


Research Article

Non-conservation of folding rates in the thioredoxin family reveals degradation of ancestral unassisted-folding

Gloria Gamiz-Arco¹, Valeria A. Risso¹, Adela M. Candel^{1,*}, Alvaro Inglés-Prieto^{1,†}, Maria L. Romero-Romero^{1,‡}, Eric A. Gaucher², Jose A. Gavira³, Beatriz Ibarra-Molero¹ and  Jose M. Sanchez-Ruiz¹

¹Departamento de Química Física, Facultad de Ciencias, Unidad de Excelencia de Química Aplicada a Biomedicina y Medioambiente (UEQ), Universidad de Granada, 18071 Granada, Spain; ²Department of Biology, Georgia State University, Atlanta, GA 30303, U.S.A.; ³Laboratorio de Estudios Cristalográficos, Instituto Andaluz de Ciencias de la Tierra, CSIC-University of Granada, Avenida de las Palmeras 4, 18100 Armilla, Granada, Spain

Correspondence: Beatriz Ibarra-Molero (beatriz@ugr.es) or Jose M. Sanchez-Ruiz (sanchezr@ugr.es)



Evolution involves not only adaptation, but also the degradation of superfluous features. Many examples of degradation at the morphological level are known (vestigial organs, for instance). However, the impact of degradation on molecular evolution has been rarely addressed. Thioredoxins serve as general oxidoreductases in all cells. Here, we report extensive mutational analyses on the folding of modern and resurrected ancestral bacterial thioredoxins. Contrary to claims from recent literature, *in vitro* folding rates in the thioredoxin family are not evolutionarily conserved, but span at least a ~100-fold range. Furthermore, modern thioredoxin folding is often substantially slower than ancestral thioredoxin folding. Unassisted folding, as probed *in vitro*, thus emerges as an ancestral vestigial feature that underwent degradation, plausibly upon the evolutionary emergence of efficient cellular folding assistance. More generally, our results provide evidence that degradation of ancestral features shapes, not only morphological evolution, but also the evolution of individual proteins.

*Present address: Eurofins
Villapharma Research. Parque
Tecnológico de Fuente Álamo.
Ctra. El Estrecho – Lobosillo
km 2.5 30320 Fuente Álamo,
Murcia, Spain

†Present address: Research
Center for Molecular Medicine
of the Austrian Academy of
Sciences. Lazarettgasse 14,
AKH BT 25.3, 1090 Vienna,
Austria

‡Present address: Max Planck
Institute for Molecular Cell
Biology and Genetics, Dresden
01307, Germany

Received: 4 October 2019
Revised: 19 November 2019
Accepted: 21 November 2019

Accepted Manuscript online:
21 November 2019
Version of Record published:
10 December 2019

Introduction

Instances of so-called imperfect (poor or suboptimal) ‘design’ have been extensively studied in records of evolutionary history, and have served as evidence that living organisms, rather than being designed, are the products of complex evolutionary forces and histories [1,2]. Glaringly questionable ‘design’, such as the recurrent laryngeal nerve in mammals, thus suggests that evolutionary tinkering with previously functional features can limit the possible outcomes of new functions in the future. Still, many examples of imperfect morphological ‘design’ are simply related to the evolutionary degradation of ancestral features that are no longer useful. Examples are abundant and include human’s limited capability to move the ears (linked to the degradation of barely used muscles) as well as the presence of a tailbone, vestigial leg bones in whales and vestigial wings in flightless birds. Evolutionary degradation is primarily a consequence of the inability of natural selection to purge mutations that impair a feature, once the feature has ceased to be useful (i.e. once it has ceased to confer a functional selective advantage). Darwin did realize that ‘rudiments’ are evidence of descent from ancestral forms and discussed many examples in the first chapter of *The Descent of Man* (1871). More recently, the discovery that genomes include large numbers of pseudogenes has provided a clear example of evolutionary degradation at the molecular level. Thus, reversing the mutations that originally led to the silencing of a given gene does not typically restore the function of the encoded protein [3] because, after a gene is silenced, it is no longer subject to purifying natural selection and quickly accumulates other degrading mutations.

Plausibly, the evolutionary degradation of useless ancestral features is widespread, not only in morphological evolution, but also during the course of molecular evolution. Other than pseudogenes [3], however, molecular examples appear to have been rarely discussed in the literature, if at all. We argue here that evolutionary analysis of protein folding processes may provide clear examples of evolutionary degradation at the molecular level. This is so because folding *in vivo* within modern organisms [4–9] is protected and assisted by a complex folding-assistance machinery, including chaperones and the chaperone functionality of the ribosome, while, on the other hand, folding studies in the test tube (*in vitro* folding) probe unassisted folding. Since evolution has no foresight, however, folding assistance cannot have arisen before protein folding itself, inasmuch as the components of the folding-assistance machinery are proteins themselves that need to be folded to be functional. It follows that the most ancient proteins could plausibly fold with little or no assistance. Therefore, unassisted folding may have been relevant at a primordial stage, prior to (or concomitantly with) the emergence of folding assistance.

We propose, therefore, that the *in vitro* folding process for modern proteins of ancient evolutionary origin may bear signatures of evolutionary degradation. Thioredoxins, general oxidoreductases that display a wide substrate scope and that are involved in a diversity of cellular processes [10,11], should provide an excellent model system to explore this possibility. They are present in all known cells (eukaryotes, bacteria and archaea) and it is thus plausible that they existed at a very early stage, even perhaps preceding the emergence of an efficient folding assistance. Indeed, thioredoxins can fold without assistance in the test tube. It has been known for many years [12], however, that thioredoxins have a ‘folding problem’ related to the presence of a proline residue in *cis* conformation at position 76 (we use *Escherichia coli* thioredoxin numbering throughout). *Cis*-prolines in native protein structures create folding kinetic bottlenecks [13–15], since isomerization is slow, the *trans* conformer is favoured in unfolded polypeptide chains and may become trapped in intermediate states in the folding landscape thus further slowing down folding. For thioredoxins, mutational escape from the problem is not possible, since position 76 is close to the catalytic disulfide bridge and the presence of a proline at that position is required for a fully functional active-site conformation [12]. Pro76 is thus strictly conserved in thioredoxins.

Here, we first study the folding *in vitro* of *E. coli* thioredoxin and two of its resurrected Precambrian ancestors. An extensive mutational analysis allows us to explain the slower folding of the modern protein in terms of a single amino acid replacement that aggravated the folding problem created by the *cis*-proline at the active site. Furthermore, the identified replacement points to a region of the thioredoxin molecule where mutations can be reasonably expected to impact the folding rate. Experimental analysis of a set of modern bacterial thioredoxins selected to represent natural sequence diversity in this region shows that, contrary to what it has been claimed in recent literature [16,17], *in vitro* folding rates are not evolutionarily conserved. In fact, *in vitro* folding for some of the studied modern thioredoxins occurs in the approximately hour time scale and is between 1 and 2 orders of magnitude slower than both the inferred ancestral folding and the folding of other modern thioredoxins. These results suggest an interpretation of *in vitro* folding as a degraded version of primordial unassisted folding. More generally, our results provide evidence that degradation shapes evolution not only at the morphological level but also at the level of individual enzymes.

Materials and methods

Protein expression and purification

E. coli, LBCA and LPBCA thioredoxins as its variants studied in this work were prepared without purification tags following procedures we have previously described in detail [18–20]. Proteins representing bacterial thioredoxins (Figure 6) were prepared with His-tags using affinity chromatography. Mutations were introduced using the QuikChange Lighting Site-Directed Mutagenesis kit (Agilent Technologies) and checked by DNA sequencing.

Protein solutions were prepared by exhaustive dialysis at 4°C against 50 mM Hepes (pH 7). Protein concentrations were determined spectrophotometrically using known values for the extinction coefficients. Solutions of guanidine in 50 mM Hepes (pH 7) were prepared as previously described [18–20]. Prior to use, urea was purified by ion-exchange chromatography as previously described [21]. Guanidine and urea concentrations were determined by refractometry.

Activity determinations

Most reported activity determinations are based on the insulin turbidimetric assay [22], as described previously [18]. Briefly, thioredoxin catalysis of the reduction in insulin by DTT is determined by following the aggregation of

the β -chain of insulin. An aliquot of a thioredoxin solution is added to 0.5 mg/ml of bovine pancreatic insulin and 1 mM DTT at pH 6.5 and the rate is calculated from the slope of a plot of absorbance versus time at the inflexion point (see Supplementary Figure S1 for illustrative examples). Values given (Figure 3c and Supplementary Table S1) are the average of at least three independent measurements.

For some selected variants, we also assayed thioredoxin activity with thioredoxin reductase coupled to the reduction in Ellman's reagent (5,5'-dithiobis(2-nitrobenzoic acid) or DTNB) at pH 8, essentially as described by Slaby and Holmgren [23]. Concentrations used were 0.02 μ M for the reductase, 0.5 mM DTNB and 0.25 mM NADPH. Thioredoxin concentrations were typically in the range 0.15–0.20 μ M. Values reported are the average of at least three independent determinations.

Unfolding and folding kinetics studied by steady-state fluorescence measurements

Kinetic data for non-mutated *E. coli* thioredoxin and LPBCA thioredoxin given in Figures 3, 4 and Supplementary Figure S10 are taken from [20]. All other kinetic data shown were obtained in this work. All experiments were performed at 25°C. Folding–unfolding kinetics were studied using procedures we have previously described in detail [20,24]. Briefly, we measured the time-dependence of the fluorescence emission at 350 nm with excitation at 276 nm, after suitable guanidine- or urea-concentration jumps. For experiments in guanidine solutions, we typically used 20-fold dilution from \sim 4 to 5 M guanidine or from zero guanidine concentration for experiments carried at denaturant concentrations approximately above or below the denaturation midpoint. For experiments in urea solutions, we typically used 20-fold dilution from \sim 10 M urea or from zero urea concentration for experiments carried at denaturant concentrations approximately above or below the denaturation midpoint. The ancestral LBCA and LPBCA thioredoxins are highly stable and are not fully denatured despite high concentrations of urea. Folding rates for these proteins in urea solutions (Figure 6) were obtained by first denaturing them in concentrated guanidine followed by a high dilution into urea solutions, in such a way that the final guanidine concentration was very low (\sim 0.1 M). Typically, the protein concentration in the fluorescence kinetic experiments was on the order of 0.05 mg/ml.

