EXTENDED PDF FORMAT
SPONSORED BY
www.usbweb.com

# Protein Dynamism and Evolvability

Nobuhiko Tokuriki, *et al.*

*Science* **324**, 203 (2009);
DOI: 10.1126/science.1169375

---

***The following resources related to this article are available online at www.sciencemag.org (this information is current as of April 10, 2009 ):***

**Updated information and services,** including high-resolution figures, can be found in the online version of this article at:
http://www.sciencemag.org/cgi/content/full/324/5924/203

This article **cites 47 articles**, 13 of which can be accessed for free:
http://www.sciencemag.org/cgi/content/full/324/5924/203#otherarticles

This article appears in the following **subject collections**:
Biochemistry
http://www.sciencemag.org/cgi/collection/biochem

Information about obtaining **reprints** of this article or about obtaining **permission to reproduce this article** in whole or in part can be found at:
http://www.sciencemag.org/about/permissions.dtl

*Science* (print ISSN 0036-8075; online ISSN 1095-9203) is published weekly, except the last week in December, by the American Association for the Advancement of Science, 1200 New York Avenue NW, Washington, DC 20005. Copyright 2009 by the American Association for the Advancement of Science; all rights reserved. The title *Science* is a registered trademark of AAAS.

level of signaling. This gap is spatial (cellular distances are on the scale of micrometers; protein movements are measured in angstroms), organizational (compartmentalization and orderly arrangements of pathways are essential to life), temporal (molecular movements occur in picoseconds to seconds, whereas cellular communication may persist substantially longer), and combinatorial (the sheer numbers of cell types, proteins, small molecules, stimuli, etc., and the multiplicity of their functional relationships with each other lead to extraordinary numbers of possibilities). The challenge to bridge this gap will require powerful new methods, interdisciplinary strategies, and creative, bold minds. We are encouraged by efforts to apply information theory to quantitatively interpret signal transmission (*53*), by multiscale modeling to bridge the molecular simulations to biochemical networks (*54*), and by whole-cell mapping of signaling protein dynamics (*55*, *56*). A holistic picture of the entire orchestra of dynamic contributions to cellular signaling can now begin to be envisioned.

## References and Notes

1. H. Yu *et al.*, *Science* **322**, 104 (2008); published online 21 August 2008 (10.1126/science.1158684).
2. M. F. Perutz, F. S. Mathews, *J. Mol. Biol.* **21**, 199 (1966).
3. K. Henzler-Wildman, D. Kern, *Nature* **450**, 964 (2007).
4. K. Linderstrom-Lang, J. Schellman, in *The Enzymes*, P. Boyer, Ed. (Academic Press, New York, 1959), vol. 1, pp. 443–510.
5. A. Ansari *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **82**, 5000 (1985).
6. H. Frauenfelder, S. G. Sligar, P. G. Wolynes, *Science* **254**, 1598 (1991).
7. G. Weber, *Biochemistry* **11**, 864 (1972).
8. P. Sherwood, B. R. Brooks, M. S. Sansom, *Curr. Opin. Struct. Biol.* **18**, 630 (2008).
9. A. Mittermaier, L. E. Kay, *Science* **312**, 224 (2006).
10. C. Bustamante, *Annu. Rev. Biochem.* **77**, 45 (2008).
11. J. F. Swain, L. M. Gierasch, *Curr. Opin. Struct. Biol.* **16**, 102 (2006).
12. S. W. Lockless, R. Ranganathan, *Science* **286**, 295 (1999).
13. E. J. Fuentes, C. J. Der, A. L. Lee, *J. Mol. Biol.* **335**, 1105 (2004).
14. A. Dhulesia, J. Gsponer, M. Vendruscolo, *J. Am. Chem. Soc.* **130**, 8931 (2008).
15. N. Ota, D. A. Agard, *J. Mol. Biol.* **351**, 345 (2005).
16. P. De Los Rios *et al.*, *Biophys. J.* **89**, 14 (2005).
17. Y. Kong, M. Karplus, *Proteins Struct. Funct. Bioinf.* **74**, 145 (2009).
18. S. Gianni *et al.*, *Structure* **14**, 1801 (2006).
19. C. N. Chi *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **105**, 4679 (2008).
20. C. J. Tsai, A. del Sol, R. Nussinov, *J. Mol. Biol.* **378**, 1 (2008).
21. V. A. Jarymowycz, M. J. Stone, *Biochemistry* **47**, 13371 (2008).
22. N. Popovych, S. Sun, R. H. Ebright, C. G. Kalodimos, *Nat. Struct. Mol. Biol.* **13**, 831 (2006).
23. K. K. Frederick, M. S. Marlow, K. G. Valentine, A. J. Wand, *Nature* **448**, 325 (2007).
24. J. Gsponer *et al.*, *Structure* **16**, 736 (2008).
25. K. P. Hoeflich, M. Ikura, *Cell* **108**, 739 (2002).
26. C. K. Johnson, *Biochemistry* **45**, 14233 (2006).
27. J. P. Junker, F. Ziegler, M. Rief, *Science* **323**, 633 (2009).
28. S. Tripathi, J. J. Portman, *Proc. Natl. Acad. Sci. U.S.A.* **106**, 2104 (2009).
29. P. Li, I. R. S. Martins, G. K. Amarasinghe, M. K. Rosen, *Nat. Struct. Mol. Biol.* **15**, 613 (2008).
30. A. K. Dunker, I. Silman, V. N. Uversky, J. L. Sussman, *Curr. Opin. Struct. Biol.* **18**, 756 (2008).
31. P. Sarkar, C. Reichman, T. Saleh, R. B. Birge, C. G. Kalodimos, *Mol. Cell* **25**, 413 (2007).
32. A. K. Dunker *et al.*, *BMC Genomics* **9** (suppl. 2), S1 (2008).
33. A. K. Dunker, V. N. Uversky, *Nat. Chem. Biol.* **4**, 229 (2008).
34. M. Fuxreiter *et al.*, *Nat. Chem. Biol.* **4**, 728 (2008).
35. R. W. Kriwacki, L. Hengst, L. Tennant, S. I. Reed, P. E. Wright, *Proc. Natl. Acad. Sci. U.S.A.* **93**, 11504 (1996).
36. C. A. Galea *et al.*, *J. Mol. Biol.* **376**, 827 (2008).
37. K. Sugase, H. J. Dyson, P. E. Wright, *Nature* **447**, 1021 (2007).
38. D. Eliezer, A. G. Palmer 3rd, *Nature* **447**, 920 (2007).
39. T. Mittag *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **105**, 17772 (2008).
40. J. E. Dueber, B. J. Yeh, R. P. Bhattacharyya, W. A. Lim, *Curr. Opin. Struct. Biol.* **14**, 690 (2004).
41. T. Pawson, P. Nash, *Science* **300**, 445 (2003).
42. L. C. van den Berk *et al.*, *Biochemistry* **46**, 13629 (2007).
43. F. C. Peterson, R. R. Penkert, B. F. Volkman, K. E. Prehoda, *Mol. Cell* **13**, 665 (2004).
44. P. Mishra *et al.*, *Cell* **131**, 80 (2007).
45. J. D. Faraldo-Gómez, B. Roux, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 13643 (2007).
46. M. Sotomayor, K. Schulten, *Biophys. J.* **94**, 4621 (2008).
47. S.-C. Sue, C. Cervantes, E. A. Komives, H. J. Dyson, *J. Mol. Biol.* **380**, 917 (2008).
48. R. Cheong, A. Hoffmann, A. Levchenko, *Mol. Syst. Biol.* **4**, 192 (2008).
49. T. Y. Besschetnova *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **105**, 12289 (2008).
50. J. P. Vilardaga *et al.*, *Nat. Chem. Biol.* **4**, 126 (2008).
51. D. Marcellino *et al.*, *Neuropharmacology* **54**, 815 (2008).
52. S. B. Padrick *et al.*, *Mol. Cell* **32**, 426 (2008).
53. T. Lenaerts *et al.*, *BMC Struct. Biol.* **8**, 43 (2008).
54. M. Stein, R. R. Gabdoulline, R. C. Wade, *Curr. Opin. Struct. Biol.* **17**, 166 (2007).
55. J. M. Ankers, D. G. Spiller, M. R. White, C. V. Harper, *Curr. Opin. Biotechnol.* **19**, 375 (2008).
56. T. D. Pollard, *Biochem. Soc. Trans.* **36**, 425 (2008).

