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.

Protein physicochemical properties must undergo complex changes during evolution, as a response to modifications in the organism environment, the result of the proteins taking up new roles or because of the need to cope with the evolution of molecular interacting partners. Recent work has emphasized the role of stability and stability–function trade-offs in these protein adaptation processes. In the present study, on the other hand, we report that combinations of a few conservative, high-frequency-of-fixation mutations in the thioredoxin molecule lead to largely independent changes in both stability and the diversity of catalytic mechanisms, as revealed by single-molecule atomic force spectroscopy. Furthermore, the changes found are evolutionarily significant, as they combine typically hyperthermophilic stability enhancements with modulations in function that span the ranges defined by the quite different catalytic patterns of thioredoxins from bacterial and eukaryotic origin. These results suggest that evolutionary protein adaptation may use, in some cases at least, the potential of conservative mutations to originate a multiplicity of evolutionarily allowed mutational paths leading to a variety of protein modulation patterns. In addition the results support the feasibility of using evolutionary information to achieve protein multi-feature optimization, an important biotechnological goal.

Key words: atomic force microscopy (AFM), mutational effect, protein evolution, protein stability, single-molecule analysis.

INTRODUCTION

Proteins must undergo adaptive changes during evolution when, for instance, the environment surrounding the organism is altered or when they are recruited for new roles [1,2]. Although these changes must involve modulation in several protein properties, recent work has mostly emphasized the role of protein stability in molecular evolution [3–6]. A common argument is that most mutations affect stability, whereas only a few are likely to affect function. Furthermore, experimental studies show that most mutations are destabilizing and, consequently, accumulation of a few mutations may compromise the so-called 'protein fitness', due to the concomitant sharp decrease in stability. In addition, it is often assumed that the evolution of biological function is 'limited' or 'constrained' by the destabilizing effects of mutations, as stability and function are generally presumed to trade-off. As a well-known example, mutations conferring the bacterial TEM-1 lactamase with resistance against third-generation antibiotics were found to be destabilizing [7]. Molecular evolution that is determined by trade-off and sign-epistasis [8] effects may be expected to be constrained to follow a few mutational paths (as mutations are bound to occur in a rather specific temporal order to avoid deleterious intermediate combinations). In fact, the possibility of 're-winding' the molecular tape of life has been suggested [9].

In the present study, we explore a point of view that differs from that often found in recent literature and summarized above. We reason that, as adaptations to new situations are common during evolution, many proteins are poised to change their properties efficiently, at least within certain evolutionarily relevant ranges; this should be particularly the case for proteins involved in several molecular tasks, which may have to cope with evolutionary changes in many interaction partners. We thus propose a mechanism for efficient adaptation based on a set of mutations with several features. First, that the mutations in the set have a high frequency of fixation during evolution. That is, they belong to the class of mutations that would be loosely described in various contexts as 'conservative', 'non-disruptive', 'neutral', 'nearly-neutral' or 'quiet'. We use the term conservative in the present paper, but we specifically refer to mutations with non-negative coefficients in substitution matrices (such as the Dayhoff et al. [10] matrix, PAM250) and, consequently, with a high frequency of fixation during evolution. Secondly, that both stability-related and function-related properties are modulated by the mutations in the set, with no strong bias for stability compared with function. Thirdly, that the effects of the several mutations in the set are roughly independent and therefore strong trade-off and sign-epistasis effects do not occur. Fourthly, and as a result, a diversity of mutational pathways leading to a variety of patterns of complex modulation become available to Darwinian evolution thus leading to efficient protein adaptation within some evolutionarily significant range of protein property values.

In the present study, we provide experimental support for the above proposal by showing that combinations of a few conservative mutations in the thioredoxin molecule lead to large, independent and evolutionarily meaningful changes in stability and the complex patterns of catalysis revealed by single-molecule AFM (atomic force microscopy). Actually, thioredoxin provides an excellent model system to investigate the issues we have raised above for the three following reasons.

First, thioredoxins [11] are present in all known organisms and, consequently, these enzymes exist for organisms that thrive in widely different environments. Temperature environment, for instance, is particularly relevant and we may expect thioredoxins from psychrophilic, mesophilic, thermophilic and hyperthermophilic organisms to show quite different thermal stabilities. Secondly, thioredoxins catalyse the reduction of target

Abbreviations used: AFM, atomic force microscopy, PCA, principal component analysis.

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

disulfide bonds, regulating a multitude of cellular processes [11]. A previous proteomic analysis [12] has identified 80 proteins associated with thioredoxin and has shown it is involved in at least 26 cellular processes in *Escherichia coli*; furthermore, additional functions and protein targets for thioredoxin have been reported in eukaryotes [13]. Clearly, a high potential for adaptation is to be expected in the case of thioredoxin, a protein which must cope with evolutionary changes in a multitude of interaction partners and the associated functional roles. Thirdly, recent single-molecule work [14,15] has indeed revealed a diversity of reduction mechanisms in thioredoxin with evolutionarily differentiated patterns of catalysis showing a well-defined correlation with the domains of life. These studies indicate the evolutionary range of variation of the chemistry of thioredoxin catalysis thus providing a clear reference framework for the assessment of the evolutionary significance of mutation-induced modulations.

Single-molecule AFM allows the application of a calibrated force to a disulfide-bonded substrate making it possible to study the catalytic mechanisms with sub-Ångström precision. When the substrate is stretched at low forces, all thioredoxins share a Michaelis–Menten-type reduction mechanism. On the other hand, at high forces there are clear-cut differences between thioredoxins of eukaryotic and bacterial origin. In the former case, disulfide-bond reduction occurs through an SET (single electron transfer) reaction, whereas thioredoxins of bacterial origin exhibit, in addition, a simple nucleophilic substitution (S_N2), thus showing a rate increase at high force that is absent in eukaryotic thioredoxins. It is relevant that the high-force S_N2 mechanism is equivalent to that observed in disulfide-bond reduction with chemicals, such as dithiothreitol, glutathione or cysteine [16,17]. The presence of this simple chemical S_N2 mechanism in bacteria may be related to their ability to survive in extreme conditions [15]. On the other hand, the simple chemical S_N2 mechanism would appear detrimental in eukaryotic cells and thioredoxins from eukaryotic origin seem to have been naturally selected to depress this pathway of reduction. A structural interpretation in terms of the narrowing of the substrate-binding groove in eukaryotic thioredoxins has been advanced [15].

