nature structural & molecular biology

Single-molecule paleoenzymology probes the chemistry of resurrected enzymes

Raul Perez-Jimenez[1], Alvaro Inglés-Prieto[2], Zi-Ming Zhao[3], Inmaculada Sanchez-Romero[2], Jorge Alegre-Cebollada[1], Pallav Kosuri[1,4], Sergi Garcia-Manyes[1], T Joseph Kappock[5], Masaru Tanokura[6], Arne Holmgren[7], Jose M Sanchez-Ruiz[2], Eric A Gaucher[3,8,9] & Julio M Fernandez[1]

It is possible to travel back in time at the molecular level by reconstructing proteins from extinct organisms. Here we report the reconstruction, based on sequence predicted by phylogenetic analysis, of seven Precambrian thioredoxin enzymes (Trx) dating back between ~1.4 and ~4 billion years (Gyr). The reconstructed enzymes are up to 32 °C more stable than modern enzymes, and the oldest show markedly higher activity than extant ones at pH 5. We probed the mechanisms of reduction of these enzymes using single-molecule force spectroscopy. From the force dependency of the rate of reduction of an engineered substrate, we conclude that ancient Trxs use chemical mechanisms of reduction similar to those of modern enzymes. Although Trx enzymes have maintained their reductase chemistry unchanged, they have adapted over 4 Gyr to the changes in temperature and ocean acidity that characterize the evolution of the global environment from ancient to modern Earth.

Experimental paleogenetics and paleobiochemistry provide an opportunity to investigate in the laboratory the molecular history of modern organisms[1–4]. The study of resurrected proteins can also reveal valuable information about the adaptation of extinct forms of life to climatic, ecological and physiological alterations[5–8]. Despite numerous experimental examples of protein resurrection in the laboratory[1], most of the resurrected proteins have reconstructed enzymes thought to exist a few million years (Myr) ago[1,2,9]. Consequently, many hypotheses about ancient life remain untested, especially for time periods associated with rapid changes in biological systems[10]. Such time traveling is largely limited by ambiguity in the historical models used for inferring ancestral sequences. For instance, uncertainties in database and sequence alignments, failures in evolutionary theories and uncertainty in the construction of phylogenetic trees are common sources of ambiguity[1]. Nonetheless, several approaches are commonly used to overcome these limitations, and some examples of reconstructed proteins from the Precambrian (4.5–0.5 Gyr ago) have been reported[1,11]. One such instance is the resurrection of an elongation factor gene family from ancient bacteria 0.5–3.5 Gyr old[9,12]. The ancestral elongation factors showed that ancient bacteria lived in a hot environment and also revealed a correlation between the thermostability of ancient life and the temperature of ancient oceans as inferred from geological records.

These reconstruction and resurrection studies have paved the way for researchers to formulate interesting questions about ancient organisms and the biomolecules that supported them. For instance, little is known about how the chemistry of primitive enzymes arose and how the environmental conditions affected the evolution of their chemistry[13]. These questions cannot be answered directly by examining fossil records. In an effort to understand the evolution of enzymatic reactions, we have reconstructed and tested Trx enzymes from extinct organisms that lived in the Precambrian.

Thioredoxins belong to a broad family of oxidoreductase enzymes that are ubiquitous in all living organisms[14]. The archetypical active site (CXXC) and the Trx fold have been conserved throughout evolution, indicating that Trx enzymes were probably present in primitive forms of life. Single-molecule force-clamp spectroscopy has been used to examine in detail the chemical mechanisms of disulfide reduction by Trx enzymes at the sub-Ångström scale[15,16]. The combination of single-molecule force spectroscopy and the resurrection of ancestral proteins might reveal new insights into the reductase activity of these sulfur-based enzymes.

RESULTS

Reconstruction of ancestral Trx enzymes

Using the large number of extant Trx sequences available, we have constructed a highly articulated phylogenetic tree encompassing more than 200 diverse Trx sequences from the three domains of life (**Supplementary Fig. 1** and **Supplementary Note** for GI numbers).

[1]Department of Biological Sciences, Columbia University, New York, New York, USA. [2]Facultad de Ciencias, Departamento de Química-Física, Universidad de Granada, Granada, Spain. [3]Georgia Institute of Technology, School of Biology, Atlanta, Georgia, USA. [4]Department of Biochemistry and Molecular Biophysics, Columbia University, New York, New York, USA. [5]Department of Biochemistry, Purdue University, West Lafayette, Indiana, USA. [6]Department of Applied Biological Chemistry, Graduate School of Agricultural and Life Sciences, The University of Tokyo, Tokyo, Japan. [7]Division of Biochemistry, Department of Medical Biochemistry and Biophysics, Karolinska Institutet, Stockholm, Sweden. [8]Georgia Institute of Technology, School of Chemistry, Atlanta, Georgia, USA. [9]Georgia Institute of Technology, Parker H. Petit Institute for Bioengineering and Bioscience, Atlanta, Georgia, USA. Correspondence should be addressed to R.P.-J. (raulpjc@biology.columbia.edu) or J.M.F. (jfernandez@columbia.edu).