57. Supported by NIH grants GM027616 and OD000945 (L.M.G.).

10.1126/science.1169377

---

REVIEW

# Protein Dynamism and Evolvability

**Nobuhiko Tokuriki and Dan S. Tawfik***

**The traditional view that proteins possess absolute functional specificity and a single, fixed structure conflicts with their marked ability to adapt and evolve new functions and structures. We consider an alternative, "avant-garde view" in which proteins are conformationally dynamic and exhibit functional promiscuity. We surmise that these properties are the foundation stones of protein evolvability; they facilitate the divergence of new functions within existing folds and the evolution of entirely new folds. Packing modes of proteins also affect their evolvability, and poorly packed, disordered, and conformationally diverse proteins may exhibit high evolvability. This dynamic view of protein structure, function, and evolvability is extrapolated to describe hypothetical scenarios for the evolution of the early proteins and future research directions in the area of protein dynamism and evolution.**

Proteins are proficient, accurate, and specific. These characteristics generally correlate with a lack of versatility; however, proteins also exhibit a marked ability to acquire new functions and structures. The evidence for the evolutionary adaptability of proteins is compelling, not only in the vast range of proteins that have presumably diverged from a few common ancestors, but also in recent evolutionary events such as the emergence of drug resistance and enzymes that degrade chemicals that appeared on this planet only a few decades ago.

What are the features that make proteins evolvable? Evolution acts by enriching preexisting diversities. Proteins conforming to the traditional view of absolute functional specificity, and only one well-defined structure are therefore not likely to readily respond to new selection pressures. However, a "new view" of proteins as an ensemble of alternative substructures, or conformers, in equilibrium with their so-called "native state" currently prevails [the new view was originally proposed by R. L. Baldwin and K. A. Dill in relation to protein folding and was later extended to describe native state ensembles (*1*)]. The new view is more consistent with evolutionary adaptability and is extended here to an avant-garde view of protein dynamism and evolvability.

Conformational variability, or dynamism, is an inherent property of any polymeric chain. The conformational diversity observed in proteins ranges from fluctuations of side chains and movements of active-site loops to secondary structure exchanges and rearrangements of the entire protein fold. Alternate structural conformers can mediate alternate folds and functions (*1*, *2*). Such structural and functional diversity is the foundation of "protein evolvability," defined as the abil-

---

Department of Biological Chemistry, Weizmann Institute of Science, Rehovot 76100, Israel.

