

# UC Berkeley

## UC Berkeley Previously Published Works

### Title

The thermostability and specificity of ancient proteins

### Permalink

<https://escholarship.org/uc/item/4b32078n>

### Authors

Wheeler, Lucas C

Lim, Shion A

Marqusee, Susan

et al.

### Publication Date

2016-06-01

### DOI

10.1016/j.sbi.2016.05.015

Peer reviewed



Published in final edited form as:

*Curr Opin Struct Biol.* 2016 June ; 38: 37–43. doi:10.1016/j.sbi.2016.05.015.

## The thermostability and specificity of ancient proteins

Lucas C Wheeler<sup>1,2,5</sup>, Shion A Lim<sup>3,4,5</sup>, Susan Marqusee<sup>3,4</sup>, and Michael J Harms<sup>1,2</sup>

<sup>1</sup>Department of Chemistry and Biochemistry, University of Oregon, Eugene, OR, United States

<sup>2</sup>Institute of Molecular Biology, University of Oregon, Eugene, OR, United States

<sup>3</sup>Department of Molecular and Cell Biology, University of California, Berkeley, Berkeley, CA, United States

<sup>4</sup>Institute for Quantitative Biosciences (QB3), University of California, Berkeley, Berkeley, CA, United States

### Abstract

Were ancient proteins systematically different than modern proteins? The answer to this question is profoundly important, shaping how we understand the origins of protein biochemical, biophysical, and functional properties. Ancestral sequence reconstruction (ASR), a phylogenetic approach to infer the sequences of ancestral proteins, may reveal such trends. We discuss two proposed trends: a transition from higher to lower thermostability and a tendency for proteins to acquire higher specificity over time. We review the evidence for elevated ancestral thermostability and discuss its possible origins in a changing environmental temperature and/or reconstruction bias. We also conclude that there is, as yet, insufficient data to support a trend from promiscuity to specificity. Finally, we propose future work to understand these proposed evolutionary trends.

### Introduction

Ancestral sequence reconstruction (ASR) has opened a window into the sequences and properties of ancient proteins [1,2]. In ASR, a multiple sequence alignment of modern protein sequences is used to construct a phylogenetic tree and the sequences of ancient proteins are inferred for specific ancestors on this tree (Figure 1a). By synthesizing the genes encoding these sequences, these reconstructed ancient proteins can be experimentally characterized. This approach has yielded an explosion of results in recent years, revealing important mechanistic insights into the evolution of protein forms and functions [3–11,12••].

One intriguing possibility is to use ASR to investigate whether ancient proteins were systematically different in the past, leading to parallel, directional changes in properties over evolutionary time (Figure 1a). Such trends are inaccessible using comparisons between modern proteins. For example, studies of the modern proteins in Figure 1b would lead one to believe the last common ancestor had a ‘blue’ trait. By allowing direct measurement of

Corresponding authors: Marqusee, Susan (marqusee@berkeley.edu) and Harms, Michael J (harms@uoregon.edu).

<sup>5</sup>LCW and SAL contributed equally to this work.

#### Conflict of interest

Nothing declared.

ancestral properties, ASR can reveal properties ('red', in this case) not evident in the modern proteins.

If the evolution of protein properties were directional, it would provide a new level at which to explain and understand these properties. This is of deep interest to evolutionary biochemists seeking to identify the general principles that shape protein evolution. Further, a trend could mean that sampling evolutionary history would provide access to qualitatively different proteins [13••] — a boon to engineers looking for proteins with desirable properties as templates for further engineering [14,15].

Recent work has suggested two trends over evolutionary time: decreasing protein stability [3] and increasing specificity [11]. Particularly for protein engineers, these trends could be extremely powerful, as high stability and broad substrate specificity are desirable traits that could be accessed using ASR. In this review, we review the evidence supporting and contradicting these trends, as well as the future work required to test and extend these conclusions.

## Reconstructed Precambrian ancient proteins exhibit elevated thermostability

We begin by evaluating evidence from ASR studies that indicate the deepest ancestors of mesophilic proteins were highly thermostable. Over billion-year timescales, reconstructed ancestral proteins display systematically higher thermostability. Reconstructed EF-Tu [3], thioredoxin [10], DNA gyrase [8], nucleotide diphosphate kinase [7], and  $\beta$ -lactamase [11] all exhibit melting temperatures ( $T_m$ ) far higher than their extant descendants. Some have argued that this is a universal trend [13••] and have interpreted this as evidence for an ancient, hot environment [3]. The evidence, however, is not completely universal, as reconstructed RNase H along a mesophilic lineage gives a relatively flat trend in stability over similar time scales [12••].

One difficulty in comparing these studies is that different proteins have different absolute requirements for stability. For example, the  $T_m$ s of EF-Tu bacterial homologs are generally  $\sim 2$  °C above the environmental temperature ( $T_{env}$ ), while the  $T_m$ s of RNase H are  $\sim 30$  °C above  $T_{env}$ . As a result,  $T_m$ s between protein families are not directly comparable. One way to overcome this challenge is to convert the measured  $T_m$  of each protein to an estimate of  $T_{env}$ , as  $T_m$  often correlates with the growth temperature of the organism from which it was derived [16]. In most cases, this correlation arises to maintain stability above some critical threshold [17]. Empirically,  $T_m$  generally rises by  $\sim 1$  °C per 1 °C of  $T_{env}$ , with an offset reflecting the required stability of the protein (e.g. 2 °C for EF-Tu and 30 °C for RNase H) [16]. This correlation has been directly established for three of the proteins above — EF-Tu, DNA gyrase and RNase H [3,7,12••] — and holds generally for many other proteins [16].

