Biochemical Journal (2020) **477** 1083–1087
https://doi.org/10.1042/BCJ20190864

PORTLAND PRESS

Commentary

Molecular origins of folding rate differences in the thioredoxin family

Athi N. Naganathan

Department of Biotechnology, Bhupat & Jyoti Mehta School of Biosciences, Indian Institute of Technology Madras, Chennai 600036, India

Correspondence: Athi N. Naganathan (athi@iitm.ac.in)

Thioredoxins are a family of conserved oxidoreductases responsible for maintaining redox balance within cells. They have also served as excellent model systems for protein design and engineering studies particularly through ancestral sequence reconstruction methods. The recent work by Gamiz-Arco et al. [Biochem J (2019) **476**, 3631–3647] answers fundamental questions on how specific sequence differences can contribute to differences in folding rates between modern and ancient thioredoxins but also among a selected subset of modern thioredoxins. They surprisingly find that rapid unassisted folding, a feature of ancient thioredoxins, is not conserved in the modern descendants suggestive of co-evolution of better folding machinery that likely enabled the accumulation of mutations that slow-down folding. The work thus provides an interesting take on the expected folding-stability-function constraint while arguing for additional factors that contribute to sequence evolution and hence impact folding efficiency.

The extant sequences of proteins are a product of innumerable mutational iterations over the evolutionary timeline. Mutations can manifest as distinct changes in the overall or local structure of the protein and contribute to various features including thermodynamic-kinetic stability, solubility and functional outputs [1–4]. One avenue to explore sequence changes that impart altered stability or function is to compare the conformational behavior of modern and ancient proteins [5]. Such a comparison would provide valuable insights for protein design and functional manipulation while enabling one to address specific questions on how mutational effects shape the conformational-functional landscape of proteins across the evolutionary timeline. The method of choice for generating the sequences of ancient proteins is 'ancestral sequence reconstruction' [6]. It involves the construction of a phylogenetic tree of modern or extant sequences (i.e. the external nodes) following which the interior nodes (extinct or ancient sequences) are inferred by maximum-likelihood or Bayesian estimates while assuming a Markovian model for amino acid substitution similar to the Dayhoff matrix construction [7,8]. Such 'resurrected' proteins are often found to be hyperstable, functionally promiscuous and flexible [9].

Modern Thioredoxins (Trx), general oxidoreductases in extant organisms, and their distant ancient cousins have been extensively studied via ancestral sequence reconstruction approaches. The small size of the protein (~108 residues), coupled with the large yield, ease of purification and enzymatic characterization, has facilitated studies which have contributed to our understanding of how sequences evolve [10], provided detailed structures of ancestral proteins [11], revealed differences in the flexibility and dynamics between ancient and modern Trx and hence catalytic differences [12], highlighted site-specific preferences of certain amino acids and their energetic origins [13], and have led to design principles for engineering hyperstable variants [14]. Given this background, the authors in this work [15] ask three specific questions: (1) Do ancient Trxs fold faster or slower than their modern counterparts? (2) If yes, what are the specific sequence positions that determine the differences in folding, their likely evolutionary origins and impact? (3) Finally, are the folding rate constants conserved across the modern Trxs?

Received: 26 December 2019
Revised: 19 February 2020
Accepted: 20 February 2020

Version of Record published:
18 March 2020

Biochemical Journal (2020) 477 1083–1087
https://doi.org/10.1042/BCJ20190864

I will briefly present a rationale as to why these are critical questions to be addressed from the view point of sequence evolution. First, one facet that has been challenging to decipher via sequence resurrection methods is the role of random mutations and their (more than) likely epistatic effects that can potentially bias and confound the identification and interpretation of mutations that alter stability and function [16,17]. In other words, it is well established that mutations occur randomly and are generally accepted in sequences as long as the function or stability (kinetic and thermodynamic) is not compromised beyond a certain threshold value. Since different sequences evolve in different backgrounds (i.e. the cellular milieu), the extant sequences would incorporate not only those sequence changes that occur randomly and impact function and stability, but also additional changes that determine its stickiness (quinary interactions), solubility, diffusivity and the interaction profiles with different translation-folding-degradation machineries (ribosome, proteasome etc.) and hence the protein half-life within a specific organism [18–20]. The observed sequence changes in the same family of proteins can have different origins that are therefore challenging to extricate. It is also possible that the folding rate is different between ancient and modern proteins of the same family due to the evolution and presence of advanced folding assistance processes in modern organisms (i.e. chaperones) [15]. Thus, modern sequences could carry an 'evolutionary baggage', i.e. specific residue alterations that were required in ancient times but that are no longer required now, or could have gotten rid of them. Or have they?

