

# Conservation of Protein Structure over Four Billion Years

Alvaro Ingles-Prieto,<sup>1,5</sup> Beatriz Ibarra-Molero,<sup>1</sup> Asuncion Delgado-Delgado,<sup>1</sup> Raul Perez-Jimenez,<sup>2,6</sup> Julio M. Fernandez,<sup>2</sup> Eric A. Gaucher,<sup>3</sup> Jose M. Sanchez-Ruiz,<sup>1,\*</sup> and Jose A. Gavira<sup>4,\*</sup>

<sup>1</sup>Departamento de Química Física, Facultad de Ciencias, Universidad de Granada, Granada 18071, Spain

<sup>2</sup>Department of Biological Sciences, Columbia University, New York, NY 10027, USA

<sup>3</sup>Georgia Institute of Technology, School of Biology, School of Chemistry and Biochemistry, and Parker H. Petit Institute for Bioengineering and Biosciences, Atlanta, GA 30332, USA

<sup>4</sup>Laboratorio de Estudios Cristalográficos, Instituto Andaluz de Ciencias de la Tierra (Consejo Superior de Investigaciones Científicas – Universidad de Granada), Avenida de las Palmeras 4, Armilla, Granada 18100, Spain

<sup>5</sup>Present address: IST Austria, Am Campus 1, 3400 Klosterneuburg, Austria

<sup>6</sup>Present address: Ikerbasque Research Foundation, CIC nanoGUNE, Tolosa Hiribidea 76, 20018 San Sebastian, Spain

\*Correspondence: [sanchezr@ugr.es](mailto:sanchezr@ugr.es) (J.M.S.-R.), [jgavira@iact.ugr-csic.es](mailto:jgavira@iact.ugr-csic.es) (J.A.G.)

<http://dx.doi.org/10.1016/j.str.2013.06.020>

## SUMMARY

Little is known about the evolution of protein structures and the degree of protein structure conservation over planetary time scales. Here, we report the X-ray crystal structures of seven laboratory resurrections of Precambrian thioredoxins dating up to approximately four billion years ago. Despite considerable sequence differences compared with extant enzymes, the ancestral proteins display the canonical thioredoxin fold, whereas only small structural changes have occurred over four billion years. This remarkable degree of structure conservation since a time near the last common ancestor of life supports a punctuated-equilibrium model of structure evolution in which the generation of new folds occurs over comparatively short periods and is followed by long periods of structural stasis.

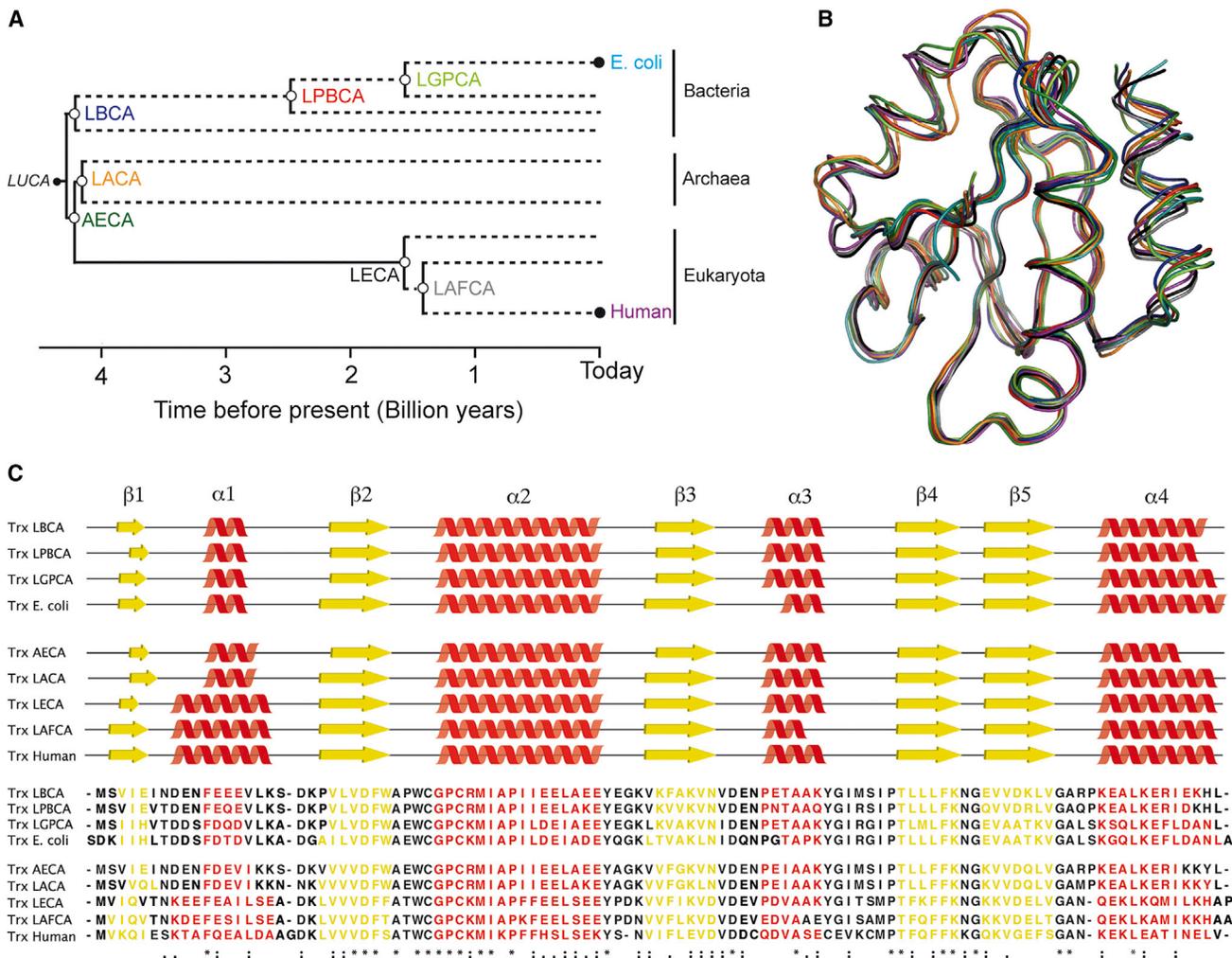
## INTRODUCTION

Little is known with certainty about the evolution of protein structures, despite the substantial number of different protein folds revealed by the structures deposited in the Protein Data Bank (PDB). As elaborated below, several facts contribute to this undesirable situation.