*To whom correspondence should be addressed. E-mail: tawfik@weizmann.ac.il

ity of proteins to rapidly adopt (i.e., within a few sequence changes) new functions within existing folds or even adopt entirely new folds.

Functional promiscuity seems to be the starting point for divergence of new functions. Mutations can shift the equilibrium toward alternative functions and structures and therefore make up the raw material on which selection acts. Here, we discuss mechanisms that enable proteins to accumulate a larger number of mutations and thereby facilitate their adaptation. We also outline possible scenarios for the evolution of the earliest proteins by drawing parallels from RNA and protein folding intermediates, and we speculate what their properties could teach us about primordial protein forms. As there are only few well-characterized examples of recent adaptations and the ancient ones are challenging to track (*3*), we outline possible mechanisms and driving forces for the divergence of new functions and folds that mostly remain in the realm of theory and need to be experimentally substantiated, as discussed in the last section.

## Local, Active-Site Flexibility Mediates Promiscuity and Evolvability

Active-sites loops are flexible, and their sampling of conformational ensembles at different time scales and magnitudes is related to catalysis and regulation (*4*, *5*). The enzymatic chemistry occurs within rigid, preorganized sites, but other steps (such as product release) may determine the enzymatic turnover rate and could be facilitated by conformational rearrangements. The same flexibility that is required for catalysis can provide the basis for functional diversity and the route to evolutionary divergence of new functions. Many proteins exhibit multiple cellular functions, and enzymes promiscuously catalyze reactions other than the ones they originally evolved to catalyze (*6*). This promiscuity and multispecificity could be ascribed to various ligands or substrates, shifting the equilibrium in favor of those minor conformations in the ensemble that bind them (*1*) (Fig. 1).

Earlier examples of multispecificity include an antibody that was shown to exist, before ligand addition, in equilibrium between several different binding site conformers. These conformers enabled the binding of two unrelated ligands, each of which could shift the equilibrium in favor of a different conformer (*7*). Recently, a nuclear magnetic resonance study of ubiquitin revealed an ensemble of conformers that are nearly identical to complexes of ubiquitin with 46 different partners (*8*).

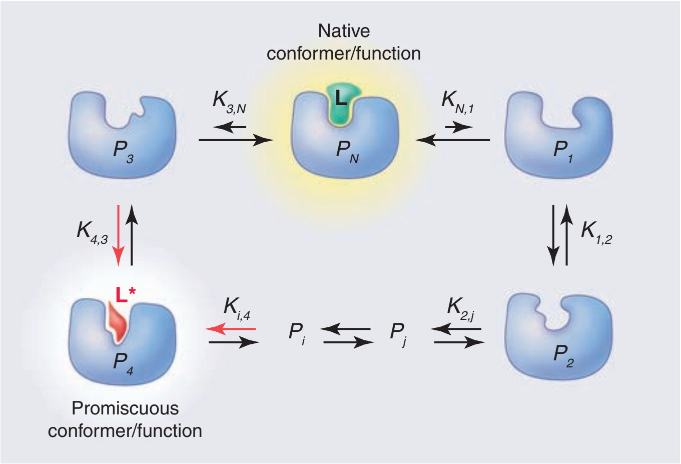

**Fig. 1.** The dynamics of protein structure and function and protein evolvability. The model assumes that proteins exist as an ensemble of conformations, the dominant one being the native state ($P_N$, interacting with the native ligand $L$). The alternative conformers relate to structural variations spanning from side-chain roatmers and active-site loop rearrangements to more profound fold transitions. Minor conformers (e.g., $P_4$) may mediate alternative functions, such as the promiscuous interaction with $L^*$ (where $L^*$ is a ligand that the protein did not evolve to bind). Mutations can gradually alter this equilibrium such that scarcely populated conformers become more favorable with substantial effects on the corresponding promiscuous function (e.g., an increase in occupancy of $P_4$ from 0.01 to 0.1 can yield a 10-fold increase in the overall level of promiscuous function). The relative occupancy of the native conformer would be hardly affected (e.g., from 0.5 to ≥0.41, leading to <20% loss of the native function). Similarly, higher specificity could evolve via mutations that reduce the occupancy of certain promiscuous conformers. This model also accounts for weak negative tradeoffs between the existing and evolving functions and the evolutionary potential of neutral mutations.

This inherent flexibility enables ubiquitin to bind multiple partners in a specific manner while fixing one out of many alternative conformations. Such conformationally plastic yet specific binding modes are also seen in T cell receptors and in "fuzzy complexes" where multiple conformations, or even complete disorder, prevail (*9*). Examples of flexible enzymes include a cytochrome P450 that displays a wide range of different active-site conformations that bind and transform a wide diversity of substrates (Fig. 2A) (*10*). Loop flexibility also enables domain repositioning in iron regulatory protein 1, where one conformer binds mRNA to repress translation or degradation and the other binds an iron-sulfur cluster and becomes an aconitase enzyme (*11*).