Received 9 August 2010; accepted 24 January 2011; published online 3 April 2011; doi:10.1038/nsmb.2020

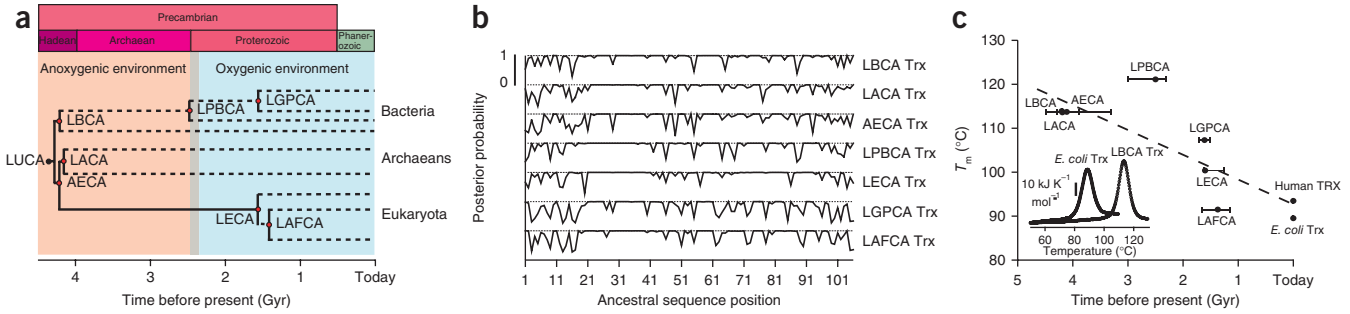

Figure 1 Phylogenetic analysis of Trx enzymes and ancestral sequence reconstruction. (**a**) Schematic phylogenetic tree showing the geological time in which different extinct organisms lived (see text). Dashed lines represent further bifurcations. Divergence times are compiled from multiple sources and are summarized in ref. 17. The figure indicates the global environment, although both aerobic and anaerobic organisms are found in modern environments. (**b**) Posterior probability distribution of the inferred amino acids across 106 sites for the interested internal nodes. The inferred amino acid at each site for the interested internal node is the residue with the highest posterior probability. (**c**) Denaturation temperatures (T_m) versus geological time for ancestral Trx enzymes. Modern *E. coli* and human Trx enzymes are also indicated. The dashed line represents a lineal fit to the data. Inset, experimental DSC thermograms for *E. coli* Trx and LBCA Trx. The instrumental uncertainty of DSC measurements is <0.5 °C.

From this tree we sampled several biologically relevant nodes for sequence reconstruction and laboratory resurrection. We applied estimates of divergence dates to nodes in the tree assuming that the root of the tree lies between bacteria and the common ancestor of archaea and eukaryotes[17].

We resurrected Trx enzymes belonging to the last bacterial common ancestor (LBCA), the last archaeal common ancestor (LACA) and the archaeal-eukaryotic common ancestor (AECA; **Fig. 1**). These organisms are thought to have inhabited Earth 4.2–3.5 Gyr ago (**Fig. 1a**) after diverging from the last universal common ancestor (LUCA)[7,17]. We also selected a node corresponding to the last eukaryotic common ancestor (LECA) that lived in the Proterozoic, about 1.60 Gyr ago. We selected two other internal nodes in the bacterial lineages; the last common ancestor of the cyanobacterial and deinococcus and thermus groups (LPBCA), which existed around 2.5 Gyr ago and represents the origin of photosynthetic bacteria, and the last common ancestor of γ-proteobacteria (LGPCA), which is around 1.61 Gyr old. Finally, we also chose the last common ancestor of animals and fungi (LAFCA), which lived around 1.37 Gyr ago (**Fig. 1a**). Although it would be interesting to observe the chemical transition between LUCA and the last common ancestors of bacteria, and archaea and eukaryotes, the reconstruction of Trx from LUCA cannot be inferred without making far-reaching assumptions about ancient life. For instance, the last common ancestor of bacteria represents an ancient population, and the last common ancestor of archaea and eukaryotes represents another ancient population. It is practical to infer that LUCA lies somewhere along the branch that connects these two ancient populations. However, we cannot identify the position of LUCA (the ancestral node) without an outgroup. We have reason to believe that LUCA existed on this branch (on the basis of ancient gene duplication events that gave rise to paralogs still present in all modern organisms), but we do not know where on the branch LUCA existed. This prevents us from inferring the sequence of LUCA (*pace* midpoint rooting).

We reconstructed the sequences of the ancestral Trx enzymes using statistical methods based on maximum likelihood[3,9]. For a given node in the tree, we calculated posterior probability values for all 20 amino acids considering each site of the inferred sequence. These values represent the probability that a certain residue occupied a specific position in the sequence at a particular point in the phylogeny. The posterior probabilities were calculated on the basis of an amino acid replacement matrix[18]. We then reconstructed the most probabilistic ancestral sequence (M-PAS) at a specific node by assigning to each site the residue with the highest posterior probability. We determined the posterior probability distribution of the inferred amino acids across 106 sites for the selected sequences (**Fig. 1b**). The M-PASs of interest are summarized in **Supplementary Figure 2**. We synthesized the genes that encode these sequences, expressed the proteins from *Escherichia coli* cells and purified them.