While it is generally admitted that structures change at a slower pace than sequences do, evidence has accumulated in recent years supporting that protein structures are not invariant and, therefore, that they may change during the course of evolution (Grishin, 2001; Murzin, 2008; Sikosek et al., 2012; Taylor, 2007; Tokuriki and Tawfik, 2009; Valas et al., 2009). In fact, due to the so-called shape-covering properties of the mapping of sequence into structure (Caetano-Anollés et al., 2009), different structures may be just a few mutational steps away in sequence space, as has been experimentally demonstrated (Cordes et al., 1999; He et al., 2012). Moreover, the possibility of convergent evolution of folds is generally accepted and,

hence, common ancestry does not necessarily follow from structural similarity (Grishin, 2001; Krishna and Grishin, 2004; Murzin, 2008; Orengo et al., 1994; Schaeffer and Daggett, 2011; Taylor, 2007). That is, transitions between folds and convergent evolution of folds may both conceivably occur during protein evolution; therefore, the identification of basic principles of structure evolution may be difficult to extract from the study of extant protein structures (Caetano-Anollés et al., 2009; Murzin, 2008). Consequently, many current fold classifications are phenetic (based on a metric of structure similarity) and the viability of phyletic classifications (based on evolutionary relationships) remains an open issue (Murzin, 2008; Valas et al., 2009). As a result, age estimates for protein folds are uncertain and based on indirect methods, such as the census of (assigned) folds in genomes (Caetano-Anollés et al., 2009; Winstanley et al., 2005). Even the usefulness of the fold concept is at stake, as several authors have discussed that fold space must be viewed as continuous rather than discrete (Honig, 2007; Sadreyev et al., 2009; Xie and Bourne, 2008).

The above observations summarize what may be viewed as a particularly clear example of the limitations of “horizontal” approaches (i.e., based on the comparison between extant proteins) to molecular evolution (Harms and Thornton, 2010). In fact, some recent work has used sequence reconstruction analyses targeting ancestral states represented by nodes in phylogenetic trees and the subsequent laboratory “resurrection” of their encoded proteins (Benner et al., 2007; Harms and Thornton, 2010) to address important issues in protein evolution, such as the role of epistasis in formation of new function (Ortlund et al., 2007), the evolution of complex biomolecular machines (Finnigan et al., 2012), the mechanisms of evolutionary innovation through gene duplication (Voordeckers et al., 2012), and the adaptation of proteins to changing environments over planetary time scales (Gaucher et al., 2008; Perez-Jimenez et al., 2011; Risso et al., 2013). Here we explore the potential of this “vertical” approach to probe the evolution of protein structures. To this end, we have obtained the three-dimensional (3D) structures of several laboratory resurrections of Precambrian enzymes dating up to approximately four billion years (Gyr) ago, i.e., up to a time close to the origin of life. In particular, we target thioredoxin enzymes corresponding to the last bacterial common ancestor



**Figure 1. Overall Structural Features of Extant Thioredoxins and Laboratory Resurrections of Precambrian Thioredoxins**  
 (A) Schematic phylogenetic tree showing the geological time (Perez-Jimenez et al., 2011) and the phylogenetic nodes targeted in this work.  
 (B) Spatial course of the polypeptide chain for the human and *E. coli* thioredoxins, as well as for the several laboratory resurrections of Precambrian thioredoxins studied in this work. The color code is that given in (A).  
 (C) Sequences (Perez-Jimenez et al., 2011) and secondary structure assignments for the extant thioredoxins and the laboratory resurrections of Precambrian thioredoxins studied in this work.  
 See also Table S1 for root-mean-square deviation and sequence identity values for all thioredoxin structure pairs.

(LBCA); the last archaeal common ancestor (LACA); the archaeal-eukaryotic common ancestor (AECA); the last eukaryotic common ancestor (LECA); the last common ancestor of fungi and animals (LAFCA); the last common ancestor of the cyanobacterial, deinococcus and thermus groups (LPBCA); and the last common ancestor of  $\gamma$ -proteobacteria (LGPCA). As briefly described subsequently, we recently “resurrected” and characterized these proteins in terms of stability and function (Perez-Jimenez et al., 2011).

We used ~200 diverse extant thioredoxin sequences encompassing the three domains of life to construct a highly articulated phylogenetic tree and subsequently perform a maximum likelihood sequence reconstruction targeting several Precambrian nodes during thioredoxin evolution (Perez-Jimenez et al., 2011). The resultant phylogenetic tree was sufficiently close to

an accepted organism phylogeny to allow us to assign the reconstructed nodes to well-defined Precambrian ancestors (see previous) and to date those nodes (see Figure 1A; Hedges and Kumar, 2009; for further details, see Perez-Jimenez et al., 2011). In the laboratory, we resurrected the proteins encoded by the reconstructed sequences and determined their stability and catalytic features. We found an increase in denaturation temperature of ~30°C when “traveling back in time” several billion years. This result afforded support for our ancestral reconstruction exercise, because it is consistent with the generally proposed thermophilic character of Precambrian life and, indeed, similar stability enhancements have been reported in Precambrian resurrection studies on other proteins systems, such as elongation factors (Gaucher et al., 2008) and  $\beta$ -lactamases (Risso et al., 2013). It is also noteworthy that some

