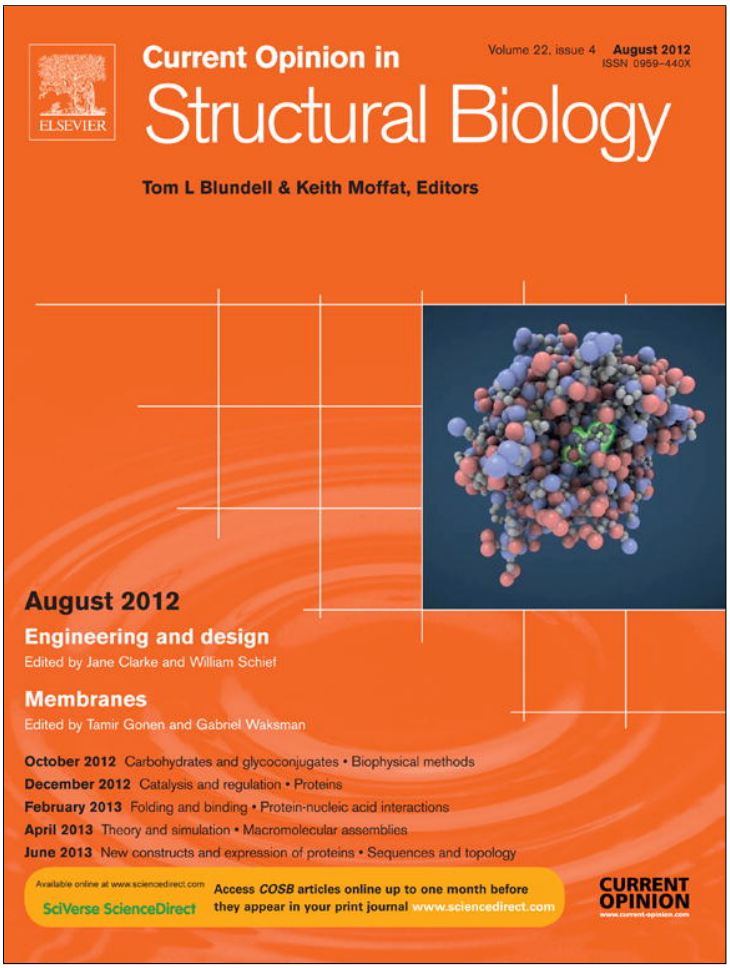


Available online at www.sciencedirect.com

SciVerse ScienceDirect

Current Opinion in Structural Biology

Designing proteins from simple motifs: opportunities in Top-Down Symmetric Deconstruction

Michael Blaber and Jihun Lee[1]

The purpose of this review is to describe the development of 'top-down' approaches to protein design. It will be argued that a diverse number of studies over the past decade, involving many investigators, and focused upon elucidating the role of symmetry in protein evolution and design, are converging into a novel top-down approach to protein design. Top-down design methodologies have successfully produced comparatively simple polypeptide 'building blocks' (typically comprising 40–60 amino acids) useful in generating complex protein architecture, and have produced compelling data in support of macro-evolutionary pathways of protein structure. Furthermore, a distillation of the experimental approaches utilized in such studies suggests the potential for method formalism, one that may accelerate future success in this field.

Address
Department of Biomedical Sciences, College of Medicine, Florida State University, Tallahassee, FL 32306-4300, United States

Corresponding author: Blaber, Michael (michael.blaber@med.fsu.edu)

[1] Current address: Celltrion Inc., 13-1 Songdo-dong, Yeonsu-gu, Incheon City 406-840, Republic of Korea.

Current Opinion in Structural Biology 2012, **22**:442–450

This review comes from a themed issue on **Engineering and design**

Edited by **Jane Clarke** and **William Schief**

For a complete overview see the Issue and the Editorial

Available online 20th June 2012

0959-440X/$ – see front matter,

http://dx.doi.org/10.1016/j.sbi.2012.05.008

'Bottom-up' and 'top-down' protein design and structural symmetry

Much of the *de novo* protein design effort in the 1990s focused upon a 'bottom-up' hierarchical approach based upon fundamental principles of non-covalent interactions and protein secondary structure [1,2]. This classic approach typically proceeds by identification of desired target architecture, design of secondary structure elements by selection of amino acids with favorable propensities, design of an appropriate hydrophobic patterning consistent with the target architecture (with the goal of achieving efficient hydrophobic core packing) [3], addition of linker regions (e.g. reverse turns) to make the desired secondary structure connectivity, and adding charged partners, compatible H-bonding groups, or disulfide bonds to stabilize structure-specific intramolecular interactions. Depending on the design parameters, specific functional residues may also be incorporated. Computational energy calculations, modeling and visualization are intrinsic to the entire design process. Once the design is finalized, the target polypeptide is expressed from a synthetic gene, and the purified protein is characterized to confirm fitness of the design principles. Helical architecture is less complicated to design than β-sheet architecture due to the complexity of inter-strand interactions in the latter [4]; and notable success has been achieved in the *de novo* design of all-α proteins [5–7], although all-β proteins have proven more difficult. A not uncommon result of bottom-up design is a '*de novo* molten globule' [2] exhibiting unsatisfactory folding cooperativity, thermostability or solubility; such problems are typically improved by subsequent redesign or mutagenesis. A significant breakthrough in computational bottom-up design was achieved by Baker and coworkers in an approach involving alternating the search for a low energy set of side chains for a defined (i.e. rigid) backbone, and subsequently, searching for a low energy backbone solution for a defined set of side chains [8]. Thus, the general aspects of the desired target architecture were initially defined, and repeated iterations of the alternating side chain/backbone computational search converged upon the detailed design solution. This approach yielded a thermostable, cooperatively folding polypeptide with an architecture that fit the initial design features, and was also a novel architecture not previously described in the structural databank.

