Title
Core-packing constraints, hydrophobicity and protein design.

Permalink
https://escholarship.org/uc/item/6k25c588

Journal
Current opinion in biotechnology, 5(4)

ISSN
0958-1669

Authors
Baldwin, EP
Matthews, BW

Publication Date
1994-08-01

DOI
10.1016/0958-1669(94)90048-5

Peer reviewed
Core-packing constraints, hydrophobicity and protein design

Enoch P Baldwin and Brian W Matthews

Howard Hughes Medical Institute, University of Oregon, Eugene, USA

Recent crystallographic studies have shown that both backbone and side-chain adjustments occur when different core-packing arrangements are accommodated in proteins. Thus, modeling methods, which have typically considered only side-chain adjustments, must now also account for backbone movements to accurately predict the energies and structures of mutated or designed proteins. The ‘plasticity’ of protein cores demonstrated by random mutagenesis simplifies protein design by increasing the likelihood of identifying alternative core sequences.

Current Opinion in Biotechnology 1994, 5:396–402

Introduction

To effectively engineer proteins, we need to understand how amino acid sequences specify structure, function and stability. Patterns of hydrophobic residues in polypeptide sequences are a key determinant of the type of fold that is adopted [1,2]. In soluble globular proteins, burial of hydrophobic residues drives compaction and helps specify secondary and tertiary structure [3,4,5]. Non-polar side chains typically pack tightly in the interior, forming the solvent-inaccessible protein 'core', whereas surface side chains are generally polar. Tight packing of buried residues has been suggested to limit the core sequences that are tolerated in a particular fold [6–9]. In computational approaches to protein design and structure analysis, packing constraints are a major consideration. Here, we briefly review the origins of these ideas and discuss recent work indicating that such constraints are, in fact, much looser than previously thought. Paradoxically, the structural basis for tolerance to different cores complicates prediction of the effects of mutation on structure and stability, but simplifies the problem of protein design.

The importance of the hydrophobic effect in determining protein structure and stability

Hydrophobicity patterns in amino acid sequences are one of the most conserved features of proteins that have the same fold [2]. Theoretical studies using highly simplified lattice models [10–12] of protein chains, as well as recent experimental work, suggest that such patterns can specify secondary and tertiary structure. For example, four-helix bundles were readily generated using polypeptide sequences based on helical repeats of hydrophobic and hydrophilic residues [13,14,15,16,17]. Coiled-coil dimers, trimers and tetramers have also been specified by simple patterns of hydrophobic residues [18,19]. In mutants containing insertions between residues that are in α-helices [19,20,21,22], the choice between 'looping out' or translocating adjacent helical residues was apparently controlled by the relative dispositions of polar and non-polar residues [19,20]. Finally, burial of a non-polar surface has been correlated with helical propensity at solvent-accessible sites in T4 lysozyme [23,24].

The importance of the hydrophobic nature of protein cores for stability is also well established [25,26,27]. Truncation of core side chains to alanine [27–31, 32–34] is much more destabilizing (up to 5 kcal mol⁻¹ for leucine residues) than alanine substitutions of most surface residues (±1 kcal mol⁻¹) [23,30,31,32–34,35,36]. Similarly, substituting non-polar buried side chains with smaller hydrophobic [27,28,37,38,39] or polar residues [40,41,42] is more destabilizing than increasing the hydrophobicity of surface-exposed side chains [23,24]. In some cases, substitution of buried polar residues with non-polar amino acids has led to stabilization [44,45,46].

Theoretical evaluation of packing constraints

The constraints imposed on protein structure and stability by core packing have not yet been clearly defined [25,26,47,48,49]. As judged from protein crystal structures, core side chains are 'well packed': they are rigid, clustered at high density, and generally in low-energy conformations. This jigsaw puzzle-like fit of side chains led to the suggestion that very few combinations of hydrophobic residues might be compatible

Abbreviation

rmsd—root mean square deviation.
with a given fold [7]. In this view, packing arrangements, and thus the identities of core residues, contain crucial information specifying protein structure.