When placed on the  $T_{env}$  scale, reconstructed proteins report an elevated environmental temperature  $\sim 3$  billion years ago, though with significant scatter. Figure 2a shows the estimated  $T_{env}$  over time for 17 ancestors of proteins found in the lineages leading to mesophilic *E. coli*. A total least-squares fit to the data reveals a highly significant negative

slope that explains 75% of the variation in the data ( $R^2 = 0.75$ ). In contrast, the estimated  $T_{\text{env}}$  over time for ancestors leading to thermophile *T. thermophilus* exhibits a slope statistically indistinguishable from 0 (Figure 2b). When taken in aggregate, these data support the hypothesis that the deepest ancestors had stabilities similar to proteins from modern thermophiles. While these data focus on the *E. coli* and *T. thermophilus* lineages, their deepest ancestors are shared both with each other and with most modern bacteria, thus suggesting a global transition away from ancient thermostability, at least along mesophilic lineages. It is not clear from these sparse, lineage-specific data whether mesophilicity evolved in parallel along many lineages or whether it evolved on a few key, early ancestral branches.

## Trends in thermostability are complex

While ASR studies suggest that the most ancient proteins were highly thermostable, they do not support a smooth trend in thermostability over time. Ancestors exhibit extensive random scatter to the proposed trend. Such variation is expected as, over more recent timescales, protein stability fluctuates in response to neutral drift or adaptation in apparently random fashion [6,9,18–20]. The observed variation may also reflect uncertainty in the reconstruction, multiple heterogeneous environments experienced by ancient organisms, or uncertainty in the map between  $T_{\text{m}}$  and  $T_{\text{env}}$ .

This scatter extends to the mechanism of stabilization. A recent study of the evolution of thermostability in RNase H revealed that the thermodynamic mechanism of stabilization for the ancestral proteins could fluctuate, even as the  $T_{\text{ms}}$  of the proteins varied smoothly [12••]. This indicates that, even while under selection to maintain stability in a given environment, proteins are free to accumulate mutations to access alternate mechanisms of stabilization. Practically, studying multiple ancestors may reveal new sequence and thermodynamic determinants of stability. Although thermostability and the mechanism of stabilization appear to change independently for RNase H, the generality of this result for other proteins remains unknown.

Finally, these ASR studies generally used small, monomeric, and well-behaved proteins. Although such simple proteins may be representative of the first proteins to arise, studies on a greater diversity of protein families will reveal whether observed trends are applicable to the entire proteome.

## Can reconstruction errors lead to inflated ancestral thermostability?

While existing data are suggestive, further work must be done to test the hypothesis of ancient thermostability. The primary concern is that ancestral proteins are statistical reconstructions that cannot be directly verified. Even with good statistical support, it is unlikely that the reconstructed ancestor will have the exact sequence of the true ancestral state. Addressing and understanding this uncertainty will be critical for establishing or refuting the hypothesis that the earliest proteins were thermostable.

High stability is unlikely to arise from random errors in the reconstruction. To account for uncertainty, ASR studies have generated different versions of ancestral sequences to assess

the robustness of the measured stability to phylogenetic errors. For example, Hart et al. measured ten alternate sequences of a ~3 billion year-old ancestor and found a  $T_m$  of  $76.7 \pm 2$  °C (compared to 68.0 °C of RNase H from *E. coli*) [12••]. Using such approaches, many sources of random error have been investigated: uncertain tree topology [3,7,21,22], alternate evolutionary models [23], choice of reconstruction method [6,22], different amino acid frequencies [3], and reconstruction ambiguity [3,7,11,12••,24]. In all such studies, the properties of the ancestors have proven robust to uncertainty.

Of bigger concern are sources of systematic error in ASR — in particular, a bias towards elevated stability for deeper ancestors [25–28]. Some have argued that ASR could be biased towards consensus sequences, which may lead to an increase in stability [26,29,30]. Simulations have also suggested that maximum likelihood (ML), the most popular form of ASR, may give rise to artificially elevated stability [25]. If different stabilizing mutations accumulate along different lineages, ML may incorrectly incorporate all of the stabilizing mutations, creating an artificially stable ancestor. There is also concern that variable amino acid distributions and mutation rates can alter reconstructions [27,28].

There have been some limited experimental tests of these computational predictions of bias. Comparisons between ancestral and consensus sequences have shown distinct statistical and functional properties [7,8,14,31••]. This suggests that any consensus bias that exists must be subtle. Other work has indirectly addressed this concern - the molecular basis of stability fluctuating over evolutionary time in the RNase H family is not consistent with bias arising from a single, convergent stabilization mechanism [12••,25].

Important experiments remain. One test would be a systematic comparison of ancestors reconstructed using both ML and an alternative, Bayesian, method. A Bayesian reconstruction averages over uncertainty; therefore, it is not expected to have the same stability bias as ML reconstructions [25]. Observing high thermostability in ancient Bayesian ancestors would be strong evidence that thermostability is not an artifact of the ML method. The experiment is not perfect, however, as Bayesian ancestors have more errors than ML ancestors as a result of incorporating uncertainty [22]. Because of this, they may not accurately reflect the ancestral state. For example, one study found that a Bayesian ancestor had fundamentally different folding properties than the ML ancestor or any modern protein in the family [6], consistent with a poor reconstruction.