A number of investigators have been interested in symmetric protein architecture, its role in protein evolution via gene duplication and fusion, and exploitation in protein design from a more top-down approach. In 'fragmentation' studies symmetric architectures such as the $(\beta\alpha)_8$-barrel or β-propeller have been fragmented into subdomains to determine whether the resulting peptides fold independently or can assemble as oligomers to reconstitute the parent architecture [9•,10–13]. 'Consensus design' is a top-down approach that involves a comparison of naturally occurring sequences of symmetric protein architecture to identify the most conserved and therefore presumably the most structurally important amino acids in the repeating motif [14,15]. Another area of investigation is to understand the practical limits of symmetry in protein folding and design. While detectable tertiary structure symmetry is a common feature in proteins, it is substantially diminished at the level of

the primary structure. Thus, although gene duplication and fusion is a compelling evolutionary mechanism, a convergent evolutionary process is also plausible [16]. Additionally, exact primary structure symmetry has been postulated to be a potential source of folding frustration [17]; furthermore, exact primary structure symmetry represents a substantial reduction in sequence complexity and may potentially violate some minimum complexity necessary for efficient folding [18]. Thus the limits of primary structure symmetry for efficient protein folding are unclear, but are essential to elucidate from the standpoint of protein design and evolutionary pathways [13,19–22].

A tantalizing possibility emerging from symmetric protein design studies is that an appropriately designed purely symmetric primary structure represents a type of idealized structural solution, yielding extreme thermostability and rigidity [15,23•,24]. In this perspective, specific function and localized dynamics are accommodated as a *structural defect* (i.e. an asymmetric feature within the background of a purely symmetric architecture, an extension of the 'function/stability' tradeoff hypothesis [25,26]). Another possible attribute of purely symmetric proteins is that they have a redundant folding pathway, and can therefore tolerate potentially deleterious mutational change to individual folding nuclei and still fold efficiently, thereby providing transient evolutionary robustness. These postulated characteristics of idealized symmetric protein architecture suggest an exceptional utility in protein engineering and design, placing increasing importance upon successful design strategies of simple peptide building blocks for such architecture.

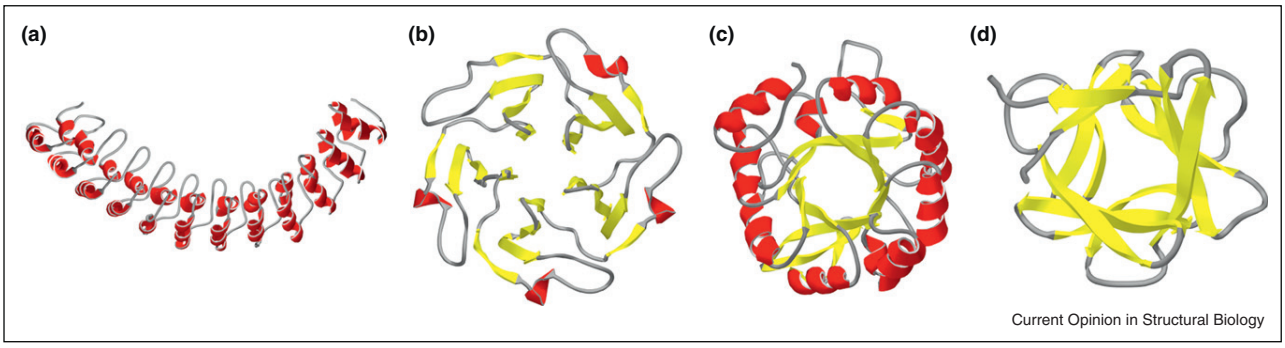
Symmetric target architecture and top-down strategies in protein design

Four broad categories of symmetric protein architecture, distinguished by their fundamentally different characteristics of key hydrophobic core/inter-domain packing interactions, have been subjected to top-down design studies. For purposes of comparison in this review, these categories include: linear repeat, circularly closed repeat, limited core circularly closed repeat and single domain globular symmetric proteins (Figure 1). Linear repeat proteins are non-globular proteins comprised of successive homologous structural units of 20–40 amino acids that stack to form elongated structures with no conserved number of repeats or cooperative contacts beyond adjacent units. Because of these unique structural features they will not be discussed in detail in this review; however, interested readers are directed to the work of Grutter, Pluckthun, Barrick, Peng and coworkers for examples of work in the area of repeat protein design [27–30].

'Circularly closed repeat' architecture

Similar to linear repeat proteins, the circularly closed repeat architecture is comprised of multiple structural repeats with adjacent packing interactions (but no cooperative central core packing group), and a potentially variable number of overall repeats. Unlike linear repeat proteins, however, circularly closed repeat architecture exhibits an overall 'closed' fold having adjacent N-termini and C-termini. One representative and extensively studied example of circularly closed repeat architecture is the 'β-propeller' characterized by a repeating structural motif comprised of four anti-parallel β-strands referred to as a 'blade'. The overall β-propeller architecture is typically comprised of four to eight blades arranged toroidally around a hollow central axis (Figure 1B). β-Propeller proteins exhibit extreme diversity in sequence, function and phylogenetic origin and the sequence similarity between blades of the same propeller, and between blades of different propellers, can vary considerably.

Figure 1



Examples of symmetric proteins utilized in top-down design studies (note: images not to scale). **(a)** Ankyrin repeat (D34 region of human Ankyrin-R, RCSB accession 1N11), a linear-repeat architecture; **(b)** five-bladed β-propeller (Tachylectin-2, RCSB accession 1TL2), a circularly closed repeat architecture; **(c)** $(\beta\alpha)_8$-barrel (HisF, RCSB accession 1THF), a limited core circularly closed architecture; **(d)** β-trefoil (FGF-1, RCSB accession 2AFG), a single domain globular symmetric architecture. Aspects of structural symmetry (as well as asymmetry) can be readily appreciated from simple ribbon diagrams.