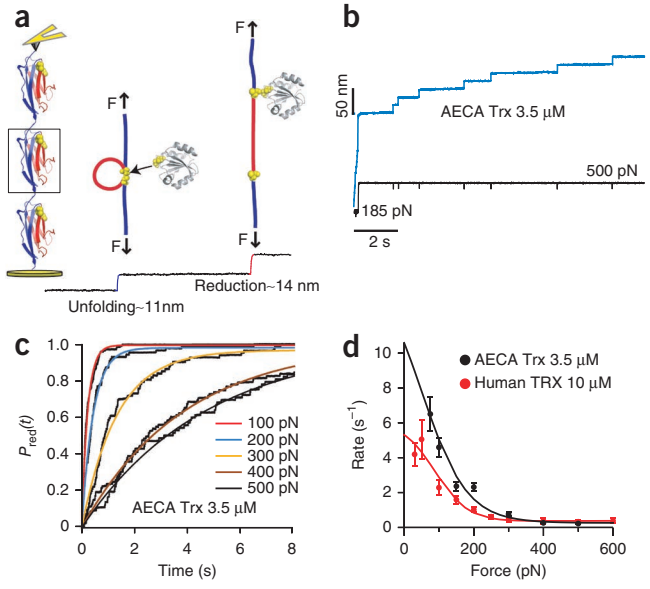

Figure 2 Single-molecule disulfide reduction assay. (**a**) Schematic representation of the single-molecule disulfide reduction assay. A first pulse of force rapidly unfolds the $I27_{G32C-A75C}$ domains (11-nm step). When the disulfide bond is exposed to the solvent a single Trx molecule can reduce it (14-nm step). (**b**) Experimental force-clamp trace showing single disulfide reductions of a $(I27_{G32C-A75C})_8$ polyprotein. The unfolding pulse was set at 185 pN for 0.2 s and the test-pulse force at 500 pN. (**c**) Probability of reduction ($P_{red}(t)$) resulted from summing and normalizing the reduction test pulse at different forces for AECA Trx (3.5 μM). (**d**) Force dependency of disulfide reduction by AECA Trx; human TRX is also shown for comparison. Both Trx enzymes show a similar pattern: a negative force dependency of the reduction rate, from 30–200 pN, consistent with a Michaelis-Menten mechanism and a force-independent mechanism, from 200 pN and up, described by an electron transfer reaction[16]. Notice the higher activity for AECA Trx (3.5 μM for AECA Trx versus 10 μM for human TRX). The lines represent fittings to the kinetic model (see Online Methods). The error bars are given by the s.e.m. obtained using the bootstrap method.

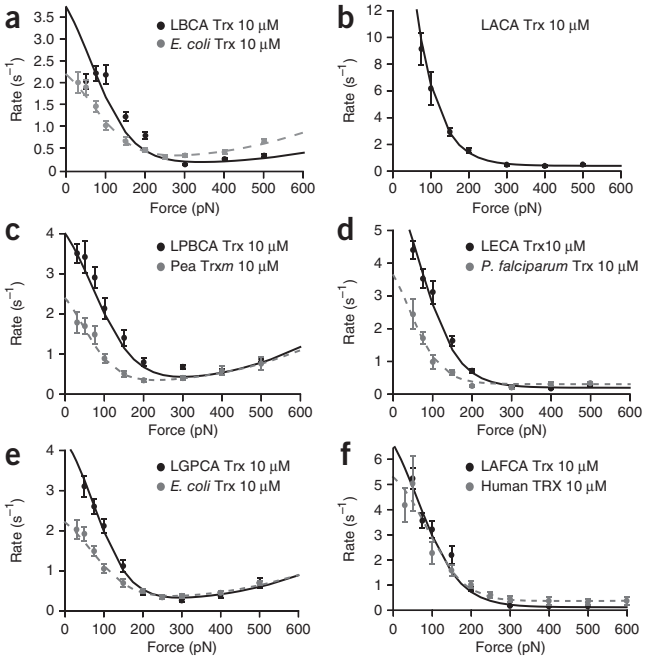

Figure 3 Force dependence of disulfide reduction by ancestral Trx enzymes. (**a**–**f**) The reduction rate at a given force is obtained by summing, averaging and fitting to single exponential numerous traces (15–80) like the one shown in **Figure 2b**. The solid lines are fitting to the kinetic model. The gray circles and dashed lines represent the rate versus force dependence for modern Trxs: Pea Trx*m* from chloroplast (**c**), *P. falciparum* Trx (**d**), *E. coli* Trx (**a**,**e**) and human TRX (**f**; all extracted from ref. 16). These modern Trxs are descendants of the ancestral Trxs in the same plot. The error bars represent the s.e.m. obtained using the bootstrap method.

Thermal stability of ancient Trx enzymes

As a first step toward investigating the physicochemical properties of these resurrected enzymes, we used differential scanning calorimetry (DSC) to measure their thermal stabilities. The denaturation temperature (T_m) can provide an idea of the temperature range in which the proteins operate. A plot of the T_m of the resurrected enzymes against geological time is shown in **Figure 1c**. The three oldest Trx enzymes from LBCA, AECA and LACA showed a similar T_m of around 113 °C. LBCA Trx maintained a highly populated native state up to ~105 °C, where the thermal transition began (**Fig. 1c**, inset). By contrast, we determined T_m for modern *E. coli* and human Trxs of 88.8 and 93.3 °C, respectively. The ΔT_m between the oldest and modern Trx is around 25 °C, a value similar to that determined for bacterial elongation factor[12], which corroborates the hypothesis of the thermophilic nature of LBCA, AECA and LACA[7]. From the data in **Figure 1c**, we determined a paleotemperature trend yielding a decrease in the T_m of 5.8 ± 1.8 K Gyr^{-1}.

Although the thermodynamic denaturation temperatures determined for the ancestral Trxs follow a cooling trend similar to that of ancient oceans, the values are about 50 °C higher than the ocean temperatures inferred from maximum $\delta^{18}O$ (ref. 12). It is possible that Trx evolution operated primarily on kinetic stability, and this could be reflected in thermodynamic stability[19]. However, other than the obvious loss of function upon denaturation, it is unclear exactly how the value of T_m is related to Trx enzyme fitness.

Force-dependent chemical kinetics of disulfide reduction

We investigated the chemical mechanisms of disulfide bond reduction used by the resurrected enzymes. Given the ancient origin of the reconstructed Trx enzymes, with some of them predating the buildup of atmospheric oxygen, it seemed logical to assume that their catalytic chemistry might be closer to that of simple sulfur-based molecules that use a straightforward collision-driven substitution nucleophilic bimolecular (S_N2) mechanism of reduction[20]. By contrast, Trx enzymes use a complex mixture of chemical mechanisms, including a crucial substrate binding and rearrangement reaction that accounts for the vast increase in the efficiency of Trx over the simpler sulfur compounds that were available in early geochemistry[15,16].

