N_V/N_I)$ es una función energía-libre que refleja la preferencia relativa de I *versus* V en todas las posiciones en los alineamientos de secuencias, y la temperatura θ puede interpretarse como una medida de la presión evolutiva sobre la estabilidad (Dokholyan y Shakhnovich, 2001). La ecuación 4.1 está obviamente sugerida por la hipótesis de pseudo-equilibrio (Godoy-Ruiz *et al.*, 2004, 2005); ésto significa que de acuerdo con la ecuación 4.1, si el ambiente de una posición determinada está fuertemente optimizado para V y consecuentemente la mutación de $I \rightarrow V$ es estabilizante, durante la evolución de proteínas V se fijará en esa posición más a menudo que I y como resultado la frecuencia de aparición de V en los alineamientos de secuencias será mayor. Del mismo modo, si el ambiente está fuertemente optimizado para I, podemos esperar que la mutación $I \rightarrow V$ sea desestabilizante y la frecuencia de aparición en los alineamientos de secuencias sea mayor para I (*versus* V). Finalmente, hay que destacar que hemos aplicado la ecuación 4.1, a las mutaciones $I \rightarrow V$ y nuevamente a las mutaciones de ácidos carboxílicos (véase Capítulo 3; Godoy-Ruiz *et al.*, 2004). En este último caso, el miembro izquierdo de la ecuación es el efecto de mutaciones en el sentido $E \rightarrow D$ sobre la estabilidad ($\Delta\Delta G_{E\rightarrow D}$) y los números de secuencias en el miembro de la

4. Un patrón de estabilidad de mutaciones hidrofóbicas...

derecha se calculan para los residuos D y E en la posición dada (N_D y N_E), tal y como indican los subíndices entre paréntesis en la ecuación 4.1.

La representación gráfica de $\Delta\Delta G_{I \rightarrow V}(\Delta\Delta G_{E \rightarrow D})$ versus $\ln(N_V/N_I)$ ($\ln(N_D/N_E)$) de las Figuras 4.5, 4.5, 4.6 y 4.7 muestra una correlación significativa entre los efectos de las mutaciones sobre la estabilidad y una función simple de las frecuencias relativas de aparición de residuos en un alineamiento de secuencias; hay solamente un residuo (mutación I \rightarrow V en posición 75) que se va de la correlación de los 27 datos de mutaciones. Cuando se usa para este cálculo una similaridad de 0.3, se están considerando un total de 247 secuencias (Figura 4.5), el coeficiente de correlación correspondiente es $r=0.84$ y la significancia estadística (probabilidad de que la correlación observada ocurra al azar) es $p=6 \cdot 10^{-8}$. El hecho de que los valores de frecuencias calculados para I75V se salgan de la correlación puede deberse a que I75V se encuentra en contacto con el puente disulfuro C32-C35 existente en tiorredoxina y por tanto, su elevada conservación podría estar relacionada con razones funcionales (más que de estabilidad). Por esta razón, no se ha tenido en cuenta para hacer el ajuste lineal por mínimos cuadrados. La correlación resultante es robusta (Figuras 4.4, 4.5, 4.6 y 4.7); es decir, se obtiene una correlación significativa cuando se utilizan secuencias con similaridad con la secuencia *query* mayor de 0.25 (491 secuencias, $r=0.85$, $p=3 \cdot 10^{-8}$, Figura 4.4), 0.3 (247 secuencias, $r=0.84$, $p=6 \cdot 10^{-8}$, Figura 4.5), 0.35 (144 secuencias, $r=0.83$, $p=2 \cdot 10^{-7}$, Figura 4.6) o 0.4 (74 secuencias, $r=0.85$, $p=5 \cdot 10^{-8}$, Figura 4.7). Además, se obtienen correlaciones significativas cuando las mutaciones de ácidos carboxílicos y mutaciones hidrofóbicas se consideran separadamente (por ejemplo, $r=0.89$ y $p=6 \cdot 10^{-5}$ para mutaciones de ácidos carboxílicos, y $r=0.80$ y $p=1 \cdot 10^{-3}$ para mutaciones hidrofóbicas, usando las 491 secuencias de similaridad con la secuencia *query* mayor de 0.25).

Como ejemplo, la ordenada en el origen en las Figuras 4.5 es cercana a cero (-0.69 ± 0.25 kJ/mol). La pendiente tiene unidades de energía y se puede considerar como $R\theta$, donde el valor de la “temperatura” θ podría ser interpretada como una medida de la presión evolutiva sobre la estabilidad o quizás, a ciertas características físicas asociadas de alguna forma con la estabilidad (véase el apartado 4.3). Esta última posibilidad (ver más adelante) parece estar apoyada por el valor encontrado para θ (180 ± 23 K), el cual es significativamente menor que la

4. Un patrón de estabilidad de mutaciones hidrofóbicas...

temperatura fisiológica. El anterior valor de θ se ha calculado utilizando secuencias que tienen con la secuencia *query* una similaridad mayor de 0.3 (Figura 4.5). Sin embargo, se han obtenido resultados cualitativamente similares usando otros cortes de similaridad: 0.25 ($\theta=211\pm 26$ K, 491 secuencias, Figura 4.4), 0.35 ($\theta=150\pm 21$ K, 144 secuencias, Figura 4.6) y 0.4 ($\theta=126\pm 16$ K, 74 secuencias, Figura 4.7). Valores recogidos en Tabla 4.2.

Para el cálculo de las frecuencias del número de apariciones de un residuo en un alineamiento de secuencias (N) se han utilizado los programas FRC-2.exe y Sec.exe (desarrollados en el grupo de investigación de J.M. Sánchez-Ruiz); estos programas utilizan el fichero del alineamiento de secuencias llamado Trx25.aln (parámetros de entrada: 491 secuencias; 1 tanda; corte 1). El cálculo de los valores de p se ha realizado con el programa *scriptp.m* creado en MATLAB (Apéndice I, Programa 9).

4. Un patrón de estabilidad de mutaciones hidrofóbicas...

Figura 4.4: Correlación entre los efectos de mutaciones $I \rightarrow V$ sobre la estabilidad de tiorredoxina (símbolos negros) y la frecuencia de aparición de residuos I y V en un alineamiento de secuencias. Los cambios en estabilidad se miden por el efecto de la mutación sobre la temperatura de desnaturalización (eje de la derecha) y sobre el cambio de la energía libre de Gibbs de desnaturalización (eje de la izquierda). Todos los cambios de estabilidad dados describen el efecto de las mutaciones $I \rightarrow V$. Los círculos negros corresponden a las mutaciones $I \rightarrow V$ originales que existen en tiorredoxina *WT*. Los cuadrados negros corresponden a las mutaciones originales $V \rightarrow I$, para las cuales el signo del cambio en estabilidad ha sido invertido, así que los valores mostrados corresponden a efectos $I \rightarrow V$ ($\Delta T_m = T_m(WT, V) - T_m(\text{variante}, I)$). Los números junto a los símbolos indican la posición mutada. Además incluimos (símbolos blancos) los efectos de las mutaciones de ácidos carboxílicos (sentido de la mutación $E \rightarrow D$) sobre la estabilidad de tiorredoxina (Godoy-Ruiz *et al.*, 2004). Los números de la esquina superior izquierda de la gráfica indican el rango de similaridad de secuencias utilizado en el cálculo del número de residuos: N_I , N_V , N_E y N_D .

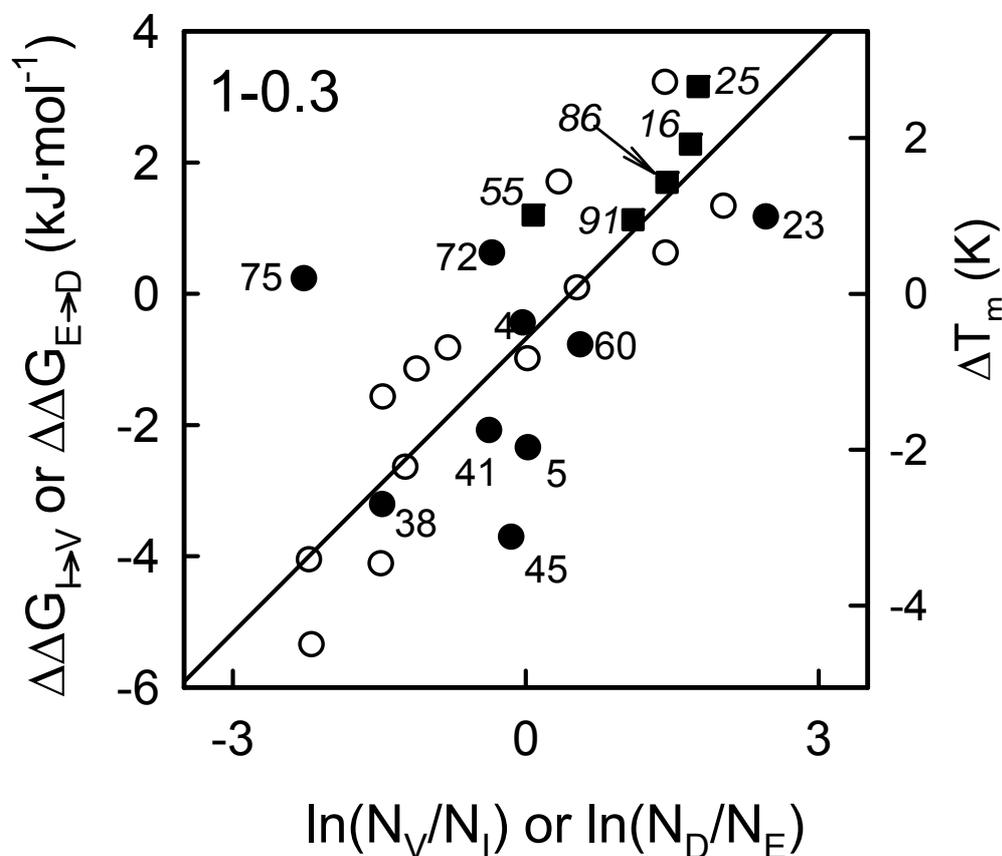


Figura 4.5: Correlación entre los efectos de mutaciones $I \rightarrow V$ sobre la estabilidad de tiorredoxina (símbolos negros) y la frecuencia de aparición de residuos I y V en un alineamiento de secuencias. Los cambios en estabilidad se miden por el efecto de la mutación sobre la temperatura de desnaturalización (eje de la derecha) y sobre el cambio de la energía libre de Gibbs de desnaturalización (eje de la izquierda). Todos los cambios de estabilidad dados describen el efecto de las mutaciones $I \rightarrow V$. Los círculos negros corresponden a las mutaciones $I \rightarrow V$ originales que existen en tiorredoxina *WT*. Los cuadrados negros corresponden a las mutaciones originales $V \rightarrow I$, para las cuales el signo del cambio en estabilidad ha sido invertido, así que los valores mostrados corresponden a efectos $I \rightarrow V$ ($\Delta T_m = T_m(WT, V) - T_m(\text{variante}, I)$). Los números junto a los símbolos indican la posición mutada. Además incluimos (símbolos blancos) los efectos de las mutaciones de ácidos carboxílicos (sentido de la mutación $E \rightarrow D$) sobre la estabilidad de tiorredoxina (Godoy-Ruiz *et al.*, 2004). Los números de la esquina superior izquierda de la gráfica indican el rango de similaridad de secuencias utilizado en el cálculo del número de residuos: N_I , N_V , N_E y N_D .

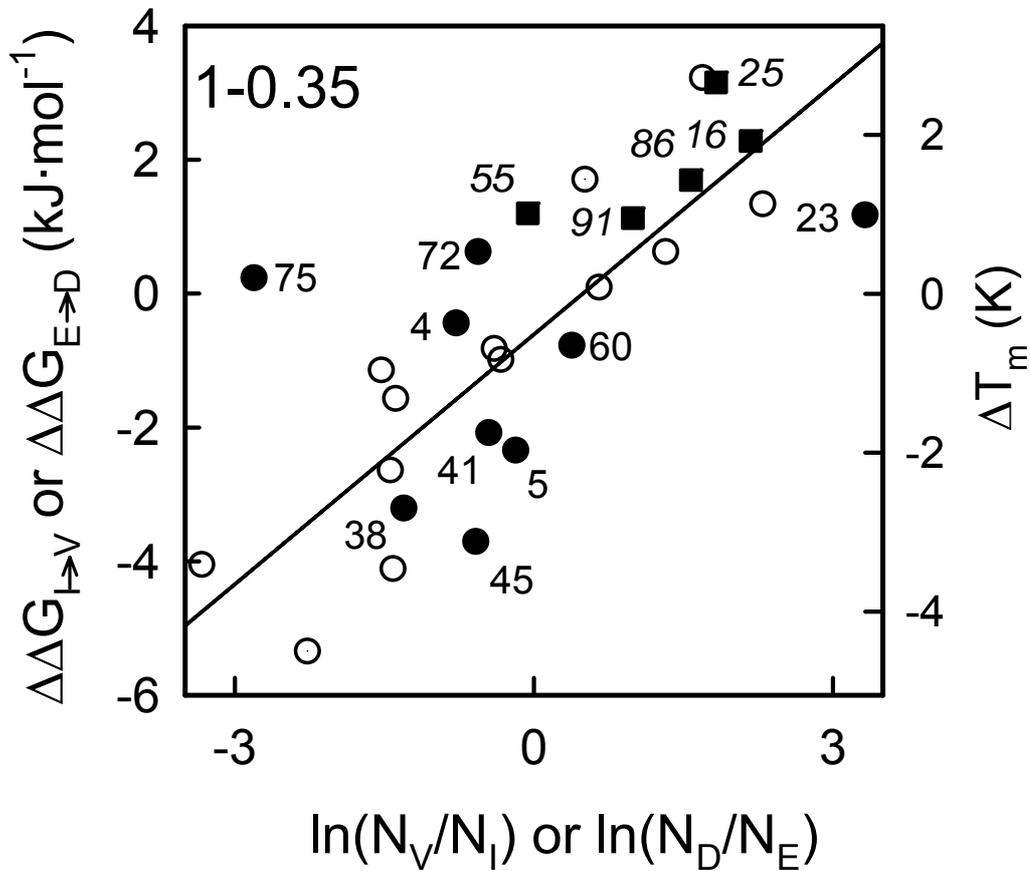


Figura 4.6: Correlación entre los efectos de mutaciones $I \rightarrow V$ sobre la estabilidad de tiorredoxina (símbolos negros) y la frecuencia de aparición de residuos I y V en un alineamiento de secuencias. Los cambios en estabilidad se miden por el efecto de la mutación sobre la temperatura de desnaturalización (eje de la derecha) y sobre el cambio de la energía libre de Gibbs de desnaturalización (eje de la izquierda). Todos los cambios de estabilidad dados describen el efecto de las mutaciones $I \rightarrow V$. Los círculos negros corresponden a las mutaciones $I \rightarrow V$ originales que existen en tiorredoxina *WT*. Los cuadrados negros corresponden a las mutaciones originales $V \rightarrow I$, para las cuales el signo del cambio en estabilidad ha sido invertido, así que los valores mostrados corresponden a efectos $I \rightarrow V$ ($\Delta T_m = T_m(WT, V) - T_m(\text{variante}, I)$). Los números junto a los símbolos indican la posición mutada. Además incluimos (símbolos blancos) los efectos de las mutaciones de ácidos carboxílicos (sentido de la mutación $E \rightarrow D$) sobre la estabilidad de tiorredoxina (Godoy-Ruiz *et al.*, 2004). Los números de la esquina superior izquierda de la gráfica indican el rango de similaridad de secuencias utilizado en el cálculo del número de residuos: N_I , N_V , N_E y N_D .

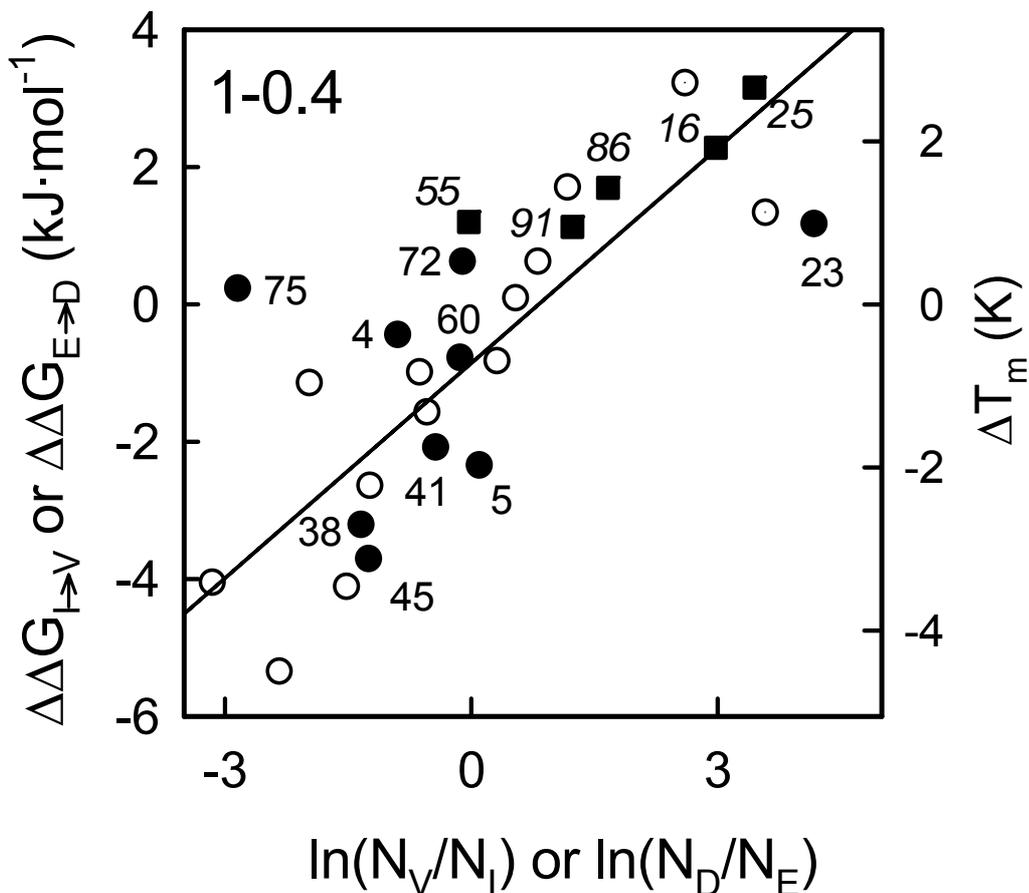


Figura 4.7: Correlación entre los efectos de mutaciones $I \rightarrow V$ sobre la estabilidad de tiorredoxina (símbolos negros) y la frecuencia de aparición de residuos I y V en un alineamiento de secuencias. Los cambios en estabilidad se miden por el efecto de la mutación sobre la temperatura de desnaturalización (eje de la derecha) y sobre el cambio de la energía libre de Gibbs de desnaturalización (eje de la izquierda). Todos los cambios de estabilidad dados describen el efecto de las mutaciones $I \rightarrow V$. Los círculos negros corresponden a las mutaciones $I \rightarrow V$ originales que existen en tiorredoxina *WT*. Los cuadrados negros corresponden a las mutaciones originales $V \rightarrow I$, para las cuales el signo del cambio en estabilidad ha sido invertido, así que los valores mostrados corresponden a efectos $I \rightarrow V$ ($\Delta T_m = T_m(\text{WT}, V) - T_m(\text{variante}, I)$). Los números junto a los símbolos indican la posición mutada. Además incluimos (símbolos blancos) los efectos de las mutaciones de ácidos carboxílicos (sentido de la mutación $E \rightarrow D$) sobre la estabilidad de tiorredoxina (Godoy-Ruiz *et al.*, 2004). Los números de la esquina superior izquierda de la gráfica indican el rango de similaridad de secuencias utilizado en el cálculo del número de residuos: N_I , N_V , N_E y N_D .

Tabla 4.2: Valores resultantes del ajuste lineal por mínimos cuadrados realizado a las correlaciones de $\Delta\Delta G_{I(E)\rightarrow V(D)}$ versus $R\theta\ln(N_{V(D)}/N_{I(E)})$.

Similaridad	A (kJ/Kmol)	RT (kJ/Kmol)	r^2	θ (K)	p
0.25	-0.72±0.25	1.76±0.22	0.725	211±28	3·10 ⁻⁸
0.30	-0.69±0.25	1.49±0.20	0.706	180±23	7·10 ⁻⁸
0.35	-0.61±0.26	1.24±0.17	0.672	150±21	2·10 ⁻⁷
0.40	-0.86±0.25	1.04±0.13	0.723	126±16	5·10 ⁻⁸

Los cálculos que se describen a continuación se han llevado a cabo usando los números de apariciones de los residuos I, V, D y E en el alineamiento de secuencias (números N_I , N_V , N_D y N_E). En el Capítulo 3, usamos números de similaridad pesados en el alineamiento, pero como dicho peso no afecta significativamente el resultado del análisis, aquí se ha omitido. Por otro lado, hemos explorado en este trabajo el efecto de realizar una “pseudocuenta” de las frecuencias de los residuos en cuestión, sobre las correlaciones estudiadas. Este tipo de pseudocuenta ya se ha utilizado previamente para corregir las frecuencias brutas de aminoácidos cuando el número de secuencias en el alineamiento es pequeño y existe el riesgo de que algunas frecuencias brutas lleguen a ser 0, hecho que puede causar problemas fatales en algunos algoritmos (véase el capítulo 6 en Ewens y G.R., 2001). La pseudocuenta utilizada se basa en la estadística Bayesiana, aunque la elección práctica de los valores de la pseudocuenta es a menudo hecha subjetivamente (capítulo 6 en Ewens y G.R., 2001). Aquí,

seguimos el procedimiento de Lawrence *et al.* (1993). Definimos las frecuencias “corregidas” como:

$$q_{ij} = \frac{N_{ij} + b_j}{N - 1 + B} \quad (4.2)$$

donde q_{ij} es la frecuencia corregida para el amino ácido j en la posición i , N_{ij} son las frecuencias brutas del aminoácido j en la posición i , N es el número total de secuencias en los alineamientos, b_j es el valor de la pseudocuenta del aminoácido j y B es la suma de los valores de las pseudocuentas para los 20 aminoácidos ($B = \sum_{j=1}^{20} b_j$). Siguiendo Lawrence *et al.* (1993) se obtienen las pseudocuentas proporcionalmente a las frecuencias base de los aminoácidos utilizando la constante de proporcionalidad \sqrt{N} . Como frecuencias base elegimos las calculadas del alineamiento total (incluyendo todas las secuencias y posiciones).

Una vez calculadas las pseudocuentas para las posiciones que estamos estudiando, usamos las frecuencias corregidas calculadas (ecuación 4.2) en lugar de los números de apariciones, sin corregir, en el análisis basado en la ecuación 4.1. Los resultados obtenidos han sido esencialmente los mismos que los obtenidos previamente usando los números de apariciones de residuos en el alineamiento de secuencias sin calcular las pseudocuentas (Tabla 4.3). Por ejemplo, usando todas las secuencias en el alineamiento hemos obtenido $\theta=216\pm 27$ K, $r=0.85$ y $p=3\cdot 10^{-8}$ (*versus* $\theta=211\pm 26$ K, $r=0.85$, $p=3\cdot 10^{-8}$ usando números de apariciones sin corregir; Figura 4.4) y usando sólo secuencias con similaridad mayor de 0.4 con la secuencia *query* hemos obtenido $\theta=143\pm 19$ K, $r=0.83$ $p=1\cdot 10^{-7}$ (*versus* $\theta=126\pm 16$, $r=0.85$ y $p=5\cdot 10^{-8}$ que se obtienen usando números de apariciones sin corregir; Figura 4.7). Este resultado no nos sorprende dado el alto número de secuencias en nuestro alineamiento (Figura 4.8).

El cálculo de las frecuencias utilizando pseudocuentas de aminoácidos se ha realizado con el programa PSCOUNTS.exe (desarrollado en grupo de investigación de J.M Sánchez-Ruiz). Este programa utiliza el fichero Trx25.aln como entrada de datos.

Tabla 4.3: Valores resultantes del ajuste lineal por mínimos cuadrados realizado a las correlaciones de $\Delta\Delta G_{I(E)\rightarrow V(D)}$ versus $R\theta \ln(q_{V(D)}/q_{I(E)})$.

Similaridad	A (kJ/Kmol)	RT (kJ/Kmol)	r^2	θ (K)	p
0.25	-0.73±0.25	1.79±0.23	0.724	216±27	$3\cdot 10^{-8}$
0.30	-0.69±0.25	1.55±0.20	0.711	187±24	$6\cdot 10^{-8}$
0.35	-0.63±0.25	1.36±0.18	0.710	163±22	$1\cdot 10^{-7}$
0.40	-0.86±0.26	1.19±0.16	0.710	143±119	$1\cdot 10^{-7}$

4. Un patrón de estabilidad de mutaciones hidrofóbicas...

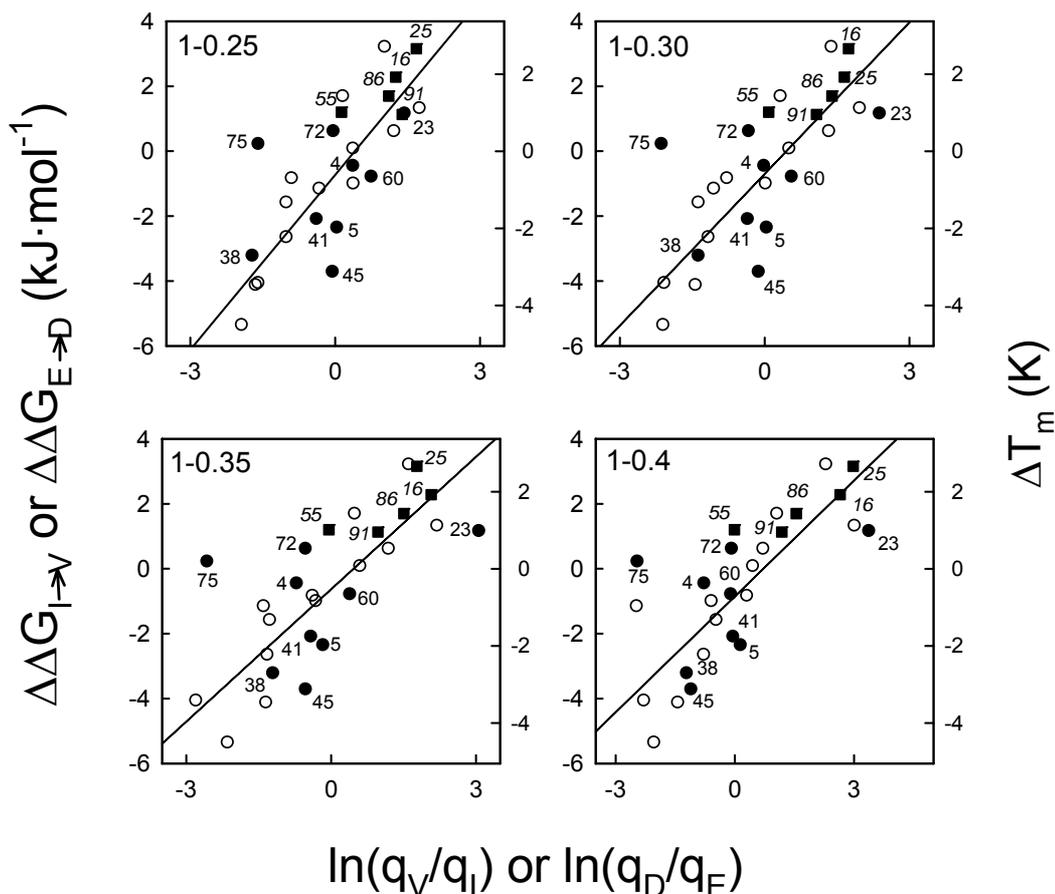


Figura 4.8: Correlación entre los efectos de mutaciones $I \rightarrow V$ sobre la estabilidad de tiorredoxina (símbolos negros) y la frecuencia de aparición, utilizando pseudocuentas, de residuos I y V en un alineamiento de secuencias. Los cambios en estabilidad se miden por el efecto de la mutación sobre la temperatura de desnaturalización (eje de la derecha) y sobre el cambio de la energía libre de Gibbs de desnaturalización (eje de la izquierda). Todos los cambios de estabilidad dados describen el efecto de las mutaciones $I \rightarrow V$. Los círculos negros corresponden a las mutaciones $I \rightarrow V$ originales que existen en tiorredoxina *WT*. Los cuadrados negros corresponden a las mutaciones originales $V \rightarrow I$, para las cuales el signo del cambio en estabilidad ha sido invertido, así que los valores mostrados corresponden a efectos $I \rightarrow V$ ($\Delta T_m = T_m(\text{WT}, V) - T_m(\text{variante}, I)$). Los números junto a los símbolos indican la posición mutada. Además incluimos (símbolos blancos) los efectos de las mutaciones de ácidos carboxílicos (sentido de la mutación $E \rightarrow D$) sobre la estabilidad de tiorredoxina (Godoy-Ruiz *et al.*, 2004). Los números de la esquina superior izquierda de la gráfica indican el rango de similitud de secuencias utilizado en el cálculo de las pseudocuentas: q_I , q_V , q_E y q_D .

4. Un patrón de estabilidad de mutaciones hidrofóbicas...

Con objeto de comprobar más profundamente la aplicabilidad de la hipótesis de pseudo-equilibrio, hemos llevado a cabo el análisis de la correlación basado en la ecuación 4.1 para diferentes subgrupos de alineamientos elegidos de acuerdo a los siguientes criterios:

- 1) Agrupación de secuencias de acuerdo a su similaridad.
- 2) Agrupación de secuencias según su clasificación taxonómica.

Teniendo en cuenta el primer criterio, hemos creado una matriz de distancias para las 491 secuencias del alineamiento, usando como “distancias” los valores de 1 menos la similaridad (distancias de Dayhoff; Dayhoff *et al.*, 1978), y hemos realizado un análisis de grupos (*clusters*) llamado *K-means* basado en esa matriz. Para ello se ha usado el programa *K-means* incluido en MATLAB, eligiendo la opción de cálculo con distancias euclídeas al cuadrado y realizando 10 réplicas para cada análisis (Programa 10 en Apéndice I). La función *K-means* origina grupos de secuencias en base a la similaridad de dichas secuencias entre si, para ello utiliza un algoritmo iterativo que minimiza la suma de las distancias existentes entre cada una de las secuencias. Como resultado se obtiene un conjunto de *clusters* de secuencias que se encuentran separados unos de otros lo mejor posible. Para encontrar el número correcto de *clusters* de secuencias que se pueden formar en nuestro alineamiento, hemos realizado el análisis de *K-means* considerando diferentes números de *clusters* de secuencias, desde 2 hasta 9 (Figuras 4.9 hasta 4.13). Para visualizar dichos *clusters* de secuencias hemos realizado una representación gráfica de su silueta, usando la función *Silhouette*, también incluida en MATLAB (Programa 10 en Apéndice I). Si el número de *clusters* de secuencias es adecuado, éstos se encontrarán bien separados y poseerán un valor alto de la función *Silhouette*.

Como ilustración de los resultados obtenidos, en la Figura 4.11 se muestra la el análisis de *K-means* para 5 *cluster* de secuencias. Hay que señalar que en las gráficas de $\Delta\Delta G_{I \rightarrow V}$ (or $\Delta\Delta G_{E \rightarrow D}$) *versus* $\ln(N_V/N_I)$ (or $\ln(N_D/N_E)$) se han obtenido correlaciones significativas ($p < 10^{-2}$) en todos los casos y para 3 de los 5 *clusters*, p es del orden de 10^{-5} . Este resultado general es bastante robusto con respecto al número de grupos especificado en el análisis; aunque para un mayor número de *clusters* algunos de ellos son de tamaño demasiado reducido como para llevar a cabo un análisis de correlación estadística fiable (Figuras 4.12 y 4.13).

4. Un patrón de estabilidad de mutaciones hidrofóbicas...

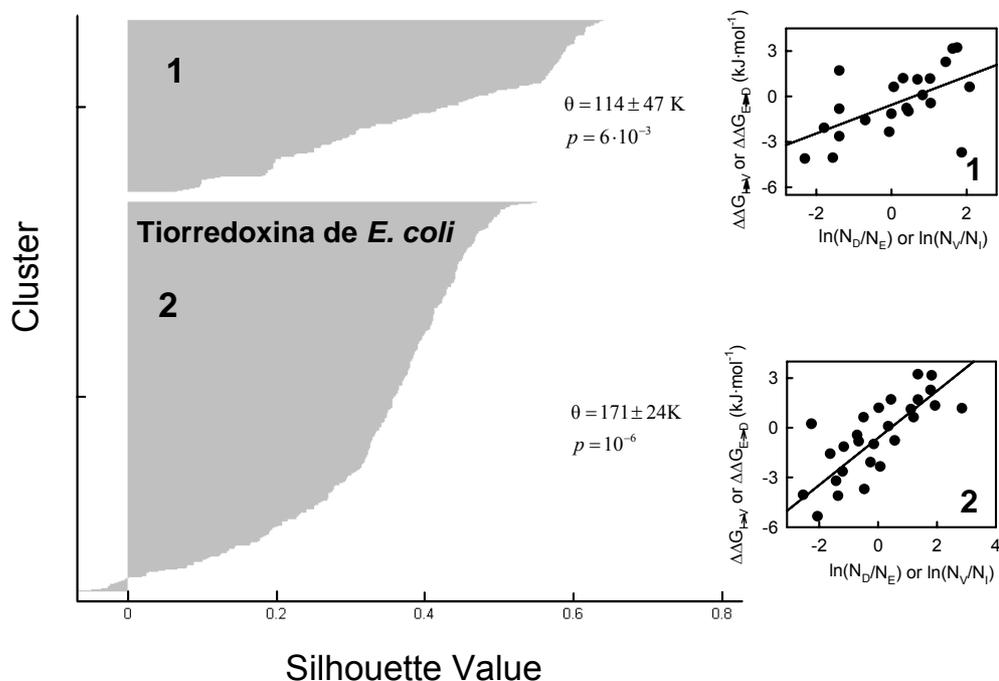


Figura 4.9: Correlación entre los efectos de una mutación sobre la estabilidad de tiorredoxina de *E. coli* y la frecuencia de aparición de los aminoácidos implicados para 2 subconjuntos de secuencias, obtenidos del alineamiento original, realizando un cálculo de *cluster* con *K-means* (programa incluido en MATLAB). La gráfica de mayores dimensiones nos da los valores de las siluetas para las secuencias en cada uno de sus *cluster*. El valor de *Silhouette* para cada secuencia puede variar desde -1 a 1 y es una medida de cómo de similar una secuencia es a secuencias en su propio *cluster* comparada con las secuencias en otros *clusters*. Los números en las gráficas de los efectos de mutación sobre la estabilidad frente al logaritmo del cociente de las apariciones sin corregir, se refieren a cada uno de los *cluster*. Los valores de θ y p (significancia estadística) derivados de los análisis de las 2 correlaciones se muestran junto a los correspondientes *cluster* en la representación de *Silhouette*.

4. Un patrón de estabilidad de mutaciones hidrofóbicas...

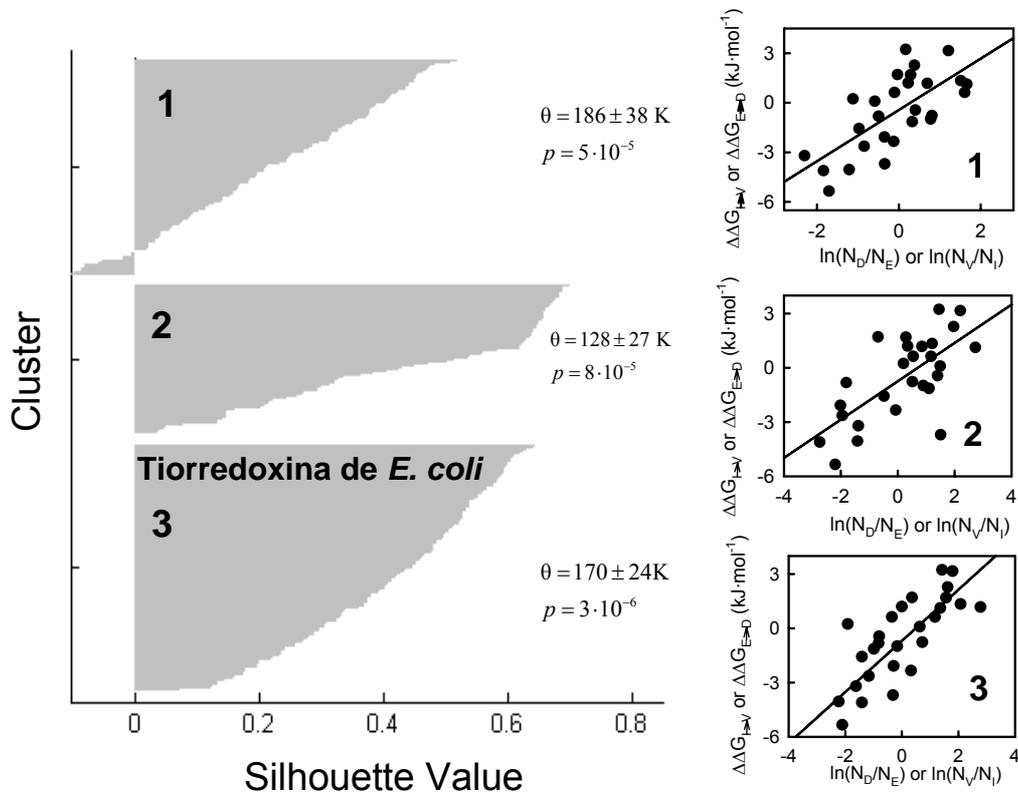


Figura 4.10: Correlación entre los efectos de una mutación sobre la estabilidad de tiorredoxina de *E. coli* y la frecuencia de aparición de los aminoácidos implicados para 3 subconjuntos de secuencias, obtenidos del alineamiento original, realizando un cálculo de *cluster* con *K-means* (programa incluido en MATLAB). La gráfica de mayores dimensiones nos da los valores de las siluetas para las secuencias en cada uno de sus *cluster*. El valor de *Silhouette* para cada secuencia puede variar desde -1 a 1 y es una medida de cómo de similar una secuencia es a secuencias en su propio *cluster* comparada con las secuencias en otros *clusters*. Los números en las gráficas de los efectos de mutación sobre la estabilidad frente al logaritmo del cociente de las apariciones sin corregir, se refieren a cada uno de los *cluster*. Los valores de θ y p (significancia estadística) derivados de los análisis de las 3 correlaciones se muestran junto a los correspondientes *cluster* en la representación de *Silhouette*.

4. Un patrón de estabilidad de mutaciones hidrofóbicas...

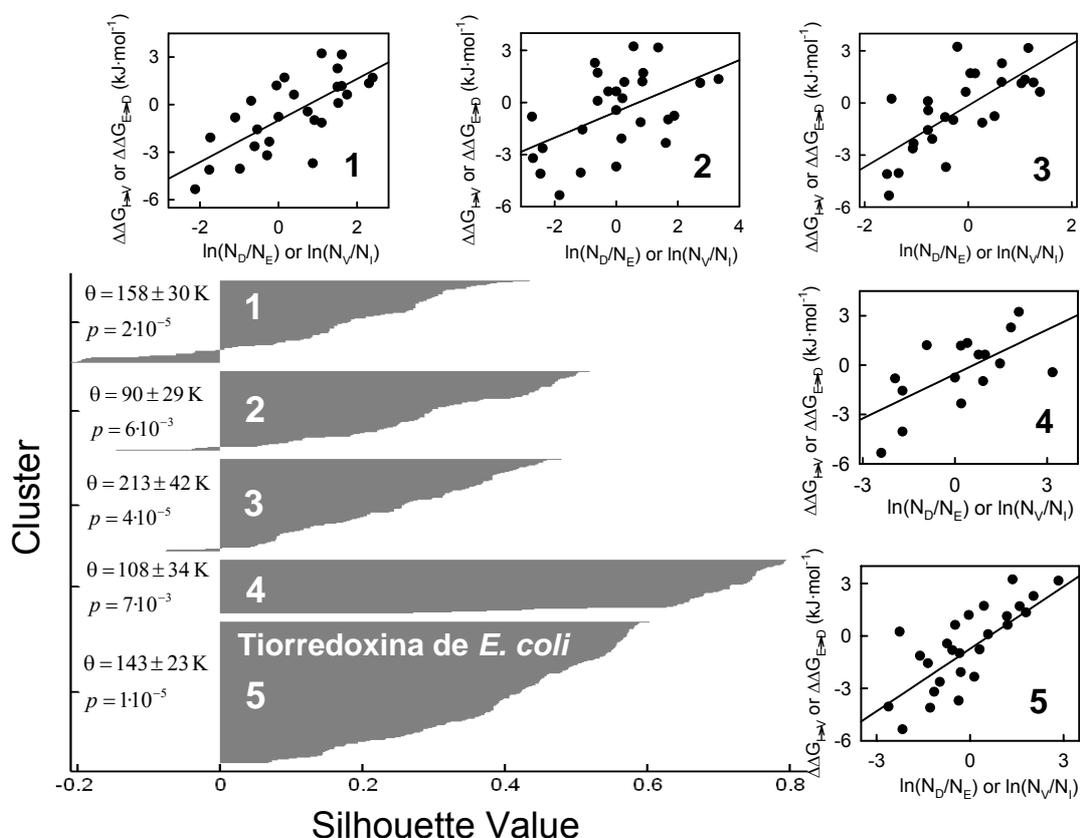


Figura 4.11: Correlación entre los efectos de una mutación sobre la estabilidad de tiorredoxina de *E. coli* y la frecuencia de aparición de los aminoácidos implicados para 5 ° subconjuntos de secuencias, obtenidos del alineamiento original, realizando un cálculo de *cluster* con *K-means* (programa incluido en MATLAB). La gráfica de mayores dimensiones nos da los valores de las siluetas para las secuencias en cada uno de sus *cluster*. El valor de *Silhouette* para cada secuencia puede variar desde -1 a 1 y es una medida de cómo de similar una secuencia es a secuencias en su propio *cluster* comparada con las secuencias en otros *clusters*. Los números en las gráficas de los efectos de mutación sobre la estabilidad frente al logaritmo del cociente de las apariciones sin corregir, se refieren a cada uno de los *cluster*. Los valores de θ y p (significancia estadística) derivados de los análisis de las 5 correlaciones se muestran junto a los correspondientes *cluster* en la representación de *Silhouette*.

4. Un patrón de estabilidad de mutaciones hidrofóbicas...

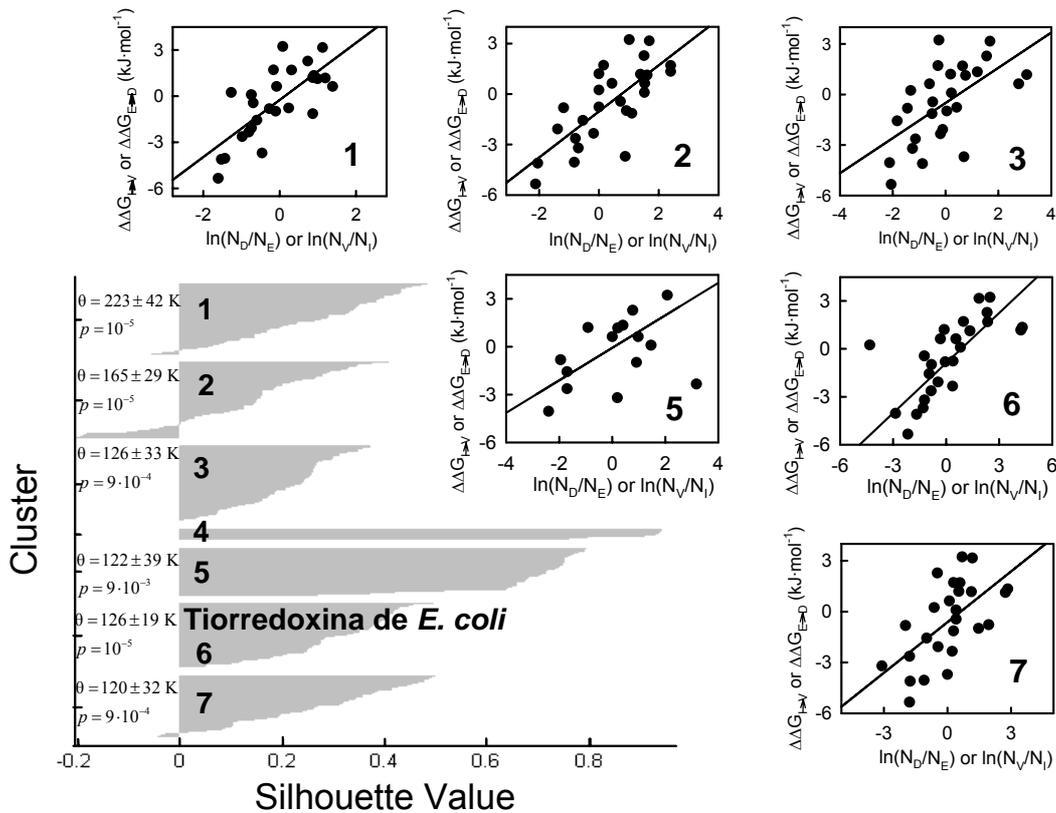


Figura 4.12: Correlación entre los efectos de una mutación sobre la estabilidad de tiorredoxina de *E. coli* y la frecuencia de aparición de los aminoácidos implicados para 7 subconjuntos de secuencias, obtenidos del alineamiento original, realizando un cálculo de *cluster* con *K-means* (programa incluido en MATLAB). La gráfica de mayores dimensiones nos da los valores de las siluetas para las secuencias en cada uno de sus *cluster*. El valor de *Silhouette* para cada secuencia puede variar desde -1 a 1 y es una medida de cómo de similar una secuencia es a secuencias en su propio *cluster* comparada con las secuencias en otros *clusters*. Los números en las gráficas de los efectos de mutación sobre la estabilidad frente al logaritmo del cociente de las apariciones sin corregir, se refieren a cada uno de los *cluster*. Los valores de θ y p (significancia estadística) derivados de los análisis de las 7 correlaciones se muestran junto a los correspondientes *cluster* en la representación de *Silhouette*. La correlación correspondiente al *cluster* 4 no se muestra ya que no posee suficientes secuencias para establecer una correlación estadística fiable.

4. Un patrón de estabilidad de mutaciones hidrofóbicas...

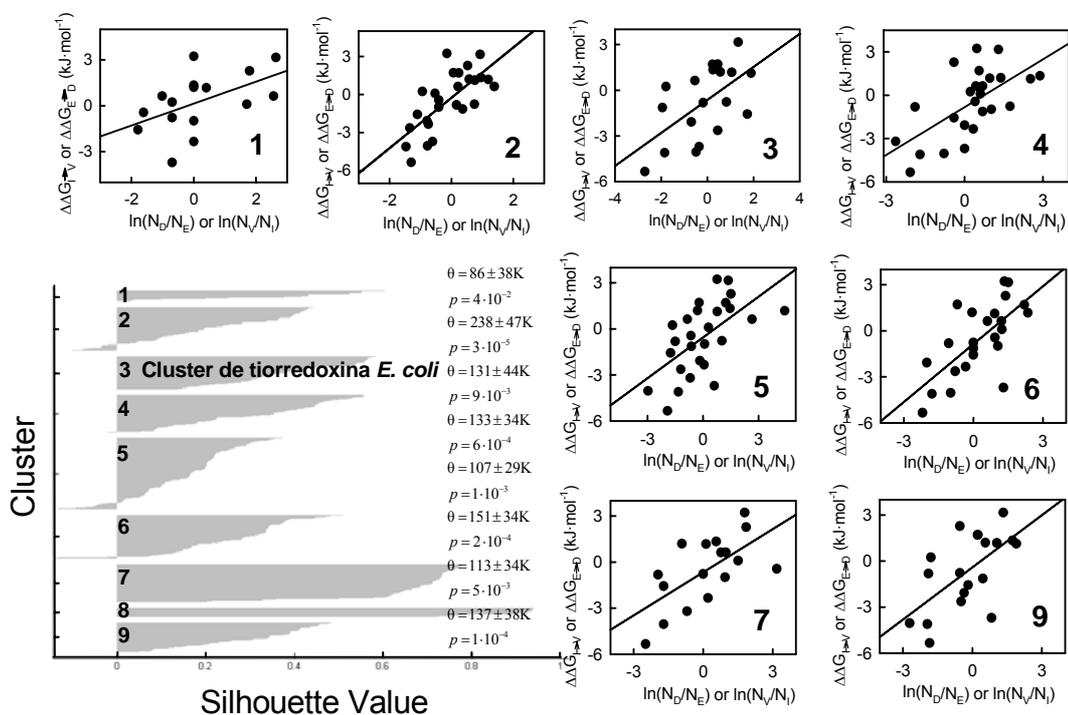


Figura 4.13: Correlación entre los efectos de una mutación sobre la estabilidad de tioredoxina de *E. coli* y la frecuencia de aparición de los aminoácidos implicados para 9 subconjuntos de secuencias, obtenidos del alineamiento original, realizando un cálculo de *cluster* con *K-means* (programa incluido en MATLAB). La gráfica de mayores dimensiones nos da los valores de las siluetas para las secuencias en cada uno de sus *cluster*. El valor de *Silhouette* para cada secuencia puede variar desde -1 a 1 y es una medida de cómo de similar una secuencia es a secuencias en su propio *cluster* comparada con las secuencias en otros *clusters*. Los números en las gráficas de los efectos de mutación sobre la estabilidad frente al logaritmo del cociente de las apariciones sin corregir, se refieren a cada uno de los *cluster*. Los valores de θ y p (significancia estadística) derivados de los análisis de las 9 correlaciones se muestran junto a los correspondientes *cluster* en la representación de *Silhouette*. La correlación correspondiente al *cluster* 8 no se muestra ya que no posee suficientes secuencias para establecer una correlación estadística fiable.

4. Un patrón de estabilidad de mutaciones hidrofóbicas...

Teniendo en cuenta el segundo criterio expuesto anteriormente, hemos realizado la agrupación de secuencias de acuerdo a su clasificación taxonómica. Se han hecho representaciones gráficas de $\Delta\Delta G_{I \rightarrow V}$ (or $\Delta\Delta G_{E \rightarrow D}$) *versus* $\ln(N_V/N_I)$ (or $\ln(N_D/N_E)$) para las secuencias pertenecientes a bacteria (219 secuencias), eucariota (249 secuencias), proteobacteria (101 secuencias) y metazoa (123 secuencias). En los cuatro casos (Figura 4.14), se obtuvieron correlaciones significativas (p del orden de 10^{-5} - 10^{-6}). Las correlaciones para viridiplantae (81 secuencias) y firmicutes (51 secuencias) fueron menos significativas ($p=1 \cdot 10^{-2}$ y $p=7 \cdot 10^{-3}$, respectivamente), aunque esto puede ser una consecuencia estadística asociada al bajo número de secuencias pertenecientes a cada una de estas taxas en nuestro alineamiento.

Los análisis realizados apoyan fuertemente la hipótesis de pseudo-equilibrio, al menos como una descripción de primer orden de la probabilidad de que se dé una mutación durante la evolución de las posiciones estudiadas. Encontramos particularmente relevantes las correlaciones entre los efectos de las mutaciones sobre la estabilidad de tiorredoxina de *E. coli* (de proteobacteria) con el logaritmo del cociente del número de apariciones en el alineamiento de los residuos en cuestión, en los subgrupos de eucariota y metazoa (ver Figura 4.14). Esto sugiere claramente que el tipo de optimización del ambiente para una posición dada puede estar altamente conservado durante la evolución (en la sección 4.3 se realiza una discusión más detallada).

4. Un patrón de estabilidad de mutaciones hidrofóbicas...

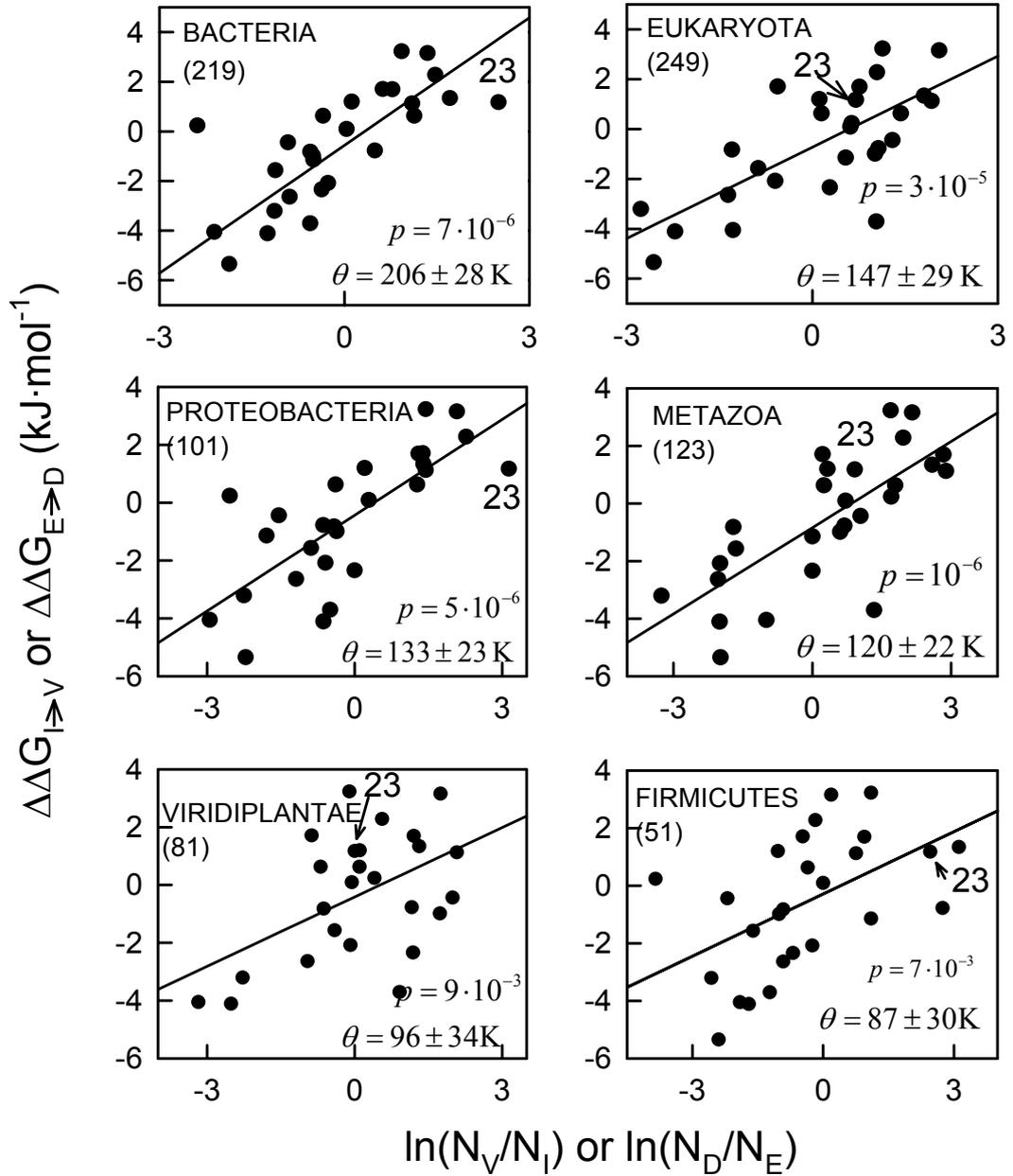


Figura 4.14: Correlación entre los efectos de la mutación sobre la estabilidad de *E.coli* y las frecuencias de aparición de los aminoácidos implicados para los cuatro subgrupos correspondientes a las taxas representadas en el alineamiento de secuencias original. En cada gráfica de los efectos de mutación sobre la estabilidad frente al logaritmo del cociente de las apariciones brutas se muestra el nombre de la taxa, el número de las secuencias (entre paréntesis), y los valores de θ y p (significancia estadística) derivados del análisis de la correlación.

4.3 EL PATRÓN EXPERIMENTAL DE ESTABILIDAD-EMPAQUETAMIENTO PARA MUTACIONES HIDROFÓBICAS REFLEJA OPTIMIZACIÓN ESTRUCTURAL EVOLUTIVA

Los resultados mostrados en las Figuras 4.4-4.14 indican que los ambientes hidrofóbicos pueden estar optimizados para interacciones estabilizantes en un alto grado de especificidad de residuo (hasta el punto de discriminar entre I y V). Más aún, esta optimización estructural evolutiva se refleja claramente en el sorprendente, y a la vez simple, patrón (Figura 4.15) que observamos cuando representamos los efectos de las mutaciones sobre la estabilidad frente a la densidad de empaquetamiento (valores de η).

Las densidades de empaquetamiento hidrofóbico se han obtenido para todas las posiciones con I o V en tiorredoxina silvestre (*WT*) de *E. coli* de acuerdo al siguiente procedimiento: se ha calculado el promedio de átomos (N_C) de carbono existente sobre la cadena lateral de I o V, en una esfera de 6 Å. Más concretamente, se ha calculado $[N_C(C_\beta) + N_C(C_{\gamma1}) + N_C(C_{\gamma2})]/3$ cuando se trata de V y $[N_C(C_\beta) + N_C(C_{\gamma1}) + N_C(C_{\gamma2}) + N_C(C_{\delta1})]/4$ cuando se trata de I. El cálculo de N_C se ha realizado con el programa METH.exe (desarrollado en el grupo de investigación de J.M. Sánchez-Ruiz).

En la Figura 4.15-A se puede observar que, para una densidad de empaquetamiento elevada, los ambientes se encuentran optimizados para V ($\Delta\Delta G_{I \rightarrow V} \gg 0$) o para I ($\Delta\Delta G_{I \rightarrow V} \ll 0$). En términos estructurales, si el ambiente de una posición dada se encuentra optimizado (se selecciona evolutivamente) para una V, una I en esa posición causará sobreempaquetamiento y originará un descenso en la estabilidad. Del mismo modo, si el ambiente está optimizado para I, una V producirá una cavidad con la consiguiente pérdida de interacciones debidas al empaquetamiento ocasionando nuevamente un descenso de la estabilidad (véase como ejemplo, Loladze *et al.*, 2002). La relación entre la selección evolutiva del ambiente que rodea al residuo y los efectos de la mutación sobre la estabilidad de la proteína se han propuesto previamente por Takano *et al.*, 2003) en base a un análisis del efecto de mutaciones Thr \leftrightarrow Val sobre las estabilidad de ribonucleasa Sa.

4. Un patrón de estabilidad de mutaciones hidrofóbicas...

Además, de acuerdo a la Figura 4.15-A, la capacidad del ambiente para discriminar entre V e I desciende cuando decrece la densidad de empaquetamiento y desaparece a una densidad de empaquetamiento alrededor de 10, donde las ramas optimizadas para V e I se unen (merece la pena prestar atención al hecho de que las valinas en *E. coli* tiorredoxina tienden a agruparse en la rama optimizada para valinas). Sólo las posiciones 5 y 91 que presentan una elevada exposición al disolvente (panel inferior de la Figura 4.1), se alejan significativamente de este patrón. Esto nos indica que su ambiente no puede ser únicamente descrito en términos de densidad de empaquetamiento hidrofóbico. El ligero valor negativo de $\Delta\Delta G_{I\rightarrow V}$ para la densidad de empaquetamiento al cual las dos ramas se unen (Figura 4.15-A) refleja una preferencia no específica para I sobre V asociada con el grupo metilo de aquel.

Todo lo expuesto hasta el momento (Figura 4.15-A) indica una dependencia parabólica de la densidad de empaquetamiento hidrofóbico con $\Delta\Delta G_{I\rightarrow V}$. En la siguiente sección (sección 4.4) desarrollamos un modelo simple que racionaliza esta dependencia en términos de su origen físico y evolutivo.

4.4 UN MODELO EVOLUTIVO SIMPLE PARA EL EFECTO DE MUTACIONES HIDROFÓBICAS SOBRE LA ESTABILIDAD DE PROTEÍNAS

La existencia de patrones de estabilidad para mutaciones que reflejan optimización evolutiva (como el que se muestra en la Figura 4.15-A) sugiere la posibilidad de que los efectos de mutaciones sobre la estabilidad de proteínas se pueden interpretar (y predecir potencialmente) en base a modelos evolutivos (además de modelos físicos). En este contexto, nos referimos específicamente a “modelo evolutivo” como un modelo para los efectos de las mutaciones sobre la estabilidad, en el cual se han evaluado ciertas características físicas de los ambientes de los residuos mutados, que considera cómo dichos ambientes se han seleccionado durante la evolución de proteínas (concepto opuesto al de “modelo físico” en el cual estas características se derivarían de algún tipo de análisis energético de las características estructurales de la proteína). En esta sección, ilustramos cómo se puede construir un modelo evolutivo usando, como ejemplo, las mutaciones hidrofóbicas I→V estudiadas.

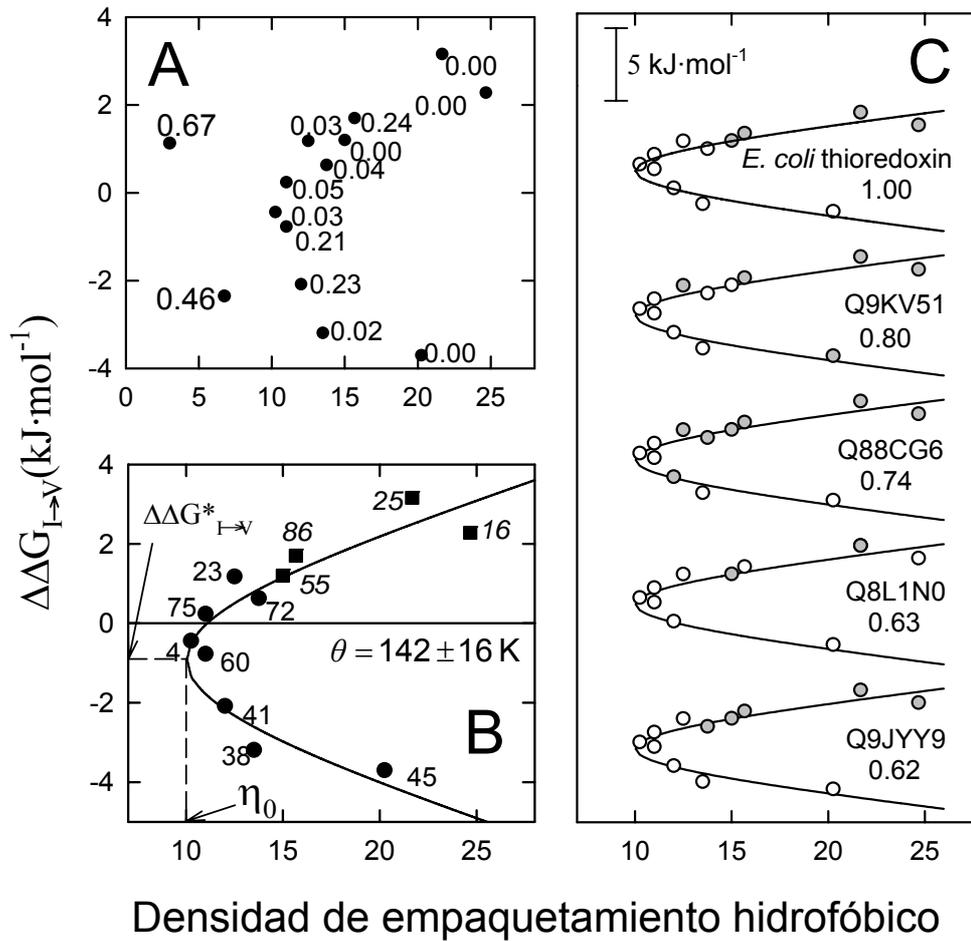


Figura 4.15: A) Efectos de mutaciones hidrofóbicas sobre la estabilidad *versus* densidad de empaquetamiento hidrofóbico. Los números junto a los puntos indican la accesibilidad de las cadenas laterales de dichos aminoácidos al disolvente en la estructura nativa. B) Ajuste de la ecuación 4.15 a los efectos experimentales de las mutaciones sobre la estabilidad de tiorredoxina. Los números junto a los símbolos indican la posición mutada. Los efectos de las mutaciones sobre la energía libre de desnaturalización se dan en la dirección I→V y el tipo de residuo presente en la forma WT se indica por el residuo usado: círculos para I y cuadrados para V. La línea continua representa el mejor ajuste ($\theta=142\pm 16 \text{ K}$) de la ecuación 4.15 a los datos experimentales de las posiciones hidrofóbicas mutadas. C) Distribución de los residuos I/V en tiorredoxina de *E. coli* y 4 secuencias del alineamiento realizado con BLAST (Trx25.aln). De arriba a abajo, esas proteínas son tiorredoxina de *E. coli* Q9KV51 (tiorredoxina de *Vibrio cholerae*), Q88CG6 (tiorredoxina de *Pseudomonas putida*), Q8L1N0 (tiorredoxina de *Buchnera aphidicola*), Q9JYY9 (tiorredoxina de *Neisseria meningitides*). Los número junto al código de la secuencia indican la similitud de esas secuencias con la secuencia *query*.

4. Un patrón de estabilidad de mutaciones hidrofóbicas...

Empezamos describiendo el efecto de una mutación I→V en una posición determinada sobre la estabilidad termodinámica como:

$$\Delta\Delta G_{I\rightarrow V} = \Delta\Delta G_{I\rightarrow V}^* - (G_V^{ENV} - G_I^{ENV}) \quad (4.3)$$

donde G_V^{ENV} and G_I^{ENV} se refieren a las energías libres asociadas con el ambiente en la estructura nativa cuando V e I se encuentran presentes en esa posición. $\Delta\Delta G_{I\rightarrow V}^*$ tiene en cuenta todas las contribuciones no relacionadas con el ambiente, sobre todo efectos del estado desnaturalizado por ejemplo. De hecho, una pequeña y no definida parte constante de la contribución del ambiente no específica a $\Delta\Delta G_{I\rightarrow V}$, se incluye en $\Delta\Delta G_{I\rightarrow V}^*$.

El ambiente de un residuo hidrofóbico dado (I o V) en la estructura nativa se describe en términos de tres características:

1. La densidad de empaquetamiento hidrofóbico alrededor del residuo (η , promedio del número de átomos de carbono, en una esfera de 6 Å de radio; véase sección 4.3). Hay que resaltar que η solo proporciona una medida global de la densidad de empaquetamiento; detalles estructurales más finos del empaquetamiento, como aquellos responsables de la optimización del ambiente para V o I, se describen por el parámetro f descrito a continuación.
2. La cuantía en la que el ambiente se encuentra optimizado para interacciones con I o V. Esta característica se describe con un único parámetro f , que toma un valor 0 cuando el ambiente se encuentra perfectamente optimizado para I y un valor de la unidad cuando el ambiente se encuentra optimizado para V. Además, la energía libre asociada a la interacción de I con el ambiente se escribe como $\alpha_0 \cdot (\eta_m - \eta) + \alpha \cdot \eta \cdot f$, donde η_m es el valor máximo de η y α_0 y α son constantes. Esto significa que nosotros asumimos una contribución no específica en la región que rodea al residuo I [$\alpha_0 \cdot (\eta_m - \eta)$] y una contribución relacionada con la optimización del ambiente [$\alpha \cdot \eta \cdot f$]. La interacción óptima en el ambiente de I ocurre cuando el empaquetamiento global es el máximo posible para I ($f=0$); en este caso, la energía libre de la interacción toma el valor mínimo posible, que es 0. La energía libre

4. Un patrón de estabilidad de mutaciones hidrofóbicas...

asociada a la interacción de V con el ambiente se escribe del mismo modo: $\alpha_0 \cdot (\eta_m - \eta) + \alpha \eta \cdot (1 - f)$ y la interacción óptima en el ambiente de V (energía libre de interacción igual a cero) ocurre cuando $\eta = \eta_m$ (empaquetamiento global máximo) y $f = 1$ (ambiente totalmente optimizado para V). Se ha usado la misma expresión para el término de la interacción no específica para V e I [$\alpha_0 \cdot (\eta_m - \eta)$], aunque esta contribución podría esperarse que fuera mayor para I (debido al grupo metilo extra). Pero, como lo que se intenta es proponer un modelo simple, se ha considerado que la pequeña diferencia entre V e I en una interacción no específica con el ambiente es aproximadamente constante y puede ser incluida en $\Delta\Delta G^*_{I \rightarrow V}$ (ecuación 4.3). Hay que señalar que, a diferencia de η , no se necesita la estructura nativa de la proteína para el cálculo de los valores de f sino que, en el modelo propuesto, la optimización del ambiente se determina en base a la hipótesis relacionada con el efecto de la selección natural (ver a continuación).

3. La “plasticidad” del ambiente. Se considera que los ambientes hidrofóbicos tienen cierta flexibilidad y como tal, pueden cambiar como respuesta a una mutación. Además, en el modelo se incluye un término de energía libre asociado a la reorganización del ambiente dado por $\beta \cdot \eta \cdot (f - f_0)^2$, donde β es una constante. La energía libre de reorganización es cero cuando $f = f_0$ y por lo tanto, el valor de f_0 se puede considerar como una medida de la preferencia “intrínseca” del ambiente por I o V. Desde luego, se pueden observar valores de f diferentes de f_0 si la interacción ambiente-residuo alcanza la energía libre de reorganización necesaria. Esta energía libre de reorganización será mayor cuanto más alto sea el valor de η ; es decir, en nuestro modelo estamos considerando que ambientes bien empaquetados son más “rígidos”.