Paoli and coworkers described formalism for top-down design of a seven-bladed β-propeller fold with the goal of developing a 40-amino acid WD repeat-based blade [14]. This formalism defined a stepwise process that proceeded according to identifying a consensus sequence among a broad category of WD repeat sequences (with added consideration for compatible local interactions), design of turn sequences, and optimization of hydrophobic packing interactions between adjacent blades. Subsequently, concatenates of 4–10 designed WD domains were expressed and characterized. This approach yielded molten globule proteins, and the authors hypothesized that divergence from exact symmetry may be essential to improve blade packing interactions. Other than the final WD design, no intermediate forms in the top-down design were evaluated (Table 1).

In a top-down design of the five-bladed β-propeller protein tachylectin-2 [13] Tawfik and coworkers utilized a design strategy comprising: first, identification of a structure-based definition for the repeating motif (selecting a domain-swapped definition); second, sequence analysis to identify which subdomain contains the greatest percentage of consensus residues for a comparison among all five subdomains; and third, construction of a pentameric repeat of the 47 amino acid subdomain. Using this approach the initial pentameric polypeptide proved to be an insoluble aggregate. Thus, a subsequent step of random mutagenesis of N-termini and C-termini regions that participate in inter-domain packing interactions was performed, and resulted in several soluble mutants with apparent target architecture. The resulting β-propeller proteins subsequently had a high degree of symmetry but

Table 1

Examples of top-down protein design and resultant peptide building blocks for symmetric target architecture.

Target architecture	Proxy	Method	Design cycle granularity	Resulting building block
β-Propeller (5 blade) [11,48•]	Tachylectin-2 • 236 amino acids • 49% internal symmetry	Fragmentation, random mutagenesis of interface regions	2	• 100 amino acids • Folds as pentamer repeat forming two 5-blade β-propellers • Lectin functionality • RCSB accession 3KIH
β-Propeller (5 blade) [13]	Tachylectin-2 • 236 amino acids • 49% internal symmetry	Fragmentation, random mutagenesis of interface regions	3	• 47 amino acids • Limited asymmetry • Folds as pentamer repeat • Reduced thermostability • Lectin functionality
β-Propeller (7 blade) [14]	G protein β-subunit • 295 amino acids • Limited consensus symmetry	Consensus design	1	• 40 amino acids • Molten globule heptamer repeat
$(\beta\alpha)_8$-Barrel [31,32,49]	HisF • 253 amino acids • 25% half-barrel symmetry • Thermophile stability	Fragmentation, mutagenesis of interface regions	3	• 120 amino acids • Folds as dimer repeat • Highly cooperative folding • Reduced thermostability • RCSB accession 3OG3
$(\beta\alpha)_8$-Barrel [33]	HisF • 253 amino acids • 15% quarter-barrel symmetry • Thermophile stability	Fragmentation	2	• 74 amino acids • Evidence of $(\beta\alpha)_4$-barrel and $(\beta\alpha)_8$-barrel oligomeric assembly • Reduced thermostability
$(\beta\alpha)_8$-Barrel [12]	TrpF • 197 amino acids • 3% internal half-barrel symmetry • Mesophile stability	Fragmentation, circular permutation	1	• 93 amino acids • Folds as homodimer and as dimer repeat • Thermophile
β-Trefoil [34•,41••]	FGF-1 • 140 amino acids • 3% internal symmetry • Mesophile stability	Top-Down Symmetric Deconstruction	14	• 42 amino acids • Folds as trimer repeat, dimer repeat, and homo-trimer • Soluble hyper-thermophile • Benign functionality • RCSB accession 3O4D, 3OGF, 3OL0
β-Trefoil [15]	Ricin • 141 amino acids • 55% internal symmetry • Thermophile stability	Consensus design	1	• 47 amino acids • Folds as trimer repeat • Lectin functionality • Soluble thermophile • RCSB accession 3PG0

retained some degree of asymmetry. The authors postulated that asymmetry in the repeating domains may be essential to achieve efficient folding, raising the question of whether pure primary symmetry in multi-bladed β-propeller design is compatible with efficient folding.

'Limited core circularly closed' architecture

The $(\beta\alpha)_8$-barrel (TIM barrel) is the most common protein fold and its structure is composed of eight modular repeats comprising a central β-strand, reverse-turn, outer α-helix, and another reverse-turn (Figure 1C). Adjacent subdomain packing interactions comprise the majority of buried hydrophobic interactions; however, in comparison to the β-propeller (which has no central hydrophobic core), each repeating motif in the $(\beta\alpha)_8$-barrel contributes one residue to a small central packing group (overall comprising ~3% of total residues). In the top-down design of a $(\beta\alpha)_8$-barrel architecture Sterner and coworkers began with a $(\beta\alpha)_4$-barrel motif derived from the C-terminus half of HisF. This motif was duplicated to create an intact $(\beta\alpha)_8$-barrel which was then subjected to mutagenesis to improve the inter-domain packing interactions [31,32]. This yielded a folded and soluble $(\beta\alpha)_8$-barrel target architecture with limited asymmetry between the two $(\beta\alpha)_4$-barrel halves. Working with a different $(\beta\alpha)_8$-barrel protein (*N*-(5′-phosphoribosyl)anthranilate isomerase) and a different approach to fragmentation (involving evaluation of circular permutation half-barrel definitions) Akanuma and coworkers successfully generated a half-barrel motif that was able to fold independently as a stable $(\beta\alpha)_4$-barrel and also dimerize to generate an intact and stable $(\beta\alpha)_8$-barrel architecture [12]. These results provide strong experimental support for the gene duplication and fusion hypothesis of evolution for the $(\beta\alpha)_8$-barrel architecture, and also identify a useful $(\beta\alpha)_4$-barrel building block. Recently, a further evolutionary pathway involving a putative ancestral $(\beta\alpha)_2$-barrel motif to generate an intact $(\beta\alpha)_8$-barrel has been studied by Sterner and coworkers [33]. These results support a two-step gene duplication and fusion process in the evolution of $(\beta\alpha)_8$-barrel architecture, and identify a useful $(\beta\alpha)_2$-barrel polypeptide building block comprising ~60 amino acids.

