Biotechnological and protein-engineering implications of ancestral protein resurrection

Valeria A Risso[1], Jose M Sanchez-Ruiz [1#] and S. Banu Ozkan[2#]

[1] Departamento de Quimica Fisica, Facultad de Ciencias University of Granada, 18071 Granada, Spain
[2]Department of Physics and Center for Biological Physics, Arizona State University, Tempe, AZ 85281

#Corresponding Authors: sanchezr@ugr.es and Banu.Ozkan@asu.edu

Summary:
Approximations to the sequences of ancestral proteins can be derived from the sequences of their modern descendants. Proteins encoded by such reconstructed sequences can be prepared in the laboratory and subjected to experimental scrutiny. These “resurrected” ancestral proteins often display remarkable properties, reflecting ancestral adaptations to intra-cellular and extra-cellular environments that differed from the environments hosting modern/extant proteins. Recent experimental and computational work has specifically discussed high stability, substrate and catalytic promiscuity, conformational flexibility/diversity and altered patterns of interaction with other sub-cellular components. In this review, we discuss these remarkable properties as well as recent attempts to explore their biotechnological and protein-engineering potential.

Introduction

Plausible approximations to words in ancient languages can be derived from their modern descendant words by using suitable models of language evolution. The common ancestor of a modern language family (an extinct Proto-language) can thus be reconstructed [1]. As a well-known example, historical linguists worked on the reconstruction of Proto-Indo-European, the common ancestor of the Indo-European language family, already in the XIX century [2]. Likewise, plausible approximations to the sequences of ancestral proteins can be derived from the sequences of their modern descendants [3], since a protein sequence can be considered as a word written using an alphabet of 20 letters. The overall procedure is called ancestral sequence reconstruction, and involves phylogenetic and statistical analyses that use simple models of sequence evolution [4]. Proteins encoded by the ancestral reconstructed sequences can be prepared in the laboratory and subjected to experimental scrutiny. Such "resurrected ancestral proteins", to use the accepted term in the field, have been extensively used to explore relevant evolutionary processes and hypothesis. This work has been covered in excellent reviews [5–9].

Besides their use over the last ~25 years as molecular tools to address important evolutionary issues, more recent literature suggests the biotechnological potential of resurrected ancestral proteins [10–24*]. The interest on practical applications arises in part because ancestral proteins are perceived as being "different" from modern/extant proteins. Ancestral proteins certainly differ from their modern counterparts in terms of sequence, in particular when "old" phylogenetic nodes are targeted. Indeed, reconstructed sequences of Precambrian proteins often show large numbers of amino acid differences with their modern descendants. More relevant, however, is the fact that ancestral proteins were adapted to intra-cellular and extra-cellular environments that likely differed from the environments hosting modern proteins. As a result, resurrected ancestral proteins could be expected display "unusual" or "extreme" properties to some extent. Experimental and computational work has specifically discussed high stability, substrate and catalytic promiscuity, conformational flexibility/diversity and altered patterns of interaction with other sub-cellular components. In this review, we summarize and discuss this recent work as well as very recent attempts to explore the biotechnological and protein-engineering potential of resurrected ancestral proteins.

Altered patterns of interaction with other sub-cellular components

The biological function of proteins involves interactions with other sub-cellular components, including, in many cases, other proteins. Modern proteins are, therefore, adapted to a substantial extent to modern cellular environments, because they have co-evolved with their interaction partners. Consequently, replacing a modern protein with a representation of one of its ancestors is expected to impair to some extent the fitness of the modern host organism [23*,25]. Nevertheless, recent work suggests that the altered patterns of interactions of ancestral proteins may be useful in biotechnological or biomedical application scenarios. Particularly, two examples in protein folding and virus-host interactions based on very recent works [23*,26] highlight the impact of utilizing ancestral reconstruction in protein biotechnology as discussed below.

Protein folding is a complex process that is assisted *in vivo* by chaperones [27]. Molecular chaperones are, of course, an outcome of evolution. Ancient proteins likely

had to fold without the assistance of chaperones or, perhaps, with the assistance of chaperones that were not as efficient as their modern counterparts are. Thus, efficient folding in ancient proteins, therefore, may have been encoded at the level of sequence to some extent. Plausibly, however, ancestral sequence determinants of efficient folding may have been lost during evolutionary history as efficient molecular chaperones evolved. Although these notions remain to be fully explored and tested, they are supported by preliminary experimental work on the folding kinetics of resurrected Precambrian thioredoxins [26]. Ancestral determinants of efficient folding may plausibly have contributed, together with other factors, to the enhanced expression levels recently reported for some resurrected ancestral proteins [15**,28]. High expression levels are certainly convenient when preparing proteins of biotechnological interest. More critically, they may enhance in vivo function of the protein drug [15**].

Viruses typically code for a rather small number of proteins. Therefore, they rely on recruiting proteins from the hosts for essential processes involved in infection and propagation. Such recruited proteins are known as proviral factors. Viruses and their hosts co-evolve. Modern viruses have, therefore, adapted to recruit modern proviral factors. It follows that replacing a modern proviral factor with a functional ancestral form may perhaps render the host resistant to virus infection. A proof of concept of this notion has been recently reported [23*] using the infection of *E. coli* by the bacteriophage T7 as a model system. Phage T7 recruits *E. coli* thioredoxin for its replisome [29]. Some resurrected Precambrian thioredoxins showed somewhat decreased, but still substantial levels of "normal" redox functionality within *E. coli*. However, these ancestral thioredoxins could not be recruited by the phage and rendered *E. coli* resistant to infection. The authors [23*] discussed the possibility of applying this approach to the important problem of the engineering of virus resistance in plants.