Recently, we developed an assay based on single-molecule force spectroscopy that measures the effect of applying a well-controlled force to a disulfide-bonded substrate on its rate of reduction by a nucleophile. This assay readily distinguishes the simple S_N2 chemistry of nucleophiles such as hydroxide, glutathione and L-cysteine from the more complex reduction chemistry of the Trx enzymes[15,16,21–24], making it ideal for probing the chemistry of the resurrected enzymes (**Fig. 2**). As a substrate, we use an engineered polyprotein made of eight repeats of the I27 immunoglobulin-like protein modified by mutating amino acids 32 and 75 to cysteine ($I27_{G32C-A75C}$)$_8$. The cysteines oxidize spontaneously, forming disulfide bonds that are hidden in each folded I27 domain in the chain. We pick up and stretch single polyproteins using an atomic force microscope (AFM) in solutions containing the desired nucleophile. In a typical experiment, we first apply a constant force to the polyprotein (175–185 pN, 0.2–0.3 s), and this rapidly unfolds the $I27_{G32C-A75C}$ modules up to the disulfide bond. The unfolding events result in a stepwise increase in the length of the polyprotein where each module contributes with ~11 nm in length (**Fig. 2a** and **Supplementary Fig. 3**). After unfolding occurs, every disulfide bond is exposed to the solvent. If active Trx enzymes are present in the solution, single reduction events of ~14 nm per module can be readily observed (**Fig. 2a**,**b** and **Supplementary Figs. 3**,**4**). All the ancestral enzymes that we resurrected could trigger staircases of reduction events (**Fig. 2b** and **Supplementary Figs. 3**,**4**), indicating that they were all active. To measure the reduction rate, we summed 15–80 reduction staircases similar to the one shown in **Figure 2b** and fit the resulting average with a single exponential. We repeated this procedure for different pulling forces (**Fig. 2c**). The resulting set of data measured the force dependency of the rate of reduction of the disulfide bond (**Fig. 2d**).

The chemical mechanisms of disulfide reduction can be distinguished by their sensitivity to the force applied to the substrate[16]. Simple thiol-reducing agents show a force dependency in which the rate always increased exponentially with the applied force[21,22]. By contrast, modern Trx enzymes show a negative force dependency in the range of 30–200 pN (ref. 16). This mechanism is consistent with a Michaelis-Menten binding reaction followed by a force-inhibited reorientation of the substrate disulfide bond that is necessary for an S_N2 reaction to occur[15]. In a second mechanism, the rate of reduction increases exponentially at forces above 200 pN. This mechanism can be described by a simple S_N2 reaction and is found only in Trx enzymes of bacterial origin. In all Trx enzymes there is a force-independent mechanism of reduction that can be ascribed to a single electron transfer reaction[16].

The same three reduction mechanisms can be observed in the ancient enzymes with similar patterns to those found in extant Trxs (**Figs. 2d** and **3**). We could easily fit the force dependency of the reduction rate measured from the resurrected enzymes using the three-state kinetic model that was previously used with modern Trxs[15,16] (see Online Methods and **Supplementary Table 1**). One might expect that Trx enzymes from primitive forms of life should have less developed chemical mechanisms. For instance, one of the main factors that controls the chemistry of Trx catalysis is the geometry of the

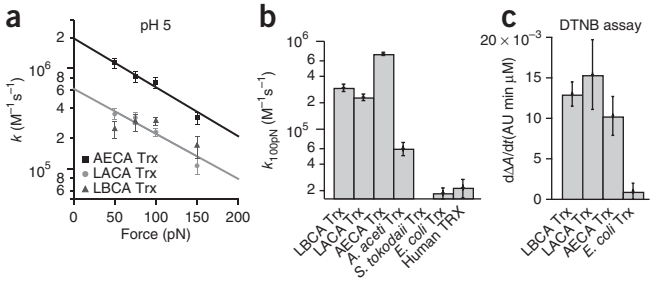

Figure 4 Rate constants of disulfide bond reduction at pH 5. (**a**) High activity for AECA (squares) and LACA (circles) Trxs at pH 5 when the substrate was pulled at low forces (50–150 pN). LBCA Trx (triangles) showed similar activity to that at pH 7.2 with a similar trend (**Fig. 3a**). The solid lines are exponential fit to the experimental data. (**b**) Rate constants for disulfide reduction by ancestral Trx enzymes at pH 5 are higher than for modern *E. coli* and human Trx. Thioredoxin from the acidophile *A. aceti* shows activity at pH 5; enzymes from the thermophilic *S. tokodaii* do not show a detectable rate of reduction at the same pH. All experiments were conducted at a pulling force of 100 pN. Error bars represent s.e.m. obtained using the bootstrap method. (**c**) Activity of ancestral Trxs and modern *E. coli* Trx measured using DTNB as substrate at pH 5 and determined by monitoring spectrophotometrically the formation of TNB at 412 nm. Error bars represent s.d. from three different measurements.

binding groove. In modern Trxs from bacteria, the binding groove is less pronounced than in eukaryotic Trxs[16]. This structural difference is responsible for the different chemical behavior observed in eukaryotic and bacterial Trxs. It seems reasonable that ancient enzymes would have had a less structured groove, making their chemistry more similar to that of simple reducing agents such as L-cysteine or TCEP[22]. However, the chemistry of Trx enzymes seems to have been established early in evolution, about 4 Gyr ago, in the same manner as is observed today. This observation suggests that the step from simple reducing compounds to well-structured and functional enzymes occurred early in molecular evolution[10].