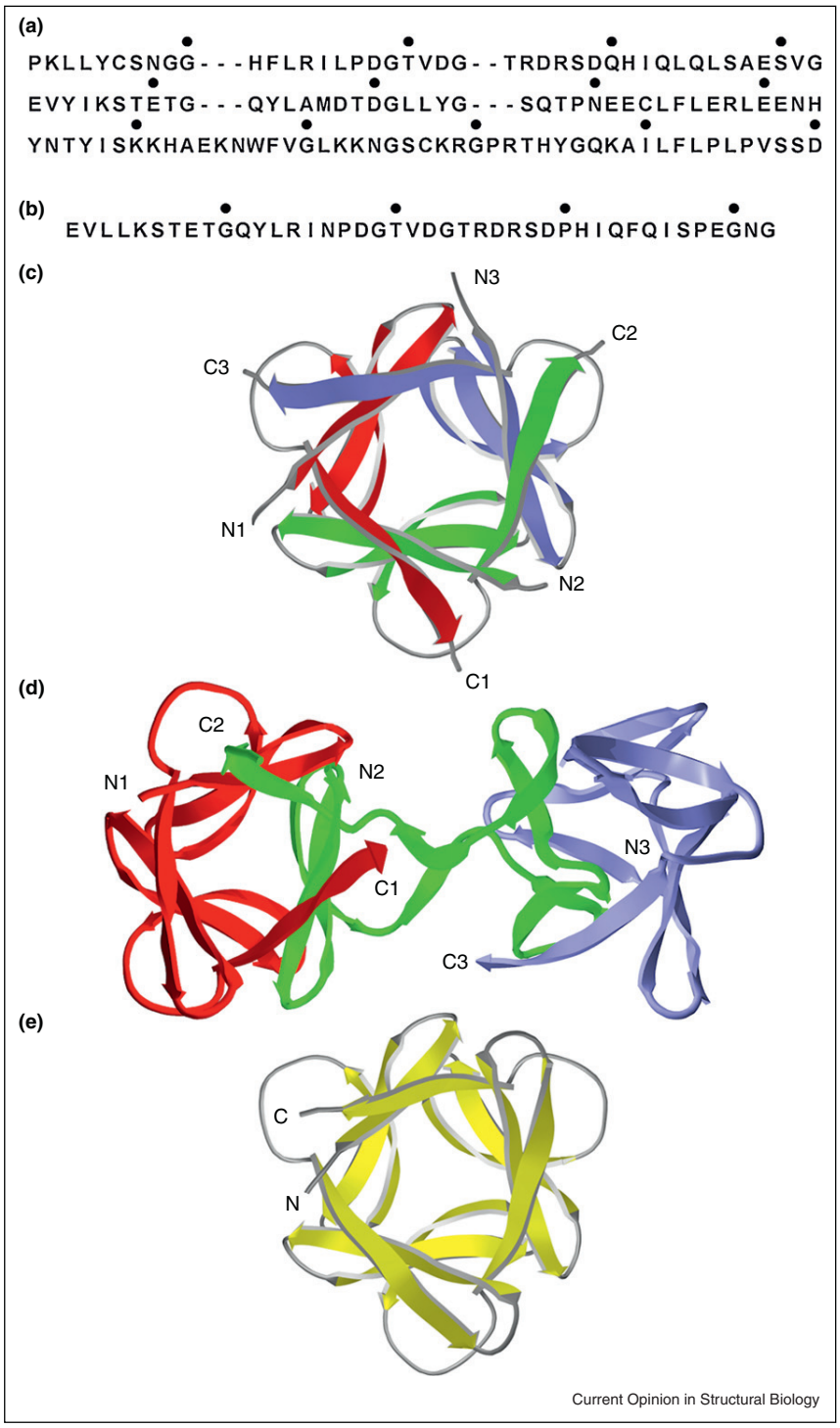
'Single domain globular symmetric' architecture

Single domain globular symmetric proteins have a central hydrophobic core that comprises the principle cooperative hydrophobic packing group in the structure. The β-trefoil is a common protein fold exhibiting threefold structural symmetry and is an example of single domain globular symmetric architecture (Figure 1D). Each repeating domain (known as a 'trefoil-fold') is 40–50 amino acids in length and comprises four β-strands in a domain-swapped arrangement of two anti-parallel β-hairpins. Each domain contributes five to six hydrophobic side chains to the central core; thus, in contrast to the $(\beta\alpha)_8$-barrel, the cooperative central hydrophobic core of the β-trefoil defines the primary core packing group and comprises ~13% of total residues. Some β-trefoil proteins (e.g. ricins, actin-bundling proteins) exhibit substantial primary structure identity between the repeating domains, while others (e.g. fibroblast growth factor (FGF-1)) exhibit identities marginally above random.

Starting with the highly asymmetric and mesophile stability FGF-1 protein (Figure 2) our lab produced a purely symmetric β-trefoil protein ('Symfoil'), via a series of 14 sequential symmetric constraint mutations upon core, reverse-turn, and β-strand secondary structure, respectively, that is soluble, cooperatively folding and thermostable [34•]. Fragmentation of a further stabilized Symfoil variant into its 42 amino acid monomer motif ('Monofoil') produces a short peptide that spontaneously folds as a homo-trimer to yield the β-trefoil architecture. Additionally, expression of a dimer repeat of this motif ('Difoil') yields an 84 amino acid polypeptide that folds as a homo-trimer yielding two intact (and interconnected) β-trefoil folds. These results support one of the two competing hypotheses for the evolution of the β-trefoil architecture (*i.e.* the 'conserved architecture' model) via gene duplication and fusion processes [35,36]. Comprehensive analysis of the purely symmetric Symfoil protein [24] shows a remarkable broad resistance to denaturation and a high structural rigidity, supporting the hypothesis that exact symmetry can be an idealized design solution [23•] (Figure 3).