Enhanced stability

A remarkable large number of studies have reported substantial stability enhancements upon ancestral protein resurrection, in particular when targeting "old" Precambrian nodes [10,14**,19,20,30–34]. In our view, the high stability of resurrected ancestral proteins most likely reflects a high-temperature environment for ancient life. Indeed, many different scenarios are consistent with a hot start for life and/or with ancient life being thermophilic. These include, for instance, the origin of life in hydrothermal vents [35], the possibility that only tough thermophilic organisms survived catastrophic extra-terrestrial impacts in the young planet (the so-called "impact bottleneck" scenarios) [36] and that the ancient oceans that hosted life were hot [37]. The primordial origin of the enhanced stability of resurrected ancestral proteins is consistent with recent work that supports site-specific amino acid preferences in proteins to be conserved to some substantial extent over evolutionary history [38–41]. Since stability is a major factor contributing to amino acid preferences, mutational effects on stability are also conserved to some substantial extent [39,40]. This supports the reliability of the reconstruction of primordial stability and rationalizes the stabilizing effect of back-to-the-predicted-ancestor mutations. Thus, while destabilizing mutations may be accepted upon cooling of the environment, the corresponding back-to-the-ancestor mutations will remain available for stabilization when this is required. This may occur when a local environment imposes again a high temperature or when other factors, such as oxidative stress or high radiation levels [42], confer stabilization with a selective advantage.

According to this interpretation, the high stability reported for some comparatively "young" resurrected ancestral enzymes [42] may be a simple recapitulation of the primordial trait.

On the other hand, the high stability of resurrected ancestral proteins can hardly be explained as an "artifact" or "bias" of the sequence reconstruction procedures, as it has been occasionally suggested. The increments in denaturation temperature obtained upon ancestral protein resurrection are often on the order of a few tens of degrees. They are, therefore, larger than computational estimates of stability biases of ancestral reconstruction, which are on the order of a few degrees [43]. They are also larger than the most denaturation temperature increments obtained through rational design or directed evolution (compare, for instance, with the experimental data reviewed in [44]).

Regardless of its origin, however, high stability is a very convenient property from a biotechnological point of view because low stability compromises many practical applications of proteins [44–47]. Also, from a protein-engineering point of view, enhanced stability may be essential as it contributes to high evolvability [48] by allowing destabilizing, but functionally benificial mutations to be accepted. Finally, enhanced stability may improve pharmacokinetics of protein drugs [12**]. Overall, we foresee that ancestral resurrection may become in the near future a common source to create stabile variants of proteins of biotechnological interest. This is all the more so as mutational comparison between ancestral nodes may lead to further stabilization (Figure 1) [49].

Several approaches for protein stabilization are also available [50] and, including the search for stable proteins in thermophilic organisms [44]. While the ancestral proteins generally show higher thermal stability compared to the extant proteins [31,33], it is also observed that the thermophilic extant protein may exhibit higher stability [51]. However, resurrected ancestral proteins, in particular those corresponding to ancient phylogenetic nodes, often show large sequence differences with their modern counterparts, including the sequences of proteins from modern thermophlic organisms. As an example, denaturation temperature values of ~113 ºC and ~103 ºC have been reported, respectively, for the resurrected thioredoxin corresponding to the last common ancestor of bacteria (an organism that existed ~4 billion years ago) and the thioredoxin from the modern thermophile *Bacillus acidocaldarius*. Still, these two proteins show only about 56% sequence identity [31,52]. Clearly, ancestral reconstruction provides an excellent approach to explore sequence space, and may yield information that would complement that obtained from other stabilization approaches. In particular, it will be of interest to determine the extent to which ancestral proteins and modern thermophilic proteins share common sequence, structure and energetics determinants of high stability. Thus, exploring the differences between molecular mechanism of stability of the ancestral and that of thermophilic extant sequences would provide additional tools for biotechnology.

Finally, promiscuity is also a common outcome of ancestral resurrection (see below) which is not incompatible with high stability. Ancestral resurrection may provide, therefore, an approach to protein scaffolds in which two properties of biotechnological interest are combined. Indeed, resurrected ancestral lactamases [10] have been shown to simultaneously display high stability (denaturation temperatures up to about 30 degrees above those of their modern mesophilic counterparts) and moderately efficient substrate

promiscuity (capability to degrade different lactam antibiotics with catalytic efficiencies similar to an average modern enzyme). Likewise, Kazlauskas, Dean and coworkers [14**] found that catalytic promiscuity of resurrected ancestral esterases and hydroxynitrile lyases did not come at a cost of stability. In fact, the denaturation temperature of the catalytically promiscuous ancestor (~80 ºC) was higher than that for its modern descendants (54-70 ºC).

Promiscuity

Although enzymes are sometimes described as efficient specialists, there appears to be no fundamental constraint to the number of tasks a protein can perform. Enzymes involved in detoxication, for instance, are highly promiscuous and can degrade a wide variety of toxics through different chemical routes [53,54]. Certainly, many enzymes carry out only one physiologically relevant function. Even in these cases, however, low-level activities with no known physiological relevance are usually observed [55,56]. This kind of promiscuity is often considered as a vestige of the proposed generalist nature of primordial enzymes [57–59].

An application may require an enzyme to catalyze a reaction that is related to, but not identical to the physiological reaction. A promiscuous, low-level activity will provide the essential starting point in the laboratory directed evolution of an efficient catalyst for the biotechnologically useful reaction. Indeed, the exponential increase in the number of papers on applications of enzymes to the transformation of non-natural products in the period 1970-1990 [60] has been linked (see chapter 10 in [61]) to the realization that enzymes are promiscuous catalysts.

Unfortunately, promiscuity is an accidental property in most modern proteins. Searching for promiscuity in Nature is, therefore, considered to be inefficient [14**]. On the other hand, promiscuity appears to be a common outcome of ancestral protein resurrection. Thornton and coworkers have recently reviewed experimental resurrection studies on 15 protein families [62]. They report that, for most families (11 out of 15), evolution involved function partitioning from a multi-functional ancestor, while *de novo* evolution of a new function was observed in only 4 protein families (see Table 1 in [62]). We suggest that the simplest, Occam-razor explanation of this result is that primordial enzymes were generalists with broad substrate scope [57,58] and, consequently, “traveling back in time” through ancestral reconstruction increases the probability of finding substantial levels of promiscuity. Still, it is also possible, as suggested by Thornton and coworkers [62], that the preponderance of function evolution trough partitioning from multi-functional ancestors (*versus de novo* evolution) is explained by higher chances of “survival” of the new function, which may become biologically significant during the pre-duplication period, when the single gene is protected from degeneration. These differences in interpretation should not distract us from the essential experimental result that many ancestral resurrection efforts have led to multifunctional (promiscuous) proteins. We foresee, therefore, that ancestral protein resurrection may become in the near future a common source of promiscuous proteins for biotechnological and protein-engineering applications.

