BJ www.biochemj.org

BJ Structure

Biochem. J. (2010) **429**, 243–249 (Printed in Great Britain) doi:10.1042/BJ20100386

online data

SUPPLEMENTARY ONLINE DATA

Role of conservative mutations in protein multi-property adaptation

David RODRIGUEZ-LARREA*, Raul PEREZ-JIMENEZ†, Inmaculada SANCHEZ-ROMERO*, Asuncion DELGADO-DELGADO*, Julio M. FERNANDEZ† and Jose M. SANCHEZ-RUIZ*[1]

*Departamento de Quimica Fisica, Facultad de Ciencias, Universidad de Granada, 18071-Granada, Spain, and †Department of Biological Sciences, Columbia University, New York, NY 10027, U.S.A.

EXPERIMENTAL

Design of the combinatorial library from sequence alignments

An alignment containing 100 sequences belonging to proteins from proteobacteria was obtained as described in the main text. Using information from [1] and [2], we examined the temperature range for growth and the optimal growth temperature for the 100 proteobacteria from which sequences are included in the alignment. All these micro-organisms are mesophiles, with only the exception of two psychrophiles. From the mesophiles, around 40 % have an optimum growth temperature of 35–37 °C, whereas the remaining grow better around 25–28 °C. None of these bacteria are able to grow at temperatures higher than 45 °C and around 25 % of them are also psychrotolerant.

For each position in the *E. coli* thioredoxin sequence, a ratio of non-wild-type to wild-type frequencies in the alignment was calculated (Figure 1A in the main text). The A22P/I23V/P68A V3 variant was used as background for a combinatorial library of thioredoxin genes which includes (besides A22P, I23V and P68A) all possible combinations of the eight next-ranked mutations (D10A, D47A, Q50A, Q62A, Y70F, G74S, E85Q and A87V), making a total of 256 different variants. Interestingly, the positions involved in the library appear spread over the structure of the protein and, in fact, many of them are not close to the active-site disulfide bridge (Figure S1).

One important issue is to what extent the frequencies of residue occurrence used in library design are robust, i.e. to what extent they depend on the alignment used. To explore this issue we have repeated the same calculation shown in Figure 1(A) in the main text with an extended alignment including all the sequences belonging to bacteria (218 sequences). As shown in Figure S2 there is a good correlation between the $\ln(f_{non-wt}/f_{wt})$ values obtained from the two alignments (bacteria and proteobacteria). In particular, there is agreement between the positions yielding high values of $\ln(f_{non-wt}/f_{wt})$, i.e. the positions used in library design and construction.

Another important issue is related to the possibility of evolutionary correlation or co-evolution between residues at the positions chosen for library constructions. To evaluate this possibility we have used the simple covariance calculation we have described recently [3]. Figure S3 shows a plot of inter-residue distance against covariance calculated from the alignment of a 100 proteobacteria sequences used in the present work. The 5778 values shown (corresponding to the 5778 position pairs in the thioredoxin molecule) reveal a wide range of covariance values. However, covariance values for the 28 pairs involving the positions employed in library construction (10, 47, 50, 62, 70, 74, 85 and 87) are comparatively low and, for the most part, cluster around zero. This lack of evolutionary coupling is a relevant result that further supports that trade-off effects are not important for the conservative mutations chosen.

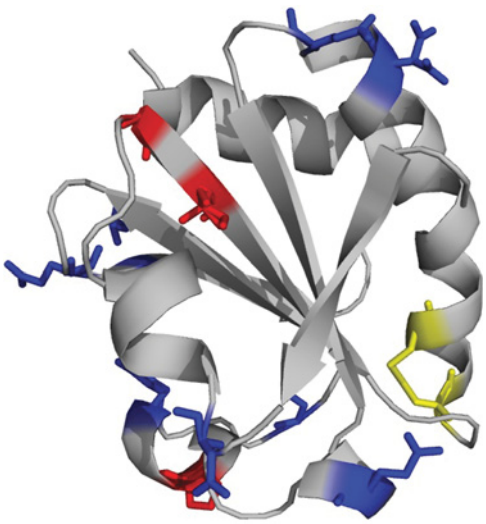

Figure S1 Positions of the *E. coli* wild-type thioredoxin molecule involved in the design of the combinatorial library

The background positions of the library (22, 23 and 68) are shown in red and the eight positions involved in the combination (10, 47, 50, 62, 70, 74, 85 and 87) are shown in blue. The cysteine residues of the active-site disulfide bridge are shown in yellow.