Nonetheless, several features of the catalytic mechanisms of some ancestral Trxs are noteworthy: for instance, the high activity of AECA and LACA Trxs when the substrate was pulled at forces below 200 pN (**Figs. 2d** and **3b**). From the fitting of the reduction rate versus force data to the three-state kinetic model, an extrapolation to zero force yielded rate constants of 30×10^5 M^{-1} s^{-1} for AECA Trx and 29×10^5 M^{-1} s^{-1} for LACA Trx. The extrapolation to zero force in the rest of the ancestral Trxs predicted rate constants from 3.7×10^5 M^{-1} s^{-1} to 6.6×10^5 M^{-1} s^{-1} (**Supplementary Fig. 5**). Although these latter values are within the range of those found in extant Trx enzymes using AFM (**Fig. 3** and ref. 16) and bulk experiments[25], there was a trend in the reconstructed enzymes to show higher reduction rates at forces below 200 pN (**Fig. 3**). We speculate that this trend might be related to the substrate specificity of the enzymes. Ancient enzymes may be less substrate specific than modern ones and therefore might be more efficient with generic substrates such as those used here.

Another notable feature is the small upward slope observed at low forces for LBCA Trx with a maximum at ~100 pN (**Fig. 3a**). Although structural information would be needed to fully address this point, it seems possible that the binding between substrate and enzyme is not optimum at zero force. A better conformation appears to be achieved by applying force to the substrate.

We also measured the activity of the ancestral enzymes using the conventional insulin assay (**Supplementary Fig. 6**). The values of insulin precipitation rates obtained with this assay are similar to those previously determined for *E. coli* Trx[26,27].

Activity of ancestral Trxs under acidic conditions

LBCA, AECA and LACA thrived in an anoxygenic environment that was probably rich in sulfur compounds and CO_2, whereas LPBCA, LECA, LGPCA and LAFCA lived in a global oxygenic environment[10] (**Fig. 1a**). The high level of CO_2 in the Hadean was partly responsible for the proposed low pH of the ancient oceans (~5.5)[28,29]. Therefore, following the hypothesis that early life lived in seawater, the natural habitat in which LBCA, AECA and LACA lived was likely to have been acidic as well as hot. This is especially important because the reactivity of Trx enzymes is due, in part, to the low pK_a value of the reactive cysteine: ~6.5 compared with 8.0 for free cysteine[14]. This low pK_a value is a consequence of complex electrostatic interactions between several residues that stabilize the deprotonated form of the reactive cysteine[30]. Thus, Trx activity is highly sensitive to pH, and modern enzymes would not work well at low pH because the catalytic thiol would be protonated and inactive.

We compared the reactivity of LACA, AECA and LBCA enzymes with the extant human and *E. coli* Trx enzymes at pH 5. The ability of the resurrected enzymes to operate in low pH environments was apparent from these experiments. We measured the force dependency of reduction for AECA, LACA and LBCA at pH 5 over the 50–150 pN force range (**Fig. 4a**). For AECA Trx, an extrapolation to zero force gave a reduction rate constant of 19×10^5 M^{-1} s^{-1} (**Fig. 4a**, solid line); similarly, for LACA we estimated a rate constant of 6.2×10^5 M^{-1} s^{-1}, whereas for LBCA Trx the reduction rates observed at pH 5 were similar to those measured at pH 7.2 (**Fig. 4a**). These are high values similar to those measured for some modern Trx enzymes at neutral pH (ref. 16). We also compared the rate of reduction measured at 100 pN for LBCA, LACA and AECA with the rates of some modern Trx enzymes also measured at pH 5 (**Fig. 4b**).

Owing to spontaneous precipitation of insulin at pH <6, we used DTNB as a substrate for disulfide reduction to verify the ability of the oldest enzymes to work at pH 5 (**Fig. 4c**). This analysis of reconstructed enzymes indicated that that ancient Trx enzymes were well adapted to function under acidic conditions and that Trx enzymes could maintain similar reduction rate constants as they evolved in more alkaline environments. A feature of the thioredoxin family of enzymes is that many of them are secreted to the extracellular environment where most disulfide-bonded proteins are found[31,32]. From this perspective, thioredoxin enzymes are perhaps one of the few types of enzymes for which a correlation can be established between their pH sensitivity and the environmental conditions found outside cells[31,32]. It is informative to compare the acid tolerance of the resurrected enzymes with that of enzymes from extant extremophiles. For example, Trx from *Sulfolobus tokodaii* (thermophilic archaea[33]), with a melting temperature of 122 °C (**Supplementary Fig. 7**), is active at pH 7 (0.12×10^5 M^{-1} s^{-1} at 50 pN) but does not show detectable activity at pH 5 (**Fig. 4b**), which is not surprising given that *Sulfolobus* regulates its cytosolic pH[34]. By contrast, Trx from *Acetobacter aceti* (acidophilic bacteria[35] that grows at pH 4) is active at pH 5 (0.6×10^5 M^{-1} s^{-1} at 100 pN), reflecting its acidic cytosol[35,36].

DISCUSSION

Our results show that the chemical mechanisms observed in modern Trx enzymes were present in reconstructed Trxs predicted to be present in Precambrian organisms. Our data support the view that ancestral Trx enzymes from LBCA, AECA and LACA, which lived in the mid to late Hadean, were highly resistant to temperature and active in relatively acidic conditions. Our results support the view that in early life Trx enzymes were present in hot environments and that these environments progressively cooled from 4 to 0.5 Gyr ago[10,12,37,38].