We note, finally, that the fact that promiscuity is a common outcome of ancestral resurrection does not rule out the possibility that, in some cases at least, ancestral proteins show enhanced levels of activity compared to their modern descendants [63]. A

particularly relevant example of this scenario has been recently reported by Gaucher and coworkers [12**]. Ancestral protein resurrection showed that uricases, the enzymes that metabolize uric acid, have progressively lost activity in the evolutionary line that leads from the last common ancestor mammals to the apes (including humans). The concomitant increases in uric acid levels plausibly allowed our ancestors to accumulate fat through the metabolism of fructose from fruit. In very simple terms, uric acid up-regulates key proteins in the metabolism of fructose which, unlike the metabolism of other sugars, increases fat stores and protects animals from starvation. This may have provided a survival advantage in periods of environmental cooling and diet changes. As an important biomedical outcome of this study, the high activity and enhanced *in vivo* stability of ancestral uricases suggest their potential therapeutic value in the treatment of gout [12**].

Conformational flexibility/diversity

We now know that proteins dynamically interconvert between conformations in the native state to achieve their function [64]. Simply, proteins possess an ensemble of conformations in their native state. It is this ensemble that it is involved in various biological functions, including allosteric signaling [65], protein-ligand recognition, and protein-protein recognition [66,67], electron transfer [68] and catalysis [69–71].

In the ensemble model, a protein samples a variety of conformations through local changes such as loop motions, side-chain rotations, or global changes through domain rearrangement. Allostery, commonly known as regulation at a distance, is a widely used emergent property of this ensemble view. Rather than forming a new structure, a ligand binding to a remote site promotes a shift in dynamics, changing the intrinsic structure-encoded dynamics and dynamic linking (i.e., the distribution of accessible conformational states in the ensemble), promoting easy access to certain conformers for allosteric regulations [65,67,72]. Furthermore, the ensemble view also agrees with the evolutionary adaptability of a protein in which the same conserved 3D native fold can adopt new functions [73*]. Mutations throughout protein evolution alter conformational dynamics, shifting the distribution of the ensemble and lead to the emergence of new functions [70,74] and adaption to different environments [75].

In recent years, computational protein design methods have been used to introduce completely novel enzymatic functions in protein scaffolds initially lacking these abilities (i.e. *de novo* enzyme design) [76–78]. Despite these notable successes, the activities of the designed enzymes are almost universally orders of magnitude lower than their naturally occurring counterparts [79,80], suggesting that our understanding of the intricacies of enzymatic processes is likely incomplete.

State of the art enzyme design algorithms based on the Pauling postulate of transition state stabilization [81] do not likely fully capture all components of enzymatic function. Design efforts are focused on sculpting artificial active sites through the introduction and stabilization of catalytic residues, often within existing cavities in proteins of known structure. Although the role of dynamics in designed enzymes function has been explored through QM/MM [82], DFT [83], and MD [84] simulations, these studies were limited to active site residues due to the computational expense associated with application of these analyses to full proteins. To date, no large-scale enzyme design

efforts have been carried out in which the ensemble of conformation and each position role in the conformational search were considered either during or after design.

Ancestral protein resurrection offers an excellent opportunity to address many of the issues raised above. As expounded in the preceding section, promiscuity is a common outcome of ancestral protein resurrection. Furthermore, it is widely accepted that enzyme promiscuity is linked to conformational flexibility/diversity [69,71,85]. In the simplest picture, enzymes exist as an ensembles of conformations, with different conformations being responsible for the different activities of a promiscuous protein. A number of recent studies [69,70,74,85–88] support a fundamental role for conformational diversity in functional evolution. In the simplest interpretation, mutations can shift the conformational equilibria towards (previously) minor conformations responsible for new enzyme functions.

For several protein families, resurrected ancestral proteins have been reported to share the same 3-D structure as their modern homologs and yet to function differently. The change of conformational dynamics as function evolves has recently been studied in three ancestral steroid receptors (the ancestors of mineralocorticoid and glucocorticoid receptor proteins) [89]. Mineralocorticoid and glucocorticoid receptors (MR and GR) arose by duplication of a single ancestor (AncCR) deep in the vertebrate lineage and then diverged in function. While AncCR are AncGR1 have a promiscuous binding showing binding affinity to both aldosterone, cortisol, AncGR2 specifically binds to cortisol. AncGR1 and AncGR2, which diverge functionally through 36 mutations, have highly similar experimental structures. However, a comparison of the conformational dynamics of the three ancestral proteins reveals AncCR and AncGR1 have a flexible binding pocket, suggesting flexibility plays a role in promiscuous binding affinity. In contrast, the mutations of AncGR2 lead to a rigid binding pocket, suggesting that, as the binding pocket becomes cortisol specific, evolution acts to shape the binding pocket toward a specific ligand [89].

Similar to the promiscuous ancestors of mineralocorticoid and glucocorticoid receptors, proteins corresponding to 2-3 billion year old Precambrian nodes in the evolution of Class A β-lactamases have been shown [10] to degrade a variety of antibiotics with catalytic efficiency levels similar to those of an average enzyme [90] (Figure 2). Consequently, ancestral lactamases can be described as moderately efficient promiscuous catalysts. Remarkably, there are only a few (and minor) structural differences (in particular, at the active-site regions) between the resurrected ancestral enzymes and penicillin- specialist modern β-lactamases [10]. This then raises the question whether the functional differences arise from the conformational dynamics of the lactamases. The dynamics of the lactamases were simulated using Molecular Dynamics and the covariance matrix was calculated and analyzed using Perturbation Response Scanning (PRS) [91] to calculate the Dynamic Flexibility Index *(DFI)* [92,93], a site specific measure to compute the contribution each position to the functionally relevant conformational dynamics. Because *DFI* is a position specific metric, it also allows us to quantify the change in flexibility per position throughout the evolution by identifying flexible and rigid position within the 3-D interaction network of protein structure. The low *DFI* sites are rigid sites (i.e. hinge sites). They are robust to perturbation occur at any part of the chain (i.e. in term of response fluctuation upon positional changes in other part of the change), yet transfer the perturbation response efficiently to rest of the protein as joints in skeleton. High *DFI* regions on the other

hand shows high response, thus these are more deformable sites.