Unfolding kinetics could be adequately fitted with a single exponential equation from which the rate constant could be easily calculated (see Supplementary Figure S2 for representative examples). Many folding profiles could also be well described by a single exponential within the time range of the manual mixing experiments. However, two exponential terms were required to achieve good fits in many other cases (see Supplementary Figure S3 for representative examples).

Finally, the long-time, equilibrium fluorescence values derived from the analyses of kinetic profiles were used to assess the thermodynamic stability of the modern thioredoxins studied in this work. Profiles of equilibrium fluorescence intensity versus urea concentration were fitted assuming a linear dependence of the unfolding free energy with denaturant concentration [25] within the narrow transition range and using linear pre- and post-transition baselines. Values of the urea midpoint concentration ($C_{1/2}$, the urea concentration at which the unfolding free energy is zero) and the denaturant-concentration dependence of the unfolding free energy ($m = -d\Delta G/d[\text{urea}]$) were obtained from the fits. These values are collected in Supplementary Table S4, where additional details of the fitting process are provided.

Using double-jump unfolding assays to determine the relevant kinetic phase of the major folding channel

Unlike unfolding, which often occurs in a single kinetic phase, protein folding is typically a complex process involving several parallel kinetic channels leading to the native state, as well as the transient population of intermediate states in many of these channels [26]. *In vitro* folding of thioredoxin is certainly known to conform to this scenario [27]. In this work, the complexity of *in vitro* thioredoxin folding is revealed by the multi-exponential folding profiles found in some cases (Supplementary Figure S3) and by clear rollovers in the folding branches of all the Chevron plots reported (Figures 3, 4, 6 and Supplementary Figure S10). In general, the folding rate of any given protein (i.e. the rate that defines the time scale of the protein folding process) could be defined in terms of the main slow phase of an experimental folding kinetic profile obtained using a suitable physical property. Still, it is absolutely essential to ascertain that this phase does indeed reflect the major kinetic channel that leads to the native protein. It would be conceivable, for instance, that most of the protein arrived to the native state in a slower phase that does not bring about a significant change in the

physical property being measured (steady-state fluorescence, in our case) and which is, therefore, not detected. Also, it would be conceivable that most of the protein arrives to the native state in a fast phase and that the slower phase detected in the kinetic folding profiles reflects a minor structural re-arrangement of the native ensemble or, alternatively, the folding of a small fraction of the protein from a kinetically trapped intermediate state. Thus, even if a single exponential phase is detected by the physical property used, there is the possibility that folding actually occurred during the dead time of the kinetic experiment. Furthermore, a very slow phase of small amplitude could just reflect instrumental drift. These and other interpretation uncertainties plagued the *in vitro* protein folding field since its beginnings. However, pioneers of the field found reliable ways around these problems on the basis of carefully designed ‘jump assays’ in which protein samples are extracted at certain times and transferred to solutions of selected composition for experimental assessment (see, for instance [13,14]). Here, we have specifically used double-jump unfolding assays, a methodology that aims at providing a direct determination of the amount of native protein [28,29]. The rationale behind this approach is that the unfolding of the native state of a protein is much slower than the unfolding of non-native or intermediate states. The amount of native state in a protein solution can then be determined from the unfolding kinetics followed in the appropriate time scale after transfer to denaturing conditions. Obviously, unfolding assays exploit the high activation free energy barrier for unfolding to determine the amount of native protein, i.e. they exploit the free energy barrier that confers kinetic stability to the native protein [30,31]. They are, therefore, particularly appropriate for this work because following folding kinetics using double-jump unfolding assays does define the time scale required for the development of kinetic stabilization. That is, they define the time span in which the unassisted folding chain is susceptible to undesirable interactions and alterations, which is a parameter of direct evolutionary significance.

For most the thioredoxin variants studied here, we have followed the folding kinetics under selected conditions by carrying out unfolding assays at different times after transfer of a denatured protein to native conditions, in such a way that folding kinetic profiles of the amount of native state versus time are obtained. In a typical experiment (see Supplementary Figure S4 for a representative example), we used a concentrated solution of unfolded protein in ~4 M guanidine (*E. coli* thioredoxin and its variants) or ~5 M guanidine (LBCA thioredoxin, LPBCA thioredoxin and their variants) and we started the folding process by a suitable dilution (within the 2–10-fold range) into a low-concentration guanidine solution to reach a final protein concentration on the order of 1 mg/ml. At given times, aliquots were extracted and transferred (20-fold dilution) to ~3 M guanidine for *E. coli* thioredoxin and its variants or to ~5 M guanidine for the ancestral thioredoxins and their variants, and the unfolding kinetics were determined by fluorescence. The fraction of native state (X_N) versus time (t) profile for folding at low denaturant concentration is easily obtained from the amplitude of the unfolding kinetic phase using a suitable control experiment (Supplementary Figure S4). Supplementary Figure S5 shows several representative examples of X_N versus t profiles that illustrate the strong effect of the S/G exchange at position 74 on the folding rate. In all cases, these profiles could be well described by single exponentials, with initial time and long-time values close to zero and unity. This indicates that these profiles probe the relevant kinetic phase of the major folding channel. Of course, it cannot be ruled out that, in several cases, small amounts of protein reach the native state through faster or slower channels, since the initial time and long-time values of the profiles actually differ somewhat from zero and unity (see also Figure 6b).

Comparison of the folding profiles from double-jump assays (X_N versus t) with those obtained using steady-state fluorescence revealed three different scenarios: (1) fluorescence profiles could be well fitted by a single exponential and the rate constant derived from such fits agreed with the value obtained from the X_N versus t profiles (see Supplementary Figure S6 for an illustrative example). (2) Two exponentials were required to fit the fluorescence folding profiles and it was the rate constant from the faster, larger amplitude phase that agreed with the value obtained from the X_N versus t profiles (see Supplementary Figure S7 for an illustrative example). (3) Two exponentials of roughly similar amplitude were required to fit the fluorescence folding profiles and it was the rate constant from the slower phase that agreed with the value obtained from the X_N versus t profiles (see Supplementary Figure S8 for an illustrative example).

Finally, it is important to note that following folding kinetics through double-jump unfolding assays is considerably time consuming. Therefore, our approach has been to carry out extensive folding kinetic studies on the basis of steady-state fluorescence measurements and to determine only a limited number of double-jump X_N versus t profiles in order to identify in the fluorescence profiles the relevant kinetic phase of the major folding channel. For the sake of simplicity and clarity, folding branches of the chevron plots given in Figures 3–5 and Supplementary Figure S10 show only the rate constants for such relevant kinetic phase and do

not differentiate between data derived from fluorescence profiles and data derived from double-jump X_N versus t profiles. However, in Supplementary Figure S9 we provide Chevron plots that include the comparison between rate constants values derived from fluorescence profiles and from double-jump X_N versus t profiles. Also, Figure 6b shows profiles of folding followed by double-jump unfolding assays for several modern bacterial thioredoxins.

Results and discussion

Modern versus ancestral thioredoxin folding

We first compared the folding of modern *E. coli* thioredoxin with that of two of its resurrected Precambrian ancestors (Figure 1a): the thioredoxins encoded by the reconstructed sequences for the last bacterial common ancestor (LBCA thioredoxin) and the last common ancestor of the cyanobacterial, deinococcus and thermus groups (LPBCA thioredoxin). These two phylogenetic nodes correspond to organisms that existed ~4 and 2.5 billion years ago, respectively [18,32,33]. We have previously characterized LPBCA and LBCA thioredoxins, as well as several other resurrected Precambrian thioredoxins, in detail [18–20,33–35]. They are properly folded, highly stable, active enzymes that share essentially an identical three dimensional (3D)-structure with *E. coli* thioredoxin (Figure 2), despite their low sequence identity with the modern protein (for sequences and structures, see Figure 1 in [33]). The structures of the three proteins under study bear a proline residue at position 76 in *cis* conformation that is strictly conserved in thioredoxins.

Figure 3 shows chevron plots of rate constant versus denaturant concentration for *E. coli* thioredoxin and the ancestral LPBCA and LBCA thioredoxins. These plots include folding and unfolding branches. We have used guanidine, a strong denaturant, for the experiments in Figure 3 in order to achieve denaturation of the highly stable ancestral thioredoxins. However, urea, a weaker denaturant, is used in other experiments reported in this work but we show that the choice of denaturant does not affect our conclusions. We specifically define folding rates in terms of the relevant kinetic phase of the major folding channel as identified by double-jump unfolding assays (see Materials and methods for details).

As it is clear from Figure 3a, the *in vitro* folding of the ancestral thioredoxins is substantially faster than the folding of their modern *E. coli* counterpart. Rate constants reported in this figure have been determined at 25°C for both the modern and ancestral proteins, while high environmental temperatures (~ 65–80°C) have been suggested for the Archaeon [36]. Experimental determinations of the temperature dependence of the rate constants (Supplementary Figure S11) show that consideration of the different living temperatures of the modern and ancient hosts does not eliminate the folding rate difference. On the contrary, the folding of the ancestral LPBCA and LBCA thioredoxins at the proposed Archaeon temperatures is actually around two orders of magnitude faster than the folding of *E. coli* thioredoxin at the optimal living temperature (37°C) of *E. coli* (Supplementary Figure S11).