There seems to be a correlation between the degree of conformational diversity and promiscuity. In P450s, for example, the relatively rigid CYP2A6 exhibits narrow substrate specificity, whereas CYP3A4, the most promiscuous CYP known, exhibits the highest flexibility (*12*). In antibodies, increased affinity for a ligand gives decreased binding-site flexibility (*13*). Directed evolution experiments yield new enzymatic specificities by introducing mutations into inherently flexible active-site loops (*14*). These mutations are generally destabilizing (*14*), suggesting an increase in configurational entropy and active-site flexibility so that the new specificity comes from increasing a latent promiscuous function (*15*). Similarly, mutations acquired by selecting for one promiscuous function often induce broad specificity and allow other unselected substrates to be accommodated (*6*), possibly by allowing new degrees of freedom to active-site loops (*15*) (Fig. 1). Though the above suggests that more flexible active sites are more promiscuous and more evolvable, a link between active-site flexibility and evolvability is yet to be established. Similarly, a role of promiscuous ligand and substrate binding, or reaction, in providing a starting point for new gene functions has been only implicated in few instances (*16*–*18*) and needs to be more broadly established.

## Global Conformational Diversity and the Evolution of New Folds

Beyond the alternative side-chain rotamers and loop conformations described above (local flexibility), global structural rearrangements and fold transitions within the same sequence have also been observed (*2*, *19*). For example, lymphotactin exists in equilibrium between two different folds (Fig. 2B) (*20*), and Mad 2 is a homodimer that adopts two different β-sheet organizations (*21*).

Theoretical studies have also addressed the issue of fold transitions by mapping networks of sequence-fold spaces, primarily in RNA where secondary structures are accurately predicted (*22*), but also with lattice protein models (*23*). Each fold makes up a neutral network—a set of sequences that adopt the same structure (and presumably the same function) and are connected by single point mutations. Individual neutral networks are connected to one another at certain transition points (sequences) and can therefore be smoothly traversed by single mutational steps. This idea was demonstrated by isolating two ribozymes with different folds and functions that are connected by single point mutations and an intermediate that equilibrates between the two (*24*). In vitro evolution showed that a few mutations could convert one function into another while inducing a new fold (*25*). Similarly drastic transitions of fold and function are rarely demonstrated with proteins, as their folds are far more complex (the RNA-protein analogy is discussed in the last section of this review). One example is a 28–amino acid cysteine-rich domain, where one mutation resulted in an equilibrium between the original fold and the fold of another naturally occurring domain; another mutation completed this transition (Fig. 2C) (*26*) [see also recent reviews (*2*, *19*)].

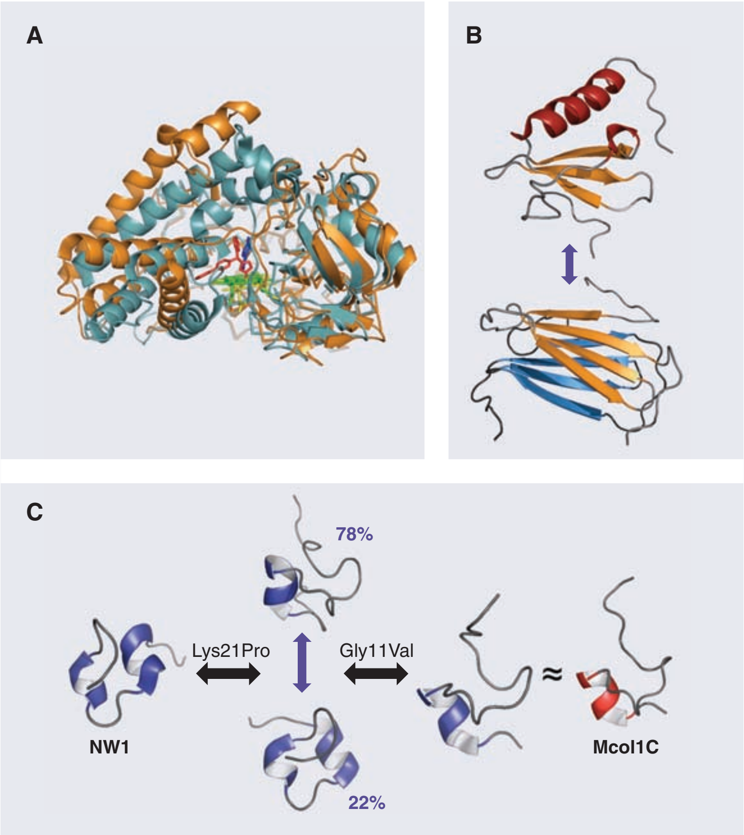

**Fig. 2.** Examples of conformational diversity. (**A**) Local conformational changes mediate an enzyme's broad substrate specificity. The open conformation of P450-CYP2B4 (orange) occurs with a large substrate (bifonazole, illustrated in red), and the closed one (light blue) occurrs with the smaller 4-(4-chlorophenyl) imidazole (darker blue) (*10*). (**B**) Metamorphic proteins. Lymphotactin exists in equilibrium between a beta-alpha mononer (top) and an all-beta dimer (bottom) (*20*). (**C**) Transition folds. Two different topologies (mediated by three different disulfide bridges) are found in two naturally occurring cysteine-rich domains (NW1 and Mcol1C) that show almost no sequence identity beyond the conserved cysteines. Conversion between these topologies was demonstrated via one mutation $\text{Lys}^{21} \rightarrow \text{Pro}^{21}$ (K21P) that afforded an intermediate that equilibrates between the two topologies, and a second mutation $\text{Gly}^{11} \rightarrow \text{Val}^{11}$ (G11V) completed the transition (*26*).