The special dynamics associated to substrate promiscuity of ancestral β-lactamases was revealed by patterns of high *DFI* values in regions close to the active site illuminating the flexibility required for the binding and catalysis of different ligands. These specific *DFI* patterns suggest that the protein native state is actually an ensemble of conformations displaying the structural variability in the active site region required for efficient binding of substrates of different sizes and shapes. On the other hand, *DFI* analysis of modern TEM-1 lactamase shows a comparatively rigid active-site region, likely reflecting adaptation for efficient degradation of a specific substrate, penicillin [94*] (Figure 2).

Thioredoxins achieved adaptation to a cooler and less acidic Earth by altering their stability and changing their catalytic rates while maintaining the same 3-D fold [31,95]. Comparison of the distribution of flexibility of residues between ancestral and extant thierodoxins reveals that the population density of very high flexible sites and rigid sites increased, as they evolved. These common feature of changing the flexibility of specific positions observed in evolution suggest a "*fine tuning*" of their native ensemble to adjust to ambient conditions in accordance with the evolution in their function [96]. These subtle changes in rigidity and flexibility of specific positions achieve required flexibility near active sites for promiscuity which is compensated by increased rigidity of distal sites, thus maintaining the stability of the ancestral proteins.

DFI analysis further reveals how functional evolution is related to changes in flexibility, specifically at hinge points (i.e. low *DFI* sites), even as the protein structure remains largely unchanged. The *DFI* analysis of reconstructed ancestral proteins of green fluorescent protein (GFP) shows the evolution of red color from a green ancestor emerged by migration of the hinge point (i.e. *low DFI region)* from the active site diagonally across the beta-barrel fold [97*]. While the flexibility of the mutational sites does not change significantly, in response to these mutations, both increase in flexibility and decrease in flexibility occurs for regions of the beta-fold that are widely separated from the mutational sites, indicating allosteric regulation in evolution. Nature introduces mutations at relatively flexible sites farther away from functionally critical sites, yet allosterically alter the flexibility of functionally critical active sites. Thus, Nature utilizes minimum perturbation maximum response as a principle through allosterically altering the dynamics of the functionally critical sites, rather than introducing mutations on these sites.

Overall, ancestral reconstruction studies provide a unique opportunity to address and understand the relation between conformational dynamics, protein evolution and protein function. At a more applied level, flexible proteins derived from ancestral resurrection may provide useful scaffolds for the engineering of new enzyme functionalities. This is so because conformational flexibility/diversity should facilitate the binding of substrates and transition states for enzyme-catalyzed reactions through the sampling of many potentially productive conformations. This notion is supported by recent work that used β-lactamases as scaffolds for the generation of new active-sites. A simple minimalist design was found to lead to substantial levels of a *de novo* Kemp-elimination activity when using flexible Precambrian proteins as scaffolds, but failed in the more rigid modern lactamases (Figure 3) [24*].

Concluding remarks

In 1963, Linus Pauling and Emile Zuckerkandl stated that it would be possible one day to infer the gene sequences of ancestral species to "synthesize these presumed components of extinct organisms...and study the physico-chemical properties of these molecules". 55 years later, the large number of sequences available in the post-genomic era, together with advances in bioinformatics and molecular biology methodologies, have contributed to make their statement true for a substantial number of protein systems. Often, resurrected ancestral proteins have been found to display high stability and enhanced promiscuity, features that are immediately advantageous in biotechnological application scenarios. Furthermore, detailed computational conformational analyses support that ancestral proteins may have evolved to new or more specific modern functions by altering their ensemble of conformational states while preserving the 3-D structure. In addition to precisely positioning amino acid residues in catalytically competent orientations within the active site, nature has evolved unique networks of interactions that enable communication between the active site and the rest of the of protein through dynamic motions These correlated dynamic motions appear to facilitate all important steps in catalytic reactions including substrate recognition, catalysis, and substrate release. Thus, efforts to develop the next generation of computational enzyme-engineering tools must not only address the precise conformation of the active site, but also the associated dynamic motion profile of the protein scaffold.

Acknowledgements

VAR and JMS-R are supported by Feder Funds, Grant BIO2015-66426-R from the Spanish Ministry of Economy and Competitiveness and Grant RGP0041/2017 from the Human Frontier Science Program. Support from NSF-MCB Award 1715591 and Scialog Fellow Award by RCS and the Gordon & Betty Moore Foundation is gratefully acknowledged by SBO.

References and recommended reading (

1. Atkinson QD: **The descent of words**. *Proc Natl Acad Sci* 2013, **110**:4159–4160.

2. Clackson J: *Indo-European Linguistics: An Introduction*. Cambridge University Press; 2007.

3. Pauling L, Zuckerkandl E: **Chemical paleogenetics**. *Acta Chem Scand* 1963, **17**:S9–S16.

4. Liberles DA: *Ancestral sequence reconstruction*. Oxford University Press on Demand; 2007.

5. Benner SA, Sassi SO, Gaucher EA: **Molecular paleoscience: systems biology from the past**. *Adv Enzym Relat Areas Mol Biol* 2007, **75**:1–132.

6. Harms MJ, Thornton JW: **Analyzing protein structure and function using ancestral gene reconstruction**. *Curr Opin Struct Biol* 2010, **20**:360–366.

7. Harms MJ, Thornton JW: **Evolutionary biochemistry: revealing the historical and physical causes of protein properties**. *Nat Rev Genet* 2013, **14**:559.

8. Merkl R, Sterner R: **Ancestral protein reconstruction: techniques and applications**. *Biol Chem* 2016, **397**:1–21.

9. Gumulya Y, Gillam EM: **Exploring the past and the future of protein evolution with ancestral sequence reconstruction: the 'retro'approach to protein engineering**. *Biochem J* 2017, **474**:1–19.

10. Risso VA, Gavira JA, Mejia-Carmona DF, Gaucher EA, Sanchez-Ruiz JM: **Hyperstability and substrate promiscuity in laboratory resurrections of Precambrian β-lactamases**. *J Am Chem Soc* 2013, **135**:2899–2902.