Table S1 Parameters derived from the fitting of a two-state kinetic model (eqns S1–S3) to the chevron plot data for wild-type thioredoxin and trx*

See Figure 4(B) in the main text for chevron plots of folding–unfolding rate. The variant codes are as given in Figure S4.

Parameter	Wild-type thioredoxin	trx*
$C_{1/2}$ (M)	1.95 ± 0.10	4.54 ± 0.06
$\ln(k_{1/2}/\text{min}^{-1})$	−6.74 ± 0.14	−8.45 ± 0.10
m_U (kJ · mol^{-1} · M^{-1})	8.23 ± 0.79	8.55 ± 0.42
m_F (kJ · mol^{-1} · M^{-1})	−4.19 (fixed)	−4.19 ± 0.42

Variant library generation

The combinatorial library of thioredoxin variant sequences was constructed by using gene-assembly mutagenesis [4]. The assembled sequences were amplified by PCR. After restriction digestion the PCR products were ligated into the NdeI and XhoI sites of the pET30a(+) vector (Novagen). The products of the ligation reaction were then transformed into *E. coli* DH10B cells. Plasmid sequencing in a significant number of colonies showed that more than 50 % of the sequences actually contained thioredoxin genes. The genes also encoded for a His_6 tag and a thrombin recognition sequence at the N-terminal end. Different variants (23) were randomly selected for protein

[1] To whom correspondence should be addressed (email sanchezr@ugr.es).

Table S2 Values of the denaturation temperature and disulfide reduction rates at high force and low force for thioredoxin variants

The variant codes are as given in Figure S4. Errors associated with denaturation temperature values are estimated to be below 1 °C. Reduction rates were determined from fittings of single experiments to the sums of approx. 40 length–time recordings. Associated standard errors were obtained by boot strapping.

Variant code	Denaturation temperature (°C)	Disuldife reduction rate (s^{-1}) at low force (75 pN)	Disulfide reduction rate (s^{-1}) at high force (500 pN)
3	102.7	2.25 ± 0.21	0.74 ± 0.08
5	98.7	2.03 ± 0.33	0.34 ± 0.03
8	97.6	3.1 ± 0.22	0.51 ± 0.03
9	98.9	2.58 ± 0.26	0.53 ± 0.06
10	96.7	3.08 ± 0.45	0.48 ± 0.1
11	103.7	1.68 ± 0.25	0.36 ± 0.05
15	105.2	1.74 ± 0.34	0.61 ± 0.06
18	99.8	2.45 ± 0.35	0.34 ± 0.04
19	99.7	3.86 ± 0.49	0.61 ± 0.08
21	97.2	2.34 ± 0.21	0.38 ± 0.04
22	100.6	3.01 ± 0.34	0.53 ± 0.07
23	99.2	1.36 ± 0.20	0.36 ± 0.04
26	99.5	2.55 ± 0.21	0.44 ± 0.06
33	97.9	2.99 ± 0.22	0.54 ± 0.03
36	103.8	2.51 ± 0.31	0.39 ± 0.04
39	103.0	1.75 ± 0.25	0.66 ± 0.07
41	103.8	1.79 ± 0.26	0.51 ± 0.07
43	101.9	3.57 ± 0.30	0.71 ± 0.06
48	101.2	2.72 ± 0.42	0.55 ± 0.09
49	98.5	2.67 ± 0.23	0.58 ± 0.08
53	99.5	1.93 ± 0.24	0.36 ± 0.05
56	103.8	1.73 ± 0.22	0.42 ± 0.08

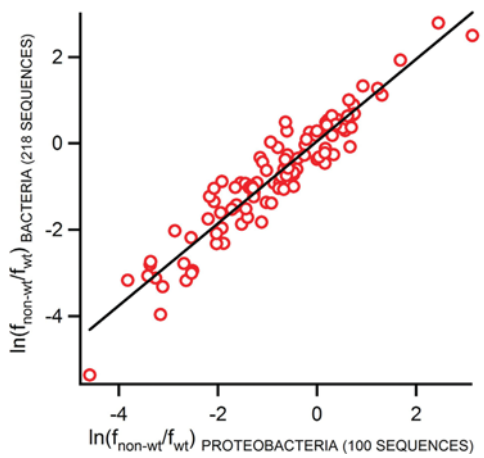

Figure S2 Correlation between the non-wild-type to wild-type frequency ratios derived from the analysis of alignments

The alignments, include 100 proteobacteria sequences (*x*-axis) and 218 bacteria sequences (*y*-axis). Both alignments were derived from a BLAST search of the UniProt/TrEMBL database using the *E. coli* thioredoxin sequence as a query.