**Table 1. Data Collection and Refinement Statistics**

|   | LPBCA (2YJ7)              | LECA (2YOI)              | AECA (3ZIV)               | LACA (2YNX)               | LAFCA (2YPM)              | LGPCA (2YN1)              | LBCA (4BA7)               |
|---|---------------------------|--------------------------|---------------------------|---------------------------|---------------------------|---------------------------|---------------------------|
| Data Collection   |                           |                          |                           |                           |                           |                           |                           |
| Space group   | P 1 2 1 1                 | C 1 2 1                  | P 1 2 1 1                 | P 1                       | P 2 1 2 1 2 1             | P 1 2 1 1                 | F 4 3 2                   |
| Cell Dimensions   |                           |                          |                           |                           |                           |                           |                           |
| a, b, c (Å)   | 55.1, 30.2, 59.0          | 58.4, 47.8, 73.8         | 37.6, 48.8, 91.1          | 32.2, 36.3, 48.1          | 37.5, 42.8, 55.9          | 36.1, 62.9, 42.9          | 192.7                     |
| $\beta$ (°)   | 117.0                     | 98.5                     | 93.2                      | 108.0                     | 90.0                      | 109.0                     | 90.0                      |
| ASU   | 2                         | 2                        | 3                         | 2                         | 1                         | 2                         | 2                         |
| Resolution (Å) <sup>a</sup>   | 48.45–1.65<br>(1.74–1.65) | 36.6–1.30<br>(1.37–1.30) | 45.51–2.65<br>(2.74–2.65) | 45.32–1.75<br>(1.84–1.75) | 20.00–2.20<br>(2.28–2.20) | 34.09–1.30<br>(1.35–1.30) | 58.10–2.45<br>(2.53–2.45) |
| R <sub>sym</sub> (%) <sup>a</sup>   | 4.5 (34.1)                | 4.7 (6.6)                | 6.5 (59.9)                | 10.5 (39.6)               | 10.3 (48.7)               | 5.6 (42.8)                | 13.0 (93.4)               |
| I/ $\sigma$ <sub>1</sub> <sup>a</sup>   | 20.10 (3.80)              | 18.40 (7.90)             | 16.21 (3.24)              | 8.50 (2.70)               | 23.26 (4.95)              | 12.47 (2.68)              | 12.92 (4.18)              |
| Completeness (%) <sup>a</sup>   | 99.8 (100.0)              | 94.4 (89.8)              | 97.8 (97.9)               | 97.6 (88.7)               | 99.0 (100.0)              | 97.2 (97.7)               | 99.9 (100.0)              |
| Unique reflections  | 21,198                    | 46,574                   | 9,592                     | 18,887                    | 4,904                     | 43,308                    | 11,808                    |
| Multiplicity <sup>a</sup>   | 3.6 (3.7)                 | 2.0 (1.8)                | 4.0 (4.0)                 | 1.0 (2.9)                 | 10.4 (10.7)               | 2.8 (2.8)                 | 7.7 (8.0)                 |
| Refinement  |                           |                          |                           |                           |                           |                           |                           |
| Resolution (Å)  | 48.45–1.65                | 36.6–1.30                | 45.51–2.65                | 45.32–1.75                | 20.00–2.20                | 34.09–1.30                | 58.10–2.45                |
| R <sub>work</sub> /R <sub>free</sub> (%)  | 18.19/21.98               | 15.43/18.47              | 18.22/27.39               | 15.76/21.52               | 16.65/23.92               | 16.76/20.42               | 15.83/21.08               |
| No. atoms   | 2,117                     | 4,371                    | 2,390                     | 3,934                     | 918                       | 4,121                     | 3,764                     |
| Protein   | 1,852                     | 1,926                    | 2,382                     | 1,817                     | 866                       | 1,825                     | 1,819                     |
| Water   | 264                       | 433                      | 8                         | 184                       | 52                        | 357                       | 69                        |
| Ligands   | 1                         | 7                        | 0                         | 2                         | 0                         | 2                         | 8                         |
| Average B-factors (Å <sup>2</sup> )   | 21.80                     | 16.20                    | 80.00                     | 19.80                     | 31.00                     | 18.30                     | 44.20                     |
| Rmsd  |                           |                          |                           |                           |                           |                           |                           |
| Bond lengths (Å)  | 0.007                     | 0.011                    | 0.008                     | 0.013                     | 0.009                     | 0.011                     | 0.015                     |
| Bond angles (°)   | 0.97                      | 1.29                     | 1.15                      | 1.39                      | 1.09                      | 1.35                      | 1.48                      |
| Ramachandran (%)  |                           |                          |                           |                           |                           |                           |                           |
| Favored   | 100.0                     | 98.0                     | 96.0                      | 98.0                      | 99.0                      | 99.0                      | 100.0                     |
| Outliers  | 0.0                       | 0.0                      | 0.0                       | 0.0                       | 0.0                       | 0.0                       | 0.0                       |
| LPBCA, last common ancestor of cyanobacterial, deinococcus, and thermus groups; LECA, last eukaryotic common ancestor; AECA, archaeal-eukaryotic common ancestor; LACA, last archaeal common ancestor; LAFCA, last common ancestor of fungi and animals; LGPCA, last common ancestor of gamma-proteobacteria; LBCA, last bacterial common ancestor; rmsd, root-mean-square deviation. See Table S4 for crystallization methods, conditions, and data collection source and molecular replacement searching model. |                           |                          |                           |                           |                           |                           |                           |
| <sup>a</sup> Values in parentheses are for the highest resolution shell.  |                           |                          |                           |                           |                           |                           |                           |

