

Exploring folding free energy landscapes using computational protein design

Brian Kuhlman¹ and David Baker²

Recent advances in computational protein design have allowed exciting new insights into the sequence dependence of protein folding free energy landscapes. Whereas most previous studies have examined the sequence dependence of protein stability and folding kinetics by characterizing naturally occurring proteins and variants of these proteins that contain a small number of mutations, it is now possible to generate and characterize computationally designed proteins that differ significantly from naturally occurring proteins in sequence and/or structure. These computer-generated proteins provide insights into the determinants of protein structure, stability and folding, and make it possible to disentangle the properties of proteins that are the consequence of natural selection from those that reflect the fundamental physical chemistry of polypeptide chains.

Addresses

¹Department of Biochemistry and Biophysics, University of North Carolina, Chapel Hill, NC 27599-7260, USA

e-mail: bkuhlman@email.unc.edu

²Department of Biochemistry, University of Washington, Seattle, WA 98195, USA

e-mail: dabaker@u.washington.edu

Current Opinion in Structural Biology 2004, **14**:89–95

This review comes from a themed issue on
Folding and binding
Edited by David Baker and William A Eaton

0959-440X/\$ – see front matter
© 2004 Elsevier Ltd. All rights reserved.

DOI 10.1016/j.sbi.2004.01.002

Introduction

Protein design is a rigorous test of our understanding of protein folding and stability [1,2]: how can you make something if you do not understand what is holding it together? Characterization of the properties of proteins generated using computational protein design methods, which identify the lowest energy amino acid sequences according to clearly defined energy functions, provides an excellent test of the accuracy of the underlying energy models [3–5]. Systematic changes in the energy function can be combined with experimental characterization of the designed proteins to identify the primary determinants of protein structure and stability. More generally, protein design can help distinguish the effects of natural selection from the fundamental physical properties of proteins. For instance, have proteins been optimized for fast folding by natural selection? This can be investigated by kinetic characterization of man-made proteins

that have not experienced natural selection and that were designed with no consideration of folding kinetics. Or, to what extent has nature sampled the viable regions of protein structure space? This question can be investigated by attempting to create proteins with structures that are not observed in nature. In this review, we will focus on recent examples in which computational protein design has been combined with experiments to help answer these and related questions.

Enhancing protein stability

Creating more stable proteins was one of the first problems that computational protein design was applied to and this subject has been reviewed previously [6]. Recent designs confirm many of the findings from earlier studies. A simple way to stabilize a protein is to increase hydrophobic surface area burial [7]. Table 1 shows the results of recent computer-based redesigns of naturally occurring proteins. In almost all the cases in which protein stability was significantly increased, there were more hydrophobic residues in the redesign than in the wild-type protein. An increase in hydrophobic surface area burial can most easily be accomplished by mutating partially buried polar residues to hydrophobic residues; these types of changes are likely to account for the increased stability of the redesigns of human growth hormone, the RNA-binding protein U1A and procarboxypeptidase [8*,9*]. A more difficult approach is to redesign a protein core so that it is packed more efficiently and contains more hydrophobic groups. Core redesigns of T4 lysozyme that mostly switched nonpolar amino acids with different nonpolar amino acids did not stabilize the protein [10**], and similar redesigns of spectrin SH3 were initially unsuccessful because of over-packing [11*].

These results do not mean that it is always advantageous to remove buried polar groups. Mayo and co-workers found that, by applying a filter based on the number of satisfied hydrogen bonds, they could identify which polar groups are best left intact in the core of thioredoxin [12*]. In another study, they demonstrated that a filter that favors positive interactions with the helix dipole and helix capping interactions led to more stable redesigns of the *Drosophila* engrailed homeodomain [13].

Redesigning protein folding pathways

Several lines of evidence suggest that the dominant protein folding pathways are those that maximize the formation of favorable native interactions while minimizing the loss of chain configurational entropy [14]. Protein

Table 1

Stability changes in computer-based redesigns of naturally occurring proteins.