Another test for bias would be to study the thermostability of reconstructed, recent ancestors of rapidly evolving proteins with known mesophilic ancestral environments. A rapidly evolving protein will accumulate similar amounts of mutations relative to the deep ancestors studied to date, albeit on a much shorter timescale. If ML reconstructions lead to biased stability, we would predict that recent ancestors of rapidly evolving proteins would exhibit erroneously elevated stability.

## **A trend from promiscuous to specific proteins is not yet established**

Another proposed trend is that proteins have, on average, changed from lower to higher specificity over deep evolutionary time [11,13••]. This stems from the idea that low

specificity proteins — particularly enzymes — were important for the ability of primordial organisms to perform diverse chemical processes with a limited proteome [32] (Figure 3a). It is also well established that increased specificity often follows gene duplication via subfunctionalization from a multi-functional or promiscuous ancestral protein [33,34] (Figure 3b). Given these considerations, proteins may, on average, increase in specificity over time.

To date, few attempts have been made to investigate the specificity of the deepest ancestors. One recent study found that an ancestral  $\beta$ -lactamase was both promiscuous and less efficient than its descendants [11]. Likewise, a study of RuBISCO found a promiscuous and inefficient ancestor, though this may be an artifact of poor reconstruction [35]. Other studies have determined the activities of ancient proteins, but not their specificity [6,7]. On the basis of these data, it is difficult to make solid conclusions about specificity trends; more measurements of ancestral specificity are warranted.

The second model — gene duplication followed by subfunctionalization — could conceivably operate continuously through evolution, leading to progressively higher specificity proteins over all evolutionary timescales (Figure 3b). Studies of the evolution of specificity for ancestors from the last ~500 million years suggest, however, that on average, proteins do not tend towards higher specificity over time. Some promiscuity-to-specificity transitions have been identified [11,36,37•,38,39]. However, other studies have found switches between two high-specificity states [40••,41•], evolution through a less-specific intermediate [42,43•,44•], and even decreased specificity over time [45].

This complexity likely arises because specificity is, at minimum, a bimolecular process that involves both the protein and its target. Further, constraints placed by the architecture of the larger system into which the proteins are embedded have been shown to shape specificity [44•,46–52]. For example, bioinformatic analyses have revealed that protein components of higher-complexity regulatory modules tend to possess lower specificity than those in simpler modules [53]. We therefore believe that it will be difficult to resolve a global evolutionary trend from lower to higher specificity.

## Conclusions

A number of ASR studies are starting to reveal a consistent pattern of elevated thermostability for the deepest ancestors. This trend of decreasing thermostability among mesophilic lineages is not smooth, involving fluctuations in both  $T_m$  and mechanism of stabilization. Whether this reflects a real evolutionary signal or simply an artifact of the reconstruction method remains to be seen. From an engineering perspective, a ML reconstruction of an ancient ancestor appears to be a reasonable strategy for generating a thermostable, thermophilic-like protein that differs from a simple consensus sequence. This approach is not guaranteed — for example, reconstructed RNase H displays non-thermophilic-like thermostability ~3 billion years in the past — however, on average, deep ancestral proteins appear to be more stable than their modern counterparts. We should also note that these are deep trends, and thus we would not predict recent ancestors to exhibit any detectable trend in stability, consistent with recent studies [6,9,19].

Information about the specificity of deep ancestral proteins remains sparse and will thus require further investigation. Studies of more recent proteins indicate that multiple modes of specificity evolution can be at play, suggesting a lack of general trends.

Protein evolution is often viewed as a random, microscopically-reversible trajectory along a fitness landscape. A global trend would suggest that the fitness landscape changed in a systematic way, even while microscopic reversibility held. Such systematic changes in fitness landscape would, in turn, shape the pathways taken by proteins and provide another level at which to understand the emergence of new properties. ASR studies are hinting at a change in fitness landscape. This may help us, at a broad brush level, gain insight into the origins of protein features and properties.

## Acknowledgments

The authors gratefully acknowledge funding sources: NIH grant GM050945 (SM), 7T32GM007759-37 (LCW). MJH is a Pew Scholar in the Biomedical Sciences, supported by The Pew Charitable Trusts. SAL is supported by the National Science Foundation Graduate Research Fellowship.

## References and recommended reading

Papers of particular interest, published within the period of review, have been highlighted as:

- of special interest
  - of outstanding interest
1. Pauling L, Zuckerkandl E. Chemical paleogenetics. Molecular “restoration studies” of extinct forms of life. *Acta Chem Scand.* 1963; 17:S9–S16.
  2. Harms MJ, Thornton JW. Analyzing protein structure and function using ancestral gene reconstruction. *Curr Opin Struct Biol.* 2010; 20:360–366. [PubMed: 20413295]
  3. Gaucher EA, Govindarajan S, Ganesh OK. Palaeotemperature trend for Precambrian life inferred from resurrected proteins. *Nature.* 2008; 451:704–707. [PubMed: 18256669]
  4. Voordeckers K, Brown CA, Vanneste K, van der Zande E, Voet A, Maere S, Verstrepen KJ. Reconstruction of ancestral metabolic enzymes reveals molecular mechanisms underlying evolutionary innovation through gene duplication. *PLoS Biol.* 2012; 10:e1001446. [PubMed: 23239941]
  5. Bridgham JT, Ortlund EA, Thornton JW. An epistatic ratchet constrains the direction of glucocorticoid receptor evolution. *Nature.* 2009; 461:515–519. [PubMed: 19779450]
  6. Hobbs JK, Shepherd C, Saul DJ, Demetras NJ, Haaning S, Monk CR, Daniel RM, Arcus VL. On the origin and evolution of thermophily: reconstruction of functional precambrian enzymes from ancestors of *Bacillus*. *Mol Biol Evol.* 2012; 29:825–835. [PubMed: 21998276]
  7. 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. [PubMed: 23776221]
  8. Akanuma S, Iwami S, Yokoi T, Nakamura N, Watanabe H, Yokobori S, Yamagishi A. Phylogeny-based design of a B-subunit of DNA gyrase and its ATPase domain using a small set of homologous amino acid sequences. *J Mol Biol.* 2011; 412:212–225. [PubMed: 21819994]
  9. Loughran NB, O’Connell MJ, O’Connor B, O’Fágáin C. Stability properties of an ancient plant peroxidase. *Biochimie.* 2014; 104:156–159. [PubMed: 24919139]
  10. 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, et al. Single-molecule paleoenzymology probes the chemistry of resurrected enzymes. *Nat Struct Mol Biol.* 2011; 18:592–596. [PubMed: 21460845]

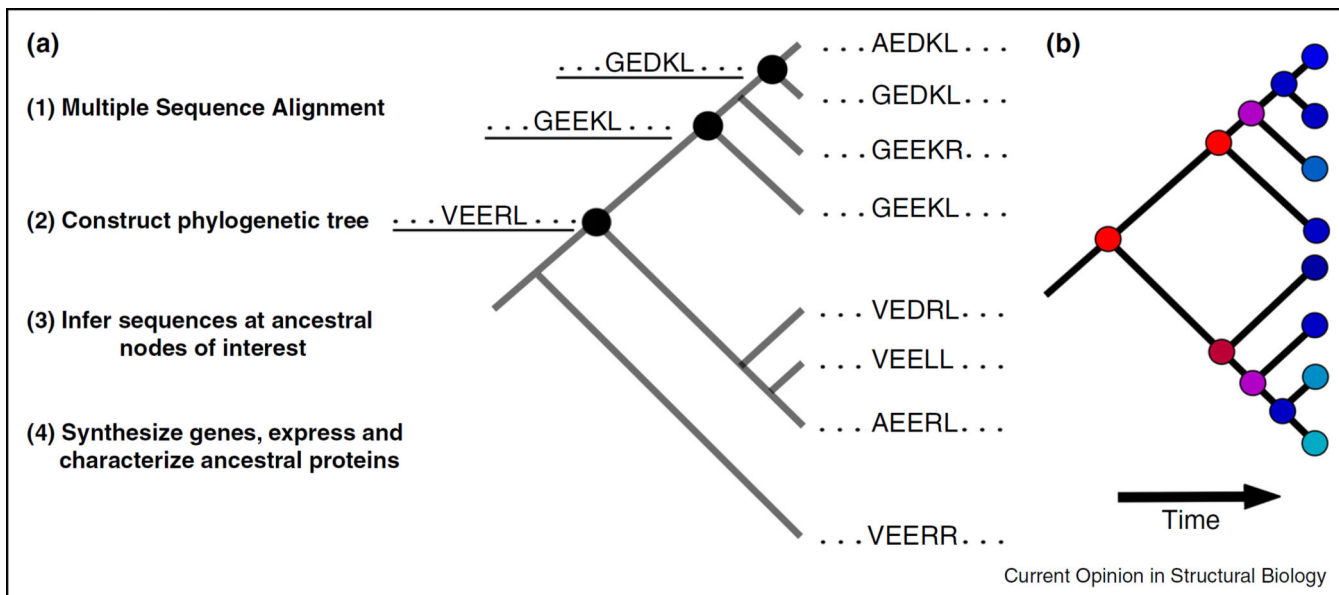
11. Risso VA, Gavira JA, Mejia-Carmona DF, Gaucher EA, Sanchez-Ruiz JM. Hyperstability and substrate promiscuity in laboratory resurrections of precambrian  $\beta$ -lactamases. *J Am Chem Soc.* 2013; 135:2899–2902. [PubMed: 23394108]
12. Hart KM, Harms MJ, Schmidt BH, Elya C, Thornton JW, Marqusee S. Thermodynamic system drift in protein evolution. *PLoS Biol.* 2014; 12:e1001994. [PubMed: 25386647] This study conducts a detailed thermodynamic characterization of ancestors of the ribonuclease H family, demonstrating that despite a clear evolutionary selection for thermostability, the underlying thermodynamic mechanism for thermostabilization was variable across the lineages.
13. Risso VA, Gavira JA, Sanchez-Ruiz JM. Thermostable and promiscuous Precambrian proteins. *Environ. Microbiol.* 2014; 16:1485–1489. [PubMed: 25009840] This article provides a summary and review of recent ASR studies that have informed our understanding of trends in protein thermostability and promiscuity.