Los términos de energía libre del ambiente en la ecuación 4.3 (G_V^{ENV} y G_I^{ENV}) contienen, por lo tanto, una contribución para las interacciones directas del residuo (I o V) con el ambiente y además otra contribución para la posible reorganización del ambiente:

4. Un patrón de estabilidad de mutaciones hidrofóbicas...

$$G_I^{ENV}(f) = \alpha_0 \cdot (\eta_m - \eta) + \alpha \cdot \eta \cdot f + \beta \cdot \eta \cdot (f - f_0)^2 \quad (4.4)$$

$$G_V^{ENV}(f) = \alpha_0 \cdot (\eta_m - \eta) + \alpha \cdot \eta \cdot (1 - f) + \beta \cdot \eta \cdot (f - f_0)^2 \quad (4.5)$$

Estas expresiones se consideran como funciones de energía libre. Es decir, los valores reales de G_V^{ENV} y G_I^{ENV} se obtienen mediante una minimización de las ecuaciones 4.4 y 4.5 respecto a f . En la Figura 4.16 se muestra gráficamente dicha minimización de energía. Los paneles A, B y C corresponden al caso de un ambiente rígido (alto η) que se encuentra optimizado para I (panel A), para V (panel B) o un ambiente que no tiene una optimización intrínseca definida ($f_0=1/2$, panel C); en el panel D se muestra la situación de un ambiente “plástico” (bajo η) y una optimización intrínseca no definida ($f_0=1/2$). Los valores óptimos de f_I y f_V son diferentes del valor de f_0 , como consecuencia de los cambios que sufre el ambiente en respuesta a la interacción con el residuo. Las expresiones generales de f_I y f_V se pueden obtener resolviendo $\partial G_I^{ENV} / \partial f = 0$ y $\partial G_V^{ENV} / \partial f = 0$, respectivamente. Los resultados son $f_I = f_0 - \alpha/2\beta$, $f_V = f_0 + \alpha/2\beta$. Y como f se encuentra en el rango $0 \leq f \leq 1$, f_0 queda restringido en nuestro modelo al rango $\alpha/2\beta \leq f_0 \leq 1 - \alpha/2\beta$. Como resultado, si se sustituye en las ecuaciones. 4.4 y 4.5 se obtienen las expresiones:

$$G_I^{ENV} = \alpha_0 \cdot (\eta_m - \eta) + \alpha \eta \cdot \left(f_0 - \frac{1}{2}\right) + \frac{\alpha \eta}{2} \left(1 - \frac{\alpha}{2\beta}\right) \quad (4.6)$$

$$G_V^{ENV} = \alpha_0 \cdot (\eta_m - \eta) - \alpha \eta \cdot \left(f_0 - \frac{1}{2}\right) + \frac{\alpha \eta}{2} \left(1 - \frac{\alpha}{2\beta}\right) \quad (4.7)$$

Para las derivaciones posteriores se necesita tener G_I^{ENV} y G_V^{ENV} referidas a sus valores más bajos posibles (que son aquellos que corresponden a $f_0 = \alpha/2\beta$ and $f_0 = 1 - \alpha/2\beta$, respectivamente):

$$\Delta G_I^{ENV} = G_I^{ENV} - G_I^{ENV}(\alpha/2\beta) = \alpha \eta \cdot \left(f_0 - \frac{1}{2}\right) + \frac{\alpha \eta}{2} \left(1 - \frac{\alpha}{\beta}\right) \quad (4.8)$$

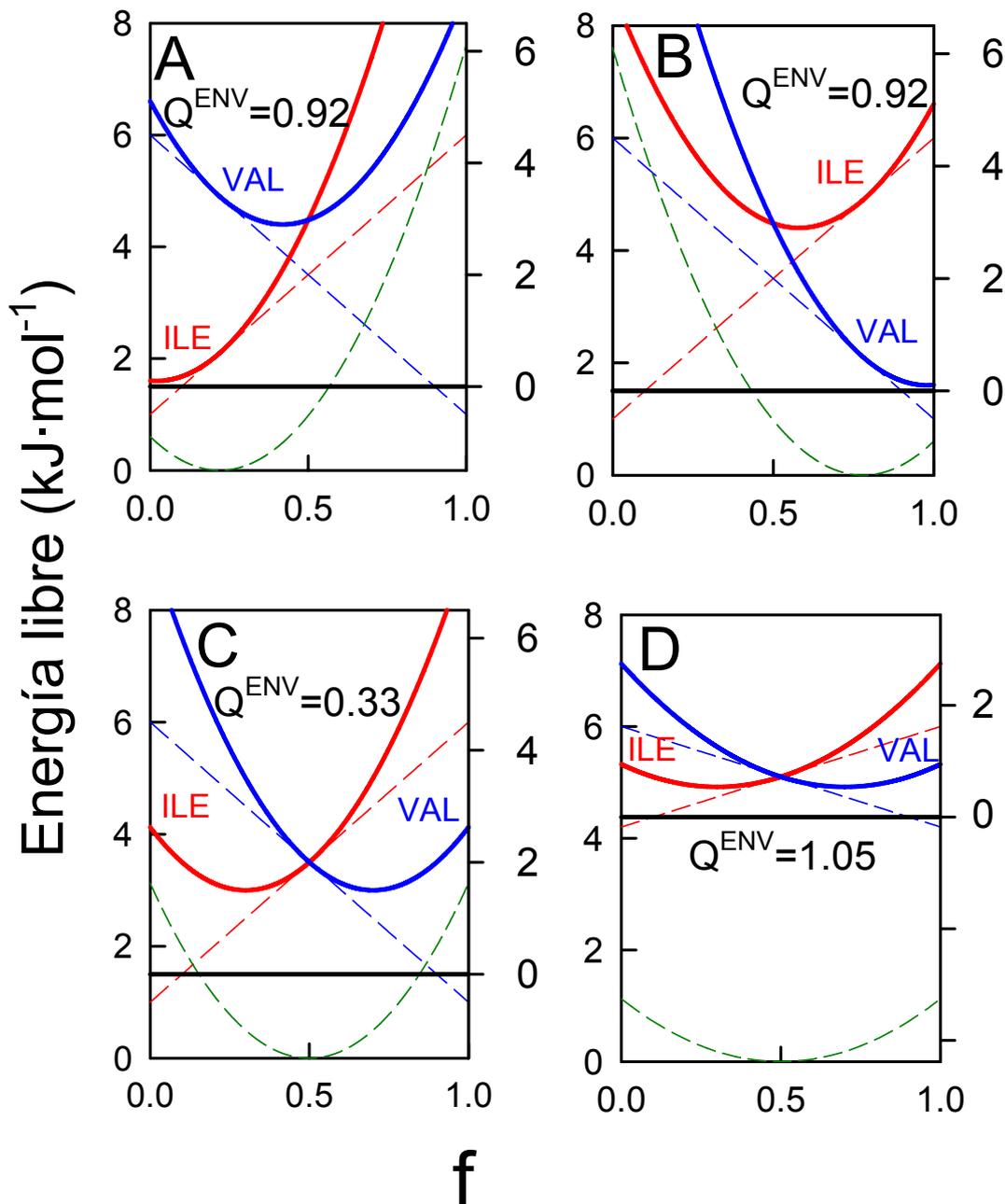


Figura 4.16: Ejemplos de energías libres asociadas al ambiente cuando I o V se encuentran presentes en una posición, calculados con el modelo simple descrito en la sección 4.4 (ecuaciones 4.3-15). Los valores de G_I^{ENV} y G_V^{ENV} se representan por líneas continuas roja y azul, respectivamente. Las líneas roja y azul discontinuas representan la contribución de G_I^{ENV} y G_V^{ENV} asociadas a la interacción residuo-ambiente (ecuaciones 4.4 y 4.5). La línea verde discontinua representa la contribución asociada a la reorganización del ambiente (ecuaciones 4.4 y 4.5). La línea negra representa el valor de G_I^{ENV} y G_V^{ENV} más bajo posible. La diferencia entre el mínimo de las líneas continuas roja y azul indica la optimización alcanzada.

4. Un patrón de estabilidad de mutaciones hidrofóbicas...

$$\Delta G_V^{ENV} = G_V^{ENV} - G_V^{ENV} (1 - \alpha / 2\beta) = -\alpha\eta \left(f_0 - \frac{1}{2}\right) + \frac{\alpha\eta}{2} \left(1 - \frac{\alpha}{\beta}\right) \quad (4.9)$$

Si se sustituyen las ecuaciones 4.6 y 4.7 en la ecuación 4.3 se obtiene que,

$$\Delta\Delta G_{I \rightarrow V} = \Delta\Delta G_{I \rightarrow V}^* + 2\alpha\eta \left(f_0 - \frac{1}{2}\right) \quad (4.10)$$

ecuación que nos da el efecto de la mutación sobre la estabilidad en términos de la preferencia del ambiente por I o V, como una medida del valor f_0 .

En este punto, planteamos una hipótesis relacionada como las restricciones impuestas por la selección natural sobre los ambientes de los residuos. Se asume que el ambiente que rodea a los residuos I/V “se acepta” durante la evolución si se alcanza la optimización para al menos un tipo de residuo (I o V). De acuerdo con esta hipótesis, la situación representada en la Figura 4.16-A se aceptaría siempre que el ambiente conduzca a una optimización clara para I (y por tanto G_I^{ENV} se encuentre cerca de su valor más bajo posible y ΔG_I^{ENV} sea próximo a cero); por supuesto, la optimización para V no se alcanza y esto plantea un pequeño problema, ya que, de acuerdo a la hipótesis de pseudo-equilibrio V raramente aparece en esa posición durante la evolución de la proteína (quizás esto ocurra sólo cuando mutaciones estabilizantes que ocurran en otra parte de la molécula compensen los efectos desestabilizantes que ocasiona el tener una V en un ambiente bien empaquetado optimizado para I). Del mismo modo, se acepta la situación que se muestra en la Figura 4.16-B donde se alcanza la optimización del ambiente que rodea a V. Sin embargo, el ambiente rígido sin preferencia intrínseca por I o V ($f_0 = 1/2$), representado en la Figura 4.16-C, se rechazará ya que no puede ser optimizado ni para I ni para V. Por otro lado, se acepta un ambiente plástico (bajo empaquetamiento) sin una preferencia clara (Figura 4.16-D), ya que se puede reorganizar para la interacción con I y V, permitiendo la optimización del ambiente para ambos tipos de residuos.

El razonamiento anterior se puede expresar en términos matemáticos considerando que las funciones de partición siguientes deben tener un valor suficientemente alto:

4. Un patrón de estabilidad de mutaciones hidrofóbicas...

$$\begin{aligned}
 Q^{ENV} &= \exp\left(-\frac{\Delta G_I^{ENV}}{R\theta}\right) + \exp\left(-\frac{\Delta G_V^{ENV}}{R\theta}\right) = \\
 &\left\{ \exp\left(-\frac{\alpha\eta}{2R\theta}(1-\alpha/\beta)\right) \right\} \left\{ \exp\left(-\frac{\alpha\eta}{R\theta}(f_0-1/2)\right) + \exp\left(\frac{\alpha\eta}{R\theta}(f_0-1/2)\right) \right\} = \quad (4.11) \\
 &2 \cdot \left\{ \exp\left(-\frac{\alpha\eta}{2R\theta}(1-\alpha/\beta)\right) \right\} \cosh\left(\frac{\alpha\eta}{R\theta}(f_0-1/2)\right)
 \end{aligned}$$

donde, en la primera igualdad hemos usado las ecuaciones 4.8 y 4.9 y en la segunda, la definición del coseno hiperbólico: $\cosh(x)=(e^x+e^{-x})/2$ (ver capítulo 3 en Warren y Gordon, 1966). Bajo esta consideración, al menos uno de los términos de energía libre (ΔG_I^{ENV} or ΔG_V^{ENV}) es próximo a cero, podemos esperar que el valor de Q^{ENV} sea próximo a la unidad (como ilustración, véase el valor de Q^{ENV} en la Figura 4.16). No se está imponiendo que la selección natural mantenga estrictamente $Q^{ENV}=1$, sino que, para obtener una razonable aproximación a los ambientes evolutivamente aceptados (como se describe por el valor de f_0), se fija $Q^{ENV}=1$ en la ecuación. 4.11 y se resuelve para f_0 :

$$f_0 = \frac{1}{2} + \frac{R\theta}{\alpha\eta} \cosh^{-1} \left\{ \frac{1}{2} \exp\left(\frac{\alpha\eta}{2R\theta}(1-\alpha/\beta)\right) \right\} \quad (4.12)$$

donde \cosh^{-1} es la inversa del coseno hiperbólico. Esta función se puede calcular fácilmente (véase capítulo 3 en Warren y Gordon, 1966) a partir de $\cosh^{-1}(x) = \ln(x \pm \sqrt{x^2-1})$ y para $x>1$, tiene dos valores que sólo se diferencian en el signo; para $x=1$ tiene sólo un valor (cero) y la función no está definida para $x<1$. Esto implica que en nuestro modelo debe haber un valor mínimo de η (que llamaremos η_0), que puede calcularse igualando el argumento de \cosh^{-1} a 1 en la ecuación 4.12. Despejando para η_0 :

$$\eta_0 = \frac{2 \cdot \ln 2 \cdot R\theta}{\alpha \cdot (1-\alpha/\beta)} \quad (4.13)$$

y usando las ecuaciones 4.12 y 4.13, podemos escribir:

4. Un patrón de estabilidad de mutaciones hidrofóbicas...

$$f_0 = \frac{1}{2} + \frac{R\theta}{\alpha\eta} \cosh^{-1} \left\{ 2^{(\eta-\eta_0)/\eta_0} \right\} \quad (4.14)$$

que si se sustituye en la ecuación. 4.10, se obtiene finalmente:

$$\Delta\Delta G_{I \rightarrow V} = \Delta\Delta G_{I \rightarrow V}^* + 2R\theta \cdot \cosh^{-1} \left\{ 2^{(\eta-\eta_0)/\eta_0} \right\} \quad (4.15)$$

De acuerdo con las ecuaciones 4.14 y 4.15, para $\eta = \eta_0$, existe sólo un valor de $f_0(1/2)$ y sólo un valor posible para el efecto de la mutación sobre la estabilidad ($\Delta\Delta G_{I \rightarrow V} = \Delta\Delta G_{I \rightarrow V}^*$). Por lo tanto, η_0 se puede interpretar por definición como la densidad de empaquetamiento más baja para la cual la reorganización del ambiente no es posible. Por otro lado, para cada valor de η mayor que η_0 , existen dos posibles valores de f_0 ($f_0 > 1/2$ y $f_0 < 1/2$) y dos posibles valores de $\Delta\Delta G_{I \rightarrow V}$ ($\Delta\Delta G_{I \rightarrow V} > \Delta\Delta G_{I \rightarrow V}^*$ y $\Delta\Delta G_{I \rightarrow V} < \Delta\Delta G_{I \rightarrow V}^*$), que obviamente corresponden a la optimización del ambiente para V e I, respectivamente.

Es importante resaltar que muchos detalles de nuestro modelo no aparecen en la ecuación 15, la cual es sorprendentemente simple. Pero utilizamos ésta como muestra de la robustez de la ecuación. Por ejemplo, se puede comprobar que si se usa una función general de η , en lugar de $\alpha_0 \cdot (\eta - \eta_m)$, o si se elimina la reorganización del ambiente (se omite el término $\beta \cdot (f - f_0)^2$ y por lo tanto la minimización de G_I^{ENV} y G_V^{ENV}), se llega a la misma ecuación 4.15. Esta última, sin embargo, habría sido inconsistente con datos estructurales experimentales que mostrasen un grado variable de colapso de la cavidad cuando se muta una cadena lateral hidrofóbica a una más pequeña. Por lo tanto, el enfoque que hemos hecho ha sido incluir la reorganización del ambiente en el análisis de una manera simple pero plausible (ecuaciones 4.4 y 4.5) ya que al final, se demuestra que los resultados obtenidos del análisis teórico (ecuación 4.15) no dependen de la reorganización del ambiente.

4.5 APLICACIÓN DEL MODELO A LOS DATOS DE MUTACIONES DE TIORREDOXINA

La ecuación 4.15 tiene sólo tres parámetros ajustables ($\Delta\Delta G^*_{I\rightarrow V}$, η_0 y θ). El ajuste no lineal por mínimos cuadrados a los datos experimentales de $\Delta\Delta G_{I\rightarrow V}$ versus η es directo.

Hemos ajustado la ecuación 4.15 a los datos experimentales de $\Delta\Delta G_{I\rightarrow V}$, sin tener en cuenta las dos mutaciones que tienen una exposición elevada al disolvente (5 y 91, véase Figura 4.15-A). El ajuste de los datos experimentales (incluyendo las 12 posiciones restantes) ha resultado ser excelente (Figura 4.15-B). Como se esperaba, los datos de los dos residuos altamente expuestos al disolvente no se ajustan, lo que indica que sus ambientes no se pueden describir sólo en términos de densidad de empaquetamiento hidrofóbico. El ajuste es bastante robusto respecto al parámetro elegido para medir el empaquetamiento: Aunque se ha usado para η el promedio del número de átomos de carbonos de la cadena lateral que rodea al residuo en una distancia de 6 Å (véase sección 4.4), también se ha comprobado que si se aumentaba la distancia hasta 8 Å o se incluían todos los átomos de las cadenas laterales (no sólo carbonos) los resultados obtenidos eran similares.

Los valores de los parámetros ajustables derivados del ajuste que se muestra en la Figura 4.15B son $\eta_0=10\pm 0.2$, $\Delta\Delta G^*_{I\rightarrow V}=-0.9\pm 0.3$ kJ/mol y $\theta=142\pm 16$ K (los errores se han calculado por el método de Monte Carlo). El valor de la temperatura θ concuerda con el obtenido del análisis de las correlaciones mostradas en este trabajo (Figuras 4.4-14). Esta coincidencia defiende claramente la validez de algunas de las características principales de nuestro modelo.

El valor bajo y negativo obtenido para $\Delta\Delta G^*_{I\rightarrow V}$ indica una preferencia general de I frente a V. Como ya se ha dicho anteriormente, la interacción no específica con el ambiente (independiente de optimización) del grupo metilo extra de isoleucina puede contribuir al valor de $\Delta\Delta G^*_{I\rightarrow V}$, así como también factores relacionados con la exposición de residuos al disolvente en el estado desnaturalizado, incluyendo estructura residual (agrupaciones de residuos hidrofóbicos). En relación con esta última posibilidad, debe señalarse que un estado desnaturalizado térmicamente fuertemente “polarizado” (un estado desnaturalizado con agrupaciones de residuos hidrofóbicos bien definidas en ciertas regiones y una alta exposición al disolvente en otras regiones), habría

ocasionado valores diferentes de $\Delta\Delta G^*_{I\rightarrow V}$ para las posiciones en diferentes regiones y el ajuste de la ecuación 4.15 a los datos experimentales con una constante $\Delta\Delta G^*_{I\rightarrow V}$ habría fallado. Para comprobar la existencia de estados de estas características, hemos empleado el enfoque propuesto por Marqusee y colaboradores (Robic *et al.*, 2003) en el cual se detecta la existencia de estructura residual en el estado desnaturalizado mediante el estudio del cambio en capacidad calorífica para distintas formas mutantes en las que se han introducido aminoácidos cargados que potencialmente destruirán tales agrupaciones.

Para comprobar esto, hemos preparado 4 variantes más de tiorredoxina, son los mutantes V16D, L42D, A67D y L99D, los cuales se encuentran en la región más enterrada de las hélices- α , 11-17, 33-49, 67-70 y 94-105, respectivamente (Figura 4.17). Y dichas regiones a su vez se encuentran a ambos lados de la lámina- β central en la estructura nativa.

4.5.1 OBTENCIÓN DE LAS PROTEÍNAS MUTADAS V16D, L42D, A67D y L99D.

Las mutaciones deseadas se han diseñado según se explicó en la sección 2.2.1.2 y se han obtenido mediante mutagénesis dirigida (véase para más detalles la sección 2.2.1.3). Para este tipo de mutaciones se ha trabajado con el gen de tiorredoxina *WT* insertado en el vector pET30a(+) (sección 2.2.2.1) y las células supercompetentes BL21-DE3. La expresión y purificación de proteína se ha explicado detalladamente en las secciones 2.2.2.5 y 2.2.2.6, respectivamente. La pureza de las proteínas purificadas se evaluó mediante una electroforesis (PAGE-SDS).

4.5.2 ESTUDIO DE LA ESTABILIDAD TERMODINÁMICA DE V16D, L42D, A67D y L99D MEDIANTE CALORIMETRÍA DIFERENCIAL DE BARRIDO.

La estabilidad termodinámica de todas las proteínas mutadas (V16D, L42D, A67D y L99D) se ha determinado por Calorimetría Diferencial de Barrido (*DSC*). Los experimentos se han llevado a cabo a una velocidad de barrido de 1.5 K/min en el intervalo de temperatura de 10-115 °C (véanse detalles de la técnica y análisis de los datos en las secciones 2.2.5.2 y 2.2.5.3, respectivamente). Las disoluciones de proteínas utilizadas en los experimentos de calorimetría se dializaron en HEPES 5 mM, pH7 y su concentración se ha determinado espectrofotométricamente según se explicó en la sección 2.2.3.1.

4. Un patrón de estabilidad de mutaciones hidrofóbicas...

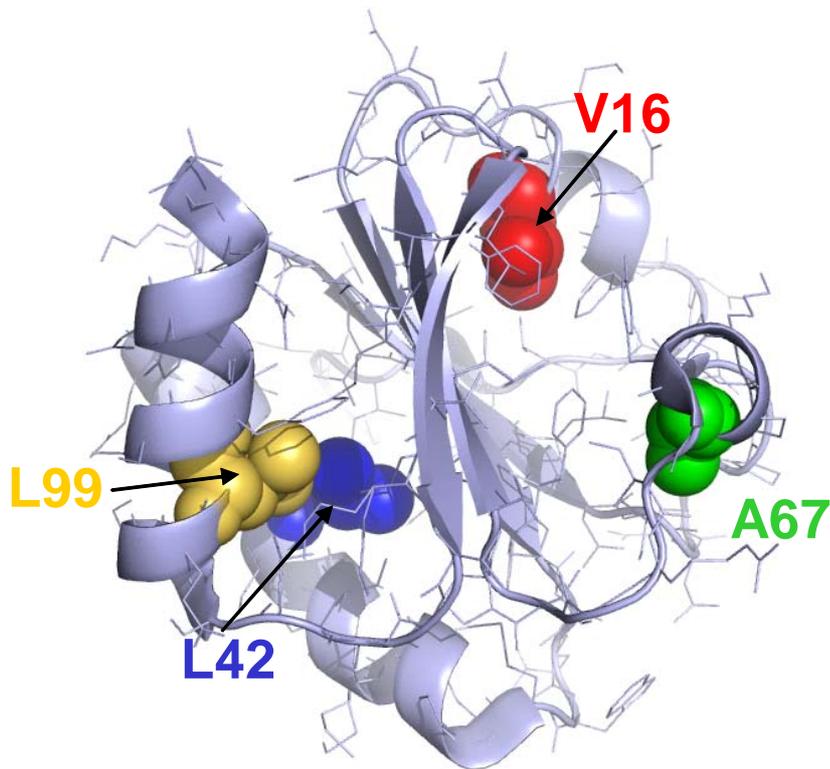


Figura 4.17: Ilustración de la posición de los residuos V16 (rojo), L42 (azul), A67 (verde) y A99 (amarillo) en la estructura de tiorredoxina de *E. coli* (2H6X.pdb). Los residuos V16, L42, A67 y L99 han sido sustituidos por aspartato, dando lugar a los mutantes V16D, L42D, A67D y L99D. Se puede observar como éstos se encuentran en la región más enterrada de las hélices- α , 11-17, 33-49, 67-70 y 94-105, respectivamente. Y dichas regiones a su vez se encuentran a ambos lados de la lámina- β central en la estructura nativa.

Variaciones en el valor de ΔC_p^0 se han relacionado con cambios en el área superficial accesible al disolvente (ASA) cuando la proteína desnatura (Makhatadze y Privalov, 1990). Además, Robic *et al.*, 2003 mostraron que ribonucleasa H termofílica tiene un valor de ΔC_p^0 significativamente menor que el de su homóloga mesofílica y esto era debido al estado desnaturado. Para ribonucleasa H termofílica, observaron un cambio abrupto de capacidad calorífica absoluta del estado desplegado en un intervalo estrecho de pH, que sugería un

4. Un patrón de estabilidad de mutaciones hidrofóbicas...

proceso extremadamente cooperativo de toma de protones en el estado desnaturalizado (formación de agrupaciones hidrofóbicas). Para demostrarlo, realizaron un estudio mutacional (basado en la sustitución de residuos hidrofóbicos por residuos cargados) que destruía dicha estructura residual haciendo que las características termodinámicas de la proteína termofílica revirtieran a las de la proteína mesofílica. De este modo, proponían cómo la formación de estructura residual en el estado desplegado, en ribonucleasa H termofílica, podría estar diseñada por la evolución.

En esta sección se ha llevado a cabo un estudio similar al realizado por Robic *et al.*, 2003. En la Figura 4.18 se muestran los perfiles de capacidad calorífica de tiorredoxina *WT* y las variantes de tiorredoxina: V16D, L42D, A67D y L99D. En la Tabla 4.4 se recogen los parámetros del ajuste de los datos experimentales al modelo de equilibrio de dos estados. Sin embargo, los resultados no fueron tan alentadores como los obtenidos por Robic *et al.* (2003). Los experimentos de DSC realizados, no nos mostraron un cambio significativo del ΔC_p^0 de las variantes de tiorredoxina: V16D, L42D, A67D y L99D, en comparación al ΔC_p^0 de tiorredoxina *WT* (Figura 4.18 y Tabla 4.4).

Para completar el estudio calorimétrico realizado y como se explica en la siguiente sección, se ha llevado a cabo también una determinación espectroscópica de la estabilidad conformacional de las variantes de tiorredoxina: V16D, L42D, A67D y L99D, mediante dicroísmo circular en el ultravioleta lejano.

4. Un patrón de estabilidad de mutaciones hidrofóbicas...

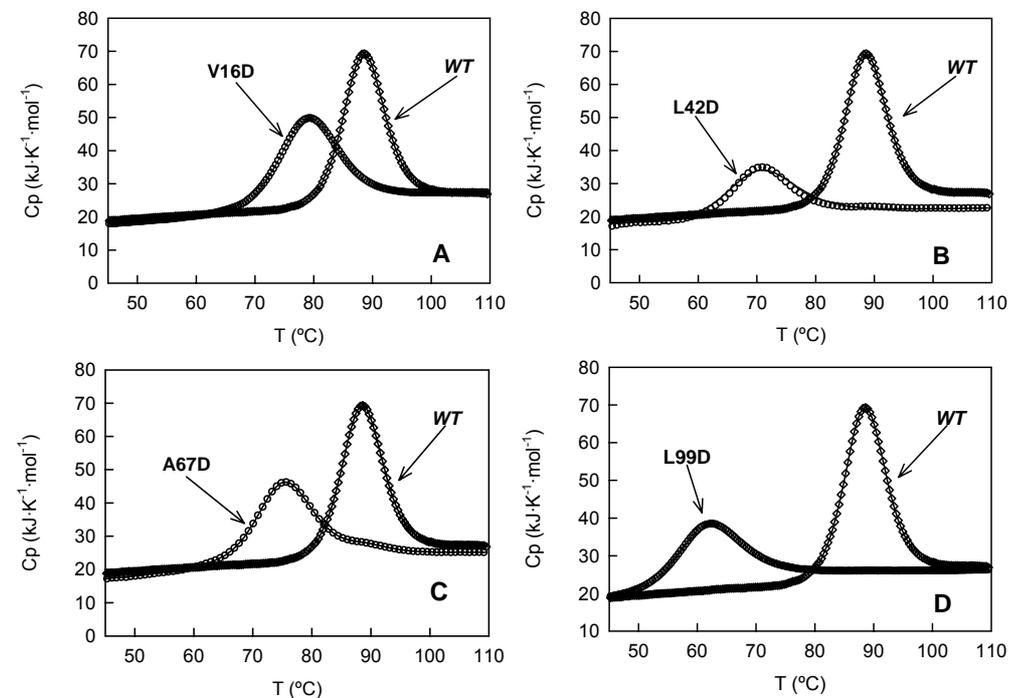


Figura 4.18: Perfiles de capacidad calorífica frente a temperatura de tioredoxina silvestre (*WT*) (rombos) y los mutantes V16D (A), L42D (B), A67D (C) y L99D (D) (círculos), obtenidos mediante Calorimetría Diferencial de Barrido (*DSC*). Los experimentos se han llevado a cabo a 1.5 K/min. Las disoluciones de proteína se dializaron en HEPES 5 mM, pH7. Las concentraciones de proteína utilizadas han sido próximas a 0.5 mg/mL. La línea continua en cada panel representa el ajuste de los datos experimentales (*WT* y variantes) al modelo de equilibrio de dos estados.

Tabla 4.4: Temperatura de desnaturalización (T_m), cambio en la temperatura de desnaturalización respecto a tiorredoxina *WT* (ecuación 3.3), cambio en la entalpía de desnaturalización (ΔH_m) derivados del ajuste al modelo de equilibrio de dos estados de las proteínas estudiadas. Se considera una desviación estándar de 0.6 kJ/K·mol para los valores de ΔC_p .

Proteína	T_m (°C)	ΔT_m (°C)	ΔH (kJ/mol)	ΔC_p (kJ/K·mol)
WT	89.05±0.05		430.0±0.5	5.8
V16D	79.11±0.03	-9.94±0.08	324.4±0.9	4.6
L42D	70.50±0.09	-18.55±0.14	242.7±1.8	4.5
A67D	75.32±0.05	-13.73±0.10	298.3±1.0	4.7
L99D	61.99±0.03	-27.06±0.08	243.7±0.4	5.7

4.5.3 CARACTERIZACIÓN ESTRUCTURAL DE V16D, L42D, A67D y L99D MEDIANTE DICROÍSMO CIRCULAR

La existencia de estructura residual en el estado desnaturalizado de V16D, L42D, A67D y L99D y de tiorredoxina silvestre (WT) también se ha evaluado mediante la observación de los espectros de dicroísmo circular (DC) de la desnaturalización térmica de tiorredoxina *WT* y variantes. Los experimentos se llevaron a cabo como se explica en la sección 2.2.6.6. Las disoluciones de proteínas se dializaron en HEPES 5 mM, pH7 y su concentración se determinó espectrofotométricamente según se explicó en la sección 2.2.3.1. En la Figura 4.19 se muestran, a modo ilustrativo, los espectros de DC en el ultravioleta lejano de tiorredoxina *WT* y V16D, L42D, A67D y L99D a diferentes temperaturas (50 y 90°C).

4. Un patrón de estabilidad de mutaciones hidrofóbicas...

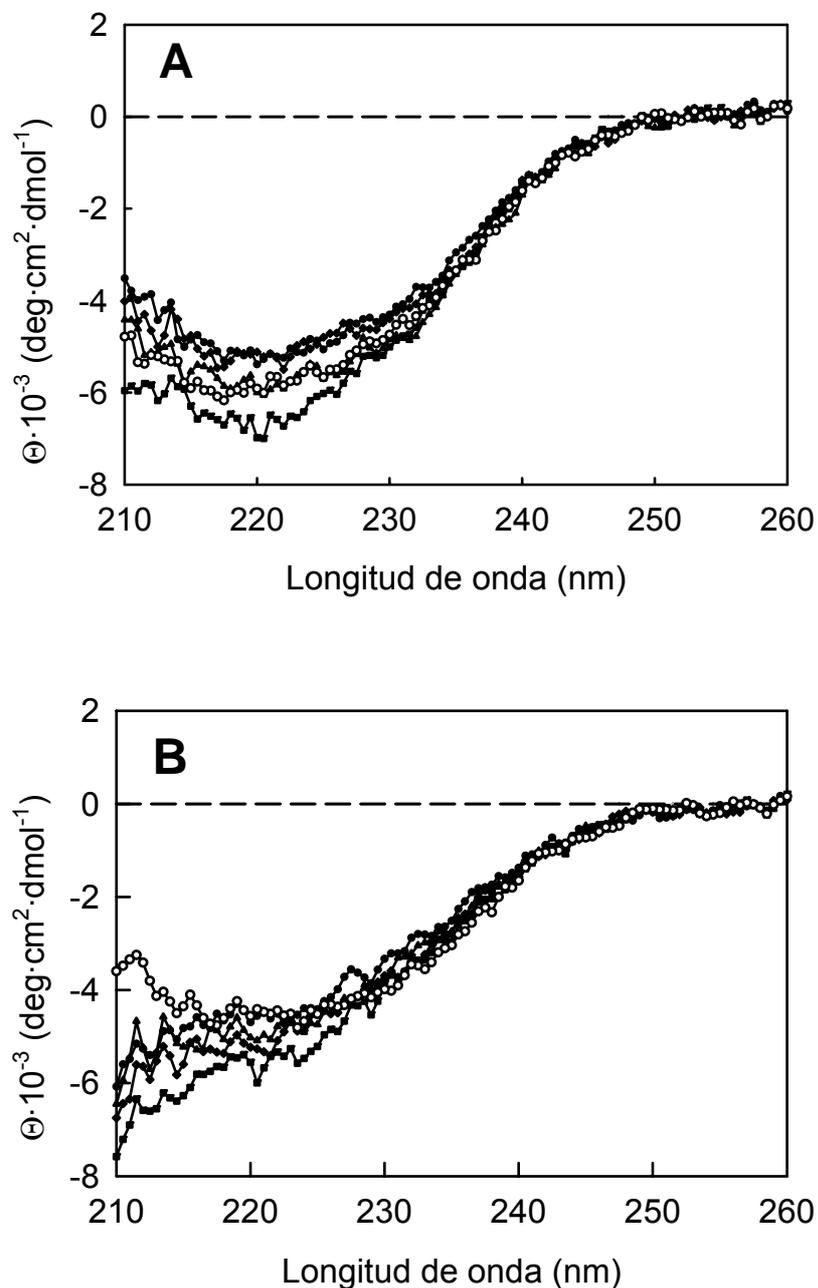


Figura 4.19: Espectros de dicroísmo circular en el UV-lejano. Se han recogido espectros a 50°C (panel A) y a 90°C (panel B). El tampón empleado ha sido HEPES 5mM, pH 7 y la concentración de proteína 0.2 mg/mL. Los círculos blancos se corresponden a tiorredoxina silvestre (WT), en círculos negros se representa el mutante V16D, en cuadrados el mutante L42D, en triángulos el mutante A67D y en rombos el mutante L99D.

4. Un patrón de estabilidad de mutaciones hidrofóbicas...

En los espectros de DC, en UV-lejano, que se muestran en la Figura 4.19, se puede observar como a temperatura de 50 °C existen pequeñas diferencias en los espectros de tiorredoxina *WT* (círculos blancos) y los de las variantes de tiorredoxina: V16D, L42D, A67D y L99D, a 2220 nm. Éstas pequeñas diferencias, pueden estar indicando diferencias entre la estructura secundaria nativa de *WT* y las variantes. A 90°C, los espectros de DC, en UV-lejano, de tiorredoxina *WT* y variantes son bastante similares.

Con estos resultados, podríamos decir que las diferencias en ΔC_p observadas mediante calorimetría diferencial de barrido parecen estar sugeridas por cambios conformacionales en el estado nativo de las variantes de tiorredoxina y no a la existencia de estructura residual en el estado desplegado. Sin embargo, sería interesante llevar a cabo un estudio mutacional y estructural (actualmente, se han cristalizado las proteínas variantes V16D, L42D, A67D y L99D) más completo; ésto nos podría ofrecer una información más detallada en relación a los cambios tanto en las características termodinámicas como estructurales de las variantes de tiorredoxina: V16D, L42D, A67D y L99D

4.6 CONSERVACIÓN DEL AMBIENTE *VERSUS* CONSERVACIÓN DE RESIDUOS

Los datos que hemos presentado (Godoy-Ruiz *et al.*, 2005) indican que muchos ambientes hidrofóbicos se encuentran optimizados en alta medida para V o para I, pero además, que durante el curso de la evolución de proteínas, I puede ser aceptado en ambientes optimizados para V (y viceversa), con frecuencias que reflejan aproximadamente las energías libres de penalización. Por ejemplo, el valor positivo de $\Delta\Delta G_{I \rightarrow V}$ obtenido para la posición 23 indica que el ambiente que rodea esa posición se encuentra optimizado para V. De hecho, si se observa el alineamiento de secuencias de tiorredoxina de *E. coli*, durante la evolución de proteínas, V aparece más frecuentemente que I en esta posición (152 V frente a 13 I en las secuencias con una similaridad mayor de 0.3 con la secuencia de tiorredoxina de *E.coli*). Esta preferencia general de V sobre I para la posición 23 se observa además cuando se analizan separadamente las tasas en las que se pueden dividir las secuencias alineadas: 107 V frente 4 I para bacteria, 37 V frente 8 I para eucariota, 52 V frente 2 I para proteobacteria y 15 V frente a 4 I para metazoa. Claramente, la posición 23 es una “posición optimizada para valina”,

4. Un patrón de estabilidad de mutaciones hidrofóbicas...

aunque hay una I en esa posición para tiorredoxina de *E.coli* como resultado de una mutación reciente (en el sentido evolutivo). Esta interpretación se encuentra en la misma línea que la reciente hipótesis evolutiva propuesta para la generación de vueltas de hélices transmembrana (Yohannan *et al.*, 2004), de acuerdo con la cual las vueltas sin prolina en proteínas de membrana se estabilizan por segundas mutaciones que se encargan de optimizar el empaquetamiento alrededor de la vuelta sin prolina. De hecho, se ha comprobado como los residuos prolina aparecen mutados frecuentemente en posiciones en las que existe una vuelta.

Merece la pena resaltar cómo los resultados presentados nos indican que los ambientes hidrofóbicos se pueden optimizar por interacciones estabilizantes con un alto grado de especificidad del residuo, hasta el punto de discriminar entre I y V, residuos que sólo se diferencian en un grupo metilo (hay que pensar que las consecuencias estructurales de mutaciones I \leftrightarrow V se esperan que sean mucho más pequeñas que aquellas debidas a la introducción de una prolina en una hélice). Esta especificidad puede ser la consecuencia de la necesidad de un buen empaquetamiento para ciertas regiones de la proteína, lo que requiere un buen empaquetamiento de las cadenas laterales de los residuos implicados. Además, si originalmente se alcanzó un buen empaquetamiento (en sentido evolutivo) alrededor una posición dada, como por ejemplo valina, el ambiente se encontrará optimizado para valina. Ciertamente, las mutaciones neutras o cuasi-neutras del tipo V \rightarrow I se aceptarán ocasionalmente durante el curso de la evolución (como ocurre en la posición 23) como resultado de una deriva al azar. Sin embargo, el tipo de optimización de ambiente tiende menos a cambiar (desde una optimización para V a una optimización para I), ya que esto requerirá posiblemente el efecto combinado de varias mutaciones de segundo orden (Baldwin *et al.*, 1996).

De todo lo anteriormente discutido, concluimos diciendo que los ambientes de los residuos hidrofóbicos se encuentran más conservados que los residuos en si mismos. Por lo tanto, ésto implica, que $\Delta\Delta G_{I\rightarrow V}$ frente al patrón de empaquetamiento (Figuras 4.15-A y 4.15-B) debería ser el mismo (o similar) para proteínas relacionadas evolutivamente, aunque la distribución de residuos I y V puede variar en diferente grado. Esta interpretación se ilustra en la Figura 4.15-C donde mostramos sobre el patrón de estabilidad/empaquetamiento para tiorredoxina, la distribución de I/V para algunas de las secuencias obtenidas en la

búsqueda realizada con BLAST usando tiorredoxina de *E.coli* como la secuencia a comparar.

4.7 RUGOSIDAD DEL PAISAJE DE ENERGÍA DE ESTABILIDAD Y SECUENCIA (ANALOGÍA CON UN JARDÍN JAPONÉS DE ROCAS)

La optimización evolutiva del ambiente de residuos puede ocurrir en diferente extensión para proteínas distintas (por ejemplo, proteínas mesofílicas frente proteínas termofílicas) y, para la misma proteína, puede ocurrir en diferentes regiones (debido por ejemplo a las restricciones funcionales o debido a interacciones existentes en la proteína). Los resultados expuestos en este trabajo de investigación sugieren que, por lo menos en algunas proteínas, para un número significativo de residuos se optimiza evolutivamente su ambiente mediante interacciones estabilizantes con residuos específicos y que residuos que no están a gusto en el ambiente aparecen raramente (con frecuencias relacionadas con la penalización de sus energías libres y asociadas a la hipótesis de pseudo-equilibrio). Sin embargo, la optimización evolutiva de algunos ambientes de residuos en proteínas no quiere decir que la estabilidad de la proteína es próxima a su máximo (lo que quizás podría encontrarse en conflicto con el hecho de que la estabilidad de proteínas es marginal). Desde nuestro punto de vista, este hecho solamente quiere decir que el paisaje de estabilidad-secuencia es rugoso, de modo que, en este sentido, la mayoría de las mutaciones sencillas son desestabilizantes. Creemos que una analogía muy ilustrativa de dicho paisaje estabilidad-secuencia puede ser un jardín de rocas japonés (como los que existen en Ryogen-in y Tofuku-ji in Kyoto, Figura 4.20 <http://phototravels.net/japan/photo-gallery/japanese-rock-gardens.html>). El nivel de la arena representa el mínimo de estabilidad necesario para un correcto plegamiento y las rocas simbolizan “máximos de estabilidad”, lo más alto de cada roca simboliza un óptimo local de estabilidad (hay que tener en cuenta que la parte superior de una roca pequeña se encuentra cerca del nivel de la arena). A cualquier secuencia de una proteína dada se le puede asignar un punto determinado sobre una roca (situado entre el nivel de la arena y la parte alta de la roca), cualquier mutación sencilla que se realice sobre dicha secuencia va a desplazar dicho punto sobre esa roca. Si el punto correspondiente a la proteína *WT* se encuentra cerca de la cima de la roca,

solamente algunas mutaciones individuales serán estabilizantes y en cualquier caso, estas mutaciones no aumentarán la estabilidad sobre el óptimo local (la cima de la roca). Desde luego, se pueden alcanzar incrementos de estabilidad mayores, pero el punto correspondiente en la roca necesitaría moverse a otra roca de mayor altura y ésto requeriría el efecto combinado de algunas mutaciones (lo que quiere decir que algunas de estas mutaciones serían desestabilizantes cada una de ellas considerada individualmente). Hoy día parece posible obtener una estabilidad elevada, mediante la generación *in vitro* de mutantes múltiples o mediante el uso de procedimientos computacionales de diseño de proteínas (Malakauskas y Mayo, 1998; Ibarra-Molero y Sanchez-Ruiz, 2002; Korkegian *et al.*, 2005).

4.8 CONCLUSIONES

En este trabajo hemos mostrado, de manera experimental experimental simple y clara, que los patrones de estabilidad de mutaciones en proteínas pueden reflejar una optimización estructural evolutiva. Pensamos que este tipo de patrones ocurren a menudo, pero no han sido detectados (por lo menos eso creemos) posiblemente porque normalmente no se realizan estudios exhaustivos de mutaciones conservativas que implican dos residuos similares (es decir, realizar por ejemplo todas las mutaciones $I \rightarrow V$ y $V \rightarrow I$ en una proteína). La existencia de estos patrones sugiere la posibilidad de racionalizar (y quizás predecir) los efectos de mutaciones sobre la estabilidad de proteínas en base a modelos evolutivos. Por “modelos evolutivos” nos referimos específicamente en este contexto, a modelos para los efectos de la mutación sobre la estabilidad en los que algunas de las características físicas que determinan los efectos de la mutación se evalúan desde el punto de vista de cómo el ambiente del residuo mutado ha sido seleccionado durante la evolución de proteínas. Para ilustrar este acercamiento e ilustrar unas pautas generales para su aplicación, hemos desarrollado aquí un modelo evolutivo que explica exitosamente el efecto de las mutaciones $I \leftrightarrow V$ sobre la estabilidad de tiorredoxina. Además, el modelo podría usarse potencialmente para predecir el efecto de mutaciones hidrofóbicas sobre otras proteínas; esta predicción requerirá un procedimiento simple para estimar la temperatura θ utilizando alineamientos de secuencias y la determinación del tipo de optimización de ambiente, realizando un análisis de la estructura nativa (hay que señalar que la optimización del

4. Un patrón de estabilidad de mutaciones hidrofóbicas...



Figura 4.20: Arriba, jardín de rocas en Zuiho-in Temple, Kyoto. Abajo, jardín de rocas en Tofuku-ji, Kyoto; ambos sacados de <http://phototravels.net/japan/photo-gallery/japanese-rock-gardens.html>.

4. Un patrón de estabilidad de mutaciones hidrofóbicas...

ambiente de una posición dada no puede deducirse con una certeza absoluta de la existencia del aminoácido presente en esa posición en la proteína de interés, porque, durante la evolución, los aminoácidos no restringidos por el ambiente se espera que sean aceptados ocasionalmente con probabilidades relacionadas a la penalización de las energías libres asociadas).

Sólo queda hacer algunos comentarios relativos al origen de la optimización energética de los ambientes de residuos en proteínas. Como ya hemos dicho anteriormente (Godoy-Ruiz *et al.*, 2004), la optimización puede ser una consecuencia de que la estabilidad de las proteínas es marginal (Taverna y Goldstein, 2000); aunque, además esta característica puede reflejar la selección natural de otras características físicas de proteínas asociadas de alguna forma a la estabilidad. Por ejemplo, una posibilidad interesante (aunque especulativa) es que la optimización de residuos se relaciona con el bajo nivel de frustración posiblemente requerido para la eficiencia del plegamiento (Wolynes *et al.*, 1995) y con el diseño evolutivo de la cooperatividad del plegamiento. Además, trabajos teóricos (Chan *et al.*, 2004) indican que para proteínas podría ser más difícil adquirir cooperatividad (y por lo tanto una barrera de plegamiento significativa) que una estructura nativa estable y, de hecho, proteínas diseñadas computacionalmente se ha visto que pliegan más rápidamente que las correspondientes naturales, aunque en el diseño de éstas no se incluya selección para una barrera de plegamiento baja (Scalley-Kim y Baker, 2004). Un análisis teórico de modelos de proteínas (Chan *et al.*, 2004) también muestra que los modelos con más interacciones específicas son más cooperativos y que la cooperatividad aumenta con el número creciente de letras en el alfabeto de aminoácidos usado.

Nuestros resultados muestran que los ambientes en proteínas se pueden optimizar con un grado remarcable de especificidad de residuos (hasta el punto de discriminar entre V e I o entre D y E). Ésto parece ser consistente con la visión sugerida por los análisis de modelos de polímeros y sugiere que la evolución puede actualmente estar tomando ventaja al alfabeto de 20 letras de aminoácidos naturales.

BIBLIOGRAFÍA

Baldwin, E., J. Xu, O. Hajiseyedjavadi, W.A. Baase and B.W. Matthews (1996). *J Mol Biol* **259**(3): 542-59.

Chan, H.S., S. Shimizu and H. Kaya (2004). *Methods Enzymol* **380**: 350-79.

Chothia, C. (1975). *Nature* **254**(5498): 304-8.

Dayhoff, M., R.M. Schwartz and B.C. Orcutt (1978). Atlas of Protein Sequence and Structure. M.e. Dayhoff. Silver Spring, MD, National Biomedical Research Foundation. **vol. 5 (3)**: 345-352.

Dokholyan, N.V. and E.I. Shakhnovich (2001). *J Mol Biol* **312**(1): 289-307.

Ewens, W.J. and G. G.R. (2001). New York, Springer.

George, D.G., W.C. Baker and L.T. Hunt (1990). *Methods Enzymol* **183**: 333-351.

Georgescu, R.E., M.M. Garcia-Mira, M.L. Tasayco and J.M. Sanchez-Ruiz (2001). *Eur J Biochem* **268**(5): 1477-85.

Godoy-Ruiz, R., R. Perez-Jimenez, B. Ibarra-Molero and J.M. Sanchez-Ruiz (2004). *J Mol Biol* **336**(2): 313-8.

Godoy-Ruiz, R., R. Perez-Jimenez, B. Ibarra-Molero and J.M. Sanchez-Ruiz (2005). *Biophys J* **89**(5): 3320-31.

Godoy-Ruiz, R., R. Perez-Jimenez, M.M. Garcia-Mira, I.M. Plaza del Pino and J.M. Sanchez-Ruiz (2005). *Biophys Chem* **115**(2-3): 263-6.

Hirsch, J.G., C.T. Nelson, M.W. Chase, H.L. Israel, R.A. Goldstein and L.E. Siltzbach (1971). *N Engl J Med* **284**(23): 1326-8.

Ibarra-Molero, B. and J.M. Sanchez-Ruiz (2002). *J. Phys. Chem. B* **106**(26): 6609-6613.

Knappenberger, J.E., Smith, J.E., Thorpe, S.H., Zitzewitz, J.Z. & Matthews, C.R. (2002). *J Mol Biol* **321**: 1-6.

Korkegian, A., M.E. Black, D. Baker and B.L. Stoddard (2005). *Science* **308**(5723): 857-60.

Ladbury, J.E., R. Wynn, H.W. Hellinga and J.M. Sturtevant (1993). *Biochemistry* **32**(29): 7526-30.

Lawrence, C.E., S.F. Altschul, M.S. Boguski, J.S. Liu, A.F. Neuwald and J.C. Wootton (1993). *Science* **262**(5131): 208-14.

4. Un patrón de estabilidad de mutaciones hidrofóbicas...

Lehmann, M., L. Pasamontes, S.F. Lassen and M. Wyss (2000). *Biochim Biophys Acta* **1543**(2): 408-415.

Loladze, V.V., D.N. Ermolenko and G.I. Makhatadze (2002). *J Mol Biol* **320**(2): 343-57.

Lumb, K.J. and P.S. Kim (1995). *Biochemistry* **34**(27): 8642-8.

Makhatadze, G.I. and P.L. Privalov (1990). *J Mol Biol* **213**(2): 375-84.

Malakauskas, S.M. and S.L. Mayo (1998). *Nat Struct Biol* **5**(6): 470-5.

Robic, S., M. Guzman-Casado, J.M. Sanchez-Ruiz and S. Marqusee (2003). *Proc Natl Acad Sci U S A* **100**(20): 11345-9.

Sanchez-Ruiz, J.M. (1995). New York, Plenum Press.

Scalley-Kim, M. and D. Baker (2004). *J Mol Biol* **338**(3): 573-83.

Shortle, D. (2003). *Protein Sci* **12**(6): 1298-1302.

Shrake, A. and J.A. Rupley (1973). *J Mol Biol* **79**(2): 351-71.

Takano, K., J.M. Scholtz, J.C. Sacchettini and C.N. Pace (2003). *J Biol Chem* **278**(34): 31790-5.

Taverna, D.M. and R.A. Goldstein (2000). *Biopolymers* **53**(1): 1-8.

Warren, J.E. and R.G. Gordon (1966). *J Phys Chem* **67**: 1524.

Wolynes, P.G., J.N. Onuchic and D. Thirumalai (1995). *Science* **267**(5204): 1619-20.

Yohannan, S., S. Faham, D. Yang, J.P. Whitelegge and J.U. Bowie (2004). *Proc Natl Acad Sci U S A* **101**(4): 959-63.

**5. Selección natural de la estabilidad
cinética: probable origen de las
correlaciones entre efectos de
mutación sobre la energética de las
proteínas y las frecuencias de
aparición de aminoácidos en
alineamientos de secuencias**

5 SELECCIÓN NATURAL DE LA ESTABILIDAD CINÉTICA: PROBABLE ORIGEN DE LAS CORRELACIONES ENTRE EFECTOS DE MUTACIÓN SOBRE LA ENERGÉTICA DE LAS PROTEÍNAS Y LAS FRECUENCIAS DE APARICIÓN DE AMINOÁCIDOS EN ALINEAMIENTOS DE SECUENCIAS

5.1 PLANTEAMIENTO INICIAL DEL PROBLEMA

Existen estudios que muestran las correlaciones entre los efectos sobre la estabilidad de proteínas y las frecuencias de aparición de aminoácidos en alineamientos de secuencias (Capítulos 3 y 4, publicados en Godoy-Ruiz *et al.*, 2004, 2005, respectivamente; Godoy-Ruiz *et al.*, 2006; Steipe *et al.*, 1994; Lehmann *et al.*, 2000; Campos *et al.*, 2004). De hecho, se han utilizado análisis estadísticos de los alineamientos de secuencias de manera exitosa para diseñar proteínas con una estabilidad mejorada (es lo que denominamos “concepto de consenso” para la estabilización de proteínas (Lehmann *et al.*, 2000)). Las relaciones existentes entre estabilidad-evolución pueden verse en términos de la llamada hipótesis de Boltzmann o hipótesis de pseudo-equilibrio (Dokholyan y Shakhnovich, 2001; Shortle, 2003; Godoy-Ruiz *et al.*, 2004, 2005). Cuando se aplica a mutaciones individuales, la hipótesis de pseudo-equilibrio implica que mutaciones neutras o cuasi-neutras se fijan en el curso de la evolución con probabilidades que reflejan aproximadamente los efectos de la mutación sobre la estabilidad. De acuerdo a esta consideración, el cociente entre el número de residuos de los tipos A y B en una posición dada en un alineamiento de secuencias debería de venir dado por una exponencial de Boltzmann del correspondiente efecto sobre la estabilidad sobre la energía libre de desplegamiento:

$$\frac{N_B}{N_A} \approx \exp\left(-\frac{\Delta\Delta G_{A \rightarrow B}}{R\theta}\right) \quad (5.1)$$

donde la “temperatura” θ se podría considerar en principio como una medida de la presión evolutiva sobre la estabilidad de la proteína (Shortle, 2003; Godoy-Ruiz *et al.*, 2004, 2005).

5. La selección natural para la estabilidad cinética...

Recientemente, hemos llevado a cabo un riguroso estudio de la aplicabilidad de la ecuación 5.1 (y consecuentemente, de la hipótesis de Boltzman) usando tiorredoxina de *E. coli* como sistema modelo (Godoy-Ruiz *et al.*, 2004, 2005; Godoy-Ruiz *et al.*, 2006). Hemos determinado el efecto de 27 mutaciones conservativas (del tipo $D \rightarrow E, E \rightarrow D, I \rightarrow V$ y $V \rightarrow I$) sobre la estabilidad termodinámica (energía libre de Gibbs de desplegamiento determinada por calorimetría diferencial de barrido) y hemos encontrado una correlación altamente significativa ($p=3 \cdot 10^{-8}$) entre $\Delta\Delta G_{A \rightarrow B}$ y $\ln(N_B/N_A)$, tal y como sugiere la ecuación 5.1. Además, la correlación es robusta, razón por la cual hemos encontrado correlaciones significativas cuando se ha realizado el cálculo de N para los subconjuntos de secuencias correspondientes a las tasas mayores representadas en el alineamiento (bacteria, proteobacteria, eucariota, metazoa; ver Capítulo 4 y Godoy-Ruiz *et al.*, 2005).

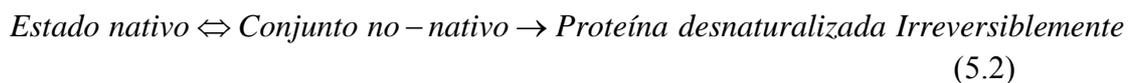
Un análisis de mecánica estadística simple indica que, para un sistema de moléculas independientes en equilibrio, se observa una distribución de Boltzmann para la energía de las moléculas cuando la energía total del sistema se mantiene esencialmente constante (véase capítulo 2 en Callen, 1985; van Holde *et al.*, 1998). Igualmente, para que se observe una distribución tipo Boltzmann para efectos de mutaciones independientes en tiempo evolutivo, la estabilidad debe estar restringida en un rango comparativamente estrecho (entre los niveles razonablemente bien definidos de estabilidad mínima y máxima (Dokholyan y Shakhnovich, 2001; Shortle, 2003). Un nivel de estabilidad máxima comparativamente bajo está posiblemente garantizado si se alcanzan mejoras moderadas de la estabilidad por acumulación de mutaciones independientes (Godoy-Ruiz *et al.*, 2005). Un nivel de estabilidad mínima comparativamente alto, podría ser la consecuencia de la selección natural si, por alguna razón, la supervivencia del organismo se ve comprometida por la baja estabilidad de la proteína. Por lo tanto, aquellas mutaciones que llevan la estabilidad de la proteína por debajo de un nivel mínimo no se aceptarán. Es decir, mutaciones fuertemente desestabilizantes se aceptarán raramente (sólo cuando un número suficiente de mutaciones estabilizantes se han aceptado previamente, de manera que la mutación desestabilizante no lleva la estabilidad de la proteína por debajo del nivel mínimo de estabilidad). Por otro lado, mutaciones moderadamente

5. La selección natural para la estabilidad cinética...

desestabilizantes (y estabilizantes) se aceptarán mucho más a menudo. Por lo tanto, la existencia de un nivel de estabilidad mínima determinado por la selección natural proporciona una explicación intuitiva de la relación estadística existente entre los efectos de las mutaciones sobre la estabilidad y la frecuencia de aparición de residuos en alineamientos de secuencias.

En este capítulo, primero usamos modelos simples para determinar el nivel de estabilidad mínima que es consistente con la correlación de efectos de mutaciones/alineamiento de secuencias, que previamente se han descrito para tiorredoxina. Sorprendentemente, hemos encontrado que el nivel de estabilidad mínima es sólo de unos pocos de kJ/mol menos que la estabilidad de tiorredoxina silvestre de *E. coli* (WT). Ésto es equivalente a un descenso de sólo algunos grados en la temperatura de desnaturalización en equilibrio (T_m) y el valor de T_m para la proteína WT es bastante alto: sobre 90 °C. Por lo tanto, parece razonable concluir que la selección natural no actúa directamente sobre la estabilidad termodinámica (por lo menos en este caso), pero sí sobre otro factor relacionado con la estabilidad termodinámica (por lo menos para las mutaciones que se han estudiado). Como exponemos a continuación, en este trabajo hemos explorado la posibilidad de que ese otro factor sea la estabilidad cinética de la proteína *in vivo*.

Estudios teóricos y un gran número de estudios experimentales sobre desnaturalización térmica irreversible de proteínas sugieren que la estabilidad termodinámica (valor positivo para la energía libre de desnaturalización) por sí sola no garantiza que la proteína permanezca en el estado nativo durante una cierta escala de tiempo, ya que las alteraciones irreversibles (incluso si ocurren en estados desnaturalizados poco poblados o estados parcialmente desplegados) pueden perjudicar el estado nativo de una forma dependiente del tiempo (Plaza del Pino *et al.*, 2000 y referencias dadas en este artículo). Esta situación se ilustra con el modelo “generalizado” de Lumry-Eyring:



donde el conjunto no-nativo incluye, por definición, los estados desplegados y parcialmente desplegados. La desnaturalización irreversible de proteínas (ocasionada por agregación, alteraciones químicas de los residuos, etc.) ocurre a

5. La selección natural para la estabilidad cinética...

menudo en experimentos *in vitro*. Sin embargo, la irreversibilidad es más propensa a ocurrir *in vivo* (Plaza del Pino *et al.*, 2000) principalmente por las siguientes razones:

a) Algunas alteraciones irreversibles de los estados no-nativos (agregación, interacciones no deseadas con otros componentes macromoleculares o proteólisis) se espera que sean rápidas en el ambiente celular o en el ambiente extracelular (para las proteínas secretadas).

b) La escala de tiempo *in vivo* (quizás relacionada con la vida media para la degradación de la proteína) se acepta que sea mayor que la escala de tiempo para los típicos experimentos de desnaturalización *in vitro* (minutos).

De acuerdo con las razones anteriormente dadas, se propuso en nuestro grupo de investigación (Plaza del Pino *et al.*, 2000) que muchas proteínas (¡aunque no necesariamente todas ellas!) son seleccionadas de forma natural para tener una estabilidad cinética significativa cuando son expuestas a los efectos destabilizantes de alteraciones irreversibles *in vivo*. De hecho, en los últimos años, se ha demostrado que la estabilidad de muchos sistemas complejos de proteínas de interés tiene un origen cinético; como ejemplos podemos encontrar el virus de la enfermedad de Newcastle (Schnyrov *et al.*, 1997), la proteasa α -lítica (Sohl *et al.*, 1998, Jaswal *et al.*, 2002), colágenos (Leikina *et al.*, 2002; Persikov y Brodsky, 2002), la dismutasa superóxido Cu/Zn (Lynch *et al.*, 2004; Manning y Colon, 2004), la proteína SHP de la cápsida vírica (Forrer *et al.*, 2004), la lipoproteína de baja densidad humana (Jayaraman *et al.*, 2005) y la lipasa de *Thermomyces lanuginosa* (Rodríguez-Larrea *et al.*, 2006). Además, la formación de fibras amiloides se puede considerar como una alteración irreversible de la proteína y, de hecho, algunos estudios sobre la inhibición de la amiloidogénesis se centran en el incremento de la estabilidad cinética del estado nativo de la proteína (Hammarström *et al.*, 2003; Petrassi *et al.*, 2005; Wiseman *et al.*, 2005), como se sugiere mediante un análisis con el modelo de Lumry-Eyring para la desnaturalización irreversible (Plaza del Pino *et al.*, 2000).

Tiorredoxina es una proteína muy conservada evolutivamente que participa en muchos procesos diferentes (reducción de puentes disulfuro, síntesis de ADN, reparación de proteínas, regulación de la apoptosis, etc; ver Tabla 5.1. De hecho, en un análisis proteómico reciente (Kumar *et al.*, 2004) de proteínas diana de tiorredoxina en *E. coli*, se identifican hasta 80 proteínas asociadas con

tiorredoxina y se sugiere la participación de tiorredoxina en al menos 26 procesos celulares distintos. Claramente, el estado nativo (funcional biológicamente) de tiorredoxina interacciona específicamente con algunos componentes macromoleculares diferentes pero, al mismo tiempo, la proteína necesita evitar interacciones no específicas de sus estados desplegados y parcialmente desplegados, además de otras alteraciones potencialmente irreversibles. Luego, parece que tiorredoxina puede ser un buen “candidato” de una proteína cinéticamente estable *in vivo*.

Una estabilidad cinética elevada (Plaza del Pino *et al.*, 2000) se puede alcanzar mediante una barrera de activación de desplegamiento suficientemente alta, que en un escenario de Lumry-Eyring (ecuación 5.2) se define como la barrera de energía libre que “separa” el estado nativo del conjunto de estados no-nativos. Las razones de la existencia de esta barrera energética de desplegamiento en estabilidad cinética son las siguientes:

1) Si las alteraciones irreversibles ocurren en estados no-nativos, el proceso irreversible global no puede ser más rápido que la velocidad de desplegamiento, sin importar cómo de rápido ocurren dichas alteraciones irreversibles; además, una velocidad de desplegamiento suficientemente baja proporciona un “mecanismo seguro” que garantiza una estabilidad cinética considerable en un ambiente agresivo (Plaza del Pino *et al.*, 2000).

2) Para una barrera de desplegamiento alta, los estados que forman la barrera tienen una energía libre elevada y no se encuentran significativamente poblados; como resultado, los procesos de alteración que ocurren en estos estados se encuentran desfavorecidos (Gruebele, 2005, Dobson, 1999).

En este capítulo mostramos que hay una diferencia fundamental entre los efectos de una mutación sobre la estabilidad termodinámica, descrita por la temperatura de desnaturalización en equilibrio (T_m) y los efectos de una mutación sobre la estabilidad cinética, descrita por la velocidad de desplegamiento. Además, la T_m cambia aproximadamente de forma lineal con $\Delta\Delta G$ de la mutación y un descenso de unos pocos kJ/mol en ΔG se traducen típicamente en un descenso de sólo algunos grados en T_m . Por otro lado, la velocidad de desplegamiento cambia de forma exponencial con el $\Delta\Delta G$ de la mutación y, consecuentemente, un descenso del orden de algunos kJ/mol lleva la vida media

de desplegamiento de una escala de tiempo que garantiza una estabilidad cinética elevada (por ejemplo, semanas) a una escala de tiempo menor (por ejemplo, horas o minutos) que podría comprometer la estabilidad cinética.

Tabla 5.1.: Funciones de tiorredoxina en distintos organismos (tomada de Elias y Holmgren, 2000).

Organismo	Función de tiorredoxina	Comentarios
Todos los organismos	- Síntesis de DNA.	- Tiorredoxina es un donador de hidrógenos para ribonucleótido-reductasa.
	- Reducción de proteínas con puentes disulfuro.	- Es fundamental para mantener los puentes disulfuro intracelulares reducidos.
Muchos organismos	- Reducción de H ₂ O ₂ .	- Muchas peroxiredoxinas, que catalizan la reducción de H ₂ O ₂ previniendo así el estrés oxidativo y la inducción de apoptosis, requieren la reducción mediante tiorredoxina.
	- Reparación de proteínas mediante reducción de sulfóxido de metionina.	- Tiorredoxina es un donador de hidrógeno para las reductasas sulfóxido de metionina.
Fagos de <i>E. coli</i> (T7, F1, M13)	- Subunidad de T7 ADN polimerasa.	- Es específica para tiorredoxina-(SH) ₂ pero no es dependiente de la actividad oxidoreductasa.
	- Participa en el ensamblaje del fago filamentoso.	- Es la única proteína requerida para el ensamblaje del fago y su exportación.
Bacterias y levaduras	- Donador de hidrógeno para la reductasa 3'-fosfoadenilsulfato (PAPS).	- Asimilación de sulfuro mediante la reducción de sulfato a sulfito.
Plantas	- Regulación de enzimas fotosintéticas del cloroplasto.	- Regulación de la fotosíntesis mediante la luz vía ferredoxina.
Mamíferos	- Regulación redox de factores de transcripción, p.e. NFkB,, AP-1.	- Diferentes factores de transcripción se encuentran regulados o inhibidos por Trx, que además puede ejercer diferentes actividades en el núcleo comparadas al citosol.
	- Regulación de la apoptosis.	- Tiorredoxina-(SH) ₂ (pero no tiorredoxina-S ₂) forma un complejo con ASK1 que previene la señalización para la apoptosis.

5. La selección natural para la estabilidad cinética...

- Inmunomodulación. - Tiorredoxina extracelular es una citoquina y una chemoquina, y una forma truncada simula a eosinófilis (células que provocan asma).
 - Embarazo. - La síntesis intracelular y extracelular de tiorredoxina desde los citotrofoblastos asiste la implantación y establecimiento del embarazo.
 - Nacimiento. - Protege de la hiperoxia en el nacimiento mediante la inducción de tiorredoxina.
 - SNC. - Tiorredoxina secretada por las células gliales promueve la supervivencia neuronal cuando ocurre isquemia/reperfusión.
-

Las observaciones realizadas apoyan nuestra hipótesis de que el nivel de estabilidad mínima (algunos kJ/mol por debajo de la estabilidad de tiorredoxina silvestre (*WT*)) consistente con la correlación entre efectos de mutación sobre la estabilidad y alineamientos de secuencias que previamente hemos descrito para tiorredoxina en los Capítulos 3 y 4 (Godoy-Ruiz *et al.*, 2004 y 2005, respectivamente), de hecho refleja selección natural para la estabilidad cinética. Para demostrar experimentalmente este argumento, hemos determinado experimentalmente las velocidades de plegamiento y desplegamiento en presencia de urea para tiorredoxina *WT* y 27 mutantes de tiorredoxina, los cuales también han sido estudiados en equilibrio por calorimetría diferencial de barrido. Los resultados obtenidos demuestran que tiorredoxina *WT* es una proteína estable cinéticamente, como indica la vida media del desplegamiento en agua (extrapolado a concentración 0 M urea) que se encuentra en torno a varias semanas. Sin embargo, hemos encontrado además que descensos moderados de la energía libre de activación originados por una mutación, llevan la vida media del desplegamiento a una escala mucho más corta. Finalmente, hemos encontrado una correlación robusta entre los efectos de mutaciones sobre la energía libre de activación del desplegamiento y las frecuencias de aparición de los residuos involucrados en un alineamiento de secuencias (dicha correlación, sin embargo, no se observa con la energía de activación del plegamiento). En resumen, estos resultados apoyan la hipótesis de que existe selección natural de la estabilidad cinética en tiorredoxina (y probablemente en otras proteínas también) y la

posibilidad de derivar información estructural y energética útil basada en la estabilidad cinética de proteínas, utilizando los alineamientos de secuencias.

5.2 DETERMINACIÓN DEL NIVEL DE ESTABILIDAD MÍNIMA PERMITIDA EVOLUTIVAMENTE PARA TIORREDOXINA

La Figura 5.1 muestra el efecto de 27 mutaciones conservativas (del tipo E→D, D→E, I→V y V→I) sobre la energía libre de desplegamiento de tiorredoxina de *E. coli*, determinada experimentalmente mediante calorimetría diferencial de barrido (ver Capítulos 3 y 4 publicados en Godoy-Ruiz *et al.*, 2004 y 2005, respectivamente). La mayoría de esas mutaciones son desestabilizantes consideradas en la dirección WT→mutante (como se muestra en la Figura 5.1). Por lo tanto, parece evidente que la elevada estabilidad de tiorredoxina sería poco probable sin la intervención de la selección natural.

En la Figura 5.2 se ilustra y explica el argumento realizado anteriormente. Se trata de una simulación de la estabilidad de la proteína realizando un simple Monte-Carlo. En cada etapa del Monte-Carlo, una de las 27 posiciones para las cuales disponemos de valores experimentales de $\Delta\Delta G$ (Figura 5.1), se selecciona aleatoriamente para ser mutada y el efecto sobre la estabilidad correspondiente se calcula como $\Delta\Delta G$ o como $-\Delta\Delta G$ (si la posición seleccionada había sido seleccionada en una etapa anterior un número impar de veces, es mutada al residuo existente en WT). La línea negra en la Figura 5.2-A muestra el caso en el que no se impone selección natural para la estabilidad en la simulación: las mutaciones desestabilizantes se acumulan rápidamente y la estabilidad tiene un valor promedio de 20 kJ/mol por debajo de la estabilidad de tiorredoxina WT, aunque se observa fluctuaciones grandes en la estabilidad (ver la distribución de probabilidades en la Figura 5.2-B). Hay que destacar que, para este caso, no existe preferencia estadística para los residuos estabilizantes en el curso de la evolución (simulada), como se indica por la representación esencialmente vertical de $\Delta\Delta G_{A\rightarrow B}$ versus $\ln(N_B/N_A)$ en la Figura 5.3-A.