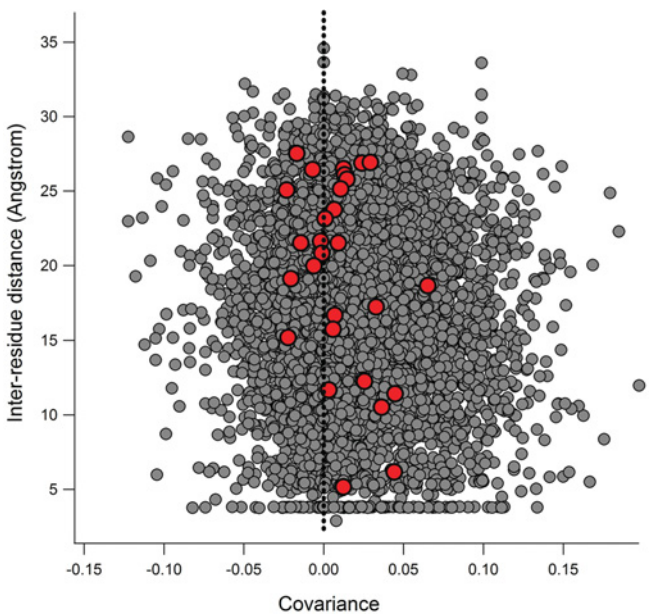

Figure S3 Plot of the inter-residue distance compared with covariance for thioredoxin

The inter-residue distance was calculated using α-carbons. Covariance values are calculated from the alignment of the 100 sequences from proteobacteria using the procedure we have described recently [3]. Data points in grey correspond to all the 5778 residue pairs in the thioredoxin molecule. Data points highlighted in red correspond to the 28 residue pairs involving the positions used in combinatorial library construction.

purification and characterization. Their sequences are given in Figure S4.

Purification of proteins from the combinatorial library

Overexpressing BL21 (DE3) cells were transformed with the amplified library. Isolated colonies were grown in LB (Luria–Bertani) medium in the presence of kanamycin and induced for protein expression with IPTG (isopropyl β-D-thiogalactoside). The His_6-tagged thioredoxin variants were purified as described below. Cell pellets were resuspended in binding buffer (20 mM sodium phosphate, 500 mM NaCl and 20 mM imidazole, pH 7.4), and sonicated (4 cycles of 45 s). Lysates were centrifuged and supernatants were applied to an affinity His-GraviTrap (GE Healthcare) column. The column was washed with excess of

VARIANT CODE	D10A	D47A	Q50A	Q62E	Y70F	G74S	E85Q	A87V
3			X		X		X	X
5	X		X				X	
8					X		X	
9	X	X			X	X	X	
10	X							
11		X	X	X	X	X		X
15	X		X		X	X	X	X
18		X		X			X	
19				X	X		X	
21		X						
22	X	X	X	X	X	X		
23		X			X	X		
26			X		X		X	
32	X		X		X		X	X
33	X				X		X	
36	X		X	X			X	X
39	X				X	X		X
41		X	X			X		X
43	X		X	X	X			X
48				X	X			X
49							X	
53	X	X			X			
56		X	X			X	X	X

Figure S4 Sequences of the 23 different variants characterized in the present study

The 23 different variants were randomly selected. X indicates that the mutation occurs in the variant. In addition, all the variants contain the A22P, I23V and P68A mutations (i.e. the V3 background).

binding buffer and the protein was eluted by applying 3 ml of elution buffer (20 mM sodium phosphate, 500 mM NaCl and 500 mM imidazole, pH 7.4). The buffer was then changed to 5 mM Hepes, pH 7 (the buffer in which all experiments were conducted) by using a Fast Desalting column (GE Healthcare). Finally, for the consensus library variants, protein solutions were heated to 75 °C for 2 h (heating at 75 °C does not affect our consensus-library thioredoxin variants as they show very high thermostability) and any aggregated material was removed by centrifugation (19000 **g** for 5 min). SDS/PAGE (15 % gels) was used to verify the purity of the proteins. Protein concentration was determined spectrophotometrically at 280 nm using a molar absorption coefficient of 14105 $M^{-1} \cdot cm^{-1}$. For those variants with different aromatic residue composition than wild-type, an appropriate absorption coefficient was calculated.