For the three proteins studied, the folding rate is determined by the presence of a *cis*-proline at position 76, as shown by the observation that the folding rate is considerably increased by mutating proline 76 to alanine (compare Figure 3a,b). Of course, replacing proline at position 76 impairs activity [12], as shown by assays based on the aggregation of insulin (Figure 3c and Supplementary Table S1) as well as assays based on the interaction of thioredoxin with thioredoxin reductase (Supplementary Table S2). Interestingly, in the latter assays, the activity of variants of the most ancient LBCA thioredoxin appear depressed (Supplementary Table S2), possibly reflecting the consequences of coevolution between thioredoxin and thioredoxin reductase [37]. Still, the overall picture is that replacing P76 impairs activity, which explains why proline is conserved at position 76 in thioredoxins. Another critical point to note here is that the folding rates are similar for the three proteins when alanine is present at position 76 and diverge when proline is present at the position. Therefore, the slower folding of *E. coli* thioredoxin is attributed to mutational differences with the ancestral proteins that aggravate the kinetic bottleneck created by proline 76. Our efforts to identify such degrading mutations are described in the next section.

Mutational basis for the slow folding of *E. coli* thioredoxin as compared with the ancestral LPBCA and LBCA thioredoxins

The sequence of *E. coli* thioredoxin and the ancestral thioredoxins studied here differ at ~40–50 amino acid positions for a protein of ~110 residues [18,33] and, in principle, many different mutations could be responsible for the slower folding of the modern protein. Still, sequence differences in the neighbourhood of pro76

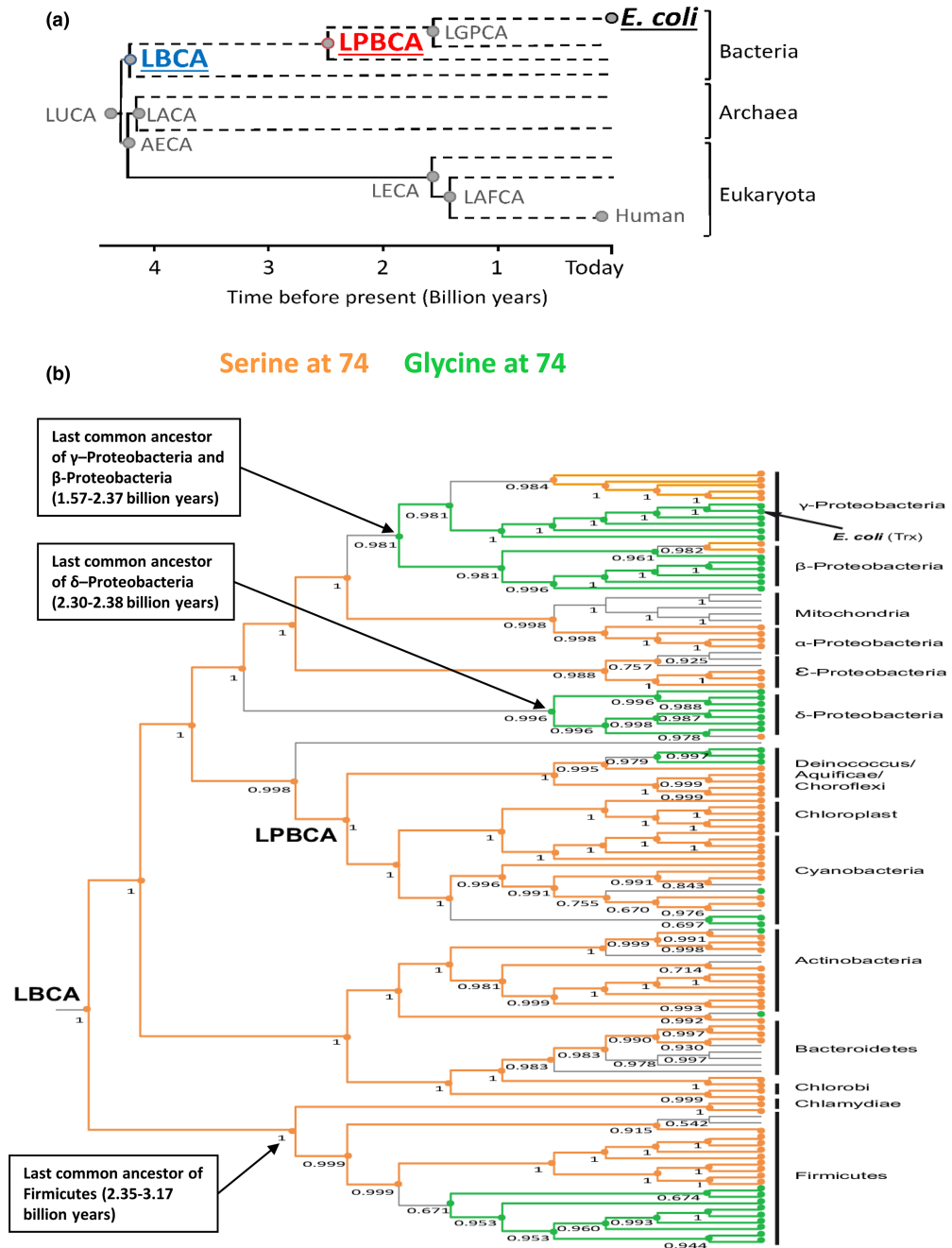


Figure 1. Thioredoxin phylogenetic tree used for ancestral sequence reconstruction [18].

(a) Schematic representation of the phylogenetic tree showing geological time. The nodes studied in this work are underlined and are defined in the text. See [18] for the definition of other nodes. (b) Bacterial section of the tree showing the evolutionary history of the amino acid (serine or glycine) present at position 74. The ages provided for some key nodes are taken from the *Timetree of Life* [32]. Numbers alongside the nodes stand for the posterior probability of the most likely residue (orange circles for serine and green circles for glycine). Most 74S residues in the modern thioredoxins are linked to the serine residue in the last common bacterial ancestor (LBCA node) through evolutionary trajectories that display serine conservation; these trajectories are highlighted in orange. Most 74G residues in modern thioredoxins can be linked to previous S74G replacements through trajectories that display glycine conservation; these trajectories are highlighted in green. The overall pattern suggests entrenchment related to coevolution with the many macromolecular partners of thioredoxin [10,11] (see Supplementary discussion in Supplementary material for details).

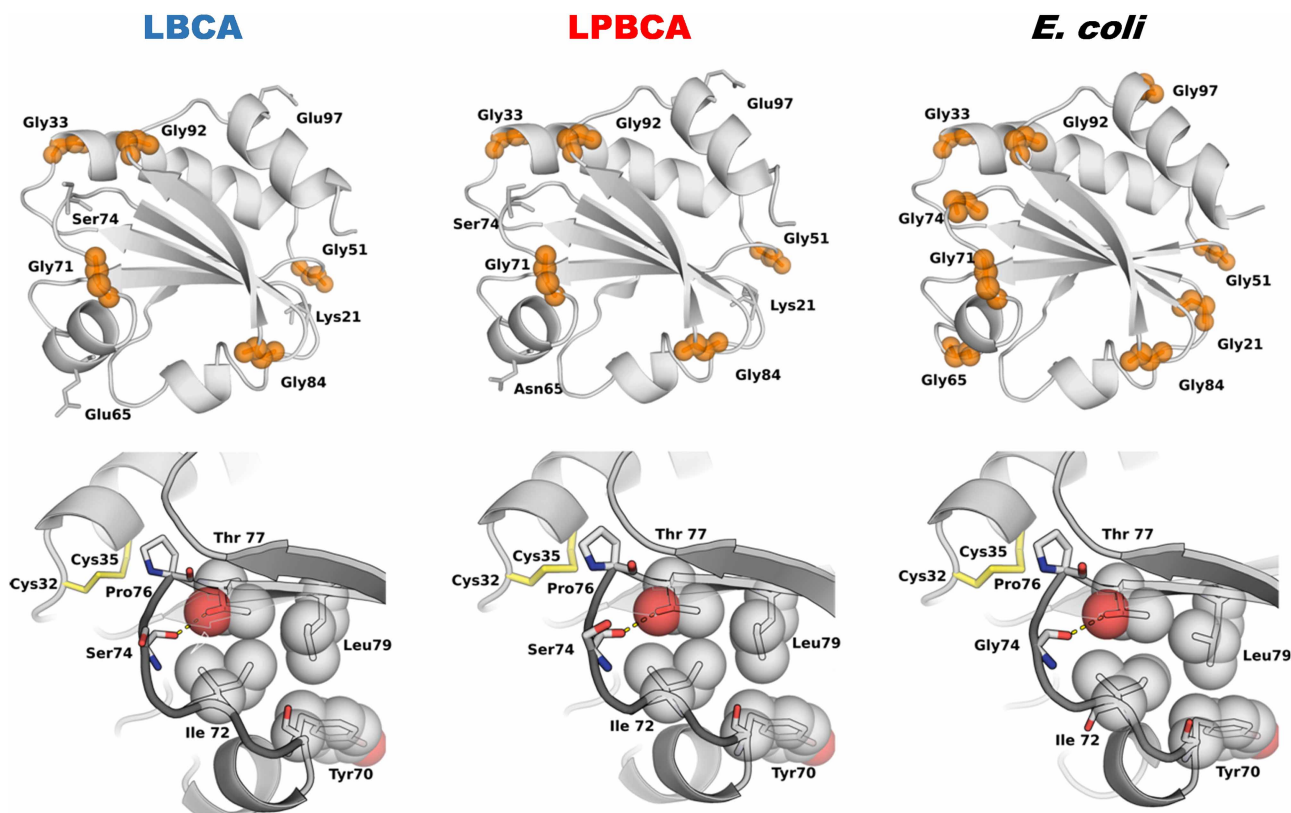


Figure 2. 3D-structures of modern and ancestral thioredoxins.