Las líneas azul y roja en la Figura 5.2-A ilustran el caso en el cual se impone un mínimo de estabilidad en la simulación (-1 kJ/mol para la línea azul y -10 kJ/mol para la línea roja). Es decir, una mutación se considera perjudicial o neutra dependiendo si lleva a la proteína por debajo del nivel de estabilidad mínima permitida asumido; en tal caso, la mutación se rechaza en la simulación.

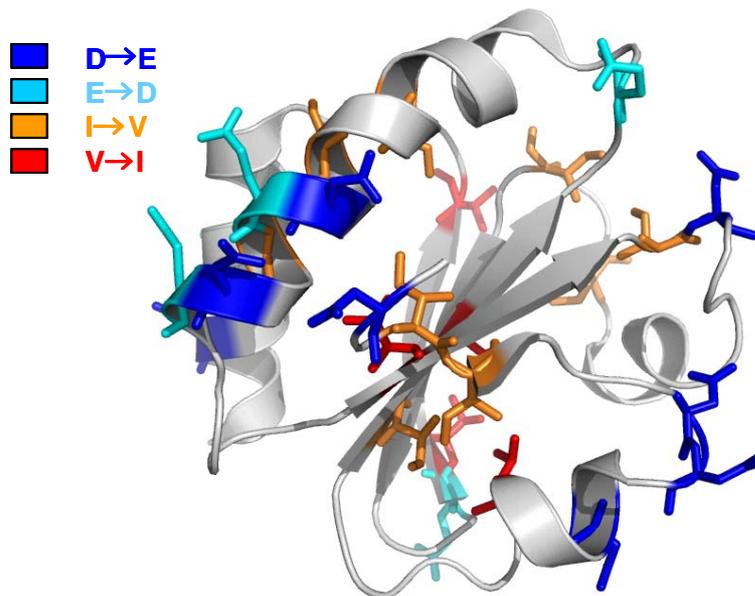
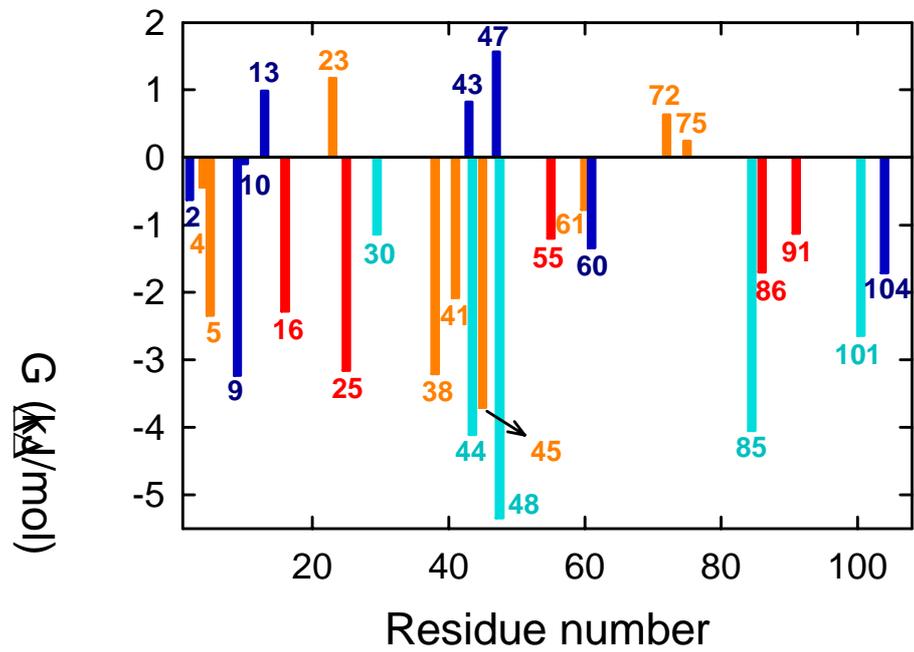


Figura 5.1: Efectos experimentales de las mutaciones sobre la estabilidad termodinámica (energía libre de desplegamiento) de tiorredoxina de *E. coli* a pH 7. Los datos se han tomado de los Capítulos 3 y 4 (Godoy-Ruiz *et al.*, 2004 y Godoy-Ruiz *et al.*, 2005, respectivamente) y derivan de los análisis de los termogramas de calorimetría diferencial de barrido. Se muestra el código de color usado para las distintas mutaciones y la ubicación de los diferentes reiduos mutados en la estructura de tiorredoxina (2TRX.pdb).

5. La selección natural para la estabilidad cinética...

Como se muestra en la Figura 5.2-A, la imposición de un mínimo de estabilidad permitido origina que la estabilidad fluctúe en rangos más estrechos.

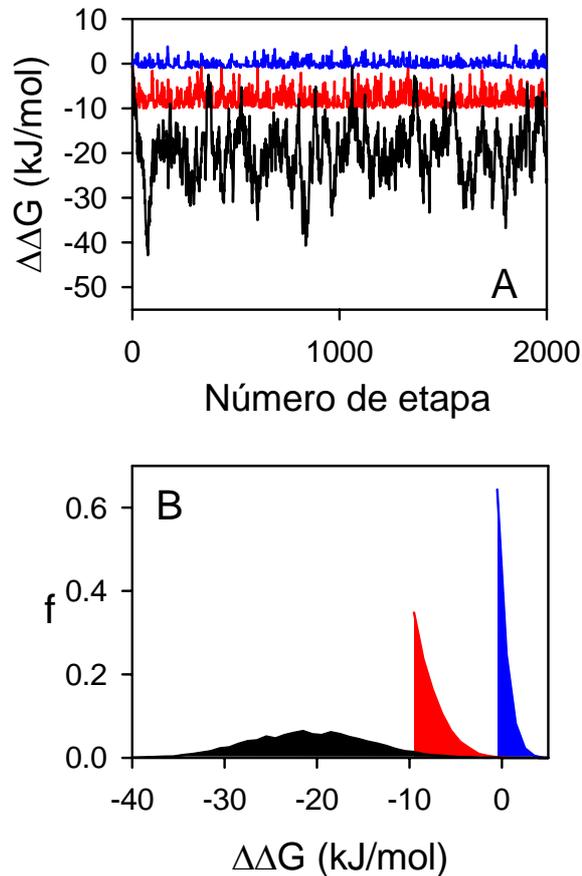


Figura 5.2: Simulaciones de Monte-Carlo de la evolución de la estabilidad de tiorredoxina. Las 27 posiciones mostradas en la Figura 5.1 se han mutado aleatoriamente y la estabilidad termodinámica (medida como la energía libre de desplegamiento tomando la proteína *WT* como referencia: $\Delta\Delta G = \Delta G - \Delta G(WT)$) se modifica de acuerdo a los valores experimentales de la mutación (Figura 5.1). A) Representación gráfica de $\Delta\Delta G$ versus el número de etapa de una simulación de Monte-Carlo en la que no se ha impuesto selección para la estabilidad (línea negra); simulación para la cual no se aceptan mutaciones que llevan el valor de $\Delta\Delta G$ por debajo de -10 kJ/mol (línea roja) o por debajo de -1 kJ/mol (línea azul). B) Probabilidad de distribución del $\Delta\Delta G$ correspondiente a las simulaciones mostradas en A). El rango de $\Delta\Delta G$ se ha dividido en intervalos de 1 kJ/mol y, para cada intervalo, el valor de f se ha calculado como el número de etapas para el cual el valor de $\Delta\Delta G$ se encuentra en el intervalo considerado, dividido por el número total de etapas en la simulación del Monte-Carlo (10000).

Además, ello ocasiona preferencias estadísticas significativas para los residuos estabilizantes como muestran las correlaciones de $\Delta\Delta G_{A \rightarrow B}$ versus $\ln(N_B/N_A)$ que se muestran en la Figura 5.3-A. Los valores de la temperatura θ en la Ecuación 5.1, se pueden derivar fácilmente de las pendientes de estas representaciones gráficas. En la Figura 5.3-B, mostramos la dependencia de los valores de θ calculados, con el mínimo de estabilidad usado en la simulación. Como el valor de θ obtenido de la correlación de los $\Delta\Delta G_{A \rightarrow B}$ experimentales de la Figura 5.1 se encuentra dentro del rango 120-211 K (dependiendo de la taxa considerada para el análisis; véase Capítulo 4 (Godoy-Ruiz *et al.*, 2005) para más detalles), la dependencia dada en la Figura 5.3-B sugiere que el mínimo de estabilidad evolutivamente permitido para tiorredoxina es del orden de unos pocos de kJ/mol por debajo de la estabilidad de la proteína *WT* de *E. coli*.

Los análisis mostrados en las Figuras 5.2 y 5.3 se confirman por los cálculos basados en los alineamientos de secuencias que se muestran en la Figura 5.4. El alineamiento (491 secuencias) se deriva de una búsqueda realizada con BLAST 2 en la base de datos UniProt/TrEMBL utilizando la proteína tiorredoxina de *E. coli* como secuencia a comparar (“secuencia *query*”; véase la sección 2.2.7 de *Materiales y métodos*). Para cada una de las secuencias, hemos calculado el valor de la energía libre de desplegamiento en base a los residuos presentes en las 27 mutaciones que se muestran en la Figura 5.1. Es decir, la estabilidad de *WT* se ha utilizado como referencia y se le asigna un valor cero; luego, si para una secuencia dada, ocurre alguna de las mutaciones que se muestran en la Figura 5.1, se corrige el valor de la energía libre de desplegamiento con el correspondiente valor experimental de $\Delta\Delta G$. Para ello, hemos considerado que el ambiente de las posiciones consideradas, y por lo tanto los valores de $\Delta\Delta G$, se encuentran significativamente conservados en el curso de la evolución de proteínas (Godoy-Ruiz *et al.*, 2005). En la Figura 5.4-A hemos representado las energías libres de desplegamiento calculadas frente a la similaridad entre secuencias y en la Figura 5.4-B, mostramos la correspondiente distribución de probabilidad. Se puede observar como para la mayoría de las secuencias los valores de energía libre de desplegamiento se encuentran en el rango existente entre 3 y -10 kJ/mol. Ésto

5. La selección natural para la estabilidad cinética...

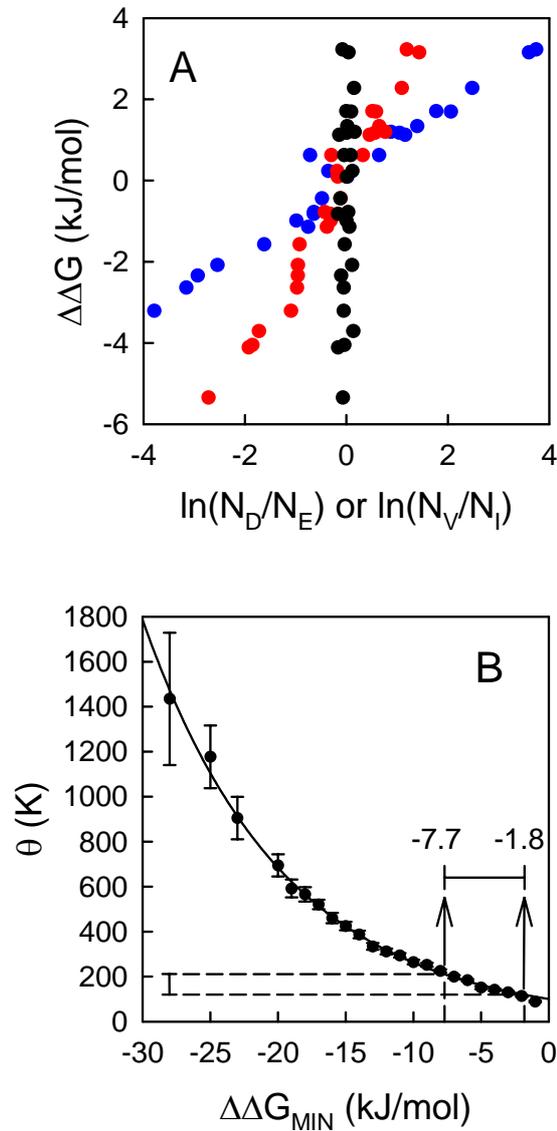


Figura 5.3: A) Correlación entre los efectos de la mutación sobre la estabilidad *versus* la frecuencia de aparición de aminoácidos (en las posiciones correspondientes) para las simulaciones de Monte-Carlo mostradas en la Figura 5.2 (el código de color es el mismo que el utilizado en la Figura 5.2). B) Valor de la temperatura θ (calculado de la pendiente de representaciones como las que se muestran en A)) (ver ecuación 5.1) para algunos valores de la estabilidad termodinámica mínima permitida en la simulación de Monte-Carlo ($\Delta\Delta G_{\text{MIN}}$). La línea continua no tiene significado teórico. El intervalo que se muestra en la esquina inferior izquierda representa el rango experimental de valores de θ para tiorredoxina, derivado de un análisis estadístico de estabilidad/alineamiento de secuencias (Godoy-Ruiz *et al.*, 2005). El rango de θ es consistente con los valores mínimos de estabilidad termodinámica (varios kJ/mol por debajo de la estabilidad de *WT*), calculados para tiorredoxina *WT* de *E. coli* ($\Delta\Delta G_{\text{MIN}}$ entre -2 kJ/mol y sobre -8 kJ/mol).

apoya nuevamente que el nivel de estabilidad mínima permitida evolutivamente para tiorredoxina de *E. coli* es del orden de unos pocos de kJ/mol por debajo de la estabilidad de la proteína *WT* de *E. coli* (algunos valores de energía libre de desplegamiento calculados que se muestran en la Figura 4-5-A son menores de -10 kJ/mol, pero esto puede ser debido a la aproximación realizada en el cálculo o, quizás, debido a la aparición de mutaciones estabilizantes en otras posiciones de la molécula que permiten que la contribución de las 27 posiciones vaya por debajo del límite de estabilidad mínima).

Las simulaciones de Monte-Carlo de la evolución de la estabilidad de tiorredoxina (Figura 5.2-A) se han realizado con el programa MUTRN4.exe (ficheros de entrada: $\Delta\Delta\text{GEQ.dat}$). La probabilidad de distribución del $\Delta\Delta\text{G}$ (Figura 5.2-B) correspondiente a las simulaciones que se muestran en la Figura 5.2-A, se han calculado STAT.exe (parámetros de entrada: límite de similaridad=0, número de etapas=10000, valor inferior de $\Delta\Delta\text{G}$ impuesto, número de intervalos=50, tamaño del intervalo=1). La correlación entre los efectos de la mutación sobre la estabilidad *versus* la frecuencia de aparición de aminoácidos en las posiciones estudiadas (Figura 5.3) para las simulaciones de Monte-Carlo mostradas en la Figura 5.2 también se ha obtenido con el programa MUTRN4.exe (utilizando los parámetros de entrada que se han dado anteriormente). El efecto de las 27 mutaciones, mostradas en la Figura 5.1, sobre la estabilidad termodinámica, calculado para todo el alineamiento y para las tasas de bacteria y proteobacteria (Figura 5.3) se ha calculado con los siguientes programas: en primer lugar, se ha utilizado el programa SEL-TAXA.exe (fichero de entrada: TRX25.ALN, 491 secuencias) para obtener el alineamiento de la taxa correspondiente, en segundo lugar, para cada conjunto de secuencias (alineamiento completo, bacteria, proteobacteria) se ha usado el programa ALIGN-G.exe (ficheros de entrada: TRX25.ALN (491 secuencias), ALNBAC.txt, ALNPRO.txt, respectivamente, y NDDGR.txt (fichero 27x3 que contiene los efectos de las 27 mutaciones sobre la estabilidad de tiorredoxina)). Como fichero de salida ALIGN-G.exe genera un fichero que contiene los efectos de las 27 posiciones sobre la estabilidad *versus* a la similaridad de cada una de las proteínas que constituyen el alineamiento utilizado. El programa FRODO.exe genera una estadística de $\Delta\Delta\text{G}$ de los alineamientos usados.

5. La selección natural para la estabilidad cinética...

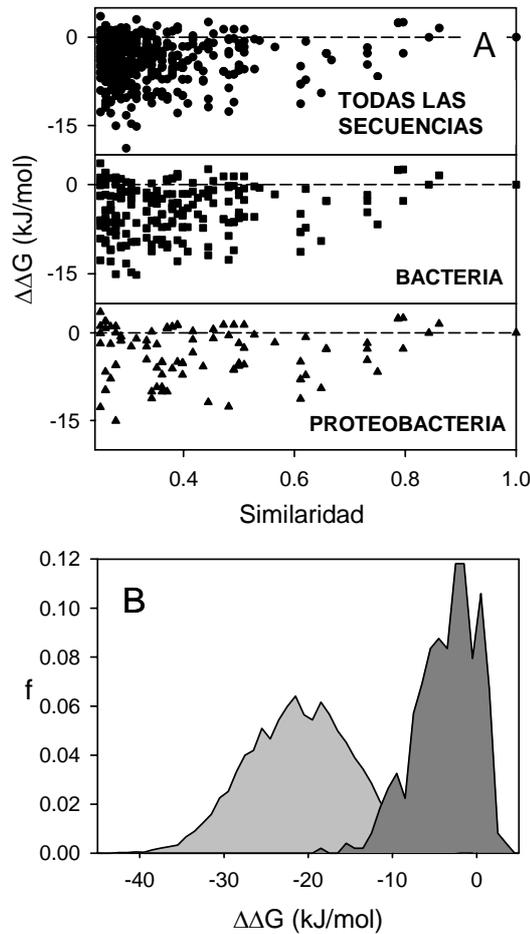


Figura 5.4: Efecto de las 27 mutaciones, mostradas en la Figura 5.1, sobre la estabilidad termodinámica calculada para las secuencias de un alineamiento, resultado de una búsqueda realizada con BLAST 2 en la base de datos UniProt/TrEMBL, utilizando la secuencia de tiorredoxina de *E. coli* como secuencia *query*. Para cada secuencia, la estabilidad se calcula asumiendo que los efectos de la mutación son los mismos que los de tiorredoxina de *E. coli* (Figura 5.1) y se considera como energía libre de desplegamiento, tomando la proteína *WT* como referencia: $\Delta\Delta G = \Delta G - \Delta G(WT)$. A) Representación de $\Delta\Delta G$ versus similaridad de secuencias con tiorredoxina de *E. coli* incluyendo las 491 secuencias del alineamiento; y la misma representación para las secuencias pertenecientes a la taxa de bacteria y proteobacteria. B) En gris oscuro, se muestra la probabilidad de distribución de $\Delta\Delta G$, derivada de los datos de estabilidad para todas las secuencias mostradas en A). El rango de $\Delta\Delta G$ se ha dividido en intervalos de 1 kJ/mol y, para cada intervalo dado, el valor de *f* se calcula como el número de secuencias para el cual el valor de $\Delta\Delta G$ se encuentra dentro del intervalo, dividido por el número total de secuencias (491). Por comparación, se muestra también la distribución derivada de una simulación de Monte-Carlo (gris claro) para la cual no se ha impuesto selección (simulación de color negro en Figura 5.2).

5.3 EFECTOS DE MUTACIONES *VERSUS* ESTABILIDAD CINÉTICA

Los análisis realizados en la sección anterior sugieren que, en el curso de la evolución de tiorredoxina, la selección natural elimina las variantes de tiorredoxina cuya estabilidad termodinámica (energía libre de desplegamiento) es sólo algunos kJ/mol menos que la de la proteína *WT* de *E. coli*. Sin embargo, una extrapolación directa de Gibbs-Helmholtz basada en los valores experimentales de la temperatura de desnaturalización, entalpía de desplegamiento y el cambio en la capacidad calorífica de desplegamiento (ver pie de Figura 5.5 para más detalles), da como resultado valores de ΔG entorno a 40 kJ/mol a 25 °C y a 37 °C para tiorredoxina *WT* de *E. coli*. Con una mutación (o una combinación de mutaciones) que cause un descenso de algunos kJ/mol, aún seguirían existiendo valores de ΔG de desplegamiento positivos significativos a temperatura fisiológica, sin comprometer en principio el plegamiento de la proteína. Probablemente, ésto se puede entender más fácilmente en términos de la temperatura de desnaturalización termodinámica (T_m , temperatura a la cual $\Delta G=0$). Un simple cálculo de Gibbs-Helmholtz (Figura 5.5) muestra que T_m cambia casi linealmente con $\Delta\Delta G$ y que un descenso de algunos kJ/mol en ΔG causa un descenso de algunos grados en T_m . Como la T_m para tiorredoxina *WT* de *E. coli* es próxima a 90°C, un descenso de algunos kJ/mol en ΔG deja a la T_m bastantes grados por encima de la temperatura fisiológica.

Consideremos ahora la estabilidad cinética, determinada por la velocidad de desplegamiento. Suponiendo que la teoría del estado de transición es aplicable, entonces:

$$\tau = \frac{1}{k_{UNF}} \propto \exp\left(\frac{\Delta G_{UNF}^\ddagger}{RT}\right) \quad (5.3)$$

donde k_{UNF} es la constante de velocidad de desplegamiento de primer orden, τ es el correspondiente tiempo de vida media y ΔG_{UNF}^\ddagger es la energía libre de activación de desplegamiento. La ecuación 5.3 indica que τ decrece con la energía libre de activación de una manera exponencial (Figura 5.5-B). Como resultado, un descenso de algunos kJ/mol en ΔG_{UNF}^\ddagger puede producir un gran cambio en τ (por

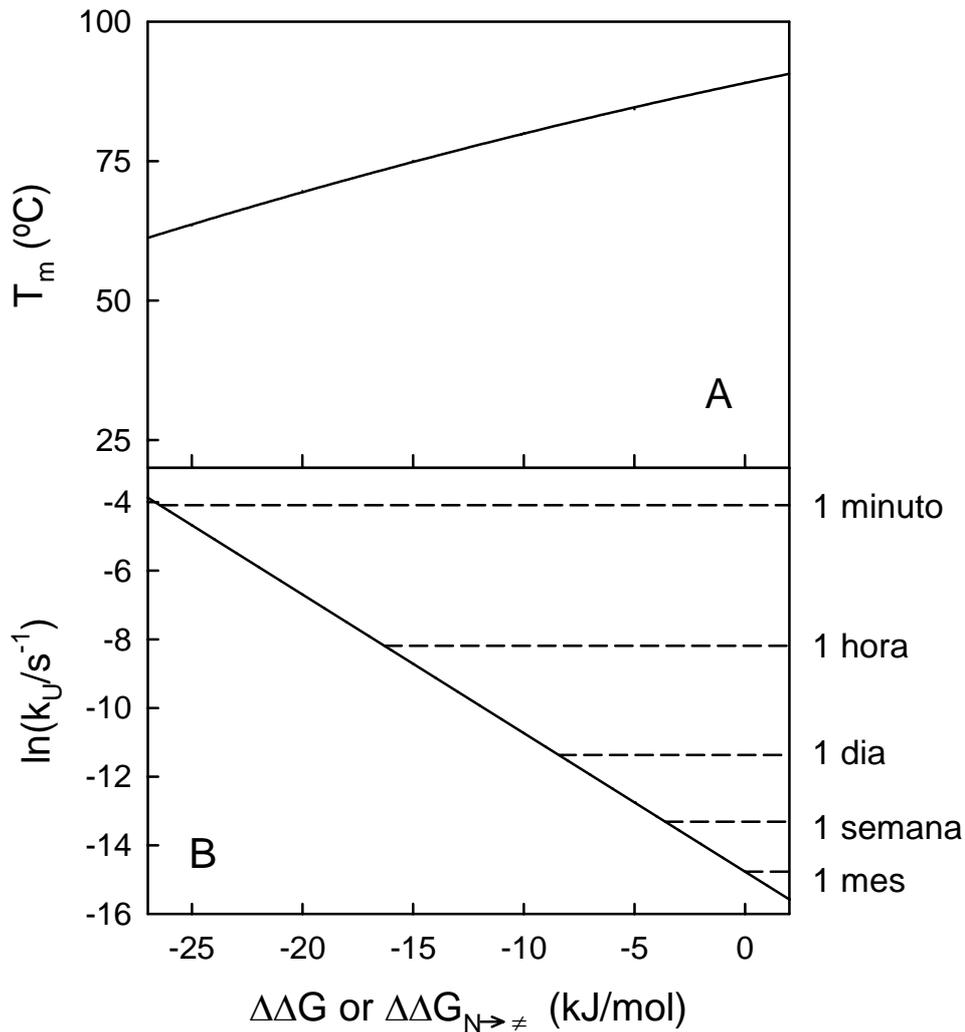


Figura 5.5: Efectos de mutación sobre la estabilidad termodinámica *versus* efectos de mutación sobre estabilidad cinética. A) Dependencia de la temperatura de desnaturalización en equilibrio (T_m) con los efectos de la mutación sobre el cambio de la energía libre ($\Delta\Delta G$). Los valores de T_m se calculan resolviendo numéricamente la ecuación de Gibbs-Helmholtz usando los parámetros termodinámicos para tiorredoxina WT de *E. coli* ($T_m(\text{WT})=89.05^\circ\text{C}$, $\Delta H=430$ kJ/mol, $\Delta C_p=5.8$ kJ·mol⁻¹·K⁻¹, valores tomados de Godoy-Ruiz *et al.*, 2004 y Godoy-Ruiz *et al.*, 2005). B) Dependencia de la constante de velocidad de desplegamiento con los efectos de la mutación sobre la energía libre de activación ($\Delta\Delta G_{N \rightarrow \neq}$). Los valores de la constante de velocidad se calculan usando la ecuación 5.3. A modo ilustrativo, en este cálculo, se ha considerado la constante de velocidad para la proteína WT ($\Delta\Delta G=0$) igual a $3.85 \cdot 10^{-7}$ s⁻¹, que corresponde a un tiempo de vida media de 1 mes. El eje de la derecha indica los valores de las constantes de velocidad correspondientes a tiempos de vida media representativos.

ejemplo, puede pasar de un valor del orden de algunas semanas a un valor de algunas horas o incluso minutos: Figura 5.5-B).

Los cálculos que se muestran en la Figura 5.5 son sólo a modo de ilustración, ya que no conocemos los valores la temperatura de desnaturalización en equilibrio o de la constante de velocidad de desplegamiento para tiorredoxina de *E. coli in vivo*. La dependencia exponencial del tiempo de vida media con ΔG_{UNF}^\ddagger apoya nuestro argumento de que, incluso descensos moderados en la energía libre de activación del desplegamiento podrían comprometer potencialmente la estabilidad cinética *in vivo*. De acuerdo con esto, la selección natural para la estabilidad cinética puede explicar una estabilidad mínima permitida, que oscilaría sobre algunos kJ/mol por debajo de la de la proteína *WT* de *E. coli*. Esta interpretación sería sólo válida si los efectos de las mutaciones sobre la energía libre de activación del desplegamiento reflejaran los efectos correspondientes sobre la estabilidad termodinámica. El estudio cinético descrito en la siguiente sección muestra como de hecho, esto es lo que ocurre para el caso de las 27 mutaciones conservadas que ilustran la Figura 5.1.

5.4 ESTUDIO EXPERIMENTAL DE LAS VELOCIDADES DE PLEGAMIENTO Y DESPLEGAMIENTO DE TIORREDOXINA DE *E. COLI* WT Y 27 MUTANTES SIMPLES

Hemos obtenido de manera muy precisa los perfiles del desplegamiento inducido por urea en equilibrio determinados mediante ensayos del desplegamiento de doble-salto y las velocidades de plegamiento y desplegamiento mediante análisis globales de las representaciones en *Chevron* (logaritmo de la constante de velocidad *versus* concentración de urea).

5.4.1 DESNATURALIZACIÓN EN EQUILIBRIO DE TIORREDOXINA ESTUDIADA POR ENSAYOS DE DESPLEGAMIENTO DE DOBLE SALTO

La preparación de tiorredoxina *WT* de *E. coli* y de las distintas formas variantes se ha explicado anteriormente en las secciones 3.1.1. y 4.1.1. Las mutaciones deseadas se diseñaron como se ha explicado en la sección 2.2.1.2 y se obtuvieron mediante mutagénesis dirigida (véase para más detalles la sección 2.2.1.3). Para este tipo de mutaciones se ha trabajado con el gen de tiorredoxina *WT* insertado en el vector pET30a(+) (sección 2.2.2.1) y las células

supercompetentes BL21-DE3. La expresión y purificación de proteína se ha explicado detalladamente en las secciones 2.2.2.5 y 2.2.2.6, respectivamente. La pureza de las proteínas purificadas se ha evaluado por electroforesis (PAGE-SDS).

Los ensayos de desplegamiento de doble salto se han utilizado para determinar la cantidad de proteína presente en el estado nativo bajo unas ciertas condiciones (concentración de urea) (Mucke y Schmid, 1994; Ibarra-Molero y Sanchez-Ruiz, 1996, 1997) mediante la observación de las cinéticas de desplegamiento que tienen lugar al transferir la proteína a condiciones fuertemente desnaturizantes. Bajo estas condiciones, los estados intermedios (estados parcialmente plegados) se espera que desnaturalicen mucho más rápido que los estados nativos. Luego, es posible seguir independientemente la cantidad de estado nativo y determinar la cantidad de estado nativo en la solución original por la amplitud de las cinéticas exponenciales observadas (Ibarra-Molero y Sanchez-Ruiz, 1996, 1997). En los ensayos de desplegamiento realizados, tiorredoxina (con una concentración próxima a 10 mg/mL) se incubado en soluciones de urea (Hepes 5 mM, pH 7 preparadas como se explica en la sección 2.2.3.2 de *Materiales y métodos*) a 25°C el tiempo suficiente, hasta asegurarnos que se había establecido el equilibrio de desplegamiento. Seguidamente, se llevaban a cabo las cinéticas de desplegamiento mediante dilución de las muestras de proteína incubadas en soluciones de urea, hasta una concentración final de proteína 0.04 mg/mL, en disoluciones de alta concentración de urea (sobre 8.5 M) en tampón Hepes 5 mM, pH 7 (sección 2.2.6.4 de *Materiales y métodos*). Bajo estas condiciones, el tiempo de vida media de la desnaturalización del estado nativo es de varios minutos. Las cinéticas de desnaturalización se monitorizaron siguiendo la dependencia de la emisión de fluorescencia a 350 nm después de haber excitado a 276 nm. Los perfiles de intensidad de fluorescencia (I) versus tiempo (Figuras 5.6 a 5.32) dan ajustes excelentes a la ecuación de primer orden 2.40 [$I = I_{\infty} - \Delta I \cdot e^{(-t/\tau)}$] (sección 2.2.6.2 de *Materiales y métodos*). Los perfiles de desnaturalización inducida por urea en el equilibrio, obtenidos usando ensayos de desplegamiento, se han descrito adecuadamente utilizando la ecuación 2.53

[$\Delta I = \frac{\Delta I^0}{1 + e^{m_{1/2}(C-C_{1/2})/RT}}$] (sección 2.2.6.4 de *Materiales y métodos*). El perfil de

desnaturalización inducida por urea en el equilibrio, determinado mediante ensayos de desplegamiento para tiorredoxina WT se muestra en la Figura 2.18 en la sección 2.2.6.4 de *Materiales y métodos*.

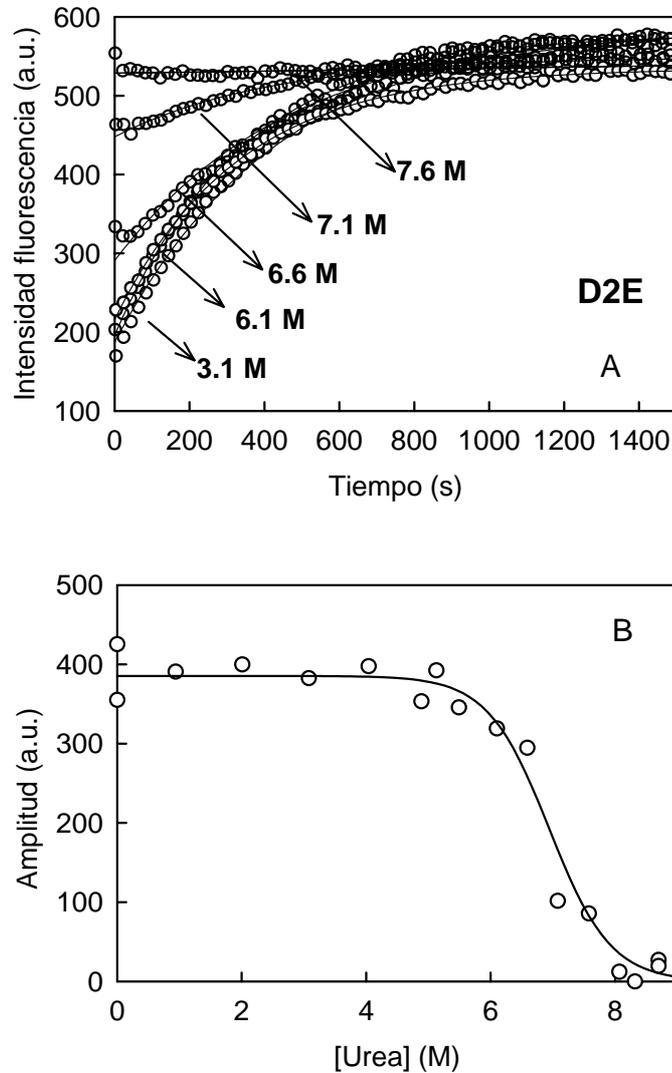


Figura 5.6: Determinación del perfil de desplegamiento inducido por urea en equilibrio, mediante ensayos de desplegamiento de doble salto, para la variante D2E de tiorredoxina de *E. coli*. A) Perfiles de intensidad de fluorescencia *versus* tiempo más representativos. Para registrarlos se ha transferido la variante de tiorredoxina (previamente incubada a 25°C y a las concentraciones de urea mostradas) hasta condiciones fuertemente desnaturizantes (8.5 M). La línea continua representa el mejor ajuste por mínimos cuadrados de la ecuación 2.40 a los datos experimentales (círculos). B) Amplitudes de desplegamiento calculadas del ajuste de la ecuación 2.40 a las cinéticas de desplegamiento mostradas en A. Estas amplitudes son proporcionales a la cantidad de proteína en estado nativo en la solución inicial y se representan frente a la concentración de urea en esas soluciones. La línea continua es el perfil predicho para la ecuación 2.53, con los parámetros dados en la Tabla 5.2.

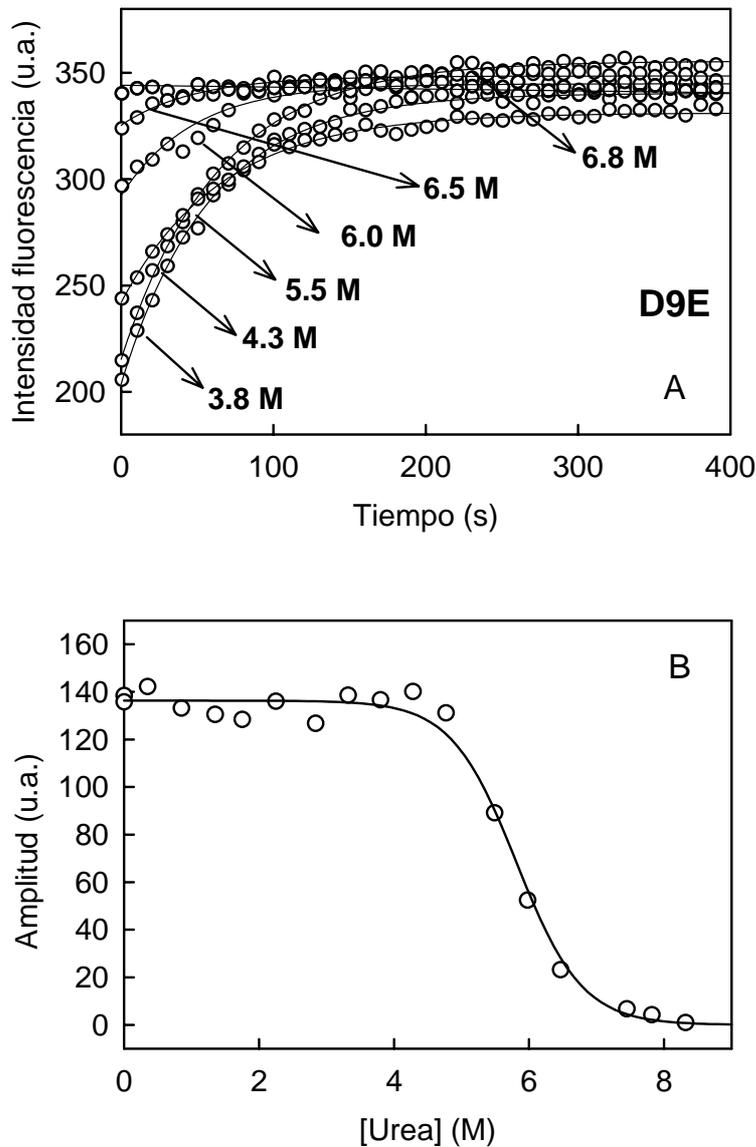


Figura 5.7: Determinación del perfil de desplegamiento inducido por urea en equilibrio, mediante ensayos de desplegamiento de doble salto, para la variante D9E de tiorredoxina de *E. coli*. A) Perfiles de intensidad de fluorescencia *versus* tiempo más representativos. Para registrarlos se ha transferido la variante de tiorredoxina (previamente incubada a 25°C y a las concentraciones de urea mostradas) hasta condiciones fuertemente desnaturizantes (8.5 M). La línea continua representa el mejor ajuste por mínimos cuadrados de la ecuación 2.40 a los datos experimentales (círculos). B) Amplitudes de desplegamiento calculadas del ajuste de la ecuación 2.40 a las cinéticas de desplegamiento mostradas en A. Estas amplitudes son proporcionales a la cantidad de proteína en estado nativo en la solución inicial y se representan frente a la concentración de urea en esas soluciones. La línea continua es el perfil predicho para la ecuación 2.53, con los parámetros dados en la Tabla 5.2.

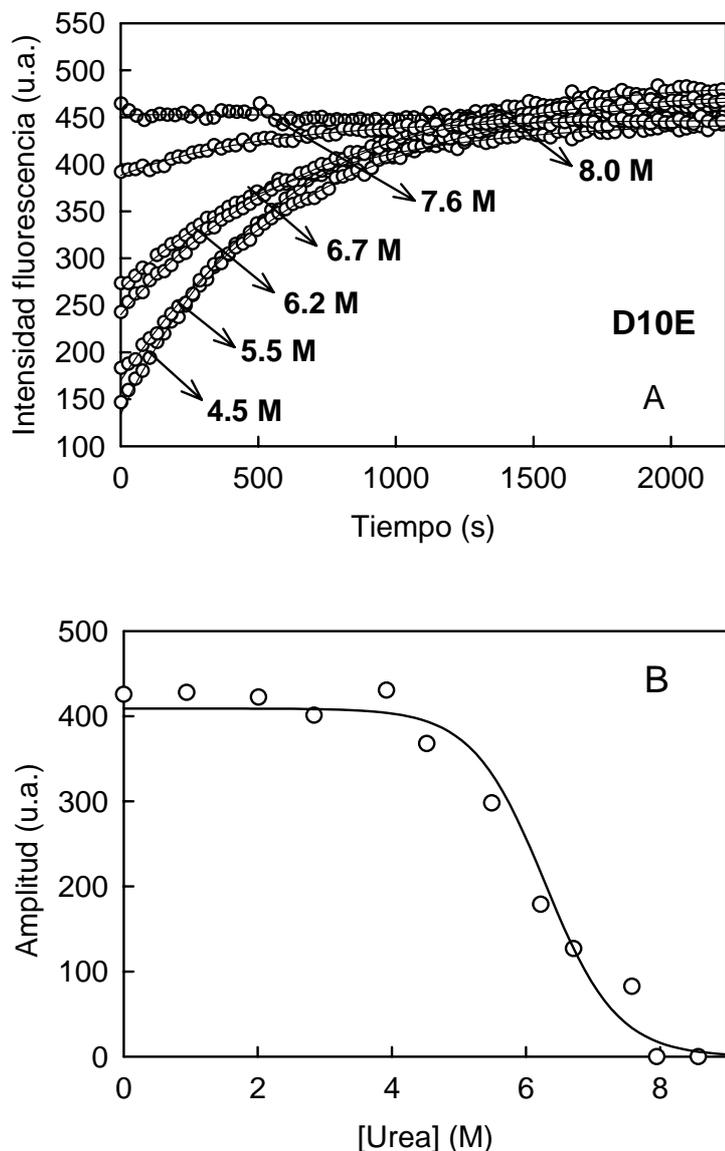


Figura 5.8: Determinación del perfil de desplegamiento inducido por urea en equilibrio, mediante ensayos de desplegamiento de doble salto, para la variante D2E de tiorredoxina de *E. coli*. A) Perfiles de intensidad de fluorescencia *versus* tiempo más representativos. Para registrarlos se ha transferido la variante de tiorredoxina (previamente incubada a 25°C y a las concentraciones de urea mostradas) hasta condiciones fuertemente desnaturalizantes (8.5 M). La línea continua representa el mejor ajuste por mínimos cuadrados de la ecuación 2.40 a los datos experimentales (círculos). B) Amplitudes de desplegamiento calculadas del ajuste de la ecuación 2.40 a las cinéticas de desplegamiento mostradas en A. Estas amplitudes son proporcionales a la cantidad de proteína en estado nativo en la solución inicial y se representan frente a la concentración de urea en esas soluciones. La línea continua es el perfil predicho para la ecuación 2.53, con los parámetros dados en la Tabla 5.2.

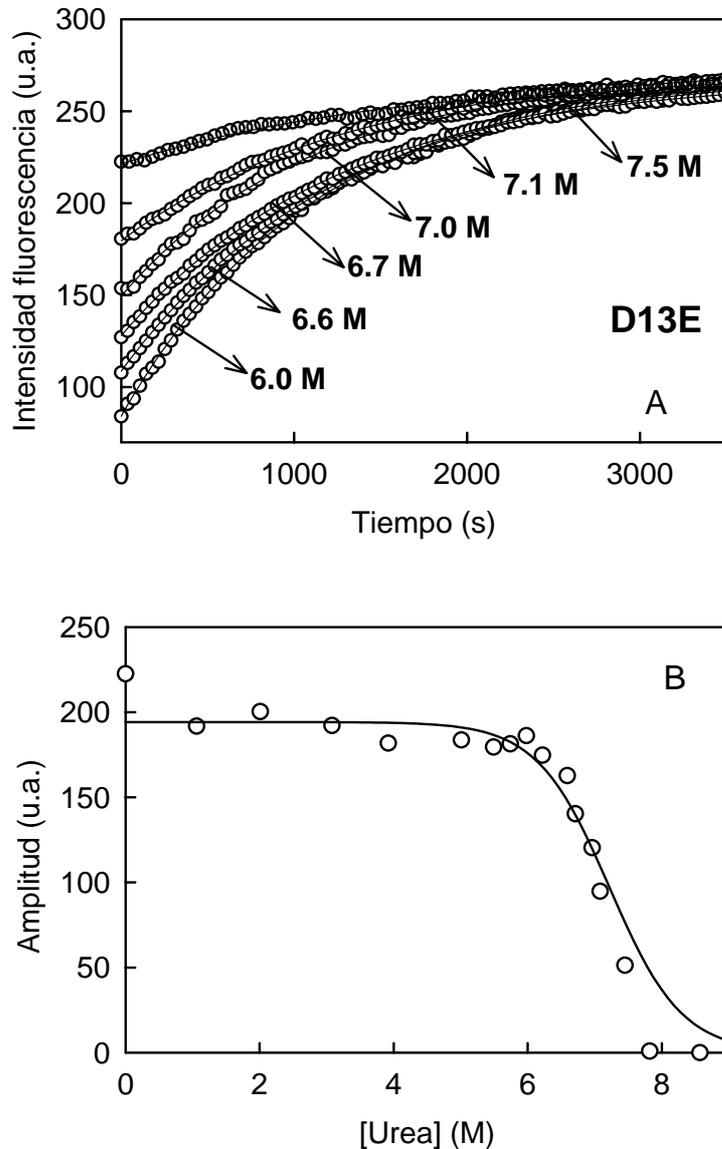


Figura 5.9: Determinación del perfil de desplegamiento inducido por urea en equilibrio, mediante ensayos de desplegamiento de doble salto, para la variante D13E de tiorredoxina de *E. coli*. A) Perfiles de intensidad de fluorescencia *versus* tiempo más representativos. Para registrarlos se ha transferido la variante de tiorredoxina (previamente incubada a 25°C y a las concentraciones de urea mostradas) hasta condiciones fuertemente desnaturalizantes (8.5 M). La línea continua representa el mejor ajuste por mínimos cuadrados de la ecuación 2.40 a los datos experimentales (círculos). B) Amplitudes de desplegamiento calculadas del ajuste de la ecuación 2.40 a las cinéticas de desplegamiento mostradas en A. Estas amplitudes son proporcionales a la cantidad de proteína en estado nativo en la solución inicial y se representan frente a la concentración de urea en esas soluciones. La línea continua es el perfil predicho para la ecuación 2.53, con los parámetros dados en la Tabla 5.2.

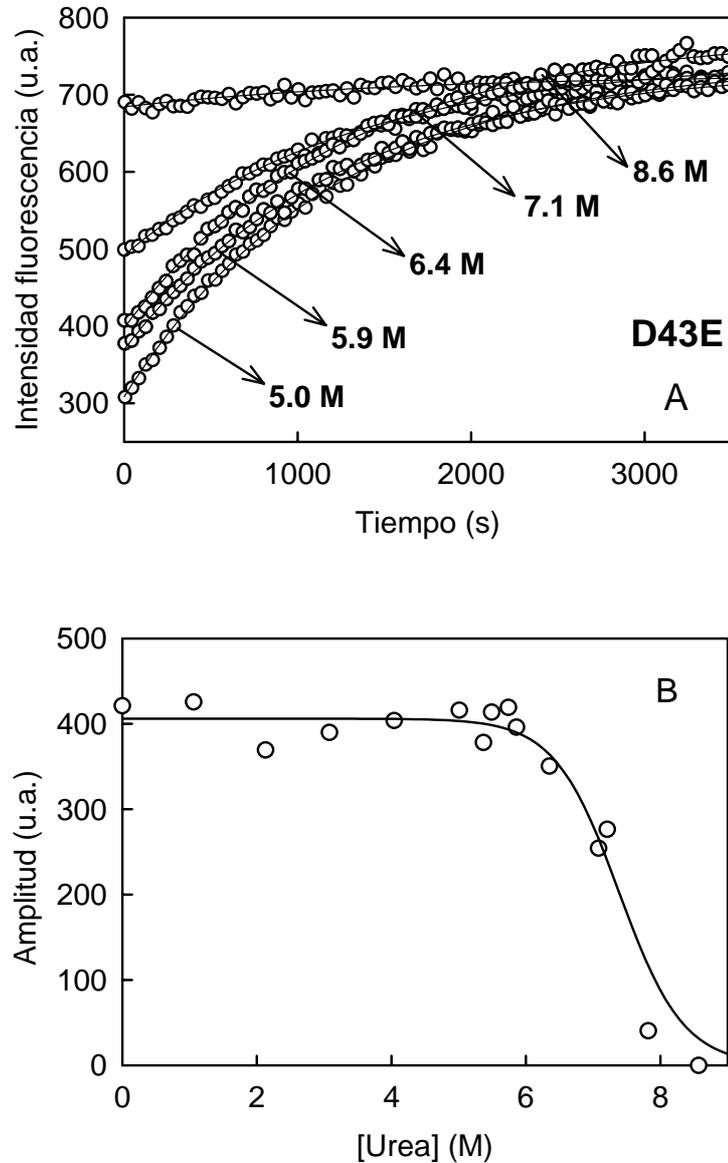


Figura 5.10: Determinación del perfil de desplegamiento inducido por urea en equilibrio, mediante ensayos de desplegamiento de doble salto, para la variante D43E de tiorredoxina de *E. coli*. A) Perfiles de intensidad de fluorescencia *versus* tiempo más representativos. Para registrarlos se ha transferido la variante de tiorredoxina (previamente incubada a 25°C y a las concentraciones de urea mostradas) hasta condiciones fuertemente desnaturizantes (8.5 M). La línea continua representa el mejor ajuste por mínimos cuadrados de la ecuación 2.40 a los datos experimentales (círculos). B) Amplitudes de desplegamiento calculadas del ajuste de la ecuación 2.40 a las cinéticas de desplegamiento mostradas en A. Estas amplitudes son proporcionales a la cantidad de proteína en estado nativo en la solución inicial y se representan frente a la concentración de urea en esas soluciones. La línea continua es el perfil predicho para la ecuación 2.53. con los parámetros dados en la Tabla 5.2.

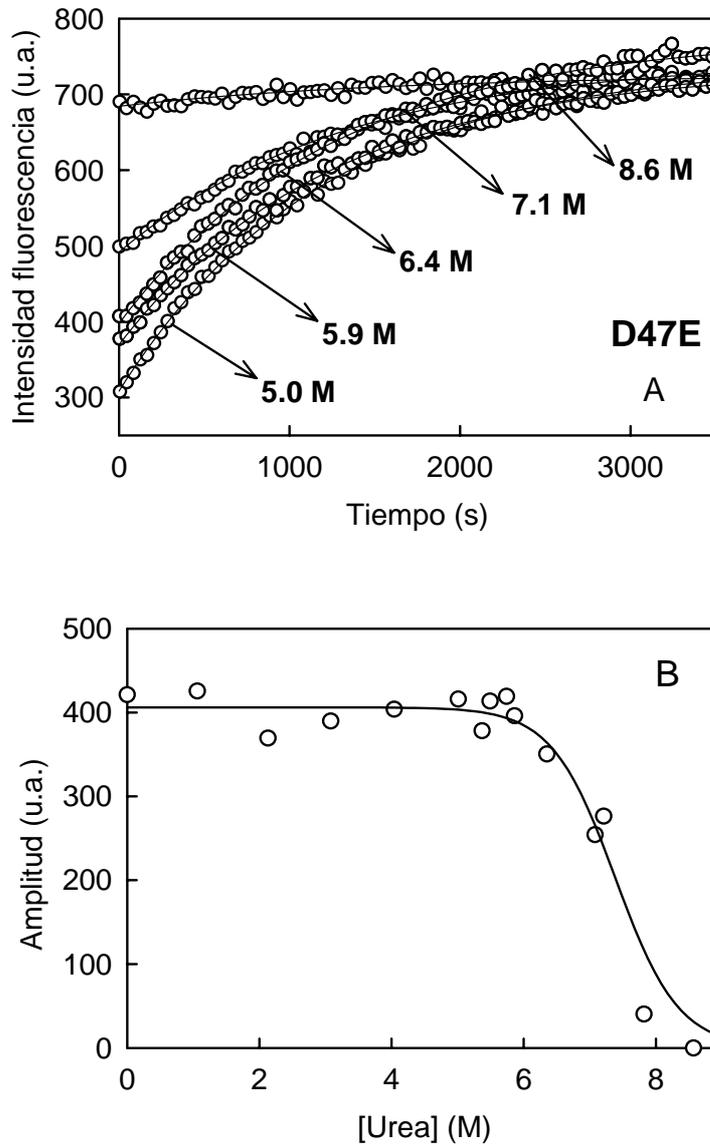


Figura 5.11: Determinación del perfil de desplegamiento inducido por urea en equilibrio, mediante ensayos de desplegamiento de doble salto, para la variante D47E de tiorredoxina de *E. coli*. A) Perfiles de intensidad de fluorescencia *versus* tiempo más representativos. Para registrarlos se ha transferido la variante de tiorredoxina (previamente incubada a 25°C y a las concentraciones de urea mostradas) hasta condiciones fuertemente desnaturalizantes (8.5 M). La línea continua representa el mejor ajuste por mínimos cuadrados de la ecuación 2.40 a los datos experimentales (círculos). B) Amplitudes de desplegamiento calculadas del ajuste de la ecuación 2.40 a las cinéticas de desplegamiento mostradas en A. Estas amplitudes son proporcionales a la cantidad de proteína en estado nativo en la solución inicial y se representan frente a la concentración de urea en esas soluciones. La línea continua es el perfil predicho para la ecuación 2.53. con los parámetros dados en la Tabla 5.2.

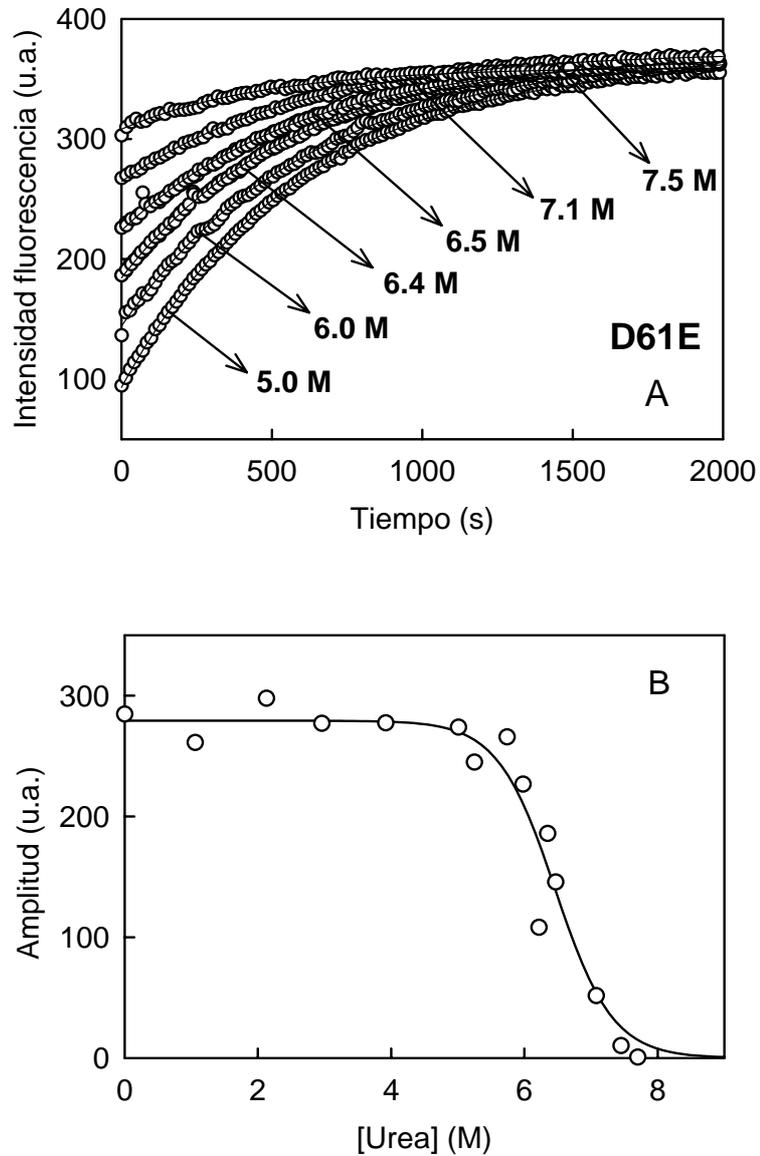


Figura 5.12: Determinación del perfil de desplegamiento inducido por urea en equilibrio, mediante ensayos de desplegamiento de doble salto, para la variante D61E de tiorredoxina de *E. coli*. A) Perfiles de intensidad de fluorescencia *versus* tiempo más representativos. Para registrarlos se ha transferido la variante de tiorredoxina (previamente incubada a 25°C y a las concentraciones de urea mostradas) hasta condiciones fuertemente desnaturizantes (8.5 M). La línea continua representa el mejor ajuste por mínimos cuadrados de la ecuación 2.40 a los datos experimentales (círculos). B) Amplitudes de desplegamiento calculadas del ajuste de la ecuación 2.40 a las cinéticas de desplegamiento mostradas en A. Estas amplitudes son proporcionales a la cantidad de proteína en estado nativo en la solución inicial y se representan frente a la concentración de urea en esas soluciones. La línea continua es el perfil predicho para la ecuación 2.53, con los parámetros dados en la Tabla 5.2.

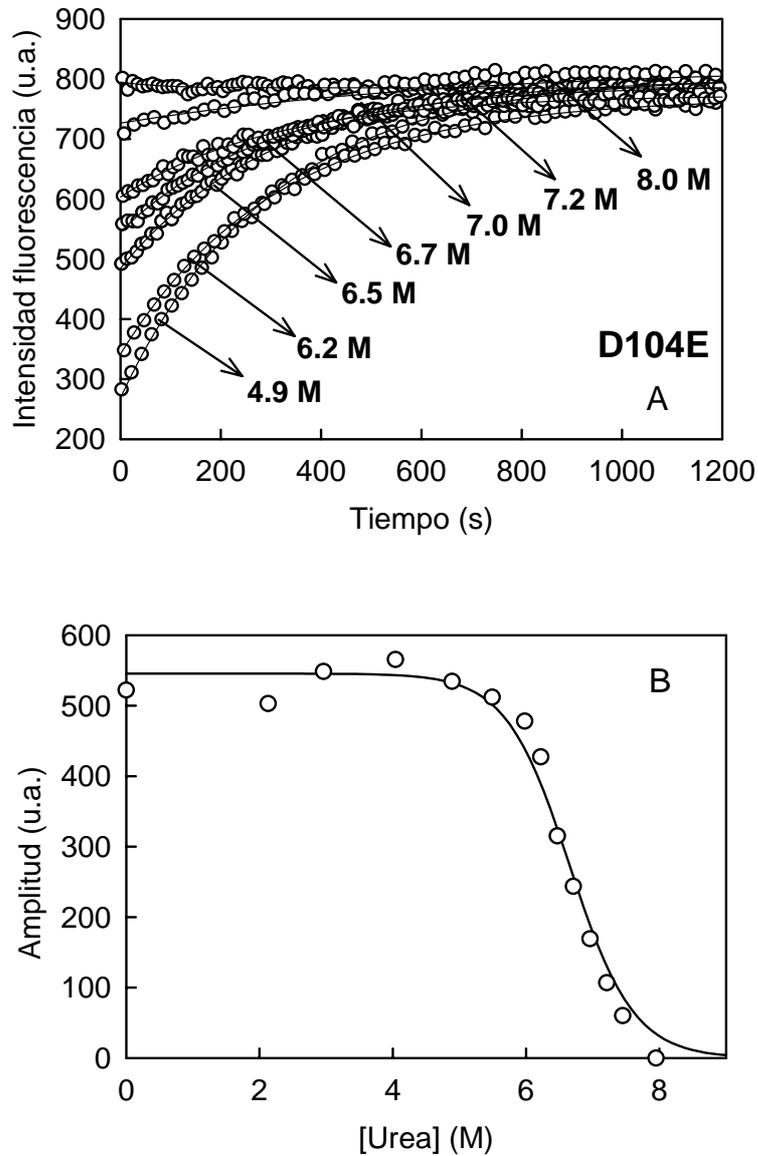


Figura 5.13: Determinación del perfil de desplegamiento inducido por urea en equilibrio, mediante ensayos de desplegamiento de doble salto, para la variante D104E de tiorredoxina de *E. coli*. A) Perfiles de intensidad de fluorescencia *versus* tiempo más representativos. Para registrarlos se ha transferido la variante de tiorredoxina (previamente incubada a 25°C y a las concentraciones de urea mostradas) hasta condiciones fuertemente desnaturizantes (8.5 M). La línea continua representa el mejor ajuste por mínimos cuadrados de la ecuación 2.40 a los datos experimentales (círculos). B) Amplitudes de desplegamiento calculadas del ajuste de la ecuación 2.40 a las cinéticas de desplegamiento mostradas en A. Estas amplitudes son proporcionales a la cantidad de proteína en estado nativo en la solución inicial y se representan frente a la concentración de urea en esas soluciones. La línea continua es el perfil predicho para la ecuación 2.53, con los parámetros dados en la Tabla 5.2.

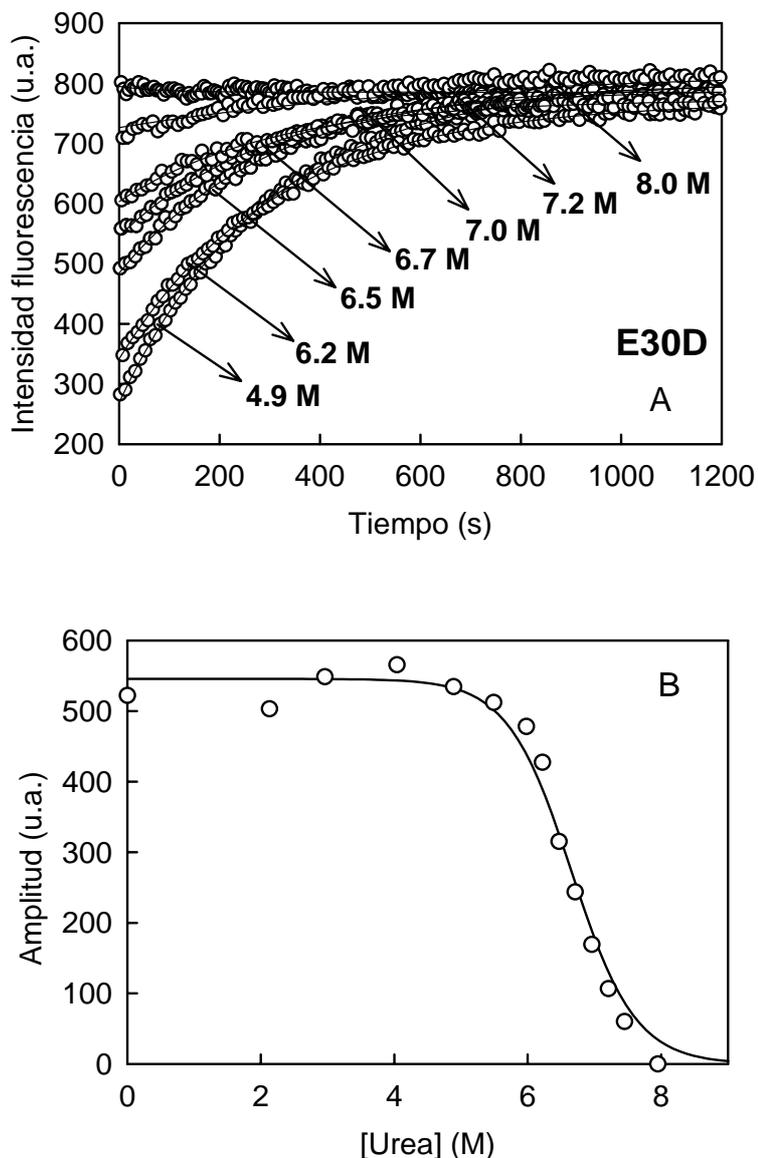


Figura 5.14: Determinación del perfil de desplegamiento inducido por urea en equilibrio, mediante ensayos de desplegamiento de doble salto, para la variante E30D de tiorredoxina de *E. coli*. A) Perfiles de intensidad de fluorescencia *versus* tiempo más representativos. Para registrarlos se ha transferido la variante de tiorredoxina (previamente incubada a 25°C y a las concentraciones de urea mostradas) hasta condiciones fuertemente desnaturizantes (8.5 M). La línea continua representa el mejor ajuste por mínimos cuadrados de la ecuación 2.40 a los datos experimentales (círculos). B) Amplitudes de desplegamiento calculadas del ajuste de la ecuación 2.40 a las cinéticas de desplegamiento mostradas en A. Estas amplitudes son proporcionales a la cantidad de proteína en estado nativo en la solución inicial y se representan frente a la concentración de urea en esas soluciones. La línea continua es el perfil predicho para la ecuación 2.53. con los parámetros dados en la Tabla 5.2.

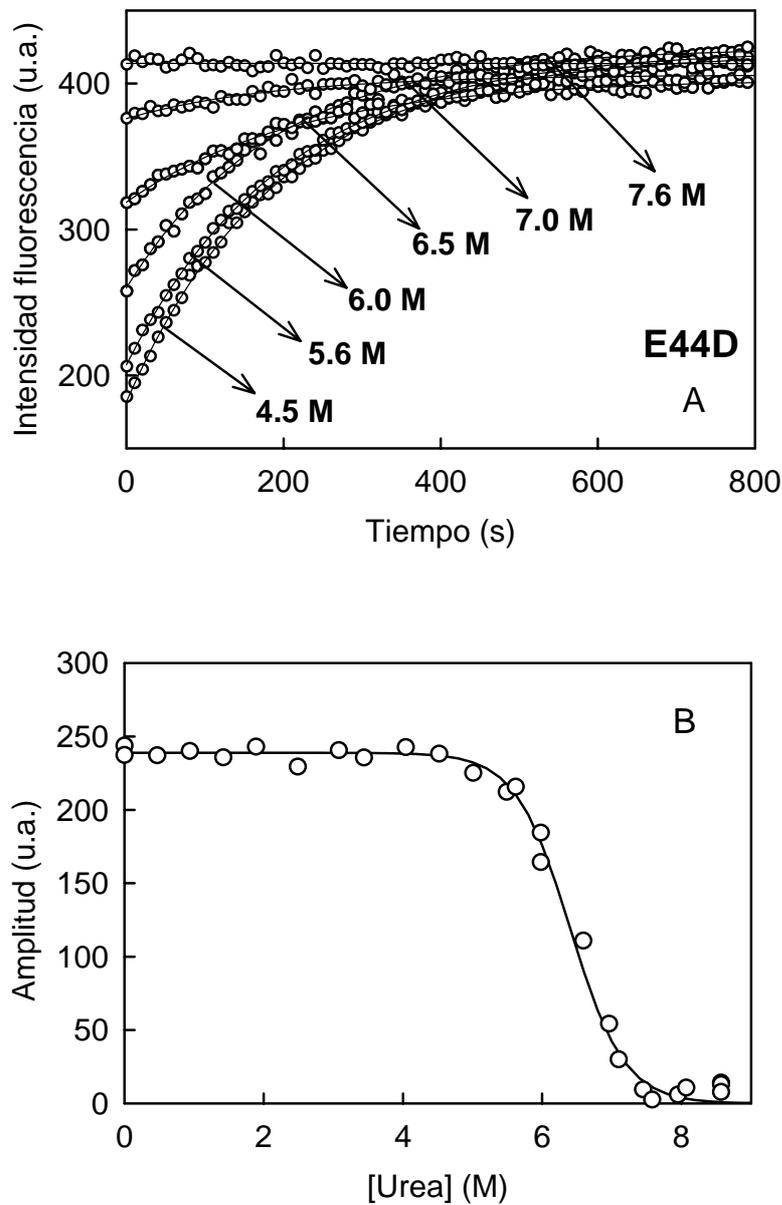


Figura 5.15: Determinación del perfil de desplegamiento inducido por urea en equilibrio, mediante ensayos de desplegamiento de doble salto, para la variante E44D de tiorredoxina de *E. coli*. A) Perfiles de intensidad de fluorescencia *versus* tiempo más representativos. Para registrarlos se ha transferido la variante de tiorredoxina (previamente incubada a 25°C y a las concentraciones de urea mostradas) hasta condiciones fuertemente desnaturalizantes (8.5 M). La línea continua representa el mejor ajuste por mínimos cuadrados de la ecuación 2.40 a los datos experimentales (círculos). B) Amplitudes de desplegamiento calculadas del ajuste de la ecuación 2.40 a las cinéticas de desplegamiento mostradas en A. Estas amplitudes son proporcionales a la cantidad de proteína en estado nativo en la solución inicial y se representan frente a la concentración de urea en esas soluciones. La línea continua es el perfil predicho para la

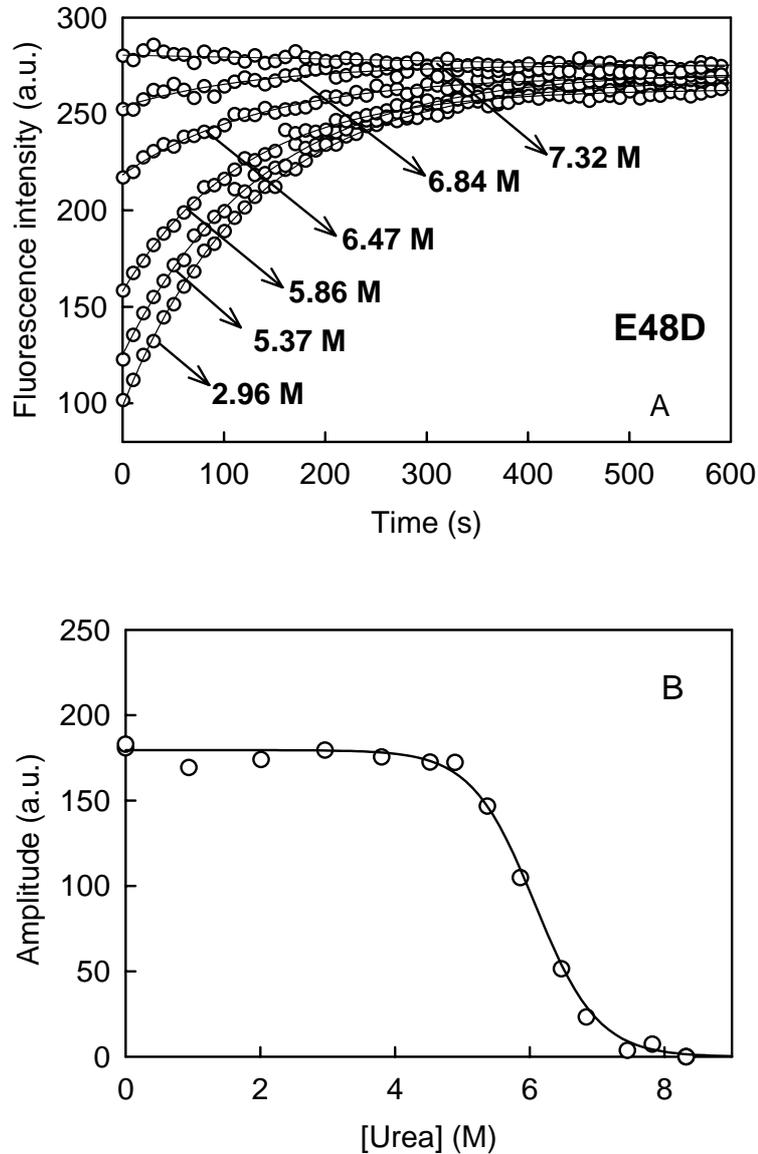


Figura 5.16: Determinación del perfil de desplegamiento inducido por urea en equilibrio, mediante ensayos de desplegamiento de doble salto, para la variante E48D de tiorredoxina de *E. coli*. A) Perfiles de intensidad de fluorescencia *versus* tiempo más representativos. Para registrarlos se ha transferido la variante de tiorredoxina (previamente incubada a 25°C y a las concentraciones de urea mostradas) hasta condiciones fuertemente desnaturizantes (8.5 M). La línea continua representa el mejor ajuste por mínimos cuadrados de la ecuación 2.40 a los datos experimentales (círculos). B) Amplitudes de desplegamiento calculadas del ajuste de la ecuación 2.40 a las cinéticas de desplegamiento mostradas en A. Estas amplitudes son proporcionales a la cantidad de proteína en estado nativo en la solución inicial y se representan frente a la concentración de urea en esas soluciones. La línea continua es el perfil predicho para la ecuación 2.53, con los parámetros dados en la Tabla 5.2.