Protein	Scope	Description	Stability	References
T4 lysozyme – Core-10	Core residues	10 mutations: 9 nonpolar to nonpolar, 1 polar to nonpolar	< WT	[10**]
Spectrin SH3 – Best 5	Core residues	5 mutations: A to V, V to L, M to I, V to I, V to L	= WT	[11*]
Spectrin SH3 – Best 5-I44V	Core residues	4 mutations: A to V, V to L, M to I, V to L	> WT	[11*]
Human growth hormone – CORE2	Core residues	8 mutations including 5 polar to nonpolar	> WT	[8*]
Granulocyte colony-stimulating factor	Core residues	10 polar to nonpolar	> WT	[43]
Thioredoxin – Prudent polar	Core residues	3 mutations: D to I, L to I, Y to F	> WT	[12*]
Thioredoxin – No polar	Core residues	5 mutations: D to I, Y to F, L to I, T to L, Y to F, T to A	= WT	[12*]
WW domain – SPANS-WW2	Complete redesign	WT: 53% polar residues. Redesign: 62% polar residues	< WT	[44*]
Protein L – pL1	Complete redesign	WT: 51% polar residues. Redesign: 54% polar residues	< WT	[9*]
Protein L – pL2	Complete redesign	WT: 51% polar residues. Redesign: 52% polar residues	= WT	[9*]
Procarboxypeptidase AYE	Complete redesign	WT: 47% polar residues. Redesign: 39% polar residues	> WT	[9*]
RNA-binding U1A – URN	Complete redesign	WT: 48% polar residues. Redesign: 44% polar residues	> WT	[9*]
Acylphosphatase – ACY	Complete redesign	WT: 50% polar residues. Redesign: 49% polar residues	> WT	[9*]
Homeodomain – ENH-FSM1	Complete redesign	WT: 63% polar residues. Redesign: 67% polar residues	> WT	[40**]
Homeodomain – NC3-Ncap	Surface residues	Added favorable N-capping and helix dipole interactions	> WT	[13]

WT, wild type.

folding pathways should thus be determined, in part, by the intrinsic stability of substructures within the protein. This hypothesis has been investigated through the characterization of truncation mutants, circular permutants and point mutants [15–17].

A particularly direct test is to stabilize specific substructures within a protein and determine whether these structures become populated earlier in folding. We used computational protein design to redesign the folding pathways of the IgG-binding domains of protein G and protein L by stabilizing hairpins that normally form late in the folding process [18**]. Protein G and protein L both consist of a single α helix packed against a four-stranded β sheet formed by symmetrically opposed β hairpins. In protein G, the second β turn is formed and the first is disrupted at the rate-limiting step of folding, whereas the reverse is true for protein L. For both proteins, computational protein sequence design, coupled with sampling alternative hairpin backbone conformations, was used to identify approximately ten mutations that stabilize the late-forming hairpin. When these mutations were combined with destabilizing mutations in the opposing hairpin, the primary folding pathways were switched (Figure 1). The crystal structures of the redesigned proteins (Figure 1d,e) were quite close to the design targets, showing that loop conformations can be altered in a predictable manner. The ability to redesign protein folding pathways shows that there has been considerable progress in our understanding of the determinants of protein folding mechanisms.