To test this idea and to evaluate packing constraints, Ponder and Richards [7] designed a prototypical packing algorithm, PROPAK, that evaluates all of the possible side-chain arrangements of a group of adjacent residues. Trial models were constructed by placing alternative side chains in the context of a fixed framework of surrounding side-chain and backbone atoms derived from crystal structures. The alternative side-chain orientations were restricted to a limited number of 'ideal' or 'preferred' conformers (rotamer libraries) that were determined from surveys of known structures [7, 24**, 50, 51, 52*]. Models were evaluated by counting the number of close van der Waals contacts. Only a minute fraction of the total number of side-chain combinations, including the wild type, satisfied the contact criteria, suggesting that viable alternative core-packing arrangements are rare.

Recently, more sophisticated algorithms have been developed both to predict the mutational effects on protein stability and to build models of proteins from known homologous structures. These methods surveyed potential side-chain configurations, either by stepwise enumeration of the possible conformations for subsets of side chains [52*, 53, 54**–56*, 57, 58, 59*] or by sampling via Monte Carlo approaches [60] or other search methods [51, 61*, 62]. The side-chain conformations were generated either from rotamer libraries [51, 52*, 53, 55*, 58, 59*, 62] (as with PROPAK) or by stepping through coarse increments of torsion angles (10–120°) [54*, 57, 60], and were often subsequently optimized using energy functions [51, 53, 54**–56*, 57, 60] that are very sensitive to interatomic distances [63, 64*, 65]. As is the procedure with PROPAK, the model backbone and surrounding framework atoms were held fixed, at least during side-chain enumeration. Where overall energy minimization was performed on a limited subset of culled structures, the backbones did not deviate far from their initial positions [53, 54*, 58]. The inherent assumption of these approaches is that movements of the backbone away from the wild-type position to avoid steric conflicts are destabilizing [66].

**Mutational studies of protein cores**

Early mutational studies supported the idea that core residues are also sensitive to hydrophobic substitutions [25*, 26*, 47*, 48*]. Temperature-sensitive mutations often occur at buried sites [67, 68*], which generally have a more limited range of permissible substitutions than surface sites [68*, 69, 70*]. Collisions with the backbone, disruption of the packing of adjacent residues, and the forcing of side chains into unfavorable conformations, are all potential sources of destabilization at tightly packed sites, especially for small to large substitutions [58, 71–74, 75*, 76*]. In an extreme example, the substitution Ala98→Val (three-letter amino acid code) in T4 lysozyme destabilized the protein by 5 kcal mol⁻¹ [71]. The mutant structure showed that the increased size of the side chain perturbed the adjacent backbone. On the other hand, small to large mutants with increased stability have also been generated [43, 44*, 74, 77, 78*], demonstrating that increased hydrophobicity, improved van der Waals contacts, and greater packing efficiency can potentially overcome unfavorable factors, such as steric interference and torsional strain. Hydrophobicity and packing can have comparable contributions to stability. Interchanging side chains of similar hydrophobicity, but different shapes (e.g. phenylalanine, leucine, isoleucine and methionine), has lead to a wide range of stability changes (+0.3 kcal mol⁻¹ to −4 kcal mol⁻¹ [28, 37**, 38, 39, 58, 76*]).

Paradoxically, although proteins can be somewhat sensitive to single substitutions, they are surprisingly tolerant of multiple adjacent interior substitutions, provided that the overall hydrophobicity of the core is roughly maintained. Libraries of core-packing variants of the DNA-binding domain of λ repressor and T4 lysozyme were evaluated for function and stability [79, 80, 81**, 82**]. As many as 70% [80] of the possible combinations of hydrophobic side chains were tolerated, albeit few with stability near that of the wild type. Some sites were more sensitive to substitution than others. Stability or function did not correlate well with volume, hydrophobicity, root mean square deviation (rmsd) from wild-type structure (where structures were determined) or packing density, although mutants with larger hydrophobic cores tended to be more stable [48*, 82**]. The ranges of tolerated side-chain volumes and hydrophobicities were within ±5–6 methylenes and ±3–4 kcal mol⁻¹, respectively, of the wild type, but repressor activity was more sensitive to changes in volume than T4 lysozyme. A triple mutant of λ repressor was actually 0.5 kcal mol⁻¹ more stable than the wild type. Unlike multiple mutants in general, which usually have additive effects on stability [83*, 84], all core mutants with adjacent substitutions that have been examined were more stable (by up to 3.9 kcal mol⁻¹) than expected by additivity [58, 71, 76*, 82**, 83*]. This non-additivity indicates that well packed residues form cooperative arrangements that can be disrupted by a single substitution [47*, 48*, 85, 86]. Further substitutions result in new interactions, although rarely as cooperative as in the wild type.