However, it is also possible that a much cooler early Earth was populated by psychrophiles, mesophiles and thermophiles and that the latter could have been the only survivors of cataclysmic events (for example, the late heavy bombardment or global glaciations on early Earth[10,39]). Our findings also support the idea that many important biochemical pathways in the modern biosphere were already established by 3.5 Gyr ago[10]. For instance, metabolism is one of the most conserved cellular processes. Important pathways such as energy production, sugar degradation, cofactor biosynthesis and amino acid processing are highly conserved from bacteria to human and were probably present in LUCA[40]. Evolution clearly operates at multiple levels of biological organization; however, enzymatic mechanisms accompanying adaptive changes seem to be highly conserved. The ability of enzymes to maintain specific chemical reactivity and mechanisms in disparate environments is necessary for the diversification of life. As we show here, this ability is exemplified by Trx enzymes, and it could also be the case for universal proteins, such as ubiquitin, RNase, ATPase or other metabolic enzymes that have been maintained in nearly all organisms throughout the history of life. We propose that the experimental resurrection of ancestors of these universal proteins together with the sensitivity of single-molecule techniques can be a powerful tool for understanding the origin and evolution of life on Earth.

METHODS

Methods and any associated references are available in the online version of the paper at http://www.nature.com/nsmb/.

Note: Supplementary information is available on the Nature Structural & Molecular Biology website.

ACKNOWLEDGMENTS

Supported by the US National Institutes of Health (HL061228 and HL066030 to J.M.F.); the Spanish Ministry of Science and Innovation (J.M.S.-R.); NASA Astrobiology (Georgia Institute of Technology); NASA Exobiology (E.A.G.); Fundación Ibercaja and Fundación Caja Madrid (R.P.-J. and S.G.-M.); and Fundación Alfonso Martín Escudero (J.A.-C.). We thank B. Ibarra-Molero (University of Granada) for assistance with bulk enzymatic assays.

AUTHOR CONTRIBUTIONS

R.P.-J., J.M.S.-R., E.A.G. and J.M.F. designed the research; Z.-M.Z. and E.A.G. conducted the phylogenetic analysis and sequence reconstruction; A.I.-P. and J.M.S.-R. expressed and purified the ancestral enzymes and conducted the calorimetric measurements and analysis; T.J.K. provided *A. aceti* Trx; M.T. provided *S. tokodaii* Trx; A.H. provided human TRX; R.P.-J., I.S.-R., J.A.-C., P.K. and S.G.-M. performed AFM experiments; R.P.-J. and I.S.-R. analyzed AFM data; R.P.-J., E.A.G. and J.M.F. wrote the paper; all authors participated in revising the manuscript.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

Published online at http://www.nature.com/nsmb/.
Reprints and permissions information is available online at http://npg.nature.com/reprints/index.html.

