Structure
Article

Evolution and Design of Protein Structure by Folding Nucleus Symmetric Expansion

Liam M. Longo,[1] Ozan S. Kumru,[2] C. Russell Middaugh,[2] and Michael Blaber[1,*]
[1]Department of Biomedical Sciences, Florida State University, Tallahassee, FL 32306-4300, USA
[2]Department of Pharmaceutical Chemistry, University of Kansas, Lawrence, KS 66047, USA
*Correspondence: michael.blaber@med.fsu.edu
http://dx.doi.org/10.1016/j.str.2014.08.008

SUMMARY

Models of symmetric protein evolution typically invoke gene duplication and fusion events, in which repetition of a structural motif generates foldable, stable symmetric protein architecture. Success of such evolutionary processes suggests that the duplicated structural motif must be capable of nucleating protein folding. If correct, symmetric expansion of a folding nucleus sequence derived from an extant symmetric fold may be an elegant and computationally tractable solution to de novo protein design. We report the efficient de novo design of a β-trefoil protein by symmetric expansion of a β-trefoil folding nucleus, previously identified by ϕ-value analysis. The resulting protein, having exact sequence symmetry, exhibits superior folding properties compared to its naturally evolved progenitor—with the potential for redundant folding nuclei. In principle, folding nucleus symmetric expansion can be applied to any given symmetric protein fold (that is, nearly one-third of the known proteome) provided information of the folding nucleus is available.

INTRODUCTION

Although a detailed understanding of protein folding, evolution, and design remains elusive, it is generally acknowledged that gene duplication and fusion is the likely evolutionary process responsible for the emergence of common symmetric protein architecture from simpler peptide motifs (McLachlan, 1972; Ohno, 1970). Symmetry is also a strategy to substantially simplify and parameterize computational methods for efficient de novo protein design (Bellesia et al., 2010; Fortenberry et al., 2011; Kuhlman et al., 2003; Lehmann and Saven, 2008; Richter et al., 2010) and exact symmetry of protein primary and tertiary structure has been experimentally shown to confer robust foldability in the face of major structural rearrangements (Longo et al., 2013). Despite the prevalence of symmetry in protein tertiary structure, a clear understanding of the role of symmetry in protein evolution, as well as the development of practical symmetric design principles (e.g., principles that can leverage the simplifying power of symmetry with the nucleation condensation mechanism of protein folding) are lacking.

The gene duplication and fusion hypothesis for the evolution of symmetric protein architectures implies that the emergence of symmetric protein folds was a consequence of major errors of DNA replication in which DNA segments encoding short peptide motifs were duplicated and concatenated (McLachlan, 1972; Ohno, 1970). Theoretical and experimental studies have supported a model of symmetric protein evolution in which the extant symmetric architecture likely emerged from a homo-oligomeric assembly of a smaller, archaic peptide motif via gene duplication and fusion events (Alsenaidy et al., 2012; Lang et al., 2000; Lee and Blaber, 2011; Ponting and Russell, 2000). From the standpoint of protein folding, for such a model to be feasible, the archaic peptide motif must have been capable of supporting protein foldability; that is, this first motif must have acted as an efficient folding nucleus. This folding nucleus was subsequently expanded by the symmetry intrinsic to the gene duplication and fusion evolutionary process, thereby obviating the need for oligomerization to form complex symmetric protein architecture and, importantly for enabling functional adaptive radiation, resulting in subsequent asymmetric divergence. Thus, it is reasonable to hypothesize that the "building block" of evolution early in the emergence of symmetric protein folds was the folding nucleus. If this hypothesis is correct, the folding nucleus represents the critical element in de novo protein design of symmetric protein architecture.