**Structural studies of core mutants**

A number of crystal structures of single and multiple core mutants have recently been determined [29, 37**, 40*, 41, 42, 44*, 58, 71, 77*, 82**, 87, 88*, 90*], revealing why proteins are tolerant of changes in packing, and allowing evaluation of the assumptions that underlie current computational methods.
New packing arrangements were typically accommodated by overall adjustments of the backbone (up to 1.0 Å) without major distortions of secondary structure, together with small changes (10–20°) in side-chain torsion angles. In mutants in which the net volume of core side chains is reduced, structural relaxation decreases the amount of unfilled space [29,37**,87,88*–90°,91], presumably reducing destabilization [29,37**,92°]. Analogously, increasing total side-chain volume results in expansion of the cores and concerted shifts of secondary structure of up to 0.8 Å [44*,58,82**,91,93,94°,95*]. Increased packing efficiency and stability has been observed in some [44*,77°,95*], but not all, cases.

In a few mutant structures, substantial side-chain rotations toward cavities [29] or away from introduced bulky side chains have been observed [44*,82**]. Mutated side chains or side chains nearby the mutation sites occasionally adopt non-ideal torsion angles [37**,44*,58,77°,90°], even in stabilized mutants [44*,77°]. The conformations of alternative side chains in multiple mutants are generally similar to each other and are within 20° of wild-type or 'ideal' torsion angles [58,71,82**,95*]. Thus, they are determined by similar constraints, including the local secondary structure and the surrounding residues that are unchanged [52°,59°]. In homologous proteins with greater than 50% sequence identity, side-chain conformations and torsion angles are also typically preserved, supporting this idea [57,96–98]. Although changes in torsion angles contribute to the reduction of unfavorable contacts, shifts of secondary structure are largely responsible for the accommodation of altered side chains, at least in the helical proteins studied. Relatively small shifts in backbone α-carbons (0.5 Å) can be accompanied by larger changes (1–2 Å) in side-chain atom positions. These adjustments assist in redistributing side-chain bulk and optimizing new sets of contacts, so that many alternative side-chain combinations can lead to a tolerably well packed core.

For the same protein in different crystal environments, where it is subject to different crystal-packing forces, the backbone shifts are similar to those observed in the repacked mutants (0.3–0.4 Å rmsd after superposition [24**,80,96,97]), suggesting that protein backbones are inherently flexible, and adjustments of this magnitude may have little penalty [82**]. In an extreme case, one triple variant of T4 lysozyme had a backbone shift of 0.63 Å, yet was destabilized only by 1.4 kcal mol⁻¹ [81**]. Clearly, even relatively large deviations of crystallographically determined backbone positions from the wild type do not necessarily translate into large destabilizations.

The infrequency of alternative rotamers in packing mutants is likely due to the close-packed nature of cores. In an unusual T4 lysozyme core-repacking variant [82*], the substitution of Ala→Trp (a large for a small residue) required an adjacent amino acid on the same helix to adopt a different rotamer to avoid a steric clash. Because this residue was also completely buried, it was expected that other potential steric conflicts would be incurred. In this case, however, a potential collision was averted because the conflicting side chain was at the surface and freely rotated away. For tightly packed residues in the core, such coordinated conformational changes are likely to be more problematic.

Implications of core-repacking studies for protein design

Computational approaches to structure and stability prediction have been reasonably successful (60–90%) in reproducing core side-chain conformations from known structures [7,51,52*–53,54*–56*,57,59,61*]. In some cases, the relative stabilities of mutants [58,61*,66], have also been predicted. These methods have, however, been less successful either in correctly predicting the conformations of mutant or homologous structures based on a known 'parent', or in predicting the finer details of known structures. The variety of structural responses to core substitutions, particularly flexibility of the backbone, suggests that precise prediction of mutant protein stability and structure still presents an enormous challenge to computational methods. Substantial deviations both from wild-type backbone positions and from 'ideal' torsion angles are routinely observed in mutant structures. Therefore, modeling based on 'ideal' torsion angles [50°] or rigid backbones is likely to lead to incorrect structures or unreliable estimates of stability changes caused by mutations. Even if such adjustments are small, they can significantly influence interatomic distances and calculated energies. Furthermore, analysis of mutant structures suggests that both backbone shifts and changes in torsion angles can have similar energetic consequences and, therefore, cannot be evaluated independently. Although present prediction methods may be in error in excluding viable sequences and configurations, they may still be of use in more qualitative applications, such as in predicting the least perturbing interior substitutions [58,60], predicting the structures of homologous proteins [99,100*,101*], and suggesting core-residue combinations for packing de novo designed structures.