1. Benner, S.A., Sassi, S.O. & Gaucher, E.A. Molecular paleoscience: systems biology from the past. *Adv. Enzymol. Relat. Areas Mol. Biol.* **75**, 1–132 (2007).
2. Thornton, J.W. Resurrecting ancient genes: experimental analysis of extinct molecules. *Nat. Rev. Genet.* **5**, 366–375 (2004).
3. Liberles, D.A. (ed.) *Ancestral Sequence Reconstruction* (Oxford Univ. Press, Oxford, New York, 2007).
4. Hall, B.G. Simple and accurate estimation of ancestral protein sequences. *Proc. Natl. Acad. Sci. USA* **103**, 5431–5436 (2006).
5. Thornton, J.W., Need, E. & Crews, D. Resurrecting the ancestral steroid receptor: ancient origin of estrogen signaling. *Science* **301**, 1714–1717 (2003).
6. Thomson, J.M. *et al.* Resurrecting ancestral alcohol dehydrogenases from yeast. *Nat. Genet.* **37**, 630–635 (2005).
7. Boussau, B., Blanquart, S., Necsulea, A., Lartillot, N. & Gouy, M. Parallel adaptations to high temperatures in the Archaean eon. *Nature* **456**, 942–945 (2008).
8. Chang, B.S., Jonsson, K., Kazmi, M.A., Donoghue, M.J. & Sakmar, T.P. Recreating a functional ancestral archosaur visual pigment. *Mol. Biol. Evol.* **19**, 1483–1489 (2002).
9. Gaucher, E.A., Thomson, J.M., Burgan, M.F. & Benner, S.A. Inferring the palaeoenvironment of ancient bacteria on the basis of resurrected proteins. *Nature* **425**, 285–288 (2003).
10. Nisbet, E.G. & Sleep, N.H. The habitat and nature of early life. *Nature* **409**, 1083–1091 (2001).
11. Pollock, D.D. & Chang, B.S.W. in *Ancestral Sequence Reconstruction* (ed. Liberles, D.A.) 85–94 (Oxford Univ. Press, Oxford, New York, 2007).
12. Gaucher, E.A., Govindarajan, S. & Ganesh, O.K. Palaeotemperature trend for Precambrian life inferred from resurrected proteins. *Nature* **451**, 704–707 (2008).
13. Zalatan, J.G. & Herschlag, D. The far reaches of enzymology. *Nat. Chem. Biol.* **5**, 516–520 (2009).
14. Holmgren, A. Thioredoxin. *Annu. Rev. Biochem.* **54**, 237–271 (1985).
15. Wiita, A.P. *et al.* Probing the chemistry of thioredoxin catalysis with force. *Nature* **450**, 124–127 (2007).
16. Perez-Jimenez, R. *et al.* Diversity of chemical mechanisms in thioredoxin catalysis revealed by single-molecule force spectroscopy. *Nat. Struct. Mol. Biol.* **16**, 890–896 (2009).
17. Hedges, S.B. & Kumar, S. *The Timetree of Life* (Oxford Univ. Press, Oxford, 2009).
18. Yang, Z., Kumar, S. & Nei, M. A new method of inference of ancestral nucleotide and amino acid sequences. *Genetics* **141**, 1641–1650 (1995).
19. Godoy-Ruiz, R. *et al.* 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 (2006).
20. Kice, J.L. Nucleophilic substitution at different oxidation states of sulfur. in *Progress in Inorganic Chemistry* (ed. Edwards, J.O.) 147–206 (John Wiley & Sons, 2007).
21. Wiita, A.P., Ainavarapu, S.R., Huang, H.H. & Fernandez, J.M. Force-dependent chemical kinetics of disulfide bond reduction observed with single-molecule techniques. *Proc. Natl. Acad. Sci. USA* **103**, 7222–7227 (2006).
22. Koti Ainavarapu, S.R., Wiita, A.P., Dougan, L., Uggerud, E. & Fernandez, J.M. Single-molecule force spectroscopy measurements of bond elongation during a bimolecular reaction. *J. Am. Chem. Soc.* **130**, 6479–6487 (2008).
23. Garcia-Manyes, S., Liang, J., Szoszkiewicz, R., Kuo, T.L. & Fernandez, J.M. Force-activated reactivity switch in a bimolecular chemical reaction. *Nat. Chem.* **1**, 236–242 (2009).
24. Liang, J. & Fernandez, J.M. Mechanochemistry: one bond at a time. *ACS Nano* **3**, 1628–1645 (2009).
25. Holmgren, A. Reduction of disulfides by thioredoxin. Exceptional reactivity of insulin and suggested functions of thioredoxin in mechanism of hormone action. *J. Biol. Chem.* **254**, 9113–9119 (1979).
26. Suarez, M. *et al.* Using multi-objective computational design to extend protein promiscuity. *Biophys. Chem.* **147**, 13–19 (2010).
27. Holmgren, A. Thioredoxin catalyzes the reduction of insulin disulfides by dithiothreitol and dihydrolipoamide. *J. Biol. Chem.* **254**, 9627–9632 (1979).
28. Walker, J.C.G. Possible limits on the composition of the archean ocean. *Nature* **302**, 518–520 (1983).
29. Russell, M.J. & Hall, A.J. The emergence of life from iron monosulphide bubbles at a submarine hydrothermal redox and pH front. *J. Geol. Soc. Lond.* **154**, 377–402 (1997).
30. Dyson, H.J. *et al.* Effects of buried charged groups on cysteine thiol ionization and reactivity in *Escherichia coli* thioredoxin: structural and functional characterization of mutants of Asp 26 and Lys 57. *Biochemistry* **36**, 2622–2636 (1997).
31. Xu, S.Z. *et al.* TRPC channel activation by extracellular thioredoxin. *Nature* **451**, 69–72 (2008).
32. Windle, H.J., Fox, A., Ni Eidhin, D. & Kelleher, D. The thioredoxin system of *Helicobacter pylori*. *J. Biol. Chem.* **275**, 5081–5089 (2000).
33. Ming, H. *et al.* Crystal structure of thioredoxin domain of ST2123 from thermophilic archaea *Sulfolobus tokodaii* strain7. *Proteins* **69**, 204–208 (2007).
34. Baker-Austin, C. & Dopson, M. Life in acid: pH homeostasis in acidophiles. *Trends Microbiol.* **15**, 165–171 (2007).
35. Starks, C.M., Francois, J.A., MacArthur, K.M., Heard, B.Z. & Kappock, T.J. Atomic-resolution crystal structure of thioredoxin from the acidophilic bacterium Acetobacter aceti. *Protein Sci.* **16**, 92–98 (2007).
36. Menzel, U. & Gottschalk, G. The internal pH of *Acetobacterium wieringae* and *Acetobacter aceti* during growth and production of acetic acid. *Arch. Microbiol.* **143**, 47–51 (1985).
37. Knauth, L.P. & Lowe, D.R. High Archean climatic temperature inferred from oxygen isotope geochemistry of cherts in the 3.5 Ga Swaziland Supergroup, South Africa. *Geol. Soc. Am. Bull.* **115**, 566–580 (2003).
38. Schulte, M. The emergence of life on earth. *Oceanography (Wash. D.C.)* **20**, 42–49 (2007).
39. Gogarten-Boekels, M., Hilario, E. & Gogarten, J.P. The effects of heavy meteorite bombardment on the early evolution—the emergence of the three domains of life. *Orig. Life Evol. Biosph.* **25**, 251–264 (1995).
40. Peregrin-Alvarez, J.M., Tsoka, S. & Ouzounis, C.A. The phylogenetic extent of metabolic enzymes and pathways. *Genome Res.* **13**, 422–427 (2003).

ONLINE METHODS

Phylogenetic analysis and ancestral sequence reconstruction. We retrieved 203 thioredoxin sequences from the three domains of life from GenBank. For GI numbers, see **Supplementary Note**. Sequences were aligned using MUSCLE[41] and further corrected manually. The multiple sequence alignment is available upon request. The phylogenetic analysis was carried out by the minimum evolution distance criterion with 1,000 bootstrap replicates using PAUP* 4.0 beta (Sinauer Associates). Ancestral sequences were reconstructed using PAML version 3.14 and incorporated the gamma distribution for variable replacement rates across sites[42]. For each site of the inferred sequences, posterior probabilities were calculated for all 20 amino acids. The amino acid residue with the highest posterior probability was then assigned at each site.

Protein expression and purification. Genes encoding the ancestral Trx enzymes were synthesized and codon-optimized for expression in *E. coli* cells (Epoch Biolabs). The genes were cloned into pQE80L vector (Qiagen) and transformed in *E. coli* BL21 (DE3) cells (Invitrogen). Cells were incubated overnight in LB medium at 37 °C, and protein expression was induced with 1 mM IPTG. Cell pellets were sonicated and the His_6-tagged proteins were loaded onto His GraviTrap affinity column (GE Healthcare). The purified protein was verified by SDS-PAGE. The proteins were then loaded into PD-10 desalting column (GE Healthcare) and finally dialyzed against 50 mM HEPES, pH 7.0 buffer. The preparation of $(I27_{G32C\text{-}A75C})_8$ was carried out as described[15,43].