3D-structures of *E. coli* thioredoxin (modern, PDB code 2TRX), and LPBCA thioredoxin (ancestral, ~2.5 billion years, PDB code 2YJ7) and LBCA thioredoxin (ancestral, ~4 billion years, PDB code 4BA7) are shown. Positions with glycine residues in the modern protein are labelled. Note that the modern and ancestral thioredoxins share the thioredoxin fold, despite low sequence identity [33]. Blow-ups of the region including the *cis*-pro76 are shown below the full structures. The active-site disulfide bridge, the Gly/Ser residues at position 74 and residues that presumably stabilize the loop that includes Pro76 are highlighted.

should provide obvious candidates and one such difference stands out (Figure 2): serine is the residue at position 74 in LBCA and LPBCA thioredoxins, while glycine is the residue at position 74 in *E. coli* thioredoxin. Mutational analyses show that the S/G exchange at position 74 indeed accounts for most of the observed folding rate difference between the modern and the ancestral proteins (Figure 4a). Replacement of the ancestral residue at position 74 (serine) with glycine thus slows down folding in the ancestral LBCA and LPBCA thioredoxins, while the back-to-the-ancestor G74S in *E. coli* thioredoxin increases the rate of folding.

Many experiments support the robustness of our identification of S74G as a folding-degrading mutation. First, the folding rate enhancement obtained upon the reverse, back-to-the-ancestor G74S replacement in *E. coli* thioredoxin is reproduced when the active-site disulfide has been reduced and also when using urea, instead of guanidine, as denaturant (Figure 4b). In addition, *E. coli* thioredoxin actually has four additional glycine residues (at positions 21, 65, 74 and 97) with respect to the ancestral LPBCA and LBCA thioredoxins (Figure 2). However, extensive mutational studies (Supplementary Figure S10) indicate that it is only the glycine/ancestral-state replacement at position 74 that affects the folding rate.

A mutation (S74G) that aggravated the folding problem created by the active site *cis*-proline occurs in the line of descent that led to *E. coli* thioredoxin

As discussed above, the effect of the G/S exchange on the thioredoxin folding rate is experimentally robust. However, while the fact that glycine is the modern residue at position 74 (the residue present in *E. coli*

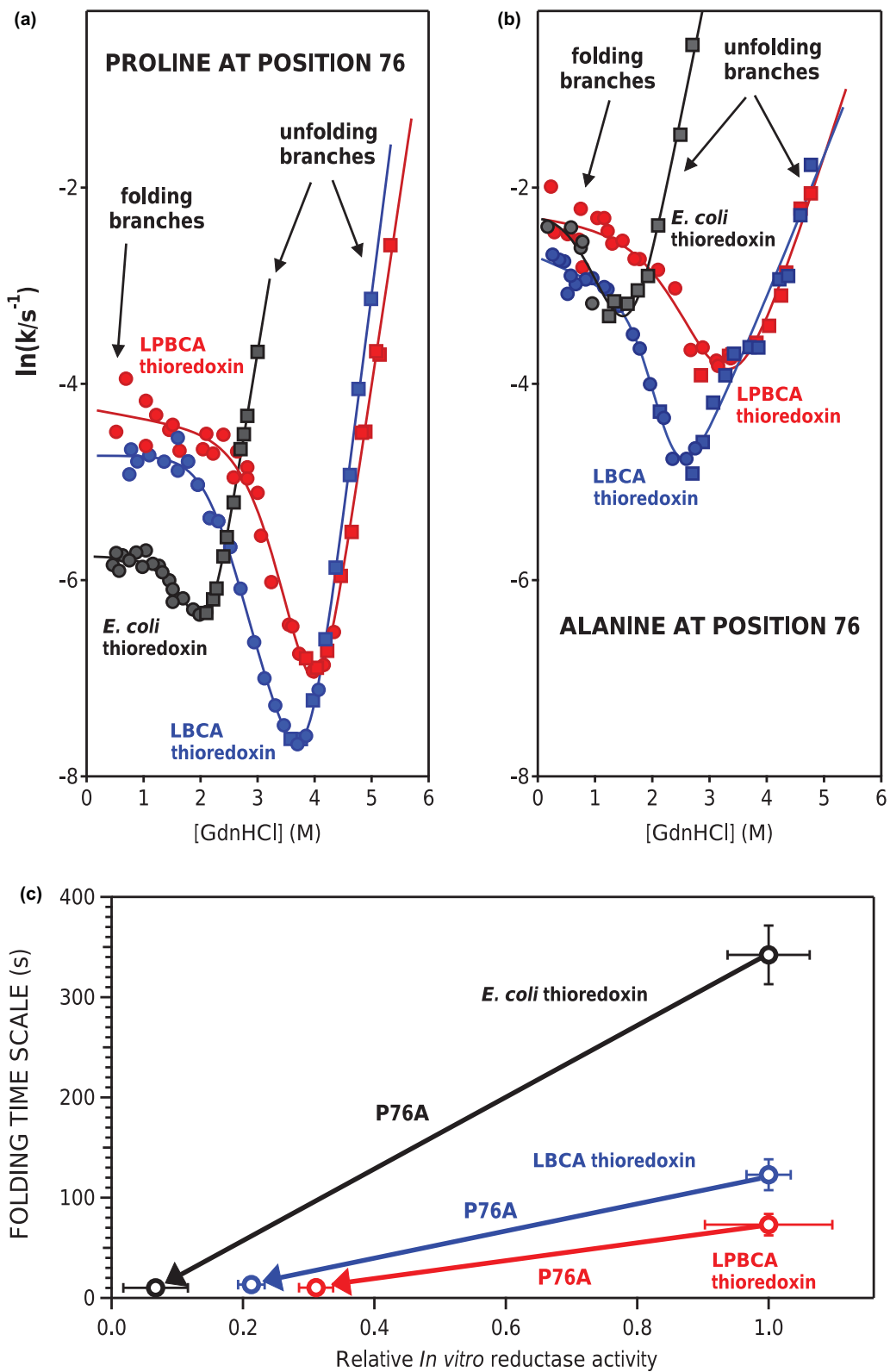


Figure 3. Folding–unfolding rates for *E. coli* thioredoxin and two resurrected Precambrian thioredoxins (see Figure 1).

Part 1 of 2

(a) Chevron plots (including folding and unfolding branches) of the logarithm of the rate constant versus guanidine concentration for the ‘wild-type’ proteins that display the conserved proline at position 76. Circles and squares refer to the data

Figure 3. Folding–unfolding rates for *E. coli* thioredoxin and two resurrected Precambrian thioredoxins (see Figure 1).

Part 2 of 2

obtained in the folding and unfolding directions, respectively. (b) same as (a) except for the P76A variants of the three proteins. (c) Plot of the time scale for folding versus *in vitro* reductase activity for the modern *E. coli* thioredoxin and the ancestral LPBCA and LBCA thioredoxins. Values of the folding time scale are calculated as the inverse of the folding rate constant extrapolated to zero denaturant concentration (see Materials and methods for details). Data for the ‘wild-type’ forms and the P76A variants included in this plot are connected by arrows in order to highlight the effect of the mutation and the function–folding trade-off: eliminating the proline at position 76 accelerates folding, but impair function. Error bars are not shown when they are smaller than the data point.

thioredoxin) is an observable result, the identification of serine as the ancestral residue is a statistical inference. There could be some doubt, therefore, that the S74G replacement actually occurred in the line of descent that led to *E. coli* thioredoxin. This is of particular interest given discussions from the literature [38–41] that ancestral sequence reconstruction can potentially be biased from uncertainties in the process. In this case, however,

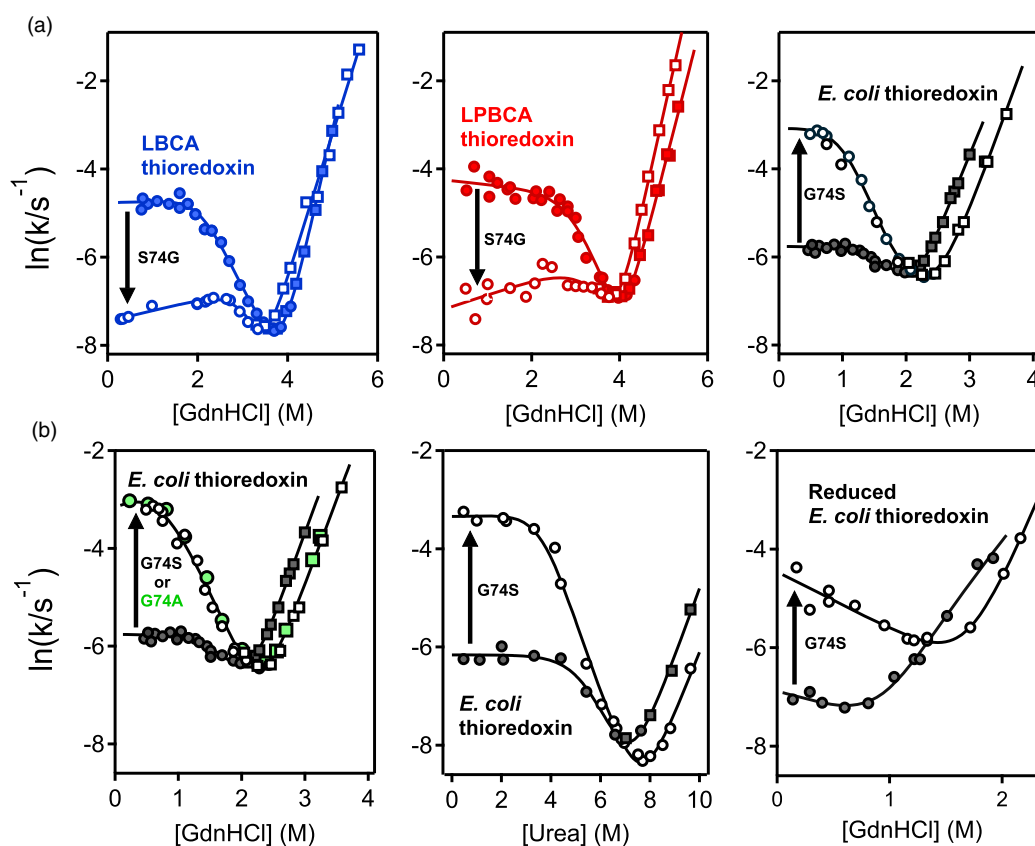


Figure 4. Effect of the G/S exchange at position 74 on thioredoxin folding rate.