Second, all that is required of a protein *in vivo* is that it folds in a reasonable time-scale to function while contributing to organismal fitness in a positive manner. However, it is well established that there always is and will be a 'tug-of-war' between folding and function that constantly 'selects' for specific sequence patterning over others. For example, certain sequence positions that are critical for function (imagine the specific orientation or active-site residues required for catalysis) can themselves slow-down folding as they are not the optimal choice for folding in a rapid manner. Such 'frustration' can have varied energetic-topological origins [21] and the energy landscape theory predicts that proteins are 'minimally frustrated' as Natural Selection has likely weeded out many of these conflicting interactions [22]. However, it is also pertinent to note that any selection would happen in a specific background or cellular milieu (*in vivo*) and not in isolation. Third, and continuing from the point above, sequence alterations can also modulate the number and nature of intermediate states, barrier heights and folding diffusion coefficients [23,24]. Thus, it is unlikely that sequences that evolve in different backgrounds exhibit similar folding rates as sequence evolution is constantly buffeted by all of the factors discussed above to different extents.

The detailed mutational analysis presented by Gamiz-Arco et al. [15] provides a well-rounded take on the issues reviewed above. Fluorescence and double-jump-unfolding kinetic experiments clearly show that the modern (*E. coli*) Trx folds slower than the LPBCA Trx (last common ancestor of the cyanobacterial, deinococcus and thermus groups, existed 2.5 billion years ago) and LBCA Trx (last bacterial common ancestor, existed 4 billion years ago). While the differences in rates are marginal at 25°C, they differ by more than two orders of magnitude when accounting for the optimal growth temperatures of the organisms (37°C for *E. coli* and >60°C for LPBCA/LBCA). This is an important and less-discussed aspect that needs to be considered when comparing proteins whose source organisms exhibit different growth temperatures [25] as the large difference in thermal energy, molecular diffusivity and the relative time-scale of molecular binding events contribute to different selection pressures [26]. When a conserved cis-proline at position 76 is mutated to alanine the folding rates speed up for all the three proteins and the differences vanish. Interestingly, the catalytic activity is also impaired thus highlighting the direct connection between folding efficiency, sequence conservation and function. Since P76A mutation allows for a greater dihedral flexibility around this position, it is expected to destabilize the proteins and this is experimentally observed. However, any thermodynamic destabilization could be naively expected to decrease the folding rate and not increase it (as experimentally observed) arguing for a large kinetic ruggedness in the folding landscape of Trx that is eliminated as a consequence of this mutation. These observations also indicate that there are additional sequence modifications in the modern Trx that slow down its folding with respect to ancient Trxs and that is potentially exacerbated by P76.

While it is challenging to pinpoint the specific sequence changes that could contribute to the observed differences, the authors identify position 74 (serine and glycine in ancient and modern Trx, respectively) as a plausible candidate through a combination of sequence alignment and structural analysis. They are able to successfully engineer the ancient protein to exhibit a folding time-scale similar to that of the modern counterpart through S74G mutation and vice versa through the reverse G74S in modern Trx. This observation is counter-intuitive as one would expect the faster folding variant (i.e. serine instead of glycine in modern Trx) to be evolutionarily selected as the slower a protein folds, the more will be the time spent during folding and this can

