



HAL
open science

Role of conservative mutations in protein multi-property adaptation

David Rodriguez-Larrea, Raul Perez-Jimenez, Inmaculada Sanchez-Romero, Asuncion Delgado-Delgado, Julio M Fernandez, Jose M Sanchez-Ruiz

► To cite this version:

David Rodriguez-Larrea, Raul Perez-Jimenez, Inmaculada Sanchez-Romero, Asuncion Delgado-Delgado, Julio M Fernandez, et al.. Role of conservative mutations in protein multi-property adaptation. *Biochemical Journal*, 2010, 429 (2), pp.243-249. 10.1042/BJ20100386 . hal-00495495

HAL Id: hal-00495495

<https://hal.science/hal-00495495>

Submitted on 28 Jun 2010

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Title: 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.

[†] Department of Biological Sciences, Columbia University, New York, NY 10027, USA.

1 Address correspondence to: Jose M Sanchez-Ruiz. Facultad de Ciencias, Departamento de Quimica Fisica. Fuentenueva s/n. 18071-Granada, Spain. Tel: 34-958243189. Fax: 34-958272879. E-mail: sanchezr@ugr.es

SHORT TITLE: Protein Multi-property Adaptation

Abbreviations: AFM, atomic force microscopy. SET, single electron transfer. Wt, wild type.

Accepted Manuscript

ABSTRACT: Protein physicochemical properties must undergo complex changes during evolution, as a response to modifications in the organism environment, as a 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. Here, 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 revealed by single-molecule atomic-force spectroscopy. Furthermore, the changes found are evolutionary 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, a potential of conservative mutations to originate a multiplicity of evolutionary-allowed mutational paths leading to a variety of protein modulation patterns. Besides, they support the feasibility of using of evolutionary information to achieve protein multi-feature optimization, an important biotechnological goal.

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]. While 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, while 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 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 (since 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].

Here, we explore a point of view which differs from that often found in recent literature and summarized in the above paragraph. We reason that, since adaptations to new situations are common during evolution, many proteins are poised to change their properties efficiently, at least within certain evolutionary-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 the following features:

a) 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 “quite”. We use the term “conservative” in this work, but we specifically refer to mutations with non-negative coefficients in substitution matrices (such as Dayhoff’s PAM250 [10]) and, consequently, with high-frequency of fixation during evolution;

b) both, stability-related and function-related properties are modulated by the mutations in the set, with no strong bias for stability versus function;

c) the effects of the several mutations in the set are roughly independent and, therefore, strong trade-off and sign-epistasis effects do not occur;

d) 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 evolutionary significant range of protein properties values.

In this work, 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 evolutionary-meaningful changes in stability and the complex patterns of catalysis revealed by single-molecule atomic-force spectroscopy. Actually, thioredoxin provides an excellent model system to investigate the issues we have raised above for the three following reasons:

1) 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.

2) Thioredoxins catalyze the reduction of target disulfide bonds, regulating a multitude of cellular processes [11]. A recent proteomic analysis [12] has identified 80 proteins associated with thioredoxin and has involved thioredoxin in at least 26 cellular processes in *E. 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.

3) Recent single-molecule work [14, 15] has indeed revealed a diversity of reduction mechanisms in thioredoxin with evolutionary-differentiated patterns of catalysis showing a well-defined correlation with the domains of life. These studies indicate the evolutionary ranges 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 atomic force microscopy (AFM) allows the application of a calibrated force to a disulfide bonded substrate making it possible the study of the catalytic mechanisms with sub-Angstrom 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 electron-transfer reaction (SET), while thioredoxins of bacterial origin exhibit in addition a simple nucleophilic substitution (S_N2), thus showing a rate increase at high force which 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].

EXPERIMENTAL

See “Supporting Information” for a detailed description.