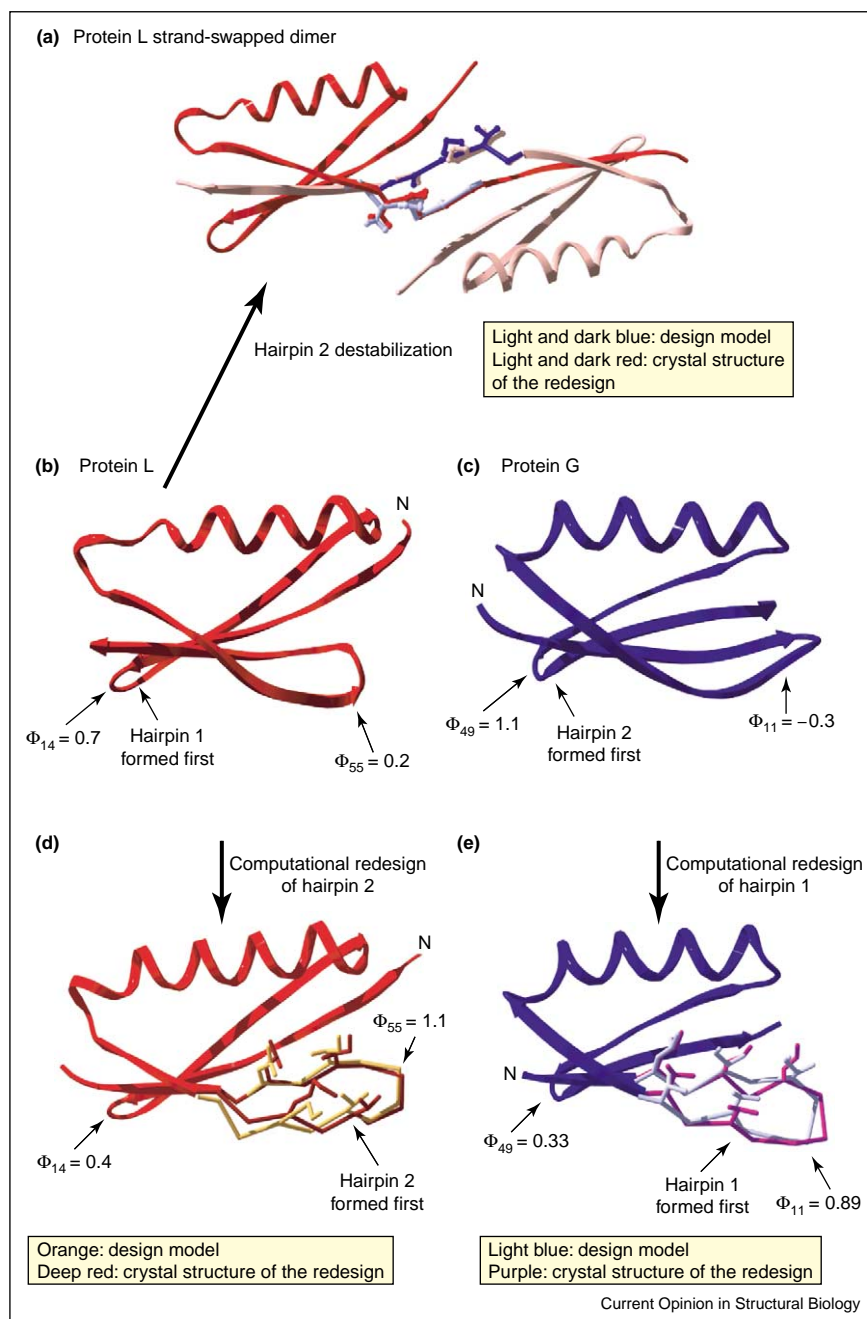
Interestingly, when mutations are made to protein L that destabilize the second β turn, instead of stabilizing it (as in the above experiment), the protein adopts a domain-swapped dimer in which the segment corresponding to the turn in the wild-type protein remains in an extended

conformation. The association constant for this dimer was dramatically enhanced by using computer-based design to identify sequences compatible with the structure of the dimer, but incompatible with the monomeric structure. A crystal structure of the redesigned dimer confirmed that it adopted the target structure (Figure 1a) [19]. Other destabilizing mutations in protein G have been shown to create domain-swapped tetramers and dimers [20,21]. These results illustrate how small amino acid sequence changes can produce large changes in quaternary structure [22].

Exploring protein fold space

There are a large but finite number of protein folds observed thus far in nature, and it is not clear whether structures not yet observed are physically unrealizable or simply have not yet been sampled by the evolutionary process or characterized by a structural biologist. Methods for the computational protein design of novel protein structures provide a route to answering this question. The primary challenge of designing novel proteins is that it is unlikely that any arbitrarily chosen protein backbone will be designable and therefore it is essential that the design procedure include a search of nearby conformational space in addition to sequence space. This was clearly demonstrated in the landmark design by Harbury *et al.* [23] of coiled-coil oligomers that adopted an unprecedented right-handed superhelical fold. In this study, the authors took advantage of the inherent symmetry of coiled coils to create an algebraic parameterization of the backbone that allowed the rapid optimization of the protein backbone for a large number of putative sequences. Another approach is to optimize the amino acid sequence for a large number of fixed backbone conformations, but this approach is restricted by the need to specify, in advance, a limited number of backbone conformations [24–27].