Do protein structures and stabilities need to be accurately predicted to effectively engineer them? A recent experiment by Hecht and coworkers [14**] suggests that de novo designed proteins can be obtained in the absence of strict design criteria [102]. They generated a library of potential four-helix bundles specified by a simple 'binary' pattern of polar and non-polar amino acids. Although previous approaches were based on fully defined sequences [13,15,16*], in this case, a potential 10⁴¹ combinations were specified using random assortments of five different non-polar amino acids and six different polar types. When expressed in Escherichia coli, 48 of 69 sequences (60%) yielded soluble products. Initial characterization of three polypeptides showed that they
were monomeric and highly helical, and two had stabilities approaching those of known proteins (3.7 kcal mol$^{-1}$ and 4.4 kcal mol$^{-1}$). Thus, a large fraction of all possible sequences folded into compact soluble structures. At least 3% were reasonably stable, suggesting that packing details are a very weak constraint in this system. Even so, a cautionary note is required: many of these bundles may be ‘molten globules’ or may lack well-defined three-dimensional structures. Characterization of native-like sequences obtained in this way should facilitate the formulation of rules for design of novel polypeptide structures.

**Conclusions**

The ‘plasticity’ of protein cores that is demonstrated by data from random mutagenesis, simplifies protein design by increasing the likelihood of identifying the alternative amino acid sequences that lead to folded functional proteins. At the same time, however, recent crystallographic studies have shown that both backbone and side-chain adjustments occur when different core-packing arrangements are accommodated in proteins. Thus, modeling methods, which have typically considered only side-chain adjustments, must now also account for backbone movements to accurately predict the energies and structures of mutated or designed proteins. This remains the challenge for the future.

**Acknowledgements**

This work was supported in part by National Institutes of Health Grants GM12989 and GM21967. The authors would like to thank Drs J Hurley and W Lim for helpful discussions and for providing access to unpublished manuscripts.

**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

15. This paper describes a combinatorial genetic approach to obtaining four-helix bundles. A degenerate gene library, which encodes polypeptides with the same pattern of residue polarity but different distinct sequences, is screened for products accumulating in soluble form in *Escherichia coli* that are judged to be ‘folded’. Preliminary physical characterization of three proteins shows that they are highly helical and compact. This paper will likely inspire many similar experiments with other folds.
19. A description of a designed lysine rich four-helix bundle that catalyzes the decarboxylation of oxaloacetate with a rate acceleration of 10^2–10^4. Data show that a relatively non-specific surface with appropriate functionality can significantly catalyze a reaction.
21. An authoritative review on designed four-helix bundles.
22. This paper describes simple rules for specifying different coiled-coil oligomers on the basis of sequences of residues at the helix interfaces.
24. A complete structural and thermodynamic analysis of T4 lysozyme mutants with one to four residues inserted into the beginning, middle, and ends of an α-helix. The pattern of hydrophobic residues in sequences adjacent to the insertion site are shown to influence the choice between ‘looping out’ or translocation of the adjacent residues toward the end of the helix.
26. Describes the structures of insertion mutants, in which either ‘looping out’ of inserted residues or translocation of neighboring helical residues occurs. When amino acids are inserted, the helix length is maintained, but the adjacent loops increase in size. See also [21,22].
28. See [22].

This paper and [21*] describe structural studies of insertion mutants of hemoglobin and staphylococcal nuclease, in which the inserted residue interrupts the local helical structure.


This paper proposes that helical propensity of a particular residue type is influenced by the amount of surface area buried upon folding.


A comprehensive structural and thermodynamic analysis of substitutions at a solvent-exposed site in the middle of a helix. It discusses the contributions of hydrophobicity and side-chain entropy to helix propensity and includes a comprehensive survey of torsion angles in highly refined X-ray structures.


See [26*].


This review and [25*] summarize recent thermodynamic and structural analyses of mutant proteins.


Review on structure and stability studies of mutant proteins.


See [34*].


This paper and [33*] describe recent studies using alanine mutagenesis to probe the role in stability of buried residues in myoglobin.