Design and purification of the trx* variant

As part of a preliminary characterization, the bulk-solvent oxidoreductase activity for the 23 library variants described in Figure S4 was determined using an insulin aggregation assay [5]. The trx* variant is identical with one of the highest stability, highest activity variants of Figure S4, except for the absence of the His_6 tag and the elimination of a mutation expected to decrease activity according to an analysis based on an independent-mutation model.

Purification of wild-type thioredoxin and the trx* variant

The His_6-tagged version of wild-type thioredoxin was obtained as explained above for the library variants. In addition, wild-type thioredoxin and trx* were also prepared without a His_6 tag as described previously [6,7].

Folding–unfolding kinetics

Folding–unfolding kinetics were studied at 25 °C by following the time-dependence of the protein fluorescence emission at 360 nm after suitable guanidine concentration jumps in a manner similar to that described previously [6,8]. Apparent folding–unfolding rate constants (k) were calculated by fitting a first-order rate equation ($I = I_\infty - \Delta I e^{-kt}$) to the experimental profiles of fluorescence intensity against time. These fits were excellent. Chevron plots (including folding and unfolding branches) were determined for trx* and wild-type thioredoxin from *E. coli*. Significant roll-over was observed in the folding branches, perhaps due to proline isomerization effects. Interestingly, there is agreement between the roll-over regions of wild-type thioredoxin and trx* (Figure 4B in the main paper). Chevron plots were fitted with a two-state kinetic model as described previously [6,8]:

$$\ln k = \ln (k_U + k_F) \tag{S1}$$

$$\ln k_F = \ln k_{1/2} + \frac{m_F}{RT}(C - C_{1/2}) \tag{S2}$$

$$\ln k_U = \ln k_{1/2} + \frac{m_U}{RT}(C - C_{1/2}) \tag{S3}$$

where k_U and k_F are the unfolding and folding rate constants, m_U and m_F describe the guanidine-concentration (C) effect on the activation free energies for unfolding and folding ($m = -\partial \Delta G / \partial C$), and $k_{1/2}$ is the value of both k_F and k_U at the midpoint guanidine concentration ($C_{1/2}$). For the purpose of these two-state kinetic fits, the roll-over sections were excluded. However, in the case of wild-type thioredoxin, a rather short folding branch is left after roll-over exclusion (Figure 4B in the main paper) and we deemed it necessary to fix the folding m-value for wild-type thioredoxin in the value obtained for trx*. The values derived from the fittings are given in Table S1.

DSC (differential scanning calorimetry)

DSC experiments were performed using a capillary VP-DSC calorimeter from MicroCal in 5 mM Hepes buffer, pH 7, at a scan rate of 2.5 K/min. Calorimetric cells were kept under an excess pressure to avoid degassing during the scan and to allow scans to proceed above 100 °C without boiling. Additional details about the DSC experiments can be found in previous publications [6,8,9]. Reheating runs were routinely performed to check reversibility. Remarkably, trx* denaturation shows a significant degree of reversibility (approx. 70 %), despite the fact that the first scan was terminated at a high temperature (120 °C). In fact, reversibility was found to be somewhat lower for the library