Figure 1

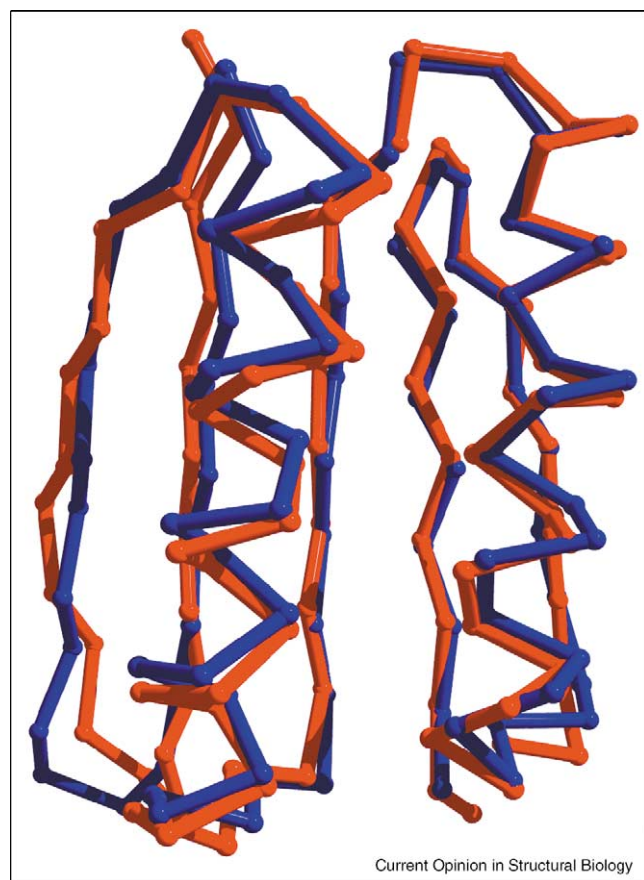


Redesigning protein folding pathways. Wild-type protein G and protein L both fold via an asymmetric transition state. **(b)** The first hairpin of protein L is formed at the rate-limiting step of folding, but the second hairpin is not [45], **(c)** whereas the reverse is true for protein G (representative Φ values are shown for each hairpin) [46]. **(d,e)** In each case, the late-forming hairpins were redesigned to increase their intrinsic stability; the redesigned proteins, when combined with destabilizing mutations in the other hairpin, folded via a switched mechanism [18^{••}]. Crystal structures of the designs (each design contained over ten mutations) confirmed that the backbone coordinates of the hairpins matched the design models fairly closely [27,47]. **(a)** When the first hairpin of protein L is destabilized, the protein forms a strand-swapped dimer [19]. Domain swapping has also been observed with protein G [21].

Recently, we used a general flexible backbone protein design strategy to create a 93-residue α/β protein, called Top7, with a topology not present in the Protein Data Bank (PDB) as an independent domain [28^{••}]. In this

study, alternating cycles of backbone and sequence optimization were used to find low free energy sequence-structure pairs [29,30]. The energy function was used to guide the search at all stages and, at each stage, only the

Figure 2



Creating a new protein fold. Superposition of the design model (red) and the crystal structure (blue) of a newly created protein (PDB code 1QYS) [28[•]]. The topology of the designed protein is not present as an independently folded domain in the PDB, but is observed as a discontinuous portion of the protein *S*-adenosylmethionine decarboxylase.

lowest energy sequence or structure identified in the previous iteration was optimized, thereby avoiding the large-scale and computationally expensive enumeration of alternative backbones or general sequences. The designed sequence has no detectable similarity with known protein sequences and a crystal structure of the design, in addition to confirming that the protein has a novel fold, is remarkably similar to the design model (rmsd = 1.2 Å; Figure 2). The very close correspondence between the design model and the experimental crystal structure suggests that the potential function guiding the design process captures much of the important physical chemistry.

Top7 is exceptionally stable, ~ 13 kcal mol⁻¹, and hence the absence of the Top7 fold as an independent domain in the PDB cannot be attributed to the basic physical chemistry of proteins. Evolution may simply not have had

the time to sample this fold (there are approximately 1000 different motifs that can be generated by five-stranded β sheets and many of these are currently not observed in the PDB [31]). The successful design of Top7 suggests that there are likely to be a large number of highly stable globular protein folds that have not been sampled by nature; many of these could perhaps be useful in a wide variety of applications.

One interesting aspect of the Top7 design procedure is that, aside from forcing polar residues at some surface positions, no explicit measures were used to destabilize alternative conformations. In comparison, several groups have noticed that it is often essential to screen against alternative folds when designing helical-bundle proteins [1,23,32,33] and designs of all- β proteins often aggregate. Specificity may be easier to achieve in α/β proteins because there is a strong intrinsic preference for right-handed $\beta\alpha\beta$ motifs [34], and short connections between the helices and strands can be used to prevent many alternative conformations. In a recent study, Takada and co-workers [35[•]] designed a three-helix-bundle protein by explicitly destabilizing a large set of alternative conformations. Although they needed to use simplified models of amino acid sidechains to complete this computationally intensive calculation, they were able to create a stable protein with a cooperative temperature melt and a well-defined NMR spectrum. It will be interesting to see if the structure of the protein matches their design target.