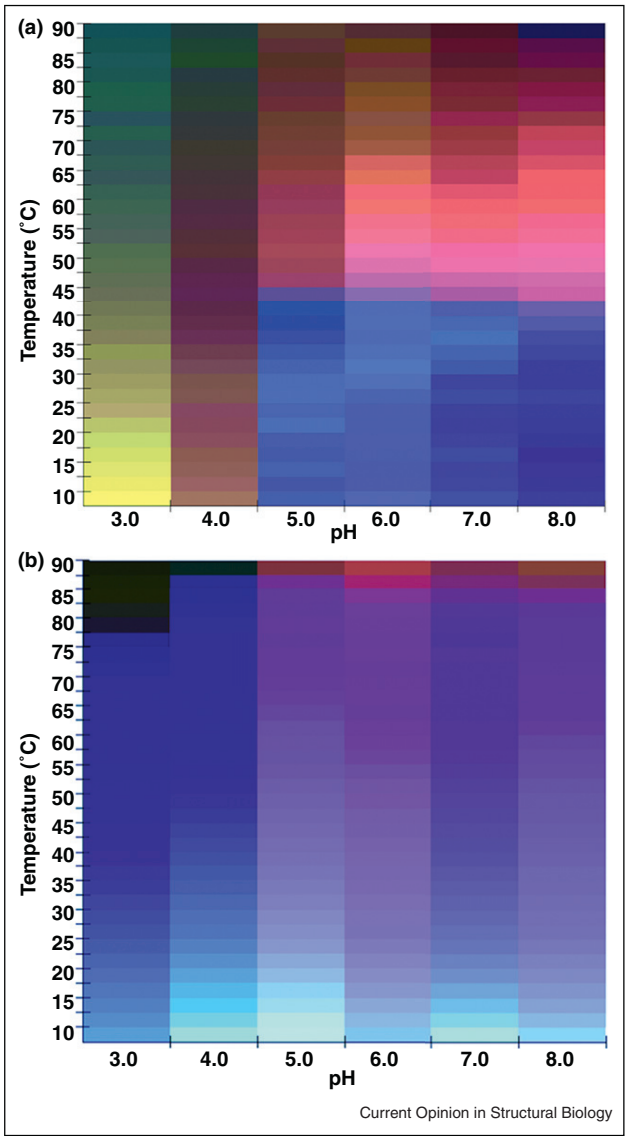
Meiering and coworkers pursued consensus formalism in the top-down design of the β-trefoil-fold starting with a thermophile member of the ricin family having 55% identity between the three repeating trefoil-fold subdomains [15]. The design formalism proceeded with identification of consensus residues among the three trefoil-fold subdomains, followed by consensus analysis of highly homologous sequences to partially fill in remaining asymmetric regions, and finally computational design [37] to complete the symmetric constraint. This formalism did not separate the design approach by particular secondary structure or packing interactions; however, a symmetric core packing group was part of the consensus sequence in step 1; additionally, step 2 involved the introduction of a symmetric constraint primarily upon two turn positions, and step 3 involved introducing a symmetric constraint primarily upon three different β-strands. The resulting 47 amino acid building blocks ('Onefoil') successfully folds into the target β-trefoil-fold as a trimer concatenate ('Threefoil'); however, attempts to fold the Onefoil peptide building block as a homo-trimer proved unsuccessful. Other than the proxy and final design, no intermediate mutant forms were characterized. While the overall primary structure identity between Symfoil and Threefoil proteins is limited, the core-packing arrangements are essentially identical (RCSB depositions 3PG0, 3Q7Y and 3O4D), suggesting that solutions for an efficient core-packing arrangement in the β-trefoil architecture may be restricted.

Figure 2



(a) 140 amino acid sequence of FGF-1 utilized as the proxy in Top-Down Symmetric Deconstruction. The sequence is aligned by the repeating trefoil-fold subdomains (dots are used to indicate every 10th amino acid). **(b)** The 42 amino acid building block (Monofoil-4P) for the β-trefoil-fold derived from the FGF-1 proxy [34•,41••]. **(c)** X-ray crystal structure of Monofoil-4P which spontaneously folds as a homo-trimer to form the β-trefoil target architecture (RCSB accession 3OL0). Panel D: X-ray crystal structure of a dimeric concatenated of the Monofoil-4P building block (Difoil-4P) which spontaneously folds as a homo-trimer to form two intact instances of the β-trefoil target architecture (RCSB accession 3OGF). Panel E: X-ray crystal structure of a trimeric concatenate of the Monofoil-4P building block (Symfoil-4P) which spontaneously folds into a hyper-thermophile β-trefoil target architecture (RCSB accession 3O4D).

Figure 3



(a) Empirical phase diagram (EPD) [24] of FGF-1 a proxy in the Top-Down Symmetric Deconstruction of the β-trefoil-fold [34•]. The blue region in the EPD indicates the natively folded regime as characterized by a comprehensive battery of analytical methodologies, including circular dichroism, intrinsic and extrinsic fluorescence spectroscopy, static light scattering, and ANS dye binding. **(b)** the EPD of a resultant purely symmetric protein design Symfoil-4P. The symmetric solution for the β-trefoil-fold is compatible with extreme thermostability.

Formalism for Top-Down Symmetric Deconstruction

''Blacksmith, I set ye a task. Take these harpoons and lances. Melt them down. Forge me new weapons that will strike deep and hold fast.''—Ahab, Moby Dick

Top-down design studies have been self-described as a process of 'reverse engineering' [38] or 'reverse approach' [11], evolutionary or structural 'reconstruction' [15,31], 'fragmentation analysis' [11,39], structural 'dissection' [10], and so on. Despite these seemingly diverse descriptions, substantially overlapping conclusions regarding protein evolution and design are emerging. For example, in fragmentation studies designed to probe evolutionary pathways, the value to protein design of the resultant peptide building block has been noted [40]; conversely, in studies designed to probe effects of enhanced symmetry in proteins, support for specific evolutionary models has emerged [41••]; additionally, consensus design studies have made note of both evolutionary and protein design implications [15]. Top-down design studies can produce simple optimized peptide motifs that can be utilized in the design of symmetric protein architecture; correspondingly, we have proposed the term 'Top-Down Symmetric Deconstruction' [34•] as an apt descriptor of this approach to protein design.