DSC experiments. Thermal stabilities of ancestral and modern Trx enzymes were measured with a VP-Capillary DSC (MicroCal). Protein solutions were dialyzed into a buffer of 50 mM HEPES, pH 7. The scan speed was set to 1.5 K min^{-1}. Several buffer-buffer baselines were first obtained for proper equilibration of the calorimeter. Concentrations were 0.3–0.7 mg ml^{-1} and were determined spectrophotometrically at 280 nm using theoretical extinction coefficients and molecular weights. The experimental traces were analyzed following the two-state thermodynamic model[44].

AFM experiments. The AFM used is a custom-made design previously described[45]. Cantilever model MLCT of silicon nitride were used (Veeco). We calibrated the cantilever using the equipartition theorem[46] giving rise to a typical spring constant of 0.02 N m^{-1}. The AFM works in the force-clamp mode with length resolution 0.5 nm. The feedback response can reach 5 ms. The buffer used in the experiment was 10 mM HEPES, pH 7.2, 150 mM NaCl, 1 mM EDTA and 2 mM NADPH. We add Trx enzymes to the desired concentration. The buffer also contains Trx reductase (50 nM) to keep Trx enzymes in their reduced state. *E. coli* Trx reductase works well with bacterial-origin Trx enzymes, whereas eukaryotic Trx reductase works with archaea and eukaryote Trx enzymes; however, similar results were obtained using DTE 200 μM to keep Trx enzymes reduced, thus showing that modern Trx reductases maintain fully reduced ancestral Trx enzymes. For the experiments at pH 5, we used 20 mM sodium acetate buffer and 200 μM DTE.

To perform the experiment, we deposited 3–6 μl of substrate ~0.1 mg ml^{-1} on a gold-covered coverslide. A drop of ~100 μl containing the Trx solution was then added. The force-clamp protocol consists of three pulses of force. In the first pulse the cantilever tip was pressed against the surface at 800 pN for 2 s. In the second pulse the attached $(I27_{G32C\text{-}A75C})_8$ was stretched at 175–185 pN for 0.2–0.3 s. The third pulse was the test force where the reduction events were captured. This pulse was applied at different forces (30–500 pN) enough times to capture all the possible reduction events.

The traces were collected and analyzed using custom-written software in Igor Pro 6.03 (Wavemetrics). The traces containing the reduction events at each force were summed, normalized and fitted with a single exponential. From the fitting we obtained a time constant τ and thus the reduction rate at a given force ($r = 1/\tau$). We used the bootstrapping method to obtain the error of the reduction rates. The bootstrapping was run 1,000 times for each reduction rate to obtain a distribution from which the s.e.m. could be calculated.

Thioredoxin bulk enzymatic measurements. Bulk-solvent oxidoreductase activity for ancestral thioredoxins was determined using the insulin precipitation assay as described[26,27,43]. To further verify the activity of ancestral Trxs enzymes at acidic pH, we used DTNB (5,5′-dithiobis-(2-nitrobenzoic acid)) as substrate at pH 5. In this assay, Trxs enzymes were preactivated by incubation with 1 mM DTT. The reaction was initiated by adding active Trx to a final concentration of 4 μM to the cuvette containing 1 mM DTNB in 20 mM sodium acetate buffer, pH 5. We followed the change in absorbance at 412 nm due to the formation of TNB during 1 min. Activity was determined from the slope $d\Delta A_{412}/dt$. A control experiment lacking Trx was registered and subtracted as baseline.

AFM data analysis. The data were fitted following a three-state kinetic model as described[15,16]. In this model three different chemical mechanisms are taken into account. The rate constants used in the kinetic model are

$$k_{01} = \alpha_0[\mathrm{Trx}]$$
$$k_{12} = \beta_0 \exp(F\Delta x_{12}/k_BT) + \lambda_0$$
$$k_{02} = \gamma_0 [\mathrm{Trx}] \exp(F\Delta x_{02}/k_BT) + \lambda_0$$
$$k_{10} = \delta_0$$

The rate constants k_{01} and k_{02} depend on Trx concentration in a linear manner. k_{12} and k_{02} exponentially depend on force. The kinetic model is solved using matrix analysis and parameters α_0, β_0, Δx_{12}, γ_0, Δx_{02}, λ and δ_0 can be obtained for each ancestral enzyme. The optimal kinetic parameters are calculated by numerical optimization using the downhill simplex method[47] (**Supplementary Table 1**). An extensive explanation of the different chemical mechanisms can be found in ref. 16, where the origin of this chemical diversity was explained on the basis of the structural features of the binding groove.

Igor Pro software was used for data analysis. All figures were generated using Igor Pro and Adobe Illustrator CS4.

41. Edgar, R.C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32**, 1792–1797 (2004).
42. Yang, Z.H. PAML: a program package for phylogenetic analysis by maximum likelihood. *Comput. Appl. Biosci.* **13**, 555–556 (1997).
43. Perez-Jimenez, R. *et al.* Force-clamp spectroscopy detects residue co-evolution in enzyme catalysis. *J. Biol. Chem.* **283**, 27121–27129 (2008).
44. Ibarra-Molero, B., Loladze, V.V., Makhatadze, G.I. & Sanchez-Ruiz, J.M. Thermal versus guanidine-induced unfolding of ubiquitin. An analysis in terms of the contributions from charge-charge interactions to protein stability. *Biochemistry* **38**, 8138–8149 (1999).
45. Fernandez, J.M. & Li, H. Force-clamp spectroscopy monitors the folding trajectory of a single protein. *Science* **303**, 1674–1678 (2004).
46. Florin, E.L. *et al.* Sensing specific molecular interactions with the atomic-force microscope. *Biosens. Bioelectron.* **10**, 895–901 (1995).
47. Nelder, J.A. & Mead, R.A. Simplex method for function minimization. *Comput. J.* **7**, 308–313 (1965).