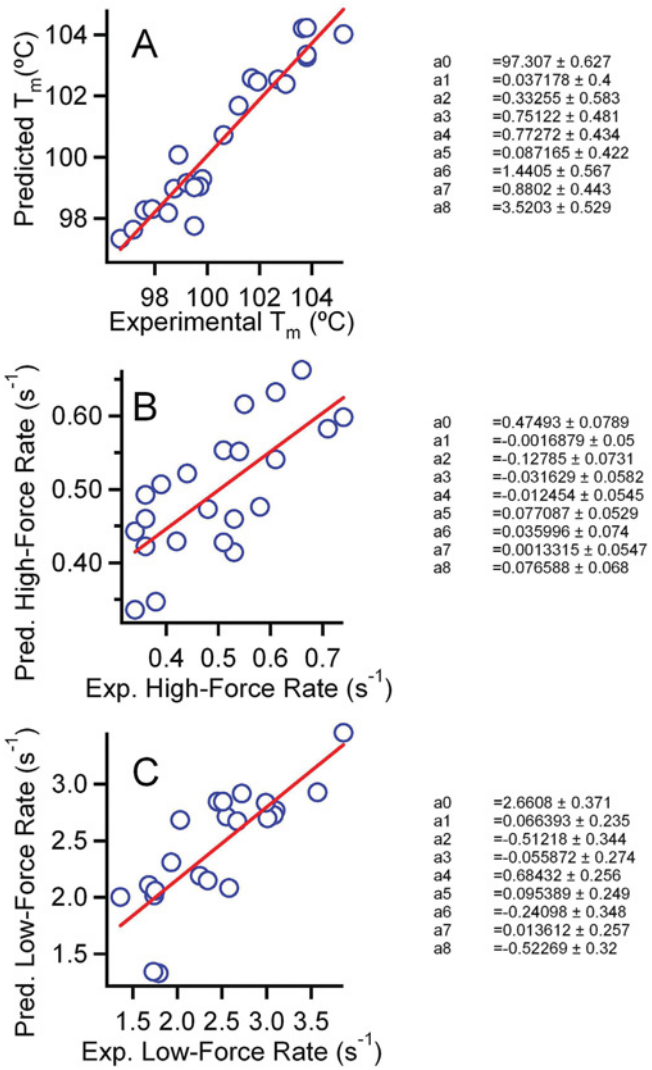

Figure S5 Correlation between experimental values and the predicted values derived from the fitting of an independent-mutation model

The fitting was to (**A**) denaturation temperatures, (**B**) rate of reduction at high force and (**C**) rate of reduction at low force. Alongside the panels we show the values of the coefficients obtained from each fit of the model $P = a + {}^{8}_{i=1}\Delta p(i) \cdot \delta_i$ (eqn S4) where a0 stands for a, a1 for ΔP(D10A), a2 for ΔP(D47A), a3 for ΔP(Q50A), a4 for ΔP(Q62E), a5 for ΔP(Y70F), a6 for ΔP(G74S), a7 for ΔP(E85Q) and a8 for ΔP(A87V).

variants, probably due to the presence of the His_6 tag. DSC profiles for the library variants were mainly used to obtain denaturation temperature values. In the case of wild-type thioredoxin and trx* (prepared without His_6 tags) we carried out two-state fits (Figure 4A in the main paper) which yielded denaturation temperature and enthalpy values of 89.3 °C and 445 kJ/mol (wild-type thioredoxin) and 107.3 °C and 535 kJ/mol (trx*). The difference between the denaturation enthalpy values indicates a denaturation heat capacity of approx. 5 $kJ \cdot K^{-1} \cdot mol^{-1}$, which is consistent with previous determinations [9]. Denaturation temperature values for library variants are given in Table S2.

Single-molecule AFM

Single-molecule AFM was performed according to the procedure described previously [10,11]. Briefly, a custom-built atomic force microscope controlled by an analogue proportional-integral-derivative feedback system [12] was employed. The buffer used in all experiments was 10 mM Hepes, pH 7.2, 150 mM NaCl and 1 mM EDTA, and contained 2 mM NADPH and 50 nM thioredoxin reductase. Single $(I27_{SS})_8$ protein molecules were stretched by first pressing the cantilever on the coverslide at a constant force of 800 pN for 3 s, then retracting to a constant force of 165 pN for 400 ms during the unfolding pulse. The indicated test-pulse force was applied for ~5 s. We summed and normalized the test-pulse portions of numerous recordings that contained only disulfide reduction events to obtain the experimental r value. Further details and a description of the procedure of fitting the model to the experimental r against force profiles can be found in Wiita et al. [10]. The values obtained for the reduction rates at low force (75 pN) and high force (500 pN) are collected in Table S2.

RESULTS AND DISCUSSION

Analysis of stability and activity data in terms of an independent-mutation-effects model

The model is based on eqn (S4):

$$P = a + \sum_{i=1}^{8} \Delta P(i) \cdot \delta_i \qquad \text{(S4)}$$

where P is the value of the property, δ_i is Kronecker's delta that takes values of 1 or 0 depending on whether mutation i is present in the variant, $\Delta P(i)$ is a measure of the sensitivity of the property to mutation i, and a is a constant. Non-linear least-squares fits of the equation to the experimental data for denaturation temperature, high-force rate and low-force rate were performed using a and the $\Delta P(i)$ as fitting parameters. Figure S5 compares experimental values with the predictions of the fits. The independent-mutation-effect model provides a good description of the denaturation temperature data (Figure S5A), supporting the hypothesis that mutation effects on stability are, to a first approximation, additive. On the other hand, fits of the model to the force-dependent activity data, although still visually acceptable (Figures S3B and S3C), are somewhat less satisfactory than that for the T_m values, suggesting the existence of some degree of coupling between mutation effects on catalysis. Accordingly, the values obtained for the individual mutation effects with high-force and low-force rates must be considered as estimates of the average mutation effects. In particular, high significance should not be attached to apparent correlations/anti-correlations observed between these estimated values. In fact, the degree of dependence between the modulations in stability and catalysis is best revealed by a principal component analysis of the library variant data, as shown in Figure 3 of the main text.