There have also been several computational design studies in which the target structure was a known fold, but the backbone coordinates were created from scratch. A 216-residue α/β -barrel protein was designed by creating an idealized backbone from simple geometric parameters. The design algorithm ORBIT was then used to find low energy sequences for the structure [36[•]]. NMR and CD spectra of the designed protein are consistent with the target structure. In another study, DeGrado and co-workers designed an A_2B_2 di-iron protein by creating an idealized four-helix bundle and searching for sequences that stabilize the target structure and destabilize alternative structures. Experimental studies verified that the protein is a helical A_2B_2 heterotetramer [37[•]].

Folding kinetics of computer-generated proteins

Because of the enormous number of conformations a protein can adopt, it has been suggested that fast folding rates may be the product of strong evolutionary pressure to fold on a biologically relevant timescale. An alternative hypothesis is that fast folding kinetics and smooth folding free energy landscapes are the consequence of selection for protein stability (similar sets of interactions may stabilize native structures and the partially folded

Table 2

Folding kinetics of computer-generated proteins.

Protein		k_f (s ⁻¹)	k_u (s ⁻¹)	ΔG (kcal mol ⁻¹)	References
Acylphosphatase	WT	0.2	0.000065	4.8	M Scalley-Kim, D Baker, unpublished
	Des	22 000	0.3	6.6	
Procarboxypeptidase	WT	408	0.5	4.0	M Scalley-Kim, D Baker, unpublished
	Des	51 000	0.000019	12.7	
Src SH3	WT	57	0.1	-0.2	M Scalley-Kim, D Baker, unpublished
	Des	6.8	9.6	3.8	
Protein L	WT	60	0.02	4.7	M Scalley-Kim, D Baker, unpublished
	Des	555	0.2	4.6	
Homeodomain – NC3-Ncap	WT	37 000	1100	1.8	[40**]
	Des	29 000	230	2.9	
Homeodomain – ENH-FSM1	WT	37 000	1100	1.8	[40**]
	Des	79 000	920	2.7	
Spectrin SH3 – Best 5	WT	3.9	0.0045	3.8	[11*]
	Des	22.8	0.0190	3.6	

Des, designed protein; WT, wild type.

conformations sampled during folding). This question can be addressed by studying the folding of proteins that have not been generated by the natural evolutionary process. One approach is to use experimental selection strategies, such as phage display, to retrieve sequences that fold from random libraries [38,39]. Although powerful, this approach is constrained by limits on the complexity of the libraries that can be generated and the selection pressures are not always clear (in phage display experiments, for example, the randomized proteins must be assembled onto the phage surface). Computational protein design is an excellent avenue for addressing this question as very large sequence changes can be achieved and the selection pressure — the optimization of the energy function used in the design process — is clear. In particular, computational design methods typically focus entirely on the stability of the native state and are completely ignorant of the kinetic folding process.

Table 2 shows the folding rates for seven computer-generated proteins; five of these were complete redesigns of naturally occurring proteins and, on average, 65% of the residues are mutated ([40**]; M Scalley-Kim, D Baker, unpublished). Of the complete redesigns, four out of five fold faster than the wild-type protein, suggesting that naturally occurring proteins are not extensively optimized for fast folding. Interestingly, three out of the five also unfold faster than their wild-type counterpart (this was also observed when five residues in the spectrin SH3 core were redesigned [11*] and when the core of the four-helix-bundle protein Rop was redesigned [41]). From an evolutionary standpoint, fast unfolding may be disadvantageous as it may be correlated with decreases in protein rigidity. Loss of rigidity may affect the protein's ability to function, as well as increase its susceptibility to protein misfolding events, such as incorporation into amyloid fibrils. The rigidity of a *de novo* designed helical-bundle protein was probed by NMR relaxation studies and the

protein was found to be more flexible than most naturally occurring proteins [42].

The folding kinetics of Top7 are of particular interest because the sequence and structure are completely novel. Unlike many small globular proteins, Top7 does not appear to fold by a simple two-state mechanism, but rather, at low concentrations of denaturant, there are two folding phases and the rate constants of both phases are independent of denaturant (M Scalley-Kim, D Baker, unpublished). These results suggest that there are partially folded intermediates or kinetic traps that slow the folding of Top7. Most naturally occurring, small-domain proteins do not exhibit such complex folding kinetics, suggesting that nature has selected for proteins that have smooth energy landscapes, perhaps to prevent the interaction of partially folded proteins with other cellular components.