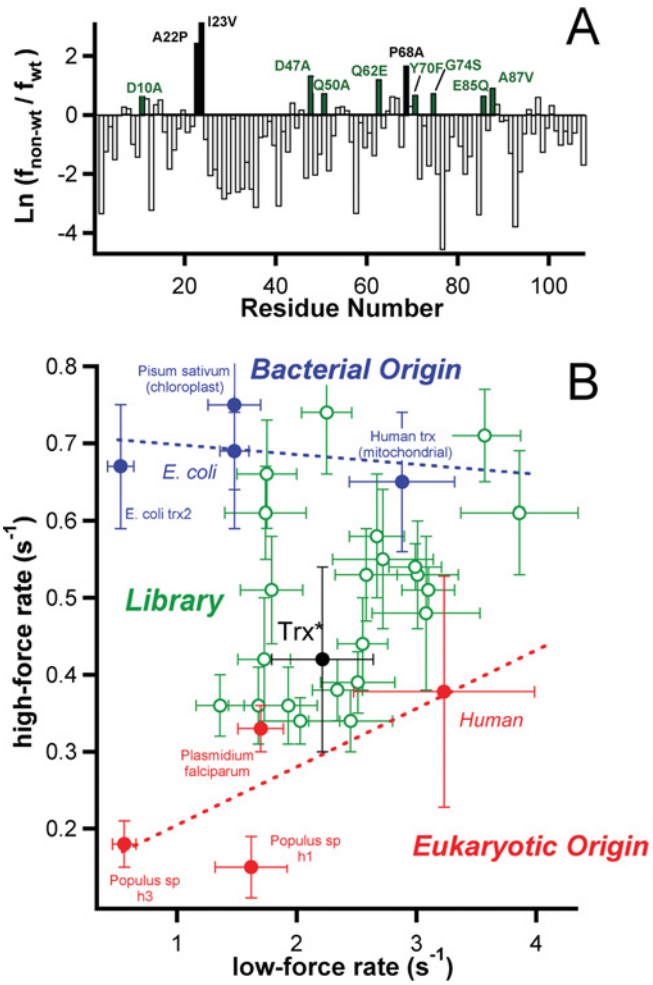

Figure 1 Single-molecule analysis of thioredoxin catalysis using a combinatorial library of conservative mutations

(**A**) Statistical analysis of an alignment of 100 sequences derived from a database search using the *E. coli* sequence as the query. All sequences in the alignment belong to proteins from proteobacteria. Mutations are selected on the basis of the $f_{non\text{-}wt}/f_{wt}$ ratio, where f_{wt} is the number of occurrences in the alignment of the residue present in the query sequence (that of wild-type thioredoxin from *E. coli*) and $f_{non\text{-}wt}$ is the number of occurrences of the highest frequency residue which is not the query-sequence amino acid. The three mutations with the three highest values of the ratio (shown in black) provide the background for a combinatorial library based on the eight mutations with the next highest values of the ratio (shown in green). (**B**) Rates of disulfide-bond reduction by thioredoxin determined by single-molecule AFM. Measurements for 23 variants from the combinatorial library (shown in green) were carried out upon application of low and high mechanical force (75 and 500 pN). The behaviour of wild-type thioredoxins of bacterial and eukaryotic origin (taken from [15]) is shown in blue and red respectively. Trx* is a library variant with an unusually high stability and an eukaryotic signature of catalysis (see the text and Figure 3).

EXPERIMENTAL

Further details are provided in the Supplementary Experimental section at http://www.BiochemJ.org/bj/429/bj4290243add.htm.

The sequence alignment used has been described in detail previously [18,19]. Briefly, BLAST2 (http://blast.wust1.edu) was used to search the sequence of *E. coli* thioredoxin with the PDB code 2TRX as the query in the UniProt database with default search options. Resulting sequences were aligned using the Smith–Waterman algorithm. Those belonging to proteobacteria and with a sequence identity with the query higher than 0.25 were used to calculate the ratios of frequencies of occurrence given in Figure 1(A). Information from [20] and [21] was used to examine the temperature range for growth and the optimal growth temperature for the 100 proteobacteria for the sequences included in the alignment. All these micro-organisms are mesophiles, with the only exception of two psychrophiles. The library of thioredoxin variant sequences was constructed by using gene assembly mutagenesis [22]. Thermal stability was determined by differential scanning calorimetry [18,23]. Single-molecule AFM experiments were carried out as described previously [14,15]. Very briefly, we used as the protein substrate a polypeptide made of eight repeats of the I27 domain of human cardiac titin with engineered cysteine residues. A custom-built AFM controlled by an analogue proportional-integral-derivative feedback system [24] was employed to follow the reduction, catalysed by thioredoxin, of individual disulfide bonds under a stretching force applied to the substrate.

RESULTS

Modulations in stability and evolutionarily significant patterns of thioredoxin catalysis can be achieved by combinations of a few conservative mutations

To investigate how evolutionarily relevant changes in thioredoxin properties may be achieved, we have determined a number of high-frequency-of-fixation mutations using an alignment of 100

sequences derived [18] from a database BLAST search with the *E. coli* thioredoxin sequence as the query. It is very important to note from the outset, that the alignment used contains exclusively sequences of proteins from proteobacteria. All proteins in the alignment are thus of bacterial origin and, furthermore, they happen to belong to mesophilic organisms or (in a few cases) to psychrophilic organisms. That is, thioredoxins from thermophilic or hyperthermophilic organisms are not represented in the alignment used. Despite this, the three highest frequency mutations in the alignment led to a stability enhancement, an increase in the denaturation temperature of approx. 10 °C with respect to the wild-type *E. coli* protein, as shown in previous work [18]. The corresponding triple-mutant variant (referred to as V3) was the starting point of the analysis in the present study. Using V3 as a background, we have constructed a combinatorial library using the next eight highest frequency mutations (Figure 1A). This library contains 256 variants from which 23 were randomly selected and subjected to single-molecule determination of the reduction rates at 75 pN (a low force at which the Michaelis–Menten mechanism is the major contribution) and 500 pN (a high force which reveals the contribution of the simple chemical S_N2 mechanism). The surprising result (Figure 1B) is that the library variants showed a considerable variability in the high-force and low-force rates, actually spanning the range defined by the quite different catalysis patterns of thioredoxins from bacterial and eukaryotic origin. We also assessed the thermodynamic stability (equilibrium denaturation temperature as measured by scanning calorimetry) for the library variants; again a large variability was found although a trend to stability enhancement is clearly apparent (Figures 2A and 2B). This trend may reflect a threshold evolutionary limit for stability, which would lead to statistical preferences for stabilizing mutations in the alignment [25]. Nevertheless, the important point to note is that a small number of mutations determine simultaneous modulations in stability and complex patterns of catalysis. It must be also be emphasized that these modulations are observed in variants extracted from a combinatorial library, a procedure that, in fact, reproduces the situations found during thioredoxin evolution: note that the mutations we use are conservative, high-frequency ones and, as a result, they appear combined very often during evolution. In fact, the distribution in the alignment used combinations of the mutations close to the binomial distribution (Figure 2C).