14. Cole MF, Gaucher EA. Utilizing natural diversity to evolve protein function: applications towards thermostability. *Curr Opin Chem Biol.* 2011; 15:399–406. [PubMed: 21470898]
15. 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. [PubMed: 26061224]
16. Gromiha MM, Oobatake M, Sarai A. Important amino acid properties for enhanced thermostability from mesophilic to thermophilic proteins. *Biophys Chem.* 1999; 82:51–67. [PubMed: 10584295]
17. Taverna DM, Goldstein RA. Why are proteins marginally stable? *Proteins Struct Funct Genet.* 2002; 46:105–109. [PubMed: 11746707]
18. Malcolm BA, Wilson KP, Matthews BW, Kirsch JF, Wilson AC. Ancestral lysozymes reconstructed, neutrality tested, and thermostability linked to hydrocarbon packing. *Nature.* 1990; 345:86–89. [PubMed: 2330057]
19. Dasmeh P, Serohijos AWR, Kepp KP, Shakhnovich EI. Positively selected sites in cetacean myoglobins contribute to protein stability. *PLoS Comput Biol.* 2013; 9:e1002929. [PubMed: 23505347]
20. Gong LI, Suchard MA, Bloom JD. Stability-mediated epistasis constrains the evolution of an influenza protein. *Elife.* 2013; 2:e00631. [PubMed: 23682315]
21. Groussin M, Hobbs JK, Szoll si GJ, Gribaldo S, Arcus VL, Gouy M. Toward more accurate ancestral protein genotype–phenotype reconstructions with the use of species tree-aware gene trees. *Mol Biol Evol.* 2015; 32:13–22. [PubMed: 25371435]
22. Hanson-Smith V, Kolaczowski B, Thornton JW. Robustness of ancestral sequence reconstruction to phylogenetic uncertainty. *Mol Biol Evol.* 2010; 27:1988–1999. [PubMed: 20368266]
23. Akanuma S, Yokobori S, Nakajima Y, Bessho M, Yamagishi A. Robustness of predictions of extremely thermally stable proteins in ancient organisms. *Evolution.* 2015; 69:2954–2962. [PubMed: 26404857]
24. Bar-Rogovsky H, Stern A, Penn O, Kobl I, Pupko T, Tawfik DS. Assessing the prediction fidelity of ancestral reconstruction by a library approach. *Protein Eng Des Sel.* 2015; 28:507–518. [PubMed: 26275856]
25. Williams PD, Pollock DD, Blackburne BP, Goldstein RA. Assessing the accuracy of ancestral protein reconstruction methods. *PLoS Comput Biol.* 2006; 2:e69. [PubMed: 16789817]
26. Bershtein S, Goldin K, Tawfik DS. Intense neutral drifts yield robust and evolvable consensus proteins. *J Mol Biol.* 2008; 379:1029–1044. [PubMed: 18495157]
27. Pollock DD, Thiltgen G, Goldstein RA. Amino acid coevolution induces an evolutionary Stokes shift. *Proc Natl Acad Sci.* 2012; 109:E1352–E1359. [PubMed: 22547823]
28. Goldstein RA, Pollard ST, Shah SD, Pollock DD. Nonadaptive amino acid convergence rates decrease over time. *Mol Biol Evol.* 2015; 32:1373–1381. [PubMed: 25737491]
29. Gaschen B, Taylor J, Yusim K, Foley B, Gao F, Lang D, Novitsky V, Haynes B, Hahn BH, Bhattacharya T, et al. Diversity considerations in HIV-1 vaccine selection. *Science.* 2002; 296:2354–2360. [PubMed: 12089434]
30. Kothe DL, Li Y, Decker JM, Bibollet-Ruche F, Zammit KP, Salazar MG, Chen Y, Weng Z, Weaver EA, Gao F, et al. Ancestral and consensus envelope immunogens for HIV-1 subtype C. *Virology.* 2006; 352:438–449. [PubMed: 16780913]