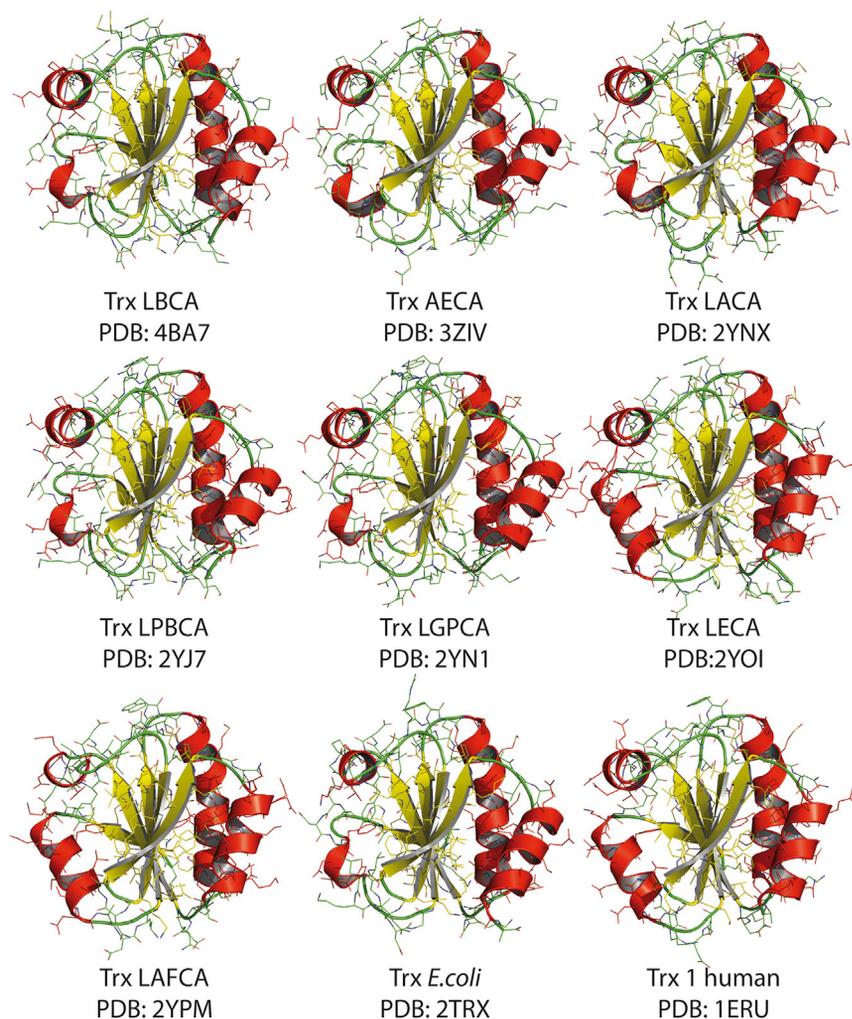
proposed scenarios for the primitive Earth environment include acidic ancestral oceans and that both single-molecule and bulk-solutions assays indicated that the oldest resurrected thioredoxins were actually well adapted to function at acidic pH (Perez-Jimenez et al., 2011). Overall, the highly enhanced stability and the catalysis at acidic pH in the older thioredoxins provided evidence of adaptation to the proposed hot and acidic conditions of the ancient oceans (Perez-Jimenez et al., 2011).

Here, we report and analyze the X-ray crystal structures of the laboratory resurrections of Precambrian thioredoxins we previously studied in terms of stability and catalysis. We find a remarkable degree of structure conservation up to a time close to the origin of life, a result that seems consistent with a punctuated-equilibrium model of structure evolution in which the generation of new folds occurs over comparatively short periods and is followed by long periods of structural stasis. Furthermore, the results and analyses reported here support that laboratory resurrection targeting Precambrian nodes followed by 3D structure

determination can be a powerful approach to explore the poorly understood evolution of protein structures.

## RESULTS AND DISCUSSION

The seven Precambrian thioredoxins were crystallized by either the counter-diffusion or the sitting drop vapor diffusion method. The X-ray data ranged from high to medium resolution limit (Table 1; Table S4 available online) and, therefore, conclusions drawn from the 3D model are well supported. A general view of the seven putative ancestral structures (Figures 1B, 1C, and 2; Table S1) shows that each displays the topology of the thioredoxin fold consisting of N-terminal  $\beta 1\alpha 1\beta 2\alpha 2\beta 3\alpha 3$  and C-terminal  $\beta 4\beta 5\alpha 4$  domains arranged in a central core of three parallel and two antiparallel strands of pleated  $\beta$  sheet surrounded by the four helices. Furthermore, no large differences emerge when the putative ancestral structures are compared among themselves (or when they are compared with the extant human and *Escherichia coli* thioredoxins), in terms of polar and apolar



**Figure 2. Ribbon Representations of the Thioresoxin Structures Studied in This Work**

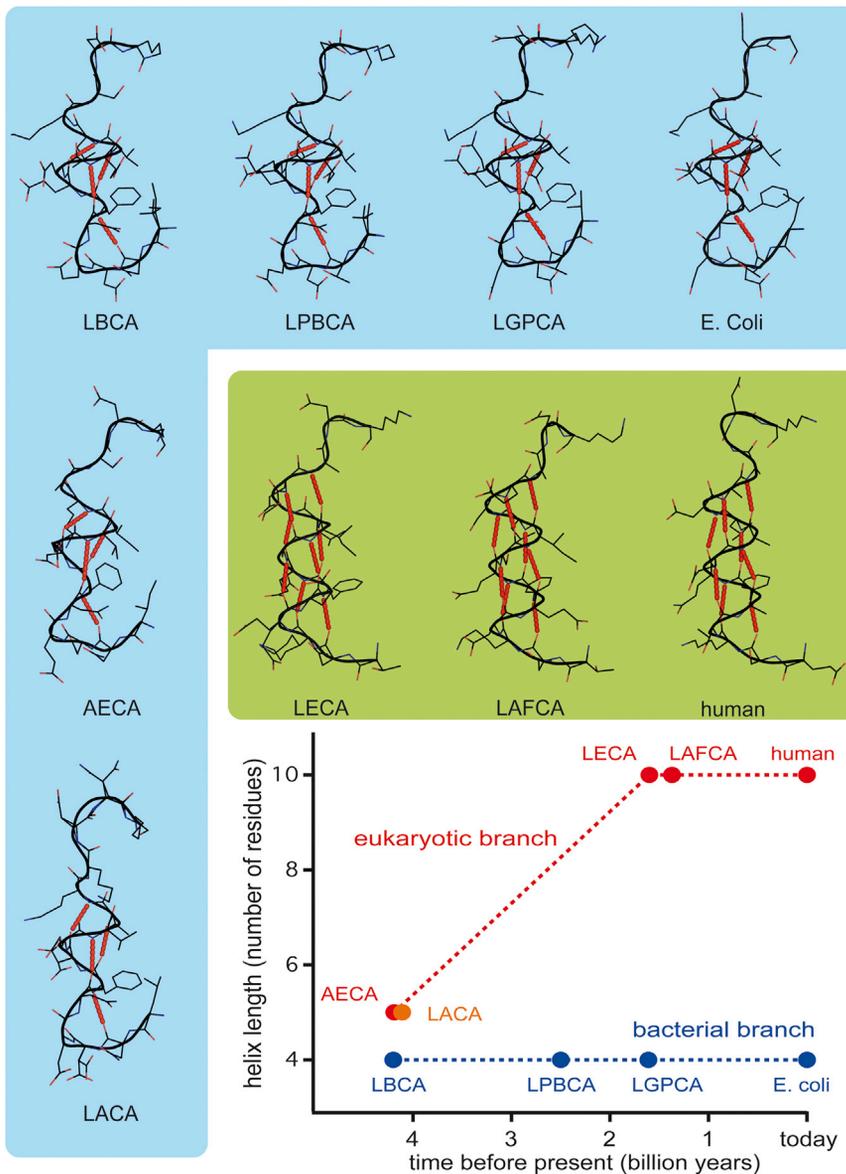
General overview of the seven laboratory resurrections of Precambrian thioresoxins and the extant *E. coli* and human thioresoxins showing the canonical fold. See Figure S1 and Table S2 for energies of charge-charge interactions, accessible surface areas, and numbers of hydrogen bonds and salt bridges for all the thioresoxin structures studied in this work.