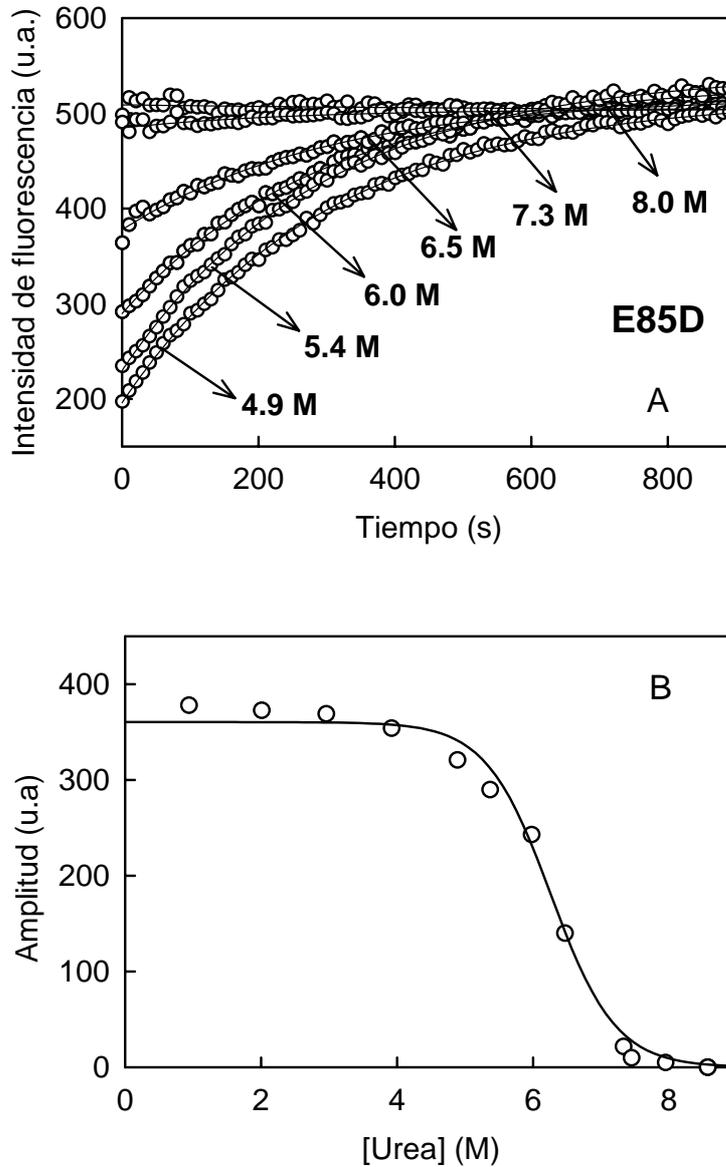


Figura 5.17: Determinación del perfil de desplegamiento inducido por urea en equilibrio, mediante ensayos de desplegamiento de doble salto, para la variante E85D de tiorredoxina de *E. coli*. A) Perfiles de intensidad de fluorescencia *versus* tiempo más representativos. Para registrarlos se ha transferido la variante de tiorredoxina (previamente incubada a 25°C y a las concentraciones de urea mostradas) hasta condiciones fuertemente desnaturalizantes (8.5 M). La línea continua representa el mejor ajuste por mínimos cuadrados de la ecuación 2.40 a los datos experimentales (círculos). B) Amplitudes de desplegamiento calculadas del ajuste de la ecuación 2.40 a las cinéticas de desplegamiento mostradas en A. Estas amplitudes son proporcionales a la cantidad de proteína en estado nativo en la solución inicial y se representan frente a la concentración de urea en esas soluciones. La línea continua es el perfil predicho para la ecuación 2.53. con los parámetros dados en la Tabla 5.2.

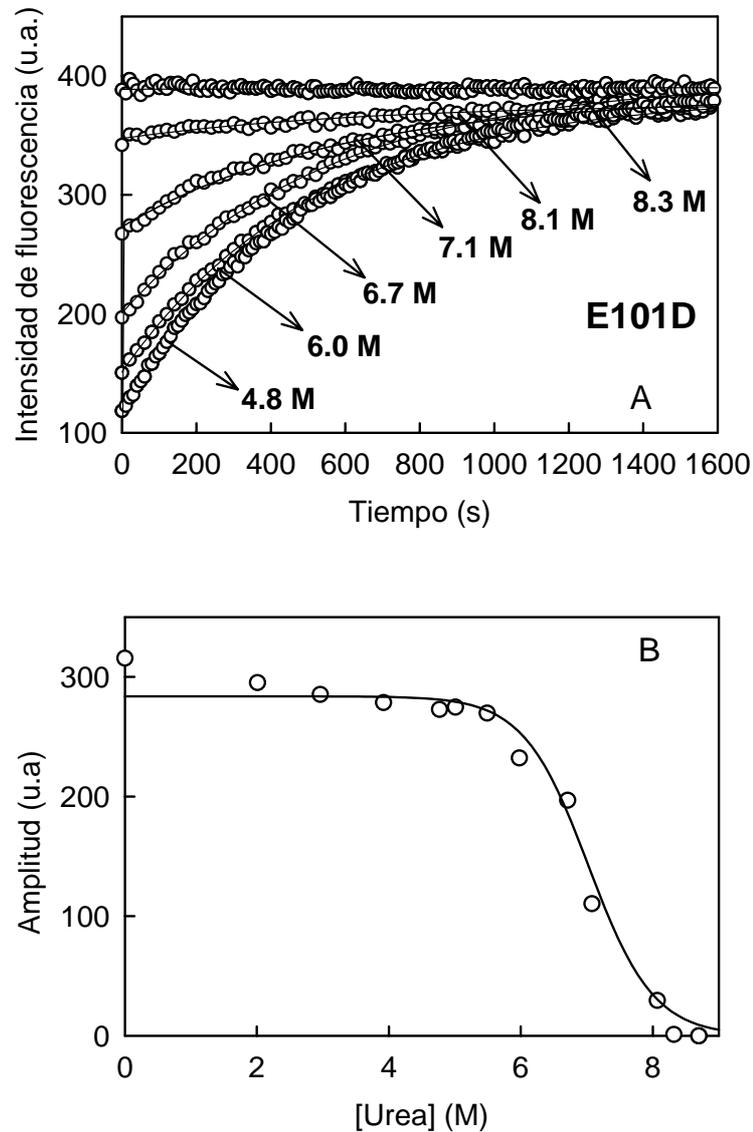


Figura 5.18: Determinación del perfil de desplegamiento inducido por urea en equilibrio, mediante ensayos de desplegamiento de doble salto, para la variante E101D de tiorredoxina de *E. coli*. A) Perfiles de intensidad de fluorescencia *versus* tiempo más representativos. Para registrarlos se ha transferido la variante de tiorredoxina (previamente incubada a 25°C y a las concentraciones de urea mostradas) hasta condiciones fuertemente desnaturalizantes (8.5 M). La línea continua representa el mejor ajuste por mínimos cuadrados de la ecuación 2.40 a los datos experimentales (círculos). B) Amplitudes de desplegamiento calculadas del ajuste de la ecuación 2.40 a las cinéticas de desplegamiento mostradas en A. Estas amplitudes son proporcionales a la cantidad de proteína en estado nativo en la solución inicial y se representan frente a la concentración de urea en esas soluciones. La línea continua es el perfil predicho para la ecuación 2.53, con los parámetros dados en la Tabla 5.2.

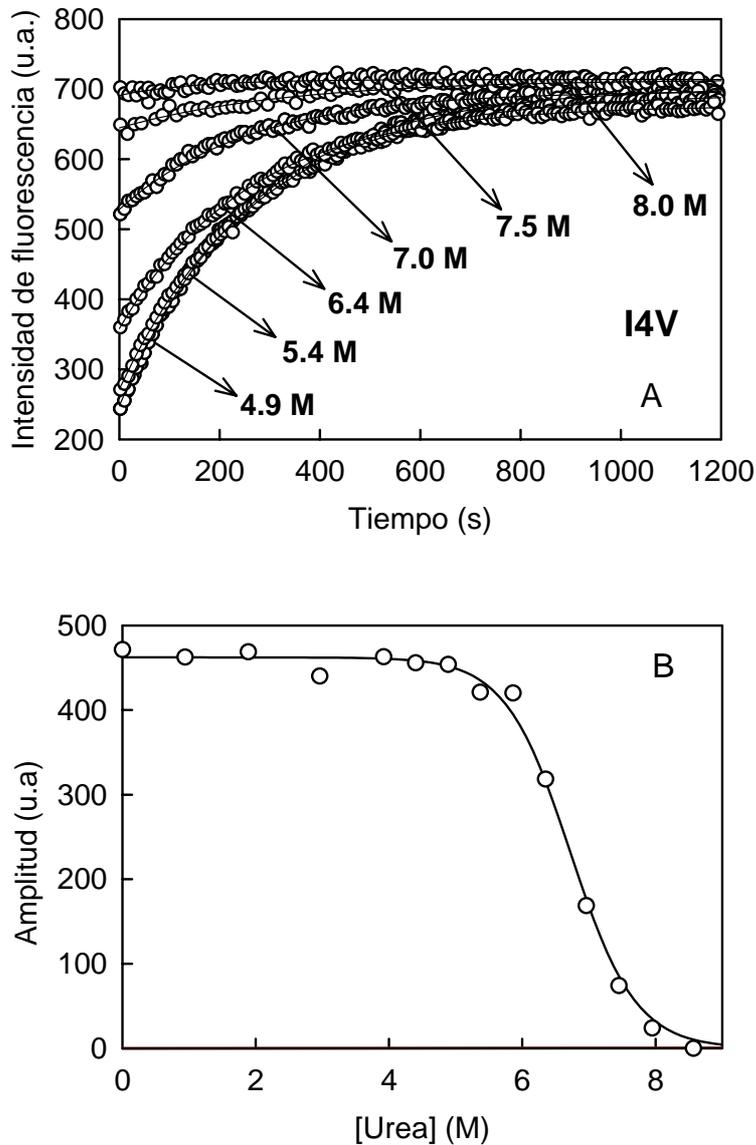


Figura 5.19: Determinación del perfil de desplegamiento inducido por urea en equilibrio, mediante ensayos de desplegamiento de doble salto, para la variante I4V de tiorredoxina de *E. coli*. A) Perfiles de intensidad de fluorescencia *versus* tiempo más representativos. Para registrarlos se ha transferido la variante de tiorredoxina (previamente incubada a 25°C y a las concentraciones de urea mostradas) hasta condiciones fuertemente desnaturizantes (8.5 M). La línea continua representa el mejor ajuste por mínimos cuadrados de la ecuación 2.40 a los datos experimentales (círculos). B) Amplitudes de desplegamiento calculadas del ajuste de la ecuación 2.40 a las cinéticas de desplegamiento mostradas en A. Estas amplitudes son proporcionales a la cantidad de proteína en estado nativo en la solución inicial y se representan frente a la concentración de urea en esas soluciones. La línea continua es el perfil predicho para la ecuación 2.53, con los parámetros dados en la Tabla 5.2.

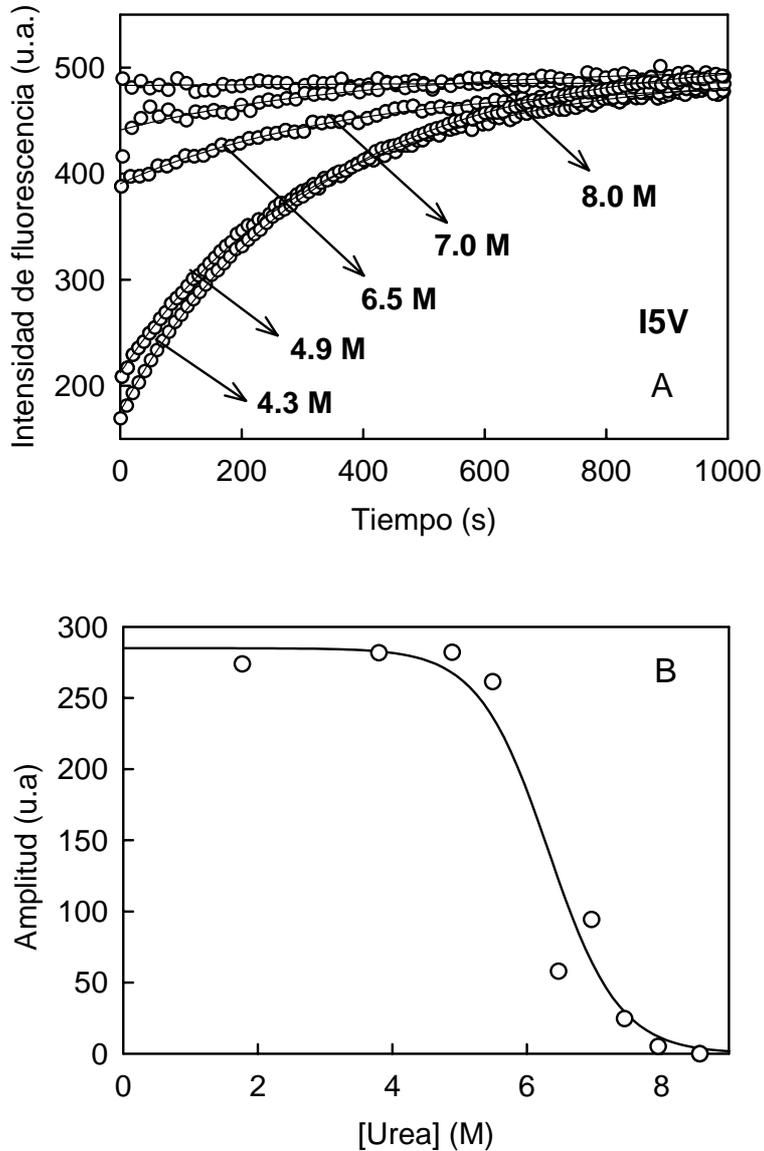


Figura 5.20: Determinación del perfil de desplegamiento inducido por urea en equilibrio, mediante ensayos de desplegamiento de doble salto, para la variante I5V de tiorredoxina de *E. coli*. A) Perfiles de intensidad de fluorescencia *versus* tiempo más representativos. Para registrarlos se ha transferido la variante de tiorredoxina (previamente incubada a 25°C y a las concentraciones de urea mostradas) hasta condiciones fuertemente desnaturizantes (8.5 M). La línea continua representa el mejor ajuste por mínimos cuadrados de la ecuación 2.40 a los datos experimentales (círculos). B) Amplitudes de desplegamiento calculadas del ajuste de la ecuación 2.40 a las cinéticas de desplegamiento mostradas en A. Estas amplitudes son proporcionales a la cantidad de proteína en estado nativo en la solución inicial y se representan frente a la concentración de urea en esas soluciones. La línea continua es el perfil predicho para la ecuación 2.53, con los parámetros dados en la Tabla 5.2.

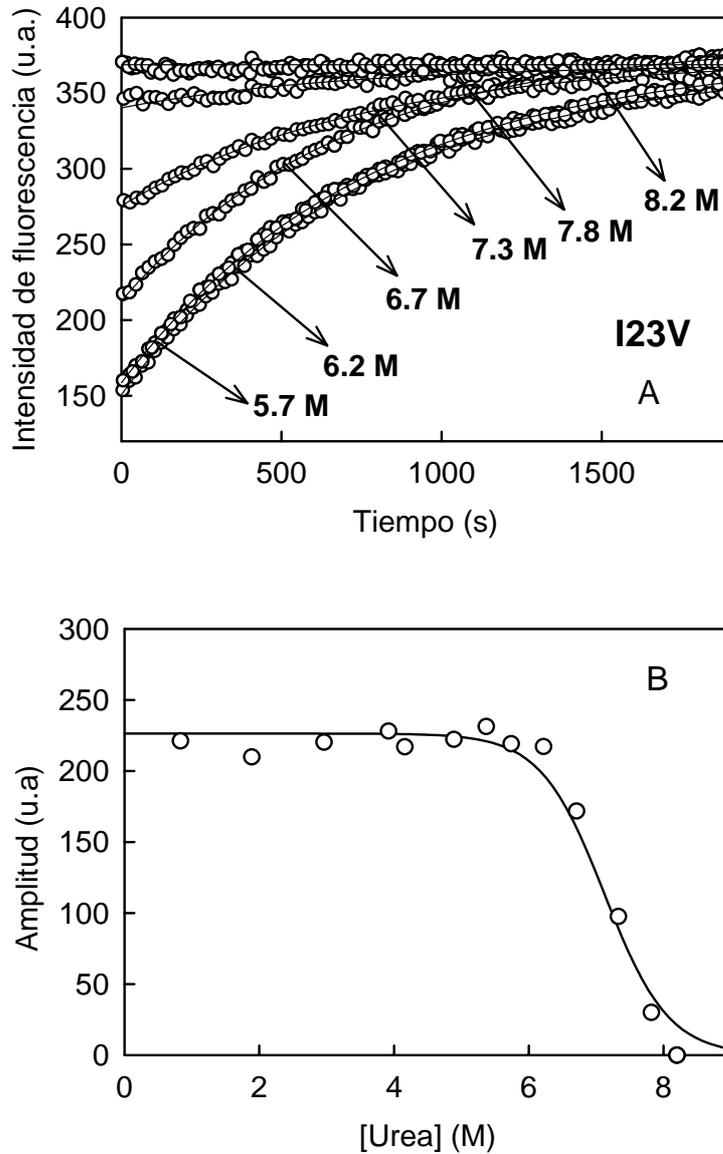


Figura 5.21: Determinación del perfil de desplegamiento inducido por urea en equilibrio, mediante ensayos de desplegamiento de doble salto, para la variante I23V de tiorredoxina de *E. coli*. A) Perfiles de intensidad de fluorescencia *versus* tiempo más representativos. Para registrarlos se ha transferido la variante de tiorredoxina (previamente incubada a 25°C y a las concentraciones de urea mostradas) hasta condiciones fuertemente desnaturizantes (8.5 M). La línea continua representa el mejor ajuste por mínimos cuadrados de la ecuación 2.40 a los datos experimentales (círculos). B) Amplitudes de desplegamiento calculadas del ajuste de la ecuación 2.40 a las cinéticas de desplegamiento mostradas en A. Estas amplitudes son proporcionales a la cantidad de proteína en estado nativo en la solución inicial y se representan frente a la concentración de urea en esas soluciones. La línea continua es el perfil predicho para la ecuación 2.53, con los parámetros dados en la Tabla 5.2.

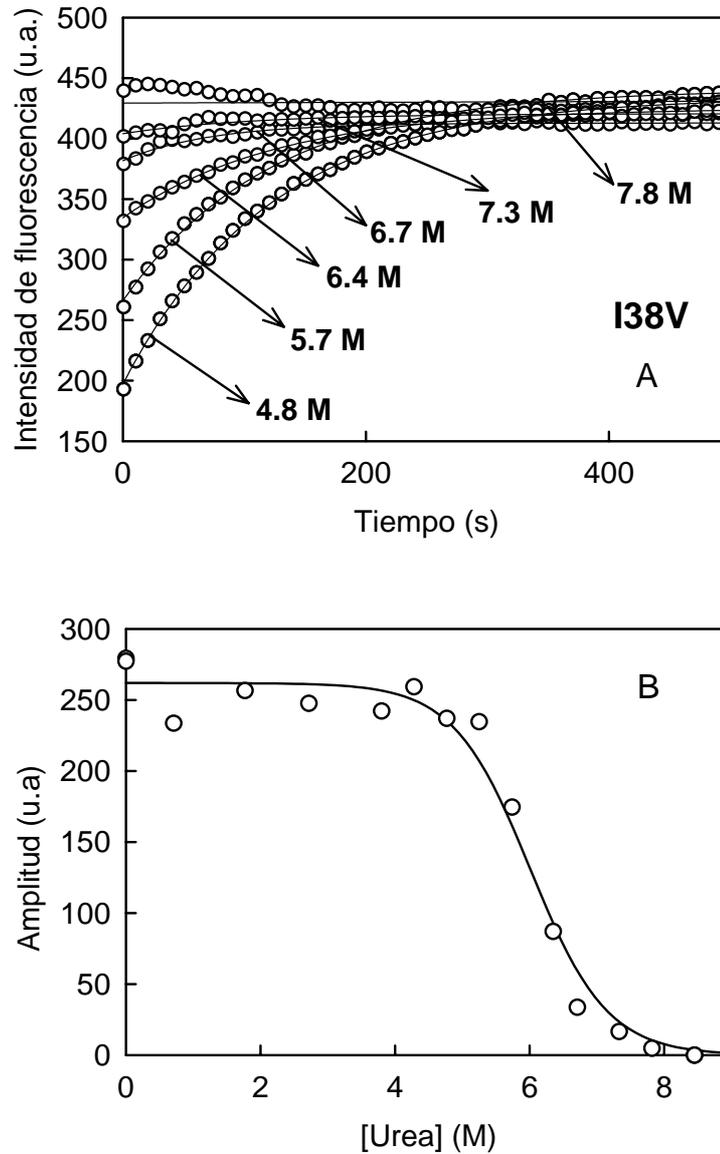


Figura 5.22: Determinación del perfil de desplegamiento inducido por urea en equilibrio, mediante ensayos de desplegamiento de doble salto, para la variante I38V de tiorredoxina de *E. coli*. A) Perfiles de intensidad de fluorescencia *versus* tiempo más representativos. Para registrarlos se ha transferido la variante de tiorredoxina (previamente incubada a 25°C y a las concentraciones de urea mostradas) hasta condiciones fuertemente desnaturizantes (8.5 M). La línea continua representa el mejor ajuste por mínimos cuadrados de la ecuación 2.40 a los datos experimentales (círculos). B) Amplitudes de desplegamiento calculadas del ajuste de la ecuación 2.40 a las cinéticas de desplegamiento mostradas en A. Estas amplitudes son proporcionales a la cantidad de proteína en estado nativo en la solución inicial y se representan frente a la concentración de urea en esas soluciones. La línea continua es el perfil predicho para la ecuación 2.53, con los parámetros dados en la Tabla 5.2.

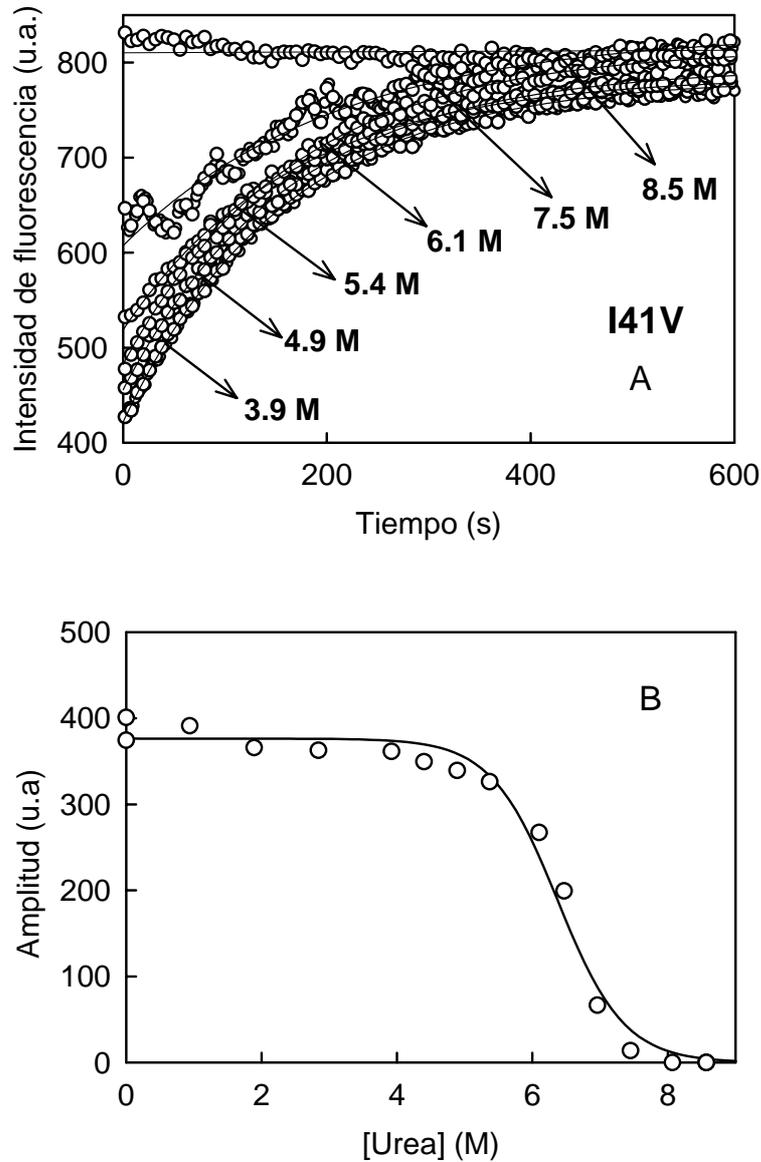


Figura 5.23: Determinación del perfil de desplegamiento inducido por urea en equilibrio, mediante ensayos de desplegamiento de doble salto, para la variante I41V de tiorredoxina de *E. coli*. A) Perfiles de intensidad de fluorescencia *versus* tiempo más representativos. Para registrarlos se ha transferido la variante de tiorredoxina (previamente incubada a 25°C y a las concentraciones de urea mostradas) hasta condiciones fuertemente desnaturizantes (8.5 M). La línea continua representa el mejor ajuste por mínimos cuadrados de la ecuación 2.40 a los datos experimentales (círculos). B) Amplitudes de desplegamiento calculadas del ajuste de la ecuación 2.40 a las cinéticas de desplegamiento mostradas en A. Estas amplitudes son proporcionales a la cantidad de proteína en estado nativo en la solución inicial y se representan frente a la concentración de urea en esas soluciones. La línea continua es el perfil predicho para la ecuación 2.53. con los parámetros dados en la Tabla 5.2.

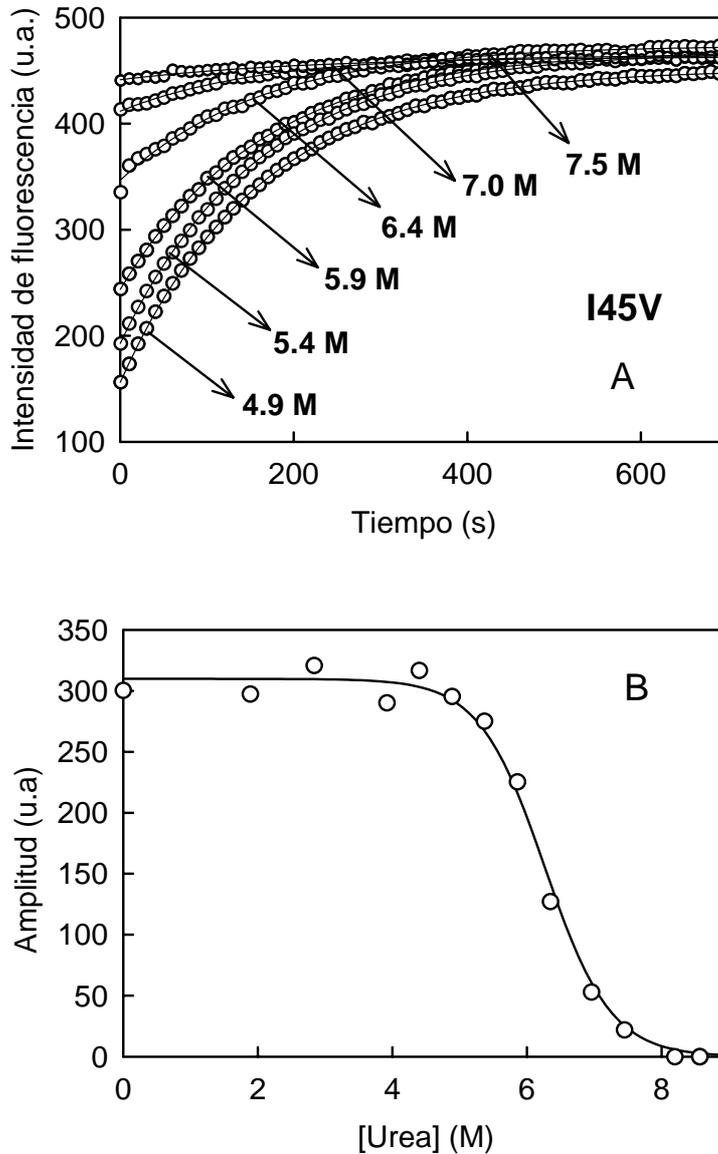


Figura 5.24: Determinación del perfil de desplegamiento inducido por urea en equilibrio, mediante ensayos de desplegamiento de doble salto, para la variante 145V de tiorredoxina de *E. coli*. A) Perfiles de intensidad de fluorescencia *versus* tiempo más representativos. Para registrarlos se ha transferido la variante de tiorredoxina (previamente incubada a 25°C y a las concentraciones de urea mostradas) hasta condiciones fuertemente desnaturizantes (8.5 M). La línea continua representa el mejor ajuste por mínimos cuadrados de la ecuación 2.40 a los datos experimentales (círculos). B) Amplitudes de desplegamiento calculadas del ajuste de la ecuación 2.40 a las cinéticas de desplegamiento mostradas en A. Estas amplitudes son proporcionales a la cantidad de proteína en estado nativo en la solución inicial y se representan frente a la concentración de urea en esas soluciones. La línea continua es el perfil predicho para la ecuación 2.53, con los parámetros dados en la Tabla 5.2.

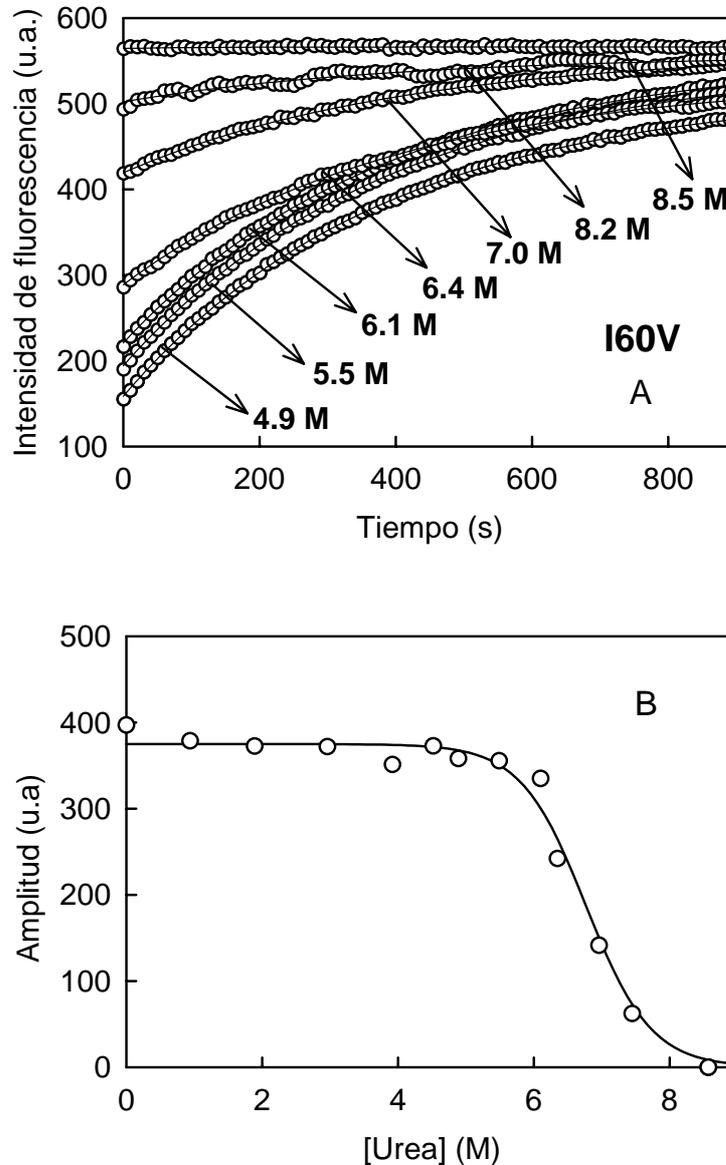


Figura 5.25: Determinación del perfil de desplegamiento inducido por urea en equilibrio, mediante ensayos de desplegamiento de doble salto, para la variante I60V de tiorredoxina de *E. coli*. A) Perfiles de intensidad de fluorescencia *versus* tiempo más representativos. Para registrarlos se ha transferido la variante de tiorredoxina (previamente incubada a 25°C y a las concentraciones de urea mostradas) hasta condiciones fuertemente desnaturalizantes (8.5 M). La línea continua representa el mejor ajuste por mínimos cuadrados de la ecuación 2.40 a los datos experimentales (círculos). B) Amplitudes de desplegamiento calculadas del ajuste de la ecuación 2.40 a las cinéticas de desplegamiento mostradas en A. Estas amplitudes son proporcionales a la cantidad de proteína en estado nativo en la solución inicial y se representan frente a la concentración de urea en esas soluciones. La línea continua es el perfil predicho para la ecuación 2.53. con los parámetros dados en la Tabla 5.2.

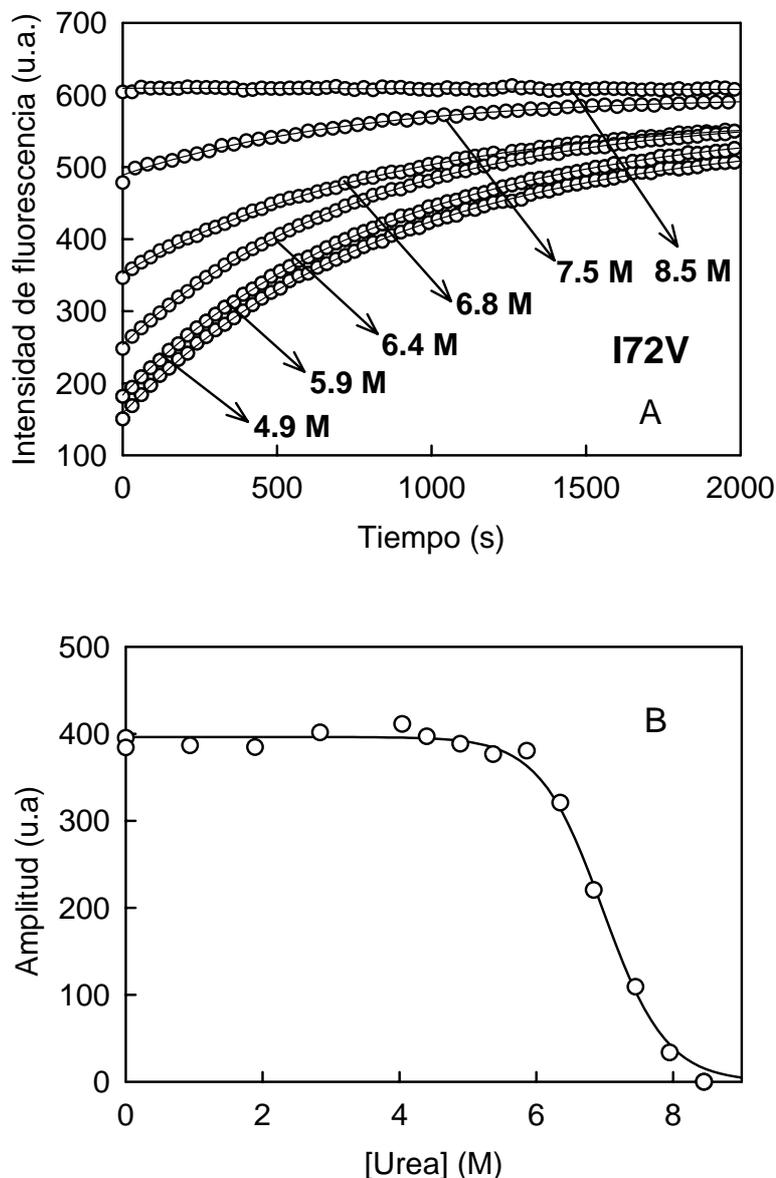


Figura 5.26: Determinación del perfil de desplegamiento inducido por urea en equilibrio, mediante ensayos de desplegamiento de doble salto, para la variante I72V de tiorredoxina de *E. coli*. A) Perfiles de intensidad de fluorescencia *versus* tiempo más representativos. Para registrarlos se ha transferido la variante de tiorredoxina (previamente incubada a 25°C y a las concentraciones de urea mostradas) hasta condiciones fuertemente desnaturalizantes (8.5 M). La línea continua representa el mejor ajuste por mínimos cuadrados de la ecuación 2.40 a los datos experimentales (círculos). B) Amplitudes de desplegamiento calculadas del ajuste de la ecuación 2.40 a las cinéticas de desplegamiento mostradas en A. Estas amplitudes son proporcionales a la cantidad de proteína en estado nativo en la solución inicial y se representan frente a la concentración de urea en esas soluciones. La línea continua es el perfil predicho para la ecuación 2.53, con los parámetros dados en la Tabla 5.2.

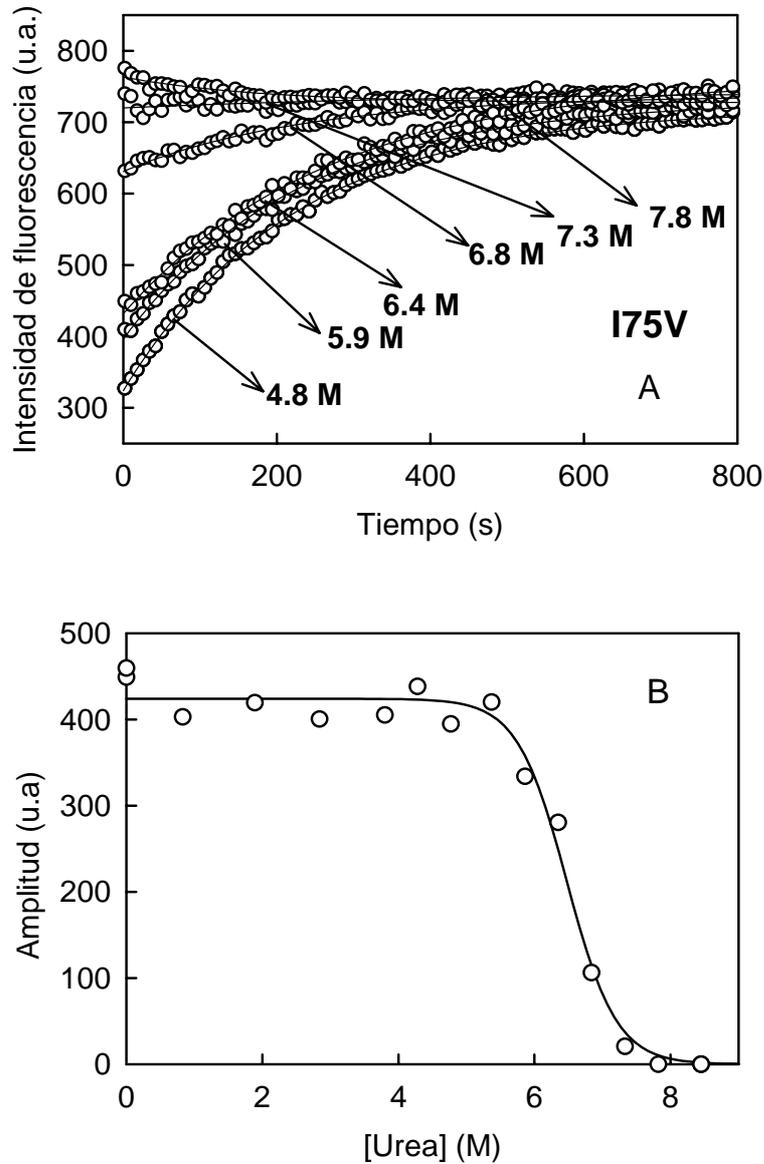


Figura 5.27: Determinación del perfil de desplegamiento inducido por urea en equilibrio, mediante ensayos de desplegamiento de doble salto, para la variante 175V de tiorredoxina de *E. coli*. A) Perfiles de intensidad de fluorescencia *versus* tiempo más representativos. Para registrarlos se ha transferido la variante de tiorredoxina (previamente incubada a 25°C y a las concentraciones de urea mostradas) hasta condiciones fuertemente desnaturizantes (8.5 M). La línea continua representa el mejor ajuste por mínimos cuadrados de la ecuación 2.40 a los datos experimentales (círculos). B) Amplitudes de desplegamiento calculadas del ajuste de la ecuación 2.40 a las cinéticas de desplegamiento mostradas en A. Estas amplitudes son proporcionales a la cantidad de proteína en estado nativo en la solución inicial y se representan frente a la concentración de urea en esas soluciones. La línea continua es el perfil predicho para la ecuación 2.53. con los parámetros dados en la Tabla 5.2.

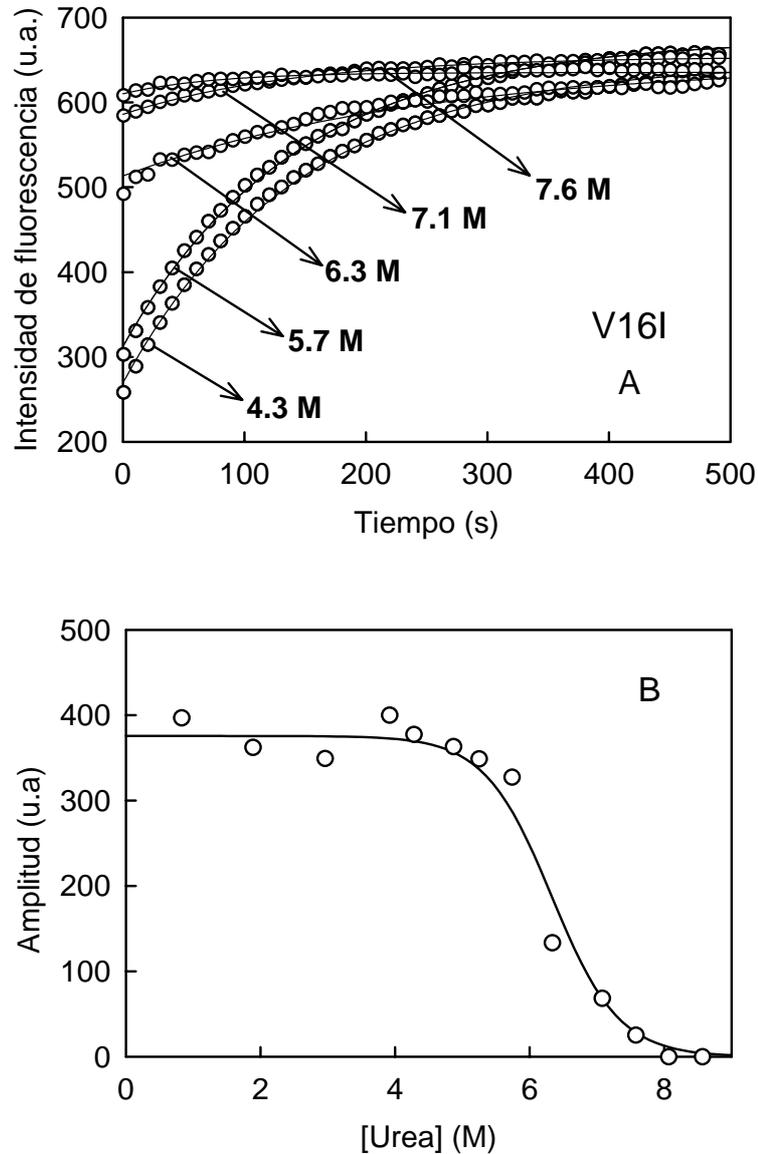


Figura 5.28: Determinación del perfil de desplegamiento inducido por urea en equilibrio, mediante ensayos de desplegamiento de doble salto, para la variante V16I de tiorredoxina de *E. coli*. A) Perfiles de intensidad de fluorescencia *versus* tiempo más representativos. Para registrarlos se ha transferido la variante de tiorredoxina (previamente incubada a 25°C y a las concentraciones de urea mostradas) hasta condiciones fuertemente desnaturizantes (8.5 M). La línea continua representa el mejor ajuste por mínimos cuadrados de la ecuación 2.40 a los datos experimentales (círculos). B) Amplitudes de desplegamiento calculadas del ajuste de la ecuación 2.40 a las cinéticas de desplegamiento mostradas en A. Estas amplitudes son proporcionales a la cantidad de proteína en estado nativo en la solución inicial y se representan frente a la concentración de urea en esas soluciones. La línea continua es el perfil predicho para la ecuación 2.53, con los parámetros dados en la Tabla 5.2.

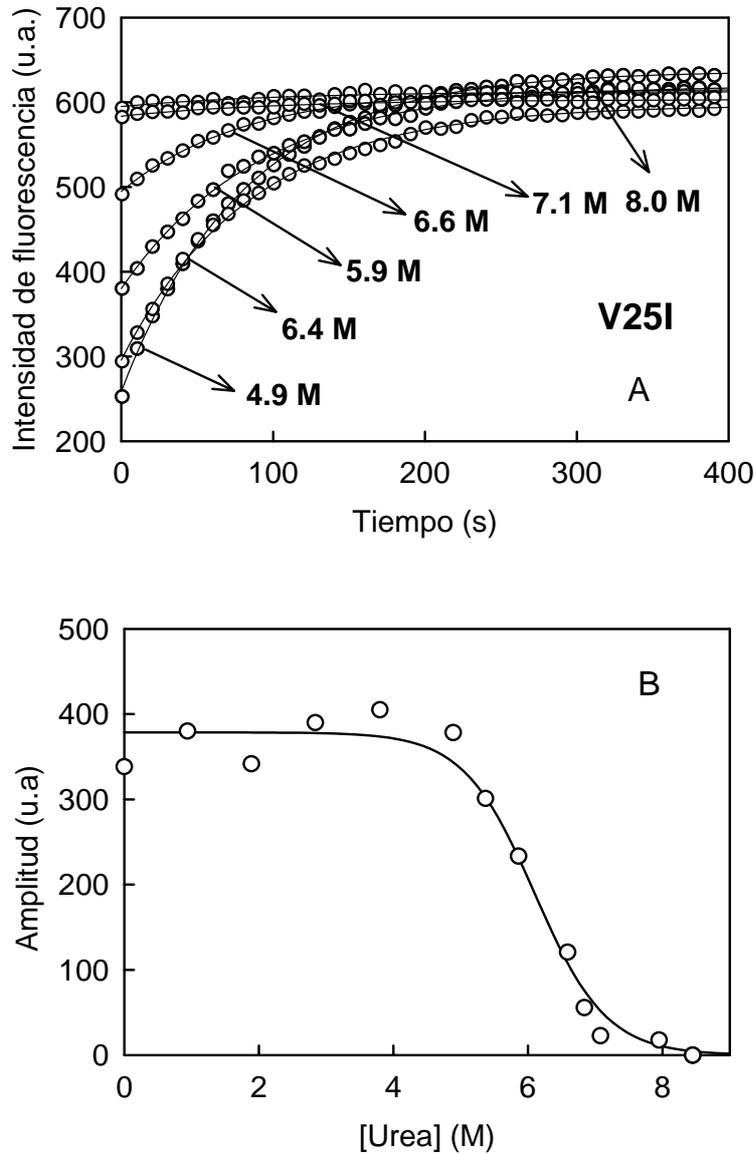


Figura 5.29: Determinación del perfil de desplegamiento inducido por urea en equilibrio, mediante ensayos de desplegamiento de doble salto, para la variante V25I de tiorredoxina de *E. coli*. A) Perfiles de intensidad de fluorescencia *versus* tiempo más representativos. Para registrarlos se ha transferido la variante de tiorredoxina (previamente incubada a 25°C y a las concentraciones de urea mostradas) hasta condiciones fuertemente desnaturalizantes (8.5 M). La línea continua representa el mejor ajuste por mínimos cuadrados de la ecuación 2.40 a los datos experimentales (círculos). B) Amplitudes de desplegamiento calculadas del ajuste de la ecuación 2.40 a las cinéticas de desplegamiento mostradas en A. Estas amplitudes son proporcionales a la cantidad de proteína en estado nativo en la solución inicial y se representan frente a la concentración de urea en esas soluciones. La línea continua es el perfil predicho para la ecuación 2.53, con los parámetros dados en la Tabla 5.2.

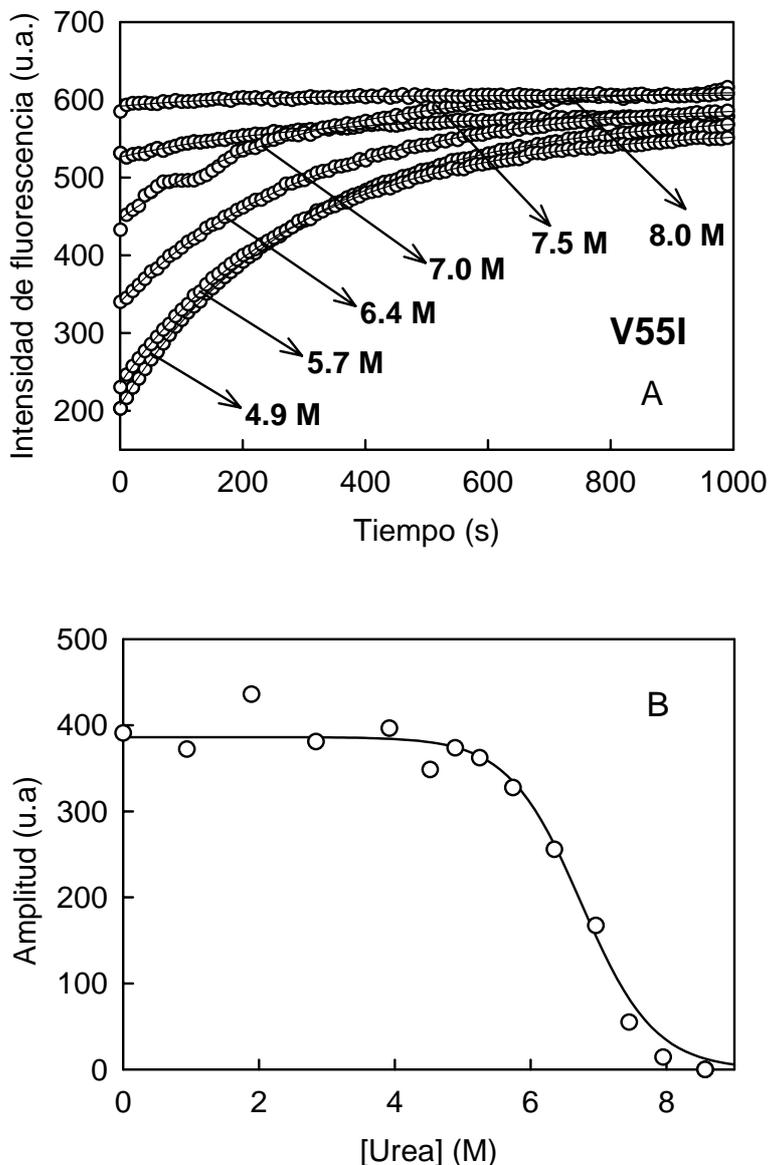


Figura 5.30: Determinación del perfil de desplegamiento inducido por urea en equilibrio, mediante ensayos de desplegamiento de doble salto, para la variante V55I de tiorredoxina de *E. coli*. A) Perfiles de intensidad de fluorescencia *versus* tiempo más representativos. Para registrarlos se ha transferido la variante de tiorredoxina (previamente incubada a 25°C y a las concentraciones de urea mostradas) hasta condiciones fuertemente desnaturizantes (8.5 M). La línea continua representa el mejor ajuste por mínimos cuadrados de la ecuación 2.40 a los datos experimentales (círculos). B) Amplitudes de desplegamiento calculadas del ajuste de la ecuación 2.40 a las cinéticas de desplegamiento mostradas en A. Estas amplitudes son proporcionales a la cantidad de proteína en estado nativo en la solución inicial y se representan frente a la concentración de urea en esas soluciones. La línea continua es el perfil predicho para la ecuación 2.53. con los parámetros dados en la Tabla 5.2.

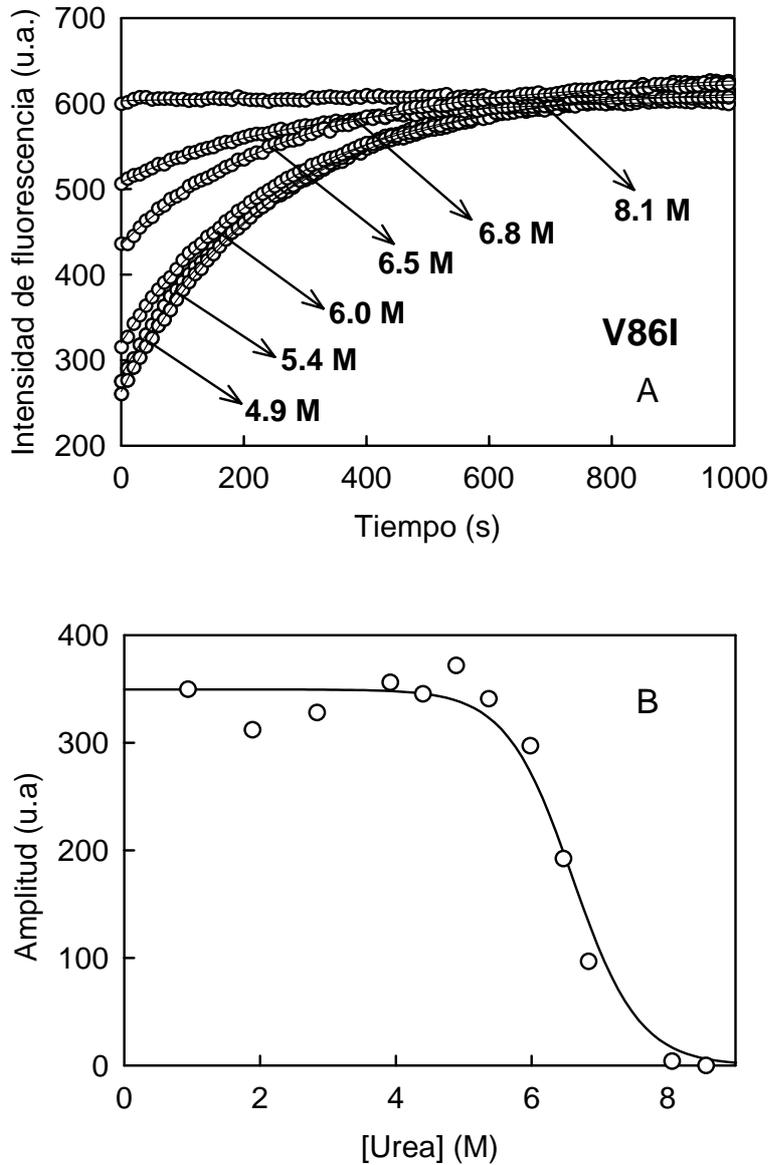


Figura 5.31: Determinación del perfil de desplegamiento inducido por urea en equilibrio, mediante ensayos de desplegamiento de doble salto, para la variante V86I de tiorredoxina de *E. coli*. A) Perfiles de intensidad de fluorescencia *versus* tiempo más representativos. Para registrarlos se ha transferido la variante de tiorredoxina (previamente incubada a 25°C y a las concentraciones de urea mostradas) hasta condiciones fuertemente desnaturizantes (8.5 M). La línea continua representa el mejor ajuste por mínimos cuadrados de la ecuación 2.40 a los datos experimentales (círculos). B) Amplitudes de desplegamiento calculadas del ajuste de la ecuación 2.40 a las cinéticas de desplegamiento mostradas en A. Estas amplitudes son proporcionales a la cantidad de proteína en estado nativo en la solución inicial y se representan frente a la concentración de urea en esas soluciones. La línea continua es el perfil predicho para la ecuación 2.53. con los parámetros dados en la Tabla 5.2.

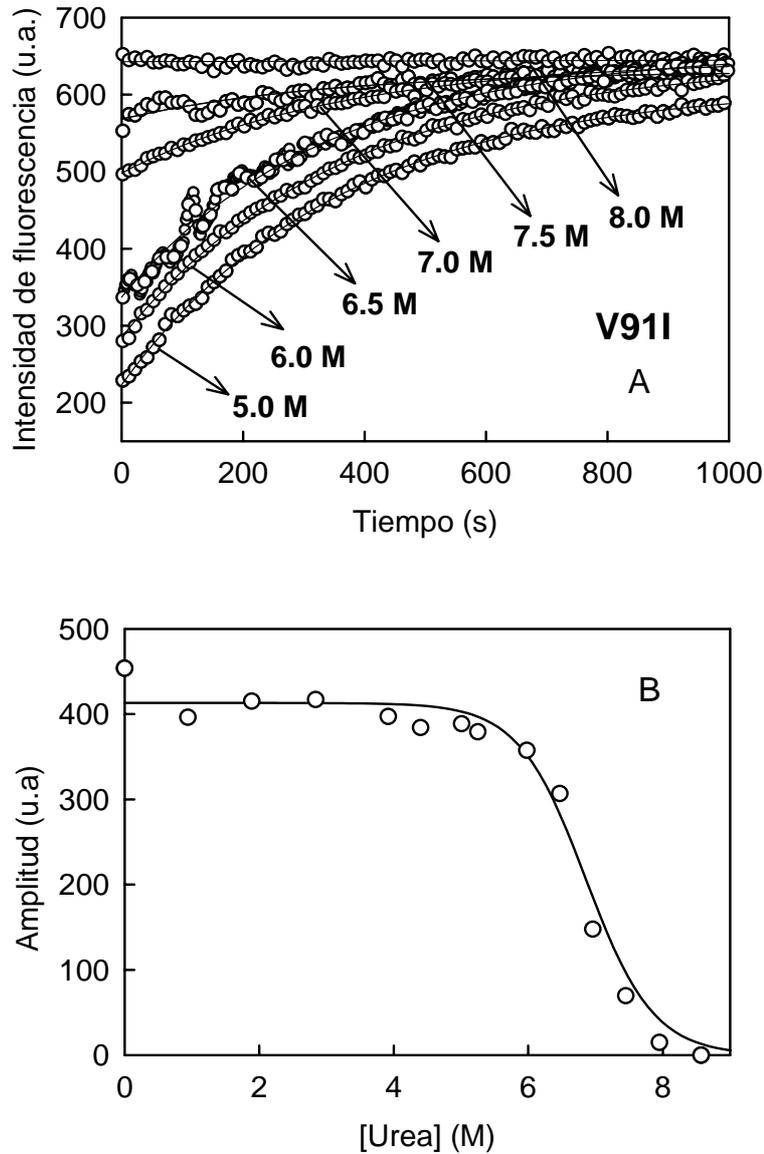


Figura 5.32: Determinación del perfil de desplegamiento inducido por urea en equilibrio, mediante ensayos de desplegamiento de doble salto, para la variante V91I de tiorredoxina de *E. coli*. A) Perfiles de intensidad de fluorescencia *versus* tiempo más representativos. Para registrarlos se ha transferido la variante de tiorredoxina (previamente incubada a 25°C y a las concentraciones de urea mostradas) hasta condiciones fuertemente desnaturizantes (8.5 M). La línea continua representa el mejor ajuste por mínimos cuadrados de la ecuación 2.40 a los datos experimentales (círculos). B) Amplitudes de desplegamiento calculadas del ajuste de la ecuación 2.40 a las cinéticas de desplegamiento mostradas en A. Estas amplitudes son proporcionales a la cantidad de proteína en estado nativo en la solución inicial y se representan frente a la concentración de urea en esas soluciones. La línea continua es el perfil predicho para la ecuación 2.53, con los parámetros dados en la Tabla 5.2.

5.4.2 CINÉTICAS DE PLEGAMIENTO Y DESPLEGAMIENTO ESTUDIADAS MEDIANTE MEDIDAS DE FLUORESCENCIA DEL ESTADO ESTACIONARIO

Las cinéticas de plegamiento y desplegamiento se han realizado a 25°C (HEPES 5 mM, pH 7), siguiendo la dependencia con el tiempo de la emisión de fluorescencia a 350 nm (excitación a 276 nm; ver sección 2.2.6.1 para más detalles) al llevar a cabo saltos en la concentración de urea (disolución de proteína a concentración final de 0.04 mg/mL, disuelta en urea de concentración 9M ó 0M para los experimentos llevados hasta concentraciones de urea menores o mayores de la $C_{1/2}$, respectivamente). Las constantes de velocidad aparentes de plegamiento-desplegamiento (k) se han calculado de los ajustes de las ecuaciones de primer orden 2.40 [$I = I_{\infty} - \Delta I \cdot e^{(-t/\tau)}$] y 2.42 [$I = I_{\infty} + \Delta I \cdot e^{(-t/\tau)}$] a los datos de intensidad de fluorescencia *versus* tiempo experimentales (para detalles experimentales sobre la realización de los perfiles cinéticos de plegamiento y desplegamiento, véase la sección 2.2.6.2 de *Materiales y métodos* y la Figura 2.14 como ejemplo ilustrativo). Los ajustes fueron siempre excelentes. En las Figuras 5.33-B, 5.34-B, 5.35-B y 5.36-B, se muestran las representaciones gráficas de los *Chevron* ($\ln k$ *versus* concentración de urea) de tiorredoxina *WT* y sus distintas formas mutantes. Se puede observar que hemos restringido la rama de plegamiento de los *Chevron* a concentraciones cercanas a la $C_{1/2}$, ya que a bajas concentraciones de urea existen complicaciones cinéticas (presencia de intermedios, isomerización de prolinas) (Wilson *et al.*, 1986; Georgescu *et al.*, 1998). Los *Chevron* de tiorredoxina *WT* y de las distintas variantes se han descrito adecuadamente según las ecuaciones 2.45 [$\ln k = \ln(k_F + k_D)$], 2.46 [$\ln k_F = \ln k_{1/2} + \frac{m_{D-\ddagger}}{RT} \cdot (C - C_{1/2})$] y 2.47 [$\ln k_D = \ln k_{1/2} + \frac{m_{N-\ddagger}}{RT} \cdot (C - C_{1/2})$] descritas en la sección 2.2.6.2 de *Materiales y métodos*.

5.4.3 ANÁLISIS GLOBALES DE LOS DATOS DE DESPLEGAMIENTO INDUCIDO POR UREA, EN EQUILIBRIO Y CINÉTICOS

Con el fin de acotar el error de los parámetros ajustables, tanto de los ajustes de los *Chevron* (programas 4, 5 y 6 de Apéndice I) como de los perfiles de equilibrio de amplitudes frente a la concentración de urea, se han llevado a cabo ajustes globales de los *Chevron* y los perfiles de equilibrio, de tiorredoxina *WT* y

5. La selección natural para la estabilidad cinética...

27 mutantes simples de tiorredoxina. El valor de la $C_{1/2}$ aparece en las descripciones matemáticas de los perfiles del desplegamiento en equilibrio [ecuación 2.53] y de los *Chevron* [ecuaciones 2.45, 2.46 y 2.47]. Además, los valores de m cinéticos en las ecuaciones 2.46 y 2.47 se relacionan con el valor de m correspondiente al equilibrio mediante la ecuación 2.51 [$m_{1/2} = m_{N-\ddagger} - m_{D-\ddagger}$]. Por lo tanto, se han obtenido, para cada variante de tiorredoxina, valores de los parámetros relevantes, en equilibrio (ΔI^0 , $C_{1/2}$, $m_{1/2}$) y cinéticos ($m_{N-\ddagger}$, $m_{1/2}$, $C_{1/2}$, $\ln k_{1/2}$), realizando ajustes globales (simultáneos) de las ecuaciones 2.45, 2.46, 2.47, 2.51 y 2.53 a los perfiles de equilibrio y cinéticos, usando el siguiente conjunto de parámetros ajustables: ΔI^0 , $C_{1/2}$, $m_{1/2}$, $\ln k_{1/2}$, $m_{N-\ddagger}$. Los ajustes obtenidos han sido excelentes en la mayoría de los casos (Figuras 5.33, 5.34, 5.35 y 5.36), lo que apoya el modelo de ajuste global utilizado. Aunque el valor de m cinético de plegamiento ($m_{D-\ddagger}$) no se ha tomado como un parámetro ajustable, se puede derivar de los valores de $m_{1/2}$ y $m_{N-\ddagger}$ utilizando la ecuación 2.51. De hecho, esta forma de cálculo, da lugar a una estimación bastante buena del valor de m de plegamiento teniendo en cuenta lo corta que es la rama de plegamiento de los *Chevron* experimentales. Además, se han obtenido otros parámetros de interés utilizando los parámetros resultantes del ajuste global. Por ejemplo, la estimación de la extrapolación lineal de la constante de desplegamiento en agua (es decir, en ausencia de urea), se puede calcular utilizando la siguiente expresión:

$$\ln k_D(0M) = \ln k_{1/2} - \frac{m_{N-\ddagger}}{RT} C_{1/2} \quad (5.4)$$

Los errores estándar asociados a los valores de los parámetros del ajuste (ΔI^0 , $C_{1/2}$, $m_{1/2}$, $\ln k_{1/2}$, $m_{N-\ddagger}$) y a los valores de los parámetros derivados ($m_{N-\ddagger}$ y $\ln k_D(0M)$), se han estimado usando un procedimiento de Monte-Carlo. Se han generado unas 50 réplicas de cada uno de los conjuntos de datos originales (perfiles de desplegamiento en equilibrio y *Chevron*) añadiendo a los valores del mejor ajuste los errores Gaussianos más las desviaciones estándar derivadas del mejor ajuste. El conjunto de réplicas se sometió al mismo procedimiento de ajuste

que el conjunto de datos originales y el análisis estadístico de los valores de los parámetros, nos proporcionó sus errores estándar.

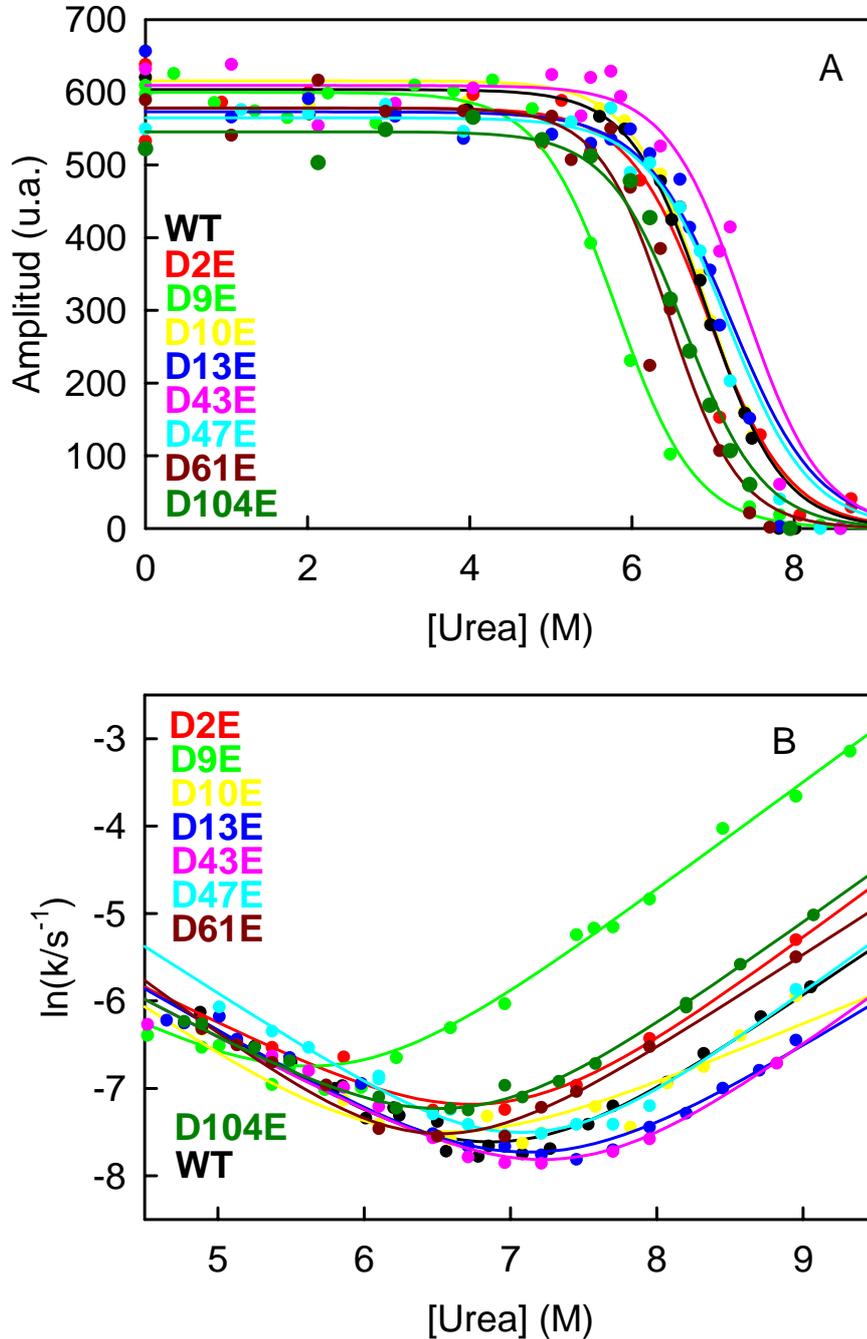


Figura 5.33: Ajustes globales, no lineales por mínimos cuadrados, de los perfiles de equilibrio obtenidos mediante ensayos de desplegamiento de doble salto (A) y cinéticos (B), para la desnaturalización inducida por urea para tiorredoxina *WT* y todos las variantes del tipo $D \rightarrow E$. Las líneas continuas representan los mejores ajustes de las ecuaciones 2.40 a 2.53, a los datos experimentales. Los resultados de los ajustes se dan en la Tabla 5.2.