Proposes a statistical method for identifying interacting pairs of residues from the frequency of corresponding pairs of substitutions in a functional population of random multiple alanine mutants.


A complete thermodynamic and crystallographic analysis of site-directed core mutants that illustrates how accommodation of substitutions is affected by context.


40. Blaber M, Lindstrom JD, Gaisser N, Xu J, Heinz DW, Matthews BW: Energetic Cost and Structural Consequences of Burying a Hydroxyl Group Within the Core of a Protein Determined from Ala→Ser and Val→Thr Substitutions in T4 Lysozyme. Biochemistry 1993, 32:11363–11373.

Describes a structural and thermodynamic survey of the responses to isosteric polar replacements of core residues. Evidence from this study shows that buried polar groups always seem to satisfy their hydrogen bonds, if necessary by sequestration of a solvent molecule.


Describes the structure of a packing mutant, containing phenylalnine in place of a buried serine, that has enhanced stability and a modest increase in packing efficiency. Backbone adjustments and side-chain conformation changes are also observed.


A review of the packing problem, with emphasis on mutational and structural analysis. Computational methods and the non-additivity of multiple-core mutants are also discussed.


A comprehensive review of theoretical and experimental studies of geometric packing in proteins. A compendium of stability data for packing mutants is also included.


This paper analyzes side-chain torsion angles and their distributions in known structures, questioning the validity of the ‘average rotamer’ concept.


The relationship between side-chain torsion angles and backbone conformation is investigated. This analysis shows that rotamer matching is improved if backbone torsional information is included.


See [55*].


This paper and [54*,60] contain excellent discussions of the limitations of homology modeling as well as what can be done about them.


This paper describes a strategy to predict packing based on arrangements observed in three-dimensional structures. A rigid backbone is assumed. This procedure is fundamentally different from the rotamer approach to packing because it incorporates context information from the database.


Describes the first method to predict side-chain configuration that incorporates rigid helical motions in energy minimization of a triplex structure. These minimization steps are performed between side-chain enumeration steps. The method is used to correctly predict the orientations of three of four helices in myohemerythrin.


Presents qualitatively correct predictions of 1 repressor mutant stabilities (see also [66,79,80]) using a rigid backbone model and a new method to sample side-chain configurations.


This paper and [63,65] contain discussions of the accuracy and/or feasibility of current energy calculations on proteins.


Complete segment survey of the relative mutability of residues in gene V protein by segment combinatorial mutagenesis.


Describes segment combinatorial mutagenesis of three residues in an interhelical loop of the four-helix bundle protein cytochrome c662. Of the 31 clones analyzed, all folded into native-like structures, as judged by their heme-binding activity in vivo.


Describes the use of unnatural side chains to probe protein core packing and stability (see also [74]). A homologous series of alkyl side chains was introduced at the same position in thioesterase by disulfide exchange.


A study of thermodynamic and kinetic effects of single and double core mutants on protein folding.


Reports the structures of two packing mutants with enhanced stability and increased packing efficiency. Substitutions (valine to leucine or leucine) are sampled by side-chain optimizations in the more stable leucine variant in a strained conformation.


See [82*].

Structural and thermodynamic characterizations of eight repacking variants containing three to five substitutions (see also [81]). The structures show that much of the accommodation of the repacked cores is due to shifts in the backbone. In contrast, relatively few changes are observed in the rotational angles of the substituted side chains (see [94]).


See [90].


See [90].


This paper and [87,88,89,91] describe the structures of variants containing single-residue replacements in their cores.


A rationalization of the stability of cavity-creating mutants based on differences in excluded volume of the folded and unfolded state.


Presents crystal structures of packing mutants from homologous replacements of a loop in cytochrome c. Substitutions that replace the core-packing residues in the mutant loop with those of the wild type result in a structure that more closely resembles wild-type cytochrome c.


Reports the crystal structure of a stabilized triple mutant of λ repressor. Both an increase in packing density and preservation of wild-type torsion angles accompanied by backbone adjustments are observed (see also [82]). Torsion angles differ from the predicted values [66].


See [101].


This paper and [99,100] review the applications of homology modeling to the prediction of structures within protein families.


EP Baldwin and BW Matthews, Institute of Molecular Biology, Howard Hughes Medical Institute and Department of Physics, University of Oregon, Eugene, Oregon 97403, USA.