The sequence alignment used has been previously described in detail [18, 19]. Briefly, BLAST2 [Gish W. (1996-2003) <http://blast.wustl.edu>] was used to search the sequence of *E. coli* thioredoxin as query in the UniProt/TrEMBL 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 1A. Information from references 20 and 21 was used to examine the temperature range for growth and the optimal growth temperature for the 100 proteobacteria from which sequences are included in the alignment. All these microorganisms 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 atomic force microscopy experiments were carried out as recently described [14, 15]. Very briefly, we used as protein substrate a polypeptide made of eight repeats of the I27 domain of human cardiac titin with engineered cysteines. A custom-built atomic force microscope controlled by an analogue proportional-integral-derivative feedback system [24] was employed to follow the reduction catalyzed by thioredoxin of individual disulfide bonds under a stretching force applied to the substrate.

RESULTS

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

To investigate how evolutionary relevant changes in thioredoxin properties may be achieved, we have determined a number of high-frequency-of-fixation mutations using an alignment of a hundred sequences derived [18] from a database BLAST search with the *E. coli* thioredoxin sequence as 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 of this, the three highest-frequency mutations in the alignment lead to stability enhancement (an increase in denaturation temperature of about 10 degrees with respect to the wt *E. coli* protein), as shown in previous work [18]. The corresponding triple-mutant variant (referred to as V3) is the starting point of the analysis reported here. Using V3 as a background, we have constructed a combinatorial library using the next 8 high-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 show 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 have also assessed the thermodynamic stability (equilibrium denaturation temperature measured by scanning calorimetry) for the library variants: again a large variability is found although a trend to stability enhancement is clearly apparent (figure 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 at this point is that a small number of mutations determine simultaneous modulations in stability and complex patterns of catalysis. It must be also 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 of combinations of the mutations selected is close to the binomial one (figure 2C).

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 property vs. property for the library variants (figures 1B, 2A and 2B) are essentially scattergrams that show little correlation. Note that 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 vs. property plots (figures 1B, 2A and 2B) reveals the important result that the several properties can be to a significant extent modulated in an independent manner by suitable combinations of conservative mutations.

To further quantify the independency between the studied properties we have carried out a Principal Component Analysis (pca), 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 data (i.e., low-force rate vs. denaturation temperature, high-force rate vs. temperature and high-force rate vs. low-force rate). For two properties that are strongly correlated, 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 (Figure 3A, 3B and 3C), the contributions of the first and second component appear both 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 3D shows the percentage of data variance explained by the resulting three principal components. While the first and second components dominate, even the third one makes a significant contribution (about 10% of the variance).

A wide variety of patterns of protein modulation become in principle possible. For instance: 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 1B, 2A and 2B. Clearly, the set of mutations we study 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-tag for ease of purification. Previous studies have shown a negligible effect of His-tags on reduction rates measured by single-molecule AFM and results reported here indicate only a small effect on stability. Still, we deemed convenient to carry 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 “Supporting Information” for further details). A comparison between *trx** and wt *trx* from *E. coli* (without His-tags in both cases) is shown in figure 4. *Trx** has a denaturation temperature of 108 °C (about 20 degrees higher than that for *E. coli* wt) and a very slow unfolding rate (about 15000 times slower than that for *E. coli* wt). Furthermore, in single-molecule experiments, it shows 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 also figure 1B). The combination of high stability and eukaryotic pattern of catalysis in *trx** is a particularly suggestive one, since, in most cases, eukaryotic organisms are not thermophiles [27].

Specific patterns of protein properties modulation may be likely 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 this work are based on conservative mutations, which are not likely to show strong trade-off or sign-epistasis effects. This is clearly supported by: 1) the analysis of the combinatorial library data in terms of an independent-mutation-effect

model which yields small values for the mutation effects on activity and catalysis (see figure S4 in Supporting Information ; 2) strong trade-off effects would cause residue co-evolution (for instance, if a mutation improves function while it decreases strongly stability, it will occur during evolution coupled with mutations that strongly increase stability). However, we have carried out a covariance analysis of the alignment used (figure S3 in Supporting Information) which indicates that co-evolution between the positions included in our library is not significant; 3) 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 evolutionary-allowed mutational paths exist. This idea is illustrated by the simple calculation we describe below:

Assume that, over some evolutionary time-span, changes in 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, 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 evolutionary 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 relation between protein fitness (f) and the values of the properties under consideration:

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

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 5 mutations has occurred) and ΔP_i stands for the total change: the final value of the property (when the 5 mutations have occurred) minus the starting values. Equation 1 leads to fitness values which are arbitrarily scaled between 0 (no mutations) and 1 (the five mutations are present). More importantly, equation 1 assumes that fitness increases linearly as the properties approach their final (5-mutations) values and that the three properties under consideration contribute equally to fitness (same weight in the fitness equation). Equation 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 data (see Supporting Information). The resulting fitness value versus number-of-accumulated-mutations profiles are shown in figure 5 for all 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 the initial value). Furthermore, even those paths that involve a decrease in fitness at some node, 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 glutamate as an intermediate amino-acid: $Q \rightarrow E \rightarrow A$; 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 Q, 20 for E and 23 for A).

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 relation between fitness *in vivo* and stability/reduction-rates may be much more complex than that suggested by equation 1. This notwithstanding, our illustrative calculation does provide some support for the notion that conservative mutations may contribute

significantly to evolutionary protein adaptation, since, not only they may 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 evolutionary 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 zero (D/A, Q/A and A/V substitutions) to 7 (Y/F substitution). Still, 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 obtain a eukaryotic signature in catalysis, while all sequences in the alignment are of bacterial origin and we obtain typically thermophilic/hyperthermophilic stability enhancement, while 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 analyzed 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 kind of mutations that are usually labelled as conservative, neutral, nearly-neutral, quite, 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 (see figure 2C). The general implication is that natural selection may operate on many different mutation combinations, a wide variety of patterns of protein modulation become available to Darwinian evolution and 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 (since 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” (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 exception [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 here 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.

FUNDING

This work was supported by NIH grants HL061228 y HL066030 to J.M.F., and Feder Funds and Spanish Ministry of Education and Science Grants BIO2009-09562 and CSD2009-00088 to J.M.S.-R.

Accepted Manuscript

THIS IS NOT THE VERSION OF RECORD - see doi:10.1042/BJ20100386

FIGURE LEGENDS

Figure 1

Single-molecule analysis of thioredoxin catalysis using a combinatorial library of conservative mutations. (A) Statistical analysis of an alignment of a hundred sequences derived from a database search using the *E. coli* sequence as query. All sequences in the alignment belong to proteins from proteobacteria. Mutations are selected on the basis of the $f_{\text{non-wt}}/f_{\text{wt}}$ ratio, where f_{wt} is the number of occurrences in the alignment of the residue present in the query sequence (that of wt thioredoxin from *E. coli*) and $f_{\text{non-wt}}$ 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 8 mutations with the next highest values of the ratio (shown in green). (B) Rates of disulfide-bond reduction by thioredoxin determined by single-molecule atomic force spectroscopy. 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 text and figure 3).

Figure 2

Disulfide-reduction rates and denaturation temperatures for thioredoxin variants from the combinatorial library. (See text and figure 1 for details on library design). (A, B) Plots of high-force rate and low-force rate vs. denaturation temperature. Denaturation temperatures are given with reference to the wt protein [i.e., $\Delta T_m = T_m(\text{variant}) - T_m(\text{wt})$]. 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 vs low-force rate in Fig. 1B) show large scatter and little correlation, indicating a significant degree of independent modulation of the three properties. (C) Distribution in the alignment used as starting point of the analysis (fig. 1A) of combinations of the mutations included in the combinatorial library. The fraction of combinations with a given number of library mutations (relative frequencies) is plotted vs. that number of mutations. For comparison, a binomial distribution with $p=q=1/2$ is also shown.

Figure 3

Principal component analysis 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 are shown in red 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, 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.

Figure 4

Very high stability and eukaryotic pattern of catalysis for a variant of *E. coli* thioredoxin. (See text and figure 1 for details on library design). This variant, termed trx*, is compared with the wild-type *E. coli* protein (labelled wt). For the sake of this comparison, both proteins were prepared devoid of a His-tag. (A) Scanning calorimetry thermograms showing that the denaturation temperature of trx* is about 20 degrees higher than that for the wild-type protein. (B) Chevron plots of folding-unfolding rate showing a hugely enhanced kinetic stability for trx* (unfolding rate is about 15000-times slower in the case of trx* as compared with wt). See "Supporting Information" for further details about chevron plot analysis. (C) Disulfide-reduction rate vs. applied mechanical force for wt and trx*, as determined from single-molecule atomic force spectroscopy experiments. Trx* shows a clearly eukaryotic pattern of catalysis, with the chemical S_N2 mechanism clearly depressed.