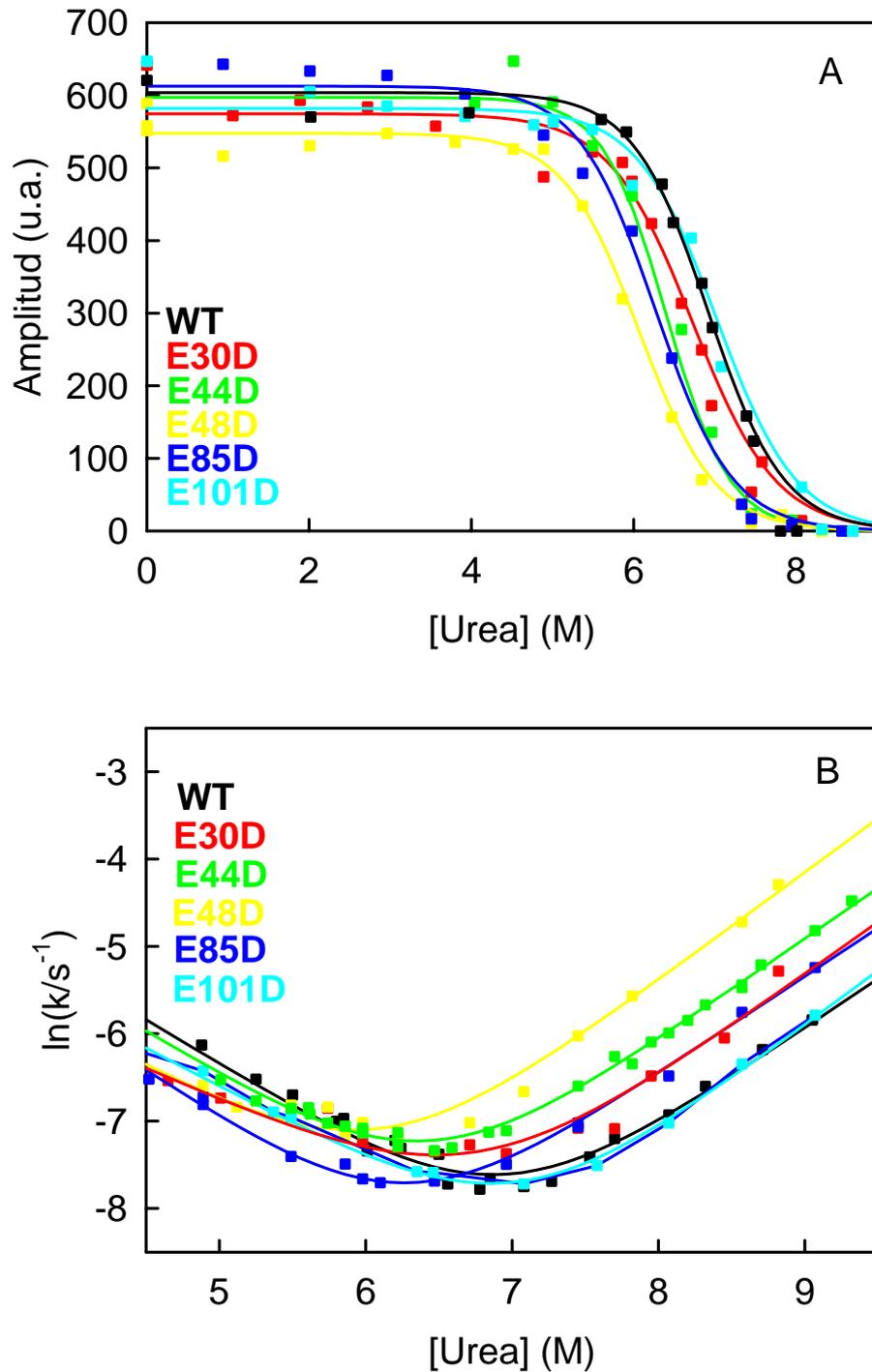


Figura 5.34: Ajustes globales, no lineales por mínimos cuadrados, de los perfiles de equilibrio obtenidos mediante ensayos de desplegamiento de doble salto (A) y cinéticos (B), para la desnaturalización inducida por urea para tiorredoxina *WT* y todas las variantes del tipo $D \rightarrow E$. Las líneas continuas representan los mejores ajustes de las ecuaciones 2.40 a 2.53, a los datos experimentales. Los resultados de los ajustes se dan en la Tabla 5.2.

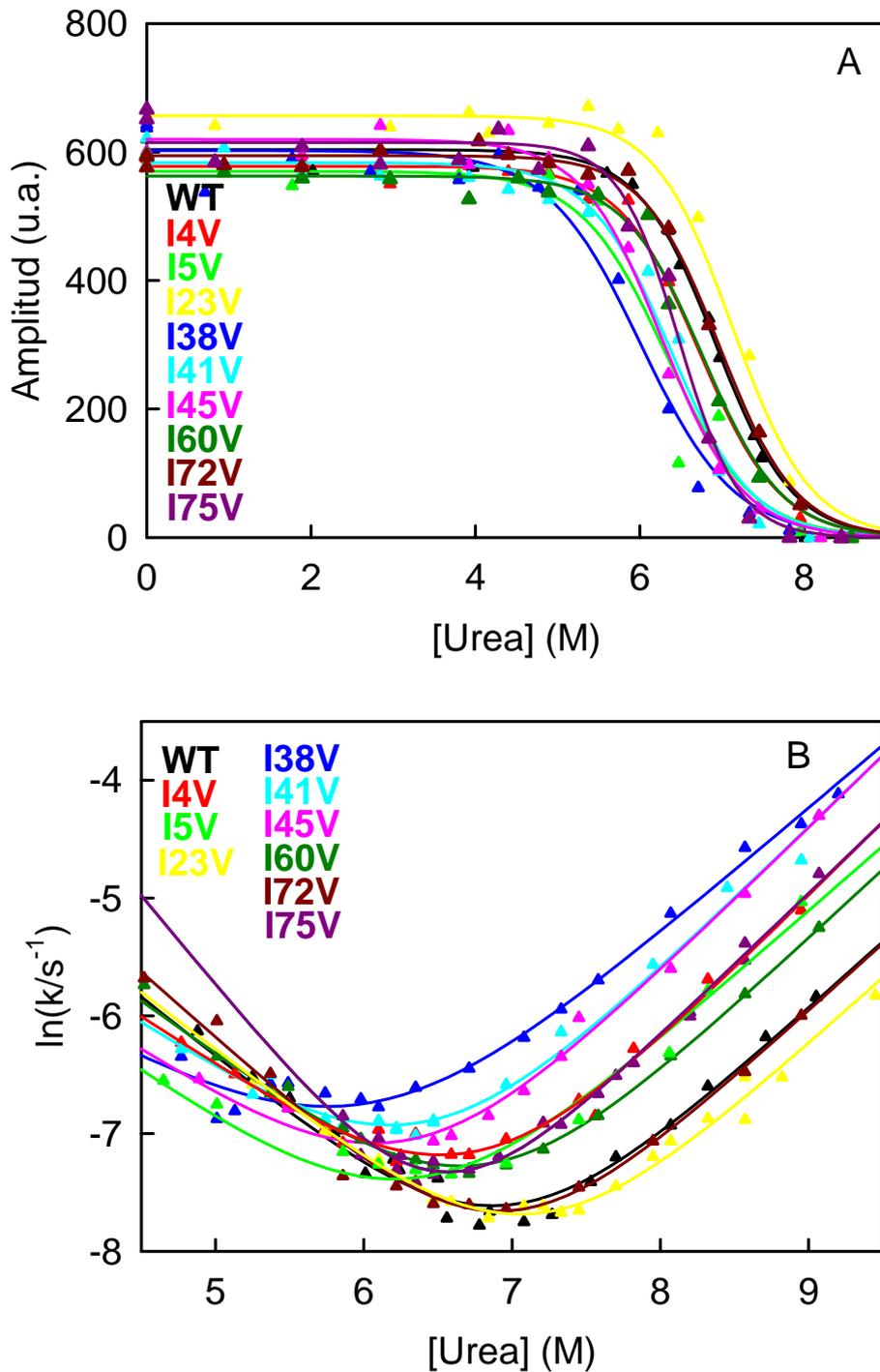


Figura 5.35: Ajustes globales, no lineales por mínimos cuadrados, de los perfiles de equilibrio obtenidos mediante ensayos de desplegamiento de doble salto (A) y cinéticos (B), para la desnaturalización inducida por urea para tiorredoxina WT y todos las variantes del tipo D \rightarrow E. Las líneas continuas representan los mejores ajustes de las ecuaciones 2.40 a 2.53, a los datos experimentales. Los resultados de los ajustes se dan en la Tabla 5.2.

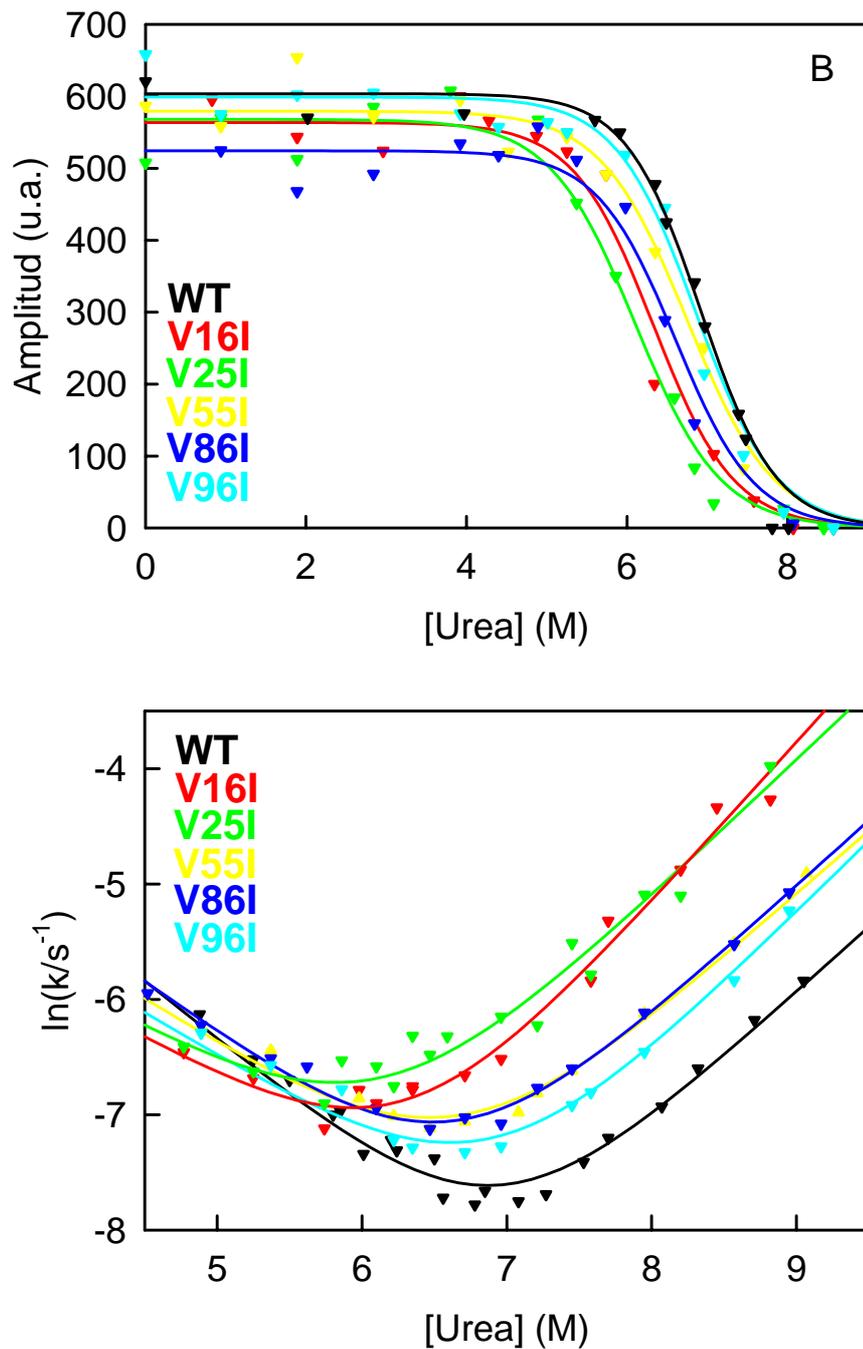


Figura 5.36: Ajustes globales, no lineales por mínimos cuadrados, de los perfiles de equilibrio obtenidos mediante ensayos de desplegamiento de doble salto (A) y cinéticos (B), para la desnaturalización inducida por urea para tiorredoxina *WT* y todos las variantes del tipo *D*→*E*. Las líneas continuas representan los mejores ajustes de las ecuaciones 2.40 a 2.53, a los datos experimentales. Los resultados de los ajustes se dan en la Tabla 5.2.

Los resultados obtenidos para tiorredoxina *WT* y las 27 variantes se encuentran resumidos en las Tablas 5.2 y 5.3.

Tabla 5.2. Parámetros cinéticos y de equilibrio para el plegamiento-desplegamiento de tiorredoxina *WT* de *E. coli* y variantes en presencia de urea.

PROTEÍNA	$C_{1/2}$ (M)	$m_{1/2}$ (kJ·mol ⁻¹ ·M ⁻¹)	$m_{N \rightarrow \ddagger}$ (kJ·mol ⁻¹ ·M ⁻¹)	$m_{U \rightarrow \ddagger}$ (kJ·mol ⁻¹ ·M ⁻¹)	$\ln(k_{1/2}/s^{-1})$
WT	6.92 ± 0.05	5.34 ± 0.10	2.82 ± 0.09	-2.52 ± 0.10	-8.30 ± 0.01
D2E	6.94 ± 0.08	5.05 ± 0.19	3.12 ± 0.19	-1.93 ± 0.12	-7.85 ± 0.05
D9E	5.82 ± 0.04	5.12 ± 0.21	3.05 ± 0.08	-2.07 ± 0.17	-7.42 ± 0.05
D10E	6.77 ± 0.09	4.84 ± 0.18	2.44 ± 0.14	-2.40 ± 0.17	-8.26 ± 0.05
D13E	7.07 ± 0.06	4.96 ± 0.17	2.45 ± 0.16	-2.51 ± 0.24	-8.43 ± 0.04
D43E	7.38 ± 0.07	5.18 ± 0.22	3.02 ± 0.23	-2.16 ± 0.11	-8.49 ± 0.04
D47E	7.12 ± 0.05	5.68 ± 0.16	3.02 ± 0.16	-2.66 ± 0.12	-8.20 ± 0.03
D61E	6.46 ± 0.06	5.75 ± 0.14	2.67 ± 0.12	-3.07 ± 0.17	-8.21 ± 0.03
D104E	6.66 ± 0.04	5.18 ± 0.09	2.97 ± 0.08	-2.21 ± 0.09	-7.91 ± 0.03
E30D	6.73 ± 0.05	4.81 ± 0.18	2.97 ± 0.14	-1.83 ± 0.09	-8.05 ± 0.05
E44D	6.42 ± 0.04	5.39 ± 0.11	2.88 ± 0.06	-2.51 ± 0.13	-7.92 ± 0.02
E48D	6.07 ± 0.07	5.28 ± 0.12	3.07 ± 0.08	-2.20 ± 0.14	-7.78 ± 0.03
E85D	6.27 ± 0.04	5.56 ± 0.11	2.79 ± 0.05	-2.78 ± 0.10	-8.41 ± 0.02
E101D	7.02 ± 0.04	5.29 ± 0.09	3.11 ± 0.09	-2.19 ± 0.07	-8.39 ± 0.03
I4V	6.71 ± 0.03	5.17 ± 0.14	3.11 ± 0.09	-2.06 ± 0.11	-7.85 ± 0.03
I5V	6.32 ± 0.11	4.90 ± 0.18	2.74 ± 0.12	-2.16 ± 0.20	-8.07 ± 0.06
I23V	7.12 ± 0.05	5.24 ± 0.09	2.81 ± 0.08	-2.43 ± 0.09	-8.37 ± 0.02
I38V	6.01 ± 0.06	4.35 ± 0.14	2.65 ± 0.08	-1.70 ± 0.11	-7.44 ± 0.04
I41V	6.37 ± 0.06	5.04 ± 0.17	3.02 ± 0.11	-2.02 ± 0.14	-7.60 ± 0.03
I45V	6.27 ± 0.04	5.07 ± 0.09	3.04 ± 0.06	-2.03 ± 0.09	-7.75 ± 0.03
I60V	6.77 ± 0.04	5.18 ± 0.07	2.91 ± 0.07	-2.27 ± 0.06	-7.96 ± 0.02
I72V	6.91 ± 0.04	5.64 ± 0.15	2.84 ± 0.15	-2.80 ± 0.10	-8.35 ± 0.04
I75V	6.47 ± 0.04	6.80 ± 0.32	3.00 ± 0.08	-3.80 ± 0.34	-8.01 ± 0.02
V16I	6.27 ± 0.07	5.17 ± 0.19	3.45 ± 0.13	-1.72 ± 0.17	-7.57 ± 0.06
V25I	6.10 ± 0.07	4.67 ± 0.23	2.95 ± 0.14	-1.71 ± 0.23	-7.38 ± 0.05
V55I	6.64 ± 0.06	4.70 ± 0.13	2.74 ± 0.11	-1.97 ± 0.11	-7.70 ± 0.03
V86I	6.60 ± 0.06	5.06 ± 0.11	2.82 ± 0.11	-2.24 ± 0.12	-7.75 ± 0.03
V91I	6.87 ± 0.06	4.95 ± 0.14	3.07 ± 0.11	-1.87 ± 0.08	-7.90 ± 0.04

Estos valores se han determinado realizando ajustes globales de las ecuaciones 2.45, 2.46, 2.47, 2.51 y 2.53 (conjunto de parámetros ajustables: ΔI^0 , $C_{1/2}$, $m_{1/2}$, $\ln k_{1/2}$, $m_{N \rightarrow \ddagger}$) a los datos cinéticos (*Chevron*) y los perfiles de equilibrio derivados de los ensayos de desplegamiento de doble salto. Los errores dados se han determinado usando el método de Monte-Carlo y describen las incertidumbres asociadas a proceso de ajuste. Los errores reales podrían ser algo mayores.

Tabla 5.3. Efecto de mutaciones en el equilibrio y cambios de la energía libre de activación para el plegamiento-desplegamiento de tiorredoxina de *E. coli* ([urea]=6.9 M). Además, valores de la constante de velocidad de desplegamiento a concentración de urea cero junto con el tiempo de vida media correspondiente.

PROTEÍNA	$\Delta\Delta G$ (kJ mol ⁻¹)	$\Delta\Delta G_{N \rightarrow \ddagger}$ (kJ mol ⁻¹)	$\Delta\Delta G_{U \rightarrow \ddagger}$ (kJ mol ⁻¹)	$\ln(k_U/s^{-1})$ a [urea]=0M	τ
WT				-16.17 ± 0.30	~ 4 months
D2E	0.1 ± 0.7	-1.1 ± 0.4	-1.2 ± 0.4	-16.59 ± 0.61	~ 6 weeks
D9E	-5.7 ± 0.6	-5.6 ± 0.3	0.1 ± 0.5	-14.58 ± 0.25	~ 3 weeks
D10E	-0.7 ± 0.7	-0.5 ± 0.3	0.2 ± 0.5	-14.92 ± 0.46	~ 5 weeks
D13E	0.8 ± 0.6	0.7 ± 0.4	-0.1 ± 0.3	-15.43 ± 0.52	~ 8 weeks
D43E	2.4 ± 0.7	1.9 ± 0.5	-0.5 ± 0.3	-17.50 ± 0.76	~ 15 months
D47E	1.2 ± 0.6	0.4 ± 0.3	-0.8 ± 0.3	-16.89 ± 0.50	~ 8 months
D61E	-2.6 ± 0.6	-1.5 ± 0.3	1.2 ± 0.5	-15.18 ± 0.36	~ 6 weeks
D104E	-1.4 ± 0.5	-1.8 ± 0.3	-0.4 ± 0.3	-15.89 ± 0.26	~ 3 months
E30D	-0.9 ± 0.5	-1.2 ± 0.3	-0.3 ± 0.3	-16.12 ± 0.45	~ 4 months
E44D	-2.7 ± 0.5	-2.4 ± 0.3	0.3 ± 0.5	-15.38 ± 0.21	~ 8 weeks
E48D	-4.5 ± 0.7	-3.9 ± 0.3	0.6 ± 0.5	-15.31 ± 0.28	~ 7 weeks
E85D	-3.6 ± 0.5	-1.6 ± 0.2	2.1 ± 0.4	-15.46 ± 0.15	~ 8 weeks
E101D	0.5 ± 0.5	0.5 ± 0.3	-0.0 ± 0.3	-17.19 ± 0.29	~ 11 months
I4V	-1.1 ± 0.4	-1.8 ± 0.3	-0.7 ± 0.3	-16.28 ± 0.29	~ 5 months
I5V	-2.9 ± 0.9	-2.2 ± 0.4	0.7 ± 0.6	-15.05 ± 0.41	~ 6 weeks
I23V	1.1 ± 0.5	0.8 ± 0.3	-0.3 ± 0.3	-16.46 ± 0.26	~ 5 months
I38V	-4.0 ± 0.6	-4.6 ± 0.3	-0.6 ± 0.4	-13.87 ± 0.27	~ 12 days
I41V	-2.7 ± 0.6	-3.4 ± 0.3	-0.7 ± 0.4	-15.37 ± 0.35	~ 8 weeks
I45V	-3.3 ± 0.5	-3.3 ± 0.2	-0.0 ± 0.3	-15.44 ± 0.22	~ 8 weeks
I60V	-0.8 ± 0.4	-1.3 ± 0.3	-0.5 ± 0.3	-15.89 ± 0.22	~ 3 months
I72V	-0.1 ± 0.5	0.1 ± 0.3	0.1 ± 0.3	-16.28 ± 0.37	~ 5 months
I75V	-3.0 ± 0.6	-2.1 ± 0.2	1.0 ± 0.3	-15.85 ± 0.24	~ 3 months
V16I	-3.4 ± 0.7	-4.1 ± 0.3	-0.7 ± 0.5	-16.31 ± 0.39	~ 5 months
V25I	-3.8 ± 0.7	-4.7 ± 0.3	-0.9 ± 0.5	-14.65 ± 0.41	~ 1 month
V55I	-1.3 ± 0.6	-2.3 ± 0.3	-0.9 ± 0.4	-15.03 ± 0.34	~ 6 weeks
V86I	-1.6 ± 0.6	-2.3 ± 0.3	-0.7 ± 0.4	-15.27 ± 0.36	~ 7 weeks
V91I	-0.3 ± 0.5	-1.2 ± 0.3	-0.9 ± 0.3	-16.40 ± 0.44	~ 5 months

Estos valores se han determinado de los ajustes globales de las ecuaciones 2.45, 2.46, 2.47, 2.51 y 2.53 (conjunto de parámetros ajustables: ΔI^0 , $C_{1/2}$, $m_{1/2}$, $\ln k_{1/2}$, $m_{N-\ddagger}$) a los datos cinéticos (*Chevron*) y los perfiles de equilibrio derivados de los ensayos de desplegamiento de doble salto. Los errores dados se han determinado

usando el método de Monte-Carlo y describen las incertidumbres asociadas al proceso de ajuste. Los errores reales podrían ser algo mayores. Los valores de las dos últimas columnas se deben de considerar como estimaciones ya que se han calculado realizando una extrapolación lineal a [urea]=0.

5.4.4 RESULTADOS

Es interesante resaltar algunos puntos en relación a los resultados obtenidos:

1) Existe una correlación buena (Figura 5.37) entre los efectos de mutaciones sobre la estabilidad termodinámica previamente determinados por DSC (Capítulos 3 y 4 publicados en Godoy-Ruiz *et al.*, 2004 y Godoy-Ruiz *et al.*, 2005, respectivamente; correspondientes a concentración de urea cero y alta temperatura: sobre 90 °C, valor de la T_m para la proteína *WT*) y los efectos de mutaciones sobre la estabilidad termodinámica derivados del análisis de experimentos de desnaturalización inducida por urea (realizados a 25°C y una concentración de urea igual al punto medio del perfil del desplegamiento en equilibrio para *WT*: 6.92 M). Esto muestra consistencia entre los estudios de DSC previamente realizados y los estudios de desnaturalización inducida por urea.

2) La velocidad de desplegamiento para tiorredoxina *WT* en agua (extrapolación a concentración de urea cero) estimada es $9.5 \cdot 10^{-8} \text{ s}^{-1}$, que resulta en un tiempo de vida media en agua alrededor de 4 meses (Tabla 5.3). Ésto quiere decir que tiorredoxina de *E. coli* es una proteína cinéticamente estable.

3) Muchas de las variantes simples de tiorredoxina han resultado tener velocidades de desplegamiento significativamente mayores que la de la forma *WT*, dando lugar tiempos de vida media del orden de algunos días en algunos casos (Tabla 5.3). Estos resultados, confirman que cambios en la energía libre de activación de unos pocos de kJ/mol pueden tener efectos sobre la estabilidad cinética significativos.

4) Existe una correlación excelente entre los efectos de mutaciones sobre la estabilidad termodinámica y los efectos de mutaciones sobre la energía libre de activación del desplegamiento (Figura 5.38). Es decir, para las mutaciones que hemos estudiado, los efectos sobre la estabilidad termodinámica se reflejan directamente en las velocidades de desplegamiento y, presumiblemente, sobre la

5. La selección natural para la estabilidad cinética...

estabilidad cinética también. Por lo tanto, la correlación que se muestra en la Figura 5.38 apoya nuestro argumento de que el nivel de estabilidad mínima permitido evolutivamente para tiorredoxina (unos pocos kJ/mol por debajo de la estabilidad termodinámica de la proteína *WT* de *E. coli*), refleja la existencia de selección natural para la estabilidad cinética.

5) La pendiente de la Figura 5.38 es próxima a la unidad. Por lo tanto, la mayoría de las mutaciones incluidas en la figura dan lugar a valores de ϕ próximos a la unidad, lo que implica que, de acuerdo con la actual interpretación de valores de ϕ (Matouschek *et al.*, 1989), la mayoría de las posiciones estudiadas se encuentran en regiones en las que se ha destruido su estructura nativa en el estado de transición del desplegamiento. Una posible explicación es que, dicho estado de transición se encuentre altamente desorganizado (y por lo tanto tenga un núcleo “plegado” comparativamente pequeño), reflejando ésto la existencia de selección natural para una elevada estabilidad cinética. Es decir, se puede considerar una energía libre de activación elevada para el desplegamiento si las interacciones en el estado nativo se encuentran significativamente optimizadas (como sugieren los resultados mostrados en la Figura 5.1) y, muchas de estas interacciones, se rompen en el estado de transición del desplegamiento. En otras palabras, uno de los factores que contribuye a la estabilidad cinética requerida es la optimización energética en el estado nativo del ambiente que rodea a residuos en regiones que se destruyen en el estado de transición del desplegamiento.

6) Existe una correlación significativa entre los efectos de mutaciones sobre la energía libre de activación del desplegamiento y la frecuencia de aparición de los residuos involucrados en alineamientos de secuencias (Figura 5.39-C). Esta correlación es significativa, sin embargo, no se observa con los efectos de mutaciones sobre la energía libre de activación del plegamiento (Figura 5.39-C). La correlación que se muestra en la Figura 5.39-B es la esperada, ya que i) existe una correlación estabilidad/alineamientos de secuencias similar para los efectos de mutaciones sobre la estabilidad termodinámica (energía libre del desplegamiento en equilibrio), como se muestra en la Figura 5.39-A; ii) los efectos de mutaciones sobre la estabilidad termodinámica muestran una correlación excelente con los efectos de mutaciones sobre la energía libre de activación del desplegamiento (Figura 5.38 y punto 4) de este apartado). En cualquier caso, las correlaciones que se muestra en la Figura 5.39 son consistentes

con el título de este capítulo, es decir, que la existencia de selección natural para estabilidad cinética es el origen probable de la existencia de correlaciones entre efectos de mutaciones sobre la energética de las proteínas y la frecuencia de aparición de aminoácidos en alineamientos de secuencias.

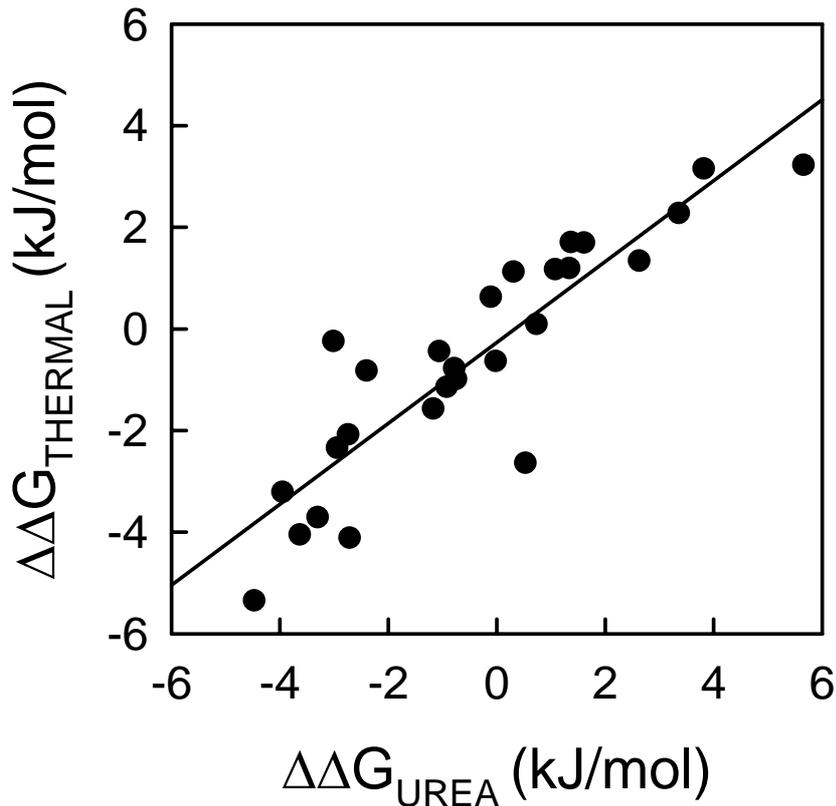


Figura 5.37: Correlación entre los efectos de mutaciones sobre el cambio de energía libre para el desplegamiento de tiorredoxina de *E. coli* determinados mediante estudios de desnaturalización térmica por DSC ($\Delta\Delta G_{\text{THERMAL}}$); datos tomados de los Capítulos 3 y 4, Godoy-Ruiz *et al.*, 2004 y Godoy-Ruiz *et al.*, 2005, respectivamente) y estudios de desnaturalización inducida por urea ($\Delta\Delta G_{\text{UREA}}$; este capítulo). Las mutaciones incluidas son las que se han mostrado en la Figura 5.1 (los efectos de la mutación se dan en las direcciones I \rightarrow V y E \rightarrow D). Hay que resaltar que los dos conjuntos de valores de $\Delta\Delta G$ corresponden a condiciones diferentes (sobre 90°C y concentración cero de urea en el caso de $\Delta\Delta G_{\text{THERMAL}}$ y 6.9 M urea para el caso de $\Delta\Delta G_{\text{UREA}}$).

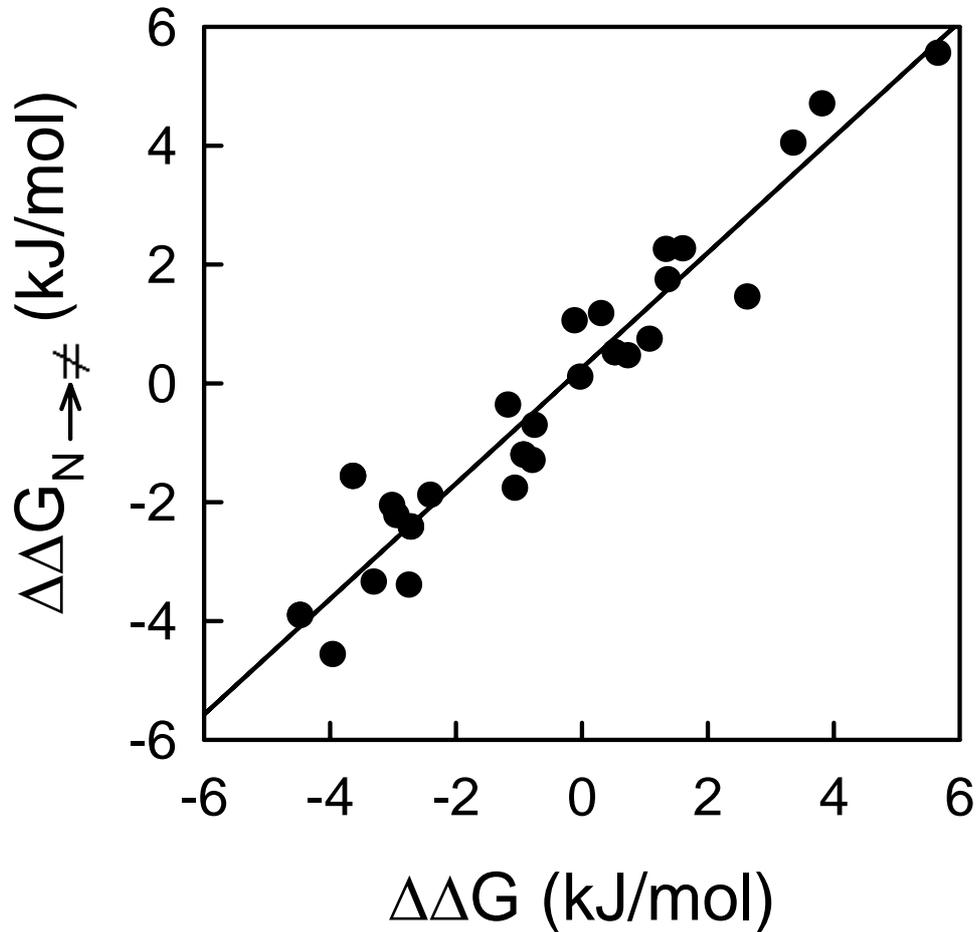


Figura 5.38: Correlación entre los efectos de mutaciones sobre el cambio de energía libre para el desplegamiento en equilibrio de tiorredoxina de *E. coli* y los efectos de mutaciones sobre la energía de activación del desplegamiento. Los datos se han determinado del análisis de los perfiles de equilibrio y cinéticos, de la desnaturalización inducida por urea (Figuras 5.33-5.36). Las mutaciones incluidas son las que se han mostrado en la Figura 5.1 (los efectos de la mutación se dan en las direcciones I→V y E→D). La pendiente de la representación es próxima a la unidad, sugiriendo que la mayoría de las posiciones mutadas se encuentran en regiones cuya estructura nativa se encuentra destruída en el estado de transición del desplegamiento.

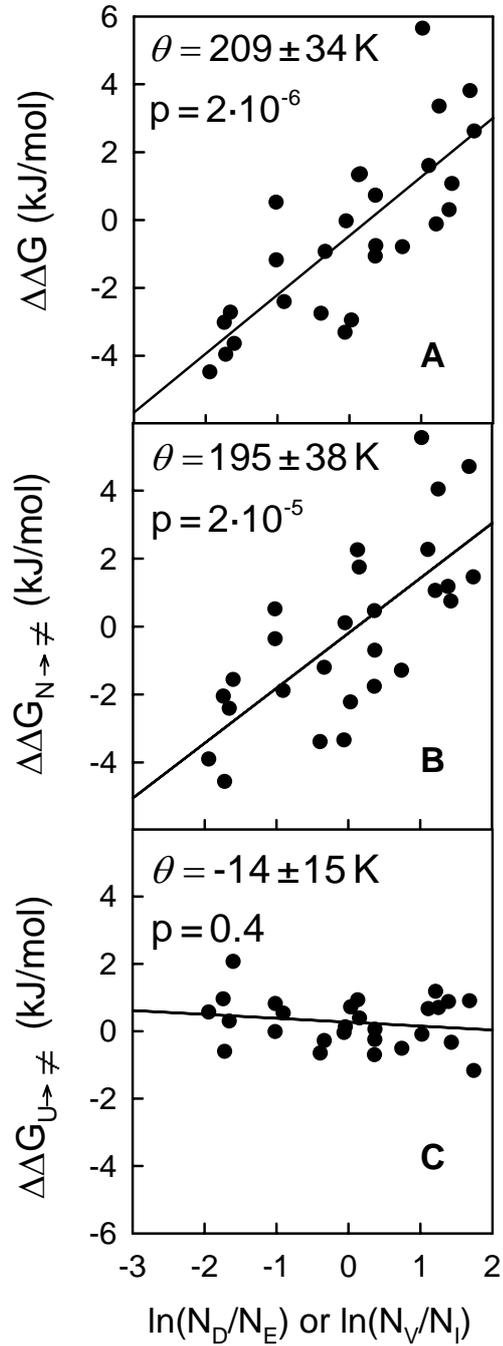


Figura 5.39: Correlaciones entre los efectos de mutaciones en equilibrio y los cambios de energía libre de activación para el plegamiento-desplegamiento de tiorredoxina de *E. coli* y la frecuencia de aparición de residuos en alineamientos de secuencias. Se dan los valores de θ [ecuación 5.1] y la significancia estadística (p). Las mutaciones incluidas son las que se han mostrado en la Figura 5.1 (los efectos de las mutaciones se dan en las direcciones $I \rightarrow V$ y $E \rightarrow D$)

5.5 CONCLUSIONES

Trabajos teóricos y experimentales recientes (Akmal y Munoz, 2004; Chan *et al.*, 2004; Kubelka *et al.*, 2004; Scalley-Kim y Baker, 2004) han sugerido que la existencia de una barrera de energía libre elevada que separe el estado nativo y desplegado, no es un requerimiento fundamental de la física del plegamiento de proteínas, sino más probablemente, el resultado de la selección natural. Se han publicado evidencias de barreras marginales en el plegamiento de algunas proteínas pequeñas (García-Mira *et al.*, 2002; Yang y Gruebele, 2003; Munoz y Sanchez-Ruiz, 2004; Yang y Gruebele, 2004; Naganathan *et al.*, 2005; Sadqi *et al.*, 2006) y se ha encontrado que proteínas diseñadas computacionalmente pliegan más rápido que las correspondientes proteínas naturales, aunque no se realice ningún tipo de selección para aumentar la eficiencia del plegamiento (Scalley-Kim y Baker, 2004). Un análisis de barrera-variable del desplegamiento inducido por temperatura en equilibrio (Munoz y Sanchez-Ruiz, 2004) y las velocidades de desplegamiento que se muestran en este trabajo (Tablas 5.2 y 5.3) indican una barrera de energía libre significativa para el desplegamiento de tiorredoxina de *E. coli*. De hecho, dicha barrera parece haber sido seleccionada naturalmente (posiblemente para garantizar una elevada estabilidad cinética *in vivo*), como indica el hecho de que los resultados de los análisis estadísticos de los alineamientos de secuencias correlacionen con los efectos de mutaciones sobre la energía libre de activación del desplegamiento experimentales (Figura 5.39).

5. La selección natural para la estabilidad cinética...

5.6 BIBLIOGRAFÍA

- Akmal, A. and V. Munoz (2004). *Proteins* **57**(1): 142-52.
- Callen, H.B. (1985). New York, John Willey & Sons.
- Campos, L.A., M.M. Garcia-Mira, R. Godoy-Ruiz, J.M. Sanchez-Ruiz and J. Sancho (2004). *J Mol Biol* **344**(1): 223-37.
- Chan, H.S., S. Shimizu and H. Kaya (2004). *Methods Enzymol* **380**: 350-79.
- Dobson, C.M. (1999). *Trends Biochem Sci* **24**(9): 329-32.
- Dokholyan, N.V. and E.I. Shakhnovich (2001). *J Mol Biol* **312**(1): 289-307.
- Elias, S.J.A. and A. Holmgren (2000). *Eur J Biochem* **267**: 6102-6109.
- Ferrer, P., C. Chang, D. Ott, A. Wlodawer and A. Plückthun (2004). *J Mol Biol* **344**: 179-193.
- Garcia-Mira, M.M., M. Sadqi, N. Fischer, J.M. Sanchez-Ruiz and V. Munoz (2002). *Science* **298**(5601): 2191-5.
- Georgescu, R.E., J.-H. Li, M.E. Goldberg, M.L. Tasayco and A.F. Chaffotte (1998). *Biochemistry* **1998**: 10286-10297.
- Godoy-Ruiz, R., R. Perez-Jimenez, B. Ibarra-Molero and J.M. Sanchez-Ruiz (2004). *J Mol Biol* **336**(2): 313-8.
- Godoy-Ruiz, R., R. Perez-Jimenez, B. Ibarra-Molero and J.M. Sanchez-Ruiz (2005). *Biophys J* **89**(5): 3320-31.
- Godoy-Ruiz, R., F. Ariza, D. Rodriguez-Larrea, R. Perez-Jimenez, B. Ibarra-Molero and J.M. Sanchez-Ruiz (2006). *J Mol Biol* **In press**.
- Gruebele, M. (2005). *C R Biol* **328**(8): 701-12.
- Hammarström, P., R.L. Wiseman, E.T. Powers and J.W. Kelly (2003). *Science* **299**: 713-716.
- Ibarra-Molero, B. and J.M. Sanchez-Ruiz (1996). *Biochemistry* **35**(47): 14689-702.
- Ibarra-Molero, B. and J.M. Sanchez-Ruiz (1997). *Biochemistry* **36**(31): 9616-24.
- Jaswal, S.S., J.L. Sohl, J.H. Dans and D.A. Agard (2002). *Nature* **415**: 343-346.
- Jayaraman, s., D. Gantz and O. Gursky (2005). *Biochemistry*(44).

5. La selección natural para la estabilidad cinética...

- Kubelka, J., J. Hofrichter and W.A. Eaton (2004). *Curr Opin Struct Biol* **14**(1): 76-88.
- Kumar, J.K., S. Tabor and C.C. Richardson (2004). *Proc Natl Acad Sci U S A* **101**(11): 3759-3764.
- Lehmann, M., L. Pasamontes, S.F. Lassen and M. Wyss (2000). *Biochim Biophys Acta* **1543**(2): 408-415.
- Leikina, E., M.V. Merts, N. Kuznetsova and S. Leikin (2002). *Proc Natl Acad Sci U S A* **99**: 1314-1318.
- Lynch, S.M., S.A. Boswell and W. Colon (2004). *Biochemistry* **43**: 16525-16531.
- Manning, M. and W. Colon (2004). *Biochemistry* **43**: 11248-111254.
- Matouschek, A., J.T. Kellis, Jr. and A.R. Fersht (1989). *Nature* **340**: 122-6.
- Mucke, M. and F.X. Schmid (1994). *Biochemistry* **33**(43): 12930-5.
- Munoz, V. and J.M. Sanchez-Ruiz (2004). *Proc Natl Acad Sci U S A* **101**(51): 17646-51.
- Naganathan, A.N., J.M. Sanchez-Ruiz and V. Munoz (2005). *J Am Chem Soc* **127**(51): 17970-1.
- Persikov, A.V. and B. Brodsky (2002). *Proc Natl Acad Sci U S A* **99**: 1101-1103.
- Petrassi, M.M., S.M. Johnson, H.E. Purkey, K.P. Chiang, T. Walkup, X. Jinag, E.T. Powers and J.W. Kelly (2005). *J Am Chem Soc* **127**: 6662-6671.
- Plaza del Pino, I.M., B. Ibarra-Molero and J.M. Sanchez-Ruiz (2000). *Proteins* **40**(1): 58-70.
- Rodriguez-Larrea, D., S. Minning, T.V. Borchert and J.M. Sanchez-Ruiz (2006). *J Mol Biol* **360**(3): 715-24.
- Sadqi, M., D. Fushman and V. Munoz (2006). *Nature*.
- Scalley-Kim, M. and D. Baker (2004). *J Mol Biol* **338**(3): 573-83.
- Schnyrova, V.L., G.G. Zhadan, C. Cobaleda, I. Muñoz-Barroso and E. Villar (1997). *Arch Biochem Biophys* **341**: 89-97.
- Shortle, D. (2003). *Protein Sci* **12**(6): 1298-1302.
- Sohl, J.L., S.S. Jaswal and D.A. Agard (1998). *Nature* **395**: 817-819.

5. *La selección natural para la estabilidad cinética...*

Steipe, B., B. Schiller, A. Plückthun and S. Steinbacher (1994). *J Mol Biol* **240**: 182-192.

van Holde, K.E., W.C. Johnson and P.S. Ho (1998). New Jersey, Simon & Schuster/ A Viacom Company.

Wilson, J., R.F. Kelley, W. Shalongo, D. Lowery and E. Stellwagen (1986). *Biochemistry* **25**(23): 7560-6.

Wiseman, R.L., S.M. Johnson, M.S. Kelker, T. Foss, I.A. Wilson and J.W. Kelly (2005). *J Am Chem Soc* **127**: 5540-5541.

Yang, W.Y. and M. Gruebele (2003). *Nature* **423**(6936): 193-7.

Yang, W.Y. and M. Gruebele (2004). *Biophys J* **87**(1): 596-608.

5. La selección natural para la estabilidad cinética...

**6. Estudio cristalográfico de
tiorredoxina silvestre y las variantes
D2E, D9E, D10E, D13E, D43E, E44D,
D47E, E48D y E85D**

6 ESTUDIO CRISTALOGRAFICO DE TIORREDOXINA DE *E. COLI* Y LAS VARIANTES D2E, D9E, D10E, D13E, D43E, E44D, D47E, E48D Y E85D

En esta sección hemos realizado un análisis de las estructuras de tiorredoxina silvestre (*WT*) y de las variantes, mutantes simples en tiorredoxina de ácidos carboxílicos, D2E, D9E, D10E, D13E, D43E, E44D, D47E, E48D y E85D, resueltas hasta el momento. Estas variantes de tiorredoxina son muy interesantes, pues cada una de ellas contiene una mutación ($D \rightarrow E$ o $E \rightarrow D$) en la que participan residuos altamente conservativos en un alineamiento de secuencias obtenido con BLAST 2, realizando una búsqueda para tiorredoxina de *E. coli*. Los residuos glutamato (E) y aspartato (D) se caracterizan por ser física-químicamente similares y pueden sustituirse fácilmente uno por el otro en el curso de la evolución de proteínas (Godoy-Ruiz *et al.*, 2004). En los Capítulos 3 y 5 hemos visto como las mutaciones $E \rightarrow D$ y $D \rightarrow E$ tienen un efecto pequeño, pero significativo sobre la estabilidad de tiorredoxina y su relación con la evolución. Un estudio cristalográfico de las interacciones en las que participan estos residuos ha sido necesario para conocer más sobre la relación entre estabilidad, estructura y evolución de proteínas, además de que nos ha ofrecido información muy útil debido al polimorfismo y elevada flexibilidad que poseen tiorredoxina y las variantes estudiadas.

El problema lo hemos abordado cristalizando y resolviendo la estructura de tiorredoxina *WT* y, hasta el momento, las 9 variantes, mutaciones de ácidos carboxílicos (D2E, D9E, D10E, D13E, D43E, E44D, D47E, E48D y E85D). La información estructural derivada de la comparación de cada variante respecto a la proteína *WT*, se ha obtenido aplicando diferentes mecanismos: análisis comparativos de las desviaciones cuadráticas medias (R.M.S.D.) de las cadenas principales y totales (cadenas principales más cadenas laterales), estudio de las interacciones (corto y largo alcance) en las que participa el residuo mutado, comparativa de las diferencias en área superficial accesible al disolvente (Δ ASA) de las variantes de tiorredoxina respecto a tiorredoxina *WT*, estudio de los factores-B de la cadena principal y por último, un análisis energético y estructural de los cambios estructurales experimentalmente observados en los mutantes de tiorredoxina, sugeridos por alineamientos de secuencias.

6.1 DESCRIPCIÓN GENERAL DE LA ESTRUCTURA DE TIORREDOXINA

En la Figura 6.1 se muestra un diagrama de la conformación del esqueleto de tiorredoxina *WT* (2H6X.pdb). La molécula está formada por un núcleo central de tres láminas paralelas y dos antiparalelas, rodeadas por 4 hélices. En las Figuras 6.1 y 6.2, se puede observar que la primera lámina (β_1) está formada por los residuos 2 al 8. La cadena continúa con una hélice (α_1) compuesta por los residuos del 11 al 18 y sigue en la lámina central (β_2), formada por los residuos 22 al 29. Los residuos 29 al 37, que incluyen el puente disulfuro del centro activo, constituyen el extremo carboxilo de β_2 . El último residuo de este conjunto, constituye el comienzo de la hélice de mayor longitud (α_2), compuesta por los residuos del 34 al 49, cuyo eje es paralelo a la lámina β_2 . La lámina β_3 , desde los residuos 53 al 58, se encuentra paralelamente situada entre β_1 y β_2 . Los residuos 59 a 72 forman un lazo que contiene un giro de α -hélice, α_3 . Hasta ahora, se ha descrito la estructura de los residuos 1 al 72, que constituyen una región de la molécula de tiorredoxina. El resto de residuos constituyen una segunda región de la molécula, dos láminas antiparalelas (β_4 y β_5) y una hélice carboxilo-terminal (α_4). Las dos regiones descritas se mantienen unidas mediante enlaces de hidrógeno entre las láminas β_4 y β_2 , quedando antiparalelamente alineadas en el conjunto de la molécula (Figura 6.2). El conjunto de láminas β presenta una rotación de aproximadamente 80° . 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. En las regiones existentes entre el conjunto de láminas β y las hélices α sólo existen residuos hidrofóbicos.

Se pueden diferenciar claramente 4 giros reversos en la estructura (Figura 6.2). De este modo, la molécula de tiorredoxina tiene más de un 75% de los residuos participando en elementos de estructura secundaria, número muy elevado si se tiene en cuenta el bajo peso molecular de la proteína. El alto contenido de estructura secundaria proporciona una estabilidad elevada a la proteína, debido al elevado número de enlaces de hidrógeno intramoleculares, compatible con la estabilidad térmica y las concentraciones de urea necesarias para que se produzca la desnaturalización reversible (Godoy-Ruiz *et al.*, 2004, Godoy-Ruiz *et al.*, 2005, Godoy-Ruiz, 2006, in press).

Tiorredoxina en $P6_1$, se empaqueta como dos moléculas independientes (denominadas A y B), muy similares en conformación, pero diferentes en detalles atómicos (RMSD de la cadena principal igual a 1.01 \AA^2). Las moléculas de metilpentanediol a menudo se empaquetan en los lazos y estabilizan su estructura.

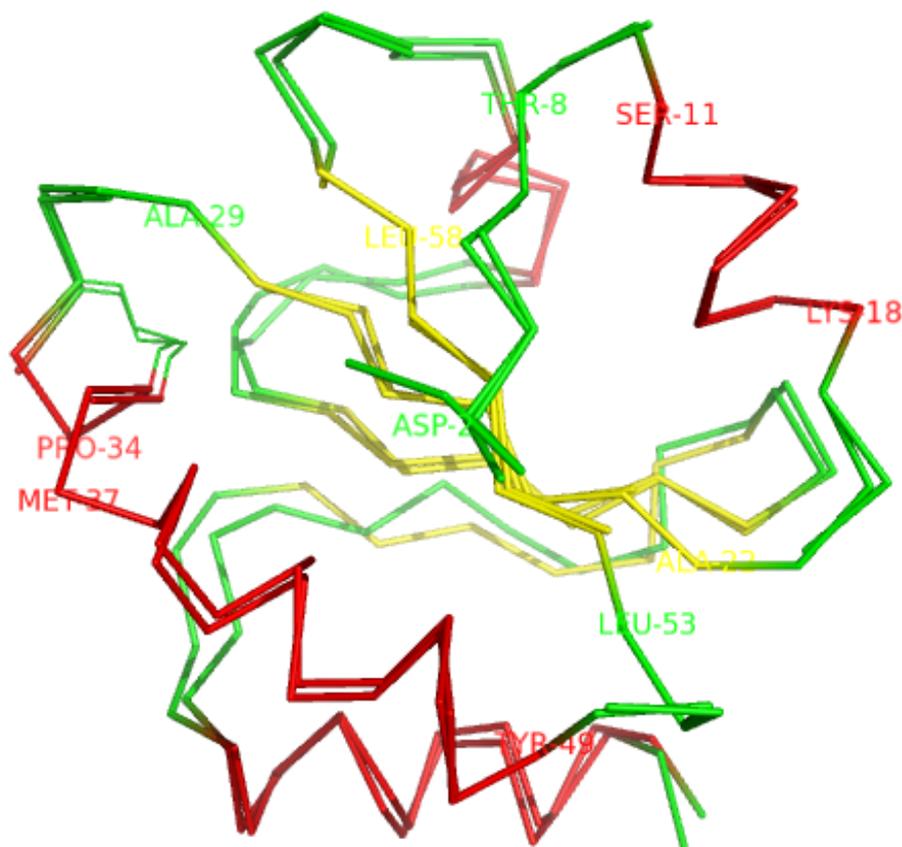


Figura 6.1: Imagen ilustrativa del esqueleto de los dos monómeros de tiorredoxina silvestre (2H6X.pdb) superpuestos. Se encuentran marcados los residuos en los que comienza y acaban los distintos elementos de estructura secundaria.

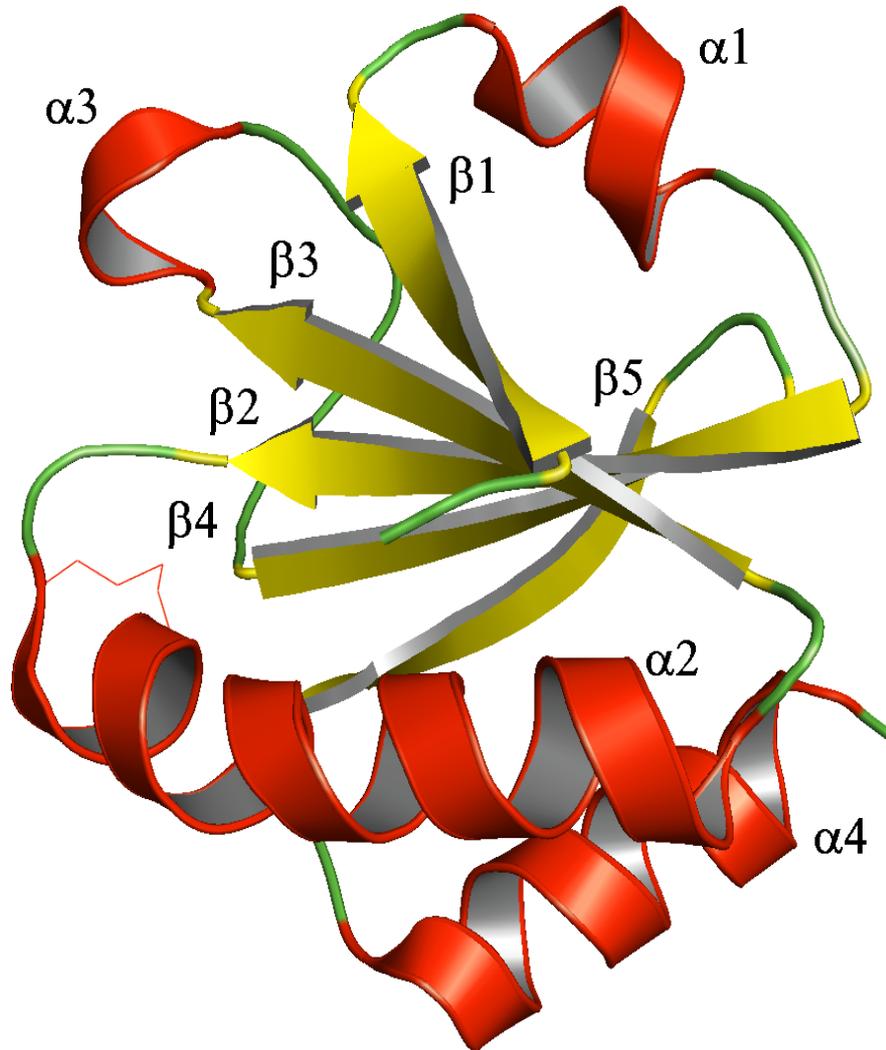


Figura 6.2: Imagen ilustrativa del esqueleto de tiorredoxina silvestre (monómero A de 2H6X.pdb) en el que se destaca los distintos elementos de estructura secundaria.

6.2 CRISTALIZACIÓN DE LA PROTEÍNA, RECOGIDA DE DATOS Y RESOLUCIÓN DE ESTRUCTURA

Hemos cristalizado y resuelto 9 estructuras de los 27 variantes mutantes de tiorredoxina que se han estudiado (Capítulos 3, 4 y 5), además de la forma silvestre (*WT*), utilizando las condiciones descritas en el apartado 2.2.8.4 de Materiales y métodos (Tabla 2.5), y empleando la técnica de contradifusión en capilar (García-Ruiz, 2003 y apartado 2.2.8.4; ver Figura 2.22 en Materiales y métodos).

Los nueve mutantes y la proteína silvestre han cristalizado en el mismo grupo espacial: P61, coincidente con el grupo espacial de otro mutante previamente depositado en el PDB (1THO.pdb), pero con distinta celdilla unidad. Tiorredoxina *WT* (2H6X.pdb) y las nueve variantes de estructura resuelta: D2E (2H74.pdb), D9E (2H70.pdb), D10E (2H76.pdb), D13E (2H75.pdb), D43E (2H73.pdb), E44D (2H6Z.pdb), D47E (2H71.pdb), E48D (2H6Y.pdb) y E85D (2H72.pdb), han cristalizado en condiciones similares a las descritas por Katti *et al.*, 1990, en un amplio rango de pH. Los resultados obtenidos se resumen en la Tabla 6.1.

Para llevar a cabo la recogida de datos de difracción sin manipulación de los cristales (Figura 2.26 en *Materiales y métodos*) se ha seguido el protocolo descrito en el apartado 2.2.9.1 de *Materiales y métodos*. Se han utilizado los programas *CNS* (Brünger *et al.*, 1998), MolRep (perteneciente al paquete de programas CCP4, Collaborative Computational Project, Number 4, <http://www.ccp4.ac.uk/main.html>, Murshudov *et al.*, 1997), para encontrar la solución deemplazamiento molecular. El refinamiento se ha llevado a cabo con los programas los programas *CNS* (Brünger *et al.*, 1998) y *REFMAC* (del paquete de CCP4). Entre cada ciclo de refinamiento, la reconstrucción manual se ha realizado con los programas gráficos XtalView (McRee, 1993) y *COOT* (apartado 2.2.9.3 de *Materiales y métodos*). Para posicionar los aminoácidos y las cadenas laterales se han utilizado los mapas de densidad electrónica $2|F_0| - |F_C|$ contorneados a 1σ y los mapas $|F_0| - |F_C|$ contorneados a 3σ . El resumen de los datos y estadísticas cristalográficas se recoge en la Tabla 6.2. Los mapas de densidad electrónica han sido siempre continuos para la cadena principal y la mayoría de las cadenas laterales. En ausencia de densidad electrónica de las cadenas laterales, se ha cortado átomo a átomo y se ha comprobado su

desaparición tras un nuevo ciclo de refinamiento. A continuación se enumeran los aminoácidos cortados, en distinta extensión en tioredoxina WT: Lys-100A, Glu-101A y Glu-44B; en D2E: Lys-52A, Lys-100A, Lys-69A, Lys-3B, Arg-73B y Lys-82B; en D9E: Lys-69B y Arg-73B; D10E está completo; en D13E hemos cortado: Lys-100A, Lys-3B, Lys-69B y Glu-85B; D43E está completo; en E44D hemos cortado: Lys-A3, Lys-52A, Lys-90A, Lys-100A, Lys-69B, Lys-96B y Lys-100B; en D47E hemos cortado: Lys-100A, Lys-69B y Lys-90B; en E48D: Arg-73B y Lys-100A; y en E85D: Lys-52A y Lys-100A. La conformación del puente disulfuro, que conecta los residuos Cys32S^γ y Cys35S^γ, siempre ha quedado claramente definida en $2|F_0| - |F_C|$ y $|F_0| - |F_C|$. Las moléculas de agua se han añadido, después de varios ciclos de refinamiento manual y reconstrucción, con el criterio de observación de por lo menos, 1σ de densidad $2|F_0| - |F_C|$, 3σ de densidad $|F_0| - |F_C|$ y si se encontraban a la distancia de enlace de hidrógeno (al menos 3.5 Å) de la proteína o de otras moléculas de agua. El modelo final contiene la proteína completa (residuos 1-108) y 90 moléculas de H₂O, para el caso de tioredoxina WT; 63 moléculas de H₂O para D2E; 61 moléculas de H₂O para D9E; 102 moléculas de H₂O para D10E; 101 moléculas de H₂O para D13E; 82 moléculas de H₂O para D43E; 102 moléculas de H₂O para E44D; 90 moléculas de H₂O para E48D y 80 moléculas de H₂O para E85D.

Tabla 6.1. Resumen de las estructuras y condiciones de cristalización de tiorredoxina y proteínas mutantes depositadas en el PDB. La zona sombreada corresponde a los datos aportados en este trabajo. Las condiciones de cristalización similares se han coloreado en rojo y azul.

PDB ID	Proteína	G.E.	a-b-c [Å]	α - β - γ [°]	Res.[Å]	Condiciones	Método	Referencia
1KEB	P40S.	P 1	29.32, 37.50,50.74	69.25, 79.71, 85.390	1.8	pH 3.8, Etanol 25% CuAc ₂ 10 mM,NaAc 100 mM; T=27 °C	Difusión- Fase vapor	Rudresh <i>et al.</i> , 2002
1SRX	Forma oxidada	C 2	89.70, 51.10, 60.30	90.00, 113.50, 90.00	2.8	pH 3.8, Etanol 25%, CuAc ₂ 10 mM, NaAc 100 mM	Difusión- Fase vapor	Holmgren <i>et al.</i> , 1975
1THO	Inserción	P 61	78.40, 78.40, 35.10	90.00, 90.00, 120.00	2.3			Nikkola <i>et al.</i> , 1993
1TXX	P33V, G34W	P 43 21 2	42.53, 42.53, 105.24	90.00	2.2	pH 4.2, PEG 2000 10% (p/v) metil-éter, NaSuc.100 mM, CuAc ₂ 2 mM	Difusión- Fase vapor	Schultz <i>et al.</i> , 1999
2TIR	K36E	P 21 21 21	50.80, 26.80, 80.70	90.00	2.0	pH 4.8, PEG 6000 17% (p/v), NaAc 15mM, CuAc ₂ 1 mM ; T=4°C	Difusión- Fase vapor	Nikkola <i>et al.</i> , 1993
2TRX	Silvestre	C 2	89.50, 51.06, 60.45	90.00, 113.50, 90.00	1.68	Rango de pH: 3.7-4.1, Butanol: 0.001-0M, MPD 48-50%, NaAc 0.01-0M, CuAc ₂ 1 mM ; T=4°C	Microdiálisis	Katti <i>et al.</i> , 1990
2FCH	G74S	P 21 21 21	79.65, 88.80, 118.08	90.00	2.6	pH 5.4, MPD 60%, NaAc 15 mM,	Contradifusión	

						CuAc ₂ 1 mM ; T=4°C	
1ZZY	L7V	P1	33.66, 36.60, 42.84	78.43, 67.38, 88.39	2.5	pH 3.8 Etanol 25%, CuAc ₂ 10 mM, NaAc 100 mM	Contradifusión
2FD3	P34H	P1 21 1	31.14, 90.04, 36.33	90.00, 112.99, 90.00	2.45	pH 8.0, MPD 60%, Hepes 15 mM, CuAc ₂ 1 mM ; T=4°C	Contradifusión
2H6X	WT		102.73, 102.73, 42.87		2.6	pH 6.9	
2H6Y	E48D		101.61, 101.61, 38.66		2.4	pH 3.5	
2H6Z	E44D		102.92, 102.92, 42.56		2.25	pH 7.0	
2H70	D9E		103.03, 103.03, 40.28		2.7	pH 7.0	
2H71	D47E	P6 ₁	102.95, 102.95, 41.84	90.00, 90.00, 120.00	2.2	pH 7.0	Contradifusión
2H72	E85D		103.51, 103.51, 41.40		2.25	pH 7.0	
2H73	D43E		102.99, 102.99, 42.45		2.45	pH 7.0	
2H74	D2E		103.50, 103.50, 103.50		2.4	pH 3.5	
2H75	D13E		103.08, 103.08, 42.59		2.2	pH 7.0	
2H76	D10E		103.17, 103.17, 41.84		2.25	pH 4.1	
						MPD 60%	
						NaAc/Hepes: 15 mM	
						CuAc ₂ : 1 mM ; T=4°C	

6.3 COMPARACIÓN DE LA ESTRUCTURA DE TIORREDOXINA SILVESTRE CON LA DE SUS VARIANTES

Las proteínas no son rígidas. Es conocido que el estado nativo de las proteínas no tiene una única conformación, sino un conjunto de conformaciones pobladas bajo condiciones fisiológicas. En muchos casos, esta flexibilidad es necesaria para desarrollar la función biológica (Hermoso Dominguez y Martínez-Ripoll, 2003, Gerstein *et al.*, 1994). Por lo tanto, caracterizar la flexibilidad de una proteína nativa nos puede permitir entender mejor la relación entre estabilidad y estructura. Dentro de este objetivo, hemos realizado en primer lugar un estudio comparativo de las estructuras de tiorredoxina silvestre (*WT*) y las nueve variantes (D2E, D9E, D10E, D13E, D43E, E44D, D47E, E48D y E85D), ha consistido en una observación de las desviaciones que presentan las cadenas principales de las diferentes estructuras resueltas. La comparación de tiorredoxina *WT* y variantes (P6₁) con la estructura de tiorredoxina (C2), nos ha permitido identificar zonas con desviaciones significativas. Seguidamente, la comparación de tiorredoxina *WT* con sus variantes (ambas en P6₁), nos ha mostrado cómo a pesar de encontrarse en el mismo grupo espacial, siguen existiendo desviaciones importantes entre sus cadenas principales.

6.3.1 R.M.S.D. DE LA ESTRUCTURA DE 2TRX (C2)

COMPARADA CON LA DE 2H74, 2H70, 2H76, 2H75, 2H73, 2H6Z, 2H71, 2H6Y Y 2H72 (P6₁)

Hemos comparado la estructura de tiorredoxina silvestre (*WT*) cristalizada en el grupo espacial C2 (2TRX.pdb) con la estructura de cada una de las variantes de tiorredoxina, mutaciones de ácidos carboxílicos, cristalizadas en el grupo espacial P6₁: D2E (2H74.pdb), D9E (2H70.pdb), D10E (2H76.pdb), D13E (2H75.pdb), D43E (2H73.pdb), E44D (2H6Z.pdb), D47E (2H71.pdb), E48D (2H6Y.pdb) y E85D (2H72.pdb). En ambos grupos espaciales, C2 y P6₁, la unidad asimétrica está compuesta por dos monómeros de tiorredoxina denominados A y B.

En la Tabla 6.3, se puede comprobar que la desviación cuadrática media (R.M.S.D.) de la cadena principal del monómero A de tiorredoxina *WT* (2TRX-A) comparado con la cadena principal del monómero B (2TRX-B), es en muchos casos superior al R.M.S.D. de la cadena principal 2TRX-A comparada con la

cadena principal de los monómeros A de las estructuras de las variantes de tiorredoxina. Las filas sombreadas en la Tabla 6.2 se representan en la Figura 6.3. En ellas, se puede observar que las diferencias existentes entre 2TRX-A y 2TRX-B son mayores que las existentes entre 2TRX-A comparado con el monómero A de cada una de las variantes. Un comportamiento similar se ha observado cuando se compara 2TRX-B con el monómero B de cada una de las variantes. Es decir, comprobamos que existen una serie de variaciones sistemáticas al comparar 2TRX-A con las cadenas A de todos los mutantes; y 2TRX-B con las cadenas B de todos los mutantes. Ésto puede ser debido al distinto empaquetamiento de 2TRX (C2) y las variantes de tiorredoxina (P6₁) que estamos estudiando.

Por lo tanto, desviaciones de las proteínas variantes de tiorredoxina, inferiores a las existentes entre los 2TRX-A y B, no son significativas. Ésta es una de las razones por las que se ha cristalizado y resuelto la estructura de tiorredoxina *WT* en el mismo grupo espacial que el resto de variantes y resolver su estructura. En el siguiente apartado, se realiza una comparativa de las estructuras de tiorredoxina *WT* cristalizada en el grupo espacial P6₁ (2H6X.pdb), con las estructuras del resto de variantes de tiorredoxina, mutaciones de ácidos carboxílicos, también cristalizadas en el grupo espacial P6₁.