Chevron plots of the logarithm of folding–unfolding rate versus guanidine concentration are shown for the ‘wild-type’ forms (closed data points) and variants (open data points). Circles and squares refer to the data obtained in the folding and unfolding directions, respectively. (a) Comparison between the modern *E. coli* thioredoxin and the ancestral LBCA and LPBCA thioredoxins. Note that glycine is the wild-type residue in *E. coli* thioredoxin, while serine is ‘wild-type’ in the ancestral thioredoxins. The effects of the mutation that replaces the wild-type are highlighted with an arrow. (b) Effect of the S74G mutation on *E. coli* thioredoxin folding and unfolding rates. Data obtained using urea and guanidine as denaturants are shown. In the latter case, data obtained using thioredoxin with a reduced active-site disulfide are also included. Note that the panel at the left also includes the effect of a G74A mutation (see text for details).

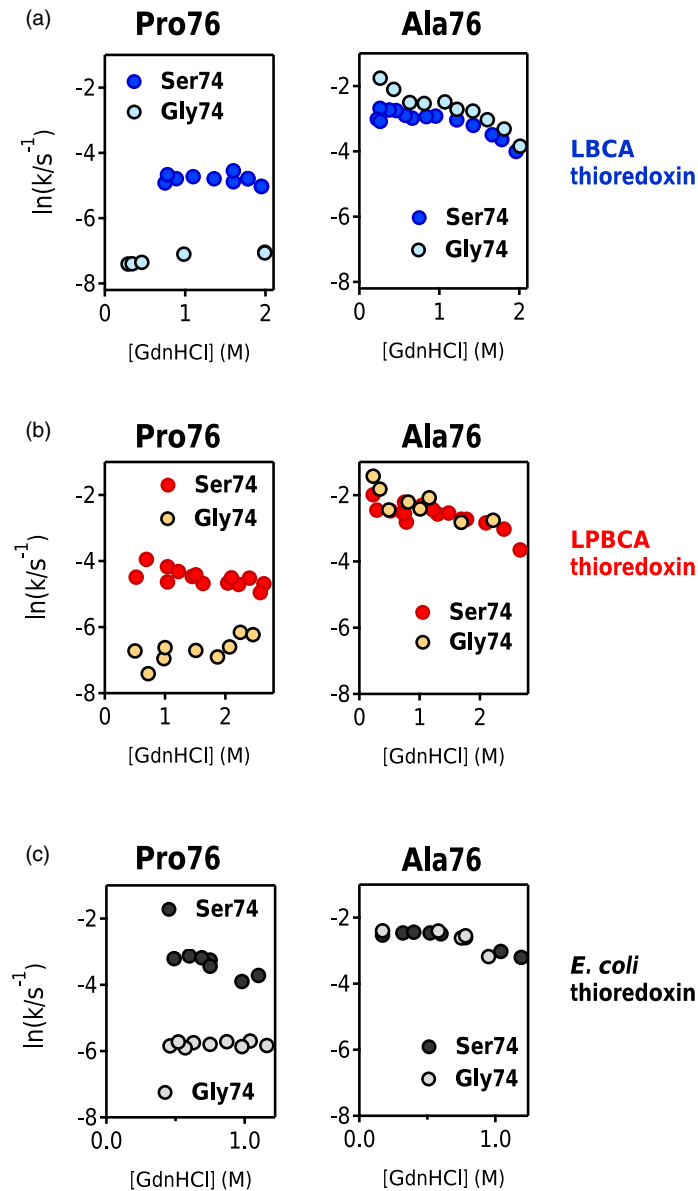


Figure 5. Double-mutant cycle analysis on the folding rate coupling between positions 74 and 76 in thioredoxins.

Folding rates for variants of: (a) ancestral LBCA thioredoxin; (b) ancestral LPBCA thioredoxin; (c) modern *E. coli* thioredoxin. Data for thioredoxin variants including P/A exchange at position 76, G/S exchange at position 74 and the combination of the two exchanges are shown. The plots shown are meant to make visually apparent the effect of the G74S mutation is substantial when proline is present at position 76 (plots at the left) but very low when alanine is present at position 76 (plots at the right). Note that glycine is the wild-type residue in *E. coli* thioredoxin, while serine is ‘wild-type’ in the ancestral thioredoxins. Proline is the wild-type residue at position 76 in all thioredoxins.

the identification of serine as the ancestral state at position 74 is quite robust. This follows first from the observation that serine is the consensus residue (i.e. the most frequent residue) at position 74 in modern thioredoxins (Figure 1b). Of course, discrepancies between consensus sequences and reconstructed ancestral sequences do exist and may have a phenotypic impact. Yet, as we have recently discussed [42], these discrepancies are typically restricted to positions at sites with a high sequence diversity and, consequently, high evolutionary rates. This does not appear to be the case for position 74 which is populated mainly by serine and glycine residues in modern bacterial thioredoxins (Figure 1b). Furthermore, the Bayesian posterior probabilities for the inferred

residues in both ancestors at position 74 is 100% (Figure 1b). We have previously shown that such sites rarely, if ever, are incorrectly inferred with such a high posterior probability [43].

Secondly, the S74G mutation decreases the folding rate of the ancestral LPBCA and LBCA thioredoxins by about one order of magnitude and the back-to-the-inferred-ancestor mutation G74S increases the folding rate of *E. coli* thioredoxin by about one order of magnitude (Figure 4 and Supplementary Figure S10). Therefore, the effect of the S/G exchange at position 74 on the folding rate is to a large extent independent of the background sequence (modern or ancestral). This implies that the mutational effect is reasonably robust against reconstruction uncertainties in other positions of the thioredoxin molecule.

Finally, the link between the effect of the S/G replacement at position 74 on folding kinetics and the *cis*-proline at position 76 is immediately revealed by a double-mutant cycle analysis of the coupling between positions 74 and 76 on the three thioredoxins studied (Figure 5). The S/G replacement at position 74 thus strongly affects the folding rate only when proline is at position 76 and not when pro76 has been replaced with alanine. Clearly, the mutation S74G did occur in the line of descent that led to *E. coli* thioredoxin and aggravated the folding problem created by the active site *cis*-proline.

Experimental study of a set of modern bacterial thioredoxins shows that folding rates are not evolutionarily conserved

The effect of the S/G exchange at position 74 on the thioredoxin folding rate is likely related to the fact that glycine residues have no side-chains which place little restriction on backbone dihedral angles and generate flexible links in polypeptide chains [44]. The 3D-structures of the modern and ancestral thioredoxins studied so far (Figure 2) reveal interactions that appear to stabilize the 70–79 segment in the conformation imposed by the *cis*-Pro76 (Figure 2, lower panel), namely hydrophobic contacts and a hydrogen bond between the backbone carbonyl of residue 74 (either G or S) and Thr77. Besides stabilizing the native protein structure with a *cis*-proline at position 76, these interactions should also favour local residual structures in the high energy regions of the folding landscape that favour the *cis* conformer, thus promoting correct folding. However, the flexible link generated by glycine residue at position 74 should allow many alternative conformations for the 70–79 segment to occur in the upper (high energy) regions of the folding landscape thus slowing down folding. This interpretation is strongly supported by the fact that replacing glycine at position 74 in *E. coli* thioredoxin with alanine brings about an increase in folding rate that is essentially identical with that produced by the G74S (left panel in Figure 4b), thus indicating folding rate enhancement is caused by the elimination of flexibility at position 74, either through the G74A mutation or through the G74S mutation.

The interpretation proposed above suggests that other amino acid replacements in the neighbourhood of position 74 could also impact the folding rate. In particular, a visual inspection of the 3D-structures (Figure 2) points to the residues at positions 70, 72, 74, 77 and 79 as being involved in interactions that could plausibly modulate the stability of the 70–79 in upper regions of the folding landscape. We, therefore, used these positions to guide the selection of a set of modern bacterial thioredoxins for experimental characterization. We performed a search in the NCBI Reference Sequence Database using the sequence of *E. coli* thioredoxin as query and considered the ~5000 top hits. A substantial fraction of these sequences displayed differences with *E. coli* thioredoxin at positions 70, 72, 74, 77 and 79. We selected a small subset of modern thioredoxins to capture this sequence diversity in a meaningful way (Supplementary Figure S12). That is, for some of the proteins in the subset, most residues at positions 70, 72, 74, 77 and 79 are the same as in the ancestral LBCA or LPBCA thioredoxins (including the presence of serine at position 74). Yet, other proteins in the subset differ at several of the selected positions from the sequences of ancestral LBCA and LPBCA thioredoxins as well as from the sequence of *E. coli* thioredoxin. In all cases, the thioredoxin selected displayed the highest sequence identity with *E. coli* thioredoxin, given the amino acid residues present at positions 70, 72, 74, 77 and 79. The selected thioredoxins show similar activities in the insulin aggregation assay (Supplementary Table S3).