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 (figure 4).** The calculations shown are meant for illustration and are based upon linear model for the relation between protein fitness and the property values (equation 1). Open symbols indicate the nodes for which library variants (Table S1 in “Supporting information) were actually prepared (note that 23 library variants- out of a total of 256- were prepared, leading to a coverage of about 10%). The initial and final nodes (0 and 5 mutations, respectively) correspond to the V3 and *trx** variants.

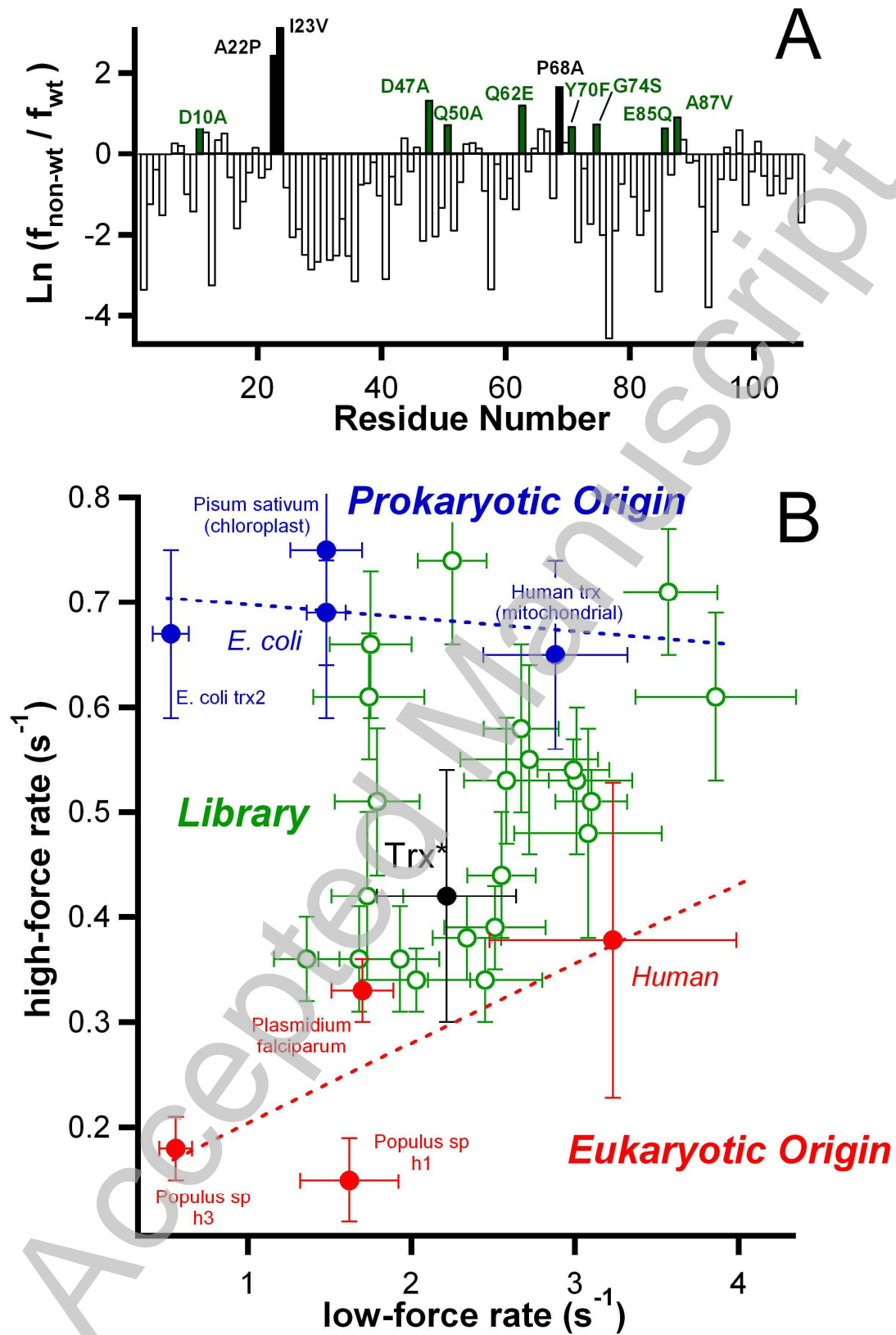
Accepted Manuscript

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. and Miginiac-Maslow, M., Decottignies P., *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., Cho, S.H., Beckwith, J., Gelhaye, E., Jacquot, J.P., Gaucher, E.A., Sanchez-Ruiz, J.M., Berne, B.J. and Fernandez, J.M. (2009) Diversity of chemical mechanisms in thioredoxin catalysis revealed by single-molecule force spectroscopy. *Nat. Struct. Mol. Biol.* **16**, 890-896
16. Koti Ainaravapu, 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., Ainaravapu, 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. Eds. (2007) *The Prokaryotes*. 3rd edition
21. Garrity, G. M., Brenner, D. J., Krieg, N. R. and Staley, J. T. Eds. (2005) *The Bergey's Manual of Systematic Bacteriology*. Second edition
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(3)**, 303-304
27. Weber, A.P.M., Horst, R.J., Barbier, G.G. and Oesterhelt, C. (2007) Metabolism and metabolomics 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 engineering. *P.E.D.S.* **20**, 69-79
30. Strausberg, S.L., Ruan, B., Fisher, K.E., Alexander, P.A., Bryan, P.N. (2005) Directed coevolution of stability and catalytic activity in calcium-free subtilisin. *Biochemistry* **44**, 3272-3279