Studies of the folding nucleus of extant protein folds indicate that the nucleus typically comprises one-third to one-half of the overall polypeptide chain of single-domain, globular proteins (Bradley and Barrick, 2006; Courtemanche and Barrick, 2008; Fowler and Clarke, 2001; Kim et al., 2000; Lindberg et al., 2006; Liu et al., 2002; Longo et al., 2012; Saeki et al., 2004; Went and Jackson, 2005). Furthermore, due to divergence subsequent to gene duplication and fusion, the folding nucleus may be a cryptic region—not defined by exon boundaries or contained neatly within one of the apparent structural repeating motifs. ϕ-value analysis is one means by which the folding nucleus can be experimentally identified via its contribution to formation of the folding transition state (Fersht and Sato, 2004). Thus, whereas the presence of an efficient folding nucleus appears to be a critical design requirement, there is no clear principle for the efficient utilization of a folding nucleus in protein design, as well as how to complete the design of the remaining majority of the polypeptide, to produce a robustly folding protein.

Folding nucleus symmetric expansion (FNSE; Figure 1) is an efficient protein design approach that exploits evolutionary principles and experimental identification of a folding nucleus to rapidly generate a completely symmetric globular protein

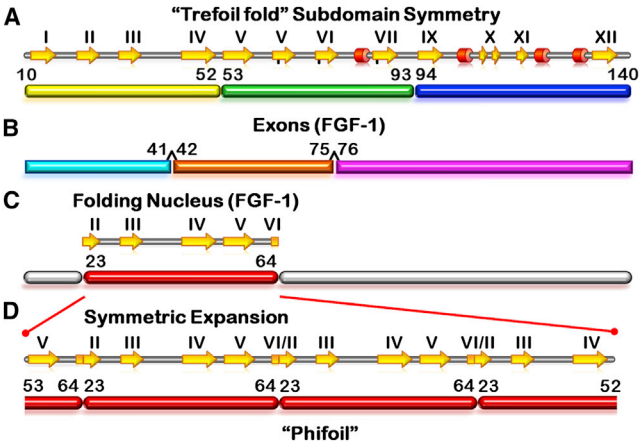

Figure 1. Scheme for Folding Nucleus Symmetric Expansion Applied to the β-Trefoil Symmetric Protein Fold

(A) Secondary structure schematic of FGF-1 with the repeating "trefoil-fold" structural subdomains indicated by different colors and associated secondary structure elements (arrows indicate β strands, cylinders helices, and gray bars loops/surface turns). The amino acid numbering scheme of the 140 amino acid form of FGF-1 is used throughout.

(B) Intron-exon structure of the FGF-1 gene.

(C) Location of critical folding nucleus of FGF-1 as determined from experimental ϕ-value analysis.

(D) Phifoil design based on the folding nucleus of FGF-1 expanded by the three-fold symmetry of the β-trefoil target architecture (and associated secondary structure elements derived from FGF-1).

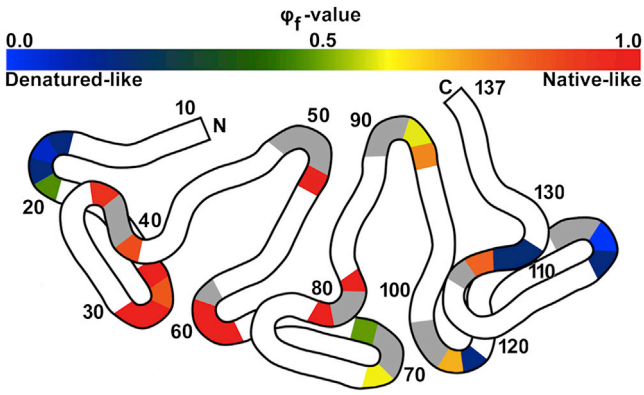

Figure 2. Heat Map of ϕ-Values in FGF-1

A previously reported study of ϕ-values of FGF-1 at all turn positions demonstrated that FGF-1 has a highly polarized folding nucleus, comprised principally of residue positions 23–64 (Longo et al., 2012).