Conclusions

Computational protein design provides a route to creating dramatically new protein sequences and structures, and thus makes it possible to investigate the extent that natural selection has shaped the properties of folding free energy landscapes and the viability of novel protein folds. As methods for protein design continue to improve, it should be possible to explore more subtle properties of proteins, for example, the interplay between dynamics, binding and catalysis, by comparing the properties of designed functional proteins with those of their naturally occurring counterparts.

Update

The recently determined high-resolution structure of a four-helix-bundle protein from a binary hydrophobic-polar patterned library challenges the notion that sequences that fold into a very well packed structure are rare [48**].

References and recommended reading

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

- of special interest
 - of outstanding interest
1. Hill RB, Raleigh DP, Lombardi A, DeGrado WF: **De novo design of helical bundles as models for understanding protein folding and function.** *Acc Chem Res* 2000, **33**:745-754.
 2. DeGrado WF, Summa CM, Pavone V, Nastri F, Lombardi A: **De novo design and structural characterization of proteins and metalloproteins.** *Annu Rev Biochem* 1999, **68**:779-819.
 3. Hellenga HW: **Computational protein engineering.** *Nat Struct Biol* 1998, **5**:525-527.
 4. Kraemer-Pecore CM, Wollacott AM, Desjarlais JR: **Computational protein design.** *Curr Opin Chem Biol* 2001, **5**:690-695.
 5. Street AG, Mayo SL: **Computational protein design.** *Structure Fold Des* 1999, **7**:R105-R109.
 6. Dahiyat BI: **In silico design for protein stabilization.** *Curr Opin Biotechnol* 1999, **10**:387-390.
 7. Malakauskas SM, Mayo SL: **Design, structure and stability of a hyperthermophilic protein variant.** *Nat Struct Biol* 1998, **5**:470-475.
 8. Filikov AV, Hayes RJ, Luo P, Stark DM, Chan C, Kundu A, Dahiyat BI: **Computational stabilization of human growth hormone.** *Protein Sci* 2002, **11**:1452-1461.
- Computational design was used to identify eight mutations that increase the thermal stability of human growth hormone by approximately 15°C. Five of the mutations replace buried polar amino acids with nonpolar amino acids.
9. Dantas G, Kuhlman B, Callender D, Wong M, Baker D: **A large scale test of computational protein design: folding and stability of nine completely redesigned globular proteins.** *J Mol Biol* 2003, **332**:449-460.
- The computer program RosettaDesign was used to identify new sequences for nine naturally occurring proteins. Four of the designed proteins have cooperative thermal melts, and NMR spectra with sharp lines and good dispersion. The designs that are more stable than the wild-type proteins have more nonpolar residues than the wild-type proteins.
10. Mooers BH, Datta D, Baase WA, Zollars ES, Mayo SL, Matthews BW: **Repacking the core of T4 lysozyme by automated design.** *J Mol Biol* 2003, **332**:741-756.
- The program ORBIT was used to redesign the core of T4 lysozyme. Several constructs were well structured, but all of them were less stable than the wild-type protein. Impressively, crystal structures were determined for 15 of the mutant proteins and, in several cases, the backbone was changed significantly relative to the wild-type structure. These results suggest that the sequence in the core of T4 lysozyme is close to optimal for its structure.
11. Ventura S, Vega MC, Lacroix E, Angrand I, Spagnolo L, Serrano L: **Conformational strain in the hydrophobic core and its implications for protein folding and design.** *Nat Struct Biol* 2002, **9**:485-493.
- Computer-based redesigns of spectrin SH3 show accelerated unfolding and folding rates relative to the wild-type protein. Further mutations indicate that the preferential stabilization of the transition state is due to strain in the folded state that is relieved in the folding transition state.
12. Bolon DN, Marcus JS, Ross SA, Mayo SL: **Prudent modeling of core polar residues in computational protein design.** *J Mol Biol* 2003, **329**:611-622.
- A filter based on the number of hydrogen bonds an amino acid forms was used to determine which polar residues are best left intact when redesigning the core of a protein.