Biochemical Journal (2020) **477** 1083–1087
https://doi.org/10.1042/BCJ20190864

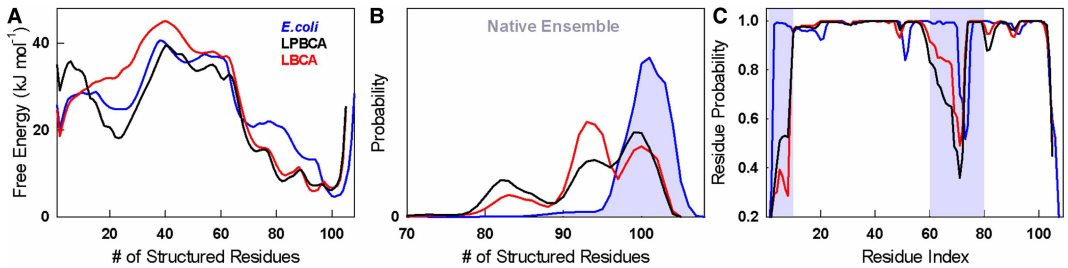

Figure 1. WSME Model Predictions.

(**A**) Free energy profiles as a function of the number of structured residues predicted by the WSME model at a fixed stability of 15 kJ mol^{-1} and at pH 7.0, 298 K for the three indicated proteins. (**B**) The native ensemble distribution at 298 K. (**C**) The probability of finding residue folded as a function of residue index at 298 K. The shaded regions represent residue stretches that display local disorder and that differ across the three proteins.

potentially contribute to larger interference from other cellular constituents (non-specific binding, co-aggregation, degradation *etc.*). However, it is possible that the evolution of better folding machinery in modern organisms mitigate such unwanted effects and this could be one reason why unassisted fast folding (as in ancient Trxs) has undergone evolutionary degradation. Thus, it appears that the modern *E. coli* Trx sequence has, at least in part, shed its evolutionary baggage answering a fundamental question on how folding-functional landscapes of proteins *in vivo* are highly co-evolved features that are at the mercy of not just folding rates but numerous other factors. True to this, the authors find that the folding rates of a set of 13 modern thioredoxins vary by a factor of 100, a majority of the orthologs exhibit slower folding compared to their ancient cousins and exhibit no correlation with thermodynamic stability. It is important to note that despite these differences, enzymatic studies point to similar reductase activities for the modern Trxs providing evidence that sequence variations are a consequence of selective forces other than functional requirements (to the extent one can infer from *in vitro* experiments).

The results of Gamiz-Arco *et al.* can also be viewed through the lens of the Wako-Saitô-Muñoz-Eaton (WSME) model [27,28] to explore how sequence-structure connection modulates the folding landscape. The WSME model treats residues as folding units and we employ an advanced native-centric description of the folding process with contributions from packing, electrostatics, excess conformational entropy and implicit solvation [24,29]. We consider Trxs from *E. coli*, LPBCA and LBCA for simplicity and as representative examples and without explicitly accounting for the trans-to-cis proline isomerization. Briefly, (1) The predicted folded state ensembles (Figure 1A,B) of ancient Trxs are quite broad compared with the *E. coli* arguing for a flexible native state and in agreement with earlier analysis [12], (2) Under conditions of fixed stability, the folding barrier heights are similar (Figure 1A) but the higher stability of the ancient cousins ($\Delta C_m \sim 2$ M GdnHCl) would translate to lower folding barriers and speed up folding in concordance with experiments of Gamiz-Arco et al. (3) The regions of proteins that show large difference in native probability are concentrated in positions 1–10 and 60–80 (Figure 1C). In fact, sequence differences within the latter stretch was considered for selecting modern Trxs that exhibited varied folding times, and (4) The folding free-energy profiles are extremely complex pointing to multiple intermediates and barriers that differ across the evolutionary timeline (Figure 1A) hinting that any sequence variation in modern Trxs would also potentially modulate the number and nature of intermediates. This feature explains why varied sequences would invariably exhibit different folding rates and the observed lack of correlation between stability and folding rates.

The work of Gamiz-Arco et al. thus answers key questions in the field of protein sequence evolution and the eventual connection to folding and function. Importantly, the non-conservation of faster folding in the modern descendants and the large variation in the folding times of modern Trxs highlights the role of organism-specific sequence variations that are likely selected for reasons other than folding efficiency. While the determining factors themselves could be very challenging to extricate, it should still be possible to relate the sequence patterns to the energetic outcomes and the eventual impact on folding and function.

Competing Interests

The author declares that there are no competing interests associated with this manuscript.