6.3.2 R.M.S.D. DE LA ESTRUCTURA DE 2H6X (P6₁) COMPARADA CON LA DE 2H74, 2H70, 2H76, 2H75, 2H73, 2H6Z, 2H71, 2H6Y Y 2H72 (P6₁)

Hemos comparado la estructura de tiorredoxina silvestre (*WT*) cristalizada en el grupo espacial P6₁ (2H6X) con tiorredoxina cristalizada en el grupo espacial C2 (2TRX). Las desviaciones cuadráticas medias (R.M.S.D.) de la cadena principal y total (cadena principal más cadenas laterales) para los monómeros A son 0.79 y 1.21, respectivamente. Para los monómeros B, son 1.07 y 1.58, respectivamente (Figura 6.4). Estos valores de R.M.S.D., nos indican la existencia de desviaciones significativas entre las dos estructuras de tiorredoxina (2TRX y 2H6X). Sin embargo, las diferencias son mucho más importantes entre los monómero A y B de 2H6X (R.M.S.D.=1.01). Debido a ello, la comparación de las estructuras de las distintas variantes de tiorredoxina con la de tiorredoxina *WT* se realiza entre el mismo tipo de monómeros (Figura 6.5 y Tablas 6.4 y 6.5). De este modo, hemos comparado la estructura de tiorredoxina *WT* (2H6X) con la

Tabla 6.2. Tabla estadística resumen de la recogida de datos y refinamiento de tioredoxina *WT* y variantes. Todas las estructuras

Código PDB/proteína	2H6X/WT	2H74/D2E	2H70/D9E	2H76/D10E	2H75/D13E	2H73/D43E	2H71/D47E	2H6Z/E44D	2H6Y/E48D	2H72/E85D
Wavelength (Å)	1.5418	1.5418	1.5418	1.5418	1.5418	1.5418	1.0724	1.5418	1.5418	1.0724
Temperature (K)	100.0	100.0	100.0	100.0	100.0	100	100.0	100	100	100.0
Unit cell constants	102.73	103.50	103.03	103.17	103.08	102.99	102.95	102.92	101.61	103.51
a, b, c (Å)	102.73	103.50	103.03	103.17	103.08	102.99	102.95	102.92	101.61	103.51
	42.87	41.78	40.28	41.83	42.59	42.45	41.84	42.56	38.66	41.40
Resolution range (Å)	2.60 (2.65)	2.40 (2.45)	2.70 (2.80)	2.25 (2.30)	2.20 (2.25)	2.45 (2.50)	2.20 (2.24)	2.25 (2.30)	2.40 (2.45)	2.25 (2.29)
Completeness (%)	99.8 (100.0)	99.7 (97.9)	99.6 (98.9)	99.9 (100.0)	99.9 (100.0)	99.9 (100.0)	99.0 (98.9)	99.9 (100.0)	93.9 (99.5)	99.1 (95.5)
R_{merg} (%)	7.87	2.42	3.66	2.41	2.58	3.19	5.60	2.20	0.048	0.059
Mean intensity $\langle I/\sigma(I) \rangle$	9.08 (2.63)	32.58 (4.36)	22.53 (3.19)	27.13 (6.46)	23.38 (4.16)	22.19 (3.93)	28.80 (3.55)	29.51 (3.59)	18.96 (3.45)	25.82 (3.69)
R value (%)	21.9	20.1	20.6	19.5	22.7	20.9	22.3	22.8	18.6	22.5
R_{free} value (%)	28.4	25.8	29.2	26.8	29.4	27.8	29.1	29.0	22.9	26.8
Nº solvent molucul.	96	63	61	102	101	82	97	102	90	80
Mean B-factor (Å ²)	43.01	20.84	52.48	26.62	43.27	2.10	35.99	43.10	23.72	42.84
Bond R.m.s.d. (Å)	0.014	0.029	0.019	0.014	0.013	0.020	0.013	0.013	0.008	0.013
Angles R.m.s.d. (°)	1.427	1.597	1.619	1.403	1.229	1.731	1.573	1.303	1.228	1.518

han sido resueltas en el grupo espacial $P6_1$ ($\alpha=\beta=90^\circ$ y $\gamma=120^\circ$).

Tabla 6.3. Tabla estadística para las desviaciones R.M.S. de la cadena principal de tioredoxina *WT* (2TRX.pdb) comparada con las distintas variantes, mutaciones puntuales de ácidos carboxílicos en tioredoxina: D2E (2H74.pdb), D9E (2H70.pdb), D10E (2H76.pdb), D13E (2H75.pdb), D43E (2H73.pdb), E44D (2H6Z.pdb), D47E (2H71.pdb), E48D (2H6Y.pdb) y E85D (2H72.pdb).

		2TRX		2H74		2H70		2H76		2H75		2H73		2H6Z		2H71		2H6Y		2H72	
		A	B	A	B	A	B	A	B	A	B	A	B	A	B	A	B	A	B	A	B
2TRX	A			0.80		0.65		0.78		0.68		0.63		0.77		0.65		0.68		0.64	
	B	0.76																			
2H74	A					0.39		0.39		0.30		0.43		0.44		0.34		0.35		0.32	
	B		1.19																		
2H70	A							0.35		0.37		0.34		0.36		0.38		0.33		0.32	
	B		1.02		0.50																
2H76	A									0.13		0.36		0.19		0.17		0.22		0.30	
	B		0.99		0.46		0.48														
2H75	A											0.28		0.15		0.18		0.25		0.30	
	B		1.06		0.53		0.46		0.30												
2H73	A													0.35		0.28		0.31		0.30	
	B		0.97		0.60		0.58		0.49		0.53										
2H6Z	A															0.19		0.24		0.30	
	B		1.06		0.54		0.51		0.25		0.32		0.54								
2H71	A																	0.26		0.32	
	B		1.04		0.37		0.46		0.33		0.27		0.43		0.37						
2H6Y	A																			0.35	
	B		1.00		0.60		0.53		0.35		0.46		0.62		0.41		0.52				
2H72	A																				
	B		0.96		0.52		0.34		0.23		0.25		0.50		0.33		0.28		0.36		

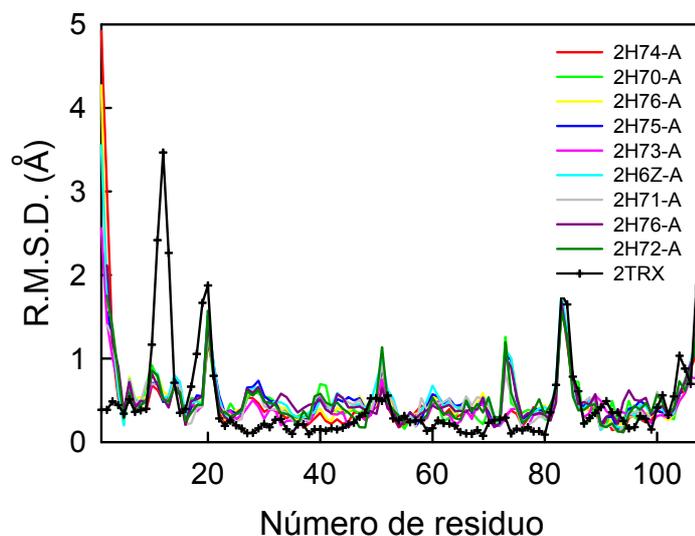


Figura 6.3: Representación de los R.M.S.D. de la cadena principal de tiorredoxina *WT* 2TRX-A respecto de 2TRX-B (línea más cruces negras); R.M.S.D. de la cadena principal de las proteínas variantes de tiorredoxina: 2H74-A (D2E-A, línea roja), 2H70-A (D9E-A, línea verde), 2H76-A (D10E-A, línea amarilla), 2H75-A (D13E-A, línea azul), 2H73-A (D43E-A, línea rosa), 2H6Z-A (E44D-A, línea azul turquesa), 2H71-A (D47E-A, línea gris), 2H76-A (E48D-A, línea violeta) y 2H72-A (E85D-A, línea verde oscuro), todas ellas comparadas con el monómero A de tiorredoxina 2TRX. En la leyenda de la figura se dan los códigos PDB de las estructuras comparadas.

estructura de cada una de las variantes de tiorredoxina, mutaciones de ácidos carboxílicos: D2E (2H74), D9E (2H70), D10E (2H76), D13E (2H75), D43E (2H73), E44D (2H6Z), D47E (2H71), E48D (2H6Y) y E85D (2H72).

En la Tablas 6.4, se dan las R.M.S.D. resultantes de la comparación del monómero A de tiorredoxina *WT* (2H6X-A) respecto de la cadena principal de su monómero B (2H6X-B) (Figura 6.4); además se dan las R.M.S.D. de los monómeros A y B de las nueve variantes de tiorredoxina comparados con 2H6X-A y 2H6X-B, respectivamente (Figura 6.6); así como las R.M.S.D. resultantes de la comparación de todos los monómeros A y B de las nueve variantes de tiorredoxina estudiadas, entre si. En la Tabla 6.5 se recogen las R.M.S.D. totales (teniendo en cuenta las cadenas laterales de los aminoácidos), resultantes de la comparación de 2H6X y las nueve variantes. Una visión global de los R.M.S.D. nos está indicando que realmente existen diferencias entre tiorredoxina *WT* y variantes (Figuras 6.5-7).

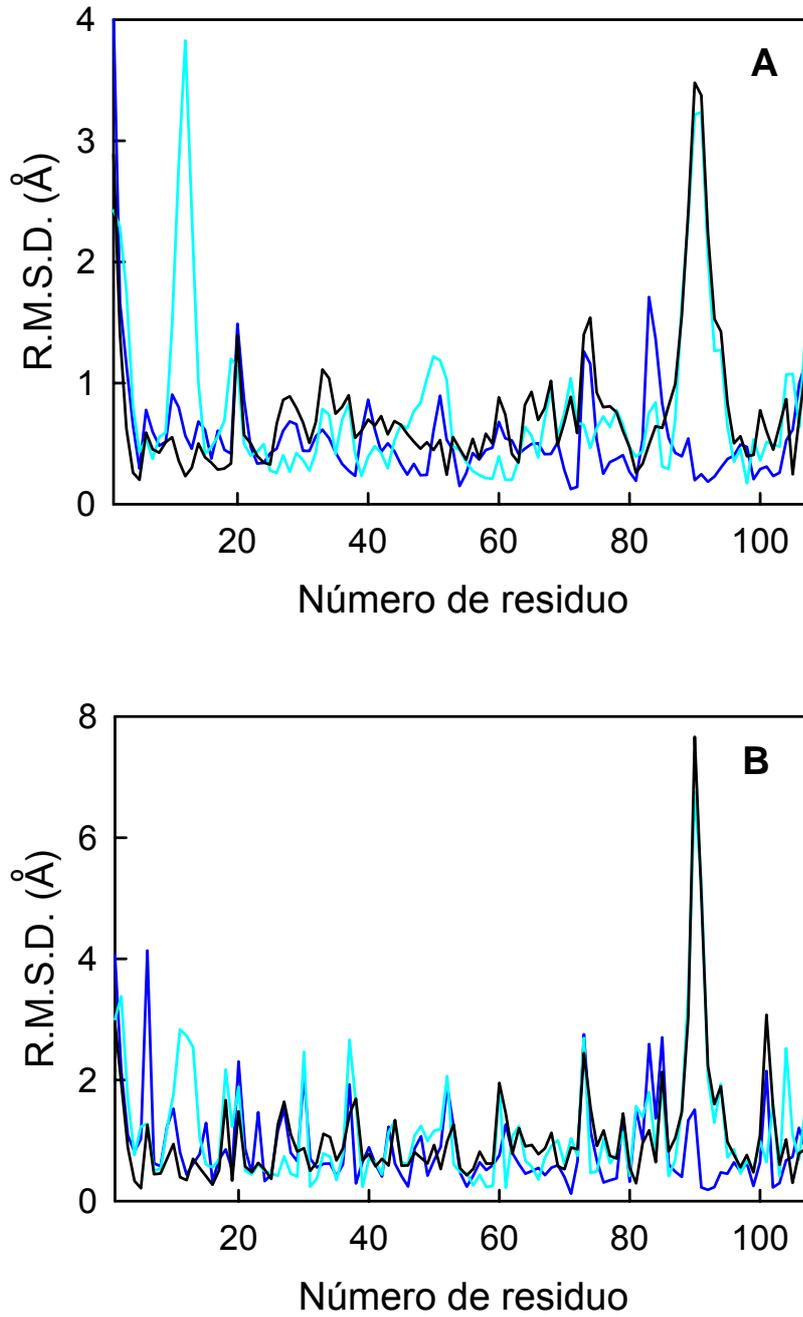


Figura 6.4: A) En negro, se representa la R.M.S.D. de la cadena principal de los monómeros A y B de tioredoxina *WT* (2H6X); en azul oscuro, R.M.S.D. de la cadena principal existente entre 2TRX-A y 2H6X-A; en azul claro, R.M.S.D. de la cadena principal existente entre 2TRX-B y 2H6X-B. B) Igual que el panel A, pero se considerado la cadena principal más las cadenas laterales de los residuos en el cálculo de R.M.S.D.

Tabla 6.4. Tabla estadística para las desviaciones R.M.S. de la cadena principal de tioredoxina *WT* (2H6X.pdb) comparada con las distintas variantes, mutaciones puntuales de ácidos carboxílicos en tioredoxina: D2E (2H74.pdb), D9E (2H70.pdb), D10E (2H76.pdb), D13E (2H75.pdb), D43E (2H73.pdb), E44D (2H6Z.pdb), D47E (2H71.pdb), E48D (2H6Y.pdb) y E85D (2H72.pdb).

		2H6X		2H74		2H70		2H76		2H75		2H73		2H6Z		2H71		2H6Y		2H72	
		A	B	A	B	A	B	A	B	A	B	A	B	A	B	A	B	A	B	A	B
2H6X	A			0.43		0.37		0.22		0.21		0.39		0.25		0.24		0.26		0.34	
	B	1.01																			
2H74	A					0.39		0.39		0.30		0.43		0.44		0.34		0.35		0.32	
	B		0.58																		
2H70	A							0.35		0.37		0.34		0.36		0.38		0.33		0.32	
	B		0.51		0.50																
2H76	A									0.13		0.36		0.19		0.17		0.22		0.30	
	B		0.31		0.46		0.48														
2H75	A											0.28		0.15		0.18		0.25		0.30	
	B		0.38		0.53		0.46		0.30												
2H73	A													0.35		0.28		0.31		0.30	
	B		0.58		0.60		0.58		0.49		0.53										
2H6Z	A															0.19		0.24		0.30	
	B		0.40		0.54		0.51		0.25		0.32		0.54								
2H71	A																	0.26		0.32	
	B		0.29		0.37		0.46		0.33		0.27		0.43		0.37						
2H6Y	A																			0.35	
	B		0.45		0.60		0.53		0.35		0.46		0.62		0.41		0.52				
2H72	A																				
	B		0.39		0.52		0.34		0.23		0.25		0.50		0.33		0.28		0.36		

Tabla 6.5. Tabla estadística para las desviaciones R.M.S. de tioredoxina *WT* (2H6X.pdb) comparada con las distintas variantes, mutaciones puntuales de ácidos carboxílicos en tioredoxina, teniendo en cuenta las cadenas laterales de los aminoácidos: D2E (2H74.pdb), D9E (2H70.pdb), D10E (2H76.pdb), D13E (2H75.pdb), D43E (2H73.pdb), E44D (2H6Z.pdb), D47E (2H71.pdb), E48D (2H6Y.pdb) y E85D (2H72.pdb).

		2H6X		2H74		2H70		2H76		2H75		2H73		2H6Z		2H71		2H6Y		2H72	
		A	B	A	B	A	B	A	B	A	B	A	B	A	B	A	B	A	B	A	B
2H6X	A			0.89		0.83		0.52		0.65		0.64		0.77		0.64		0.78		0.78	
	B	1.48																			
2H74	A					0.84		0.84		0.83		0.93		0.85		0.78		0.81		0.76	
	B		0.93																		
2H70	A							0.74		0.78		0.80		0.84		0.76		0.56		0.55	
	B		0.98		1.08																
2H76	A									0.43		0.44		0.68		0.52		0.68		0.65	
	B		0.63		0.88		0.89														
2H75	A											0.55		0.67		0.58		0.70		0.60	
	B		0.78		0.96		0.87		0.65												
2H73	A													0.79		0.60		0.76		0.82	
	B		0.82		0.99		0.95		0.53		0.80										
2H6Z	A															0.66		0.76		0.67	
	B		0.76		0.82		0.99		0.65		0.69		0.80								
2H71	A																	0.67		0.61	
	B		0.88		0.85		0.99		0.75		0.76		0.81		0.70						
2H6Y	A																			0.75	
	B		0.85		1.05		0.82		0.70		0.82		0.82		0.86		0.92				
2H72	A																				
	B		0.72		0.93		0.81		0.47		0.64		0.66		0.68		0.76		0.72		

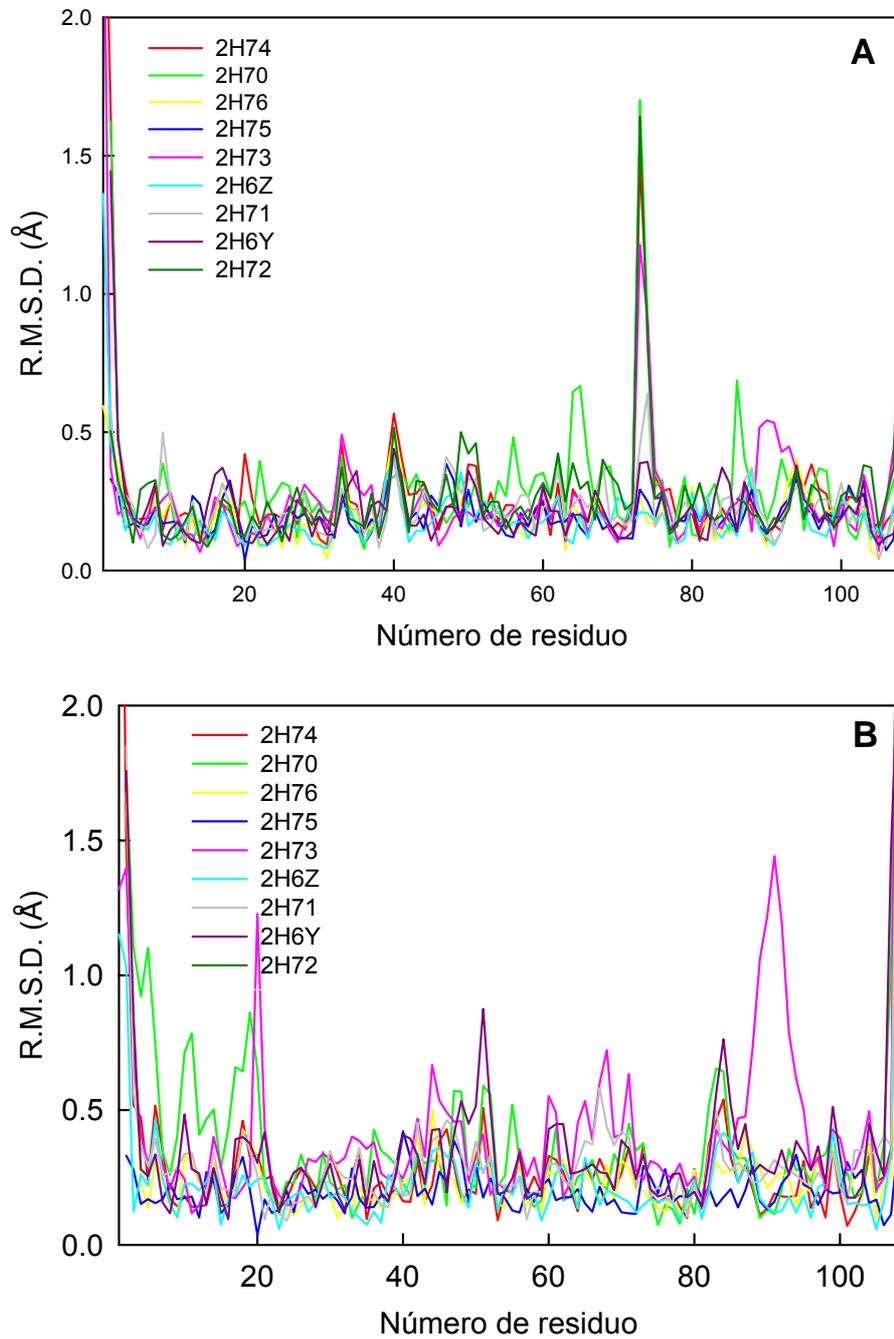


Figura 6.5: R.M.S.D. de la cadena principal, de los monómeros A (panel A) y B (panel B), de las variantes de tiorredoxina, mutaciones de ácidos carboxílicos: D2E (2H74), D9E (2H70), D10E (2H76), D13E (2H75), D43E (2H73), E44D (2H6Z), D47E (2H71), E48D (2H6Y) y E85D (2H72), comparadas con los monómeros A y B, respectivamente, de tiorredoxina WT (2H6X). Valores recogidos en Tabla 6.4.

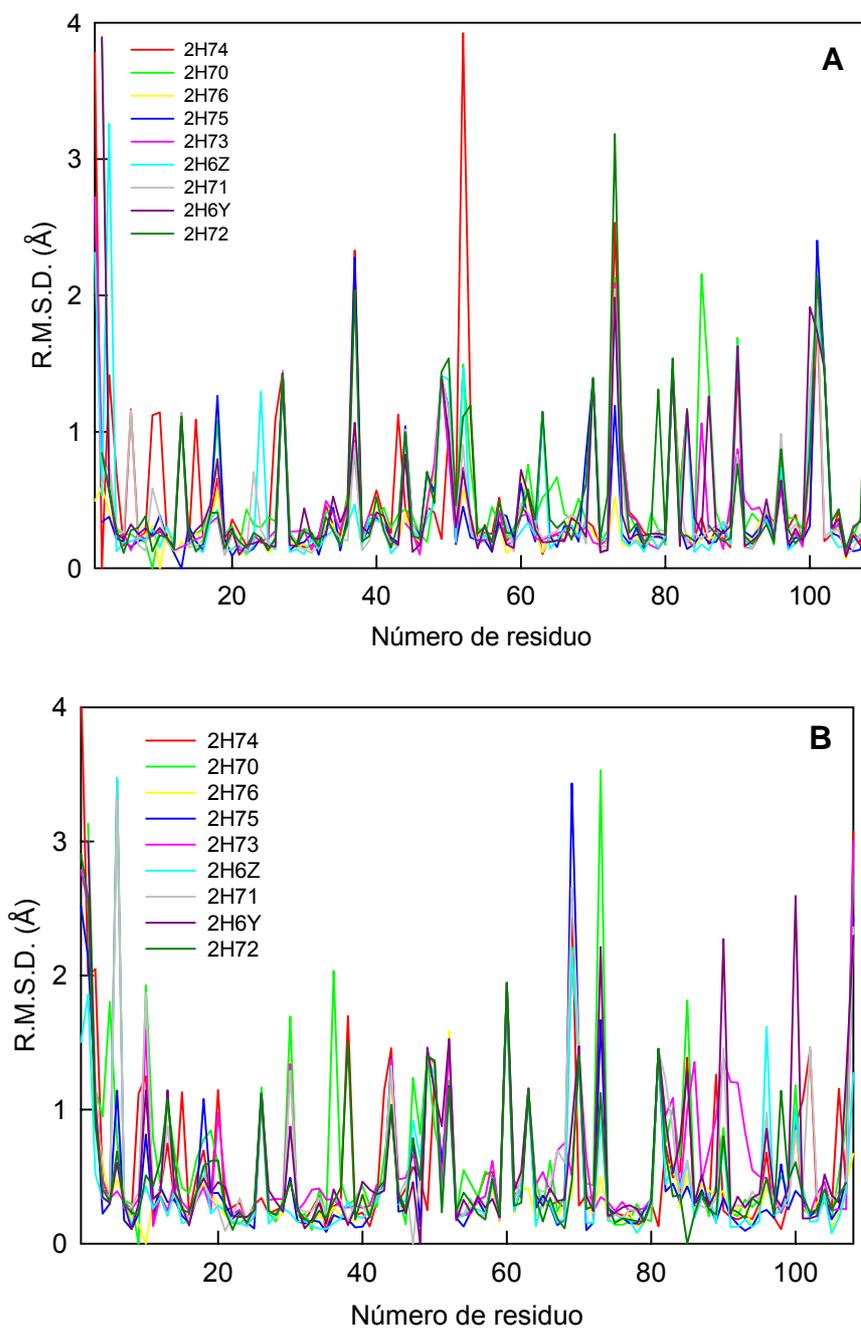


Figura 6.6: R.M.S.D. total (cadena principal más cadenas laterales), de los monómeros A (panel A) y B (panel B), de las variantes de tioredoxina, mutaciones de ácidos carboxílicos: D2E (2H74), D9E (2H70), D10E (2H76), D13E (2H75), D43E (2H73), E44D (2H6Z), D47E (2H71), E48D (2H6Y) y E85D (2H72), comparadas con los monómeros A y B, respectivamente, de tioredoxina *WT* (2H6X). Valores recogidos en Tabla 6.5.

Además, las desviaciones observadas son mayores cuando se tienen en cuenta las cadenas laterales, probablemente por la mayor libertad de movimiento que estas tienen. Por esta razón, también se han estudiado los cambios en área superficial accesible (Δ ASA) de las estructuras de las distintas variantes de tiorredoxina respecto de tiorredoxina *WT* (Figura 6.8).

Este primer análisis nos ha mostrado dos características muy peculiares de tiorredoxina. Por una parte, hemos podido observar el polimorfismo de tiorredoxina. 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. Sin embargo, en algunos casos es posible obtener distintos polimorfos en condiciones químico-físicas similares. El elevado polimorfismo de los cristales de proteínas se explica por el gran número y debilidad de los contactos entre las moléculas de proteína que forman el cristal. Una mutación puntual puede ser suficiente para impedir la formación del cristal, deteriorar su calidad ó provocar la formación de un nuevo polimorfo. Nosotros hemos cristalizado, y resuelto la estructura, de tiorredoxina *WT* y 9 variantes, mutaciones de ácidos carboxílicos (D2E, D9E, D10E, D13E, D43E, E44D, D47E, E48D y E85D). Todos ellas han cristalizado en un grupo espacial ($P6_1$) diferente del de otras estructuras de tiorredoxina previamente cristalizadas (por ejemplo, tiorredoxina 2TRX está en C2). Ésto ha sido posible modificando las condiciones de cristalización (velocidad y rango de sobresaturación barrido) empleando medios con transporte de masa difusivo. Y nos ha permitido observar cómo la obtención de polimorfos puede ir acompañada de cambios estructurales en la proteína, que se pueden entender como reflejo de la flexibilidad que presentan en solución.

Por otro lado, la movilidad de la cadena principal que revelan las estructuras de tiorredoxina ha podido ser adoptada por la proteína para acomodar el efecto de la mutación introducida. Ésto indica que el esqueleto de tiorredoxina muestra un grado significativo de flexibilidad. Comportamientos de este tipo han sido previamente observados y se han utilizado para diseñar proteínas realizando una parametrización de la estructura supersecundaria de la proteína (Su y Mayo, 1997). Además, si se tiene en cuenta que todas las posiciones mutadas son residuos muy conservados en la evolución, estos resultados pueden indicar que la

flexibilidad de la cadena principal observada puede estar relacionada con la selección natural de dichos residuos.

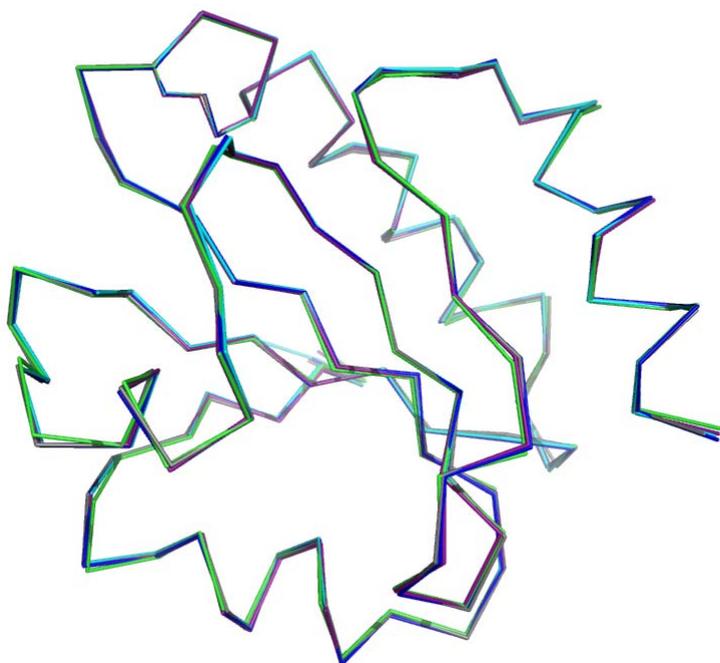


Figura 6.7: Ilustración gráfica de la desviación de la cadena principal (monómero A) de tiorredoxina *WT* y 5 variantes, mutaciones de ácidos carboxílicos: D9E (verde), D13E (azul), E44D (azul turquesa), E48D (violeta) y E85D (gris).

El alineamiento y cálculo de las R.M.S.D. se ha realizado con el programa NOC (Chenmengen y H.X. Cang, IBP-CAS; <http://noc.ibp.ac.cn/>). Se ha utilizado la opción de alinear sólo los residuos iguales en las secuencias. Los cálculos de ASA se han realizado también con el programa NOC. En el cálculo el programa usa una sonda de disolvente de 1.4 Å radio.

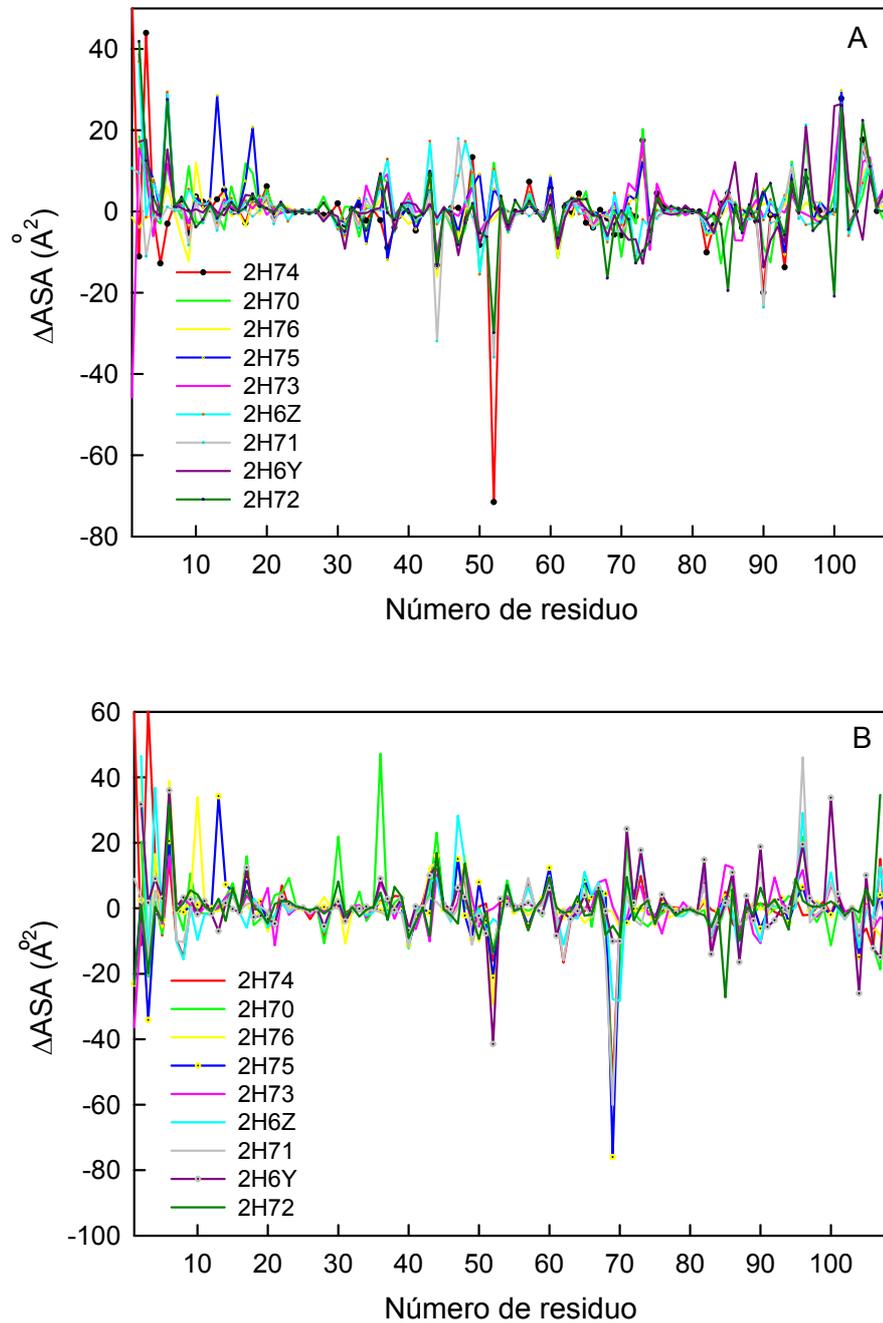


Figura 6.8: Cambios en área superficial accesible (ΔASA) de los monómeros A (panel A) y B (panel B), de las variantes de tiorredoxina, mutaciones de ácidos carboxílicos: D2E (2H74), D9E (2H70), D10E (2H76), D13E (2H75), D43E (2H73), E44D (2H6Z), D47E (2H71), E48D (2H6Y) y E85D (2H72), respecto del ASA de los monómeros A y B, respectivamente, de tiorredoxina WT (2H6X) [$\Delta\text{ASA}=\text{ASA}_{\text{mutante}}-\text{ASA}_{\text{WT}}$].

6.4 ESTUDIO LOCAL DEL EFECTO DE LA MUTACIÓN SOBRE LA ESTRUCTURA DE TIORREDOXINA

Los residuos glutamato (E) y aspartato (D) son física-químicamente similares y pueden sustituirse fácilmente uno por el otro en el curso de la evolución de proteínas, como muestran sus correspondientes coeficientes en las matrices de PAM (George *et al.*, 1990) y BLOSUM (Henikoff y Henikoff, 1992). La observación de dichas matrices nos indica que los residuos E y D son muy conservativos (véase sección 2.2.7.2 en *Materiales y métodos*). Por otro lado, los residuos aspartato son más sensibles que los residuos glutamato a las interacciones con el ambiente que los rodea (Forsyth *et al.*, 2002). Por lo tanto, el hecho de que las mutaciones E→D y D→E tengan un efecto pequeño, pero significativo sobre la estabilidad de la proteína (Figura 3.1), junto al hecho de que E y D son residuos que se encuentran frecuentemente sobre la superficie (lo que significa que no se encuentran restringidos estructural o funcionalmente), nos ha permitido estudiar la relación entre estabilidad y evolución (Capítulo 3 publicado en Godoy-Ruiz *et al.*, 2004). Mediante un estudio local de las interacciones existentes en la zona de la mutación, se pretende determinar qué cambios conformacionales en el residuo mutado y en el ambiente que lo rodea, están contribuyendo a los cambios de estabilidad de la proteína. Y al mismo tiempo podremos conocer más sobre la relación estructura y flexibilidad. Para ello, se ha analizado cada zona de la estructura de tiorredoxina *WT* en la que se ha introducido una mutación y se ha comparado con la misma región en la estructura de la variante correspondiente. Se han calculado las interacciones polares (interacciones electrostáticas y enlaces de hidrógeno) existentes en una esfera de 3.5 Å centrada en el residuo mutado, tanto en la proteína *WT* como en la variante correspondiente. En este análisis, no se han tenido en cuenta los enlaces de hidrógeno formados con las moléculas de agua asignadas a cada estructura, puesto que se considera que el número moléculas de agua que se han asignado es dependiente de la resolución a la que se han resuelto las estructuras.

El análisis local de las interacciones que mantiene el residuo mutado, se ha completado con una observación de las R.M.S.D. de las cadenas principales, de las Δ ASAS de las variantes de tiorredoxina, en comparación a tiorredoxina *WT* (2H6X) y de los factores-B.

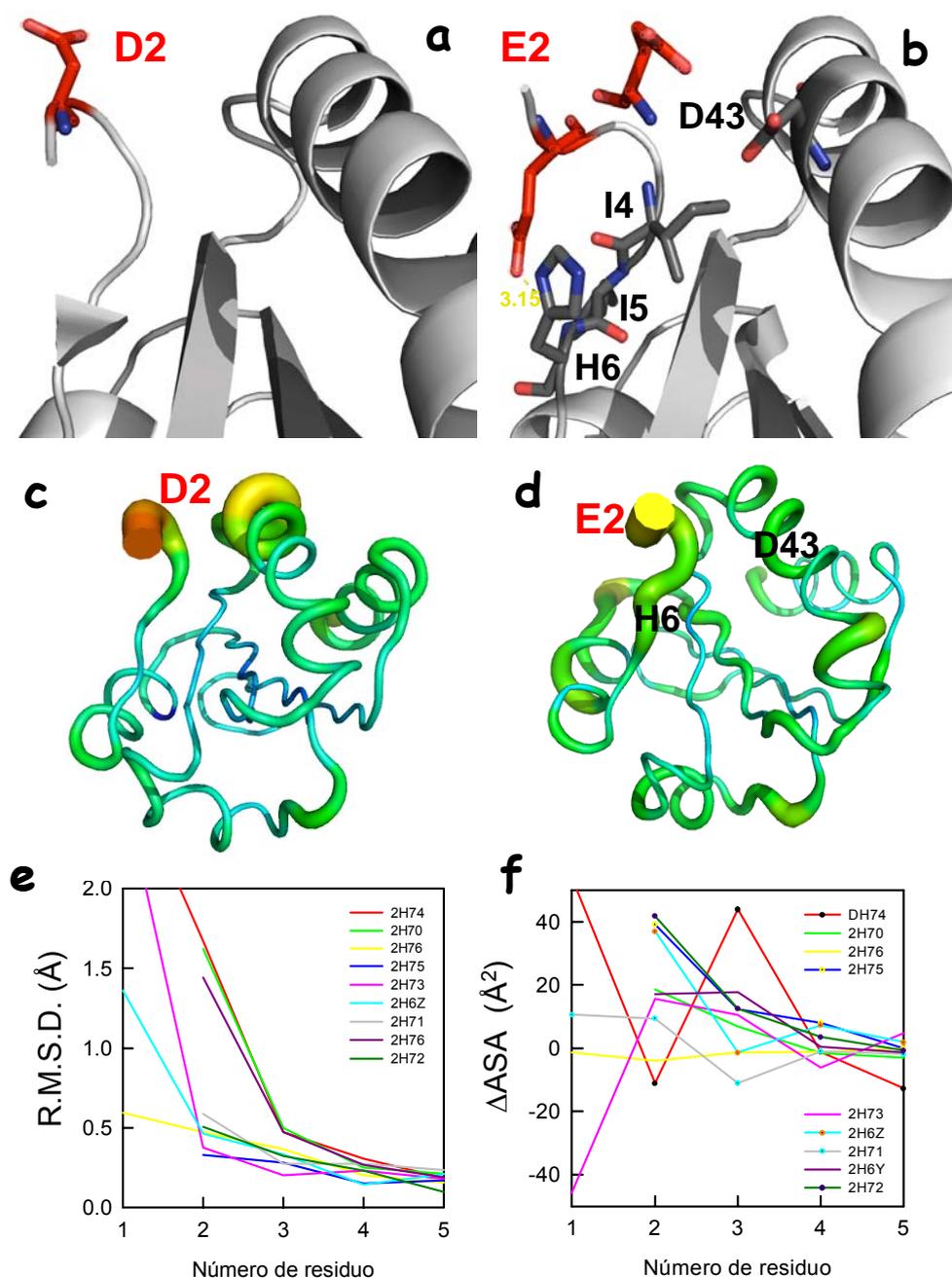


Figura 6.9: a) Estructura tridimensional del monómero A de tiorredoxina WT (2H6X.pdb; resolución 2.60 Å) centrada en el residuo D2 (rojo). b) Estructura tridimensional de la variante de tiorredoxina D2E (2H74.pdb; resolución 2.40 Å), centrada en el residuo E2 (rojo). Se muestra la doble conformación encontrada para la cadena lateral de E2; en gris oscuro, los residuos con los que el residuo 2 mantiene interacciones polares en una esfera de 3.5 Å. c) y d) representan los factores térmicos B de 2H6X y 2H74, respectivamente. e) RMSD de la cadena principal del monómero A de las variantes de tiorredoxina, en comparación al monómero A de tiorredoxina WT (2H6X), en la zona de la mutación. f) Área superficial accesible al disolvente de las variantes de tiorredoxina respecto de 2H6X, en la zona de la mutación [$\Delta ASA = ASA_{mutante} - ASA_{WT}$].

Una representación de la estructura en función de los factores-B refleja la movilidad de la cadena principal de las proteínas y se ha usado para medir la flexibilidad de la proteínas plegadas (Karplus y Schultz, 1985, 1995). Zonas de alta flexibilidad se representarán en color rojo (zonas calientes), mientras que zonas de baja flexibilidad se representan en azul intenso (zonas frías). En la estructura cristalina de macromoléculas, los factores-B reflejan la incertidumbre de las posiciones atómicas en el modelo y a menudo representan los efectos combinados de las vibraciones térmicas y desorden estático (Rhodes, 1993).

A continuación se comienza el análisis de las nueve posiciones mutadas:

Asp 2 (D2) en 2H6X tiene una accesibilidad al disolvente del 81.4% en comparación al 64.2% que presenta Glu 2 (E2) en la variante 2H74 (Figura 6.9). Ésto, puede estar probablemente provocado por el hecho de que la cadena lateral de E2 presenta doble conformación y el confórmero más probable se encuentra estabilizado mediante un enlace de hidrógeno con His 6 (3.15 Å). Esta ciclación del extremo N-terminal hace que los tres primeros residuos de la variante (Ser 1, Glu 2 y Lys 3) tengan doble conformación (Figura 6.9 a y b). La representación en función de los factores-B, muestra como los C α de E2 pierden libertad de movimiento y por el contrario las regiones de His 6 y Asp 43, ganan flexibilidad. En la Figura 6.9-f, se puede observar el enterramiento del residuo E2 de 2H74 en comparación a D2 de 2H6X; es decir, E2 posee un $\Delta ASA < 0$ ($\Delta ASA = ASA_{mutante} - ASA_{WT}$) en la estructura de 2H74.

Asp 9 (D9) en 2H6X tiene una accesibilidad al disolvente del 34.4% en comparación al 33.1% que presenta Glu 9 (E9) en la variante 2H70 (Figura 6.10). D9 mantiene en 2H6X numerosas interacciones (electrostáticas y enlaces de hidrógeno) con la secuencia de residuos: Asn 63 (3.20 Å), Pro 64 (3.47 Å), Gly 65 (2.83 Å) y Thr 66 (3.04 y 2.53 Å), que se encuentran en el lazo previo a $\alpha 3$. En la estructura de 2H70, la mayoría de estas interacciones se mantienen a distancias más cortas: P64 (3.13 Å), G65 (3.38 Å) y T66 (2.75 Å), lo cual tiene sentido pues la cadena lateral de E9 tiene un carbono más que la de D9. Ésto causa un aumento de la flexibilidad del lazo en el que se encuentran esos cuatro residuos y de $\alpha 3$ (desde Gly 65 hasta Arg 73); al mismo tiempo podemos observar que aumentan los factores-B de otras zonas de la proteína que se encuentran a distancias

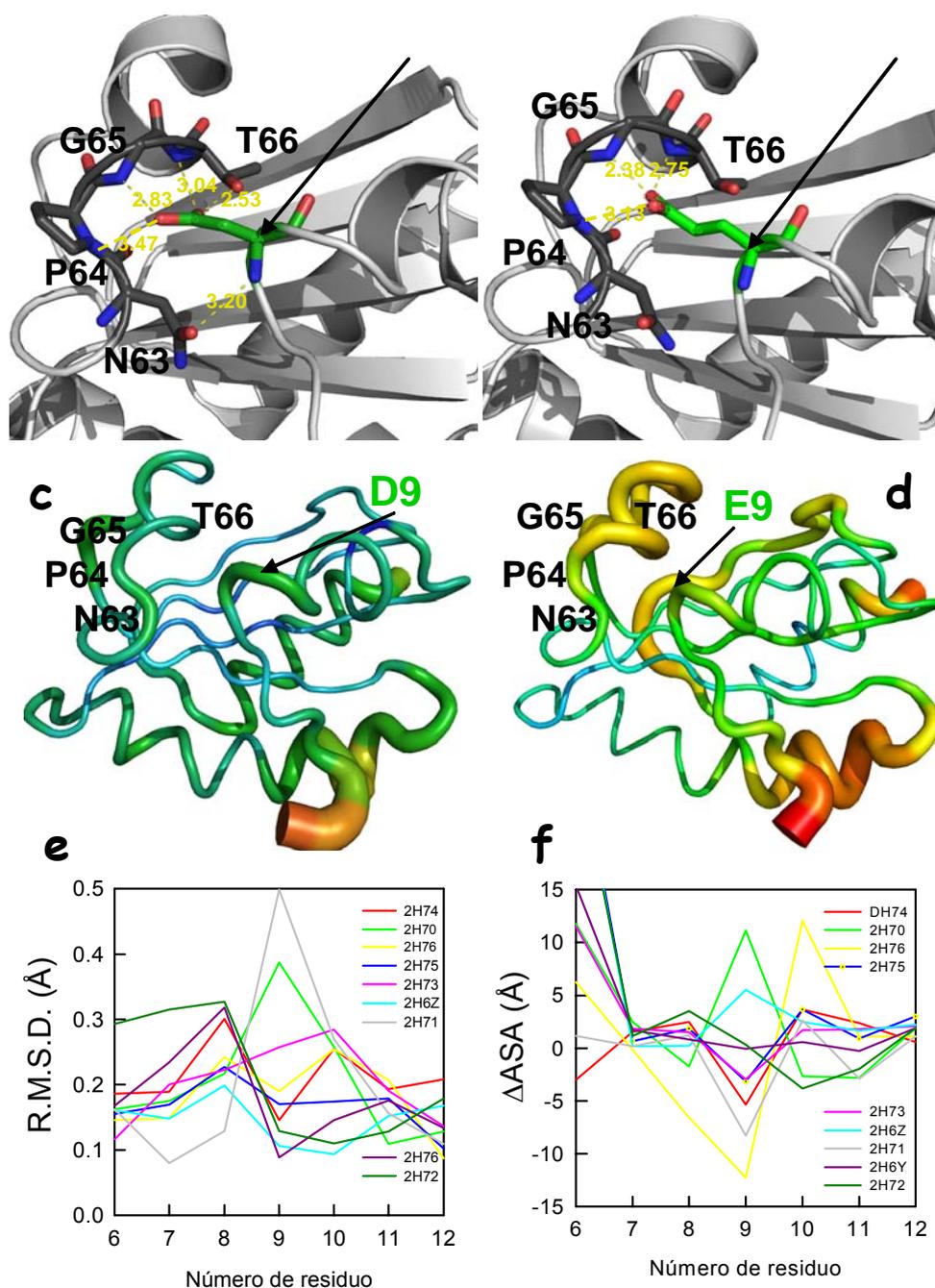


Figura 6.10: a) Estructura tridimensional del monómero A de tiorredoxina *WT* (2H6X.pdb; resolución 2.60 Å) centrada en el residuo D9 (verde). b) Estructura tridimensional de la variante de tiorredoxina D9E (2H70.pdb; resolución 2.70 Å), centrada en el residuo E9 (verde). En gris oscuro, se muestran los residuos con los que el residuo 9 mantiene interacciones polares en una esfera de 3.5 Å. c) y d) representan los factores térmicos B de 2H6X y 2H70, respectivamente. e) RMSD de la cadena principal del monómero A de las variantes de tiorredoxina, en comparación al monómero A de tiorredoxina *WT* (2H6X), en la zona de la mutación. f) Área superficial accesible al disolvente de las variantes de tiorredoxina respecto de 2H6X, en la zona de la mutación [Δ ASA=ASA_{mutante}-ASA_{WT}].

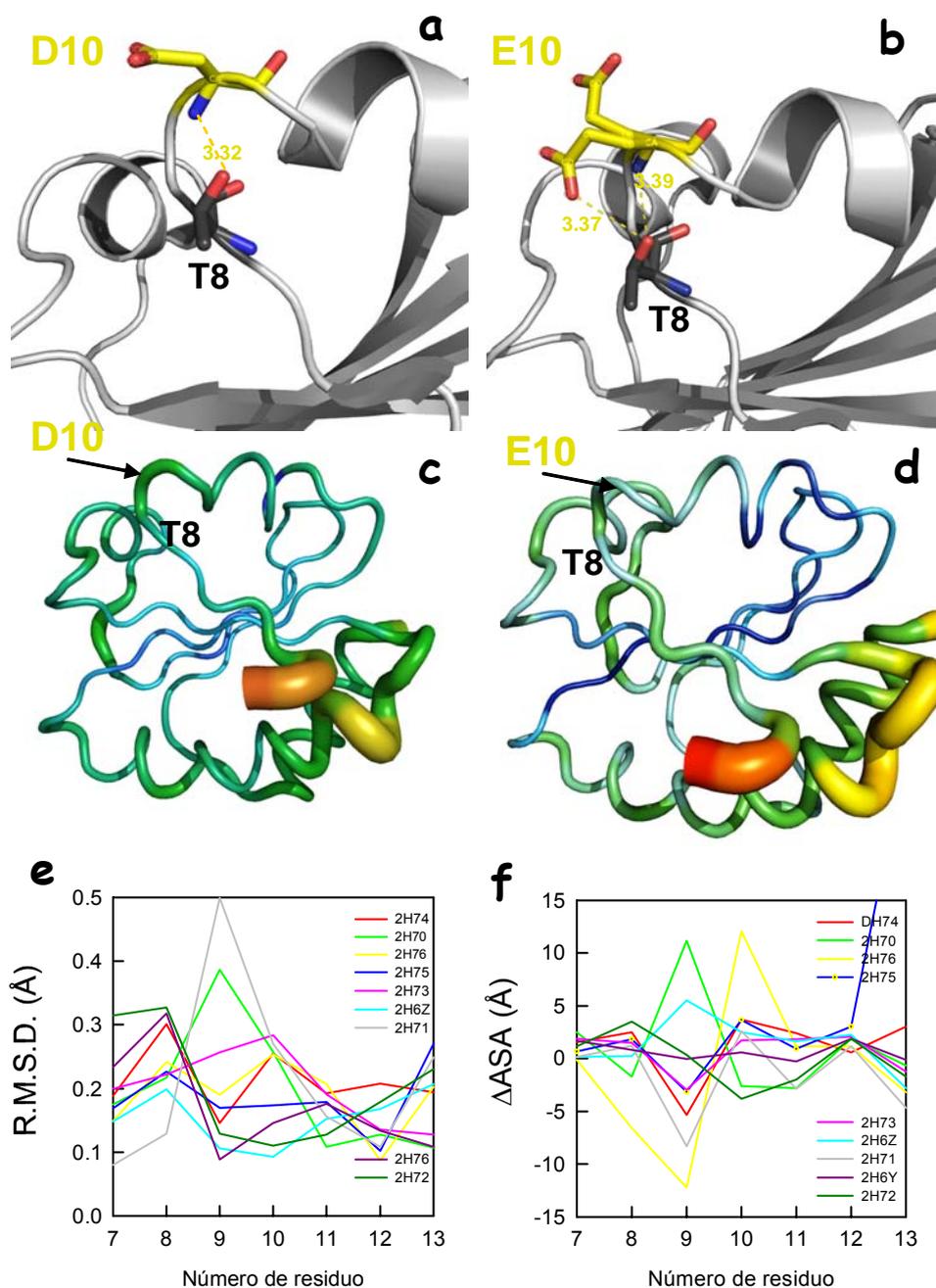


Figura 6.11: a) Estructura tridimensional del monómero A de tiorredoxina WT (2H6X.pdb; resolución 2.60 Å) centrada en el residuo D10 (amarillo). b) Estructura tridimensional de la variante de tiorredoxina D10E (2H76.pdb; resolución 2.25 Å), centrada en el residuo E10 (amarillo). Se muestra la doble conformación encontrada para la cadena lateral de E10; en gris oscuro, los residuos con los que el residuo 10 mantiene interacciones polares en una esfera de 3.5 Å. c) y d) representan los factores térmicos B de 2H6X y 2H76, respectivamente. e) RMSD de la cadena principal del monómero A de las variantes de tiorredoxina, en comparación al monómero A de tiorredoxina WT (2H6X), en la zona de la mutación. f) Área superficial accesible al disolvente de las variantes de tiorredoxina respecto de 2H6X, en la zona de la mutación

mayores de 3.5 Å (aminoácidos desde Lys 90 hasta Leu 94 en $\beta 5$ y aminoácidos desde Asp 43 hasta Gly 51 en $\alpha 2$). Al mismo tiempo, se puede observar una exposición del residuo E9 de 2H70 en comparación a D9 de 2H6X. Este mutante nos muestra como la introducción de un residuo polar (E9 en 2H70), no satisface todos los enlaces de hidrógeno y produce, como consecuencia, una desestabilización del estado nativo de la proteína. Es decir, un residuo polar puede incrementar la estabilidad termodinámica (D9 en 2H6X) cuando forma enlaces de hidrógeno con átomos polares que son inaccesibles al disolvente (residuos T66 y T77). Ésta podría ser una evidencia experimental de la elevada conservación del residuo D9 en tiorredoxina, ya que la naturaleza de las interacciones polares permite a los residuos polares que forman parte del interior de las proteínas, determinar el plegamiento de una forma específica (Bowie *et al.*, 1990, Bolon y Mayo, 2001).

Asp 10 (D10) en 2H6X tiene una accesibilidad al disolvente del 83.2% en comparación al 89.2% que presenta Glu 10 (E10) en la variante 2H76 (Figura 6.11). Este aumento del área superficial accesible de E10 puede ser debido a la doble conformación que presenta en 2H76. D10 mantiene una interacción con Thr 8 (3.32 Å) en 2H6X; dicha interacción la mantiene también E10 en 2H76 (3.37 y 3.39 Å). Los factores-B se ven poco alterados en esta proteína. Merece la pena destacar el aumento de la flexibilidad de aminoácidos que se encuentran a distancias mayores de 3.5 Å respecto de E10 (desde Asp 46 hasta Lys 52 en $\alpha 2$).

Asp 13 (D13) en 2H6X tiene una accesibilidad al disolvente del 51.7% en comparación al 42.4% que presenta Glu 13 (E13) en la variante 2H75 (Figura 6.12). En 2H6X, D13 mantiene una interacción electrostática (3.40 Å) y forma enlaces de hidrógeno con Lys 18 y Leu 7 (2.92 Å y 3.08 Å, respectivamente). En 2H75, la interacción electrostática se mantiene a distancias mayores de 3.5 Å. En la Figura 6.12-b se puede observar la doble conformación de la cadena lateral asignada a Lys 18. Merece la pena destacar el incremento del ASA (30 Å² aproximadamente) de 2H75 en la zona de la mutación. Además, en la Figura 6.12-f se puede apreciar una disminución, en general de los valores de los factores-B para todas las posiciones de la variante 2H75, incluido el comienzo de $\alpha 2$ (aminoácidos desde Asp 47 a Lys 52).

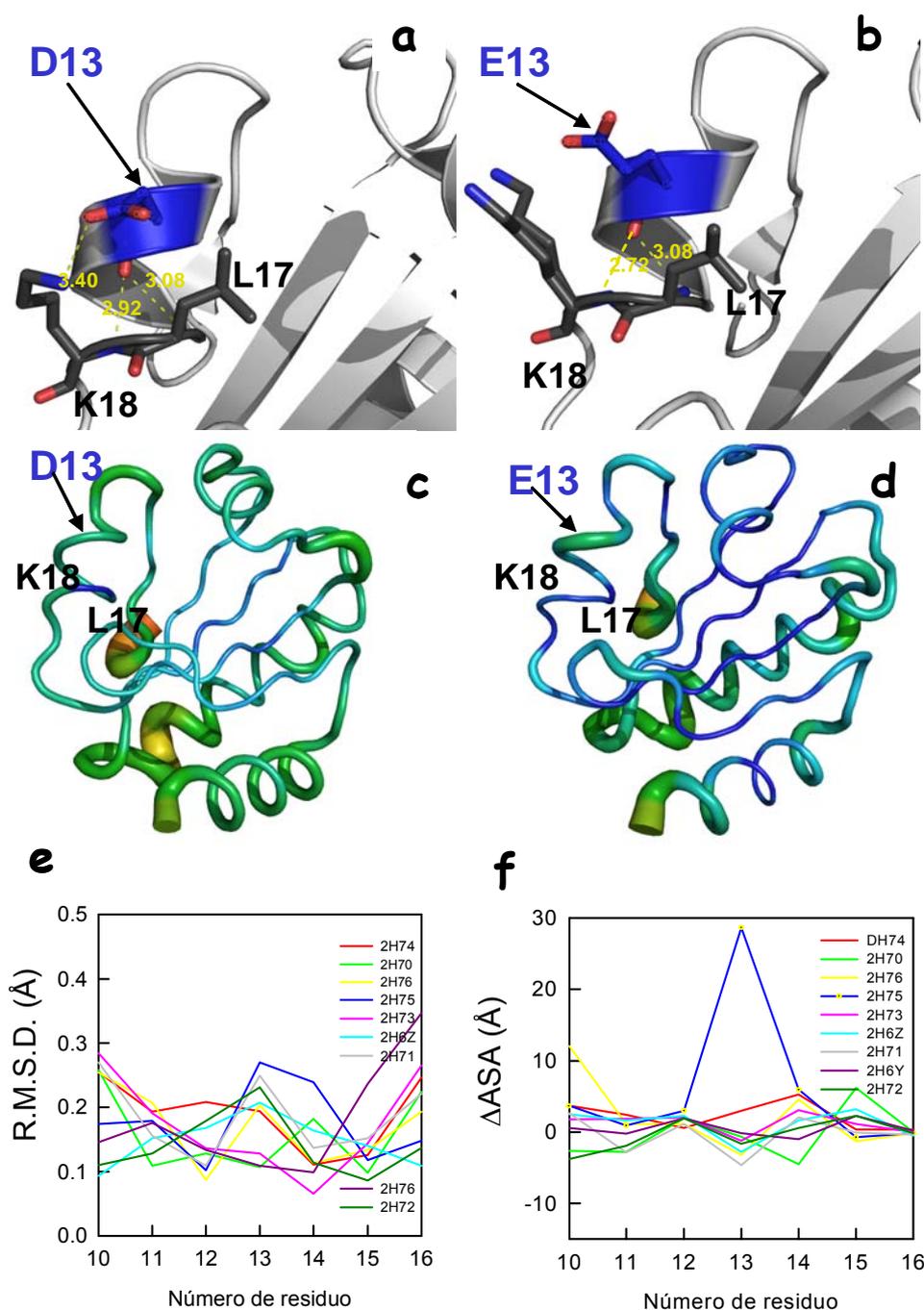


Figura 6.12: a) Estructura tridimensional del monómero A de tiorredoxina WT (2H6X.pdb; resolución 2.60 Å) centrada en el residuo D13 (azul). b) Estructura tridimensional de la variante de tiorredoxina D13E (2H75.pdb; resolución 2.20 Å), centrada en el residuo E13 (azul). En gris oscuro se muestran los residuos con los que el residuo 13 mantiene interacciones polares en una esfera de 3.5 Å. c) y d) representan los factores térmicos B de 2H6X y 2H75, respectivamente. e) RMSD de la cadena principal del monómero A de las variantes de tiorredoxina, en comparación al monómero A de tiorredoxina WT (2H6X), en la zona de la mutación. f) Área superficial accesible al disolvente de las variantes de tiorredoxina respecto de 2H6X, en la zona de la mutación [$\Delta ASA = ASA_{mutante} - ASA_{WT}$].

En 2H75 hemos observado como la introducción de la mutación ($T_m=0.83^\circ\text{C}$ más que la de tiorredoxina *WT*) aumenta ligeramente la movilidad atómica de los átomos de la cadena principal en la posición mutada. Ésto se compensa enormemente con una bajada significativa de los factores-B en el resto de la estructura.

Asp 43 (D43) en 2H6X tiene una accesibilidad al disolvente del 46.0% en comparación al 40.7% que presenta Glu 43 (E43) en la variante 2H73 (Figura 6.13). D43, se encuentra en $\alpha 1$ y mantiene, en 2H6X, enlaces de hidrógeno a distancias menores de 3.5 Å, con Ala 39 (3.34 Å) y Pro 40 (3.23 Å). E43, en 2H73, conserva esas mismas interacciones a distancias de 3.21 y 3.32 Å, respectivamente. Al igual que ocurría con la variante D13E, para la variante D43E la introducción de la mutación causa una disminución general de los factores-B en la toda la estructura de la proteína. Más concretamente, se trata de una estabilización de $\alpha 4$ y de la secuencia de aminoácidos desde Asp 47 hasta Glu 48 en $\alpha 2$ (Figura 6.13-f). Hay que recordar que la variante D43E es 0.69°C más estable que tiorredoxina *WT*.

Glu 44 (E44) en 2H6X tiene una accesibilidad al disolvente del 37.6% en comparación al 49.3% que presenta Asp 44 (D44) en la variante 2H6Z (Figura 6.14). En 2H6X, E44 se encuentra formando numerosos enlaces de hidrógeno, a distancias menores de 3.5 Å, con los aminoácidos que le rodean en $\alpha 2$. Más concretamente, está interaccionando con Pro 40 (3.35 Å), Ile 41 (3.18 Å), Asp 47 (3.20 Å) y Glu 48 (3.08 Å). Los residuos E44 y E48 mantienen una interacción electrostática con Lys 96, a 2.66 y 2.96, respectivamente. En 2H6Z, D44 mantiene enlaces de hidrógeno con Pro 40 (3.27 Å), Asp 47 (2.34 Å) y Glu 48 (2.94 Å), además los residuos D44 y E48 conservan las interacciones electrostáticas existentes en 2H6X a distancias 2.95 y 2.94 Å, respectivamente. Resulta interesante prestar atención, cómo a pesar de que la mutación E44D causa un descenso en la $T_m=3.46^\circ\text{C}$ respecto de la T_m de tiorredoxina *WT*, probablemente debido a que no se satisfacen tantos enlaces de hidrógeno como en tiorredoxina *WT*, ésto da lugar a una bajada de los factores-B en la estructura en general.

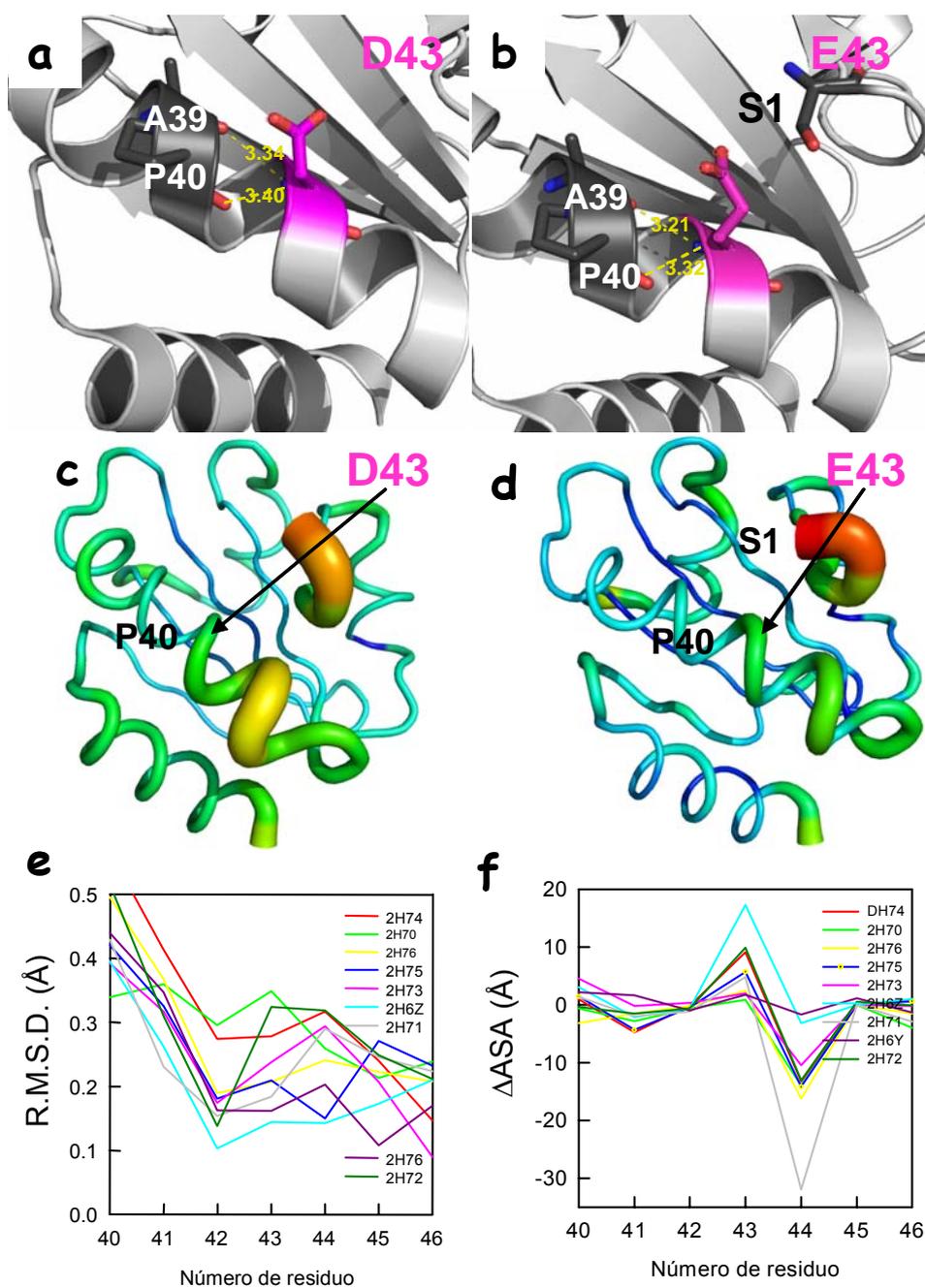


Figura 6.13: a) Estructura tridimensional del monómero A de tiorredoxina WT (2H6X.pdb; resolución 2.60 Å) centrada en el residuo D43 (rosa). b) Estructura tridimensional de la variante de tiorredoxina D43E (2H73.pdb; resolución 2.45 Å), centrada en el residuo E43 (rosa). En gris oscuro se muestran los residuos con los que el residuo 13 mantiene interacciones polares en una esfera de 3.5 Å. c) y d) representan los factores térmicos B de 2H6X y 2H73, respectivamente. e) RMSD de la cadena principal del monómero A de las variantes de tiorredoxina, en comparación al monómero A de tiorredoxina WT (2H6X), en la zona de la mutación. f) Área superficial accesible al disolvente de las variantes de tiorredoxina respecto de 2H6X, en la zona de la mutación [Δ ASA=ASA_{mutante}-ASA_{WT}].

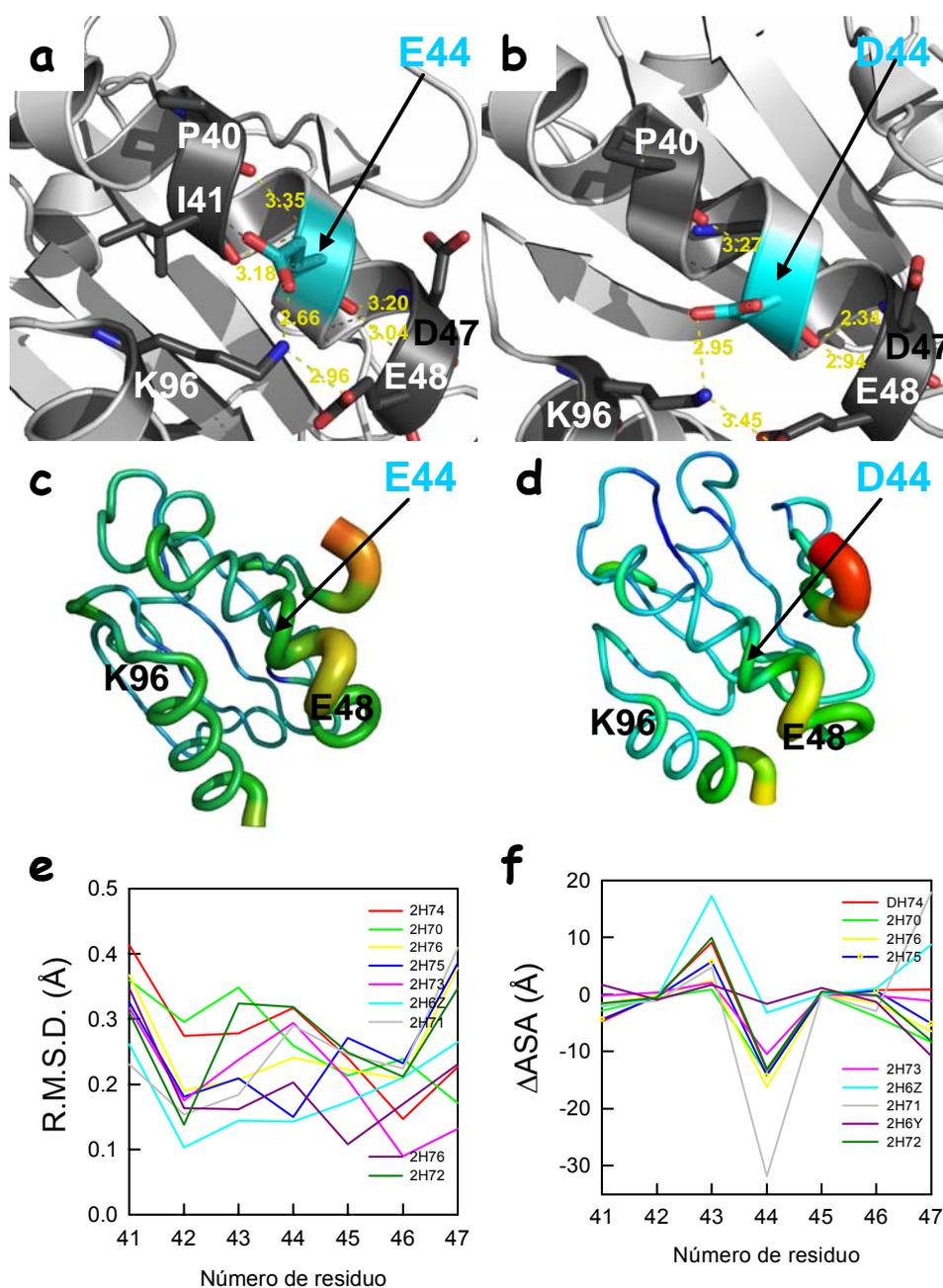


Figura 6.14: a) Estructura tridimensional del monómero A de tiorredoxina WT (2H6X.pdb; resolución 2.60 Å) centrada en el residuo E44 (azul turquesa). b) Estructura tridimensional de la variante de tiorredoxina E44D (2H6Z.pdb; resolución 2.25 Å), centrada en el residuo D44 (azul turquesa). En gris oscuro se muestran los residuos con los que el residuo 13 mantiene interacciones polares en una esfera de 3.5 Å. c) y d) representan los factores térmicos B de 2H6X y 2H6Z, respectivamente. e) RMSD de la cadena principal del monómero A de las variantes de tiorredoxina, en comparación al monómero A de tiorredoxina WT (2H6X), en la zona de la mutación. f) Área superficial accesible al disolvente de las variantes de tiorredoxina respecto de 2H6X, en la zona de la mutación [$\Delta\text{ASA} = \text{ASA}_{\text{mutante}} - \text{ASA}_{\text{WT}}$].