solvent-accessible surface areas, numbers of hydrogen bonds and salt bridges, and surface charge distributions (Table S2; Figure S1).

Despite the overall structural conservation, our experimental results do support some changes in the thioresoxin structure over the four-billion-year period, in particular in length of helix  $\alpha 1$  (Figures 1C and 3). Although the posterior probability values for the reconstructed sequences are comparatively low at some positions of the helix  $\alpha 1$  region (Perez-Jimenez et al., 2011), we are confident that the observed structural features are robust to alternative sequence inferences (Hanson-Smith et al., 2010), in particular because the change observed in helix  $\alpha 1$  shows a clearly defined evolutionary pattern that can be understood by the fact that many extant thioresoxin-like fold proteins are known to differ in the first  $\alpha$ -layer (Qi and Grishin, 2005). For instance, the structures of the extant human and *E. coli* thioresoxins greatly differ in the length of helix  $\alpha 1$  (Figures 1C and 3), a result that is robust against different methods to ascertain helix length (Table S3). Furthermore, an analysis of the structures deposited in the PDB indicates a shorter helix  $\alpha 1$  for most bacterial thioresoxins as compared with eukaryotic thioresoxins (Figure 4). This leads to one obvious question:

which of the structural features (long helix versus short helix) is ancestral and which is derived? This kind of evolutionary question cannot be readily addressed by using a “horizontal” approach (i.e., the comparative analysis of the extant structures summarized in Figure 4). However, the “vertical” approach based on the laboratory resurrection of putative ancestral proteins followed by structure determination does suggest an immediate answer. Figure 3A includes a plot of helix length versus geological time for modern human and *E. coli* thioresoxins as well as for the ancestral resurrections studied in this work. The structures for the resurrections corresponding to organisms that inhabited Earth approximately four billion years ago display a short helix  $\alpha 1$ . This suggests that the short helix in *E. coli* thioresoxin (and most bacterial thioresoxins) is very likely the ancestral structural feature (present in the thioresoxin of LUCA, we might speculate), while the long helix in human thioresoxin (and most eukaryotic thioresoxins) is a derived feature that was acquired (perhaps in a switch-like manner) along the evolution from AECA thioresoxin to LECA thioresoxin. Finally, we have found only four structures of thioresoxins from archaea in the PDB. They show a helix  $\alpha 1$  length of seven to eight residues, a value somewhat higher than that determined for the laboratory resurrection corresponding to the last archaeal common ancestor (five residues; see LACA in Figure 3). However, the poor statistical basis provided by the small number of available structures for archaeal thioresoxins, together with the fact that our recent thioresoxin resurrection study (Perez-Jimenez et al., 2011) targeted only one archaeal ancestor (the approximate four billion-year-old LACA), prevents us from analyzing in detail the change in helix  $\alpha 1$  length along the archaeal branch.

The putative ancestral structures reported here are consistent with the thioresoxin fold being an approximate four billion-year-old molecular fossil of sorts and confirms that protein structures can evolve slowly. We anticipate that additional Precambrian resurrection studies may help define structural prototypes, despite the likely geometrically



**Figure 3. Changes in the Size of Helix  $\alpha 1$  in Thioredoxins over Approximately Four Billion Years as Inferred from Laboratory Resurrections of Precambrian Proteins**

Canonical  $\alpha$  helix hydrogen bonds are shown in red to highlight the changes in helix length. Different color backgrounds are used for short helices (blue) and long helices (green). A plot of helix length versus geological time is also included. See Table S3 for calculations supporting the robustness of the differences found in helix  $\alpha 1$  length.

time scale over which we may expect significant changes in protein structure to occur. We have furthermore shown that critical evolutionary issues regarding fold definition, fold age, and the identification of ancestral and derived structural features can be readily addressed based on putative ancestral structures. The results and analyses reported here thus support that laboratory resurrection targeting Precambrian nodes followed by 3D structure determination can be a powerful approach to explore the poorly understood evolution of protein structures.