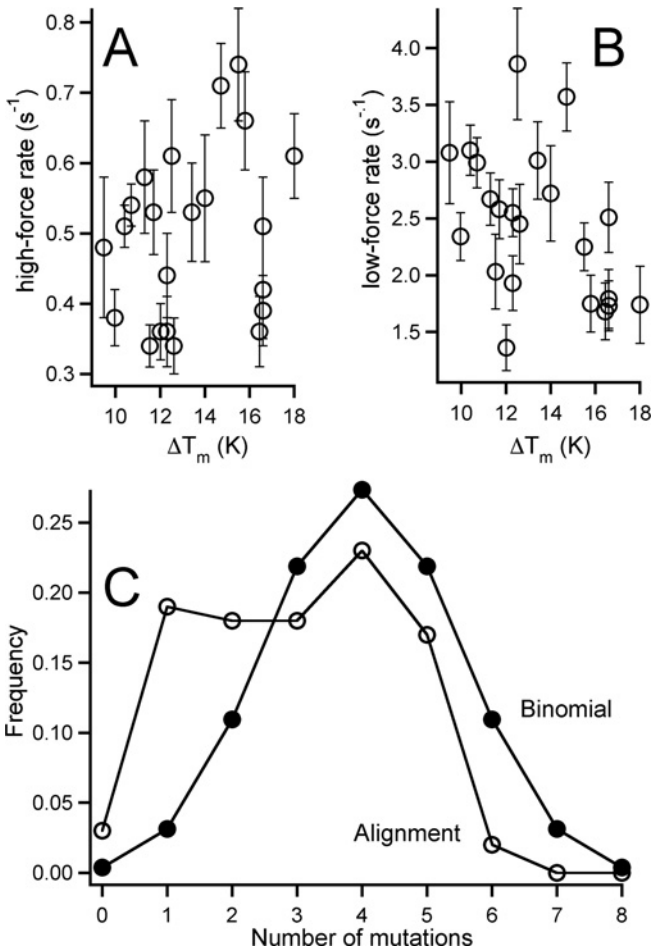

Figure 2 Disulfide-reduction rates and denaturation temperatures for thioredoxin variants from the combinatorial library

See the text and Figure 1 for details on the library design. (**A** and **B**) Plots of high-force rate and low-force rate against denaturation temperature. Denaturation temperatures are given with reference to the wild-type protein [i.e. $\Delta T_m = T_m(\text{variant}) - T_m(\text{wild-type})$]. Error bars are not shown when they are smaller than the size of the data point. These plots (as well as the plot of high-force rate against low-force rate shown in Figure 1B) have a large scatter and little correlation, indicating a significant degree of independent modulation of the three properties. (**C**) Distribution in the alignment used as the starting point of the analysis (Figure 1A) for combinations of the mutations included in the combinatorial library. The fraction of combinations with a given number of library mutations (relative frequencies) is plotted against that number of mutations. For comparison, a binomial distribution with $P = q = 1/2$ is also shown.

The modulations found in stability and different pathways of catalysis are largely independent of each other

The modulations found in stability and catalysis appear to be largely independent of each other, as is qualitatively shown by the fact that the plots of the properties for the library variants (Figures 1B, 2A and 2B) are essentially scattergrams that show little correlation, whereas non-independence between two properties (because of the existence of trade-offs or because of the two properties reflecting a common underlying feature) would result in correlated plots. Therefore the absence of clear correlations in the experimental property plots (Figures 1B, 2A and 2B) revealed the important result that several properties can be modulated to a significant extent in an independent manner by suitable combinations of conservative mutations.

To further quantify the independence between the studied properties we have carried out a PCA (principal component analysis), a mathematical procedure to identify the directions along which sample variation is maximal and to reduce the dimensionality of data sets [26]. In a first step, we applied PCA to the three two-property sets that can be derived from our results (i.e. low-force rate compared with denaturation temperature, high-force rate compared with temperature and high-force rate compared with low-force rate). For two properties that are strongly correlated, the PCA should reveal one major component together with a very minor one, indicating that the dimensionality of the set can be reduced to one. However, for our three two-property sets (Figures 3A–3C), the contributions of the first and second component both appear significant, supporting very weak correlations and therefore efficient independent modulation. In a second step, we performed a PCA analysis with the whole three-property set and Figure 3(D) shows the percentage of data variance explained by the resulting three principal components. Whereas the first and second components dominate, even the third one makes a significant contribution (approx. 10 % of the variance).

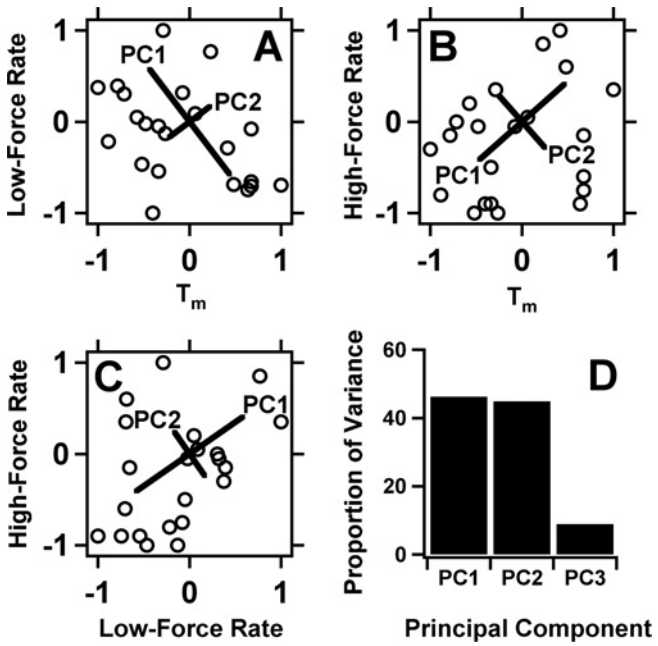

Figure 3 PCA of thioredoxin multi-protein modulation

(**A**–**C**) PCA analyses of two-property sets. Prior to the analysis, data are normalized to the (−1,1) interval. The directions of the two principal components (PC) are shown and the length of the lines is proportional to the component contribution to the variance of the data. (**D**) Results of the PCA analysis of the three-property set (denaturation temperature, high-force rate and low-force rate). The percentages of the data variance explained by each of the three principal components are shown. In all cases, PCA analyses were carried out in the MATLAB or XLminer environments.