Folding rates for the 14 modern proteins in the subset determined using urea as denaturant (Figure 6a) span a ~100-fold range, a result which is confirmed by double-jump unfolding experiments that directly probe the amount of native protein (Figure 6b, see Materials and methods for details). Of course, we cannot rule out that a substantial part of this observed folding rate variation is due to mutational changes outside the five positions we have used to guide the sequence selection. This, however, would not affect in the least the main implication of the data, namely that, contrary to what has been claimed in recent literature [16,17], *in vitro* thioredoxin

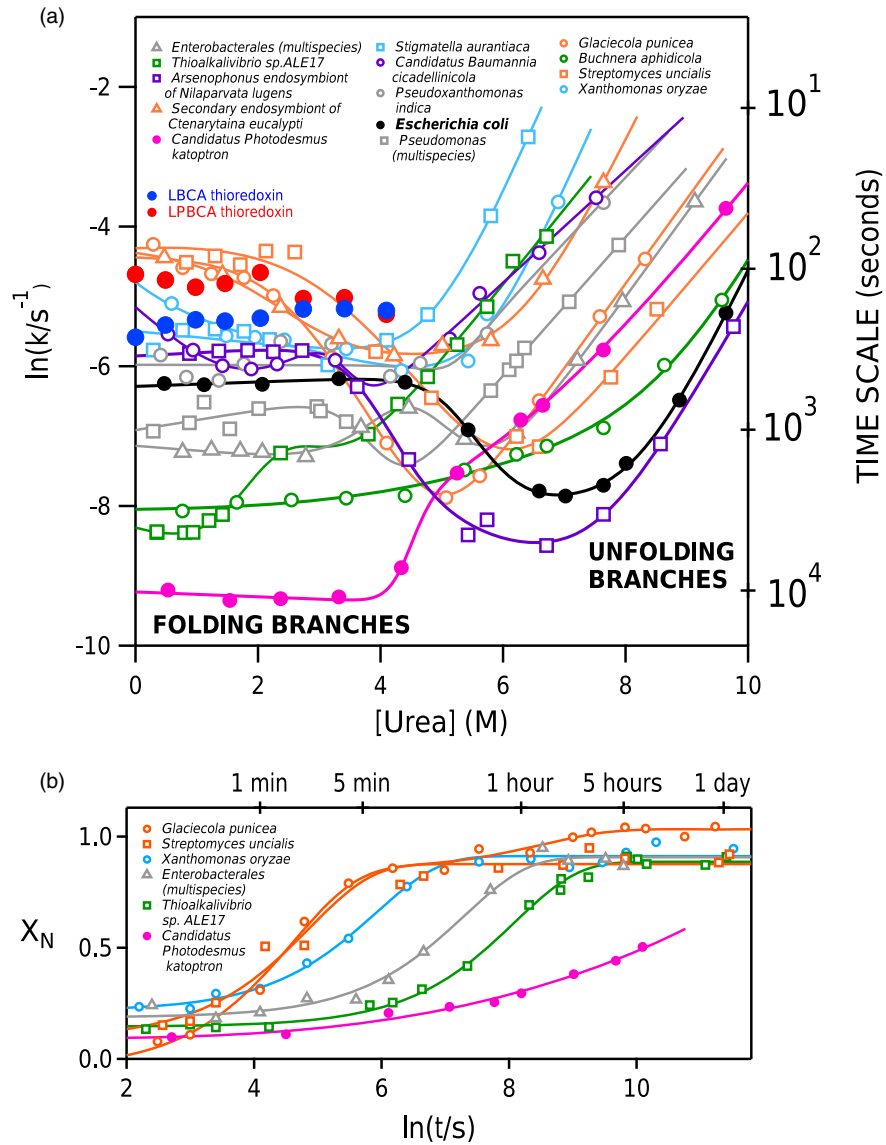


Figure 6. Folding rates are not conserved in the thioredoxin family.

(a) Chevron plots of folding–unfolding rate constant versus urea concentration for a set of 14 modern bacterial thioredoxins. Folding rates given correspond to the major slow phase of the fluorescence kinetic profiles (see Materials and methods for details). The time scale shown in the right axis is calculated as the inverse of the rate constant. For comparison, experimental folding data for LBCA and LPBCA thioredoxins are also included. Given the high stability of the ancestral proteins, their folding rates in urea solutions were obtained by denaturation in guanidine followed by high-dilution transfer to urea solutions, in such a way that the final guanidine concentration was low (~ 0.1 M). Note that the folding rate data in this plot span about two orders of magnitude. (b) Folding in 1 M urea as followed by double-jump unfolding assays, a methodology that allows a determination of the fraction of native protein. The profiles shown reveal the major folding channel (see Materials and methods for details) and are consistent with the range of folding times of the data shown in panel (a).

folding rates are not conserved, with some modern thioredoxins folding substantially faster and substantially slower than *E. coli* thioredoxin (closed black data points in Figure 6a).

We have also included in Figure 6a rate data for the ancestral LBCA and LPBCA thioredoxins in urea solutions. Clearly, while the ancestral proteins and some of the modern proteins studied fold in the approximately minute time scale, folding of some other modern thioredoxins occurs in the much slower approximately hour time scale.

Folding on-rates do not correlate with stability in the thioredoxin family

Unfolding rates, as given by the unfolded branches of the chevron plots in Figure 6a, already provide a useful metric of kinetic stability [30,31]. To assess the thermodynamic stability of the modern proteins studied we fitted equilibrium profiles of fluorescence intensity versus urea concentration on the basis of the linear extrapolation model (see Material and methods for details). Fits were visually excellent (Figure 7a) and allowed us to derive values of a known metric of thermodynamic stability: the urea concentration at which unfolding free

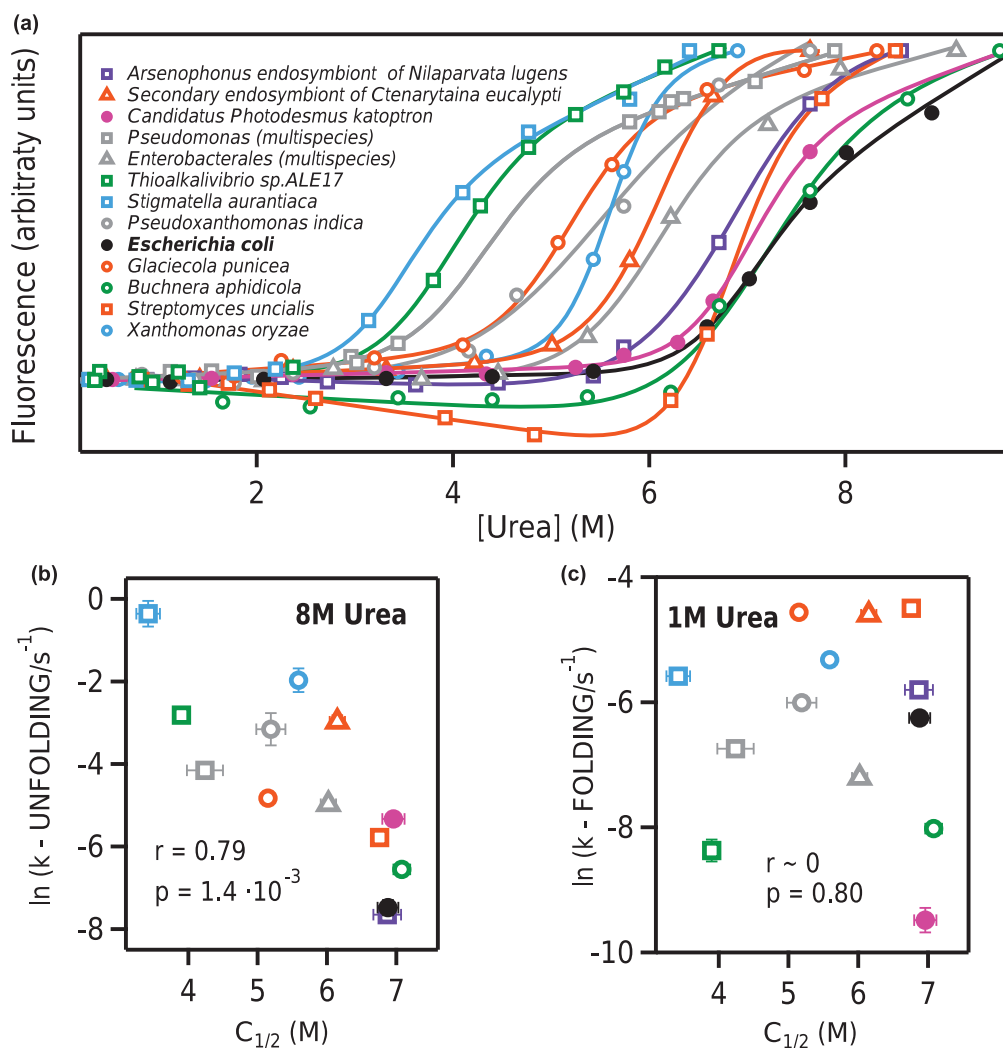


Figure 7. Folding on-rates do not correlate with stability in the thioredoxin family.

(a) Equilibrium profiles of fluorescence intensity versus urea concentration for modern thioredoxins. Lines represent the best fits of a model that assumes a linear dependence of the unfolding free energy with denaturant concentration within the narrow transition range and uses linear pre- and post-transition baselines (see Materials and methods for details and Supplementary Table S4 for the values obtained for the fitting parameters). (b and c) Plots of the logarithm of the unfolding rate constant (b) and folding rate constant (c) versus urea midpoint concentration (the urea concentration at which the unfolding free energy is zero). The values of the folding and unfolding rate constants are interpolated (for 1 M and 8 M urea concentration, respectively) from the corresponding branches of the Chevron plots. Associated errors are not shown when they are smaller than the size of the data points. Values of the correlation coefficient (r) and the probability that the correlation observed occurs by chance (p) are shown. Unfolding rates (kinetic stability) correlate with thermodynamic stability (b) but folding rates do not (c) (see the text for details).

energy is zero. These $C_{1/2}$ values, as well as the m slopes that measure the urea-concentration dependence of the unfolding free energy, are given in Supplementary Table S4 for the modern thioredoxins.