13. Marshall SA, Morgan CS, Mayo SL: **Electrostatics significantly affect the stability of designed homeodomain variants.** *J Mol Biol* 2002, **316**:189-199.
 14. Baker D: **A surprising simplicity to protein folding.** *Nature* 2000, **405**:39-42.
 15. Wright CF, Lindorff-Larsen K, Randles LG, Clarke J: **Parallel protein-unfolding pathways revealed and mapped.** *Nat Struct Biol* 2003, **10**:658-662.
 16. Lindberg M, Tangrot J, Oliveberg M: **Complete change of the protein folding transition state upon circular permutation.** *Nat Struct Biol* 2002, **9**:818-822.
 17. Nguyen H, Jager M, Moretto A, Gruebele M, Kelly JW: **Tuning the free-energy landscape of a WW domain by temperature, mutation, and truncation.** *Proc Natl Acad Sci USA* 2003, **100**:3948-3953.
 18. Nauli S, Kuhlman B, Baker D: **Computer based redesign of a protein folding pathway.** *Nat Struct Biol* 2001, **8**:602-605.
- The preferred folding pathway of protein G was switched by using computational protein design to stabilize a β hairpin that normally is not structured in the folding transition state. Unlike the wild-type protein, the first β hairpin is formed and the second is disrupted in the folding transition state.
19. Kuhlman B, O'Neill JW, Kim DE, Zhang KY, Baker D: **Conversion of monomeric protein L to an obligate dimer by computational protein design.** *Proc Natl Acad Sci USA* 2001, **98**:10687-10691.
 20. Kirsten Frank M, Dyda F, Dobrodumov A, Gronenborn AM: **Core mutations switch monomeric protein GB1 into an intertwined tetramer.** *Nat Struct Biol* 2002, **9**:877-885.
 21. Byeon IJ, Louis JM, Gronenborn AM: **A protein contortionist: core mutations of GB1 that induce dimerization and domain swapping.** *J Mol Biol* 2003, **333**:141-152.
 22. Liu Y, Eisenberg D: **3D domain swapping: as domains continue to swap.** *Protein Sci* 2002, **11**:1285-1299.
 23. Harbury PB, Plecs JJ, Tidor B, Alber T, Kim PS: **High-resolution protein design with backbone freedom.** *Science* 1998, **282**:1462-1467.
 24. Reina J, Lacroix E, Hobson SD, Fernandez-Ballester G, Rybin V, Schwab MS, Serrano L, Gonzalez C: **Computer-aided design of a PDZ domain to recognize new target sequences.** *Nat Struct Biol* 2002, **9**:621-627.
 25. Su A, Mayo SL: **Coupling backbone flexibility and amino acid sequence selection in protein design.** *Protein Sci* 1997, **6**:1701-1707.
 26. Larson SM, England JL, Desjarlais JR, Pande VS: **Thoroughly sampling sequence space: large-scale protein design of structural ensembles.** *Protein Sci* 2002, **11**:2804-2813.
 27. Kuhlman B, O'Neill JW, Kim DE, Zhang KY, Baker D: **Accurate computer-based design of a new backbone conformation in the second turn of protein L.** *J Mol Biol* 2002, **315**:471-477.
 28. Kuhlman B, Dantas G, Ireton G, Varani G, Stoddard B, Baker D: **Design of a novel globular protein fold with atomic level accuracy.** *Science* 2003, **302**:1364-1368.
- A novel protein fold was created by using a design procedure that iterates between sequence design and structure prediction. A crystal structure of the protein strongly resembles the design model (rmsd = 1.2 Å).
29. Desjarlais JR, Handel TM: **Side-chain and backbone flexibility in protein core design.** *J Mol Biol* 1999, **290**:305-318.
 30. Irbäck A, Peterson C, Potthast F, Sandelin E: **Design of sequences with good folding properties in coarse-grained protein models.** *Structure Fold Des* 1999, **7**:347-360.
 31. Ruczinski I, Kooperberg C, Bonneau R, Baker D: **Distributions of beta sheets in proteins with application to structure prediction.** *Proteins* 2002, **48**:85-97.
 32. Hill CP, Anderson DH, Wesson L, DeGrado WF, Eisenberg D: **Crystal structure of alpha 1: implications for protein design.** *Science* 1990, **249**:543-546.
 33. Havranek JJ, Harbury PB: **Automated design of specificity in molecular recognition.** *Nat Struct Biol* 2003, **10**:45-52.
 34. Sternberg MJ, Thornton JM: **On the conformation of proteins: the handedness of the connection between parallel beta-strands.** *J Mol Biol* 1977, **110**:269-283.
 35. Jin W, Kambara O, Sasakawa H, Tamura A, Takada S: **De novo design of foldable proteins with smooth folding funnel. Automated negative design and experimental verification.** *Structure (Camb)* 2003, **11**:581-590.