A wide variety of patterns of protein modulation become possible in principle: combination of hyperthermophilic stability enhancement with eukaryotic signature of catalysis

The fact that stability-related and catalysis-related properties can be modified in an independent manner implies that different patterns of protein modulation are, in principle, possible. This is evident in the wide variety of property combinations for the studied library variants shown in Figures 1(B), 2(A) and 2(B). Clearly, the set of mutations studied has a high potential for protein multi-feature modulation, as is dramatically illustrated by the combination of stability enhancement and eukaryotic signature of catalysis we describe below.

The library variants were prepared with a His_6 tag for ease of purification. Previous studies have shown a negligible effect of His_6 tags on reduction rates measured by single-molecule AFM and the results reported in the present study indicated only a small effect on stability. However, we still carried out a detailed analysis of a particularly interesting library variant, referred to as trx*, prepared without the tag. trx* is a variant of *E. coli* thioredoxin with the following mutations: A22P, I23V and P68A (i.e. the V3 background) plus D10A, Q50A, G74S, E85Q and A87V (see the Supplementary Results and discussion section at http://www.BiochemJ.org/bj/429/bj4290243add.htm for further information). A comparison between trx* and wild-type thioredoxin from *E. coli* (without His_6-tags in both cases) is shown in Figure 4. trx* had a denaturation temperature of 108 °C (approx. 20 °C higher than wild-type *E. coli* thioredoxin) and a very slow unfolding rate (approx. 15000-fold slower than wild-type *E. coli* thioredoxin). Furthermore, in single-molecule experiments, it showed an enhanced low-force reduction rate and a diminished high-force rate, indicating a depressed simple S_N2 mechanism. In fact, the pattern of catalysis for trx* approaches that of a eukaryotic thioredoxin (see Figure 1B). The combination of high stability and eukaryotic pattern of catalysis in trx* is a particularly suggestive one, as, in most cases, eukaryotic organisms are not thermophiles [27].

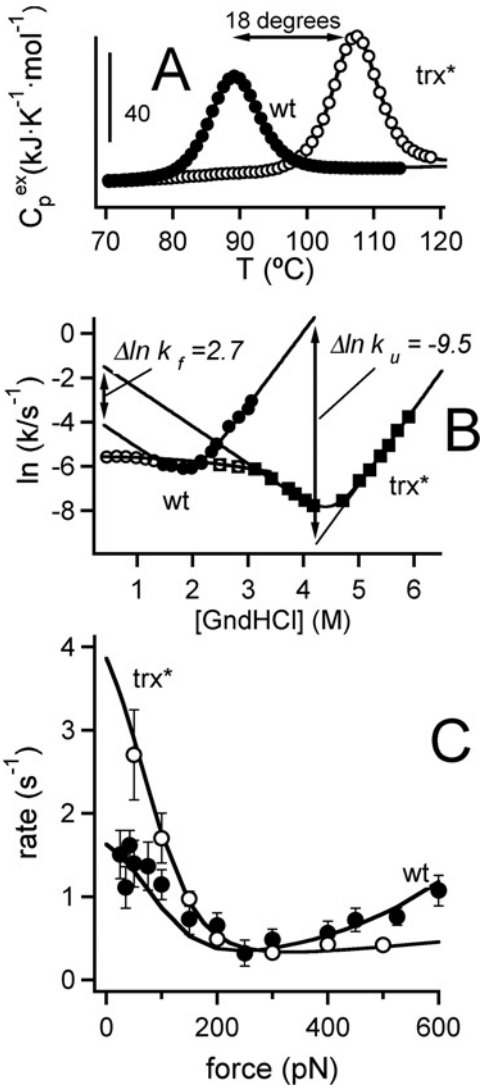

Figure 4 Very high stability and eukaryotic pattern of catalysis for a variant of *E. coli* thioredoxin

See the text and Figure 1 for details on library design. This variant, termed trx*, is compared with the wild-type *E. coli* protein (wt). For the sake of this comparison, both proteins were prepared without a His_6 tag. (**A**) Scanning calorimetry thermograms showing that the denaturation temperature of trx* is approx. 20 °C higher than that for the wild-type protein. (**B**) Chevron plots of folding–unfolding rate showing a hugely enhanced kinetic stability for trx* (the unfolding rate is approx. 15000-fold slower in the case of trx* as compared with wild-type thioredoxin). Further details on the chevron plot analysis are provided in the Supplementary section on folding–unfolding kinetics and in Supplementary Table S2 (at http://www.BiochemJ.org/bj/429/bj4290243add.htm). (**C**) Disulfide-reduction rate plotted against the applied mechanical force for wild-type thioredoxin and trx*, as determined from single-molecule AFM experiments. trx* shows a clearly eukaryotic pattern of catalysis, with the chemical S_N2 mechanism depressed.

Specific patterns of protein properties modulation are likely to be achieved through many mutational paths

The fact that a certain combination of mutations leads to a pattern of protein properties that enhances fitness does not necessarily

imply that such a pattern is accessible through Darwinian evolution. Evolutionary accessibility requires that a mutational path exists that leads to the desired combination without involving deleterious intermediate combinations. However, the protein multi-property modulations described in the present study are based on conservative mutations, which are not likely to show strong trade-off or sign-epistasis effects. This is clearly supported by three points. First, that the analysis of the combinatorial library data in terms of an 'independent mutation effect model' yields small values for the mutation effects on activity and catalysis (Supplementary Figure S5 at http://www.BiochemJ.org/bj/429/bj4290243add.htm). Secondly, the fact that strong trade-off effects would cause residue co-evolution (for instance, if a mutation improves function while it strongly decreases stability it will occur during evolution coupled with mutations that strongly increase stability). However, we have carried out a co-variance analysis of the alignment used (Supplementary Figure S3 at http://www.BiochemJ.org/bj/429/bj4290243add.htm) which indicates that co-evolution between the positions included in our library is not significant. Thirdly, that the principal component analyses shown in Figure 3 do not reveal strong correlations. Accordingly, we may expect that, for each specific modulation pattern, a significant number of evolutionarily allowed mutational paths exist. This idea is illustrated by the simple calculation we describe below.