Other examples of conformational switches in natural proteins include prions that exist in a soluble form and aggregated amyloid form, and a myriad of intermediate oligomeric forms. Intrinsically disordered proteins (IDPs) make up another class of proteins where large structural transitions occur. Typically, order and tight packing are observed only when the ligand is bound (coupled binding-folding), and even then, other parts may remain disordered (*9*, *27*). IDPs are considered as a separate class of proteins with specific sequence compositions and functions that correlate with their unusual properties. Nevertheless, coupled binding and folding might have played a role in the evolution of the first protein forms and may be important in major fold transitions. Notably, order-disorder is not an all-or-nothing property. Short disordered segments are commonly observed within otherwise ordered proteins [and are often involved in function (*9*)], and relatively low degrees of order seem to characterize certain protein classes such as viral proteins (*28*). In fact, partial order may endow high tolerance to sequence changes and higher evolvability.

## Protein Evolvability and the Effects of Mutations

Evolution is the fixation of sequence changes, or mutations, that are often associated with functional changes and adaptation to new environments (adaptive mutations). However, mutations could also occur, and even fixate, with no apparent effects, as described by Kimura's neutral theory. A related issue is that evolvability has two contradictory components. Organisms, genes, and their encoded proteins are constantly exposed to mutations, and proteins whose neutrality or robustness (i.e., the ability to accommodate mutations without loss of structure and function) is limited may cause fitness losses at the organismal level. Yet because mutations generally accumulate one at a time, to adapt, the function and structure of a protein should change in response to few mutations (plasticity). Neutrality implies that mutations have no effect, whereas plasticity demands large effects of mutations on protein function and structure. Can the neutrality-plasticity dichotomy be reconciled (*29*)? Here we discuss this conundrum in relation to the functional and structural dynamism of proteins.

Contrary to the above dichotomy, comparisons of protein folds indicate that more neutral protein folds (folds that show higher sequence diversity) also exhibit higher functional diversification (*30*). How could that be? Gene duplication, relief of selection from the redundant copy, and accumulation of a large number of mutations before the acquisition of a new function is one possibility, but the abundance of deleterious mutations leading to rapid nonfunctionalization make this scenario unfavorable (*31*). Another explanation is that mutations can exhibit substantial effects on promiscuous protein functions but have minor effects on the native one (*6*). Likewise, mutations that initially appeared as neutral in a given environment (as polymorphism, or even fixed in a population) can facilitate future adaptations under changing circumstances (*29*) by altering latent promiscuous functions and conformers (Fig. 1) (*23*, *32*). Promiscuous conformers and functions can be also regarded as phenotypic variations. As is the case with regulatory networks, physiological adaptations such as the use of a promiscuous function in response to environmental changes, might correlate with changes that occur in response to mutations (evolutionary adaptations) (*33*). That said, tradeoffs between existing and evolving protein functions (as well as between correlated physiological and evolutionary adaptations) are far from being fully understood, and many other mechanisms operate in addition to the simplistic models described here.

Another aspect that relates to the neutrality-plasticity dichotomy is the structural characteristics of highly evolvable proteins. Traditionally, neutrality correlates with high thermodynamic stability (*34*). More than 80% of deleterious mutations relate to loss of protein stability, and the concomitant decline in the levels of soluble, functional protein. Thus, high stability (and, in particular, high thermodynamic stability owing to a more stable native state) correlates with well-packed, highly compact structures with increased tolerance to mutations (Fig. 3) (*35*). In cases such as the immunoglobulin fold, tight packing and highly robust scaffolds are combined with pronounced function and sequence diversity achieved primarily via changes in surface loops (e.g., antibody complementarity determining region loops). This separation of a tightly packed scaffold and floppy active site is also seen in many enzyme families (e.g., TIM barrels) and could simultaneously promote neutrality (via a robust scaffold) and plasticity (via loop changes).

However, most proteins exhibit limited stability (*34*), even (or especially) when placed under high mutational rates. Proteins from RNA viruses, where mutations rates are $\sim 10^6$-fold higher than in bacteria and eukarya, exhibit structural features that correlate with low stability:

namely, loose packing, low compactness, and a tendency to local disorder. These features may indicate an alternative mode of tolerating mutations because individual mutations lead to smaller stability losses due to weak interresidue contacts (*28*). This alternative mode of protein robustness is also supported by the observation that folds exhibiting the highest diversity in sequence and function show a higher tendency for disorder (*30*). Many IDPs show high rates of sequence diversification, and RNAs highly evolvable character arises from a small number of long-range tertiary contacts. Hence, higher evolvability might also correlate with loose packing (low degree of tertiary interactions), low compactness, and disorder (local or global). However, although higher neutrality of disordered regions has been recorded (*36*), their ability to rapidly evolve new functions and structures (plasticity) is yet to be established. The inevitable outcome of low stability could be partially compensated for by coupling function to folding or by destabilization of the unfolded state (negative design), rather than by high stability of the folded, native state (Fig. 3).

## The Evolution of Early Protein Folds—Speculations and Future Directions

As described above, the concepts of conformational diversity and functional promiscuity can be applied to describe how new proteins diverged from existing ones, gradually (through sequential mutations) and smoothly (through functional intermediates) (*1*, *2*, *23*). These concepts can be extended to the evolution of primordial protein precursors, an issue that is still in the realm of speculation and hypotheses.