using a remarkably simple design principle. FNSE is based on the hypothesis that a purely symmetric protein (that is, a protein with both primary and tertiary structure symmetry) can be created by expanding the primary structure of the folding nucleus according to the intrinsic structural symmetry of the target architecture—independent of structural repeat definitions. Thus, with this simple design rule, a nucleation-condensation design principle is neatly combined with a solution to the completion of the entire polypeptide sequence. Furthermore, the properties of the resulting purely symmetric sequence may include folding that is robust to subsequent diverse mutational change due to the potential presence of redundant folding nuclei. In this regard, such a protein would be an ideal scaffold for functional engineering studies, an approach that has demonstrated feasibility (Farid et al., 2013; MacDonald et al., 2010; Parmeggiani et al., 2008).

As a test of the FNSE method, we applied it to the de novo design of a β-trefoil architecture (a common single-domain globular protein fold having threefold symmetric tertiary structure). Previously, the folding nucleus of fibroblast growth factor-1 (FGF-1; a β-trefoil protein) was identified (Longo et al., 2012) by ϕ-value analysis (Fersht and Sato, 2004; Goldenberg et al., 1989; Serrano et al., 1992; Figure 2). The primary structure of the FGF-1 folding nucleus was subsequently internally propagated using a three-fold symmetry operator to all equivalent positions throughout the entire β-trefoil tertiary structure to generate a purely symmetric protein scaffold in a single design step (Figure 3). The resulting protein (dubbed Phifoil for "ϕ-value analysis derived β-trefoil") is an efficiently folding polypeptide that correctly adopts the β-trefoil target architecture. Furthermore, Phifoil is more thermostable and significantly less aggregation prone during thermal unfolding than FGF-1. Phifoil, unlike FGF-1, is exceptionally well described by cooperative two-state models of protein folding. Notably, the exact three-fold symmetry of the primary structure within Phifoil provides the potential for redundant (i.e., two intact and one interrupted, or three circularly permuted) folding nuclei—any one of which may be sufficient for foldability.

A key finding of the present study is that FNSE can provide a remarkably simple and straightforward design strategy to efficiently produce robustly folding complex protein architecture. Taken together, the results highlight an underlying connectedness between the folding nucleus and protein evolution, symmetry, and design.

RESULTS

Primary Structure Characteristics

The sequence of Phifoil was extracted directly, with no mutational change, from the folding nucleus region of wild-type FGF-1 based on a previously reported ϕ-value analysis (Figure 2; Longo et al., 2012). As a consequence of the FNSE design (Figure 1), all three structural subdomains ("trefoil-folds") of Phifoil exhibit 100% sequence identity (Figure 3); conversely, there is only one symmetry-related position in FGF-1 in which all three subdomains share the same amino acid—a Gly residue at positions 29, 71, and 115. Although unintended, the size of the amino acid alphabet used by Phifoil is reduced: only 15 of the 20 amino acid types present in FGF-1 are contained within the Phifoil sequence (Asn, Cys, Met, Phe, and Trp are excluded). Thus, Phifoil contains both exact primary structure symmetry and a reduced amino acid alphabet.

Isothermal Equilibrium Denaturation

Chemical denaturation by guanidinium hydrochloride (GuHCl) of both FGF-1 and Phifoil is well described by a two-state unfolding model (Figure 4A; Table S1 available online). The unfolding transition of both proteins is highly cooperative, although Phifoil exhibits an ~10% reduction in *m*-value compared to FGF-1, presumably reflecting its smaller size ($L_{Phifoil}$ = 126 residues,