There are several powerful advantages associated with the top-down approach to protein design, foremost among these is that the top-down process *begins* with a foldable polypeptide; thus, as long as the design cycle maintains the polypeptide within foldable sequence space the design will converge upon a solution. A second advantage is that failure points in the design cycle (i.e. mutational changes that move the polypeptide out of foldable sequence space) can be readily identified and corrected. However, this latter advantage depends upon *granularity* of the design cycle. In the extreme (although common) case where only the final mutant design (involving numerous mutational changes) is evaluated (i.e. a granularity of 1), the result is often a non-foldable polypeptide requiring secondary mutations to move back into foldable sequence space (and in this case the basis for the lack of folding is unclear since many mutations were simultaneously combined). In contrast, although increasing the overall effort, intermediate forms involving few mutations can more readily identify specific design problems. Thus, granularity of the design cycle in top-down design appears critical when engineering knowledge is incomplete or imprecise.

The proxy

Top-down design begins with the selection of an appropriate *proxy* (i.e. foldable example) of the target architecture. The proxy has typically been chosen for having the greatest primary and tertiary structure symmetry [15,42], or being a primitive form (i.e. as unchanged from a presumed formative duplication/fusion event as possible) [11], or as the smallest/simplest representative of the target architecture [43]. A high degree of symmetry in the proxy reduces the number of mutational changes to achieve a symmetric solution, reducing the required granularity in the design cycle. Additionally, a thermophile proxy increases the ability to accommodate a potentially greater number of

mutations and remain within foldable sequence space. Other top-down design studies have started with computer-generated consensus-based proxies, the folding and stability properties of which were undetermined [14]. Lack of specific functionality may be an additional design goal, yielding a peptide building block with a 'blank slate' as regards function; alternatively, a generic functionality might be a key aspect of the design of a more specific-purpose scaffold (e.g. non-specific lectin affinity). If functionality is dependent upon correct folding, it can be utilized as a screen or selection to keep the polypeptide within foldable sequence space during the design cycle [13]; however, if functionality requires asymmetric attributes of structure or dynamics, it may preclude identification of a symmetric solution.

Top-Down Symmetric Deconstruction formalism

Top-Down Symmetric Deconstruction of β-trefoil target architecture was divided into discrete 'transforms' (using the nomenclature of the conceptually related Retrosynthetic Analysis of Corey and Cheng [44]) that sequentially targeted the hydrophobic core, turn secondary structure, and β-strand secondary structure, respectively [34•]. In contrast to a number of other top-down approaches, efficient hydrophobic packing with a symmetric constraint was the initial focus of design. After successful symmetric design, but before fragmentation to yield the repeating peptide motif, a final transform to enhance thermostability was included. In contrast to other studies, experimental validation was performed within each transform to ensure that stability and especially *folding cooperativity* was enhanced or maintained throughout the entire design cycle. It was observed that while thermostability was a highly modifiable property, folding cooperativity could, at best, only be maintained (and was easily diminished by mutation). Thus, *folding cooperativity is a key parameter in the design cycle*.

Granularity

Top-down protein design studies begin with a foldable proxy, and utilize some type of stepwise formalism in their symmetric deconstruction; however, very few studies have characterized intermediate forms in the overall design cycle (Table 1). Thus, a typical result is that the design moves out of foldable sequence space and 'backtracking' mutagenesis (typically targeting hydrophobic packing interactions) is performed to move the design back into foldable sequence space [10,13,31]. Recent unpublished ϕ-value studies of FGF-1 in our laboratory indicate that key residues contributing to the folding transition state are asymmetrically distributed within the FGF-1 sequence. Furthermore, these key positions for efficient folding were effectively retained in our Top-Down Symmetric Deconstruction; thus, the design cycle never deviated from foldable sequence space and key residues contributing to the folding transition state were retained. Comprehensive circularly permuted fragmentation studies of *N*-(5′-phosphoribosyl)anthranilate isomerase (a $(\beta\alpha)_8$-barrel protein) also demonstrated an asymmetric distribution of critical folding nuclei, such that only certain $(\beta\alpha)_4$-barrel definitions folded independently [43]; additionally, only certain $(\beta\alpha)_2$ fragment definitions of HisF are able to oligomerize to recapitulate the $(\beta\alpha)_8$-barrel target architecture [33]. Critical folding nuclei in symmetric protein architecture therefore appear to be non-redundantly distributed; thus, random fragmentation of the wild-type proxy may be deleterious to foldability. Additionally, depending on the set of sequences utilized in consensus approaches to symmetric design, residues critical to formation of the folding transition state may similarly be lost. Computational approaches to Top-Down Symmetric Deconstruction may therefore achieve greater success if key residues contributing to the folding transition state can be identified and retained in the design cycle.

Conclusions

One advantage of symmetric protein architecture in protein design is the potential for a redundant folding pathway that can permit greater freedom to introduce specific function in protein design. Top-down design offers a novel and potentially efficient means to identify simple peptide building blocks with which to construct symmetric target architecture. Diverse top-down design approaches may be unified through a shared formalism that begins with a foldable proxy of the target architecture, and proceeds through a design cycle that remains within foldable sequence space. A design cycle is proposed that proceeds by: proxy selection that maximizes initial symmetry; design of symmetric hydrophobic core/inter-domain packing, followed by symmetric constraint mutations within fundamental secondary structure elements, with an emphasis upon retaining residues critical to efficient folding. Although different proteins may share a common architecture, their folding pathways can differ substantially (e.g. the β-trefoil proteins interleukin-1β [45], FGF-1 [46] and hisactophilin [47]). Thus, for a given architecture the key folding transition state residues are likely dependent upon proxy selection. This suggests a folding pathway redundancy inherent to symmetric primary structure, a redundancy that is subsequently lost due to sequence divergence among the repeating subdomains. ϕ-Value analysis, initial fragmentation studies, or improvements in molecular dynamics simulations may enable identification of such residues. Top-Down Symmetric Deconstruction taken to completion yields a purely symmetric primary structure; subsequent fragmentation of this solution can yield a simple 40–60 amino acid peptide building block useful in design of the symmetric target architecture. Design cycle granularity, although time consuming, assists with success; however, improvements in computation (especially identification of residues key to formation of the folding

transition state) should reduce the required experimental granularity considerably.