Assume that, over some evolutionary time-span, changes in the organism environment and biological function are taking place in such a way that increased fitness is brought about by a thioredoxin enzyme of high stability, an enhanced rate of catalysis through the 'specific' Michaelis–Menten mechanism and depressed rate of catalysis through the 'chemical-like' S_N2 type of mechanism. This multi-property modulation is in fact achieved by the five mutations, D10A, Q50A, G74S, E85Q and A87V, present in the trx* variant described in the preceding section. Five mutations can occur in 5!=120 different temporal arrangements. The question is how many of these 120 mutational paths are evolutionarily allowed in the sense that they do not involve deleterious intermediate combinations (i.e. combinations involving strongly decreased fitness). We have assumed a simple linear model for the relationship between protein fitness (f) and the values of the properties under consideration in eqn (1):

$$f = \frac{1}{3}\left(\frac{\delta P_1}{\Delta P_1} + \frac{\delta P_2}{\Delta P_2} + \frac{\delta P_3}{\Delta P_3}\right) \quad (1)$$

where P_1, P_2 and P_3 refer to the values of the denaturation temperature (a measure of stability), the rate of disulfide reduction in single-molecule AFM experiments at low force (a measure of the contribution of the Michaelis–Menten mechanism) and the rate at high force (a measure of the chemical S_N2 mechanism). The changes labelled δP_i are the value of the property P_i with respect to the 'starting' value (the value when none of the five mutations has occurred) and ΔP_i refers to the total change, i.e. the final value of the property (when the five mutations have occurred) minus the starting values. Eqn (1) leads to fitness values that are arbitrarily scaled between 0 (no mutations) and 1 (the five mutations are present). More importantly, eqn (1) assumes that fitness increases linearly as the properties approach their final (five-mutation) values and that the three properties under consideration contribute equally to fitness (the same weight in the fitness equation). Eqn (1) allows us to calculate a fitness value for each node in any of the 120 mutational paths, provided that the mutation effects on the properties are known. For the illustrative purposes of this calculation, we have estimated these effects from the least-squares fit of a linear model to the experimental results (see the Supplementary Results and discussion section). The resulting fitness value compared with the 'number of accumulated mutations' profiles are shown in Figure 5 for all of the 120 mutational paths. Clearly, most of the paths involve increases in fitness with no deleterious intermediate combinations (i.e. no combinations with decreased fitness with respect to the initial value). Furthermore, even those paths that involve a decrease in fitness at some nodes, do so by a only a comparatively small amount. It is also worth noting again that the five mutations involved in the calculation are derived from an statistical analysis of a sequence alignment; that is, these mutations occur with high frequency during thioredoxin evolution. In the case of the Q50A (which cannot be achieved by a single-base substitution) the mutation likely occurs over an evolutionary time-scale through a glutamate residue as an intermediate amino-acid (Q→E→A); actually the three amino acids involved in this sequence occur with high frequency at position 50 of the alignment used (the number of occurrences are 11 for a glutamine residue, 20 for a glutamate residue and 23 for an alanine residue).

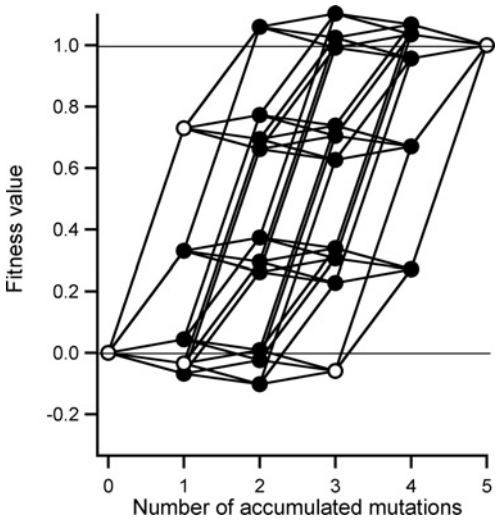

Figure 5 Fitness profiles for the 120 mutational paths leading to a variant with the five mutations related to the multi-property modulation observed in trx*

The calculations shown are for illustration and are based upon linear model for the relation between protein fitness and the property values (eqn 1). White symbols indicate the nodes for which library variants (Supplementary Figure S4 at http://www.BiochemJ.org/bj/429/bj4290243add.htm) were actually prepared. Note that 23 library variants, out of a total of 256, were prepared, leading to coverage of approx. 10 %. The initial and final nodes (zero and five mutations respectively) correspond to the V3 and trx* variants.

Finally, we must emphasize that the calculation described above and shown in Figure 5 is based on an artificial and hypothetical fitness function and is meant only for illustration. In particular, the actual relationship between fitness *in vivo* and stability/reduction-rates may be much more complex than that suggested by eqn (1). This notwithstanding, our illustrative calculation does provide some support for the notion that conservative mutations may contribute significantly to evolutionary protein adaptation, as not only may they originate a variety of patterns of multi-property modulation, but, in addition, it is likely that a multiplicity of mutational paths leading to each modulation pattern are accessible to Darwinian evolution.

DISCUSSION

Overall, we have shown evolutionarily significant modulation in protein properties related to stability and catalysis on the basis of a small number of mutations. The mutations used are fixed frequently during evolution, as deduced not only from alignment analysis (Figure 1A), but also from their non-negative coefficient values in the PAM250 substitution matrix, which range from 0 (D/A, Q/A and A/V substitutions) to 7 (the Y/F substitution). However, the modulations found by combining these mutations take thioredoxin properties clearly away from the values expected for the proteins in the sequence alignment used as a starting point. We thus obtained a eukaryotic signature in catalysis, whereas all sequences in the alignment are of bacterial origin, and we obtain typically thermophilic/hyperthermophilic stability enhancements although the sequences in the alignment belong to mesophilic or psychrophilic organisms. Furthermore, the combinatorial library screened is very small (256 variants), only about 20 variants were analysed and significant modulations are found in essentially all of them. Clearly, *in vitro* selection plays no role in the results obtained and we conclude that we have simply tapped a fundamental strategy for protein adaptation during evolution. A plausible interpretation of this strategy is summarized below.