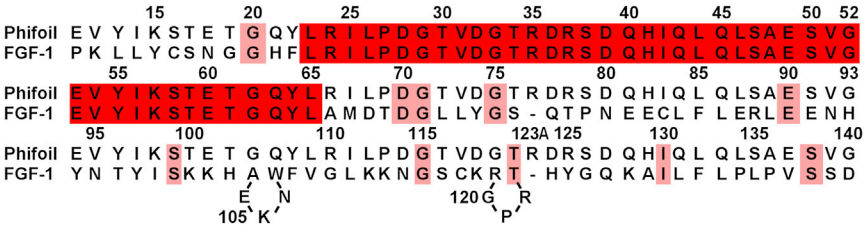

Figure 3. Primary Structures of Phifoil and FGF-1

The primary structures of Phifoil and FGF-1 (single-letter code) are arranged according to the three repeating trefoil-fold structural subdomains (Figure 1A). The numbering scheme is based upon FGF-1 (relative gaps or insertions are indicated). The red shading of contiguous amino acid positions 23–64 identify the folding nucleus of FGF-1 that was expanded by the three-fold symmetry of the β-trefoil architecture to construct the Phifoil protein (Figure 1C). Additional pink shaded positions identify identical amino acids between FGF-1 and Phifoil in the remaining regions of the protein—showing that the primary structure of the folding nucleus exhibits minimal identity with the other symmetry-related positions in FGF-1.

$L_{FGF\text{-}1}$ = 140 residues). Indeed, the predicted *m*-value (Geierhaas et al., 2007) for Phifoil based on the number of ordered residues in the crystal structure is consistent with, albeit slightly lower than, the experimentally determined value (m_{pred} = 16.1 kJ/mol/M, m_{obs} = 17.5 kJ/mol/M), suggesting that the Phifoil unfolding reaction spans a well-ordered native state and a highly unstructured unfolded state. Both proteins possess essentially identical stability at 25°C in the absence of denaturant, with ΔG_{unf} = 21.9 ± 0.3 kJ/mol (Blaber et al., 1999) and 20.7 ± 0.3 kJ/mol for FGF-1 and Phifoil, respectively.

Differential Scanning Calorimetry

The thermal unfolding of Phifoil is markedly dissimilar to that of FGF-1 (Figure 4B; Table S2) as assessed by differential scanning calorimetry (DSC). Unfolding of FGF-1 is noncooperative and characterized by significant aggregation near the unfolding transition, manifest as an exothermic signal subsequent to unfolding (apparent negative ΔC_p) and with visible precipitation in recovered samples. As such, thermodynamic parameters describing the thermal denaturation of FGF-1 without denaturant cannot be directly determined (Blaber et al., 1999). In contrast, Phifoil denaturation is well described by a two-state model of protein unfolding, with $\Delta H_{van't\ Hoff}/\Delta H_{cal}$ equal to unity. Furthermore, Phifoil unfolding is associated with a significant positive ΔCp, as expected for exposure of hydrophobic residues to solvent upon denaturation (Privalov, 1963). The predicted value of ΔC_p (T_m) for Phifoil unfolding based on the *m*-value from isothermal equilibrium denaturation (Geierhaas et al., 2007) is in good agreement with the experimentally determined constant pressure heat capacity ($\Delta C_{p,pred}$ = 10.1 kJ/mol/K, $\Delta C_{p,obs}$ = 9.5 kJ/mol/K), again highlighting the two-state nature of the Phifoil unfolding reaction. Although it is not possible to accurately compare the unfolding temperatures of FGF-1 and Phifoil (due to FGF-1 precipitation), the temperature of the excess enthalpy peak of the Phifoil endotherm occurs ~20°C higher than that of FGF-1, indicating that Phifoil is substantially more robust to thermal unfolding.

Previously, it was demonstrated that the thermal unfolding of FGF-1 is two-state in the presence of 0.7 M GuHCl, thereby permitting calculation of thermodynamic parameters (Blaber et al., 1999). Under this condition, both proteins exhibit agreement with a two-state model of protein unfolding and $\Delta H_{van't\ Hoff}/\Delta H_{cal}$ is close to unity. In 0.7 M GuHCl, the melting temperature (T_m) of Phifoil is ~8.5°C higher than that of FGF-1, again indicating that Phifoil is more robust to thermal unfolding. The enthalpy of unfolding of Phifoil is smaller than that of FGF-1 by ~16% in 0.7 M GuHCl. ΔC_p (T_m) for FGF-1 in the presence of 0.7 M GuHCl and higher (where two-state folding is observed) is 9.3 kJ/mol/K—essentially identical to that of Phifoil (in the absence of denaturant), suggesting that the unfolded states of the two proteins are equally unfolded.