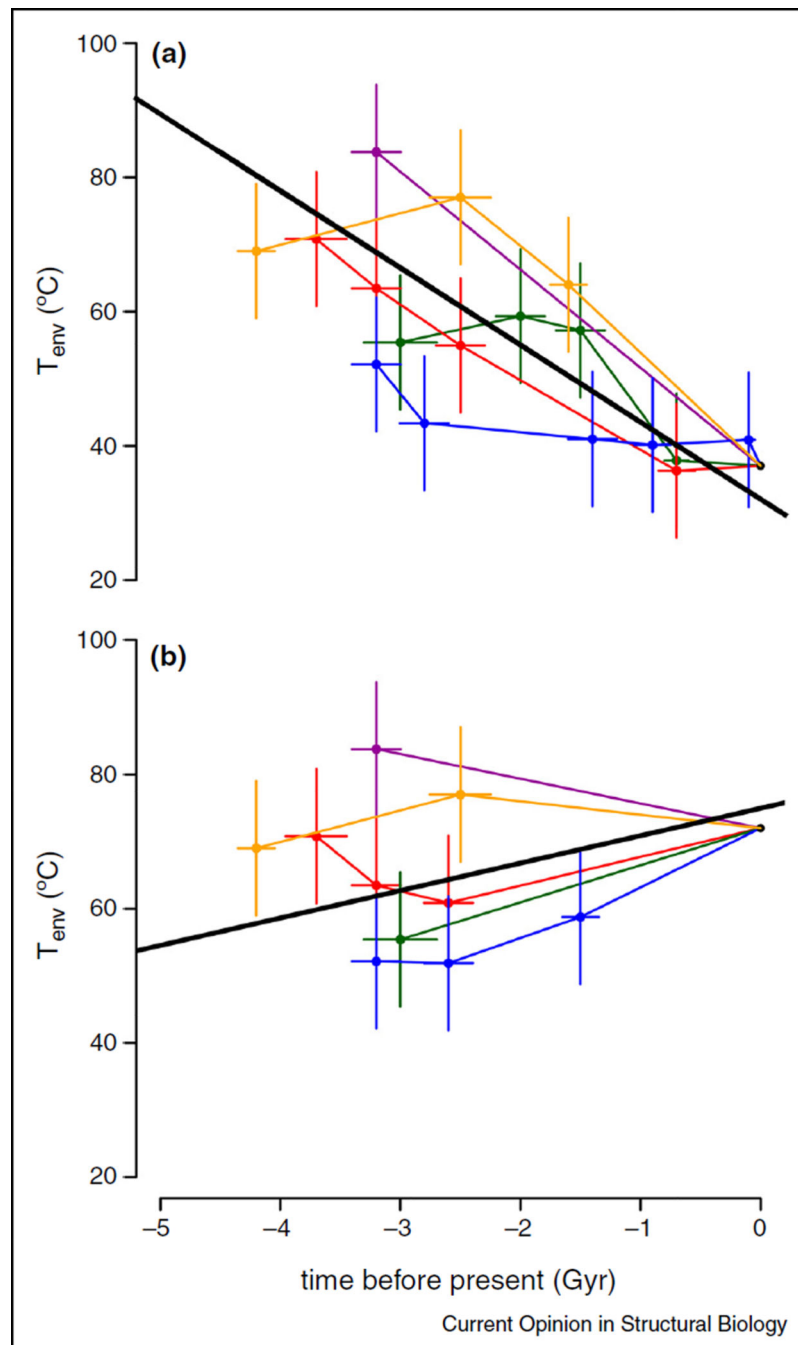
31. Risso VA, Gavira JA, Gaucher EA, Sanchez-Ruiz JM. Phenotypic comparisons of consensus variants versus laboratory resurrections of Precambrian proteins. *Proteins Struct Funct Bioinform.* 2014; 82:887–896. This study compared the properties of ancestral and consensus  $\beta$ -lactamases and found phenotypic differences in stability and activity of proteins generated by the two methods. This work demonstrates that ancestral and consensus sequences are distinct approaches and questions the interpretation that consensus sequences are a proxy of the ancestral state.
32. Jensen RA. Enzyme recruitment in evolution of new function. *Annu Rev Microbiol.* 1976; 30:409–425. [PubMed: 791073]
33. Lynch M, Force A. The probability of duplicate gene preservation by subfunctionalization. *Genetics.* 2000; 154:459–473. [PubMed: 10629003]
34. Conant GC, Wolfe KH. Turning a hobby into a job: how duplicated genes find new functions. *Nat Rev Genet.* 2008; 9:938–950. [PubMed: 19015656]
35. Ma S, Martin-Laffon J, Mininno M, Gigarel O, Brugiere S, Bastien O, Tardif M, Ravel S, Alban C. Molecular evolution of the substrate specificity of chloroplastic aldolases/rubisco lysine methyltransferases in plants. *Mol. Plant.* 2016; 9:569–581. [PubMed: 26785049]
36. Eick GN, Colucci JK, Harms MJ, Ortlund EA, Thornton JW. Evolution of minimal specificity and promiscuity in steroid hormone receptors. *PLoS Genet.* 2012; 8:e1003072. [PubMed: 23166518]
37. Boucher JJ, Jacobowitz JR, Beckett BC, Classen S, Theobald DL. An atomic-resolution view of neofunctionalization in the evolution of apicomplexan lactate dehydrogenases. *Elife.* 2014; 3:1–25. Using ASR, the authors determine the evolutionary history and structural basis of convergent evolution of an LDH from an MDH ancestor, demonstrating that neofunctionalization of specificity occurred via an insertion that shifted the position of the canonical specificity-determining residue. The study provides a detailed atomic-resolution understanding of a switch in specificity.
38. Wouters MA, Liu K, Riek P, Husain A. A despecialization step underlying evolution of a family of serine proteases. *Mol Cell.* 2003; 12:343–354. [PubMed: 14536074]
39. Wilson C, Agafonov RV, Hoemberger M, Kutter S, Zorba A, Halpin J, Buosi V, Otten R, Waterman D, Theobald DL, et al. Using ancient protein kinases to unravel a modern cancer drug's mechanism. *Science.* 2015; 347:882–886. [PubMed: 25700521]
40. McKeown AN, Bridgman JT, Anderson DW, Murphy MN, Ortlund EA, Thornton JW. Evolution of DNA specificity in a transcription factor family produced a new gene regulatory module. *Cell.* 2014; 159:58–68. [PubMed: 25259920] An excellent, well-worked out example of an evolutionary switch between two high-specificity binding modes. Using ASR, the authors dissect the evolution of DNA-binding specificity of a steroid hormone receptor that gave rise to a new regulatory module. They highlight the importance of permissive substitutions that altered the cooperativity of binding, in addition to the modulation of negative binding determinants
41. Clifton BE, Jackson CJ. Ancestral protein reconstruction yields insights into adaptive evolution of binding specificity in solute-binding proteins. *Cell Chem Biol.* 2016; 23:1–10. [PubMed: 26933730] Using ASR, X-ray crystallography, and isothermal titration calorimetry the authors determine in detail the biophysical mechanism by which glutamine-binding proteins evolved from ancestral arginine binding proteins. The switch occurred from a promiscuous ancestor via conformational selection for an alternative low-energy binding conformation. Notably, the authors find no evidence in this system for subfunctionalization from an ancestral generalist.
42. Sayou C, Monniaux M, Nanao MH, Moyroud E, Brockington SF, Thévenon E, Chahtane H, Warthmann N, Melkonian M, Zhang Y, et al. A promiscuous intermediate underlies the evolution of LEAFY DNA binding specificity. *Science.* 2014; 343:645–648. [PubMed: 24436181]
43. Aakre CD, Herrou J, Phung TN, Perchuk BS, Crosson S, Laub MT. Evolving new protein–protein interaction specificity through promiscuous intermediates. *Cell.* 2015; 163:594–606. [PubMed: 26478181] The authors probe protein coevolution in the ParD–ParE bacterial toxin–antitoxin system. Using a high-throughput screen of binding interface mutants, they demonstrate the utility and abundance of promiscuous intermediates for the coevolution of specificity.
44. Howard CJ, Hanson-Smith V, Kennedy KJ, Miller CJ, Lou HJ, Johnson AD, Turk BE, Holt LJ. Ancestral resurrection reveals evolutionary mechanisms of kinase plasticity. *Elife.* 2014; 3:e04126. The authors study the evolutionary history of specificity in the CMGC protein kinases. Using