Figure 1



THIS IS NOT THE VERSION OF RECORD - see doi:10.1042/BJ20100386

Figure 2

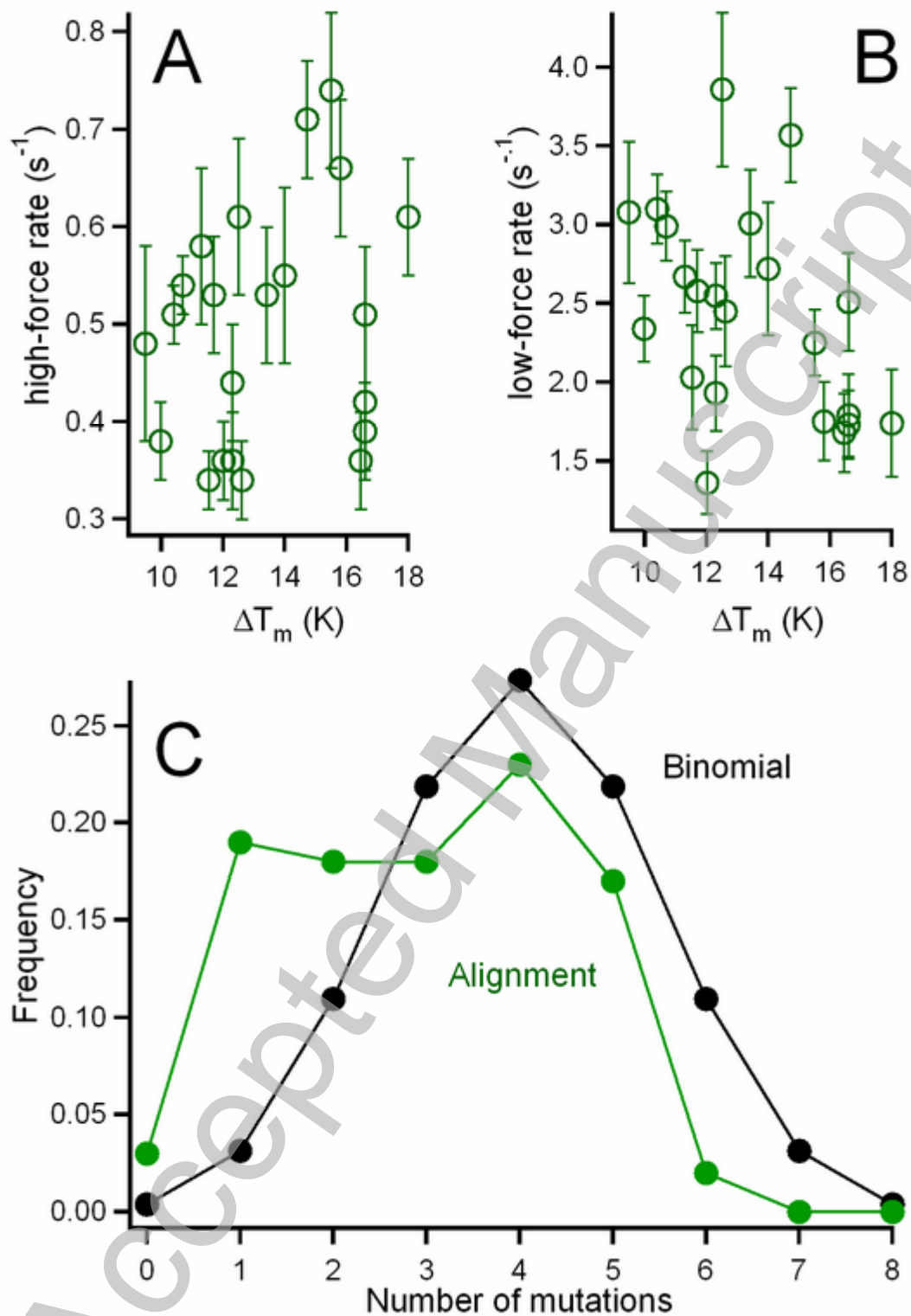
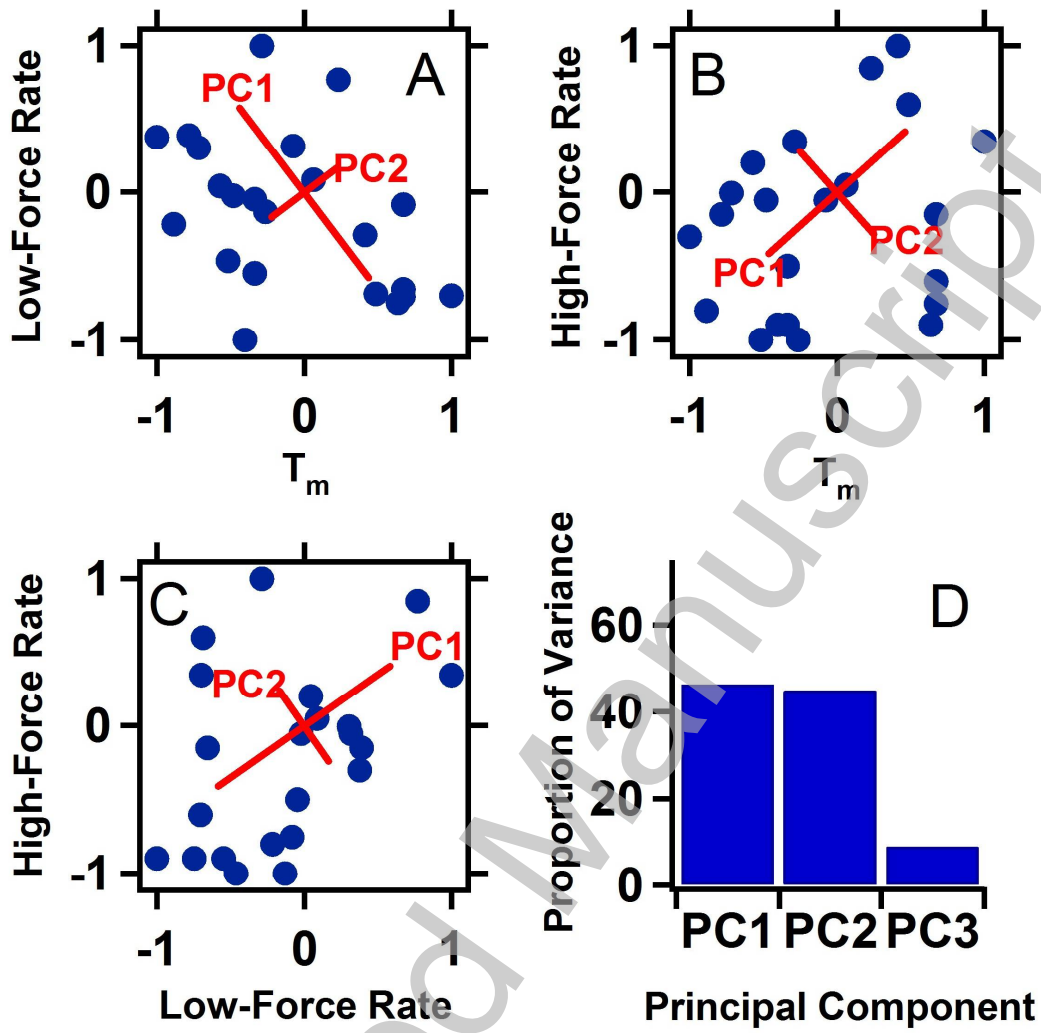