Mutations that are fixed often during evolution (the type of mutations that are usually labelled as conservative, neutral, nearly-neutral, quiet, etc.) can actually modulate a variety of relevant protein properties. These mutations are non-disruptive and are not expected to involve strong trade-offs or sign-epistasis effects [7,9]. Different combinations of these high-frequency mutations will thus be available. For instance, combinations of the mutations included in our combinatorial library do occur in the alignment used with a distribution which does not depart dramatically from the binomial one (Figure 2C). The general implication is that natural selection may operate on many different mutational combinations, that a wide variety of patterns of protein modulation become available to Darwinian evolution and that the protein will easily cope with new situations. For instance, the variant of *E. coli* thioredoxin we have termed trx* displays a thermophilic/hyperthermophilic stability enhancement and a typically eukaryotic signature of catalysis. This combination of features may perhaps occur rarely in natural thioredoxins (as eukaryotic organisms are not usually thermophiles), but still thioredoxins should easily achieve this modulation pattern during evolution provided it brings about an adaptive advantage.

Of course, stability and pattern of catalysis in single-molecule AFM are just the 'tip of the iceberg' (as two important biophysical features we can assess in *in vitro* experiments). Many other features, related to the multitude of roles and interactions of thioredoxin *in vivo*, may be expected to be modulated by high-frequency mutations. This suggests the possibility of protein multi-feature optimization, an important biotechnological goal [28], which is not easily achieved with traditional methods. Thus, for instance, consensus approaches based on sequence-alignment statistics have been mostly used to enhance a single protein feature (stability in most cases, with some notable exceptions [29]) and studies on both stability and catalysis usually emphasize trade-offs (mutations that enhance activity often have deleterious effects on stability). Certainly, multi-feature optimization has been reported in the literature, but it seems to require screening of large libraries or extensive prior knowledge on the effects of the mutations employed [30]. By contrast, the results reported in the present paper suggest that screening of very small combinatorial libraries designed using sequence-alignment information could easily lead to simultaneous modulation/optimization of several relevant protein properties.

AUTHOR CONTRIBUTION

David Rodriguez-Larrea prepared the variants from the combinatorial library and designed, performed and analysed the experiments addressed at determining their stability, activity and folding/unfolding kinetics. Raul Perez-Jimenez, Inmaculada Sanchez-Romero and Julio Fernandez designed, performed and analysed the single-molecule AFM experiments. Inmaculada Sanchez-Romero carried out the principal component analyses. Asuncion Delgado-Delgado analysed the sequence alignment used in terms of the biological classification and the optimum growth temperature of the corresponding organisms. Jose Sanchez-Ruiz designed the research and wrote the paper. All authors discussed the manuscript.

FUNDING

This work was supported by the National Institutes of Health [grant numbers HL061228 and HL066030 (to J.M.F.)]; and by federal funds from the Spanish Ministry of Education [grant numbers BIO2009–09562 and CSD2009-00088 (to J.M.S.-R)].

REFERENCES

1 Benner, S. A., Caraco, M. D., Thomson, J. M. and Gaucher, E. A. (2002) Planetary biology: paleontological, geological and molecular histories of life. Science **296**, 864–868
2 Gaucher, E. A., Govindarajan, S. and Ganesh, O. K. (2008) Paleotemperature trend for precambrian life inferred from resurrected proteins. Nature **451**, 704–707
3 Bershtein, S., Segal, M., Bekerman, R., Tokuriki, N. and Tawfik, D. S. (2006) Robustness-epistasis shapes the fitness landscape of a randomly-drifting protein. Nature **444**, 929–932
4 Zeldowich, K. B., Chen, P. and Shakhnovich, E. I. (2007) Protein stability imposes limits on organism complexity and speed of molecular evolution. Proc. Natl. Acad. Sci. U.S.A. **104**, 16125–16157
5 Tokuriki, N. and Tawfik, D. S. (2009) GroEL/GroES chaperonin overexpression promoted genetic variation and enzyme evolution. Nature **459**, 668–673
6 Tokuriki, N. and Tawfik, D. S. (2009) Stability effects of mutations and protein evolvability. Curr. Opin. Struct. Biol. **19**, 569–604
7 Wang, X., Minasov, G. and Shoichet, B. K. (2002) Evolution of an antibiotic resistance by stability and activity trade-offs. J. Mol. Biol. **320**, 85–95
8 Weinreich, D. M., Watson, R. A. and Chao, L. (2005) Sign epistasis and genetic constraint on evolutionary trajectories. Evolution **59**, 1165–1174
9 Weinreich, D. M., Delaney, N. F., Depristo, M. A. and Hartl, D. L. (2006) Darwinian evolution can follow only very few mutational paths to fitter proteins. Science **312**, 111–114
10 Dayhoff, M. O., Schwarz, R. M. and Orcutt, B. C. (1978) Atlas of Protein Sequence and Structure, National Biomedical Research Foundation, Washington DC
11 Holmgren, A. (1985) Thioredoxin. Annu. Rev. Biochem. **254**, 237–271
12 Kumar, J. K., Tabor, S. and Richardson, C. C. (2004) Proteomic analysis of thioredoxin-targeted proteins in Escherichia coli. Proc. Natl. Acad. Sci. U.S.A. **101**, 3759–3764
13 Lemaire, S. D., Guillon, B., Maréchal, P., Keryer, E., Miginiac-Maslow, M. and Decottignies, P. (2004) New thioredoxin targets in the unicellular photosynthetic eukaryote *Chlamydomonas reinhardtii*. Proc. Natl. Acad. Sci. U.S.A. **101**, 7475–7480
14 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
15 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
16 Koti Ainavarapu, S. R., Wiita, A. P., Dougan, L., Uggerud, E. and Fernandez, J. M. (2008) Single-molecule force spectroscopy measurements of bond elongation during a biomolecular reaction. J. Am. Chem. Soc. **130**, 6479–6487
17 Wiita, A. P., Ainavarapu, S. R., Huang, H. H. and Fernandez, J. M. (2006) Force-dependent chemical kinetics of disulfide-bond reduction observed with single-molecule techniques. Proc. Natl. Acad. Sci. U.S.A. **103**, 7222–7227
18 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
19 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