ASR, they show that specificity switched along the Ime2 branch via a bifunctional intermediate from a common ancestor that possessed a preference for +1 proline/arginine.

45. Chinen A, Naito Y, Handa N, Kobayashi I. Evolution of sequence recognition by restriction-modification enzymes: selective pressure for specificity decrease. *Mol Biol Evol.* 2000; 17:1610–1619. [PubMed: 11070049]
46. Peleg O, Choi J-M, Shakhnovich EI. Evolution of specificity in protein–protein interactions. *Biophys J.* 2014; 107:1686–1696. [PubMed: 25296322]
47. Kim J, Copley SD. Inhibitory cross-talk upon introduction of a new metabolic pathway into an existing metabolic network. *Proc Natl Acad Sci.* 2012; 109:E2856–E2864. [PubMed: 22984162]
48. Hong J, Gresham D. Molecular specificity convergence and constraint shape adaptive evolution in nutrient-poor environments. *PLoS Genet.* 2014; 10:e1004041. [PubMed: 24415948]
49. de Vos MGJ, Dawid A, Sunderlikova V, Tans SJ. Breaking evolutionary constraint with a tradeoff ratchet. *Proc Natl Acad Sci.* 2015; 112:14906–14911. [PubMed: 26567153]
50. Ernst A, Gfeller D, Kan Z, Seshagiri S, Kim PM, Bader GD, Sidhu SS. Coevolution of PDZ domain–ligand interactions analyzed by high-throughput phage display and deep sequencing. *Mol Biosyst.* 2010; 6:1782–1790. [PubMed: 20714644]
51. Lukatsky DB, Afek A, Shakhnovich EI. Sequence correlations shape protein promiscuity. *J Chem Phys.* 2011; 135:65104.
52. Stiffler MA, Chen JR, Grantcharova VP, Lei Y, Fuchs D, Allen JE, Zaslavskaja LA, MacBeath G. PDZ domain binding selectivity is optimized across the mouse proteome. *Science.* 2007; 317:364–369. [PubMed: 17641200]
53. Stewart AJ, Plotkin JB. The evolution of complex gene regulation by low-specificity binding sites. *Proc R Soc B.* 2013; 280:20131313.
54. Battistuzzi FU, Feijao A, Hedges SB. A genomic timescale of prokaryote evolution: insights into the origin of methanogenesis, phototrophy, and the colonization of land. *BMC Evol Biol.* 2004; 4:44. [PubMed: 15535883]