Figure 3



THIS IS NOT THE VERSION OF RECORD - see doi:10.1042/BJ20100386

Accepted Manuscript

Figure 4

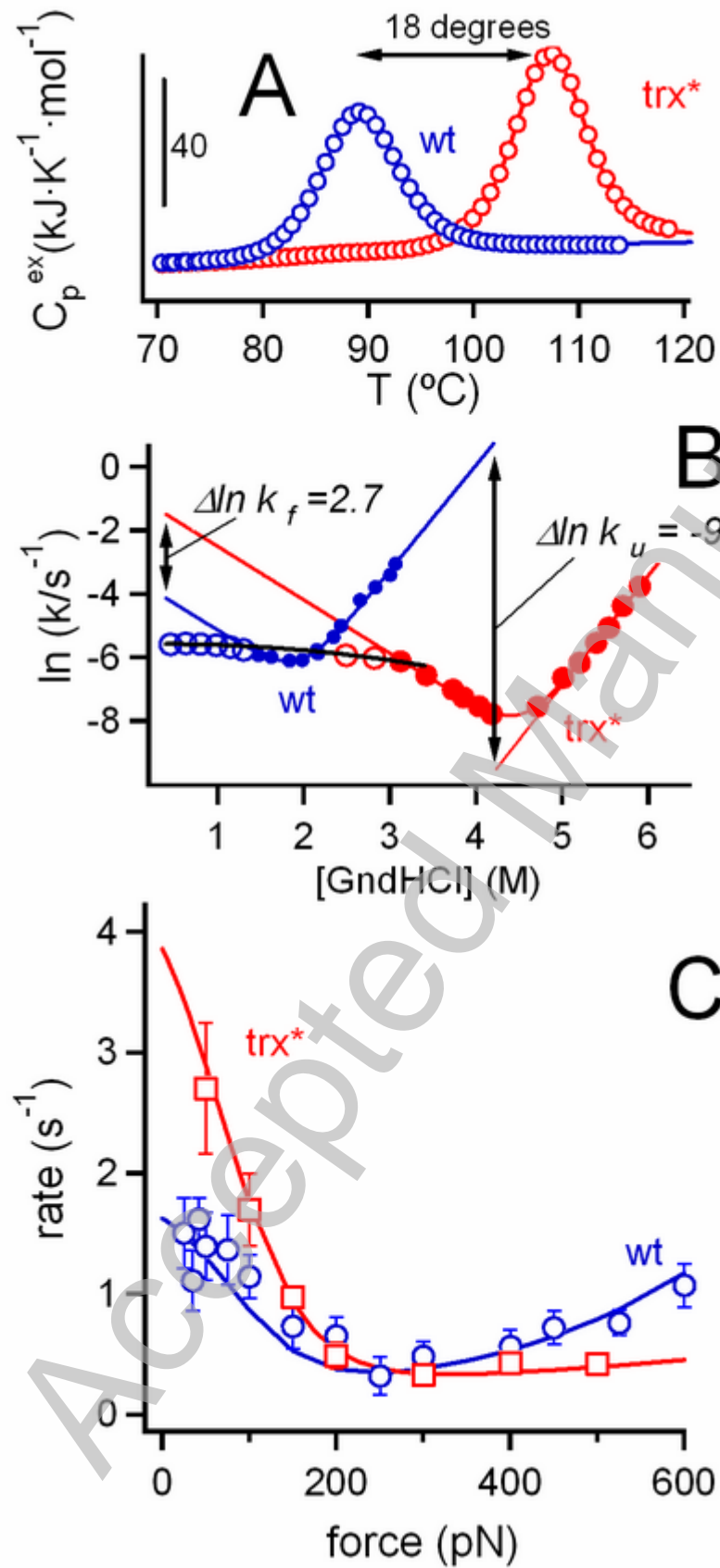