Asp 47 (D47) en 2H6X tiene una accesibilidad al disolvente del 50.6% en comparación al 53.3% que presenta Glu 47 (E47) en la variante 2H71 (Figura 6.14). En términos de ASA, esto supone una exposición de E47 aproximadamente unos 20 Å² más que D47 en 2H6X. D47, forma enlaces de hidrógeno con Asp 43 (3.41 Å), Glu 44 (3.20 Å), Tyr 49 (3.23 Å) y Gln 50 (>3.5 Å); y una interacción electrostática con Lys 3 (3.30 Å). E47, en 2H71, forma enlaces de hidrógeno con Asp 43 (3.35 Å), Glu 44 (3.19 Å) y Gln 50 (2.95 y 3.35 Å); y una interacción electrostáticas con Lys 3 (3.30 Å). Nuevamente, merece la pena destacar cómo el hecho de introducir un aspartato en la posición 47 ($T_m=90.37^\circ\text{C}$, 1.32°C más que que la T_m de *WT*) disminuye el factor-B de la zona de la mutación considerablemente y de los residuos con los que éste interacciona E47. Al mismo tiempo, esta mutación aumenta ligeramente los factores-B de la secuencia de aminoácidos desde Cis 32 hasta Ile 38, región que comprende el centro catalítico de tiorredoxina. Además, el comienzo de $\alpha 4$ sufre cambios muy interesantes: se produce un leve aumento de los factores-B de los residuos Val 91 hasta Leu 94 y una disminución significativa del factor-B de Lys 96. Es decir, la introducción de la mutación D47E origina un aumento de las interacciones con el nuevo residuo. Probablemente, esta estabilización local puede producirse gracias a que se transmite movilidad a otras regiones de la estructura.

Glu 48 (E48) en 2H6X tiene una accesibilidad al disolvente del 53.9% en comparación al 47.3% que presenta Asp 47 (E47) en la variante 2H6Y (Figura 6.15). La variación de ASA respecto a 2H6X es pequeña. E48 en 2H6X forma enlaces de hidrógeno con Glu 44 (2.66 Å) e Ile 45 (3.04 Å); e interacción electrostática con Lys 96 (2.96 Å). D48 en 2H6Y forma también enlace de hidrógeno con Glu 44 (2.98 Å); D48 y Glu 44 forma interacciones electrostáticas con Lys 96, 2.89 y 2.98, respectivamente. La mutación E48D da lugar a una leve disminución de los factores-B en general, siendo ésta un poco mayor en la zona de la mutación. Además, se puede distinguir un aumento de la flexibilidad de los aminoácidos comprendidos entre Gly 65 hasta Gly 74, superior al aumento que ocurre en el resto de la proteína.

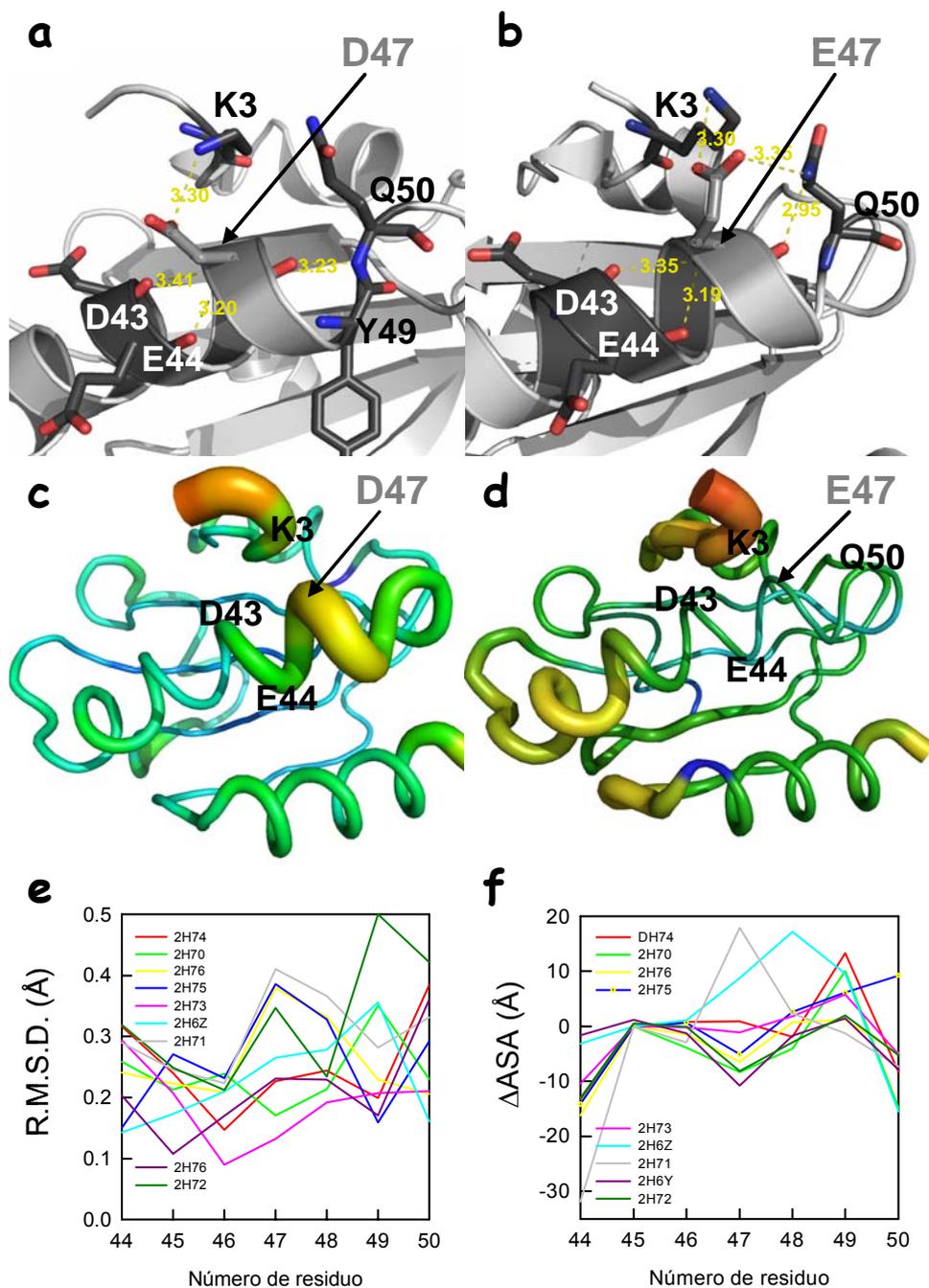


Figura 6.15: a) Estructura tridimensional del monómero A de tiorredoxina WT (2H6X.pdb; resolución 2.60 Å) centrada en el residuo D47 (gris). b) Estructura tridimensional de la variante de tiorredoxina D47E (2H71.pdb; resolución 2.20 Å), centrada en el residuo E47 (gris). En gris oscuro se muestran los residuos con los que el residuo 47 mantiene interacciones polares en una esfera de 3.5 Å. c) y d) representan los factores térmicos B de 2H6X y 2H71, respectivamente. e) RMSD de la cadena principal del monómero A de las variantes de tiorredoxina, en comparación al monómero A de tiorredoxina WT (2H6X), en la zona de la mutación. f) Área superficial accesible al disolvente de las variantes de tiorredoxina respecto de 2H6X, en la zona de la mutación [$\Delta\text{ASA} = \text{ASA}_{\text{mutante}} - \text{ASA}_{\text{WT}}$].

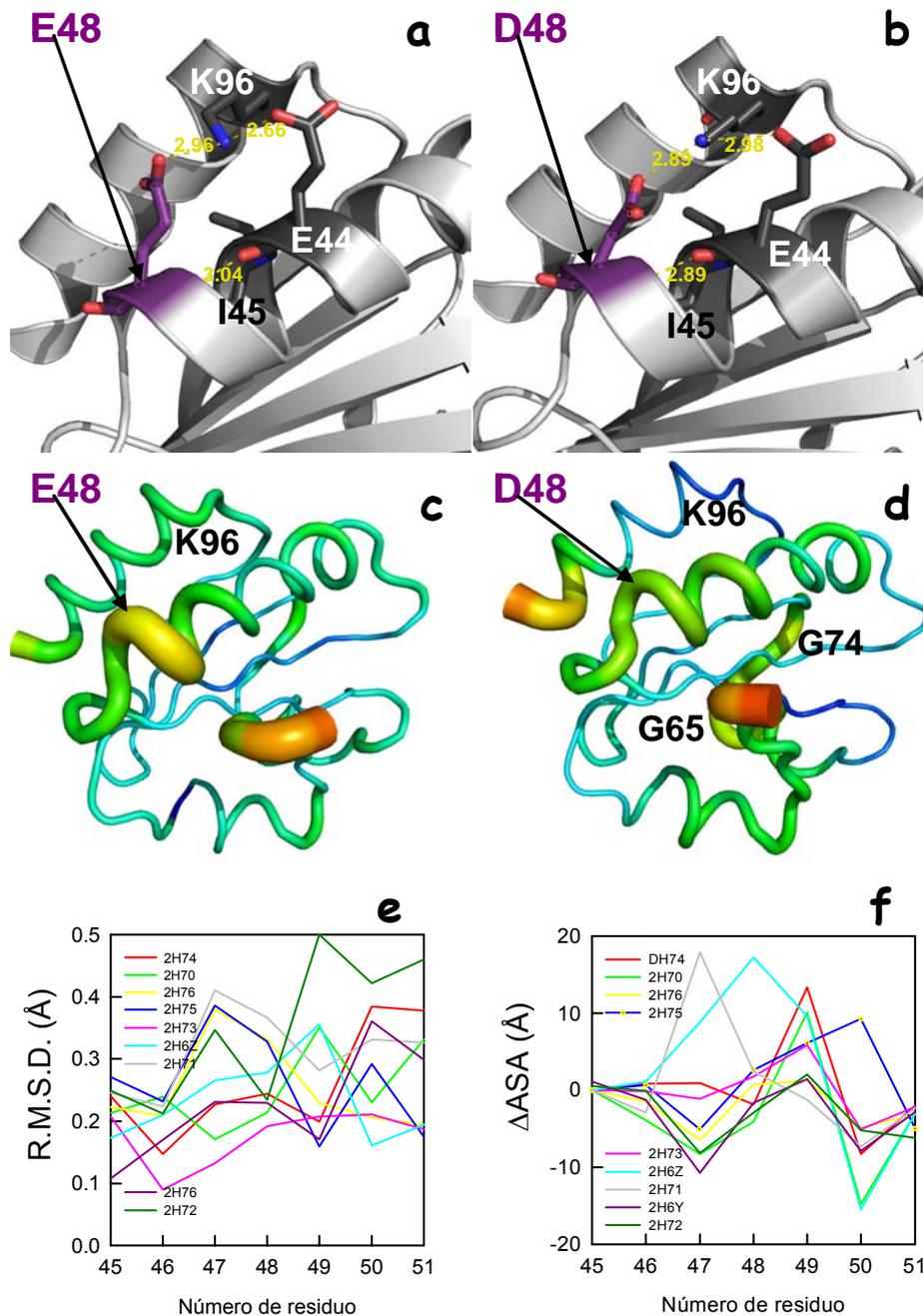


Figura 6.16: a) Estructura tridimensional del monómero A de tiorredoxina WT (2H6X.pdb; resolución 2.60 Å) centrada en el residuo E48 (violeta). b) Estructura tridimensional de la variante de tiorredoxina E48D (2H6Y.pdb; resolución 2.40 Å), centrada en el residuo D48 (violeta). En gris oscuro se muestran los residuos con los que el residuo 47 mantiene interacciones polares en una esfera de 3.5 Å. c) y d) representan los factores térmicos B de 2H6X y 2H6Y, respectivamente. e) RMSD de la cadena principal del monómero A de las variantes de tiorredoxina, en comparación al monómero A de tiorredoxina WT (2H6X), en la zona de la mutación. f) Área superficial accesible al disolvente de las variantes de tiorredoxina respecto de 2H6X, en la zona de la mutación [Δ ASA=ASA_{mutante}-ASA_{WT}].

Glu 85 (E85) en 2H6X tiene una accesibilidad al disolvente del 66% en comparación al 64.6% que presenta Asp 85 (D85) en la variante 2H72 (Figura 6.16). Ésto supone una exposición del nuevo residuo de más de 10 \AA^2 . E85 forma enlaces de hidrógeno con Lys 82 y Asn 83 (2.54 y 2.95, respectivamente). Estos enlaces se mantienen, probablemente a mayores distancias en la variante 2H72. A diferencia del resto de las mutaciones que hemos estudiado, el hecho de introducir la mutación E85D, da lugar a que el nuevo residuo (Asp) forme un contacto cristalino con Gln 98 del monómero B de tiorredoxina. Este nuevo contacto cristalino junto con el hecho de que E85 disminuye el número de interacciones intramoleculares, parece ser que contribuyen a aumentar la flexibilidad de la proteína.

En resumen, hemos comprobado la elevada flexibilidad intrínseca de tiorredoxina. Hemos observado la existencia de regiones que muestran un comportamiento muy peculiar cuando se realiza una mutación. Es decir, existen regiones que independientemente del tipo de mutación que se introduzca en tiorredoxina (estabilizante o desestabilizante), muestran una elevada flexibilidad. La secuencia de aminoácidos comprendidos desde el Asp 43 hasta la Lys 52 (hélice $\alpha 2$) es muy sensible a mutaciones de ácidos carboxílicos estudiadas. Ésto se refleja en las fluctuaciones de los factores-B de esta región. Además, curiosamente, hemos observado que si la mutación se realiza en $\alpha 2$, afecta muy poco a los factores-B del resto de la proteína. Esto podría interpretarse como si $\alpha 2$ tuviera la flexibilidad suficiente para acomodar una mutación conservativa del tipo Asp \leftrightarrow Glu (estabilizante o desestabilizante), para que afecte lo mínimo posible al conjunto de la estructura de la proteína.

Dunker *et al.*, 2002b han propuesto que las proteínas nativas pueden existir en conformaciones ordenadas (plegadas, estructuradas) y desordenadas (desplegadas, desestructuradas), donde la existencia de desorden se determina por la dinámica global de la proteína más que por su estructura secundaria local. Además, los tres tipos de estructura secundaria que caracterizan cadenas ordenadas, α -hélice, β -láminas y *coil*, pueden aparecer también en regiones de desorden intrínseco. Dada la estrecha relación entre desorden y función (Dunker *et al.*, 2002b), las proteínas que presentan desorden se hacen muy interesantes de

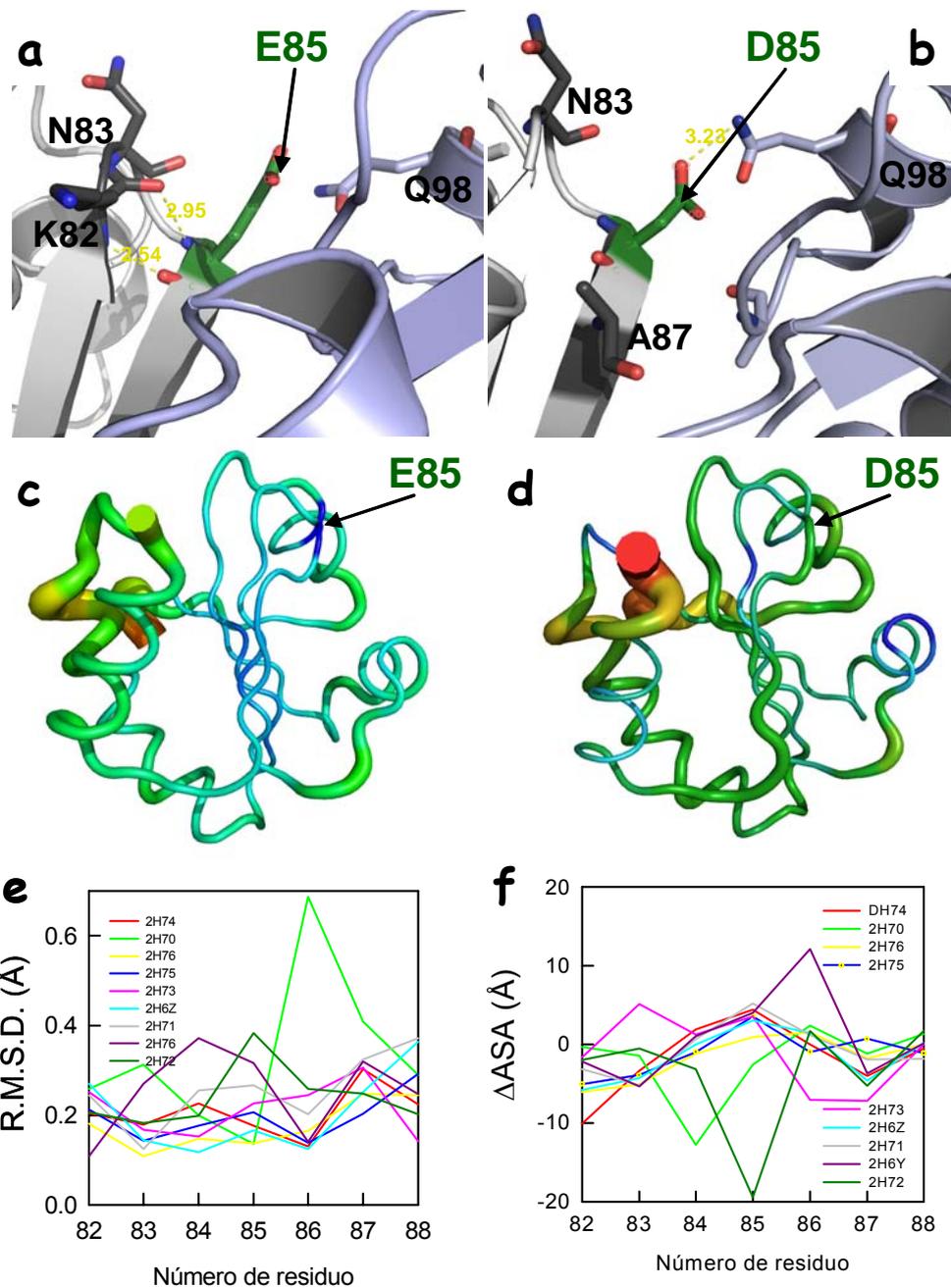


Figura 6.17: a) Estructura tridimensional del monómero A de tiorredoxina WT (2H6X.pdb; resolución 2.60 Å) centrada en el residuo E85 (verde oscuro). b) Estructura tridimensional de la variante de tiorredoxina E85D (2H72.pdb; resolución 2.25 Å), centrada en el residuo D85 (verde oscuro). En gris oscuro se muestran los residuos con los que el residuo 47 mantiene interacciones polares en una esfera de 3.5 Å. c) y d) representan los factores térmicos B de 2H6X y 2H72, respectivamente. e) RMSD de la cadena principal del monómero A de las variantes de tiorredoxina, en comparación al monómero A de tiorredoxina WT (2H6X), en la zona de la mutación. f) Área superficial accesible al disolvente de las variantes de tiorredoxina respecto de 2H6X, en la zona de la mutación [$\Delta\text{ASA} = \text{ASA}_{\text{mutante}} - \text{ASA}_{\text{WT}}$].

estudiar (Wright y Dyson, 1999; Dunker *et al.*, 2002b; Dyson y Wright, 2002; Uversky, 2002). La predictibilidad de regiones desordenadas a partir de los alineamientos de secuencias (Obradovic *et al.*, 2003), la composición observada de bases de esas regiones (Romero *et al.*, 2001), las velocidades de evolución típicamente mayores (Brown *et al.*, 2002) y los patrones de sustitución de aminoácidos distintivos durante la evolución (Radivojac *et al.*, 2002), todas estas características combinadas indican fuertemente que el desorden intrínseco de las proteínas generalmente se encuentra codificado por la secuencia de aminoácidos (Dunker *et al.*, 2002a). Las regiones de desorden intrínseco y regiones de alto factor-B pueden ser asociadas a vibraciones térmicas de átomos individuales y a una elevada flexibilidad intramolecular (Smith *et al.*, 1986; Ringe y Petsko, 1986; Rhodes 1993), así que es importante examinar si regiones de elevado factor-B se parecen más a regiones de desorden o de bajo factor-B. Radivojac *et al.*, 2004, abordan el problema clasificando la flexibilidad de la proteínas en cuatro categorías: (1) regiones ordenadas de factor-B bajo, (2) regiones ordenadas de factor-B alto, (3) regiones con bajo grado de desorden y (4) regiones con elevado grado de desorden. Como resultado, encuentran que regiones ordenadas de elevado factor-B presentan un índice de flexibilidad promedio, un promedio de hidrofiliidad, un promedio de carga neta (absoluta) y una carga total mayores que otras regiones que presentan un bajo grado o alto de desorden. Por otro lado, las regiones ordenadas de factor-B bajo son ricas en residuos hidrofóbicos y carecen de residuos cargados en comparación a las otras tres clases. Finalmente, las regiones de elevado desorden se diferencian notablemente de las regiones con bajo desorden o las regiones ordenadas de alto factor-B, siendo su carga total relativamente alta, pero su carga neta (absoluta) muy variable. Sin embargo, la conexión conceptual entre desorden y flexibilidad hoy día no es tan clara. Ya que, claramente, los métodos que predicen desorden no pueden predecir la rigidez de los residuos del sitio activo (Schelessinger *et al.*, 2006). Además, experimentalmente los factores-B dependen de la resolución experimental, de los contactos cristalinos y de los procesos de refinamiento (Sheriff *et al.*, 1985, Tronrud, 1996, Parthasarathy y Murthy, 2000). Por ello, los factores-B son una aproximación razonable para inferir fluctuaciones atómicas de los átomos de la cadena principal, pero se puede cometer mayor error si se intenta predecir el grado de movimiento de los átomos de las cadenas laterales (Doniach y Eastman, 1999,

Kumar *et al.*, 2001). Estas influencias pueden minimizarse realizando un análisis como el que se explica a continuación en la Sección 6.6. Los contactos se han calculado con el programa CONTACT del paquete de programas CCP4 y con NOC. Ringe y Petsko, 1986; Suel *et al.*, 2003, al

6.5 DETERMINACIÓN DE INTERACCIONES EXISTENTES ENTRE AMINOÁCIDOS

Ya hemos visto cómo las interacciones que mantienen los residuos aspartato y glutamato (muy conservados en los alineamientos de secuencias obtenidos para tiorredoxina de *E. coli*) pueden tener efectos significativos sobre la estabilidad de tiorredoxina (Capítulo 5, Godoy-Ruiz *et al.*, 2006). Como respuesta a la mutación, tiorredoxina se comporta como una proteína altamente flexible, como hemos podido deducir de los parámetros térmicos atómicos (factores-B).

En este apartado, planteamos la posibilidad de poder mapear las interacciones energéticas existentes en tiorredoxina de *E. coli* (WT). Para ello, hemos aplicado el método estadístico desarrollado por Lockless y Ranganathan, 1999, al alineamiento de secuencias obtenido para tiorredoxina de *E. coli* y hemos estudiado las posiciones de tiorredoxina mutadas (D2, D9, D10, D13, D43, E44, D47, E48 y E85). Este análisis ha predicho un conjunto de posiciones acopladas energéticamente (Sección 6.5.1) con las mutaciones superficiales de ácidos carboxílicos, y se ha comparado con las interacciones a corta y larga distancia detectadas estructuralmente (Jain y Ranganathan, 2004) en cada una de las variantes de tiorredoxina (D2E, D9E, D10E, D13E, D43E, E44D, D47E, E48D y E85D).

6.5.1 ESTUDIO DE LOS ACOPLAMIENTOS ENTRE LOS AMINOÁCIDOS CONSERVADOS EVOLUTIVAMENTE

La comunicación entre posiciones distantes en proteínas es fundamental para su función y a menudo define el papel biológico de una proteína dentro de la familia a la que pertenece. En un intento por predecir interacciones energéticas entre posiciones alejadas, hemos realizado un análisis estadístico de alineamientos de secuencias. La base de este análisis consiste en que, la existencia de un acoplamiento entre dos aminoácidos en una proteína, por razones estructurales o

funcionales, puede haber causado que esas dos posiciones de aminoácidos hayan co-evolucionado (Lockless y Ranganathan, 1999).

La evolución del plegamiento de una proteína es el resultado de un gran número de mutaciones aleatorias, sujetas a las constricciones impuestas por la estructura o la función de la proteína. La teoría descrita por Lockless y Ranganathan, 1999, se basa en dos hipótesis, derivadas de la observación directa de secuencias de proteínas: 1) la conservación de una determinada posición en un alineamiento de secuencias se define como la desviación de la frecuencia del aminoácido en esa posición de su valor medio (calculado de la base de datos de proteínas Swiss-Prot); y 2) el acoplamiento estadístico de dos posiciones, i y j , se define como el grado en el que la frecuencias de aminoácidos en la posición i cambian en respuesta a una perturbación de las frecuencias en otra posición, j . Ésto debería constreñir mutuamente la evolución de esas dos posiciones, y por lo tanto, reflejarse en el acoplamiento estadístico de las distribuciones de los aminoácidos implicados.

Hemos utilizado el alineamiento de secuencias obtenido con BLAST, utilizando la secuencia de tiorredoxina de *E.coli* como la secuencia a comparar (aproximadamente 500 secuencias; véase el apartado 2.2.6.2 en *Materiales y métodos*). Cada posición en el alineamiento se puede describir como un vector de 20 elementos (\vec{P}_i^x) de las frecuencias individuales de todos los aminoácidos que existen en esa posición. Cada elemento del vector es transformado a la probabilidad del aminoácido correspondiente con la función de densidad binomial siguiente:

$$P(x) = \frac{N!}{n_x!(N-n_x)!} p_x^{n_x} (1-p_x)^{N-n_x} \quad (6.1)$$

donde N es número total de secuencias del alineamiento, n_x es el número de secuencias con el aminoácido x , y p_x es la frecuencia media del aminoácido x en todas las proteínas existentes en una base de datos. Este vector, únicamente representa todos los cambios en la distribución de un aminoácido en relación a la conservación en una posición determinada. La probabilidad de que exista un aminoácido x (aspartato (D) o glutamato (E)) en una posición i relativa a la

probabilidad de que exista en otra posición j , se relaciona con la energía libre estadística, que separa las dos posiciones i y j ($\Delta G_{i \rightarrow j}^x$) para el aminoácido x , en una distribución de Boltzmann (Lockless y Ranganathan, 1999), mediante la siguiente expresión:

$$\frac{P_i^x}{P_j^x} = e^{\frac{\Delta G_{i \rightarrow j}^x}{kT^*}} \quad (6.2)$$

donde kT^* es una unidad arbitraria de energía, P_i^x es la probabilidad de que exista un aminoácido x en una posición i y viene dada por la probabilidad binomial del número observado de aminoácidos x , dada su frecuencia media en todas las proteínas del alineamiento (ecuación 6.1).

Si se considera una posición j como estado de referencia para todas las posiciones (posición donde todos los aminoácidos poseen una frecuencia media en el alineamiento de secuencias) y se utiliza la ecuación 6.2, se puede transformar \vec{P}_i^x en un vector de energías estadísticas (ΔG_i^{stat}) que representan las constricciones evolutivas de la posición i y viene dado por la siguiente ecuación:

$$\Delta G_i^{stat} = kT^* \sqrt{\sum_x \left(\ln \frac{P_i^x}{P_A^x} \right)^2} \quad (6.3)$$

Finalmente, para medir el acoplamiento funcional existente entre posiciones, se mide el vector energía estadística en una posición dada i , en dos condiciones: 1) el alineamiento completo (ΔG_i^{stat}) y 2) el subconjunto de secuencias del alineamiento, que representa una perturbación de las frecuencias de aminoácidos en otra posición j ($\Delta G_{i/\delta j}^{stat}$). La diferencia entre estos dos vectores de energía da como resultado la energía de acoplamiento estadística ($\Delta \Delta G_{i,j}^{stat}$) entre las posiciones i y j :

$$\Delta \Delta G_{i,j}^{stat} = kT^* \sqrt{\sum_x \left(\ln \frac{P_{i/\delta j}^x}{P_{A/\delta j}^x} \ln \frac{P_i^x}{P_A^x} \right)^2} \quad (6.4)$$

que representa cuantitativamente el grado en el que la probabilidad de un aminoácido en existente en la posición i depende de una perturbación en j .

Con objeto de simplificar los cálculos, se han escrito programas para llevar a cabo proceso de cálculo de las energías de acoplamiento (Ranga1-5.exe), que funcionan en entorno MS-DOS (los programas han sido desarrollados por el grupo de investigación de José M. Sanchez-Ruiz).

Siguiendo el procedimiento descrito, se han calculado las energías de acoplamiento ($\Delta\Delta G_{i,j}^{stat}$) entre las posiciones i en las que existen residuos ácidos carboxílicos en tiorredoxina de *E. coli* (D2, D9, D10, D13, D43, E44, D47, E48 y E85) y cualquier otra posición j . Las energías de acoplamiento se representan en una gráfica frente al número de residuo. Además, se han mapeado las energías de acoplamiento sobre la estructura de tiorredoxina.

Hemos podido comprobar que en la mayoría de los casos, cuando realizamos una perturbación, la mayoría de las posiciones en tiorredoxina no muestran acoplamiento con la posición perturbada (posiciones coloreadas en azul oscuro en las Figuras 6.18-26). Sin embargo, los acoplamientos resultantes los podemos clasificar en dos categorías categorías:

1) Agrupaciones de residuos que se encuentran en un ambiente próximo a la mutación perturbada. Este hallazgo es consistente con la esperada propagación de energía desde el sitio en el que introducimos la perturbación. Acoplamientos de este tipo los podemos encontrar cuando perturbamos tiorredoxina en las posiciones 9, 43 y 47. En la Figura 6.18-B, podemos apreciar cómo al perturbar la posición 9 las posiciones 28, 37, 63 y 66 resultan con una energía de acoplamiento mayor de 4 kT*. Estos resultados concuerdan con la caracterización de las interacciones locales que se ha hecho previamente para la variante D9E (Figura 6.10) en la Sección 6.4. Comprobamos cómo al introducir la mutación E9 (mutación 2.72°C más inestable que tiorredoxina *WT*) aumentaba la movilidad de la secuencia de residuos desde Asn 63 hasta Thr 66.

En la Figura 6.19-B, podemos ver otro ejemplo de transmisión de energía entre posiciones acopladas al perturbar la posición 43 de tiorredoxina. Observamos cómo al perturbar la posición 43 en tiorredoxina, las posiciones 9, 29, 37, 57, 59, 61, 63 y 75 muestran una energía de acoplamiento mayor de 4 kT*.

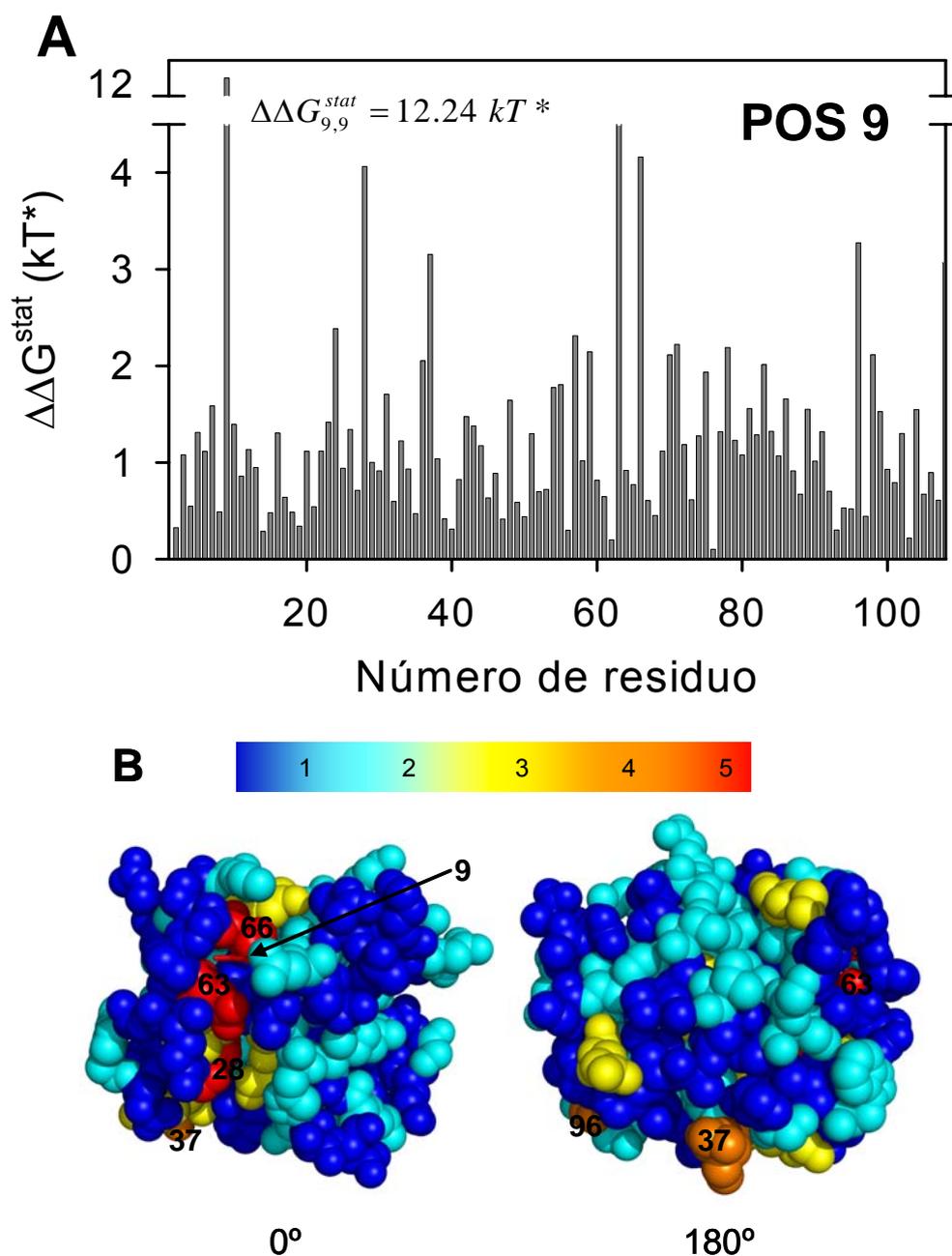


Figura 6.18: Acoplamiento estadístico para la posición 9 en el alineamiento de secuencias de tiorredoxina de *E. coli*. A) Mapa completo de $\Delta\Delta G_{i,j}^{stat}$ para la posición 9 (D) de tiorredoxina de *E. coli*. B) Posiciones acopladas de tiorredoxina de *E. coli* con la posición 9. Estas posiciones acopladas describen interacciones de la posición 9 en posiciones cercanas o alejadas de tiorredoxina. Se muestran dos orientaciones de la molécula de tiorredoxina en las que se representan las posiciones acopladas según la escala de color del azul (1 kT*) al rojo (5 kT*). Valores de $\Delta\Delta G_{i,j}^{stat} < 1$ se han considerado ruido estadístico.

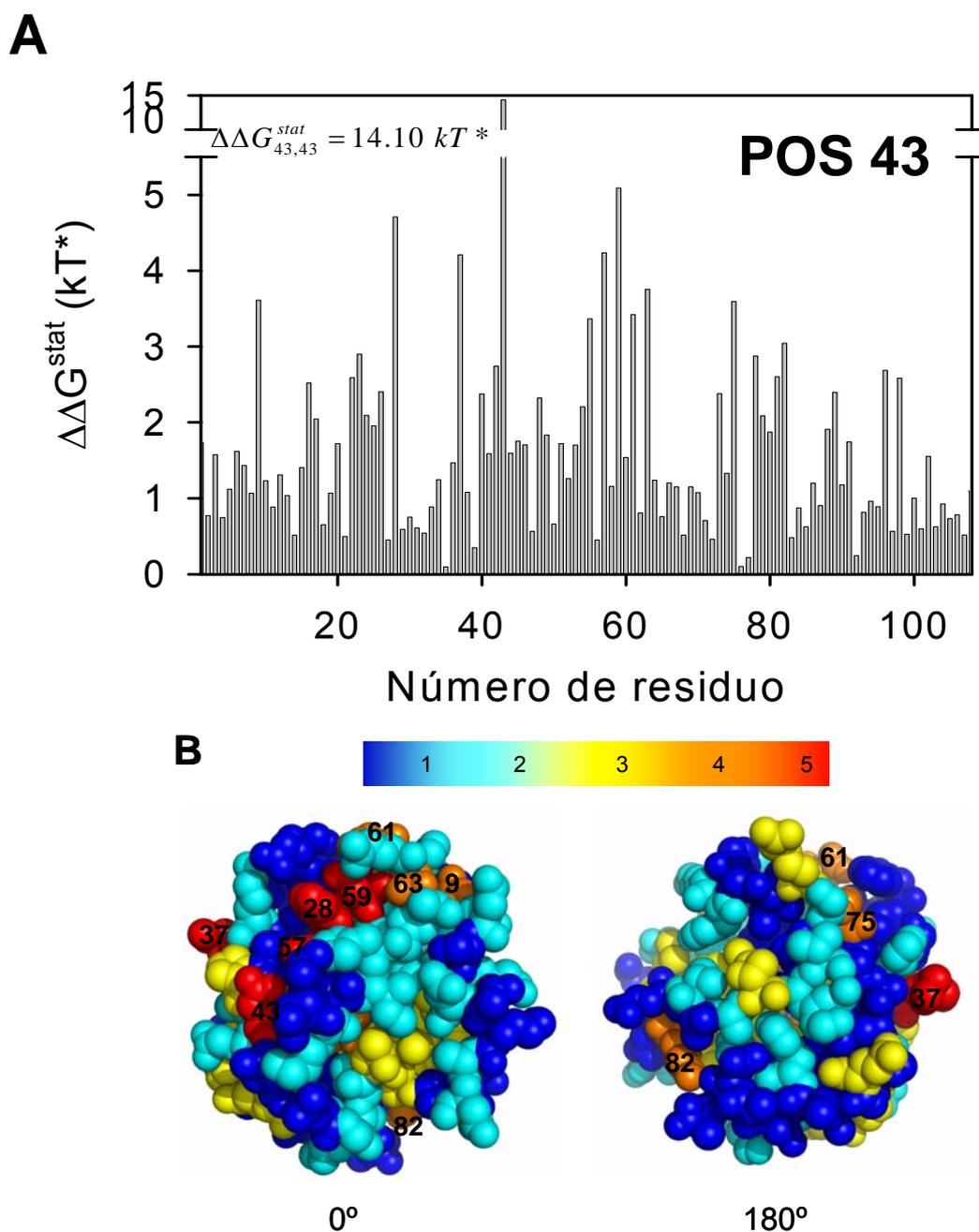


Figura 6.19: Acoplamiento estadístico para la posición 43 en el alineamiento de secuencias de tiorredoxina de *E. coli*. A) Mapa completo de $\Delta\Delta G_{i,j}^{stat}$ para la posición 43 (D) de tiorredoxina de *E. coli*. B) Posiciones acopladas de tiorredoxina de *E. coli* con la posición 43. Estas posiciones acopladas describen interacciones de la posición 43 en posiciones cercanas o alejadas de tiorredoxina. Se muestran dos orientaciones de la molécula de tiorredoxina en las que se representan las posiciones acopladas según la escala de color del azul (1 kT*) al rojo (5 kT*). Valores de $\Delta\Delta G_{i,j}^{stat} < 1$ se han considerado ruido estadístico.

En la Sección anterior, hemos visto que, en la estructura de la variante D47E, los residuos Asp 9, Ala 29, Asn 63 e Ile 75 se encuentran en lazos que disminuyen su flexibilidad al introducir la mutación D43E (0.69°C más estable que tiorredoxina *WT*). El residuo Met 37 (α 2) presenta doble conformación de la cadena lateral en la variante D43E. Además, Ile 75 se encuentra interaccionando con Cis 35, que forma parte del puente disulfuro de tiorredoxina, por lo que una perturbación del residuo 43 podría llegar a afectar a la función del tiorredoxina.

En la Figura 6.20-B se mapean, sobre la estructura de tiorredoxina, las posiciones (2, 9, 31, 35 y 57) que muestran energía de acoplamiento mayor de 4 kT* al perturbar la posición 47. Nuevamente, parece ser que los residuos acoplados a la posición 47 se encuentran en regiones de la variante D47E, que experimentalmente hemos comprobado, aumentan su flexibilidad al introducir la mutación D47E (la variante D47E es 1.32 °C más estable que tiorredoxina *WT*). Al igual que ocurría en el caso anterior, al perturbar la posición D47, el residuo Cis 35 muestra una elevada energía de acoplamiento.

2) Hemos observado un acoplamiento inesperado entre residuos alejados de la posición que se perturba. Como ejemplos, se mostramos el mapeado de las energías de interacción resultante de realizar una perturbación en las posiciones 2 (Figura 6.21), 13 (Figura 6.22), 44 (Figura 6.23), 48 (Figura 6.24) y 85 (Figura 6.25). Para estos acoplamientos a larga distancia, hasta el momento no hemos encontrado un patrón que infiera el tipo de interacción que tiene lugar.

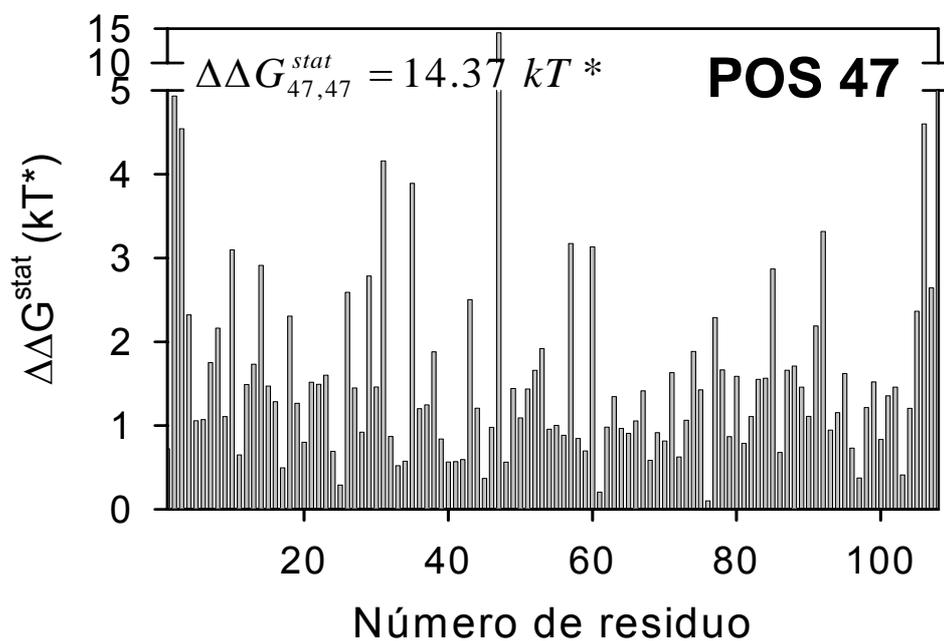
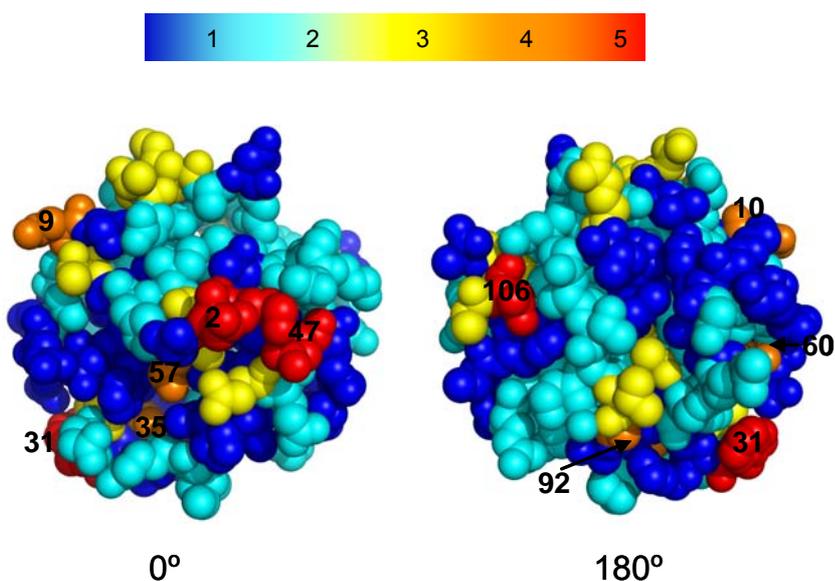
A**B**

Figura 6.20: Acoplamiento estadístico para la posición 47 en el alineamiento de secuencias de tiorredoxina de *E. coli*. A) Mapa completo de $\Delta\Delta G_{i,j}^{stat}$ para la posición 47 (D) de 7tiorredoxina de *E. coli*. B) Posiciones acopladas de tiorredoxina de *E. coli* con la posición 44. Estas posiciones acopladas describen interacciones de la posición 47 en posiciones cercanas o alejadas de tiorredoxina. Se muestran dos orientaciones de la molécula de tiorredoxina en las que se representan las posiciones acopladas según la escala de color del azul (1 kT*) al rojo (5 kT*). Valores de $\Delta\Delta G_{i,j}^{stat} < 1$ se han considerado ruido estadístico.

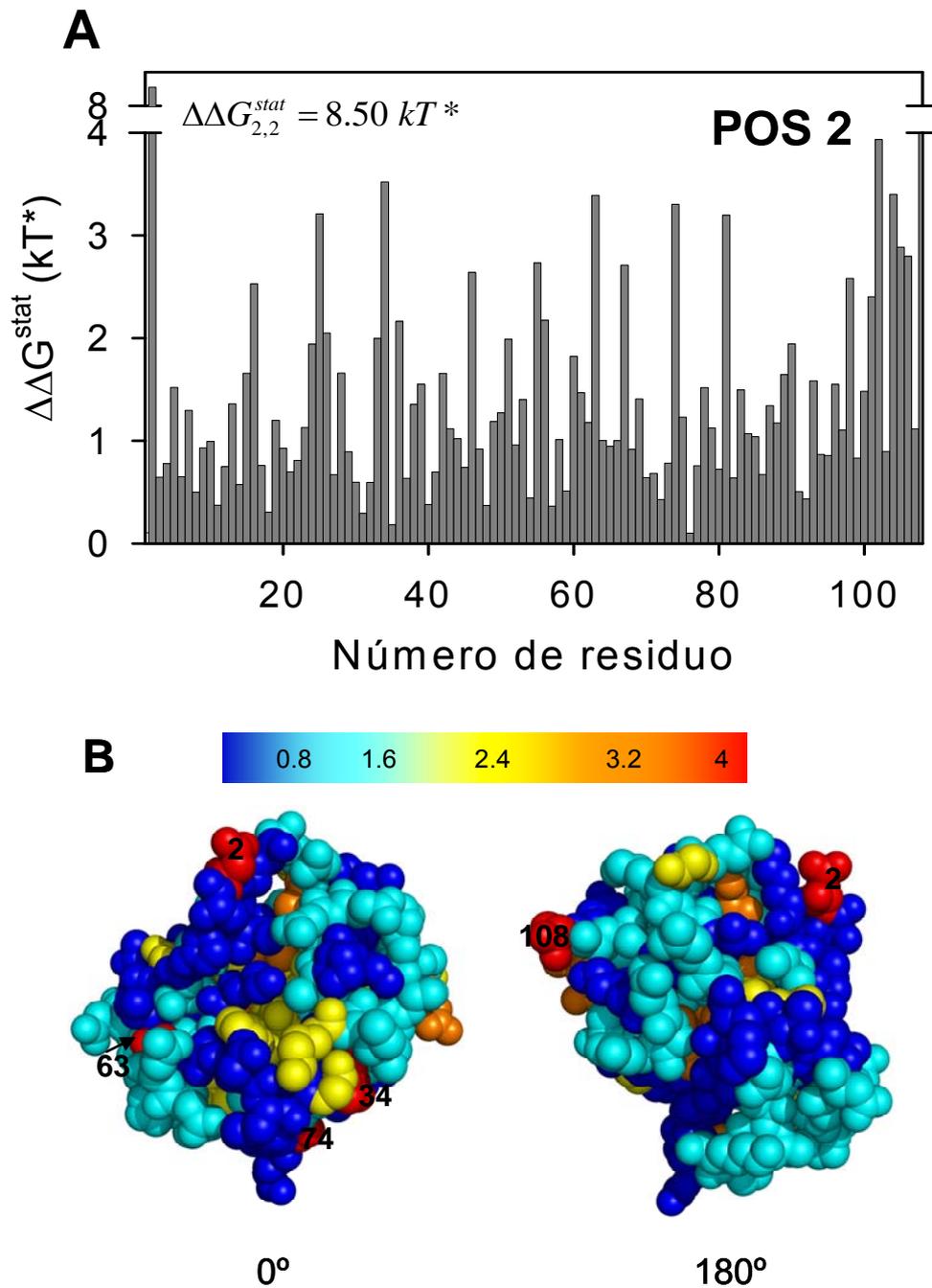


Figura 6.21: Acoplamiento estadístico para la posición 2 en el alineamiento de secuencias de tiorredoxina de *E. coli*. A) Mapa completo de $\Delta\Delta G_{i,j}^{stat}$ para la posición 2 (D) de tiorredoxina de *E. coli*. B) Posiciones acopladas de tiorredoxina de *E. coli* con la posición 2. Estas posiciones acopladas describen interacciones de la posición 2 en posiciones cercanas o alejadas de tiorredoxina. Se muestran dos orientaciones de la molécula de tiorredoxina en las que se representan las posiciones acopladas según la escala de color del azul (0.8 kT^*) al rojo (4 kT^*). Valores de $\Delta\Delta G_{i,j}^{stat} < 0.8$ se han considerado ruido estadístico

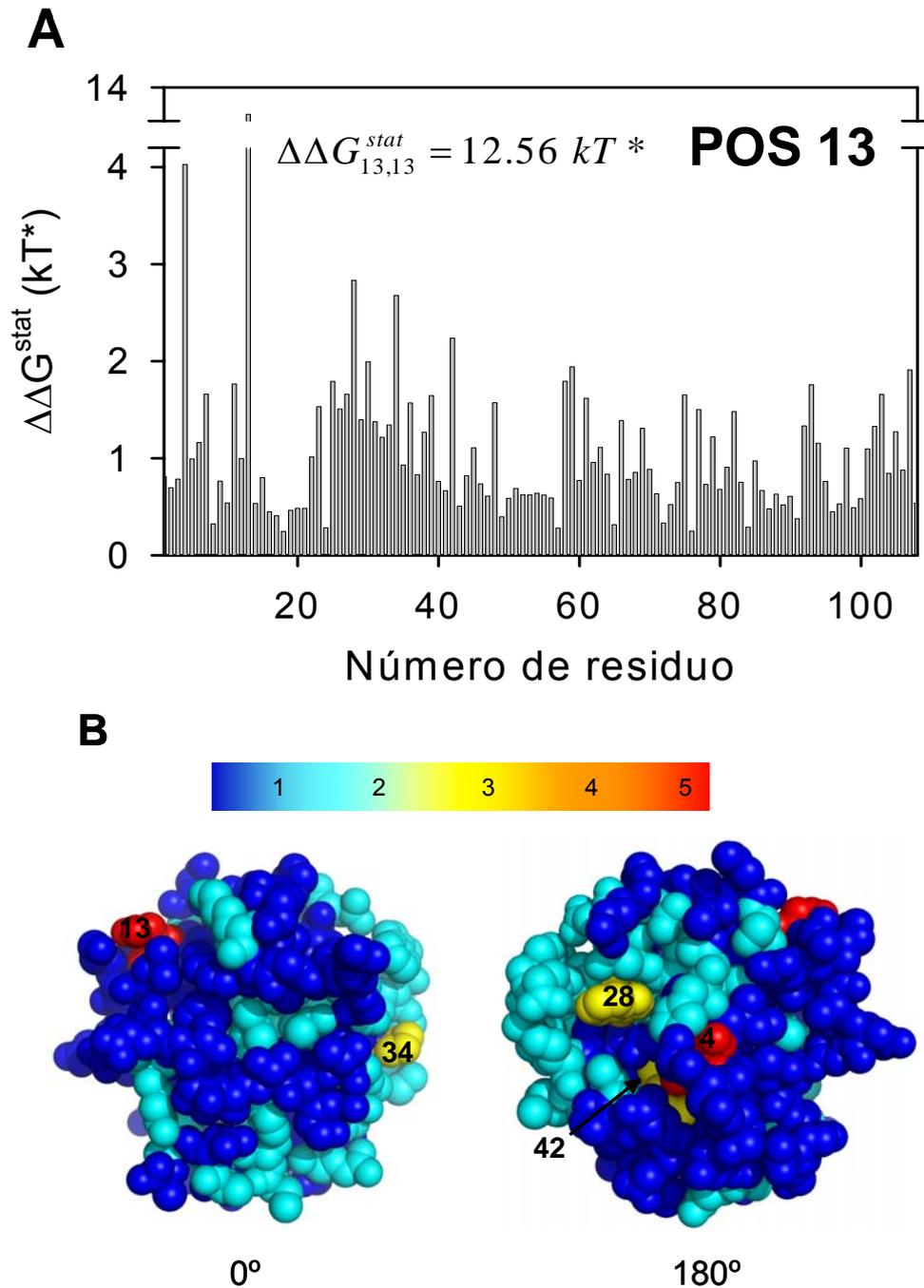


Figura 6.21: Acoplamiento estadístico para la posición 13 en el alineamiento de secuencias de tiorredoxina de *E. coli*. A) Mapa completo de $\Delta\Delta G_{i,j}^{stat}$ para la posición 13 (D) de tiorredoxina de *E. coli*. B) Posiciones acopladas de tiorredoxina de *E. coli* con la posición 13. Estas posiciones acopladas describen interacciones de la posición 13 en posiciones cercanas o alejadas de tiorredoxina. Se muestran dos orientaciones de la molécula de tiorredoxina en las que se representan las posiciones acopladas según la escala de color del azul (1 kT*) al rojo (5 kT*). Valores de $\Delta\Delta G_{i,j}^{stat} < 1$ se han considerado ruido estadístico.

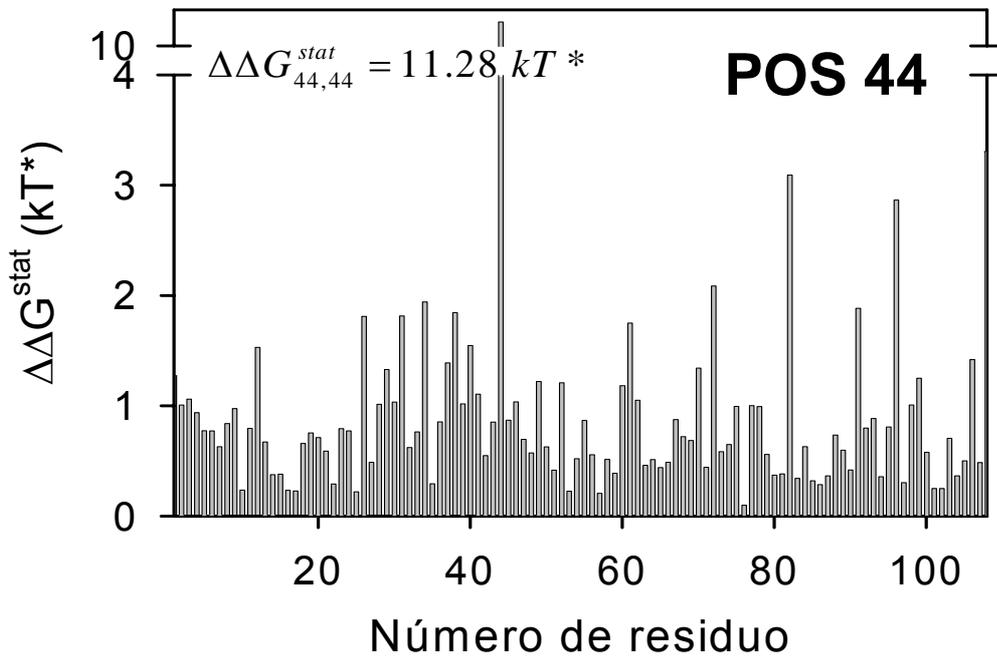
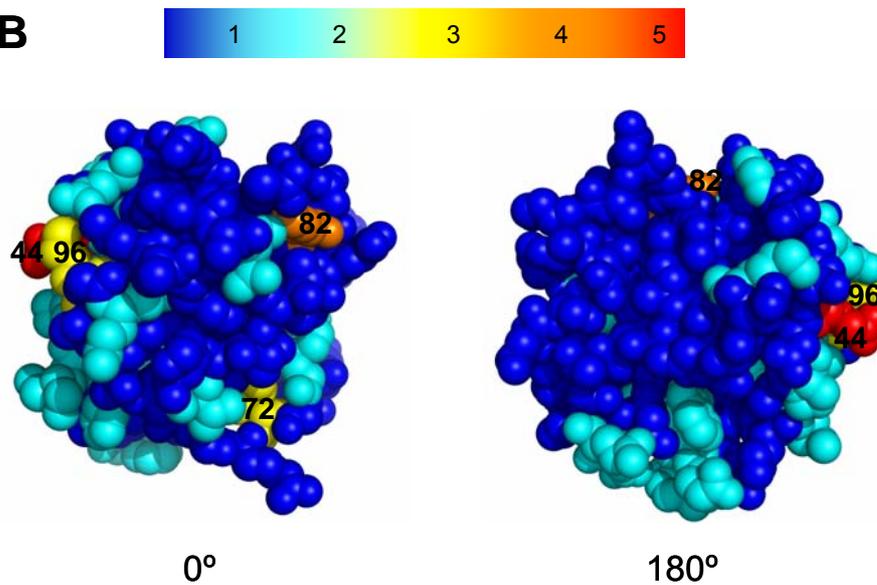
A**B**

Figura 6.23: Acoplamiento estadístico para la posición 44 en el alineamiento de secuencias de tiorredoxina de *E. coli*. A) Mapa completo de $\Delta\Delta G_{i,j}^{stat}$ para la posición 44 (D) de tiorredoxina de *E. coli*. B) Posiciones acopladas de tiorredoxina de *E. coli* con la posición 44. Estas posiciones acopladas describen interacciones de la posición 44 en posiciones cercanas o alejadas de tiorredoxina. Se muestran dos orientaciones de la molécula de tiorredoxina en las que se representan las posiciones acopladas según la escala de color del azul (1 kT*) al rojo (5 kT*). Valores de $\Delta\Delta G_{i,j}^{stat} < 1$ se han considerado ruido estadístico.

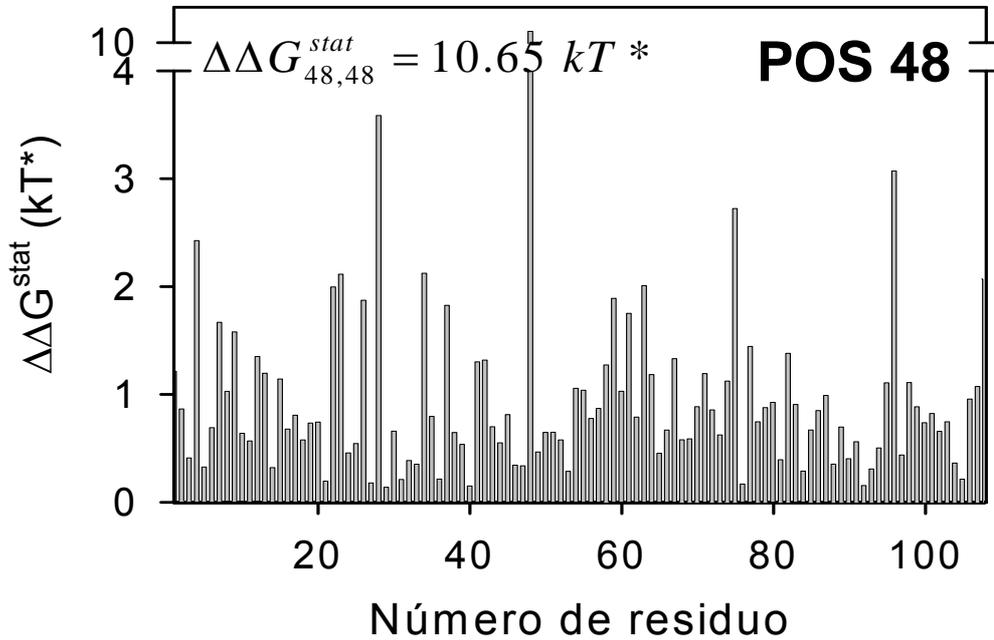
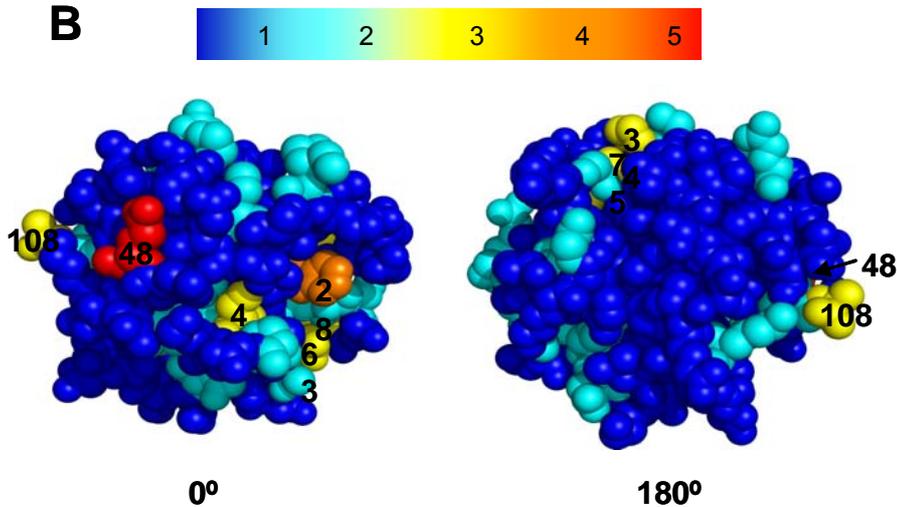
A**B**

Figura 6.25: Acoplamiento estadístico para la posición 48 en el alineamiento de secuencias de tiorredoxina de *E. coli*. A) Mapa completo de $\Delta\Delta G_{i,j}^{stat}$ para la posición 48 (D) de tiorredoxina de *E. coli*. B) Posiciones acopladas de tiorredoxina de *E. coli* con la posición 48. Estas posiciones acopladas describen interacciones de la posición 48 en posiciones cercanas o alejadas de tiorredoxina. Se muestran dos orientaciones de la molécula de tiorredoxina en las que se representan las posiciones acopladas según la escala de color del azul (1 kT*) al rojo (5 kT*). Valores de $\Delta\Delta G_{i,j}^{stat} < 1$ se han considerado ruido estadístico.

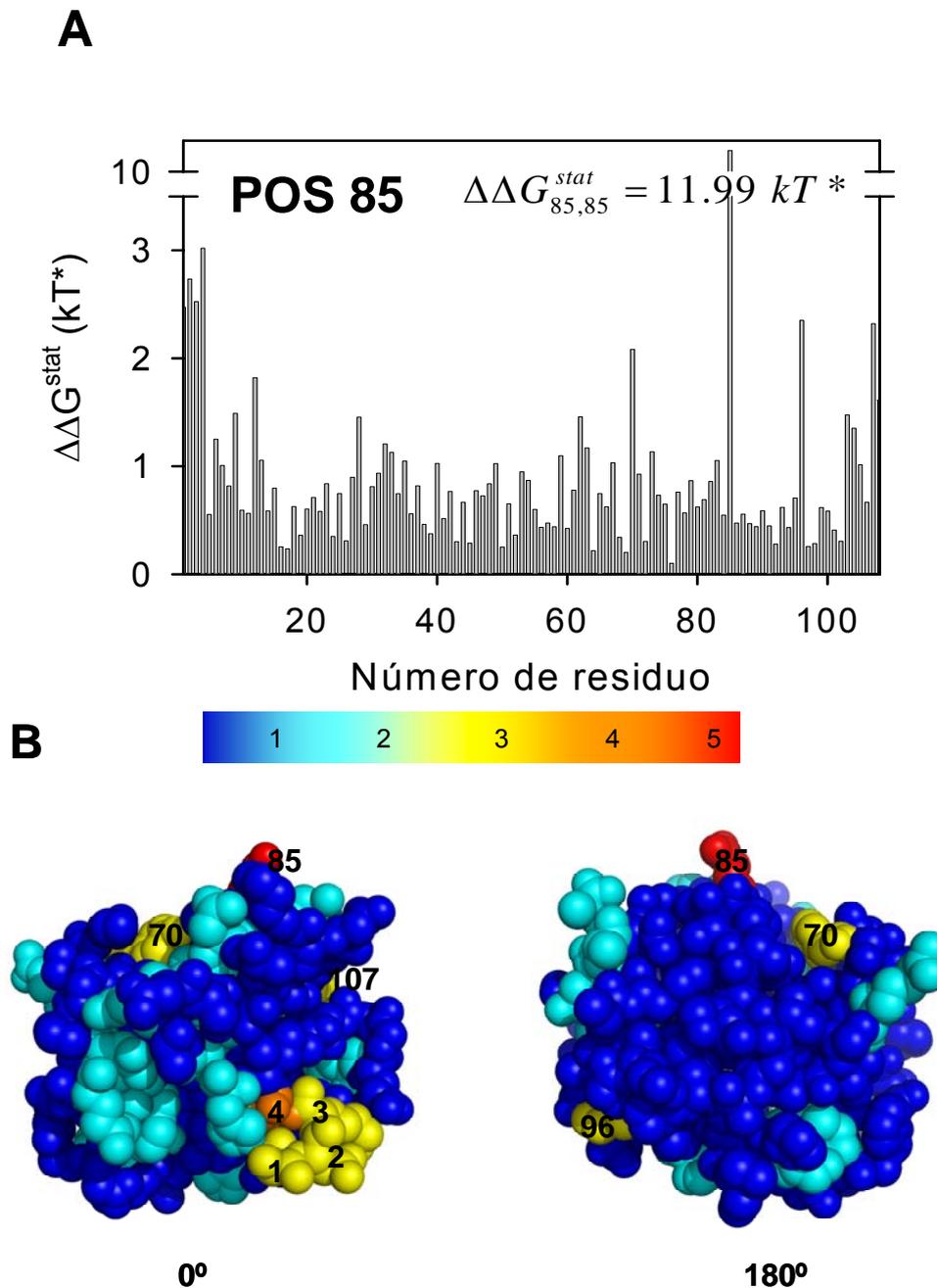


Figura 6.26: Acoplamiento estadístico para la posición 85 en el alineamiento de secuencias de tiorredoxina de *E. coli*. A) Mapa completo de $\Delta\Delta G_{i,j}^{stat}$ para la posición 85 (D) de tiorredoxina de *E. coli*. B) Posiciones acopladas de tiorredoxina de *E. coli* con la posición 85. Estas posiciones acopladas describen interacciones de la posición 85 en posiciones cercanas o alejadas de tiorredoxina. Se muestran dos orientaciones de la molécula de tiorredoxina en las que se representan las posiciones acopladas según la escala de color del azul (1 kT*) al rojo (5 kT*). Valores de $\Delta\Delta G_{i,j}^{stat} < 1$ se han considerado ruido estadístico.

6.5.2 INTERACCIONES ENERGÉTICAS LOCALES PREDICHAS POR ANÁLISIS ESTRUCTURALES

En principio, el valor de la energía libre termodinámica total de un residuo i viene dado por el conjunto jerárquico de interacciones estructurales:

$$\Delta G_i = \Delta G_i^{\text{intrinsic}} + \sum_j \Delta \Delta G_{i,j} + \sum_k \sum_j \Delta^3 G_{i,j,k} + \dots \quad (6.5)$$

donde $\Delta^3 G$ representa los n-acoplamiento entre residuos; para el caso de tres residuos, correspondería al triple acoplamiento entre los residuos i , j y k (Jain y Ranganathan, 2004). Para el conjunto total de residuos que componen una proteína, esta relación (ecuación 6.5) hace que las proteínas lleguen a tener un alto grado de conexión entre residuos, de modo que la introducción de un cambio de energía libre realizando, por ejemplo, una mutación puede afectar a muchas de interacciones que mantiene el residuo mutado.

Hemos aplicado el método de Jain y Ranganathan, 2004, para predecir interacciones energéticas locales, mediante un análisis estructural, de las nueve variantes de tiorredoxina de estructura resuelta: D2E (2H74.pdb), D9E (2H70.pdb), D10E (2H76.pdb), D13E (2H75.pdb), D43E (2H73.pdb), E44D (2H6Z.pdb), D47E (2H71.pdb), E48D (2H6Y.pdb) y E85D (2H72.pdb). La resolución a la que se han resuelto dichas estructuras, las R.M.S.D., los valores de R y R_{free} se recogen en la Tabla 6.2. En primer lugar hemos superpuesto la estructura de cada variante, mutación de ácido carboxílico, con la de tiorredoxina de *E. coli* (2H6X.pdb). Se han alineado sólo los $C\alpha$ utilizando el programa NOC (Chenmengen y H.X. Cang, IBP-CAS; <http://noc.ibp.ac.cn/>).