References and recommended reading

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

- • of special interest
- •• of outstanding interest

1. Bryson JW, Betz SF, Lu HS, Suich DJ, Zhou HX, O'Neil KT, DeGrado WF: **Protein design: a hierarchic approach**. *Science* 1995, **270**:935-941.

2. Desjarlais JR, Handel TM: **New strategies in protein design**. *Curr Opin Biotechnol* 1995, **6**:460-466.

3. Kamtekar S, Schiffer JM, Xiong H, Babik JM, Hecht MH: **Protein design by binary patterning of polar and nonpolar amino acids**. *Science* 1993, **262**:1680-1685.

4. Hecht MH: **De novo design of beta-sheet proteins**. *Proc Natl Acad Sci U S A* 1994, **91**:8729-8730.

5. Regan L, DeGrado WF: **Characterization of a helical protein designed from first principles**. *Science* 1988, **241**:976-978.

6. Hecht MH, Richardson JS, Richardson DC, Ogden RC: **De novo design, expression, and characterization of Felix: a four-helix bundle protein of native-like sequence**. *Science* 1990, **249**:884-891.

7. Schafmeister CE, LaPorte SL, Miercke LJW, Stroud RM: **A designed four helix bundle protein with native-like structure**. *Nat Struct Biol* 1997, **4**:1039-1046.

8. Kuhlman B, Dantas G, Ireton GC, Varani G, Stoddard BL, Baker D: **Design of a novel globular protein fold with atomic-level accuracy**. *Science* 2003, **302**:1364-1368.

9. • Lang D, Thoma R, Henn-Sax M, Sterner R, Wilmanns M: **Structural evidence for evolution of the beta/alpha barrel scaffold by gene duplication and fusion**. *Science* 2000, **289**:1546-1550.

One of the more compelling structure-based arguments for a specific evolutionary mechanism of gene duplication and fusion in the emergence of complex symmetric architecture.

10. Hocker B, Beismann-Driemeyer S, Hettwer S, Lustig A, Sterner R: **Dissection of a $(\beta\alpha)_8$-barrel enzyme into two folded halves**. *Nat Struct Biol* 2001, **8**:32-36.

11. Yadid I, Tawfik DS: **Reconstruction of functional β-propeller lectins via homo-oligomeric assembly of shorter fragments**. *J Mol Biol* 2007, **365**:10-17.

12. Akanuma S, Yamagishi A: **Experimental evidence for the existence of a stable half-barrel subdomain in the $(\beta/\alpha)_8$-barrel fold**. *J Mol Biol* 2008, **382**:458-466.

13. Yadid I, Tawfik DS: **Functional β-propeller lectins by tandem duplications of repetitive units**. *Protein Eng Des Sel* 2011, **24**:185-195.

14. Nikkhah M, Jawad-Alami Z, Demydchuk M, Ribbons D, Paoli M: **Engineering of β-propellor protein scaffolds by multiple gene duplication and fusion of an idealized WD repeat**. *Biomol Eng* 2006, **23**:185-194.

15. Broom A, Doxey AC, Lobsanov YD, Berthin LG, Rose DR, Howell PL, McConkey BJ, Meiering EM: **Modular evolution and the origins of symmetry: reconstruction of a three-fold symmetric globular protein**. *Structure* 2012, **20**:1-11.

16. Wolynes PG: **Symmetry and the energy landscapes of biomolecules**. *Proc Natl Acad Sci U S A* 1996, **93**:14249-14255.

17. Wright CF, Teichmann SA, Clarke J, Dobson CM: **The importance of sequence diversity in the aggregation and evolution of proteins**. *Nature* 2005, **438**:878-881.

18. Romero P, Obradovic Z, Li X, Garner EC, Brown CJ, Dunker AK: **Sequence complexity of disordered proteins**. *Proteins* 2001, **42**:38-48.

19. Houbrechts A, Moreau B, Abagyan R, Mainfroid V, Preaux G, Lamproye A, Poncin A, Goormaghtigh E, Ruysschaert JM, Martial JA: **Second-generation octarellins: two new *de novo* (beta/alpha)8 polypeptides designed for investigating the influence of beta-residue packing on the alpha/beta-barrel structure stability**. *Protein Eng* 1995, **8**:249-259.

20. Brych SR, Blaber SI, Logan TM, Blaber M: **Structure and stability effects of mutations designed to increase the primary sequence symmetry within the core region of a β-trefoil**. *Protein Sci* 2001, **10**:2587-2599.

21. Heidary DK, Jennings PA: **Three topologically equivalent core residues affect the transition state ensemble in a protein folding reaction**. *J Mol Biol* 2002, **316**:789-798.

22. Brych SR, Kim J, Logan TM, Blaber M: **Accommodation of a highly symmetric core within a symmetric protein superfold**. *Protein Sci* 2003, **12**:2704-2718.

23. Brych SR, Dubey VK, Bienkiewicz E, Lee J, Logan TM, Blaber M:
• **Symmetric primary and tertiary structure mutations within a symmetric superfold: a solution, not a constraint, to achieve a foldable polypeptide**. *J Mol Biol* 2004, **344**:769-780.

Culmination of an early series of top-down design studies of the protein core. One of the first high-granularity design cycle studies of symmetric deconstruction, and one of the first to present experimental evidence that structural symmetry is compatible with enhanced protein thermostability and foldability.

24. Alsenaidy MA, Wang T, Kim JH, Joshi SB, Lee J, Blaber M, Volkin DB, Middaugh CR: **An empirical phase diagram approach to investigate conformational stability of 'second-generation' functional mutants of acidic fibroblast growth factor (FGF-1)**. *Protein Sci* 2012, **21**:418-432.