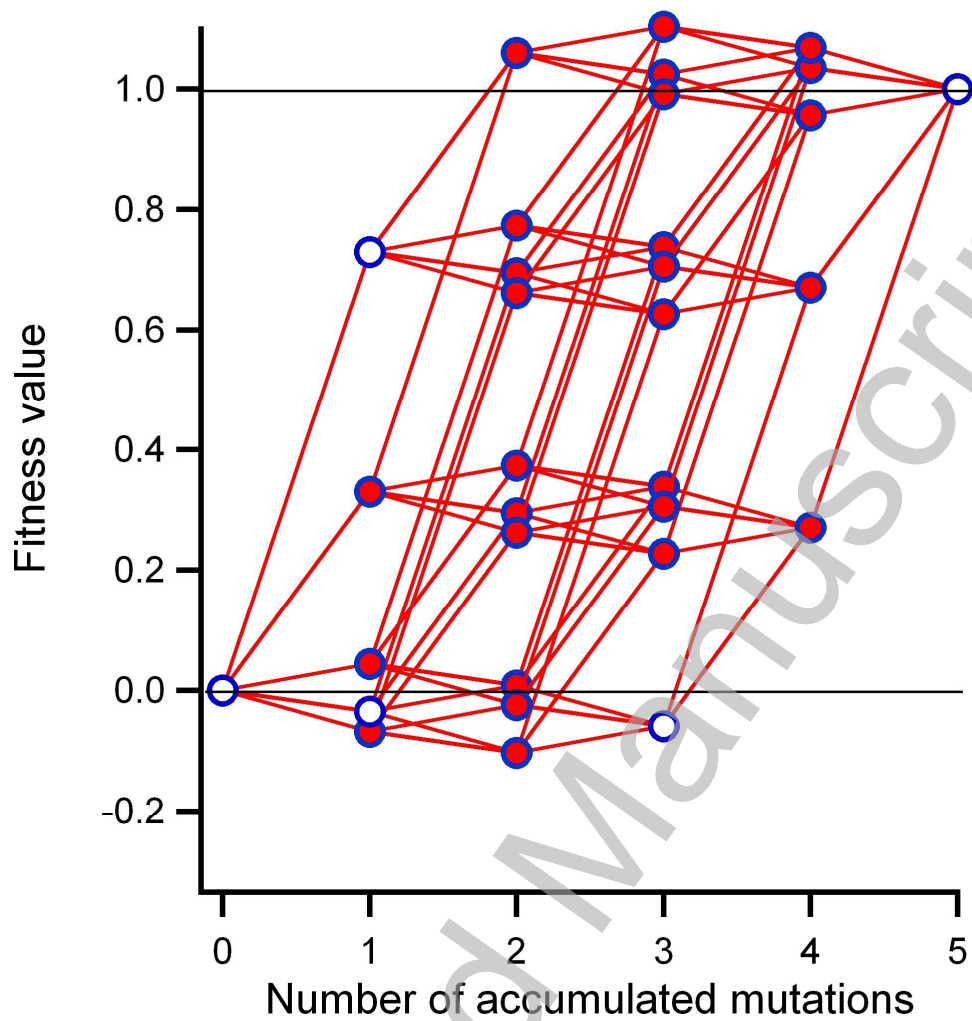


Figure 5



THIS IS NOT THE VERSION OF RECORD - see doi:10.1042/BJ20100386