A computer-based procedure was used to explicitly screen for sequences that were optimal for the target structure, but incompatible with alternative structures. NMR and CD data indicate that the designed sequence has a well-defined structure.

36. Offredi F, Dubail F, Kischel P, Sarinski K, Stern AS,
• Van de Weerd C, Hoch JC, Prosperi C, Francois JM, Mayo SL *et al.*: **De novo backbone and sequence design of an idealized alpha/beta barrel protein: evidence of stable tertiary structure.** *J Mol Biol* 2003, **325**:163-174.

An idealized backbone for an α/β protein was built from simple geometric parameters and the automated sequence selection program ORBIT was used to identify low energy sequences for the target structure. Experimental data indicate that the protein adopts a folded α/β structure.

37. Summa CM, Rosenblatt MM, Hong JK, Lear JD, DeGrado WF:
• **Computational de novo design, and characterization of an A(2)B(2) diiron protein.** *J Mol Biol* 2002, **321**:923-938.

A computer-based design procedure was used to identify sequences compatible with the target structure, but incompatible with alternative structures. Experimental results indicate that the designed protein folds into a A_2B_2 heterotetramer as intended and has ferroxidase activity.

38. Riddle D, Santiago J, Grantcharova V, Baker D: **Functionally rapidly folding proteins from simplified amino acid sequences.** *Nat Struct Biol* 1997, **4**:805-809.
39. Kim DE, Gu H, Baker D: **The sequences of small proteins are not extensively optimized for rapid folding by natural selection.** *Proc Natl Acad Sci USA* 1998, **95**:4982-4986.
40. Gillespie B, Vu DM, Shah PS, Marshall SA, Dyer RB, Mayo SL,
• Plaxco KW: **NMR and temperature-jump measurements of de novo designed proteins demonstrate rapid folding in the absence of explicit selection for kinetics.** *J Mol Biol* 2003, **330**:813-819.

The folding kinetics were determined for two redesigns of the engrailed homeodomain. One of the redesigns has only 12 residues in common with the wild-type protein and yet folds just as fast. These results suggest that native protein sequences are not explicitly optimized for fast folding.

41. Munson M, Anderson KS, Regan L: **Speeding up protein folding: mutations that increase the rate at which Rop folds and unfolds by over four orders of magnitude.** *Fold Des* 1997, **2**:77-87.
42. Walsh ST, Lee AL, DeGrado WF, Wand AJ: **Dynamics of a de novo designed three-helix bundle protein studied by ^{15}N , ^{13}C , and 2H NMR relaxation methods.** *Biochemistry* 2001, **40**:9560-9569.
43. Luo P, Hayes RJ, Chan C, Stark DM, Hwang MY, Jacinto JM, Juvvadi P, Chung HS, Kundu A, Ary ML *et al.*: **Development of a cytokine analog with enhanced stability using computational ultrahigh throughput screening.** *Protein Sci* 2002, **11**:1218-1226.
44. Kraemer-Pecore CM, Lecomte JT, Desjarlais JR: **A de novo redesign of the WW domain.** *Protein Sci* 2003, **12**:2194-2205.
A new sequence was designed for a WW domain by searching for sequences compatible with an ensemble of closely related polypeptide backbones. NMR studies show that the designed protein folds into a structure closely resembling a WW domain.
45. Kim DE, Fisher C, Baker D: **A breakdown of symmetry in the folding transition state of protein L.** *J Mol Biol* 2000, **298**:971-984.
46. McCallister EL, Baker D: **Role of beta turn formation, local interactions, and topology in protein G folding.** *Nat Struct Biol* 2000, **7**:669-673.
47. Nauli S, Kuhlman B, Le Trong I, Stenkamp RE, Teller D, Baker D: **Crystal structures and increased stabilization of the protein G variants with switched folding pathways NuG1 and NuG2.** *Protein Sci* 2002, **11**:2924-2931.
48. Wei Y, Kim S, Fela D, Baum J, Hecht MH: **Solution structure of a de novo protein from a designed combinatorial library.** *Proc Natl Acad Sci USA* 2003, **100**:13270-13273.

A high-resolution structure was solved for a four-helix-bundle protein from a binary hydrophobic-polar patterned library. This result suggests that sequences that fold into well-packed structures are not extremely rare.