25. Beadle BM, Shoichet BK: **Structural basis of stability—function tradeoffs in enzymes**. *J Mol Biol* 2002, **321**:285-296.

26. Tokuriki N, Stricher F, Serrano L, Tawfik DS: **How protein stability and new functions trade off**. *PLoS Comput Biol* 2008, **4**:e1000002.

27. Mosavi LK, Minor DL Jr, Peng ZY: **Consensus-derived structural determinants of the ankyrin repeat motif**. *Proc Natl Acad Sci U S A* 2002, **99**:16029-16034.

28. Binz HK, Stumpp MT, Forrer P, Amstutz P, Pluckthun A: **Designing repeat proteins: well-expressed, soluble and stable proteins from combinatorial libraries of consensus ankyrin repeat proteins**. *J Mol Biol* 2003, **332**:489-503.

29. Tripp KW, Barrick D: **Rerouting the folding pathway of the notch ankyrin domain by reshaping the energy landscape**. *J Am Chem Soc* 2008, **130**:5681-5688.

30. Kramer MA, Wetzel SK, Pluckthun A, Mittl PR, Grutter MG: **Structural determinants for improved stability of designed ankyrin repeat proteins with a redesigned C-capping module**. *J Mol Biol* 2010, **404**:381-391.

31. Hocker B, Claren J, Sterner R: **Mimicking enzyme evolution by generating new (beta-alpha)8-barrels from (beta-alpha)4-half-barrels**. *Proc Natl Acad Sci U S A* 2004, **101**:16448-16453.

32. Seitz T, Bocola M, Claren J, Sterner R: **Stabilization of a (beta-alpha)8-barrel protein designed from identical half barrels**. *J Mol Biol* 2007, **372**:114-129.

33. Richter M, Bosnali M, Carstensen L, Seitz T, Durchschlag H, Blanquart S, Merkl R, Sterner R: **Computational and experimental evidence for the evolution of a $(\beta\alpha)_8$-barrel protein from an ancestral quarter-barrel stabilized by disulfide bonds**. *J Mol Biol* 2010, **398**:763-773.

34. Lee J, Blaber SI, Dubey VK, Blaber M: **A polypeptide 'building**
• **block' for the β-trefoil fold identified by 'Top-Down Symmetric Deconstruction'**. *J Mol Biol* 2011, **407**:744-763.

A top-down design approach to protein design that provides one of the first attempts at a detailed and generally applicable formalism and utilization of high-granularity in the design cycle.

35. Mukhopadhyay D: **The molecular evolutionary history of a winged bean α-chymotrypsin inhibitor and modeling of its mutations through structural analysis**. *J Mol Evol* 2000, **50**:214-223.

36. Ponting CP, Russell RB: **Identification of distant homologues of fibroblast growth factors suggests a common ancestor for all beta-trefoil proteins**. *J Mol Biol* 2000, **302**:1041-1047.

37. 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.

38. Fortenberry C, Bowman EA, Proffitt W, Dorr B, Combs S, Harp J, Mizoue L, Meiler J: **Exploring symmetry as an avenue to the computational design of large protein domains**. *J Am Chem Soc* 2011, **133**:18026-18029.

39. Akanuma S, Yamagishi A: **Roles for the two N-terminal (β/α) modules in the folding of a $(\beta/\alpha)_8$-barrel protein as studied by fragmentation analysis**. *Proteins* 2011, **79**:221-231.

40. Hocker B, Schmidt S, Sterner R: **A common evolutionary origin of two elementary enzyme folds**. *FEBS Lett* 2002, **510**:133-135.

41. •• Lee J, Blaber M: **Experimental support for the evolution of symmetric protein architecture from a simple peptide motif**. *Proc Natl Acad Sci U S A* 2011, **108**:126-130.

A top-down design study of the β-trefoil architecture that was the first to successfully identify a purely symmetric hyperthermophile solution and also provide support for one of the two competing hypotheses of evolution via gene duplication and fusion processes.

42. Hocker B, Jurgens C, Wilmanns M, Sterner R: **Stability, catalytic versatility and evolution of the $(\beta\alpha)_8$-barrel fold**. *Curr Opin Biotechnol* 2001, **12**:376-381.

43. Akanuma S, Yamagishi A: **Identification and characterization of key substructures involved in the early folding events of a $(\beta/\alpha)_8$-barrel protein as studied by experimental and computational methods**. *J Mol Biol* 2005, **353**:1161-1170.

44. Corey EJ, Cheng X-M: *The Logic of Chemical Synthesis*. New York: John Wiley & Sons, Inc.; 1989.

45. Jennings P, Roy M, Heidary D, Gross L: **Folding pathway of interleukin-1 beta**. *Nat Struct Biol* 1998, **5**:11.

46. Samuel D, Kumar TKS, Balamurugan K, Lin W-Y, Chin D-H, Yu C: **Structural events during the refolding of an all β-sheet protein**. *J Biol Chem* 2001, **276**:4134-4141.

47. Liu C, Gaspar JA, Wong HJ, Meiering EM: **Conserved and nonconserved features of the folding pathway of hisactophilin, a β-trefoil protein**. *Protein Sci* 2002, **11**:669-679.

48. • Yadid I, Kirshenbaum N, Sharon M, Dym O, Tawfik DS: **Metamorphic proteins mediate evolutionary transitions of structure**. *Proc Natl Acad Sci U S A* 2010, **107**:7287-7292.

Demonstrates both a robust evolutionary mechanism in the accommodation of novel protein architecture in response to gene duplication/fusion events and also successfully identifies a simplified polypeptide building block for β-propeller target architecture.

49. Hocker B, Lochner A, Seitz T, Claren J, Sterner R: **High-resolution crystal structure of an artificial (betaalpha)(8)-barrel protein designed from identical half-barrels**. *Biochemistry* 2009, **48**:1145-1147.