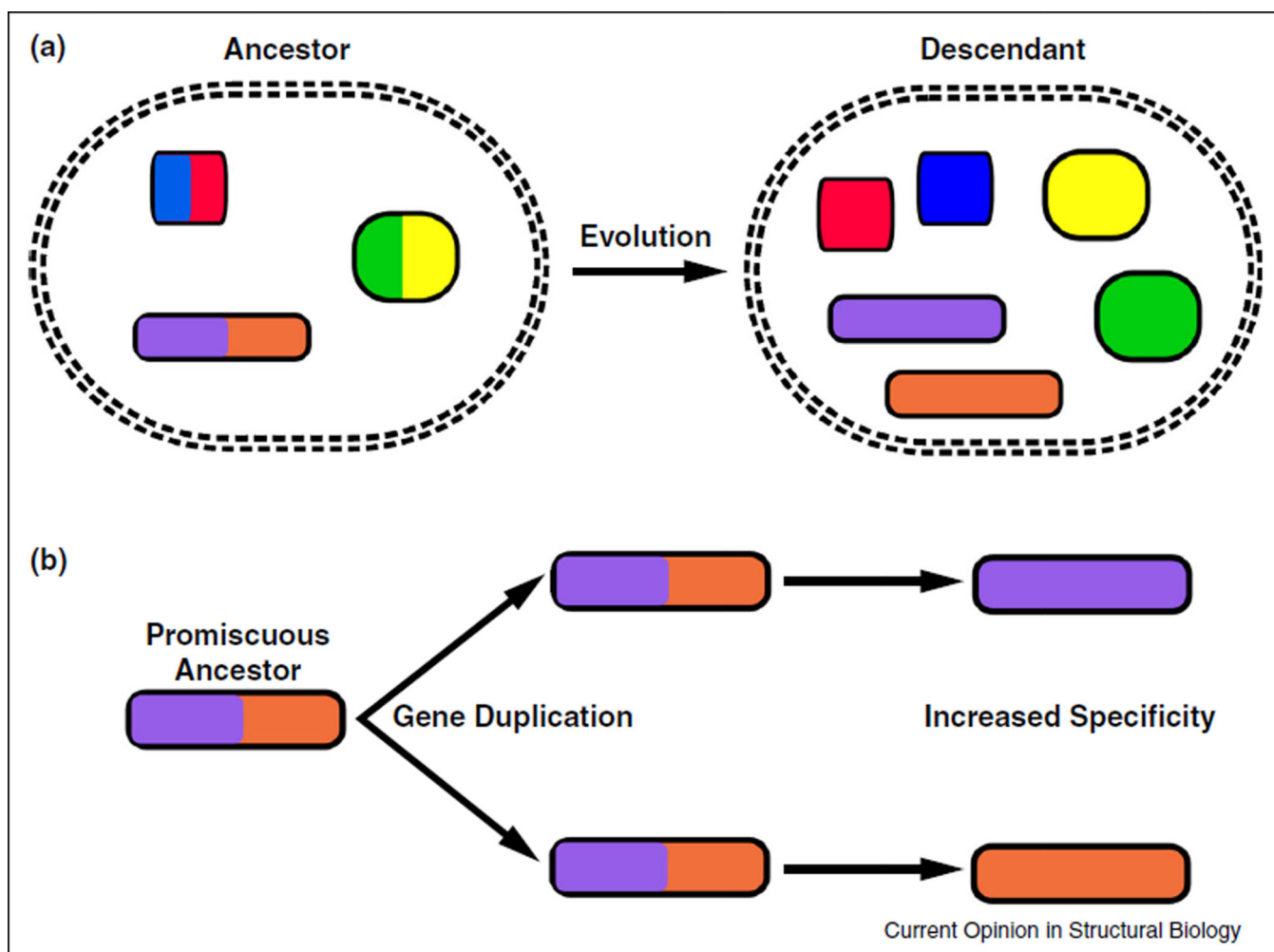
**Figure 1.**

Ancestral Sequence Reconstruction (ASR) can be used to trace the history of evolving proteins. **(a)** The ASR pipeline. A multiple sequence alignment (MSA) of extant sequences of a protein family is generated using an alignment tool. The MSA is then used to estimate an appropriate model of sequence evolution and to estimate a phylogenetic tree. The sequences at ancestral nodes of interest (filled black circles) are then inferred (underlined) based on the tree and a phylogenetic evolutionary model. The maximum likelihood sequences are those with the highest likelihood of generating the known sequences of modern proteins given the tree and phylogenetic model. Genes encoding the inferred ancestral proteins can be synthesized, expressed, and purified using standard molecular biology tools. The properties of the ancestral proteins can then be experimentally characterized. **(b)** A phylogenetic tree showing the evolution of a protein that can vary between two properties — red and blue. The last common ancestor was red, but the modern proteins are blue because of parallel changes along the lineages. This red ancestor can only be accessed using an approach like ASR.



**Figure 2.** Ancient reconstructed ancestors exhibit elevated thermostability. Estimated environmental temperatures experienced by proteins on lineages leading to (panel **a**) *E. coli* or (panel **b**) *T. thermophilus*. Point/line series indicate individual protein families: EF-Tu (red), thioredoxin (orange),  $\beta$ -lactamase (green), RNase H (blue), and nucleotide diphosphate kinase (purple). Measured melting temperatures for ancestors that give rise to *E. coli* proteins were mined from published literature [3,7,10,11,12••]. These were then converted to estimates of  $T_{env}$  using measured relationships [3,7,12••] or by adding an offset determined by the difference

in  $T_m$  and  $T_{env}$  for the *E. coli* (panel a) or *T. thermophilus* homologs (panel B). Time estimates were drawn from original publications or estimated from Battistuzzi et al. [54]. Time errors are standard errors.  $T_{env}$  standard errors were set to  $\pm 10$  °C to account for uncertainty in  $T_m$  and the  $T_m/T_{env}$  correlation. (This is a conservative estimate: when measured for NDK, RNase H, and EF-Tu [3,7,12••], the  $T_m$  the standard error was  $<5$  °C and the  $T_m$  to  $T_{env}$  variance was  $<5$  °C.) Black line is a fit determined by total linear regression. To find the standard deviation of fit slopes, we generated 1000 pseudo datasets sampled from the time and  $T_{env}$  uncertainties. For *E. coli*, the fits reject a slope = 0 ( $p = 3 \times 10^{-8}$ ). For *T. thermophilus*, the fits fail to reject a zero slope ( $p = 0.45$ ).



**Figure 3.** Models for increased specificity of proteins over time. **(a)** Large dotted ellipses denote cells. Small ellipses are proteins, colored by their specificity. Because early proteomes were presumably smaller than modern proteomes, it has been proposed that ancient proteins had to be promiscuous to achieve all the necessary chemistry. As organisms evolved, their proteomes expanded, allowing each protein to become more specific. **(b)** Higher specificity (subfunctionalization) is one of the possible outcomes of a gene duplication event. A gene encoding a low-specificity ancestral protein duplicates. Its descendants can then gain specificity and lose the promiscuous trait.