#### EXPERIMENTAL PROCEDURES

##### Protein Expression and Purification

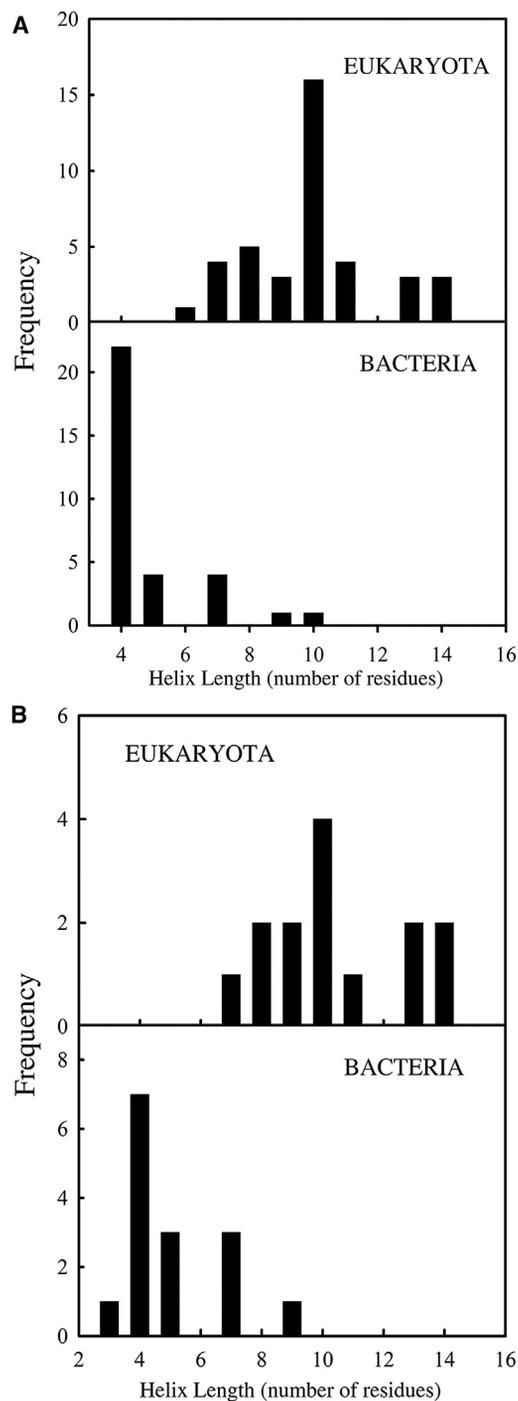
Ancestral thioredoxin open reading frames were PCR-amplified from pQE80L-derived vectors containing them (Perez-Jimenez et al., 2011). We designed 5'-end oligonucleotides to introduce an *Nde*I restriction site (CATATG) in which the ATG sequence codes for the first methionine codon; 3'-end oligonucleotides were designed to create a *Xho*I site after the stop codon of each open reading frame. PCR fragments were digested with *Nde*I and *Xho*I (New England Biolabs) and cloned between the same sites in vector pET-30a(+). In these constructs, the open reading frames from ancestral thioredoxins are expressed from the first methionine to the stop codon with no additional amino acid. Sequencing analysis confirmed that vectors corresponded to their design. Each ancestral thioredoxin gene cloned in the vector pET-30a(+) was expressed in BL21(DE3) *E. coli* bacterial strain. Cells were grown in lysogeny broth medium containing 100  $\mu$ g/ml ampicillin at 37°C and induced with 0.4 mM IPTG at an optical density 600 (OD<sub>600</sub>) 0.7. After 7 hr of incubation at 37°C, cells were harvested and resuspended in 30 mM Tris and 1 mM EDTA buffer at pH 8.3. The lysate was first applied to ion-exchange chromatography on a Fractogel EMD DEAE (M) column using a 0–1 M NaCl gradient in 30 mM Tris-EDTA buffer at pH 8.3. Fractions containing thioredoxin were pooled and subsequently applied to gel filtration chromatography on a HiLoad Superdex 75 preparative grade column. The protein was exhaustively dialyzed in 10 mM HEPES at pH 7.0. Prior to crystallization, protein was concentrated by centrifugation at 14,000 rpm using Centricon centrifugal filter units (Sartorius).

##### Crystallization, Data Collection, and Structure Determination

Crystals were grown in capillaries using the counter-diffusion technique (CCD; Otálora et al., 2009) or in nanodrops using the vapor diffusion technique. Initial

continuous nature of protein structure space (Sadreyev et al., 2009). From a more general point of view, we may speculate that the evolution of protein structures may be sometimes described as a type of punctuated equilibrium (Gould and Eldredge, 1993), with long periods of stasis while switch-like structural transitions occur over comparatively short periods.

To summarize, we have shown that protein 3D structure determination can be reliably carried out with laboratory resurrections corresponding to Precambrian nodes dating up to approximately four billion years ago, i.e., close to the origin of life. This result is remarkable, given the large number of sequence differences (up to ~50%) between the extinct and extant proteins, and demonstrates the possibility of incorporating a time scale of several billion years to expand the sequence space for 3D structure determination studies, i.e., a



**Figure 4. Statistical Distribution of Length of Helix  $\alpha 1$  for Extant Thioredoxin Structures Taken from the PDB**

(A) Query details are as follows: text search for “thioredoxin and X-ray as experimental method” was used obtaining a total of 494 structure hits. From these, all thioredoxin-related structures were discarded (i.e., thioredoxin reductases, glutaredoxins, etc.); thioredoxins from chloroplast and mitochondria as well as thioredoxins from archaea were not considered either. A total of 39 thioredoxin structures from eukaryota and 32 from bacteria were used in our analysis. Note that in some cases, the same protein structure might be overrepresented; this is the case, for instance, when different structures corresponding to mutants of the same protein are deposited in the PDB.

crystallization screenings were set up in CCD using the 24 crystallization screening kit (GSK24) and the mix of PEGs 400, 4k, and 8K kits at six different pH levels (PEG448-49; Triana S&T) in capillaries of 0.1 mm inner diameters and 50 mm length (approximately 400 nl of protein solution) at 20°C and 4°C. When crystallization failed, the next round of screening was performed using the sitting drop configuration of the vapor diffusion technique set up with a Hamilton Start-Plus robotic system with a 1:1 ratio of protein and reservoir. The drops of 200 nl were equilibrated against the reservoir filled with 50  $\mu$ l of the PEGion I or PEGRx screening kits (Hampton Research). Optimization experiments, when needed, were set up in CCD by varying the pH of the precipitant cocktail in GCBs ordered “a la carte” (Triana S&T) in capillaries of 0.1, 0.2, and 0.3 mm inner diameter. The crystallization methodologies and conditions are summarized in Table S4.