Biochemical Journal (2020) **477** 1083–1087
https://doi.org/10.1042/BCJ20190864

Funding

This work was supported by the grant BT/PR26099/BID/7/811/2017 (Department of Biotechnology, Ministry of Science and Technology, India) and Institute Research and Development Award (Early Career Level, Indian Institute of Technology Madras, India) to A.N.N.

Acknowledgements

A.N.N. is a Wellcome Trust/DBT India Alliance Intermediate Fellow

Abbreviations

C_m, chemical denaturation midpoint; GdnHCl, guanidinium hydrochloride; LBCA, last bacterial common ancestor; LPBCA, last common ancestor of the cyanobacterial, deinococcus and thermus groups; Trx, thioredoxin; WSME, Wako-Saitô-Muñoz-Eaton.

References

1 Tokuriki, N. and Tawfik, D.S. (2009) Stability effects of mutations and protein evolvability. *Curr. Opin. Struct. Biol.* **19**, 596–604 https://doi.org/10.1016/j.sbi.2009.08.003
2 Liberles, D.A., Teichmann, S.A., Bahar, I., Bastolla, U., Bloom, J., Bornberg-Bauer, E. et al. (2012) The interface of protein structure, protein biophysics, and molecular evolution. *Protein Sci.* **21**, 769–785 https://doi.org/10.1002/pro.2071
3 Sikosek, T. and Chan, H.S. (2014) Biophysics of protein evolution and evolutionary protein biophysics. *J. R. Soc. Interface* **11**, 20140419 https://doi.org/10.1098/rsif.2014.0419
4 Bershtein, S., Serohijos, A.W. and Shakhnovich, E.I. (2017) Bridging the physical scales in evolutionary biology: from protein sequence space to fitness of organisms and populations. *Curr. Opin. Struct. Biol.* **42**, 31–40 https://doi.org/10.1016/j.sbi.2016.10.013
5 Hochberg, G.K.A. and Thornton, J.W. (2017) Reconstructing ancient proteins to understand the causes of structure and function. *Annu. Rev. Biophys.* **46**, 247–269 https://doi.org/10.1146/annurev-biophys-070816-033631
6 Thornton, J.W. (2004) Resurrecting ancient genes: experimental analysis of extinct molecules. *Nat. Rev. Genet.* **5**, 366–375 https://doi.org/10.1038/nrg1324
7 Jones, D.T., Taylor, W.R. and Thornton, J.M. (1992) The rapid generation of mutation data matrices from protein sequences. *Comput. Appl. Biosci.* **8**, 275–282 https://doi.org/10.1093/bioinformatics/8.3.275
8 Yang, Z., Kumar, S. and Nei, M. (1995) A new method of inference of ancestral nucleotide and amino acid sequences. *Genetics* **141**, 1641–1650 PMID:8601501
9 Risso, V.A., Sanchez-Ruiz, J.M. and Ozkan, S.B. (2018) Biotechnological and protein-engineering implications of ancestral protein resurrection. *Curr. Opin. Struct. Biol.* **51**, 106–115 https://doi.org/10.1016/j.sbi.2018.02.007
10 Petrovic, D., Risso, V.A., Kamerlin, S.C.L. and Sanchez-Ruiz, J.M. (2018) Conformational dynamics and enzyme evolution. *J. R. Soc. Interface* **15**, 20180330 https://doi.org/10.1098/rsif.2018.0330
11 Ingles-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
12 Modi, T., Huihui, J., Ghosh, K. and Ozkan, S.B. (2018) Ancient thioredoxins evolved to modern-day stability-function requirement by altering native state ensemble. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **373**, 20170184 https://doi.org/10.1098/rstb.2017.0184
13 Risso, V.A., Manssour-Triedo, F., Delgado-Delgado, A., Arco, R., Barroso-delJesus, A., Ingles-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
14 Romero-Romero, M.L., Risso, V.A., Martinez-Rodriguez, 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
15 Gamiz-Arco, G., Risso, V.A., Candel, A.M., Ingles-Prieto, A., Romero-Romero, M.L., Gaucher, E.A. et al. (2019) Non-conservation of folding rates in the thioredoxin family reveals degradation of ancestral unassisted-folding. *Biochem. J.* **476**, 3631–3647 https://doi.org/10.1042/BCJ20190739
16 Naganathan, A.N. (2019) Modulation of allosteric coupling by mutations: from protein dynamics and packing to altered native ensembles and function. *Curr. Opin. Struct. Biol.* **54**, 1–9 https://doi.org/10.1016/j.sbi.2018.09.004
17 Horovitz, A., Fleisher, R.C. and Mondal, T. (2019) Double-mutant cycles: new directions and applications. *Curr. Opin. Struct. Biol.* **58**, 10–17 https://doi.org/10.1016/j.sbi.2019.03.025
18 Zou, T., Williams, N., Ozkan, S.B. and Ghosh, K. (2014) Proteome folding kinetics is limited by protein halflife. *PLoS ONE* **9**, e112701 https://doi.org/10.1371/journal.pone.0112701
19 Mu, X., Choi, S., Lang, L., Mowray, D., Dokholyan, N.V., Danielsson, J. et al. (2017) Physicochemical code for quinary protein interactions in *Escherichia coli*. *Proc. Natl Acad. Sci. U.S.A.* **114**, E4556–E4563 https://doi.org/10.1073/pnas.1621227114
20 Davis, C.M., Gruebele, M. and Sukenik, S. (2018) How does solvation in the cell affect protein folding and binding? *Curr. Opin. Struct. Biol.* **48**, 23–29 https://doi.org/10.1016/j.sbi.2017.09.003
21 Ferreiro, D.U., Komives, E.A. and Wolynes, P.G. (2014) Frustration in biomolecules. *Q. Rev. Biophys.* **47**, 285–363 https://doi.org/10.1017/S0033583514000092
22 Bryngelson, J.D., Onuchic, J.N., Socci, N.D. and Wolynes, P.G. (1995) Funnels, pathways, and the energy landscape of protein-folding—a synthesis. *Proteins* **21**, 167–195 https://doi.org/10.1002/prot.340210302
23 Halskau, O., Perez-Jimenez, R., Ibarra-Molero, B., Underhaug, J., Muñoz, V., Martinez, A. et al. (2008) Large-scale modulation of thermodynamic protein folding barriers linked to electrostatics. *Proc. Natl Acad. Sci. U.S.A.* **105**, 8625–8630 https://doi.org/10.1073/pnas.0709881105