The size of sequence space of biopolymers ($n^L$, where $n$ is the number of monomer types, and $L$ is chain length) can be daunting ($20^{100} \approx 10^{130}$ sequence permutations for a 100–amino acid polypeptide). It makes the emergence of sequences with function a highly improbable event, despite considerable redundancy (many sequences giving the same structure and function). The higher evolvability of RNA compared with that of proteins is partly due to a vastly smaller sequence space ($n = 4$, as opposed to 20). However, functional proteins can be obtained with minimal sets of as few as nine amino acids (*37*). Short polypeptides ($L < 30$) assembled as homo- or hetero-oligomers, can enhance function by avidity effects. Oligomerization, or even ordered β sheet–based aggregates (*38*), could also promote the stability and solubility of emerging folds by burying exposed hydrophobic surfaces. Oligomeric interfaces may have composed the first binding and active sites (as is the case in numerous modern proteins). Duplication and fusion could then have resulted in larger, single-domain, monomeric proteins. The internal symmetry of about half of the known folds and the isolation of putative oligomeric precursors for highly symmetrical folds such as β propellers support this hypothesis (*39*).

Another obstacle regarding the evolution of the first proteins (and enzymes in particular) is that function depends on a relatively structured native state, but structure itself provides no selective advantage. A scenario that function was selected for, and structure coevolved, is therefore likely (*40*) (Fig. 4A). Assuming that partially ordered polypeptides made up the first evolutionary intermediates, two scenarios could be proposed (Fig. 4B). In a manner similar to Haeckel's principle that ontogeny recapitulates phylogeny, folding intermediates may reflect the nature of evolutionary intermediates. Proteins fold by diverse pathways. In some cases, a nucleus of secondary structure (that can partly exist in the unfolded state) is followed by tertiary long-range contacts. Alternatively, folding may begin with a seed of hydrophobic core via long-range interactions, followed by the formation of secondary structure (*41*). There are potential pros and cons and anecdotal evidence supporting these scenarios. Molten globules that resemble advanced folding intermediates can exhibit enzymatic function (*42*). Secondary structure elements such as α helices and β strands seem to evolve by simple patterning of polar and nonpolar amino acids (*37*). The sequence constraints for β-strand formation in particular are very minimal (*38*). A key to RNA evolvability seems to be a range of stable, easily exchangeable secondary structure elements, with only few and much weaker long-range tertiary interactions. On the other hand, tertiary interactions can be crucial because elements of tertiary structure in the form of 20– to 30–amino acid–long loops closed by hydrophobic interactions were identified as the potential seeds of protein folds (*43*).

Overall, the dynamism of protein structure and function provides the grounds for evolutionary adaptations (whether they are new functions in existing folds or the emergence of completely new proteins) with the variety of models described here. However, key insights regarding protein evolution are still needed. How did the early protein forms evolve, and how do substantial fold transitions occur? Does global structural flexibility (and partial or complete disorder) provide higher evolvability of fold (Fig. 4)? And, conversely, does local flexibility of active-site loops, combined with a robust well-packed scaffold, promote functional changes within the same fold? High flexibility and large ensembles of alternative conformers are likely to result in lower activity, so do evolvability and activity tradeoff? Are, for example, viral enzymes, more evolvable and

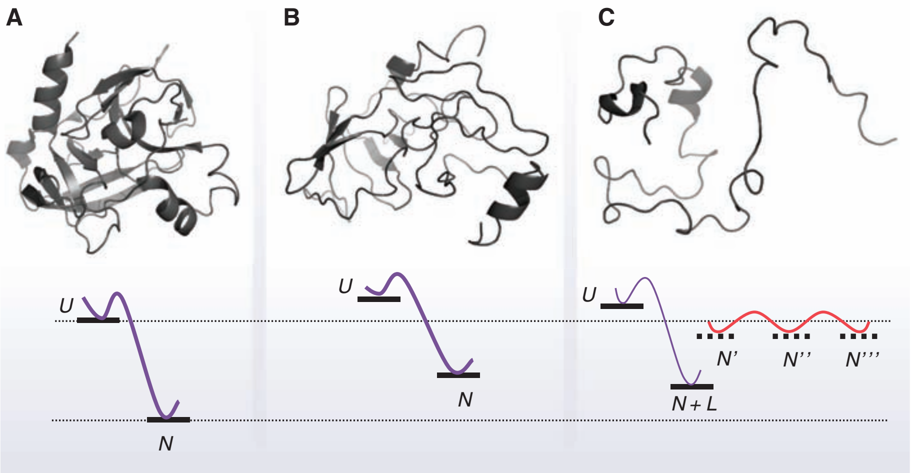

**Fig. 3.** Protein structure, stability, and evolvability. The figure depicts three stereotypes along an entire spectrum of packing orders. **(A)** Tightly packed, highly ordered, and compact proteins. An intense network of interresidue contacts makes the native state (*N*) highly favored and provides high stability [large energy difference between *U* (unfolded state) and *N*]. Such folds tolerate destabilizing mutations owing to an excess of stability that could be sacrificed without compromising their structural integrity. **(B)** Loosely packed proteins show lower degrees of interresidue contacts, fewer well-defined secondary structure elements (strands and helices), and higher fractions of loops and disordered segments. Therefore, the native state exhibits higher free energy, although the overall stability could be partly regained by destabilization of the unfolded state. Despite low stability, mutations are tolerated because sequence changes in weakly interacting residues cause smaller stability losses (*28*). **(C)** Disordered proteins adopt a defined structure only when in complex ($N + L$), but even then, their bound states are fairly loose with few long-range tertiary interactions. Their native state is composed of an ensemble of different conformers of similar energy ($N'$, $N''$, etc.), and they are highly amenable to sequence changes (*27*).