REFERENCES

1 Falkow, S., Rosenberg, E., Schleifer, K. H., Stackebrandt, E. and Dworkin, M. (2007) The Prokaryotes: A Handbook on the Biology of Bacteria, 3rd Edn, Springer, New York
2 Garrity, G. M., Brenner, D. J., Krieg, N. R. and Staley, J. T. (2005) The Bergey's Manual of Systematic Bacteriology, 2nd Edn, Springer, New York
3 Perez-Jimenez, R., Wiita, A. P., Rodriguez-Larrea, D., Kosuri, P., Gavira, J. A., Sanchez-Ruiz, J. M. and Fernandez, J. M. (2008) Force-clamp spectroscopy detects residue co-evolution in enzyme catalysis J. Biol. Chem. **283**, 27171–27129
4 Bessette, P. H., Mena, M. A., Nguyen, A. W. and Daugherty, P. S. (2003) Construction of designed protein libraries using gene assembly mutagenesis. In Directed Evolution Library Creation. Methods and Protocols (Arnold, F. H. and Georgiu, G., eds), pp. 29–38, Humana Press, Totowa, New Jersey

5 Holmgren, A. (1979) Thioredoxin catalyzes the reduction of insulin disulfides by dithiotreitol and duhydrolipoamide. J. Biol. Chem. **254**, 9627–9632
6 Pey, A. L., Rodriguez-Larrea, D., Bomke, S., Dammers, S., Godoy-Ruiz, R., Garcia-Mira, M. M. and Sanchez-Ruiz, J. M. (2008) Engineering proteins with tunable thermodynamic and kinetic stabilities. Proteins **71**, 165–174
7 Godoy-Ruiz, R., Perez-Jimenez, R., Ibarra-Molero, B. and Sanchez-Ruiz, J. M. (2005) A stability pattern of protein hydrophobic mutations that reflects evolutionary structural optimization. Biophys. J. **89**, 3320–3331
8 Godoy-Ruiz, R., Ariza, F., Rodríguez-Larrea, D., Perez-Jimenez, R., Ibarra-Molero, B. and Sanchez-Ruiz, J. M. (2006) Natural selection for kinetic stability is a likely origin of correlations between mutational effects on protein energetics and frequencies of amino acid occurrences in sequence alignments. J. Mol. Biol. **362**, 966–978
9 Georgescu, R. E., Garcia-Mira, M. M., Tasayco, M. L. and Sanchez-Ruiz, J. M. (2001) Heat capacity analysis of oxidized Escherichia coli thioredoxin fragments (1–73, 74–108) and their non-covalent complex: evidence for the burial of apolar surface in protein unfolded states. Eur. J. Biochem. **268**, 1477–1485
10 Wiita, A. P., Perez-Jimenez, R., Walter, K. A., Gräter, F., Berne, B. J., Holmgren, A., Sanchez-Ruiz, J. M. and Fernandez, J. M. (2007) Probing the chemistry of thioredoxin catalysis with force. Nature **450**, 124–127
11 Perez-Jimenez, R., Li, J., Kosuri, P., Sanchez-Romero, I., Wiita, A. P., Rodriguez-Larrea, D., Chueca, A., Holmgren, A., Miranda-Vizuete, A., Becker, K. et al. (2009) Diversity of chemical mechanisms in thioredoxin catalysis revealed by single-molecule force spectroscopy. Nat. Struct. Mol. Biol. **16**, 890–896
12 Schlierf, M., Li, H. and Fernandez, J. M. (2004) The unfolding kinetics of ubiquitin captured with single-molecule force-clamp techniques. Proc. Natl. Acad. Sci. U.S.A. **101**, 7299–7304

Received 16 March 2010/21 April 2010; accepted 6 May 2010
Published as BJ Immediate Publication 6 May 2010, doi:10.1042/BJ20100386