Unfolding rates do show some correlation with $C_{1/2}$ values (Figure 7b). This association between kinetic and thermodynamic stabilities is to be expected from the fact that the transition state for thioredoxin unfolding is substantially unstructured, as we have previously noted and discussed [20,24]

Plotting the logarithm of folding rate versus $C_{1/2}$ produces a scattergram (Figure 7c). This lack of correlation, however, should not come as a surprise, as kinetic complexity and the presence of kinetic folding intermediates should disconnect folding rates from stability. Uncoupling of kinetics and thermodynamics was already advanced by Agard and co-workers in their seminal work on the kinetic stability of α -lytic proteases [45]. The thioredoxin family provides an example of a phenomenon, folding rate/stability uncoupling, that is likely to be widespread, given that the *in vitro* folding of many proteins, even small ones, is kinetically complex, as it has been known for many years [46].

Conclusions

It has been recently claimed that thioredoxin folding rates as determined *in vitro* are evolutionarily conserved [16,17]. This supposed conservation was furthermore taken as the first experimental evidence of a cornerstone of protein folding theory: the principle of minimal frustration. In very simple terms, the folding landscape was optimized (minimally frustrated) at a very early stage and remained so over billions of years of evolutionary history leading to folding rate conservation among modern proteins. As elaborated below in some detail, however, this proposal is inconsistent not only with well-known principles of evolutionary theory, but also with our current understanding of folding processes *in vivo*.

Proteins do not evolve in isolation because they are involved *in vivo* in a wide diversity of interactions [47]. In particular, protein folding within modern organisms relies on exceedingly complex intermolecular interactions that guide and assist the process [4–9]. Folding *in vivo* occurs co-translationally, local folding events may already take place within the ribosome exit tunnel and folding may be coupled to translation kinetics. Nascent chains are involved in many interactions as they emerge from the ribosomal tunnel, including interactions with the trigger factor, a protein that binds to exposed hydrophobic segments. The trigger factor is the first of many specialized molecules (folding chaperones) that, together with the modern ribosome, assist protein folding *in vivo*. As a result of the numerous intermolecular interactions involved, the conformational space explored by a folding chain within a modern organism *in vivo* may differ substantially from the conformational space that the folding chain explores *in vitro* [8]. Therefore, unassisted protein folding, as probed by *in vitro* studies, does not necessarily correlate with the biologically-relevant assisted folding that takes place within modern organisms.

Of course, it is conceivable that at a very early evolutionary stage, prior to the emergence of folding assistance, folding efficiency relied on fast folding that minimized the time scale the polypeptide chain spent in partially folded states which are susceptible to aggregation and other undesirable interactions. That is, fast unassisted folding, linked to a landscape with low (perhaps minimal) frustration, may have been required at a primordial stage. However, once folding assistance was available, mutations that impaired unassisted folding could be accepted. That is, as it is common in morphological evolution, a feature that it is no longer useful undergoes evolutionary degradation. However, while degradation at the morphological level may often be visually apparent, degradation of unassisted folding can only be revealed by *in vitro* folding experiments, since folding within modern cells is assisted.

Indeed, the *in vitro* experiments reported here are consistent with the evolutionary degradation of unassisted folding. We have unambiguously identified a mutation that substantially slows down *in vitro* thioredoxin folding and that was accepted in the line of descent that led to *E. coli* thioredoxin. Furthermore, we have shown that, while resurrected Precambrian thioredoxins and some modern thioredoxins fold *in vitro* in the approximately minute time scale, other modern thioredoxins approach the approximately hours time scale. Such variation in the folding rate, indicating different degrees of degradation, should not come as surprise. As Darwin already advanced in the first chapter of *The Descent of Man*, the variability of superfluous features that are not under natural selection should be a common observation in morphological studies. Indeed, such type of variability is often illustrated with the widely variable size of the human appendix [1] while here we have provided an example at the molecular level. It is also worth noting that folding in the approximately hours time scale is hardly of any biological significance and that assisted folding *in vivo* is likely to be much faster. Overall, it is clear that *in vitro* thioredoxin folding rates are not evolutionarily conserved. As we have previously

noted [20], recent claims to the contrary [16,17] are probably related to the use of destabilizing conditions which buffer the effect of landscape ruggedness on *in vitro* folding experiments, as it has been known for many years [48] and it is visually apparent in our data of Figure 6a. Note that the folding rates are roughly similar for most thioredoxins at 4 M urea, while they diverge as the denaturant concentration becomes lower and the solvent becomes less destabilizing. Furthermore, the very acidic pH (=2) employed in previous studies [16] brings about extensive protonation of residues and alterations of ionic interactions that are unlikely to be physiologically and evolutionary relevant [20].

Finally, beyond clearing up a relevant and consequential controversy in recent literature, this work has general implications of wide interest. It points to a simple evolutionary interpretation of *in vitro* protein folding as a degraded version of primordial unassisted folding and thus may contribute to clarify the much-debated issue of the relation between protein folding *in vivo* and protein folding *in vitro*. More generally, this work provides evidence that degradation shapes evolution not only at the morphological level, but also at the level of individual enzymes.

Abbreviations

DTNB, 5,5'-dithiobis(2-nitrobenzoic acid); LPBCA, last common ancestor of the cyanobacterial, deinococcus and thermus groups; LBCA, last bacterial common ancestor.

Authors Contributions

G.G.-A. purified the several variants of modern and ancestral thioredoxins; she also performed and analyzed the experiments aimed at determining their folding–unfolding kinetics. V.A.R. performed bioinformatics analyses and provided essential input for the evolutionary interpretation of the data. A.I.-P. performed preliminary experiments that pointed to a crucial role of glycine residues. A.M.C. and M.L.R.-R. performed preliminary experiments that pointed to the enhanced folding and unfolding rates for the ancestral thioredoxins. E.A.G. assessed the robustness of the reconstructed ancestral states and provided essential input for the evolutionary interpretation of the data. J.A.G. provided essential input for the structural analysis of the mutational effects on folding rates. B.I.-M. designed and supervised the folding–unfolding kinetic experiments. B.I.-M. and J.M.S.-R. directed the project. J.M.S.-R. provided the general evolutionary interpretation and wrote the paper. All authors read the manuscript and provided useful comments.

Funding

This research was supported by FEDER Funds, grant BIO2015-66426-R from the Spanish Ministry of Economy and Competitiveness (J.M.S.-R.), grant RGP0041/2017 from the Human Frontier Science Program (J.M.S.-R. and E.A.G.) and National Institutes of Health 1R01AR069137 (E.A.G.), Department of Defence MURI W911NF-16-1-0372 (E.A.G.).

Competing Interests

The authors declare that there are no competing interests associated with the manuscript.

References

- 1 Coyne, J.A. (2009) *Why Evolution is True*, Ch. 3, Oxford Univ. Press, New York, NY
- 2 Dawkins, R. (2009) *The Greatest Show on Earth: the Evidence for Evolution*, Ch. 11, Bantam Press, London, U.K.
- 3 Kratzer, J.T., Lanaspá, M.A., Murphy, M.N., Cicerchi, C., Graves, C.L., Tipton, P.A. et al. (2014) Evolutionary history and metabolic insights of ancient mammalian uricases. *Proc. Natl. Acad. Sci. U.S.A.* **111**, 3763–3768 <https://doi.org/10.1073/pnas.1320393111>
- 4 Kaiser, C.M., Goldman, D.H., Chodera, J.D., Tinoco, I. and Bustamante, C. (2011) The ribosome modulates nascent protein folding. *Science* **334**, 1723–1727 <https://doi.org/10.1126/science.1209740>
- 5 Oh, E., Becker, A.H., Sandikci, A., Huber, D., Chaba, R., Gloge, F. et al. (2011) Selective ribosome profiling reveals the cotranslational chaperone action of trigger factor *in vivo*. *Cell* **147**, 1295–1308 <https://doi.org/10.1016/j.cell.2011.10.044>
- 6 Zhang, G. and Ignatova, Z. (2011) Folding at the birth of the nascent chain: coordinating translation with co-translational folding. *Curr. Opin. Struct. Biol.* **21**, 25–31 <https://doi.org/10.1016/j.sbi.2010.10.008>
- 7 Kim, Y.E., Hipp, M.S., Bracher, A., Hayer-Hartl, M. and Hartl, F.U. (2013) Molecular chaperone functions in protein folding and proteostasis. *Annu. Rev. Biochem.* **82**, 323–355 <https://doi.org/10.1146/annurev-biochem-060208-092442>
- 8 Balchin, D., Hayer-Hartl, M. and Hartl, F.U. (2016) *In vivo* aspects of protein folding and quality control. *Science* **353**, aac4354 <https://doi.org/10.1126/science.aac4354>
- 9 Thommen, M., Holtkamp, W. and Rodnina, M.V. (2017) Co-translational protein folding: progress and methods. *Curr. Opin. Struct. Biol.* **42**, 83–89 <https://doi.org/10.1016/j.sbi.2016.11.020>