Data collection was performed at the European Synchrotron Radiation Facility using beam lines ID14-1, ID29, ID23-1, and ID23-2 from crystals cryo-cooled at 100 K. In the case of LPBCA, thioredoxin crystals were cryo-protected by supplementing the crystallization mother solution with 15% (v/v) glycerol. Crystals were extracted from the capillary, fished with a loop, and flash-cooled in liquid nitrogen. Crystals of LGPCA were kept in the capillary in which the crystal was grown. A portion of the capillary containing the selected crystal was flash-cooled in liquid nitrogen for storage before data collection began. Data were indexed and integrated with either XDS (Kabsch, 2010) and scaled and merged with Scala (Evans, 2006) of the CCP4 program suite (CCP4, 1994) or using the programs Denzo and Scalepack of the HKL2000 suite (Otwinowski and Minor, 1997).

Coordinates from the *E. coli* (Protein Data Bank [PDB] code 2TRX chain A; EcTrx from now on) or human (PDB code 1ERV, the C73S mutant, HTrx from now on) thioredoxin were used as the search model for molecular replacement using Molrep (Vagin and Teplyakov, 2010). Refinement was initiated in phenix.refine of the PHENIX suite (Adams et al., 2010) including cycles of simulated annealing followed by manual building and water inspection in Coot (Emsley et al., 2010). The latest refinement steps were run including titration-libration-screw parameterization applied to group domains with similar mobility. The final refined model was checked with Procheck (Laskowski et al., 1993) and Molprobity (Chen et al., 2010). Table 1 summarizes crystallographic data statistics and final model characteristics. The coordinates and the experimental structure factors have been deposited in the PDB and the corresponding accession codes are listed in Table 1.

Secondary structural elements were determined with DSSP (Kabsch and Sander, 1983) and Stride (Frishman and Argos, 1995). Hydrogen bonds were determined with PFIS (Hebert et al., 1998) and WHAT IF (Vriend, 1990). Accessible surface areas were calculated using a modification of the Shake-Rupley algorithm that randomly places 2,000 points on the expanded van der Waals sphere representing each atom (Ibarra-Molero et al., 1999). Charge-charge interactions were estimated using the Tanford-Kirkwood algorithm as we have previously described (Ibarra-Molero et al., 1999). The number of salt bridges was determined with a threshold of 4.0 Å by ESBRI software (Costantini et al., 2008) and WHAT IF (Vriend, 1990). The visualization and comparison of the 3D structural models were done using Pymol v1.3 (Schrödinger) and COOT (Emsley and Cowtan, 2004).

#### ACCESSION NUMBERS

The PDB accession codes for the coordinates and structure factor files reported in this paper are 2YJ7, 2YNX, 2YPM, 2YN1, 2YOI, 3ZIV, and 4BA7.

#### SUPPLEMENTAL INFORMATION

Supplemental Information includes one figure and four tables and can be found with this article online at <http://dx.doi.org/10.1016/j.str.2013.06.020>.

(B) The search in the PDB was filtered to avoid overrepresentation indicated previously. In particular, a single thioredoxin structure for each microorganism was selected (i.e., wild-type protein), resulting in a total of 14 thioredoxin structures from eukaryota and 15 from bacteria.

## ACKNOWLEDGMENTS

This work was supported by grants BIO2009-09562, BIO2012-34937, CSD2009-00088 (to J.M.S.-R.) and BIO2010-16800, "Factoría Española de Cristalización," Consolider-Ingenio 2010 (to J.A.G.) from the Spanish Ministry of Science and Innovation, FEDER Funds (to J.M.S.-R. and J.A.G.), grants from NASA Exobiology (NNX08AO12G), NASA Astrobiology Institute and Georgia Institute of Technology (to E.A.G.), and NIH grants HL066030 and HL061228 (to J.M.F.). A.I.-P. was a recipient of a FPU fellowship from the Spanish Ministry of Education. We thank Dr. Valeria A. Risso for helpful comments on the manuscript, Dr. Nick Pace for kindly providing us with the PFIS software, and the staff at the ID14-1, ID29, ID23-1, and ID23-2 beamlines of the European Synchrotron Radiation Facility, Grenoble, France, for support during data collection.