20 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
21 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
22 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
23 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
24 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
25 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
26 Ringnér, M. (2008) What is principal component analysis? Nat. Biotechnol. **26**, 303–304
27 Weber, A. P. M., Horst, R. J., Barbier, G. G. and Oesterhelt, C. (2007) Metabolism and metabolobics of eukaryotes living under extreme conditions. Int. Rev. Cytol. **256**, 1–34
28 Marshall, S. A., Lazar, G. A., Chirino, A. J. and Dejarlais, J. R. (2003) Rational design and engineering of therapeutic proteins. Drug Discovery Today **8**, 212–221
29 Dai, M., Fisher, H. E., Temirov, J., Kiss, C., Phipps, M. E., Pavlik, P., Werner, J. H. and Bradbury, A. R. M. (2007) The creation of a novel fluorescent protein by guided consensus ingeniering. P.E.D.S. **20**, 69–79
30 Strausberg, S. L., Ruan, B., Fisher, K. E., Alexander, P. A. and Bryan, P. N. (2005) Directed coevolution of stability and catalytic activity in calcium-free subtilisin. Biochemistry **44**, 3272–3279

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

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.

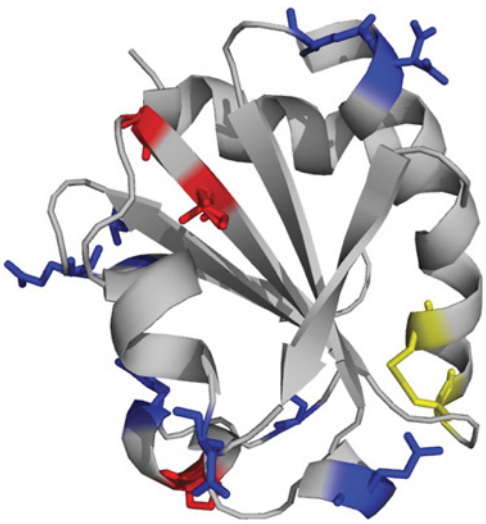

Figure S1 Positions of the *E. coli* wild-type thioredoxin molecule involved in the design of the combinatorial library

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

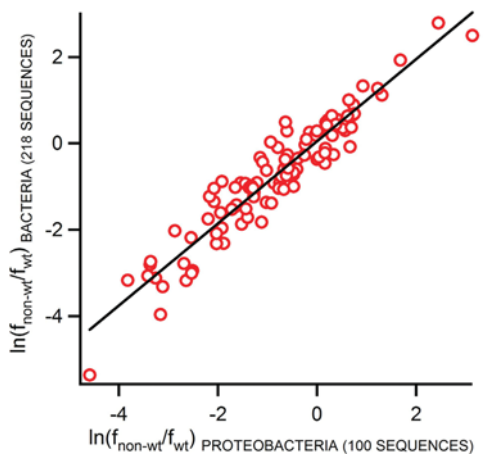

Figure S2 Correlation between the non-wild-type to wild-type frequency ratios derived from the analysis of alignments

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.

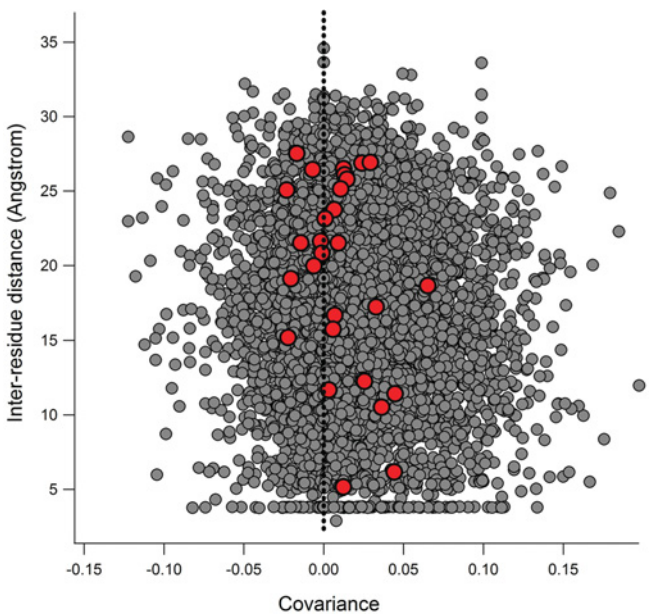

Figure S3 Plot of the inter-residue distance compared with covariance for thioredoxin

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

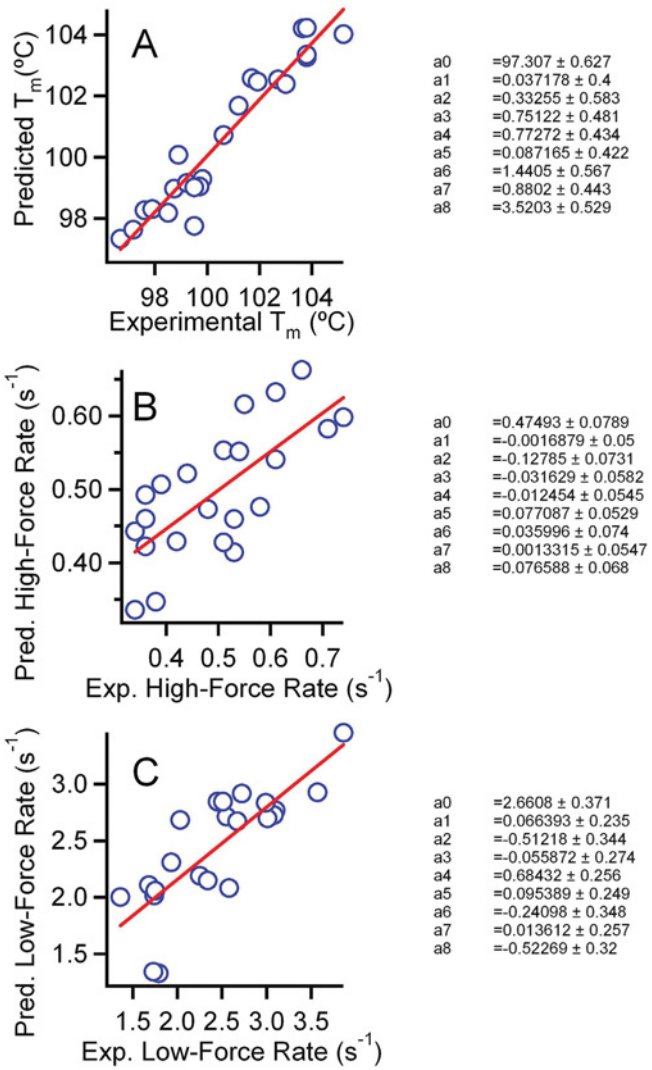

Figure S5 Correlation between experimental values and the predicted values derived from the fitting of an independent-mutation model

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

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.