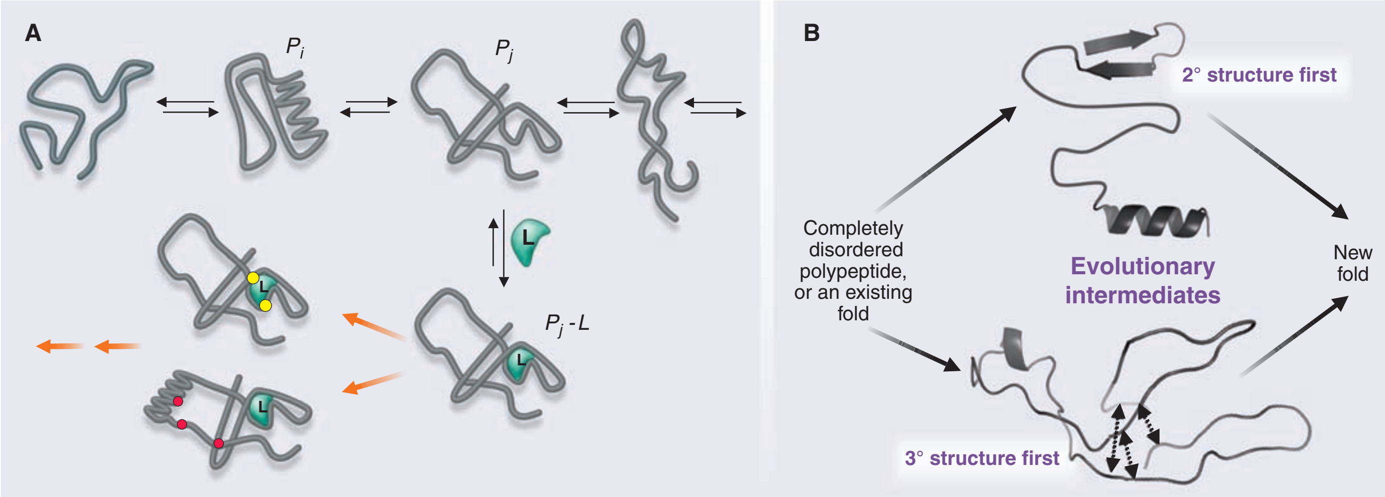
A
$P_i$
$P_j$
L
$P_j$-L
B
2° structure first
Completely disordered polypeptide, or an existing fold
Evolutionary intermediates
New fold
3° structure first

**Fig. 4.** Putative scenarios for the evolution of the primordial proteins. (**A**) Coevolution of fold and function via conformational selection from a repertoire of disordered polypeptides ($P_i$, $P_j$, etc.). Binding of a ligand (*L*) or substrate shifts the equilibrium in favor of a given conformer ($P_j$). Subsequent mutations (orange arrows) provide higher levels of function by stabilizing the functional conformer (thus evolving a fold, shown with red dots) and by altering the ligand contacting residues (yellow dots). (**B**) Evolutionary intermediates leading from a disordered protein to a folded one may resemble folding intermediates: 2° structure first involves the emergence of some secondary structure elements followed by the evolution of tertiary, long-range contacts and a fully folded protein. 3° structure first involves the formation of a rudimentary hydrophobic core by virtue of closing loops with a length of 20 to 30 amino acids (*43*), followed by the evolution of secondary structure.

hence less proficient than their highly ordered, well-packed orthologs? Is evolvability, therefore, an evolvable trait (*44*)? Traits such as functional promiscuity and conformational flexibility are inherent and need not be (and probably never were) selected for. Other traits such as robustness to mutations may have resulted from environmental, rather than genetic, pressures (*45*). It could be, however, that by virtue of their evolutionary history, certain classes of enzymes are more evolvable than others. Secondary metabolism is constantly responding to environmental changes, whereas core metabolism remained largely unchanged. Are secondary metabolism enzymes more flexible, more evolvable, and generally less proficient? Theory and simulations, perhaps of close-to-actual protein structures rather than lattice models, could help to address these questions, as can experiments, including in vitro evolution experiments that enable the actual reconstruction of evolutionary processes. The latter can use entirely random sequences as starting points (*46*) but could be better guided by bioinformatics (*47*), including phylogenies of large and prevalent superfamilies and folds that may have appeared first. The experimental reconstruction of the emergence of these rudimentary protein folds (and of function) from relatively simple and short polypeptides and insights regarding the intermediates along the route present a major challenge for future research.