11. Cox VE, Gaucher EA: **Engineering proteins by reconstructing evolutionary adaptive paths**. In *Directed Evolution Library Creation*. . Springer; 2014:353–363.

12. Kratzer JT, Lanaspa MA, Murphy MN, Cicerchi C, Graves CL, Tipton PA, Ortlund EA, Johnson RJ, Gaucher EA: **Evolutionary history and metabolic insights of ancient mammalian uricases**. *Proc Natl Acad Sci* 2014, **111**:3763–3768.

** The authors show that uricases, the enzymes that metabolize uric acid, have progressively lost activity since the last common ancestor of mammals, likely because this allowed our ancestors to accumulate fat from the metabolism of fructose. As an important biomedical implication, the high activity and enhanced *in vivo* stability of ancestral uricases suggest their potential therapeutic value in the treatment of gout.

13. Whitfield JH, Zhang WH, Herde MK, Clifton BE, Radziejewski J, Janovjak H, Henneberger C, Jackson CJ: **Construction of a robust and sensitive arginine biosensor through ancestral protein reconstruction**. *Protein Sci* 2015, **24**:1412–1422.

14. Devamani T, Rauwerdink AM, Lunzer M, Jones BJ, Mooney JL, Tan MAO, Zhang Z-J, Xu J-H, Dean AM, Kazlauskas RJ: **Catalytic promiscuity of ancestral esterases and hydroxynitrile lyases**. *J Am Chem Soc* 2016, **138**:1046–1056.

** Promiscuity is a biotechnologically important property. Search for promiscuity is, however, inefficient because it is an accidental property in modern proteins. This work supports that ancestral reconstruction targeting pre-duplication nodes may provide a convenient route to promiscuous proteins.

15. Zakas PM, Brown HC, Knight K, Meeks SL, Spencer HT, Gaucher EA, Doering CB: **Enhancing the pharmaceutical properties of protein drugs by ancestral sequence reconstruction**. *Nat Biotechnol* 2017, **35**:35.

** The authors use ancestral reconstruction to engineer variants of coagulation factor VIII with enhanced pharmaceutical properties. The work provides strong support for the biotechnological potential of ancestral protein resurrection.

16. Alcalde M: **When directed evolution met ancestral enzyme resurrection**. *Microb Biotechnol* 2017, **10**:22–24.

17. Clifton BE, Whitfield JH, Sanchez-Romero I, Herde MK, Henneberger C, Janovjak H, Jackson CJ: **Ancestral Protein Reconstruction and Circular Permutation for Improving the Stability and Dynamic Range of FRET Sensors**. *Synth Protein Switch Methods Protoc* 2017,

18. Holinski A, Heyn K, Merkl R, Sterner R: **Combining ancestral sequence reconstruction with protein design to identify an interface hotspot in a key metabolic enzyme complex**. *Proteins Struct Funct Bioinforma* 2017, **85**:312–321.

19. Babkova P, Sebestova E, Brezovsky J, Chaloupkova R, Damborsky J: **Ancestral Haloalkane Dehalogenases Show Robustness and Unique Substrate Specificity**. *ChemBioChem* 2017, **18**:1448–1456.

20. Ayuso-Fernández I, Martínez AT, Ruiz-Dueñas FJ: **Experimental recreation of the evolution of lignin-degrading enzymes from the Jurassic to date**. *Biotechnol Biofuels* 2017, **10**:67.

21. Wilding M, Peat TS, Kalyaanamoorthy S, Newman J, Scott C, Jermiin LS: **Reverse engineering: transaminase biocatalyst development using ancestral sequence reconstruction**. *Green Chem* 2017, **19**:5375–5380.

22. Blanchet G, Alili D, Protte A, Upert G, Gilles N, Tepshi L, Stura EA, Mourier G, Servent D: **Ancestral protein resurrection and engineering opportunities of the mamba aminergic toxins**. *Sci Rep* 2017, **7**:2701.

23. Delgado A, Arco R, Ibarra-Molero B, Sanchez-Ruiz JM: **Using resurrected ancestral proviral proteins to engineer virus resistance**. *Cell Rep* 2017, **19**:1247–1256.

* Proviral factors are host proteins recruited by viruses for processes essential for propagation. The authors show that replacing a suitable modern proviral factor with an ancestral form of the same protein renders *E. coli* resistant to the bacteriophage T7. The authors also discuss how this approach could be applied to the engineering of virus resistance in plants.

24. Risso VA, Martinez-Rodriguez S, Candel AM, Krüger DM, Pantoja-Uceda D, Ortega-Munoz M, Santoyo-Gonzalez F, Gaucher EA, Kamerlin SC, Bruix M: **De novo active sites for resurrected Precambrian enzymes**. *Nat Commun* 2017, **8**:16113.

* A minimalist design that generates a *de novo* enzymatic activity in several resurrected Precambrian β-lactamases fails when using ten different modern β-lactamases as scaffolds. Enhanced conformational flexibility is shown to play a crucial role in the capability of the ancestral proteins to generate a new function. This work supports the potential of ancestral reconstruction in enzyme engineering.

25. Hobbs JK, Prentice EJ, Groussin M, Arcus VL: **Reconstructed ancestral enzymes impose a fitness cost upon modern bacteria despite exhibiting favourable biochemical properties**. *J Mol Evol* 2015, **81**:110–120.

26. Candel AM, Romero-Romero ML, Gamiz-Arco G, Ibarra-Molero B, Sanchez-Ruiz JM: **Fast folding and slow unfolding of a resurrected Precambrian protein**. *Proc Natl Acad Sci* 2017,

27. Hartl FU, Hayer-Hartl M: **Converging concepts of protein folding in vitro and in vivo**. *Nat Struct Mol Biol* 2009, **16**:574.

28. Manteca A, Schönfelder J, Alonso-Caballero A, Fertin MJ, Barruetabeña N, Faria BF, Herrero-Galán E, Alegre-Cebollada J, De Sancho D, Perez-Jimenez R: **Mechanochemical evolution of the giant muscle protein titin as inferred from resurrected proteins**. *Nat Struct Mol Biol* 2017, **24**:652.