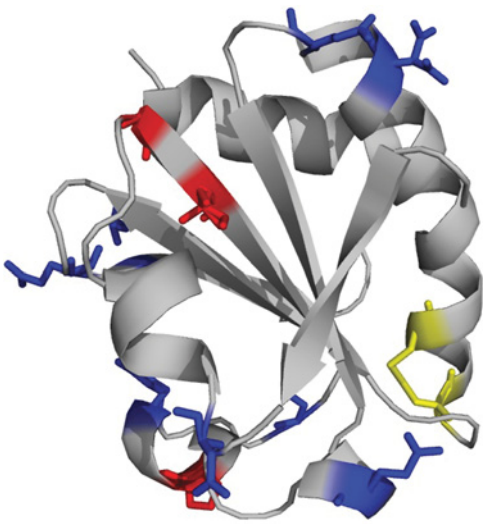

Figure S1 Positions of the *E. coli* wild-type thioredoxin molecule involved in the design of the combinatorial library

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

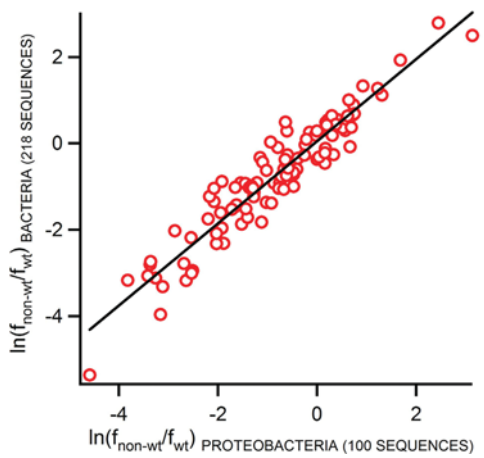

Figure S2 Correlation between the non-wild-type to wild-type frequency ratios derived from the analysis of alignments

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.

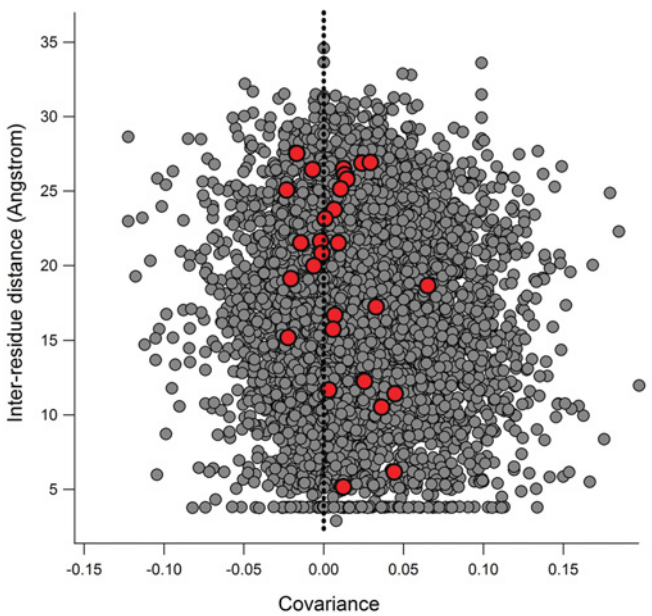

Figure S3 Plot of the inter-residue distance compared with covariance for thioredoxin

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

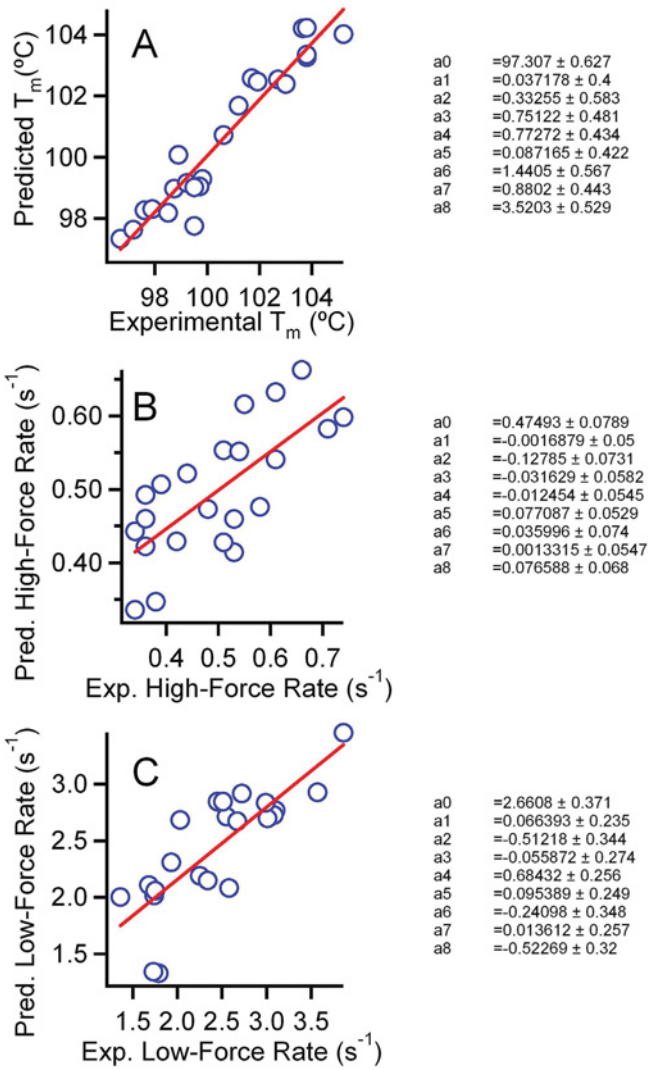

Figure S5 Correlation between experimental values and the predicted values derived from the fitting of an independent-mutation model

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