### References and Notes

1. L. C. James, D. S. Tawfik, *Trends Biochem. Sci.* **28**, 361 (2003).
2. S. Meier, S. Özbek, *Bioessays* **29**, 1095 (2007).
3. A. M. Dean, J. W. Thornton, *Nat. Rev. Genet.* **8**, 675 (2007).
4. K. A. Henzler-Wildman *et al.*, *Nature* **450**, 913 (2007).
5. N. M. Goodey, S. J. Benkovic, *Nat. Chem. Biol.* **4**, 474 (2008).
6. O. Khersonsky, C. Roodveldt, D. S. Tawfik, *Curr. Opin. Chem. Biol.* **10**, 498 (2006).
7. L. C. James, P. Roversi, D. S. Tawfik, *Science* **299**, 1362 (2003).
8. O. F. Lange *et al.*, *Science* **320**, 1471 (2008).
9. P. Tompa, M. Fuxreiter, *Trends Biochem. Sci.* **33**, 2 (2008).
10. B. K. Muralidhara, L. Sun, S. Negi, J. R. Halpert, *J. Mol. Biol.* **377**, 232 (2008).
11. W. E. Walden *et al.*, *Science* **314**, 1903 (2006).
12. J. Skopalík, P. Anzenbacher, M. Otyepka, *J. Phys. Chem. B* **112**, 8165 (2008).
13. I. F. Thorpe, C. L. Brooks, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 8821 (2007).
14. N. Tokuriki, F. Stricher, L. Serrano, D. S. Tawfik, *PLoS Comput. Biol.* **4**, e1000002 (2008).
15. Y. H. Chen, J. Delmas, J. Sirot, B. K. Shoichet, R. Bonnet, *J. Mol. Biol.* **348**, 349 (2005).
16. L. Afriat, C. Roodveldt, G. Manco, D. S. Tawfik, *Biochemistry* **45**, 13677 (2006).
17. J. T. Bridgham, S. M. Carroll, J. W. Thornton, *Science* **312**, 97 (2006).
18. D. W. Brown, M. R. Schaab, W. R. Birmingham, R. N. Armstrong, *Biochemistry* **48**, 1847 (2009).
19. A. Andreeva, A. G. Murzin, *Curr. Opin. Struct. Biol.* **16**, 399 (2006).
20. R. L. Tuinstra *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **105**, 5057 (2008).
21. M. Mapelli, L. Massimiliano, S. Santaguida, A. Musacchio, *Cell* **131**, 730 (2007).
22. L. W. Ancel, W. Fontana, *J. Exp. Zool.* **288**, 242 (2000).
23. R. Wroe, H. S. Chan, E. Bornberg-Bauer, *HFSP J.* **1**, 79 (2007).
24. E. A. Schultes, D. P. Bartel, *Science* **289**, 448 (2000).
25. E. A. Curtis, D. P. Bartel, *Nat. Struct. Mol. Biol.* **12**, 994 (2005).
26. S. Meier *et al.*, *Curr. Biol.* **17**, 173 (2007).
27. H. J. Dyson, P. E. Wright, *Nat. Rev. Mol. Cell Biol.* **6**, 197 (2005).
28. N. Tokuriki, C. J. Oldfield, V. N. Uversky, I. N. Berezovsky, D. S. Tawfik, *Trends Biochem. Sci.* **34**, 53 (2009).
29. A. Wagner, *Nat. Rev. Genet.* **9**, 965 (2008).
30. E. Ferrada, A. Wagner, *Proc. Biol. Sci.* **275**, 1595 (2008).
31. S. Bershtein, D. S. Tawfik, *Mol. Biol. Evol.* **25**, 2311 (2008).
32. G. Amitai, R. D. Gupta, D. S. Tawfik, *HFSP J.* **1**, 67 (2007).
33. R. Milo, J. H. Hou, M. Springer, M. P. Brenner, M. W. Kirschner, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 16998 (2007).
34. M. A. DePristo, D. M. Weinreich, D. L. Hartl, *Nat. Rev. Genet.* **6**, 678 (2005).
35. E. Bornberg-Bauer, H. S. Chan, *Proc. Natl. Acad. Sci. U.S.A.* **96**, 10689 (1999).
36. C. J. Brown *et al.*, *J. Mol. Evol.* **55**, 104 (2002).
37. Y. Wei, M. H. Hecht, *Protein Eng. Des. Sel.* **17**, 67 (2004).
38. A. Corazza *et al.*, *Proteins* **62**, 64 (2006).
39. I. Yadid, D. S. Tawfik, *J. Mol. Biol.* **365**, 10 (2007).
40. T. Yomo, S. Saito, M. Sasai, *Nat. Struct. Biol.* **6**, 743 (1999).
41. V. Daggett, A. R. Fersht, *Trends Biochem. Sci.* **28**, 18 (2003).
42. K. Vamvaca, I. Jelesarov, D. Hilvert, *J. Mol. Biol.* **382**, 971 (2008).
43. E. N. Trifonov, I. N. Berezovsky, *Curr. Opin. Struct. Biol.* **13**, 110 (2003).
44. D. J. Earl, M. W. Deem, *Proc. Natl. Acad. Sci. U.S.A.* **101**, 11531 (2004).
45. J. A. de Visser *et al.*, *Evol. Int. J. Org. Evol.* **57**, 1959 (2003).
46. A. D. Keefe, J. W. Szostak, *Nature* **410**, 715 (2001).
47. V. Anantharaman, L. Aravind, E. V. Koonin, *Curr. Opin. Chem. Biol.* **7**, 12 (2003).
48. We gratefully acknowledge financial support from the Israel Science Foundation and the European Union MiFEM network. We thank M. Fuxreiter, L. James, and R. Milo for their insightful comments.

10.1126/science.1169375