- 10 Holmgren, A. (1985) Thioredoxin. *Annu. Rev. Biochem.* **54**, 237–271 <https://doi.org/10.1146/annurev.bi.54.070185.001321>
- 11 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 <https://doi.org/10.1073/pnas.0308701101>
- 12 Kelley, R.F. and Richards, F.M. (1987) Replacement of proline-76 with alanine eliminates the slowest kinetic phase in thioredoxin folding. *Biochemistry* **26**, 6765–6774 <https://doi.org/10.1021/bi00395a028>
- 13 Brandts, J.F., Halvorson, H.R. and Brennan, M. (1975) Consideration of the possibility that the slow step on protein denaturation reactions is due to cis-trans isomerism of proline residues. *Biochemistry* **14**, 4953–4963 <https://doi.org/10.1021/bi00693a026>
- 14 Schmid, F.X. and Baldwin, R.L. (1978) Acid catalysis of the formation of the slow-folding species of RNase A: evidence that the reaction is proline isomerization. *Proc. Natl. Acad. Sci. U.S.A.* **75**, 4764–4768 <https://doi.org/10.1073/pnas.75.10.4764>
- 15 Schmidpeter, P.A. and Schmid, F.X. (2015) Prolyl isomerization and its catalysis in protein folding and protein function. *J. Mol. Biol.* **427**, 1609–1631 <https://doi.org/10.1016/j.jmb.2015.01.023>
- 16 Tzul, F.O., Vasilchuk, D. and Makhatadze, G.I. (2017) Evidence for the principle of minimal frustration in the evolution of protein folding landscapes. *Proc. Natl. Acad. Sci. U.S.A.* **114**, E1627–E1632 <https://doi.org/10.1073/pnas.1613892114>
- 17 Tzul, F.O., Vasilchuk, D. and Makhatadze, G.I. (2017) Evidence for the evolutionary conservation of folding kinetics in the thioredoxin protein family. *Proc. Natl. Acad. Sci. U.S.A.* **114**, E4124 <https://doi.org/10.1073/pnas.1704669114>
- 18 Perez-Jimenez, R., Inglés-Prieto, A., Zhao, Z.M., Sanchez-Romero, I., Alegre-Cebollada, J., Kosuri, P. et al. (2011) Single-molecule paleoenzymology probes the chemistry of resurrected enzymes. *Nat. Struct. Mol. Biol.* **18**, 592–596 <https://doi.org/10.1038/nsmb.2020>
- 19 Romero-Romero, M.L., Risso, V.A., Martínez-Rodríguez, S., Ibarra-Molero, B. and Sanchez-Ruiz, J.M. (2016) Engineering ancestral protein hyperstability. *Biochem. J.* **473**, 3611–3620 <https://doi.org/10.1042/BCJ20160532>
- 20 Candel, A.M., Romero-Romero, M.L., Gamiz-Arco, G., Ibarra-Molero, B. and Sanchez-Ruiz, J.M. (2017) Fast folding and slow unfolding of a resurrected Precambrian protein. *Proc. Natl. Acad. Sci. U.S.A.* **114**, E4122–E4123 <https://doi.org/10.1073/pnas.1703227114>
- 21 Acevedo, O., Guzman-Casado, M., Garcia-Mira, M.M., Ibarra-Molero, B. and Sanchez-Ruiz, J.M. (2002) pH corrections in chemical denaturant solutions. *Anal. Biochem.* **306**, 158–161 <https://doi.org/10.1006/abio.2002.5668>
- 22 Holmgren, A. (1979) Thioredoxin catalyzes the reduction of insulin disulfides by dithiothreitol and dihydroliipoamide. *J. Biol. Chem.* **254**, 9627–9632 PMID: 385588
- 23 Slaby, I. and Holmgren, A. (1975) Reconstitution of *Escherichia coli* thioredoxin from complementing peptide fragments obtained by cleavage at methionine-37 or arginine-73. *J. Biol. Chem.* **250**, 1340–1347 PMID: 803502
- 24 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 <https://doi.org/10.1016/j.jmb.2006.07.065>
- 25 Greene, R.F. and Pace, C.N. (1974) Urea and guanidine hydrochloride denaturation of ribonuclease, lysozyme, α -chymotrypsin, and β -lactoglobulin. *J. Biol. Chem.* **249**, 5388–5393 PMID: 4416801
- 26 Radford, S.E., Dobson, C.M. and Evans, P.A. (1992) The folding of hen lysozyme involves partially structured intermediates and multiple pathways. *Nature* **358**, 302–307 <https://doi.org/10.1038/358302a0>
- 27 Georgescu, R.E., Li, J.H., Goldberg, M.E., Tasayco, M.L. and Chaffotte, A.F. (1998) Proline isomerization-independent accumulation of an early intermediate and heterogeneity of the folding pathways of a mixed α/β protein, *Escherichia coli* thioredoxin. *Biochemistry* **37**, 10286–10297 <https://doi.org/10.1021/bi9805083>
- 28 Mücke, M. and Schmid, F.X. (1994) A kinetic method to evaluate the two-state character of solvent-induced protein denaturation. *Biochemistry* **33**, 12930–12935 <https://doi.org/10.1021/bi00209a025>
- 29 Ibarra-Molero, B. and Sanchez-Ruiz, J.M. (1997) Are there equilibrium intermediates in the urea-induced unfolding of hen-egg white lysozyme? *Biochemistry* **36**, 9616–9624 <https://doi.org/10.1021/bi9703305>
- 30 Sanchez-Ruiz, J.M. (2010) Protein kinetic stability. *Biophys. Chem.* **148**, 1–15 <https://doi.org/10.1016/j.bpc.2010.02.004>
- 31 Colon, W., Church, J., Sen, J., Thibeault, J., Trasatti, H. and Xia, K. (2017) Biological roles of protein kinetic stability. *Biochemistry* **56**, 6179–6186. <https://doi.org/10.1021/acs.biochem.7b00942>
- 32 Hedges, S.B. and Kumar, S. (eds.) 2009. *The Timetree of Life*, Oxford Univ. Press, New York, NY
- 33 Inglés-Prieto, A., Ibarra-Molero, B., Delgado-Delgado, A., Perez-Jimenez, R., Fernandez, J.M., Gaucher, E.A. et al. (2013) Conservation of protein structure over four billion years. *Structure* **21**, 1690–1697 <https://doi.org/10.1016/j.str.2013.06.020>
- 34 Risso, V.A., Manssour-Triedo, F., Delgado-Delgado, A., Arco, R., Barroso-delJesus, A., Inglés-Prieto, A. et al. (2015) Mutational studies on resurrected ancestral proteins reveal conservation of site-specific amino acid preferences throughout evolutionary history. *Mol. Biol. Evol.* **32**, 440–455 <https://doi.org/10.1093/molbev/msu312>
- 35 Delgado, A., Arco, R., Ibarra-Molero, B. and Sanchez-Ruiz, J.M. (2017) Using resurrected ancestral proviral proteins to engineer virus resistance. *Cell Rep.* **19**, 1247–1256 <https://doi.org/10.1016/j.celrep.2017.04.037>
- 36 Garcia, A.K., Schopf, J.W., Yokobory, S., Akanuma, A. and Yamagishi, A. (2017) Reconstructed ancestral enzymes suggest long-term cooling of Earth's photic zone since the Archean. *Proc. Natl. Acad. Sci. U.S.A.* **114**, 4619–4624 <https://doi.org/10.1073/pnas.1702729114>
- 37 Napolitano, S., Reber, R.J., Rubini, M. and Glockshuber, R. (2019) Functional analyses of ancestral thioredoxins provide insight into their evolutionary history. *J. Biol. Chem.* **294**, 14105–14118 <https://doi.org/10.1074/jbc.RA119.009718>
- 38 Williams, P.D., Pollock, D.D., Blackburne, B.P. and Goldstein, R.A. (2006) Assessing the accuracy of ancestral protein reconstruction methods. *PLoS Comput. Biol.* **2**, e69 <https://doi.org/10.1371/journal.pcbi.0020069>
- 39 Groussin, M., Hobbs, J.K., Szöllösi, G.J., Gribaldo, S., Arcus, V.L. and Gouy, M. (2015) Toward more accurate ancestral protein genotype-phenotype reconstructions with the use of species tree-aware gene trees. *Mol. Biol. Evol.* **32**, 13–22 <https://doi.org/10.1093/molbev/msu305>
- 40 Eick, G.N., Bridgman, J.T., Anderson, D.P., Harms, M.J. and Thornton, J.W. (2017) Robustness of reconstructed ancestral protein functions to statistical uncertainty. *Mol. Biol. Evol.* **34**, 247–261 <https://doi.org/10.1093/molbev/msw223>
- 41 Vialle, R.A., Tamuri, A.U. and Goldman, N. (2018) Alignment modulates ancestral sequence reconstruction accuracy. *Mol. Biol. Evol.* **35**, 1783–1797 <https://doi.org/10.1093/molbev/msy055>

- 42 Risso, V.A., Gavira, J.A., Gaucher, E.A. and Sanchez-Ruiz, J.M. (2014) Phenotypic comparisons of consensus variants versus laboratory resurrections of Precambrian proteins. *Proteins* **82**, 887–896 <https://doi.org/10.1002/prot.24575>
- 43 Randall, R.N., Radford, C.E., Roof, K.A., Natarajan, D.K. and Gaucher, E.A. (2016) An experimental phylogeny to benchmark ancestral sequence reconstruction. *Nat. Commun.* **7**, 12847 <https://doi.org/10.1038/ncomms12847>
- 44 Scott, K.A., Alonso, D.O., Sato, S., Fersht, A.R. and Daggett, V. (2007) Conformational entropy of alanine versus glycine in protein denatured states. *Proc. Natl. Acad. Sci. U.S.A.* **104**, 2661–2666 <https://doi.org/10.1073/pnas.0611182104>
- 45 Jaswal, S.S., Sohl, J.L., Davis, J.H. and Agard, D.A. (2002) Energetic landscape of α -lytic protease optimizes longevity through kinetic stability. *Nature* **415**, 343–346 <https://doi.org/10.1038/415343a>
- 46 Kim, P.S. and Baldwin, R.L. (1982) Specific intermediates in the folding reactions of small proteins and the mechanism of protein folding. *Annu. Rev. Biochem.* **51**, 459–489 <https://doi.org/10.1146/annurev.bi.51.070182.002331>
- 47 Gun, D. and Gruebele, M. (2019) Weak chemical interactions that drive protein evolution: crowding, sticking, and quinary structure in folding and function. *Chem. Rev.* **119**, 10691–10717 <https://doi.org/10.1021/acs.chemrev.8b00753>
- 48 Matouscheck, A., Kellis, J.T., Serrano, L., Bycroft, M. and Fersht, A.R. (1990) Transient folding intermediates characterized by protein engineering. *Nature* **346**, 440–445 <https://doi.org/10.1038/346440a0>
- 49 Anderson, D.W., McKeown, A.N. and Thornton, J.W. (2015) Intermolecular epistasis shaped the function and evolution of an ancient transcription factor and its DNA binding sites. *eLife* **4**, e07864 <https://doi.org/10.7554/eLife.07864>