24 Naganathan, A.N. (2012) Predictions from an Ising-like statistical mechanical model on the dynamic and thermodynamic effects of protein surface electrostatics. *J. Chem. Theory Comput.* **8**, 4646–4656 https://doi.org/10.1021/ct300676w
25 Romero-Romero, M.L., Risso, V.A., Martinez-Rodriguez, S., Gaucher, E.A., Ibarra-Molero, B. and Sanchez-Ruiz, J.M. (2016) Selection for protein kinetic stability connects denaturation temperatures to organismal temperatures and provides clues to Archaean life. *PLoS ONE* **11**, e0156657 https://doi.org/10.1371/journal.pone.0156657
26 Morcos, F., Schafer, N.P., Cheng, R.R., Onuchic, J.N. and Wolynes, P.G. (2014) Coevolutionary information, protein folding landscapes, and the thermodynamics of natural selection. *Proc. Natl Acad. Sci. U.S.A.* **111**, 12408–12413 https://doi.org/10.1073/pnas.1413575111
27 Wako, H. and Saito, N. (1978) Statistical mechanical theory of protein conformation 2. Folding pathway for protein. *J. Phys. Soc. Jpn* **44**, 1939–1945 https://doi.org/10.1143/JPSJ.44.1939
28 Muñoz, V. and Eaton, W.A. (1999) A simple model for calculating the kinetics of protein folding from three-dimensional structures. *Proc. Natl Acad. Sci. U.S.A.* **96**, 11311–11316 https://doi.org/10.1073/pnas.96.20.11311
29 Rajasekaran, N., Gopi, S., Narayan, A. and Naganathan, A.N. (2016) Quantifying protein disorder through measures of excess conformational entropy. *J. Phys. Chem. B* **120**, 4341–4350 https://doi.org/10.1021/acs.jpcb.6b00658