29. Hamdan SM, Richardson CC: **Motors, switches, and contacts in the replisome**. *Annu Rev Biochem* 2009, **78**:205–243.

30. Gaucher EA, Govindarajan S, Ganesh OK: **Palaeotemperature trend for Precambrian life inferred from resurrected proteins**. *Nature* 2008, **451**:704.

31. Perez-Jimenez R, Inglés-Prieto A, Zhao Z-M, Sanchez-Romero I, Alegre-Cebollada J, Kosuri P, Garcia-Manyes S, Kappock TJ, Tanokura M, Holmgren A: **Single-molecule paleoenzymology probes the chemistry of resurrected enzymes**. *Nat Struct Mol Biol* 2011, **18**:592.

32. Akanuma S, Nakajima Y, Yokobori S, Kimura M, Nemoto N, Mase T, Miyazono K, Tanokura M, Yamagishi A: **Experimental evidence for the thermophilicity of ancestral life**. *Proc Natl Acad Sci* 2013, **110**:11067–11072.

33. Nguyen V, Wilson C, Hoemberger M, Stiller JB, Agafonov RV, Kutter S, English J, Theobald DL, Kern D: **Evolutionary drivers of thermoadaptation in enzyme catalysis**. *Science* 2016,

34. Garcia AK, Schopf JW, Yokobori S, Akanuma S, Yamagishi A: **Reconstructed ancestral enzymes suggest long-term cooling of Earth's photic zone since the Archean**. *Proc Natl Acad Sci* 2017, **114**:4619–4624.

35. Martin W, Baross J, Kelley D, Russell MJ: **Hydrothermal vents and the origin of life**. *Nat Rev Microbiol* 2008, **6**:805.

36. Nisbet EG, Sleep NH: **The habitat and nature of early life**. *Nature* 2001, **409**:1083.

37. Robert F, Chaussidon M: **A palaeotemperature curve for the Precambrian oceans based on silicon isotopes in cherts**. *Nature* 2006, **443**:969.

38. Ashenberg O, Gong LI, Bloom JD: **Mutational effects on stability are largely conserved during protein evolution**. *Proc Natl Acad Sci* 2013, **110**:21071–21076.

39. Risso VA, Manssour-Triedo F, Delgado-Delgado A, Arco R, Barroso-delJesus A, Ingles-Prieto A, Godoy-Ruiz R, Gavira JA, Gaucher EA, Ibarra-Molero B: **Mutational studies on resurrected ancestral proteins reveal conservation of site-specific amino acid preferences throughout evolutionary history**. *Mol Biol Evol* 2014, **32**:440–455.

40. Doud MB, Ashenberg O, Bloom JD: **Site-specific amino acid preferences are mostly conserved in two closely related protein homologs**. *Mol Biol Evol* 2015, **32**:2944–2960.

41. Chan YH, Venev SV, Zeldovich KB, Matthews CR: **Correlation of fitness landscapes from three orthologous TIM barrels originates from sequence and structure constraints**. *Nat Commun* 2017, **8**:14614.

42. Trudeau DL, Kaltenbach M, Tawfik DS: **On the potential origins of the high stability of reconstructed ancestral proteins**. *Mol Biol Evol* 2016, **33**:2633–2641.

43. Williams PD, Pollock DD, Blackburne BP, Goldstein RA: **Assessing the accuracy of ancestral protein reconstruction methods**. *PLoS Comput Biol* 2006, **2**:e69.

44. Wijma HJ, Floor RJ, Janssen DB: **Structure-and sequence-analysis inspired engineering of proteins for enhanced thermostability**. *Curr Opin Struct Biol* 2013, **23**:588–594.

45. Sanchez-Romero I, Ariza A, Wilson KS, Skjøt M, Vind J, De Maria L, Skov LK, Sanchez-Ruiz JM: **Mechanism of protein kinetic stabilization by engineered disulfide crosslinks**. *PLoS One* 2013, **8**:e70013.

46. Moore EJ, Zorine D, Hansen WA, Khare SD, Fasan R: **Enzyme stabilization via computationally guided protein stapling**. *Proc Natl Acad Sci* 2017, **114**:12472–12477.

47. **Understanding Enzymes: Function, Design, Engineering, and Analysis**. *CRC Press* 2016,

48. Bloom JD, Labthavikul ST, Otey CR, Arnold FH: **Protein stability promotes evolvability**. *Proc Natl Acad Sci* 2006, **103**:5869–5874.

49. Romero-Romero ML, Risso VA, Martinez-Rodriguez S, Ibarra-Molero B, Sanchez-Ruiz JM: **Engineering ancestral protein hyperstability**. *Biochem J* 2016, **473**:3611–3620.

50. Magliery TJ: **Protein stability: computation, sequence statistics, and new experimental methods**. *Curr Opin Struct Biol* 2015, **33**:161–168.

51. Hart KM, Harms MJ, Schmidt BH, Elya C, Thornton JW, Marqusee S: **Thermodynamic system drift in protein evolution**. *PLoS Biol* 2014, **12**:e1001994.

52. Romero-Romero ML, Risso VA, Martinez-Rodriguez S, Gaucher EA, Ibarra-Molero B, Sanchez-Ruiz JM: **Selection for Protein Kinetic Stability Connects Denaturation**

Temperatures to Organismal Temperatures and Provides Clues to Archaean Life. *PLOS ONE* 2016, **11**:e0156657.

53. Zhang W, Dourado DF, Fernandes PA, Ramos MJ, Mannervik B: **Multidimensional epistasis and fitness landscapes in enzyme evolution**. *Biochem J* 2012, **445**:39–46.

54. Sheehan D, MEADE G, FOLEY VM: **Structure, function and evolution of glutathione transferases: implications for classification of non-mammalian members of an ancient enzyme superfamily**. *Biochem J* 2001, **360**:1–16.

55. Copley SD: **An evolutionary biochemist's perspective on promiscuity**. *Trends Biochem Sci* 2015, **40**:72–78.

56. Copley SD: **Shining a light on enzyme promiscuity**. *Curr Opin Struct Biol* 2017, **47**:167–175.

57. Jensen RA: **Enzyme recruitment in evolution of new function**. *Annu Rev Microbiol* 1976, **30**:409–425.

58. O'Brien PJ, Herschlag D: **Catalytic promiscuity and the evolution of new enzymatic activities**. *Chem Biol* 1999, **6**:R91–R105.