La medida del efecto que ocasiona una mutación en cada estructura se realiza cuantificando el desplazamiento atómico de la estructura de la proteína variante respecto de *WT* ($|\overline{\Delta r_i}|$); para evitar que en este análisis influyan los errores experimentales asociados a la determinación de ambas estructuras, $|\overline{\Delta r_i}|$ se corrige por el error asociado a la determinación de ambas estructuras usando el método de Stroud y Fauman, 1995.

Por lo tanto, $\Delta r_{norm,i}$ es un parámetro que da la significancia del cambio estructural:

$$\Delta r_{norm,i} = \frac{|\vec{r}_2 - \vec{r}_1|}{\sqrt{\sigma_1^2 + \sigma_2^2}} \quad (6.6)$$

donde \vec{r}_1 y \vec{r}_2 representa la posición del átomo i en las estructuras alineadas 1 y 2, y σ_1 y σ_2 son los errores asociados calculados según el método de Stroud y Fauman, 1995, como se explica a continuación.

Debido a que la precisión de las coordenadas atómicas de una estructura determinada por cristalografía de rayos-X depende de factores como la resolución del análisis, la conectividad entre átomos por la unidad asimétrica, los factores-B y el número de reflexiones únicas utilizadas, se definen el error σ_R como la desviación estándar tridimensional de las diferencias en la posición atómica entre la estructura “verdadera” y la observada:

$$\sigma_R = \sqrt{\frac{3}{2}} \varepsilon_x \quad (6.7)$$

donde ε_x es la desviación estándar monodimensional de las diferencias en la posición atómica entre la estructura “verdadera” y la observada. ε_x es el error estimado a cada factor-B, función del cociente entre el número de átomos por unidad asimétrica y el número de reflexiones únicas tomadas ($\varepsilon_x(B, ATOM / REFL)$) y se define por la siguiente ecuación:

$$\varepsilon_x(B, ATOM / REFL) = Intercept(B) + Slope(B) * e^{(-2 * ATOM / REFL)} \quad (6.8)$$

donde $Intercept(B)$ y $Slope(B)$ se definen por las siguientes ecuaciones:

$$Slope(B) = k1 + k2 * e^{(B/k3)} \quad (6.9)$$

donde $k_1=-0.687$, $k_2=-0.00223$ y $k_3=6.16$.

$$\text{Intercept}(B) = k_4 + k_5 * e^{(B/k_6)} \quad (6.10)$$

donde $k_4=0.642$, $k_5=0.00852$ y $k_6=7.88$.

En las Figuras 6.27-6.35 hemos respresentado los resultados del análisis que hemos descrito arriba (ecuaciones 6.6-6.10). En cada figura (gráfica de barras y representación colorimétrica) se han representado los desplazamientos atómicos pesados (factores-B, número de reflexiones únicas, número de átomos en la celdilla unidad) de las posiciones atómicas de la estructura de cada variante superpuesta con la estructura de tiorredoxina *WT*. En las gráficas de barras se han identificado con números las posiciones atómicas de las cadenas laterales que muestran un $\Delta r_{\text{norm}} > 2.5 \text{ \AA}$ y se han señalado en color los desplazamientos atómicos de los $C\alpha$ que se observan en el panel b) de cada figura (Tabla 6.6).

Hemos intentado establecer una correlación entre los acoplamientos energéticos predichos por los alineamientos de secuencias y los acoplamientos entre residuos en las estructuras mutadas. Sin embargo, no ha sido posible establecer un patrón de comportamiento para los desplazamientos atómicos. La no existencia de una correlación aparente puede ser debido a difrentes razones. Nosotros, hemos considerado el hecho de que las posiciones mutadas se encuentran mayormente en la superficie de las proteínas. Esto hace que los residuos mutados no se encuentren tan restringidos estructuralmente y por lo tanto, puede que la transmisión de los efectos de la mutación no sea tan directa como en el caso de residuos que se encuentran enterrados en el interior de la proteína y participan en la unión específica de un ligando (Jain y Ranganathan, 2004).

Por otro lado, la determinación de los desplazamientos atómicos corregidos por los factores-B, el número de reflexiones únicas y el número de átomos en la celdilla unidad, corrobora las observaciones previas realizadas en relación a la flexibilidad de tiorredoxina. En la Tabla 6.6 se recogen las “regiones calientes” de las estructuras de las nueve variantes de tiorredoxina. Estos resultados apoyan nuestras observaciones de los factores térmicos. Observamos

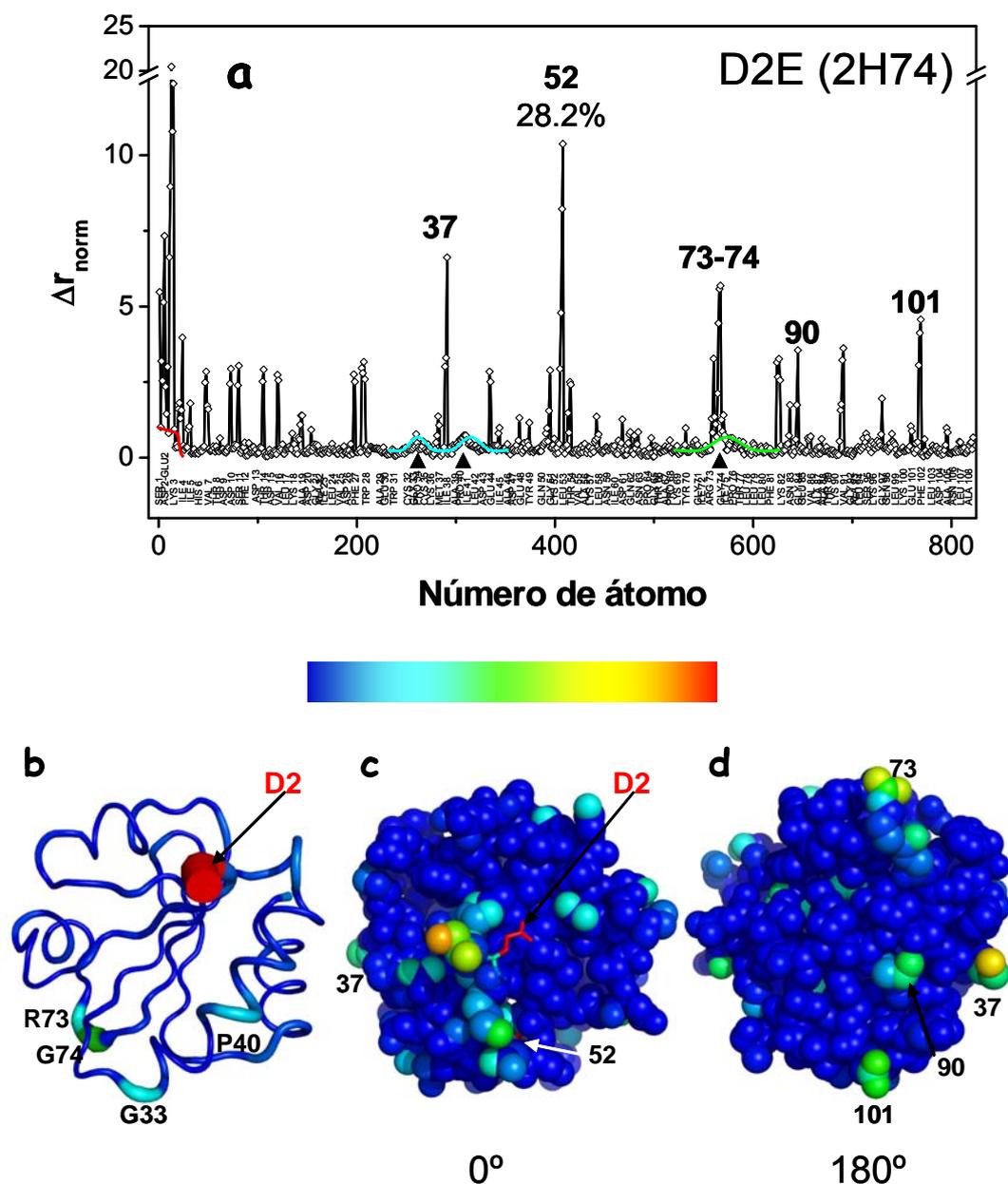


Figura 6.27: Efectos estructurales específicos del mutante D2E. a) Gráfico de barras que muestra el desplazamiento atómico de la estructura de D2E relativo a tiorredoxina *WT*. b) Representación colorimétrica de la magnitud del desplazamiento atómico de la cadena principal de la estructura del mutante D2E relativo a *WT*. Las zonas de mayor desplazamiento atómico se marcan también en el gráfico de barras (a). c) y d) Representación colorimétrica de la magnitud de desplazamiento de los átomos en la estructura del mutante D2E relativo a *WT*. Δr_{norm} es el valor del desplazamiento atómico pesado por los errores propagados en los centroides (ecuaciones 6.7-10).

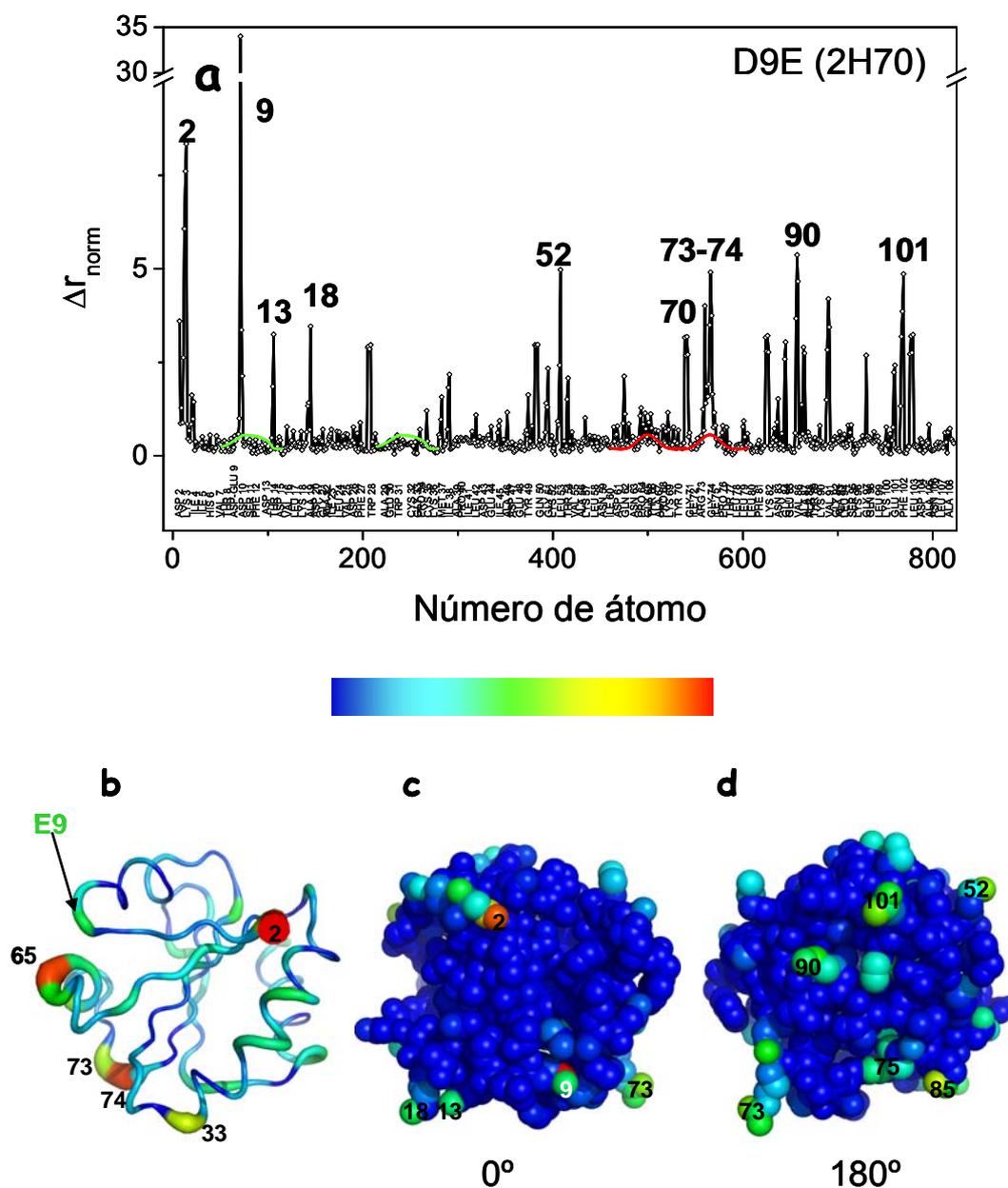


Figura 6.28: Efectos estructurales específicos del mutante D9E. a) Gráfico de barras que muestra el desplazamiento atómico de la estructura de D9E relativo a tiorredoxina *WT*. b) Representación colorimétrica de la magnitud del desplazamiento atómico de la cadena principal de la estructura del mutante D9E relativo a *WT*. Las zonas de mayor desplazamiento atómico se marcan también en el gráfico de barras (a). c) y d) Representación colorimétrica de la magnitud de desplazamiento de los átomos en la estructura del mutante D9E relativo a *WT*. Δr_{norm} es el valor del desplazamiento atómico pesado por los errores propagados en los centroides (ecuaciones 6.7-10).

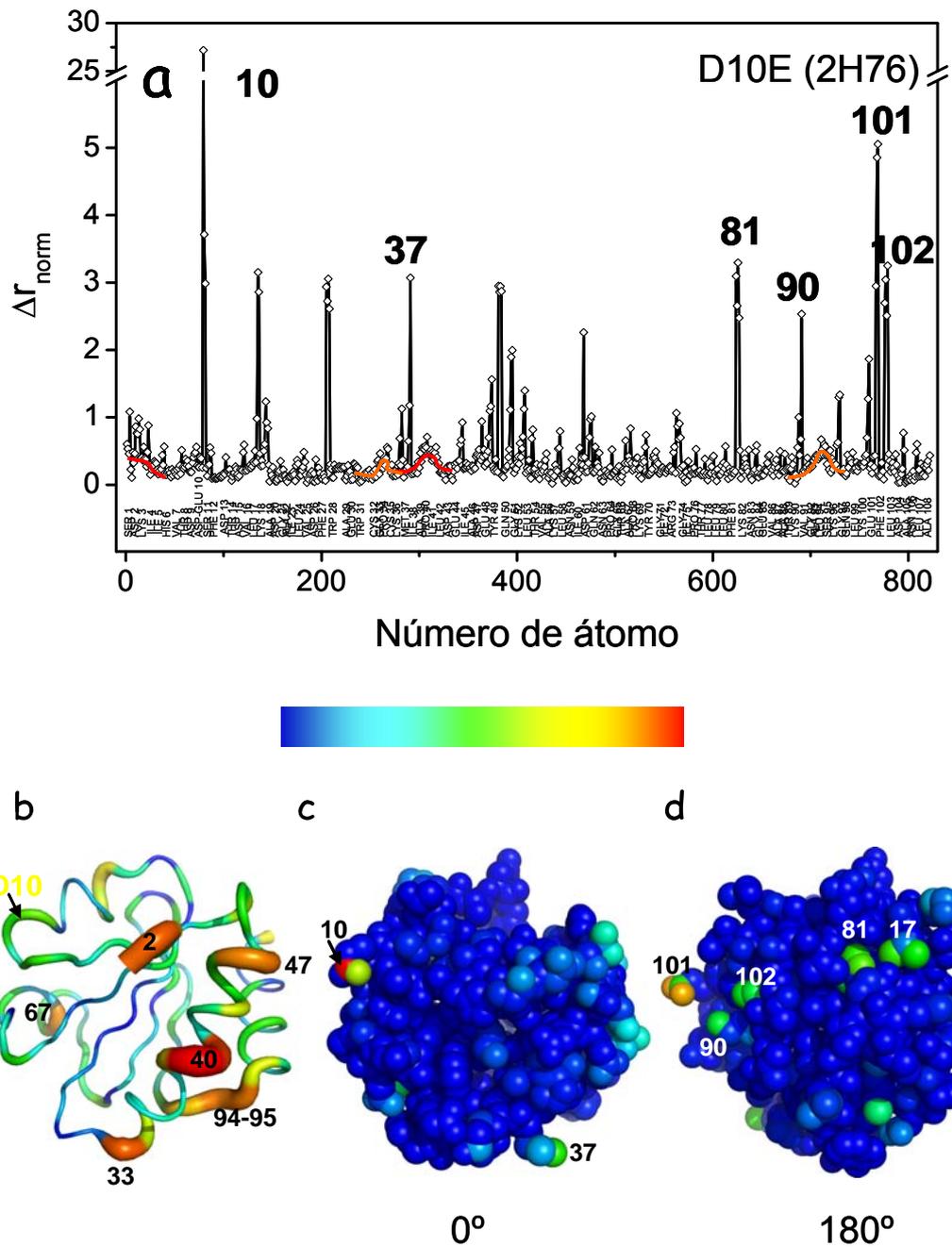


Figura 6.29: Efectos estructurales específicos del mutante D9E. a) Gráfico de barras que muestra el desplazamiento atómico de la estructura de D9E relativo a tiorredoxina *WT*. b) Representación colorimétrica de la magnitud del desplazamiento atómico de la cadena principal de la estructura del mutante D9E relativo a *WT*. Las zonas de mayor desplazamiento atómico se marcan también en el gráfico de barras (a). c) y d) Representación colorimétrica de la magnitud de desplazamiento de los átomos en la estructura del mutante D9E relativo a *WT*. Δr_{norm} es el valor del desplazamiento atómico pesado por los errores propagados en los centroides (ecuaciones 6.7-10).

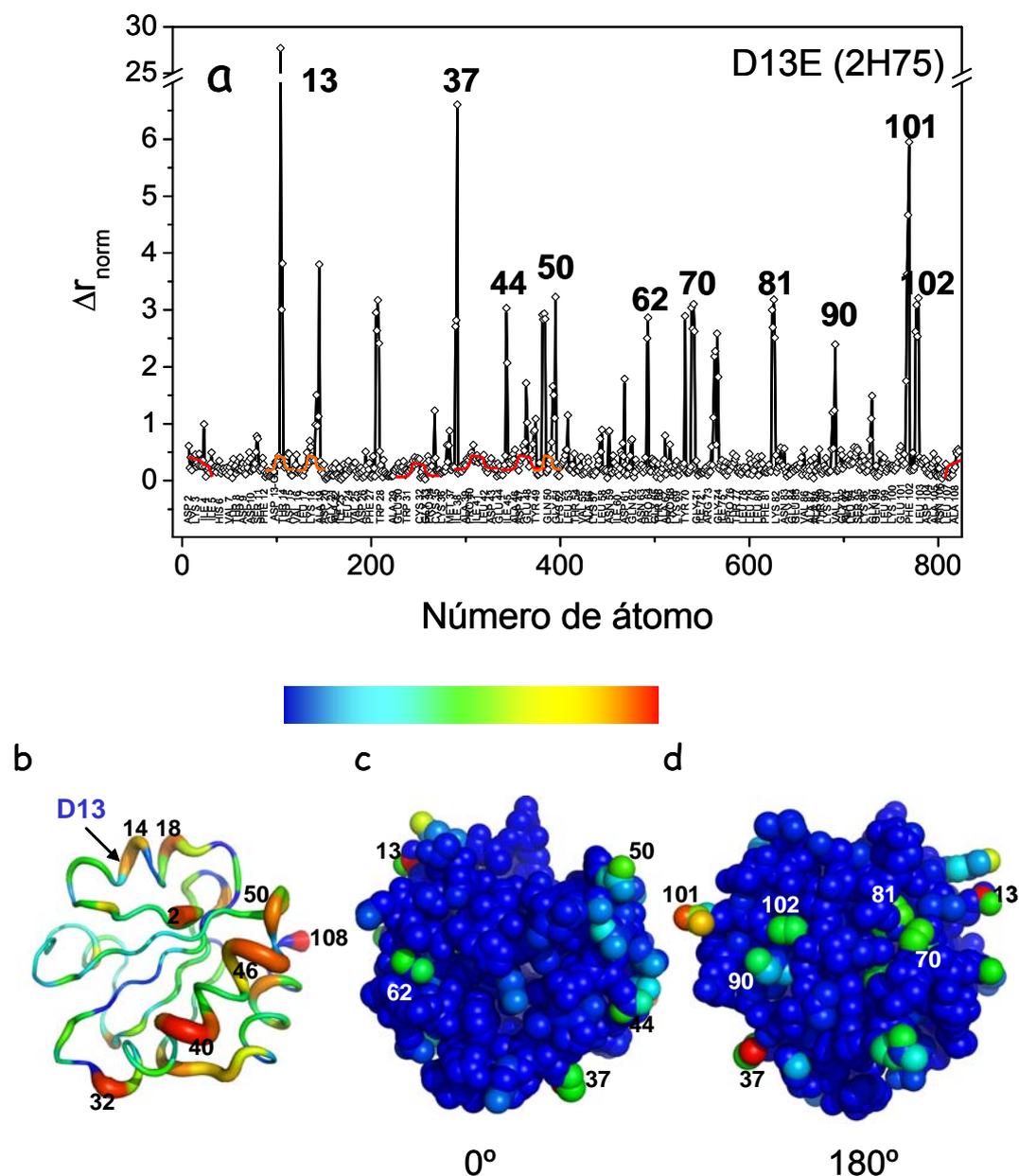


Figura 6.30: Efectos estructurales específicos del mutante D13E. a) Gráfico de barras que muestra el desplazamiento atómico de la estructura de D13E relativo a tiorredoxina *WT*. b) Representación colorimétrica de la magnitud del desplazamiento atómico de la cadena principal de la estructura del mutante D13E relativo a *WT*. Las zonas de mayor desplazamiento atómico se marcan también en el gráfico de barras (a). c) y d) Representación colorimétrica de la magnitud de desplazamiento de los átomos en la estructura del mutante D13E relativo a *WT*. Δr_{norm} es el valor del desplazamiento atómico pesado por los errores propagados en los centroides (ecuaciones 6.7-10).

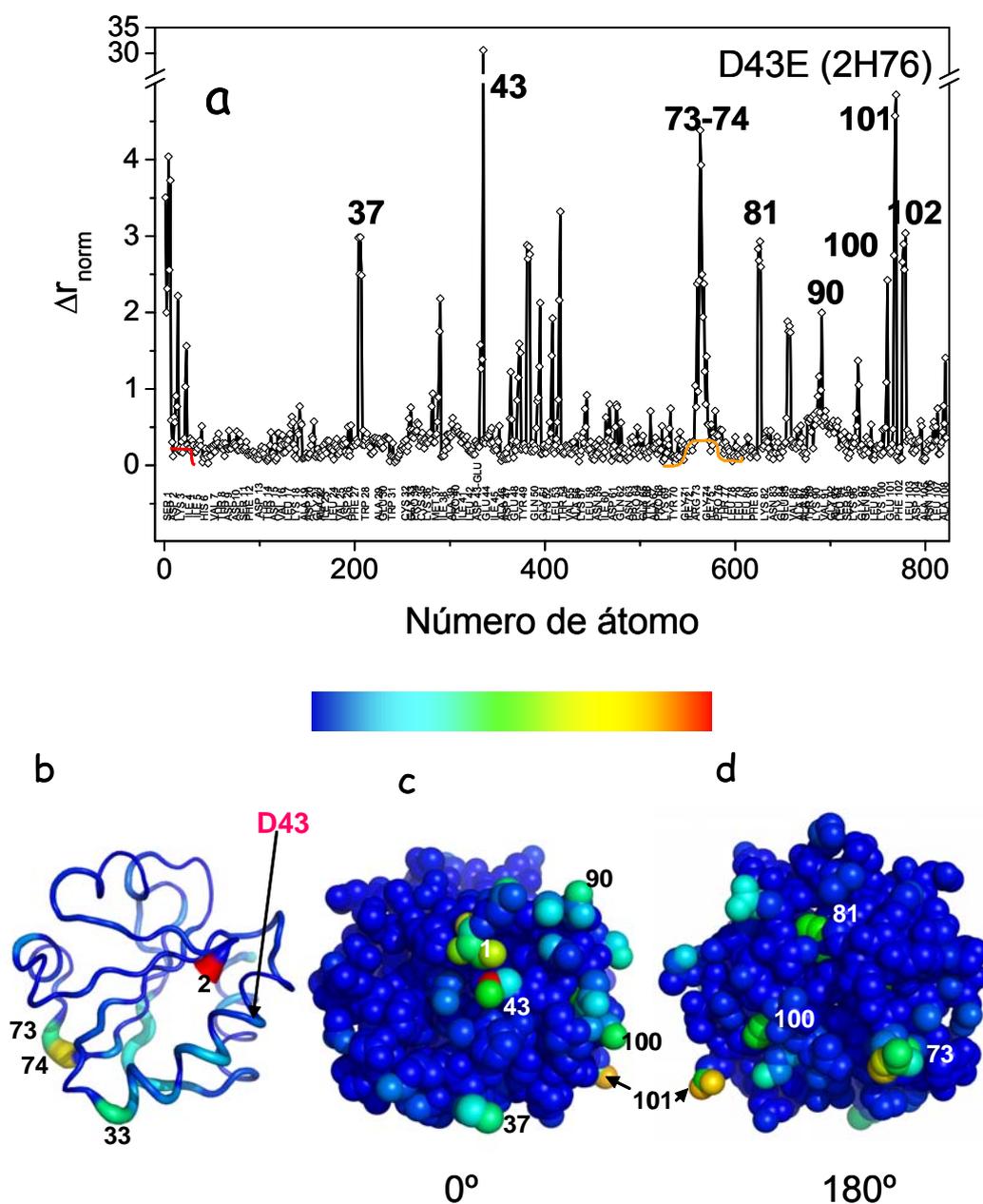


Figura 6.31: Efectos estructurales específicos del mutante D43E. a) Gráfico de barras que muestra el desplazamiento atómico de la estructura de D43E relativo a tiorredoxina *WT*. b) Representación colorimétrica de la magnitud del desplazamiento atómico de la cadena principal de la estructura del mutante D43E relativo a *WT*. Las zonas de mayor desplazamiento atómico se marcan también en el gráfico de barras (a). c) y d) Representación colorimétrica de la magnitud de desplazamiento de los átomos en la estructura del mutante D13E relativo a *WT*. Δr_{norm} es el valor del desplazamiento atómico pesado por los errores propagados en los centroides (ecuaciones 6.7-10).

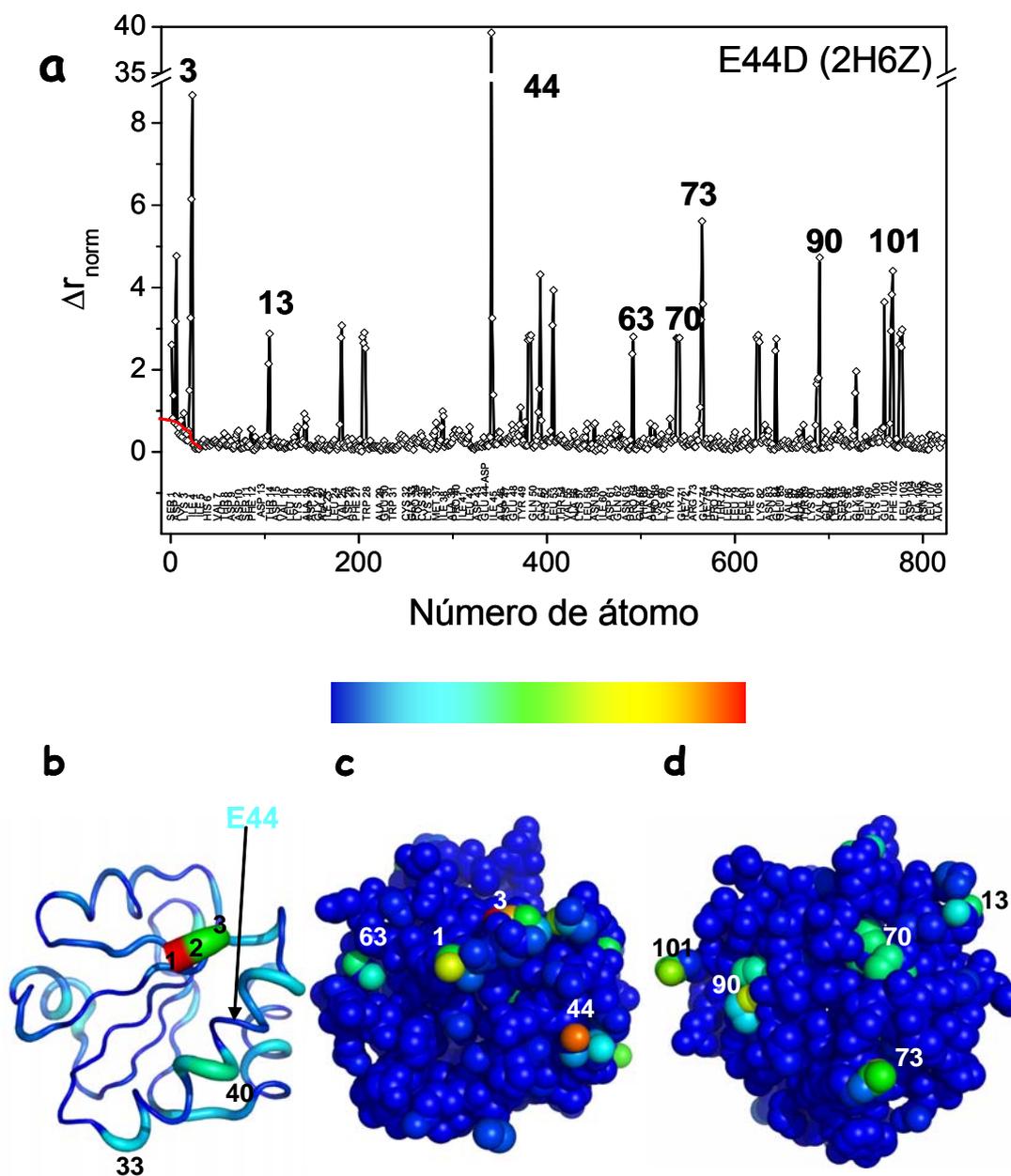


Figura 6.32: Efectos estructurales específicos del mutante E44D. a) Gráfico de barras que muestra el desplazamiento atómico de la estructura de D13E relativo a tioredoxina *WT*. b) Representación colorimétrica de la magnitud del desplazamiento atómico de la cadena principal de la estructura del mutante E44D relativo a *WT*. Las zonas de mayor desplazamiento atómico se marcan también en el gráfico de barras (a). c) y d) Representación colorimétrica de la magnitud de desplazamiento de los átomos en la estructura del mutante E44D relativo a *WT*. Δr_{norm} es el valor del desplazamiento atómico pesado por los errores propagados en los centroides (ecuaciones 6.7-10).

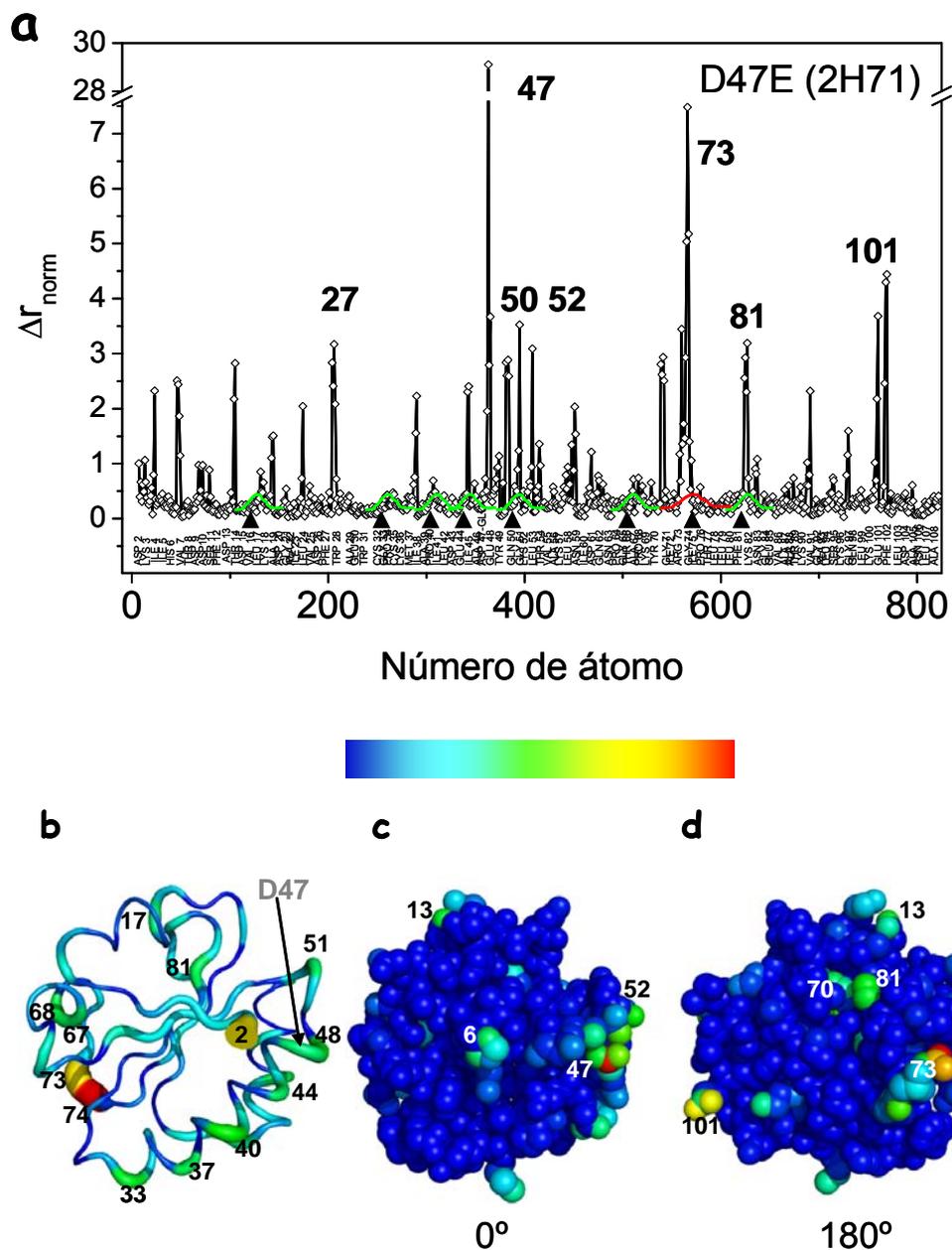


Figura 6.33: Efectos estructurales específicos del mutante D47E. a) Gráfico de barras que muestra el desplazamiento atómico de la estructura de D47E relativo a tiorredoxina *WT*. b) Representación colorimétrica de la magnitud del desplazamiento atómico de la cadena principal de la estructura del mutante D47E relativo a *WT*. Las zonas de mayor desplazamiento atómico se marcan también en el gráfico de barras (a). c) y d) Representación colorimétrica de la magnitud de desplazamiento de los átomos en la estructura del mutante D47E relativo a *WT*. Δr_{norm} es el valor del desplazamiento atómico pesado por los errores propagados en los centroides (ecuaciones 6.7-10).

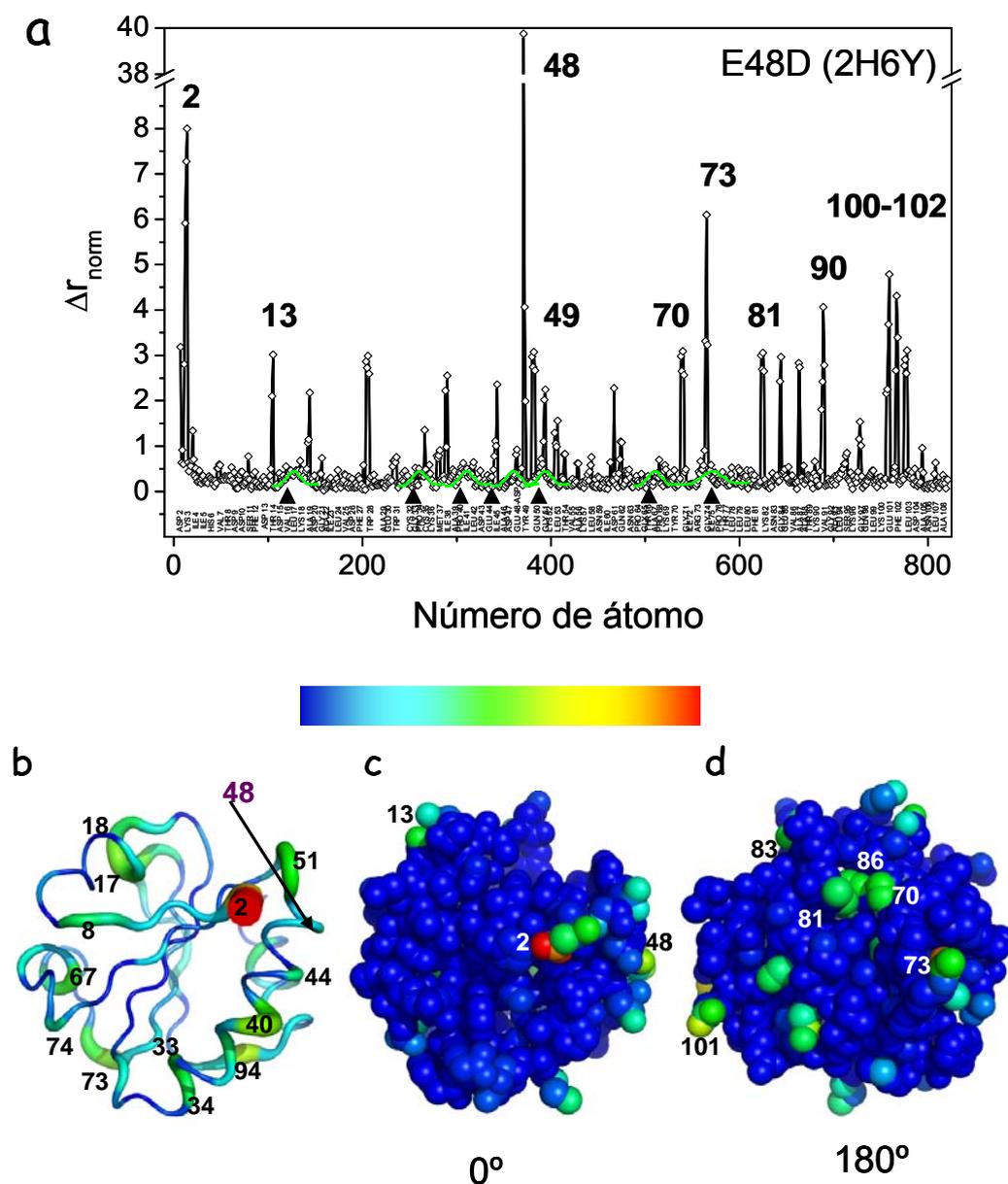


Figura 6.34: Efectos estructurales específicos del mutante E48D. a) Gráfico de barras que muestra el desplazamiento atómico de la estructura de E48D relativo a tioredoxina *WT*. b) Representación colorimétrica de la magnitud del desplazamiento atómico de la cadena principal de la estructura del mutante E48D relativo a *WT*. Las zonas de mayor desplazamiento atómico se marcan también en el gráfico de barras (a). c) y d) Representación colorimétrica de la magnitud de desplazamiento de los átomos en la estructura del mutante E48D relativo a *WT*. Δr_{norm} es el valor del desplazamiento atómico pesado por los errores propagados en los centroides (ecuaciones 6.7-10).

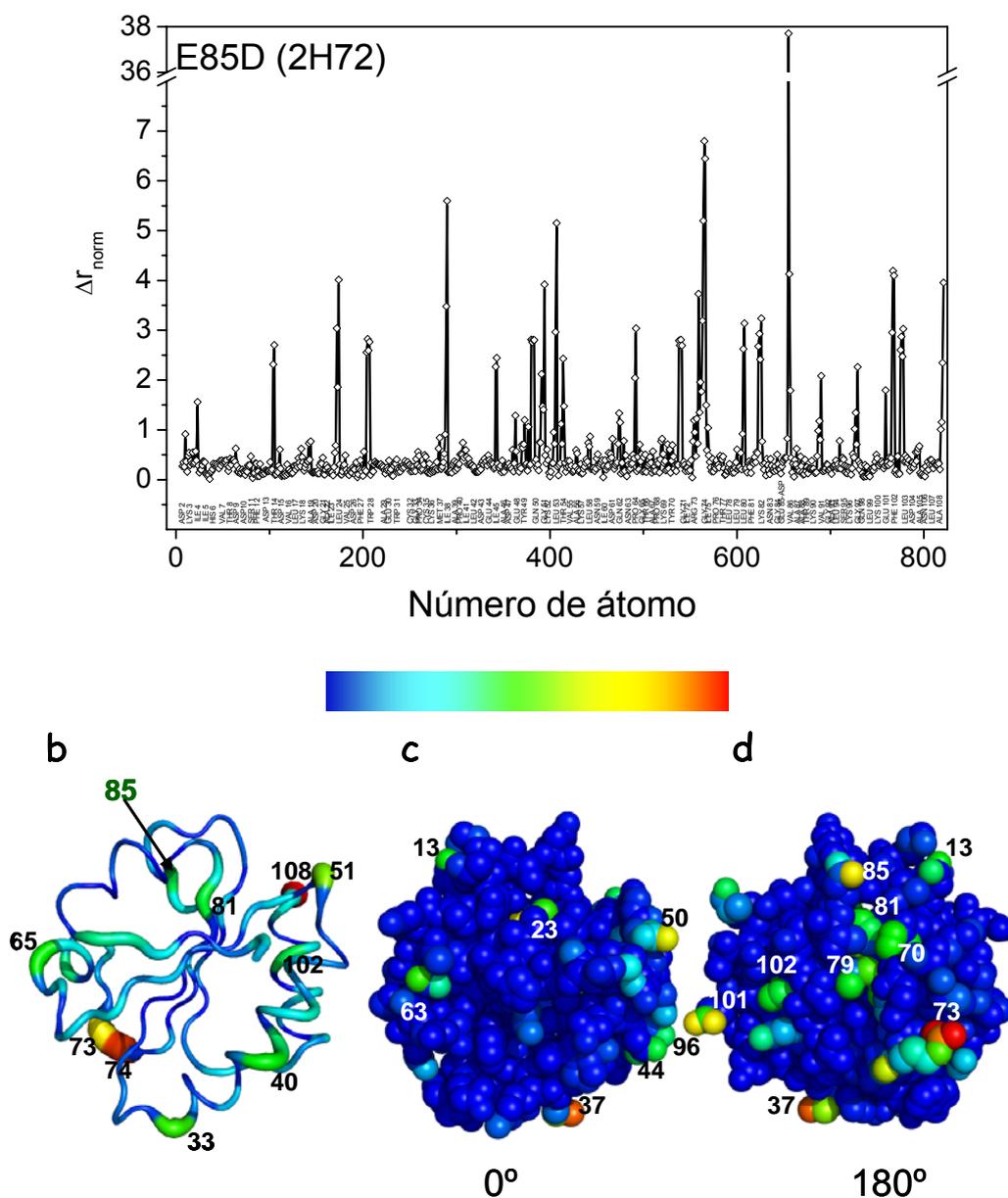


Figura 6.35: Efectos estructurales específicos del mutante E85D. a) Gráfico de barras que muestra el desplazamiento atómico de la estructura de E85D relativo a tiorredoxina *WT*. b) Representación colorimétrica de la magnitud del desplazamiento atómico de la cadena principal de la estructura del mutante E85D relativo a *WT*. Las zonas de mayor desplazamiento atómico se marcan también en el gráfico de barras (a). c) y d) Representación colorimétrica de la magnitud de desplazamiento de los átomos en la estructura del mutante E85D relativo a *WT*. Δr_{norm} es el valor del desplazamiento atómico pesado por los errores propagados en los centroides (ecuaciones 6.7-10).

Tabla 6.6: Zonas de mayor flexibilidad de las estructuras D2E, D9E, D10, D13E, D43E, E44D, D47E, E48D y E85D.

Residuo	1	2	3	4	5	6	7	8	9	10	11	13	13	14
D2E	2			33		40					73-74			
D9E	2			33		40	44	48	65		73-74			
D10E	2			33		40		47	67				94	
D13E	2		14 18	32		40		46	50				94	
D43E	2			33		40								
E44D	1 3			33		40					73-74			
D47D	2		17	33	37	40	44	47	51	67-68	73-74	81		
E48D	2	8	17 18	33	34	40	44		51	67	73-74			
E85D	2			33		40			51	65	73-74	81		102-103

que independientemente de la mutación realizada, existen residuos de tiorredoxina que responden de manera muy flexible a la mutación introducida. Se trata del extremo N-terminal (Asp 2) de tiorredoxina y de la hélice α_2 (residuos Gly 33, Pro 40, Glu 44, y Asp 47-Glu 48). Estos resultados son muy interesantes pues podemos observar cómo la Gly 33, situada entre el puente disulfuro Cys 32-X-X-Cys 35, responde muy flexiblemente a los efectos de la mutación. Estos cambios estructurales, probablemente pueden explicar los cambios en estabilidad causados por la introducción de la mutación.

6.7 CONCLUSIONES

Hemos mostrado cómo una mutación puntual puede inducir la formación de nuevos polimorfos en condiciones similares a las empleadas para la cristalización de la proteína silvestres. Lo cual nos lleva a proponer que, en igualdad de condiciones, es posible inducir la formación de nuevos polimorfos modificando la velocidad y rango de sobresaturación barridos, empleando para ello medios con transporte de masa difusivo. La obtención de polimorfos puede ir

acompañada de cambios estructurales en la proteína y se pueden entender como reflejo de la flexibilidad que presentan en solución.

Un estudio de los factores térmicos de tiorredoxina *WT* y variantes, junto con la observación de los desplazamientos atómicos de las diferentes estructuras, pesados por los factores-B, el número de reflexiones únicas necesarias para difractar el cristal de la proteína correspondiente y el número de átomos en la celdilla unidad, aporta evidencias experimentales que muestran la flexibilidad de tiorredoxina. Razón por la cual la introducción o eliminación de un grupo metileno en tiorredoxina es suficiente para provocar cambios locales que, aunque no afectan al empaquetamiento cristalino, son capaces de producir cambios estructurales a larga distancia. Estas variaciones estructurales pueden explicar los efectos de la mutación sobre la estabilidad de la proteína.

6.8 BIBLIOGRAFÍA

- Bolon, D.N. and S.L. Mayo (2001). *Biochemistry* **40**(34): 10047-53.
- Bowie, J.U., J.F. Reidhaar-Olson, W.A. Lim and R.T. Sauer (1990). *Science* **247**(4948): 1306-10.
- Brown, C.J., S. Takayama, A.M. Campen, P. Vise, T.W. Marshall, C.J. Oldfield, C.J. Williams and A.K. Dunker (2002). *J Mol Evol* **55**(1): 104-10.
- Brünger, A.T., P.D. Adams, G.M. Clore, P. Gros, R.W. Grosse-Kunstleve, J.-S. Jiang, J. Kuszewski, M. Nilges, N.S. Pannu, R.J. Read, L.M. Rice, T. Simonson and G.L. Warren (1998). *Acta Cryst.* **D54**(905-921).
- Doniach, S. and P. Eastman (1999). *Curr Opin Struct Biol* **9**(2): 157-63.
- Dunker, A.K., C.J. Brown and Z. Obradovic (2002a). *Adv Protein Chem* **62**: 25-49.
- Dunker, A.K., C.J. Brown, J.D. Lawson, L.M. Iakoucheva and Z. Obradovic (2002b). *Biochemistry* **41**(21): 6573-82.
- Dyson, H.J. and P.E. Wright (2002). *Curr Opin Struct Biol* **12**(1): 54-60.
- Forsyth, W.R., J.M. Antosiewicz and A.D. Robertson (2002). *Struct Funct Genet* **48**(388-403).
- García-Ruiz, J.M. (2003). *Methods in Enzymology* **368**: 130-154.
- George, D.G., W.C. Baker and L.T. Hunt (1990). *Methods Enzymol* **183**: 333-351.
- Gerstein, M., A.M. Lesk and C. Chothia (1994). *Biochemistry* **33**(22): 6739-49.
- Godoy-Ruiz, R., R. Perez-Jimenez, B. Ibarra-Molero and J.M. Sanchez-Ruiz (2004). *J Mol Biol* **336**(2): 313-8.
- Godoy-Ruiz, R., R. Perez-Jimenez, B. Ibarra-Molero and J.M. Sanchez-Ruiz (2005). *Biophys J* **89**(5): 3320-31.
- Godoy-Ruiz, R., F. Ariza, D. Rodriguez-Larrea, R. Perez-Jimenez, B. Ibarra-Molero and J.M. Sanchez-Ruiz (2006). *J Mol Biol* **In press**.
- Henikoff, S. and T.G. Henikoff (1992). *Proc Natl Acad Sci U S A* **89**: 10915-10919.
- Hermoso Dominguez, J.A. and M. Martínez-Ripoll (2003). Estructura de proteínas. S.A. Editorial Ariel. Barcelona, España.

- Holmgren, A., B.O. Soderberg, H. Eklund and C.I. Branden (1975). *Proc Natl Acad Sci U S A* **72**(6): 2305-9.
- Jain, R.K. and R. Ranganathan (2004). *Proc Natl Acad Sci U S A* **101**(1): 111-6.
- Katti, S.K., D.M. LeMaster and H. Eklund (1990). *J. Mol. Biol.* **212**: 167-184.
- Kumar, S., H.J. Wolfson and R. Nussinov (2001). *IBM J RES & DEV* **45**(3/4): 499-512.
- Lockless, S.W. and R. Ranganathan (1999). *Science* **286**(5438): 295-9.
- McRee, D.E. (1993). New York.
- Murshudov, G.N., A.A. Vagin and E.J. Dodson (1997). *Acta Cryst.* **D53**(240-255).
- Nikkola, M., F.K. Gleason, J.A. Fuchs and H. Eklund (1993). *Biochemistry* **32**(19): 5093-8.
- Obradovic, Z., K. Peng, S. Vucetic, P. Radivojac, C.J. Brown and A.K. Dunker (2003). *Proteins* **53 Suppl 6**: 566-72.
- Parthasarathy, S. and M.R. Murthy (2000). *Protein Eng* **13**(1): 9-13.
- Radivojac, P., Z. Obradovic, C.J. Brown and A.K. Dunker (2002). *Pac Symp Biocomput*: 589-600.
- Radivojac, P., Z. Obradovic, D.K. Smith, G. Zhu, S. Vucetic, C.J. Brown, J.D. Lawson and A.K. Dunker (2004). *Protein Sci* **13**(1): 71-80.
- Ringe, D. and G.A. Petsko (1986). *Methods Enzymol* **131**: 389-433.
- Romero, P., Z. Obradovic, X. Li, E.C. Garner, C.J. Brown and A.K. Dunker (2001). *Proteins* **42**(1): 38-48.
- Rudresh, J., V. Dani, A. Mitra, S. Srivastava, S.P. Sarma, R. Varadarajan and S. Ramakumar (2002). *Protein Eng* **15**: 627.
- Schultz, L.W., P.T. Chivers and R.T. Raines (1999). *Acta Cryst.* **D55**: 1533-1538.
- Sheriff, S., W.A. Hendrickson, R.E. Stenkamp, L.C. Sieker and L.H. Jensen (1985). *Proc Natl Acad Sci U S A* **82**(4): 1104-7.
- Smith, J.L., W.A. Hendrickson, R.B. Honzatko and S. Sheriff (1986). *Biochemistry* **25**(18): 5018-27.
- Stroud, R.M. and E.B. Fauman (1995). *Protein Sci* **4**(11): 2392-404.

Su, A. and S.L. Mayo (1997). *Protein Sci* **6**(8): 1701-7.

Suel, G.M., S.W. Lockless, M.A. Wall and R. Ranganathan (2003). *Nat Struct Biol* **10**(1): 59-69.

Tronrud, D.E. (1996). *J Appl Cryst* **29**: 100-104.

Uversky, V.N. (2002). *Protein Sci* **11**(4): 739-56.

Wright, P.E. and H.J. Dyson (1999). *J Mol Biol* **293**(2): 321-31.

7. RESUMEN Y CONCLUSIONES

RESUMEN Y CONCLUSIONES:

1. Hemos estudiado el efecto de 27 mutaciones conservativas (Glutamato (E) ↔ Aspartato (D) y Valina (V) ↔ Isoleucina(I)) en la estabilidad termodinámica de tiorredoxina de *E. coli*.
2. La mayoría de estas mutaciones son desestabilizantes, indicando que los ambientes de los grupos implicados pueden estar optimizados con un alto grado de especificidad, hasta el punto de discriminar entre D, E, V e I.
3. Las correspondientes interacciones estabilizantes podrían quizás detectarse en alineamientos de secuencias. Ésto podría ser útil en la derivación de información estructural a partir de alineamientos de secuencias.
4. Los resultados anteriores nos llevan a sugerir la posibilidad de racionalizar y quizás predecir efectos de mutaciones en la estabilidad de proteínas en base a “modelos evolutivos”. Por “modelo evolutivo” nos referimos a un modelo para los efectos de la mutación sobre la estabilidad, en el cual se evalúan ciertas características de los ambientes del residuo mutado, desde el punto de vista de cómo esos ambientes han sido seleccionados durante la evolución; concepto opuesto al de “modelo puramente físico”, en el cual esas características se determinarían de algún tipo de análisis energético de la estructura nativa. Hemos demostrado esta posibilidad desarrollando un modelo evolutivo que explica exitosamente el efecto de mutaciones hidrofóbicas en la estabilidad de tiorredoxina de *E coli*.
5. Hemos encontrado una correlación altamente significativa entre los efectos de mutaciones sobre la estabilidad de tiorredoxina y la frecuencia de aparición de residuos en alineamientos de secuencias.
6. Dicha correlación indica que mutaciones neutras o quasi-neutras en posiciones que no están limitadas estructural o funcionalmente se aceptan durante la evolución con probabilidades relacionadas con sus efectos en la estabilidad, de acuerdo a una ecuación similar a la ley de distribución de Boltzmann.
7. Este tipo de resultado implica que la estabilidad de la proteína debe fluctuar en un rango estrecho durante la evolución; y nuestros modelos computacionales simples de la evolución de la estabilidad sugieren un límite inferior a ésta del

orden de unos pocos kJ/mol por debajo de la estabilidad de tiorredoxina silvestre.

8. Hemos sugerido, por tanto, que la correlación estabilidad/alineamientos no refleja en este caso la selección natural de la estabilidad termodinámica por sí misma sino de otro factor que está asociado a ésta. Proponemos que dicho factor es la estabilidad cinética de tiorredoxina *in vivo* dado que: i) la estabilidad cinética está relacionada a una desnaturalización irreversible, ii) la velocidad de la desnaturalización irreversible en un ambiente celular (o en un ambiente extracelular más duro) esta probablemente determinada por la velocidad de desplegamiento, iii) la vida media del desplegamiento cambia de una manera exponencial con la energía libre y, consecuentemente, efectos de energía libre comparativamente pequeños pueden tener consecuencias perjudiciales para la estabilidad cinética.
9. La anterior propuesta está apoyada por un estudio cinético de la forma silvestre y de los 27 mutantes simples de tiorredoxina, basado en el análisis global de representaciones de *Chevron* y perfiles del equilibrio de desplegamiento determinados mediante los ensayos de desplegamiento de doble salto.
10. Este estudio cinético sugiere que uno de los factores que contribuye a la estabilidad cinética de tiorredoxina, es de hecho la optimización de los ambientes de los residuos en el estado N en aquellas regiones que se encuentran desorganizadas en el estado de transición para el desplegamiento.
11. Hemos llevado a cabo un estudio cristalográfico de un número significativo de mutantes con objeto de delinear las bases estructurales de dicha optimización de dichos ambientes.
12. Hemos cristalizado y resuelto la estructura de tiorredoxina silvestre y 9 variantes de ácidos carboxílicos en el mismo grupo espacial. El elevado polimorfismo presente en tiorredoxina y variantes, analizados en este trabajo y descritos en la bibliografía, va acompañado de cambios estructurales como demuestran los estudios de RMSD y factores-B. Esto nos permite calificar a la tiorredoxina como una proteína altamente flexible. La introducción o eliminación de un grupo metileno es suficiente para provocar cambios locales que, aunque no afectan al empaquetamiento cristalino, son capaces de producir

cambios estructurales a larga distancia. Estas variaciones estructurales pueden explicar los efectos de la mutación sobre la estabilidad de la proteína.

13. Los acoplamientos estructurales a larga distancia no parecen mostrar una correlación con los acoplamientos energéticos detectados en los alineamientos de secuencias. Sugerimos que esta falta de correlación es debida a una definición incompleta del modelo utilizado para predecir los acoplamientos estructurales, probablemente debido a que las mutaciones estudiadas se encuentran en regiones expuestas al solvente, con un elevado grado de flexibilidad y que no participan en contactos específicos.

APÉNDICE

PROGRAMA 1

MODELO DE EQUILIBRIO DE DOS ESTADOS PARA DSC.

```
FUNCTION CP(T) =  
CPN(T)+(CPU(T)-  
CPN(T))*(K(T)/(1+K(T)))+(DH^2/(R*T*T))*(K(T)/(1+K(T))^2)
```

```
FUNCTION K(T) = EXP((-DH/R)*(1/T-1/TM))
```

```
FUNCTION CPU(T) = C+D*(T-330)
```

```
FUNCTION CPN(T) = A+B*(T-330)
```

```
R=0.008314
```

```
FIT(A,B,C),I TO M
```

Z=POINTS(Cp, valor inicial:valor final de temperatura en °C:intervalo entre valores consecutivos)

Antes de la instrucción FIT es necesario dar valores a los parámetros ajustables (a,b,c,d,Tm,ΔH) próximos a los óptimos.

PROGRAMA 2

AJUSTE DE LAS CINÉTICAS EN LA DIRECCIÓN DE DESPLEGAMIENTO
A UNA FUNCIÓN MONOEXPONENCIAL.

Necesitamos tener en el mismo directorio `mexponential.m` y
`mexponentialfit.m`.

%mexponential.m

```
function [funct]=mexponential(v,time)
```

```
ampf=v(1);
```

```
amp1=v(2);
```

```
tau1=v(3);
```

```
nonnegt=find((time) >= 0.0);
```

```
funct=(ampf)*ones(size(time));
```

```
funct(nonnegt,:)=(ampf-amp1*exp(-((time(nonnegt)/tau1))));
```

%mexponentialfit.m

```
clear;
```

```
run='name';
```

```
path='D:\THIOREDOXIN\Fluorescence\.....';
```

```
index1=1; %number of trace
```

```
%loading data:
```

```
eval(['load ',path,run,int2str(index1),'.td;']);
```

```
eval(['dataini=',run,int2str(index1),':(:,2);']);
```

```
eval(['time=',run,int2str(index1),':(:,1);']);
```

```

%dataini=dataini.*100; %signal multiplied by 100 to decrease fitting error
%deadtime=.....; %assuming ..... milliseconds
%time=time+deadtime;

[p,q]=size(dataini);
v=[300 250 80];
w=ones(p,1);
%w(1:65,1)=0;
%w(i-45:i-1,1)=0;

[data,p,errorP,record,covariance,corrCoef]=marquardt(time,dataini,w,'mexponenti
al(p,x)',v,[]);

plot(data(:,2),data(:,3),'bo',data(:,2),data(:,5),'r')
title ('unfolding experiment at X M urea and 25 °C')
xlabel ('seconds')
ylabel ('fluorescence intensity')

%v=[data(:,[2:3,5]),sini];
pt=[p,errorP'];

save name_CnM.fit v -ascii;
save name_CnMura.par pt -ascii;

```

PROGRAMA 3

AJUSTE DE LAS CINÉTICAS EN LA DIRECCIÓN DE REPLEGAMIENTO A UNA FUNCIÓN MONOEXPONENCIAL.

Necesitamos tener en el mismo directorio `mexponential.m` y `mexponentialfit.m`.

%mexponential.m

```
function [funct]=mexponential(v,time)
```

```
ampf=v(1);
```

```
amp1=v(2);
```

```
tau1=v(3);
```

```
nonnegt=find((time) >= 0.0);
```

```
funct=(ampf)*ones(size(time));
```

```
funct(nonnegt,:)=(ampf+amp1*exp(-((time(nonnegt)/tau1))));
```

%mexponentialfit.m

```
clear;
```

```
run='name';
```

```
path='D:\THIOREDOXIN\Fluorescence\.....';
```

```
index1=1; %number of trace
```

```
%loading data:
```

```
eval(['load ',path,run,int2str(index1),'.td;']);
```

```
eval(['dataini=',run,int2str(index1),':(:,2);']);
```

```
eval(['time=',run,int2str(index1),':(:,1);']);
```

```

%dataini=dataini.*100; %signal multiplied by 100 to decrease fitting error
%deadtime=.....; %assuming ..... milliseconds
%time=time+deadtime;

[p,q]=size(dataini);
v=[300 250 80];
w=ones(p,1);
%w(1:65,1)=0;
%w(i-45:i-1,1)=0;

[data,p,errorP,record,covariance,corrCoef]=marquardt(time,dataini,w,'mexponenti
al(p,x)',v,[]);

plot(data(:,2),data(:,3),'bo',data(:,2),data(:,5),'r')
title ('refolding experiment at X M urea and 25 °C')
xlabel ('seconds')
ylabel ('fluorescence intensity')

%v=[data(:,[2:3,5]),sini];
pt=[p,errorP'];

save name_CnM.fit v -ascii;
save name_CnMura.par pt -ascii;

```

PROGRAMA 4

Fichero de MatLab necesario para construir una matriz $M_{N \times 2}$ ($\ln k$, C).

%Transformrate.m

*%This makes a matrix $M_{N \times 2}$ ($\ln(\text{rate})$ vs urea concentration). The *.par files are the outcome files of monoexponential fit done with mexponentialfit.m. *.par files contain the fitted parameters: ΔI , I and τ .*

```
load name1_C1Murea.par;
load name2_C2Murea.par;
load name3_C3Murea.par;
:
load namen_CMurea.par;
```

```
tau = [name1_C1Murea(3,1),name2_C2Murea(3,1),name3_C3Murea(3,1),...,namen_CnMurea(3,1)];
```

```
rate=1./tau;
```

```
%tempC=25;
```

```
%temp=tempC+273.15;
```

```
ureaconcen=[C1, C2, C3, ..., Cn];
```

```
lnrate=log(rate);
```

```
mutantCHEV=[ureaconcen' lnrate'];
```

```
save mutantCHEV.tab mutantCHEV -ASCII -TABS;
```

```
%figure;
```

```
plot(mutantCHEV(:,1),mutantCHEV(:,2),'co'),grid;
```

```
title ('Chevron mutant');
```

```
xlabel ('[Urea]');
```

```
ylabel ('ln(rate)');
```

PROGRAMA 5

ANÁLISIS SEGÚN EL MODELO CINÉTICO DE DOS ESTADOS.
Necesitamos tener en el mismo directorio Chevron.m y Chevronfit.m.

%Chevron.m

% $k_{1/2}$, $C_{1/2}$, $m_{N\neq}$, y $m_{N\neq}$ are the fitting parameters.

```
function [funct,b,m1,m2,U]=chevron(v,C)
```

```
T=298.15;%K
```

```
R=0.008314;%
```

```
b=v(1);
```

```
m1=v(2);
```

```
m2=v(3);
```

```
U=v(4);
```

```
%a=exp(log(km));a=exp(b);
```

```
%log(kf)=b+((m2./(R.*T)).*(C-U));
```

```
%log(ku)=b+((m1./(R.*T)).*(C-U));
```

```
kf=exp(b).*exp((m2./(R.*T)).*(C-U));
```

```
ku=exp(b).*exp((m1./(R.*T)).*(C-U));
```

```
nonnegt=find((C) >= 0.0);
```

```
%funct=(ampf)*ones(size(C));
```

```
%funct(nonnegt,:)=(ampf+amp1*exp(-((time(nonnegt)/tau1))));
```

```
%kaparente=ku+kf;
```

```
funct(nonnegt,:)=log(ku+kf);
```

%Chevronfit.m

```
clear;
run='datarate';
path='D:\THIOREDOXIN\Fluorescence\...\';

index1=2; %number of trace

%loading data:
eval(['load ',path,run,int2str(index1),'.TAB;']);
eval(['dataini=',run,int2str(index1),':(,2);']);
eval(['C=',run,int2str(index1),':(,1);']);

%dataini=dataini.*100; %signal multiplied by 100 to decrease fitting error.

[p,q]=size(dataini);

v=[-7.8 -1.8 3.0 6.5];
v=v';
w=ones(p,1);
%w(1:65,1)=0;
%w(i-45:i-1,1)=0;
[data,p,errorP,record,covariance,corrCoef]=marquardt(C,dataini,w,'chevron(p,x)',
v,[]);

plot(data(:,2),data(:,3),'bo',data(:,2),data(:,5),'b'),grid;
title ('Chevron mutant');
xlabel ('[Urea]');
ylabel ('ln(rate)');
v=[data(:,[2:3,5])];
pt=[p,errorP'];

save name.fit v -ascii;
save name.par pt -ascii;
```

PROGRAMA 6

ANÁLISIS SEGÚN EL MODELO CINÉTICO DE DOS ESTADOS.

M=READ(“nombre del fichero”, número de filas, número de columnas)

FUCTION L(X)=ln(kd(X)+kf(X))

FUCTION kd(X)=a·exp((m1/RT)·(X-C))

FUCTION kf(X)=a·exp((m2/RT)·(X-C))

R=0.008314

T=temperature de trabajo en Kelvin.

FIT(a, m1, m2, C), L to M

Z=POINTS(L, valor inicial para la concentración de urea: valor final: intervfo
de valores consecutivos)

PROGRAMA 7

Fichero de MatLab necesario para construir una matriz $M_{N \times 2}$ (ΔI , C).

Fichero de entrada de Programa 8.

%TransformAmpSalto.m

%This makes a file of 2 columns: AmpTotal vs urea concentration.

```
load name_DsaltoC1.par;
```

```
load name_DsaltoC2.par;
```

```
load name_DsaltoC3.par;
```

```
:
```

```
load name_DsaltoCn.par;
```

```
Amptotal=[abs(name_DsaltoC1(2,1)),abs(name_DsaltoC2(2,1)),abs(name_DsaltoC3(2,1))...  
abs(name_DsaltoCn(2,1))];
```

```
%tempC=25;
```

```
%temp=tempC+273.15;
```

```
ureaconcen=[C1 C2 C3....Cn];
```

```
nameSALT=[ureaconcen' Amptotal'];
```

```
save nameSALT.tab nameSALT -ASCII -TABS;
```

```
%figure;
```

```
plot(nameSALT(:,1),nameSALT(:,2),'bo');
```

```
title ('Unfolding name');
```

```
xlabel ('[Urea]');
```

```
ylabel ('Amp');
```

PROGRAMA 8

ANÁLISIS DE LOS PERFILES DE ΔI VERSUS C SEGÚN EL MODELO DE DOS ESTADOS.

La subrutina de ajuste utilizada en el programa MLAB fue:

M=read(“nombre del fichero”, número de filas, número de columnas)

R=0.008314

T=temperatura en Kelvin

FUNCTION G(X)=-F·(X-H)

FUNCTION K(X)=exp(-G(X)/RT)

FUNCTION H(X)=K(X)/(1+K(X))

FUNCTION I(X)=C·(1-H(X))

FIT(A,B,C),I TO M

Z=POINTS(I, valor inicial para la concentración de urea:valor final:intervalo entre valores consecutivos)

Antes de la instrucción FIT es necesario dar valores a los parámetros ajustables (ΔI_0 , $C_{1/2}$ y $m_{1/2}$) próximos a los óptimos.

Los errores asociados a cada parámetro son los elementos diagonales de la matriz de covarianza. Representan un límite inferior de manera que probablemente la incertidumbre asociada a dichos parámetros sea algo mayor.

PROGRAMA 9:

Scriptp.m

```
%What I have is a matrix A=[26,2] corresponding to my experimental data.  
%I want to calculate the correlation coefficient and statistical significance of my data.  
clear;  
load DATAGAMMAP.tab  
[r,p]=corrcoef(DATAGAMMAP)  
[i,j]=find(p<0.05)  
[i,j]
```

PROGRAMA 10:

Kmeans.m

```
%  
clear;  
  
load trx25.vct;  
J=trx25';  
J=squareform(J);  
IDX=kmeans(J,2,'replicate',10,'display','iter');  
silhouette(J,IDX);  
IDX2=IDX';  
csvwrite('D:\THIOREDOXIN\Evolution_Hydrophobic Environments\Biophysical  
Journal\Reviewers\IDX2',IDX2);
```

Appendix II. Row numbers of occurrences of the 20 aminoacids at several positions considered in this work in the sequences given in the alignment of Appendix I with similarity with the query higher than 0.25. Bold numbers indicate the residue present in WT *E. coli* thioredoxin.

RES. POS.	G	A	V	L	I	S	T	C	M	P	D	N	E	Q	K	R	H	F	Y	W
4	-	-	122	13	85	1	2	-	8	3	-	1	1	-	-	-	-	-	4	-
5	-	5	87	70	85	-	11	-	3	1	-	-	1	3	20	2	2	11	1	-
16	-	34	209	55	59	1	35	-	6	1	-	-	2	1	4	-	-	11	2	-
23	-	20	250	32	59	7	17	2	11	-	-	-	-	-	1	1	-	1	-	44
25	-	15	333	29	61	2	3	4	1	-	-	-	-	-	1	-	-	19	1	-
38	2	5	24	175	144	-	5	-	57	-	1	-	4	18	-	-	2	50	3	-
41	-	29	90	24	135	9	41	1	8	-	9	1	77	6	13	9	4	19	4	4
45	-	86	68	197	73	3	4	2	24	-	1	-	-	-	-	-	-	19	5	4
55	-	10	103	130	91	-	4	5	3	-	-	-	2	0	1	0	2	119	4	5
60	4	72	154	9	73	8	46	69	4	-	5	1	0	0	3	11	5	4	1	0
72	-	5	226	18	237	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-
75	-	1	35	39	181	-	37	-	49	-	-	-	1	2	1	-	-	55	85	1
86	4	27	99	46	32	4	10	-	19	39	3	2	45	28	47	8	-	5	2	-
91	4	29	148	14	36	32	13	1	11	14	11	17	13	19	7	6	9	4	3	-