Received: April 10, 2013

Revised: June 7, 2013

Accepted: June 26, 2013

Published: August 8, 2013

## REFERENCES

- Adams, P.D., Afonine, P.V., Bunkóczi, G., Chen, V.B., Davis, I.W., Echols, N., Headd, J.J., Hung, L.-W., Kapral, G.J., Grosse-Kunstleve, R.W., et al. (2010). PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 213–221.
- Benner, S.A., Sassi, S.O., and Gaucher, E.A. (2007). Molecular paleoscience: systems biology from the past. *Adv. Enzymol. Relat. Areas Mol. Biol.* **75**, 1–132, xi.
- Caetano-Anollés, G., Wang, M., Caetano-Anollés, D., and Mittenthal, J.E. (2009). The origin, evolution and structure of the protein world. *Biochem. J.* **417**, 621–637.
- Chen, V.B., Arendall, W.B., 3rd, Headd, J.J., Keedy, D.A., Immormino, R.M., Kapral, G.J., Murray, L.W., Richardson, J.S., and Richardson, D.C. (2010). MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 12–21.
- CCP4 (Collaborative Computational Project, Number 4). (1994). The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D Biol. Crystallogr.* **50**, 760–763.
- Cordes, M.H., Walsh, N.P., McKnight, C.J., and Sauer, R.T. (1999). Evolution of a protein fold in vitro. *Science* **284**, 325–328.
- Costantini, S., Colonna, G., and Facchiano, A.M. (2008). ESBRI: a web server for evaluating salt bridges in proteins. *Bioinformatics* **3**, 137–138.
- Emsley, P., and Cowtan, K. (2004). Coot: model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.* **60**, 2126–2132.
- Emsley, P., Lohkamp, B., Scott, W.G., and Cowtan, K. (2010). Features and development of Coot. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 486–501.
- Evans, P. (2006). Scaling and assessment of data quality. *Acta Crystallogr. D Biol. Crystallogr.* **62**, 72–82.
- Finnigan, G.C., Hanson-Smith, V., Stevens, T.H., and Thornton, J.W. (2012). Evolution of increased complexity in a molecular machine. *Nature* **481**, 360–364.
- Frishman, D., and Argos, P. (1995). Knowledge-based protein secondary structure assignment. *Proteins* **23**, 566–579.
- Gaucher, E.A., Govindarajan, S., and Ganesh, O.K. (2008). Palaeotemperature trend for Precambrian life inferred from resurrected proteins. *Nature* **451**, 704–707.
- Gould, S.J., and Eldredge, N. (1993). Punctuated equilibrium comes of age. *Nature* **366**, 223–227.
- Grishin, N.V. (2001). Fold change in evolution of protein structures. *J. Struct. Biol.* **134**, 167–185.
- Hanson-Smith, V., Kolaczowski, B., and Thornton, J.W. (2010). Robustness of ancestral sequence reconstruction to phylogenetic uncertainty. *Mol. Biol. Evol.* **27**, 1988–1999.
- Harms, M.J., and Thornton, J.W. (2010). Analyzing protein structure and function using ancestral gene reconstruction. *Curr. Opin. Struct. Biol.* **20**, 360–366.
- He, Y., Chen, Y., Alexander, P.A., Bryan, P.N., and Orban, J. (2012). Mutational tipping points for switching protein folds and functions. *Structure* **20**, 283–291.
- Hebert, E.J., Giletto, A., Sevcik, J., Urbanikova, L., Wilson, K.S., Dauter, Z., and Pace, C.N. (1998). Contribution of a conserved asparagine to the conformational stability of ribonucleases Sa, Ba, and T1. *Biochemistry* **37**, 16192–16200.
- Hedges, S.B., and Kumar, S., eds. (2009). *The Timetree of Life* (New York: Oxford University Press).
- Honig, B. (2007). Protein structure space is much more than the sum of its folds. *Nat. Struct. Mol. Biol.* **14**, 458.
- Ibarra-Molero, B., Loladze, V.V., Makhatadze, G.I., and Sanchez-Ruiz, J.M. (1999). 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.
- Kabsch, W. (2010). Xds. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 125–132.
- Kabsch, W., and Sander, C. (1983). Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* **22**, 2577–2637.
- Krishna, S.S., and Grishin, N.V. (2004). Structurally analogous proteins do exist! *Structure* **12**, 1125–1127.
- Laskowski, R.A., MacArthur, M.W., Moss, D.S., and Thornton, J.M. (1993). {PROCHECK}: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.* **26**, 283–291.
- Murzin, A.G. (2008). Biochemistry. Metamorphic proteins. *Science* **320**, 1725–1726.
- Orengo, C.A., Jones, D.T., and Thornton, J.M. (1994). Protein superfamilies and domain superfolds. *Nature* **372**, 631–634.
- Ortlund, E.A., Bridgman, J.T., Redinbo, M.R., and Thornton, J.W. (2007). Crystal structure of an ancient protein: evolution by conformational epistasis. *Science* **317**, 1544–1548.
- Otálora, F., Gavira, J.A., Ng, J.D., and García-Ruiz, J.M. (2009). Counterdiffusion methods applied to protein crystallization. *Prog. Biophys. Mol. Biol.* **101**, 26–37.
- Otwinowski, Z., and Minor, W. (1997). Processing of X-ray Diffraction Data Collected in Oscillation Mode. In *Methods in Enzymology*, C.W. Carter and R.M. Sweet, eds. (New York: Academic Press), pp. 307–326.
- Perez-Jimenez, R., Inglés-Prieto, A., Zhao, Z.M., Sanchez-Romero, I., Alegre-Cebollada, J., Kosuri, P., Garcia-Manyes, S., Kappock, T.J., Tanokura, M., Holmgren, A., et al. (2011). Single-molecule paleoenzymology probes the chemistry of resurrected enzymes. *Nat. Struct. Mol. Biol.* **18**, 592–596.
- Qi, Y., and Grishin, N.V. (2005). Structural classification of thioredoxin-like fold proteins. *Proteins* **58**, 376–388.
- Risso, V.A., Gavira, J.A., Mejía-Carmona, D.F., Gaucher, E.A., and Sanchez-Ruiz, J.M. (2013). Hyperstability and substrate promiscuity in laboratory resurrections of Precambrian  $\beta$ -lactamases. *J. Am. Chem. Soc.* **135**, 2899–2902.
- Sadreyev, R.I., Kim, B.H., and Grishin, N.V. (2009). Discrete-continuous duality of protein structure space. *Curr. Opin. Struct. Biol.* **19**, 321–328.
- Schaeffer, R.D., and Daggett, V. (2011). Protein folds and protein folding. *Protein Eng. Des. Sel.* **24**, 11–19.
- Sikosek, T., Bornberg-Bauer, E., and Chan, H.S. (2012). Evolutionary dynamics on protein bi-stability landscapes can potentially resolve adaptive conflicts. *PLoS Comput. Biol.* **8**, e1002659.
- Taylor, W.R. (2007). Evolutionary transitions in protein fold space. *Curr. Opin. Struct. Biol.* **17**, 354–361.
- Tokuriki, N., and Tawfik, D.S. (2009). Protein dynamism and evolvability. *Science* **324**, 203–207.
- Vagin, A., and Teplyakov, A. (2010). Molecular replacement with MOLREP. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 22–25.

Valas, R.E., Yang, S., and Bourne, P.E. (2009). Nothing about protein structure classification makes sense except in the light of evolution. *Curr. Opin. Struct. Biol.* *19*, 329–334.

Voordeckers, K., Brown, C.A., Vanneste, K., van der Zande, E., Voet, A., Maere, S., and Verstrepen, K.J. (2012). Reconstruction of ancestral metabolic enzymes reveals molecular mechanisms underlying evolutionary innovation through gene duplication. *PLoS Biol.* *10*, e1001446.

Vriend, G. (1990). WHAT IF: a molecular modeling and drug design program. *J. Mol. Graph* *8*, 52–56, 29.

Winstanley, H.F., Abeln, S., and Deane, C.M. (2005). How old is your fold? *Bioinformatics* *21*(Suppl 1), i449–i458.

Xie, L., and Bourne, P.E. (2008). Detecting evolutionary relationships across existing fold space, using sequence order-independent profile-profile alignments. *Proc. Natl. Acad. Sci. USA* *105*, 5441–5446.