59. Tawfik OK, S D: **Enzyme promiscuity: a mechanistic and evolutionary perspective**. *Annu Rev Biochem* 2010, **79**:471–505.

60. Faber K: **Biotransformation of non-natural compounds: State of the art and future development**. *Pure Appl Chem* 1997, **69**:1613–1632.

61. Grunwald P: *Biocatalysis: biochemical fundamentals and applications*. Imperial College Press; 2009.

62. Siddiq MA, Hochberg GK, Thornton JW: **Evolution of protein specificity: insights from ancestral protein reconstruction**. *Curr Opin Struct Biol* 2017, **47**:113–122.

63. Newton MS, Arcus VL, Patrick WM: **Rapid bursts and slow declines: on the possible evolutionary trajectories of enzymes**. *J R Soc Interface* 2015, **12**:20150036.

64. Bahar I, Lezon TR, Yang L-W, Eyal E: **Global dynamics of proteins: bridging between structure and function**. 2010,

65. Guo J, Zhou H-X: **Protein allostery and conformational dynamics**. *Chem Rev* 2016, **116**:6503–6515.

66. Kumar A, Butler BM, Kumar S, Ozkan SB: **Integration of structural dynamics and molecular evolution via protein interaction networks: a new era in genomic medicine**. *Curr Opin Struct Biol* 2015, **35**:135–142.

67. Nussinov R, Tsai C-J: **Allostery in disease and in drug discovery**. *Cell* 2013, **153**:293–305.

68. Martin DR, Ozkan SB, Matyushov DV: **Dissipative electro-elastic network model of protein electrostatics**. *Phys Biol* 2012, **9**:036004.

69. Tokuriki N, Tawfik DS: **Protein dynamism and evolvability**. *Science* 2009, **324**:203–207.

70. Campbell E, Kaltenbach M, Correy GJ, Carr PD, Porebski BT, Livingstone EK, Afriat-Jurnou L, Buckle AM, Weik M, Hollfelder F: **The role of protein dynamics in the evolution of new enzyme function**. *Nat Chem Biol* 2016, **12**:944.

71. Pabis A, Risso VA, Sanchez-Ruiz JM, Kamerlin SC: **Cooperativity and flexibility in enzyme evolution**. *Curr Opin Struct Biol* 2018, **48**:83–92.

72. Woldeyes RA, Sivak DA, Fraser JS: **E pluribus unum, no more: from one crystal, many conformations**. *Curr Opin Struct Biol* 2014, **28**:56–62.

73. Haliloglu T, Bahar I: **Adaptability of protein structures to enable functional interactions and evolutionary implications**. *Curr Opin Struct Biol* 2015, **35**:17–23.

*This excellent review summarizes recent studies that explain the role of conformational dynamics in protein interaction and allosteric regulations, particularly the relationship between evolutionary selection of protein structure and corresponding intrinsic dynamics.

74. Bhabha G, Ekiert DC, Jennewein M, Zmasek CM, Tuttle LM, Kroon G, Dyson HJ, Godzik A, Wilson IA, Wright PE: **Divergent evolution of protein conformational dynamics in dihydrofolate reductase**. *Nat Struct Mol Biol* 2013, **20**:1243.

75. Isaksen GV, \AAqvist J, Brandsdal BO: **Enzyme surface rigidity tunes the temperature dependence of catalytic rates**. *Proc Natl Acad Sci* 2016, **113**:7822–7827.

76. Röthlisberger D, Khersonsky O, Wollacott AM, Jiang L, DeChancie J, Betker J, Gallaher JL, Althoff EA, Zanghellini A, Dym O: **Kemp elimination catalysts by computational enzyme design**. *Nature* 2008, **453**:190.

77. Jiang L, Althoff EA, Clemente FR, Doyle L, Röthlisberger D, Zanghellini A, Gallaher JL, Betker JL, Tanaka F, Barbas CF: **De novo computational design of retro-aldol enzymes**. *science* 2008, **319**:1387–1391.

78. Siegel JB, Zanghellini A, Lovick HM, Kiss G, Lambert AR, Clair JLS, Gallaher JL, Hilvert D, Gelb MH, Stoddard BL: **Computational design of an enzyme catalyst for a stereoselective bimolecular Diels-Alder reaction**. *Science* 2010, **329**:309–313.

79. Kiss G, Çelebi-Ölçüm N, Moretti R, Baker D, Houk KN: **Computational enzyme design**. *Angew Chem Int Ed* 2013, **52**:5700–5725.

80. Korendovych IV, DeGrado WF: **Catalytic efficiency of designed catalytic proteins**. *Curr Opin Struct Biol* 2014, **27**:113–121.

81. Pauling L: **Molecular architecture and biological reactions**. *Chem Eng News* 1946, **24**:1375–1377.

82. Alexandrova AN, Röthlisberger D, Baker D, Jorgensen WL: **Catalytic mechanism and performance of computationally designed enzymes for Kemp elimination**. *J Am Chem Soc* 2008, **130**:15907–15915.

83. Kiss G, Röthlisberger D, Baker D, Houk KN: **Evaluation and ranking of enzyme designs**. *Protein Sci* 2010, **19**:1760–1773.

84. Ruscio JZ, Kohn JE, Ball KA, Head-Gordon T: **The influence of protein dynamics on the success of computational enzyme design**. *J Am Chem Soc* 2009, **131**:14111–14115.

85. James LC, Tawfik DS: **Conformational diversity and protein evolution–a 60-year-old hypothesis revisited**. *Trends Biochem Sci* 2003, **28**:361–368.

86. Klinman JP, Kohen A: **Evolutionary aspects of enzyme dynamics**. *J Biol Chem* 2014, **289**:30205–30212.

87. Petrović D, Frank D, Kamerlin SCL, Hoffmann K, Strodel B: **Shuffling Active Site Substate Populations Affects Catalytic Activity: The Case of Glucose Oxidase**. *ACS Catal* 2017, **7**:6188–6197.

88. Clifton BE, Jackson CJ: **Ancestral protein reconstruction yields insights into adaptive evolution of binding specificity in solute-binding proteins**. *Cell Chem Biol* 2016, **23**:236–245.

89. Glembo TJ, Farrell DW, Gerek ZN, Thorpe MF, Ozkan SB: **Collective dynamics differentiates functional divergence in protein evolution**. *PLoS Comput Biol* 2012, **8**:e1002428.

90. Bar-Even A, Noor E, Savir Y, Liebermeister W, Davidi D, Tawfik DS, Milo R: **The moderately efficient enzyme: evolutionary and physicochemical trends shaping enzyme parameters**. *Biochemistry (Mosc)* 2011, **50**:4402–4410.

91. Atilgan C, Gerek ZN, Ozkan SB, Atilgan AR: **Manipulation of conformational change in proteins by single-residue perturbations**. *Biophys J* 2010, **99**:933–943.

92. Butler BM, Gerek ZN, Kumar S, Ozkan SB: **Conformational dynamics of nonsynonymous variants at protein interfaces reveals disease association**. *Proteins Struct Funct Bioinforma* 2015, **83**:428–435.

93. Nevin Gerek Z, Kumar S, Banu Ozkan S: **Structural dynamics flexibility informs function and evolution at a proteome scale**. *Evol Appl* 2013, **6**:423–433.

94. Zou T, Risso VA, Gavira JA, Sanchez-Ruiz JM, Ozkan SB: **Evolution of conformational dynamics determines the conversion of a promiscuous generalist into a specialist enzyme**. *Mol Biol Evol* 2014, **32**:132–143.

* This work provides computational evidence that the conversion from ancestral generalists to modern specialists, likely to have occurred often during protein evolution, is linked to changes in conformational dynamics.

95. Ingles-Prieto A, Ibarra-Molero B, Delgado-Delgado A, Perez-Jimenez R, Fernandez JM, Gaucher EA, Sanchez-Ruiz JM, Gavira JA: **Conservation of protein structure over four billion years**. *Structure* 2013, **21**:1690–1697.

96. Modi T, Huihui J, Ghosh K, Ozkan B: **Ancient Thioredoxins Evolved to Modern Day Stability-Function Requirement by Altering Native State Ensemble**. *Biophys J* 2018, **114**:231a.

97. Kim H, Zou T, Modi C, Dörner K, Grunkemeyer TJ, Chen L, Fromme R, Matz MV, Ozkan SB, Wachter RM: **A hinge migration mechanism unlocks the evolution of green-to-red photoconversion in GFP-like proteins**. *Structure* 2015, **23**:34–43.

* This integrated experimental and computational work highlights changes of conformational dynamics through allostery in the evolution of GFP-like proteins from green to red color.

LEGENDS TO THE FIGURES

Figure 1. High stability of resurrected Precambrian thioredoxins. A) Schematic phylogenetic tree used for the reconstruction of thioredoxin ancestral sequences [31]. Only the bacterial branch is shown. LBCA and LPBCA stand, respectively, for the last common ancestor of bacteria and the last common ancestor of the cyanobacterial, *Deinococcus* and *Thermus* groups. B) 3D-structures of LBCA thioredoxin and LPBCA thioredoxin [95]. Mutational differences and experimental denaturation temperature values are shown. C) Mutational comparison between LBCA thioredoxin and LPBCA thioredoxin reveals three mutations that further stabilize the LPBCA protein [49]. The triple-mutant variant of LPBCA thioredoxin has a denaturation temperature about 40 degrees above that of the modern *E. coli* thioredoxin, as shown by experimental differential scanning calorimetry (DSC) profiles [49]. Note that overpressure is customarily applied in DSC experiments to prevent boiling above 100 ºC.

Figure 2. Evolution of conformational dynamics determines the conversion of a promiscuous generalist into a specialist enzyme. A) Schematic phylogetic tree used for the reconstruction of ancestral sequences of β-lactamases [10]. ENCA, GNCA and PNCA stand, respectively, for the last common ancestor of enterobacteria, the last common ancestor of various Gram-negative bacteria and the last common ancestor of various Gram-positive and Gram-negative bacteria. B) The “oldest” resurrected Precambrian β-lactamases can promiscuously catalyze the degradation of several lactam antibiotics [10], including benzylpenicillin (BZ) and the third generation antibiotics cefotaxime (CTX) and ceftazidime (CAZ). By contrast, the modern TEM-1 β-lactamase is a penicillin specialist. Catalytic efficiencies are shown on the distribution for modern

proteins [90]. Note that GNCA and PNCA β-lactamases are efficient promiscuous enzymes that degrade several antibiotics with catalytic efficiencies that compare well with a modern average enzyme. C) DFI profiles of extant (TEM-1) and ancestral β-lactamases [94] mapped on 3-D protein structures using a color coded scheme with a spectrum from red to blue. Lowest DFI regions are denoted with blue and flexible regions are red. The oldest and most promiscuous ancestors GNCA and PNCA exhibit higher flexibility near the active site. β-lactam specific TEM-1 shows less flexibility near the active site. B) A cladogram of SVD distances for β-lactamases determined from their DFI profiles, showing that dynamics based clustering captures the promiscuity of two ancestral enzymes and cluster them together [94].

Figure 3. *De novo* enzyme functionality in ancestral β-lactamase scaffolds linked to conformational flexibility [24]. A) NMR relaxation studies on the modern TEM-1 β-lactamase and the ancestral GNCA β-lactamase (see legend to figure 2 for definitions). Red color is used to highlight the residues with relaxation rates that suggest a conformational exchange contribution. The residue targeted for new active-site generation (W229) is highlighted in blue. B) A new active site capable of catalyzing Kemp elimination is generated in ancestral β-lactamases (but not in modern β-lactamases) by a single W229D mutation. Here, a blowup of the new active site generated in GNCA β-lactamase is shown with a transition state analogue bound. C) The 3D-structure of the W229D variant of GNCA β-lactamase with a transition-state analogue bound is shown superimposed with that of the GNCA β-lactamase background. It is apparent that transition-state binding (and, consequently, the generation of a *de novo* activity) relies on conformational re-arrangements, in particular, on the shift of the α-helices h1 and h11.