Empirical Phase Diagrams

pH versus temperature empirical phase diagrams (EPDs; Fan et al., 2007) for FGF-1 and Phifoil were determined (Figure 5; the FGF-1 EPD is a new diagram that includes prior published data; Alsenaidy et al., 2012). Probes were chosen to monitor several key aspects of protein structure: circular dichroism (CD) was selected to probe secondary structure formation, 8-anilinonaphthalene-1-sulfonic acid (ANS) binding was included as a probe of partially folded or molten globule states, and static light scattering (SLS) was used as a probe of protein aggregation. Taken together, these data provide a comprehensive view of the structural state occupied by the protein at any given pH and temperature.

Phifoil, unlike FGF-1, exhibits a cooperative unfolding transition as seen by CD at every pH tested. Under conditions in which FGF-1 also exhibits a cooperative unfolding CD transition, Phifoil is significantly more thermostable in each case. The unfolding of Phifoil appears two-state over a much wider range of pH compared to FGF-1; for example, from pH 6 to pH 8, the unfolding transition of Phifoil is not associated with either aggregation or ANS binding. In contrast, there is no pH at which FGF-1 undergoes a two-state transition; unfolding of FGF-1 is always associated with both ANS binding and aggregation. In cases in which Phifoil does exhibit ANS binding and aggregation, these signals occur subsequent to or concurrent with the unfolding event (Figure 5; Figure S1) whereas with FGF-1, ANS binding (and to a lesser extent, aggregation) are observed prior to secondary structure melting. Aggregation of Phifoil is observed only at pH 4 and pH 5 near the calculated isoelectric point ($pI_{Phifoil}$ = 4.63). In total, the EPDs reveal that Phifoil is better described by a two-state unfolding process, is more thermostable, less aggregation-prone, and less likely to adopt partially folded or molten globule states across a wide range of pH and temperature compared to FGF-1.

X-ray Crystallography

An X-ray structure of Phifoil was solved to 2.15 Å (Figure 6; see Table 1 for refinement statistics); the crystal structure for human FGF-1 has been previously reported (Bernett et al., 2004; Blaber et al., 1996). Phifoil adopts an idealized (i.e., purely symmetric)

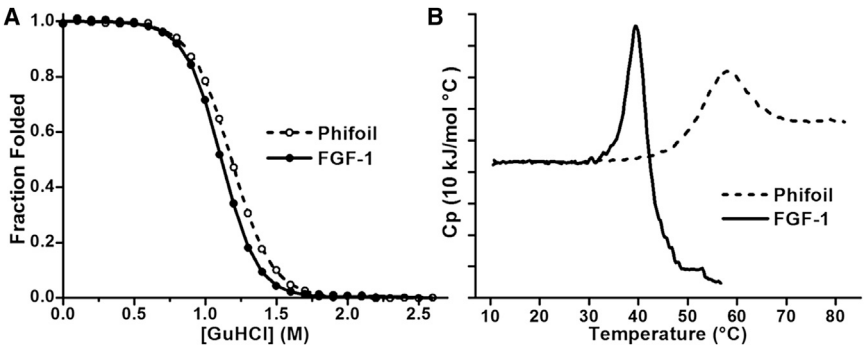

Figure 4. Unfolding of FGF-1 and Phifoil

(A) Isothermal equilibrium denaturation by GuHCl of FGF-1 (solid line) and Phifoil (dashed line). The lines are the two-state model fit to the indicated experimental data points.

(B) Differential scanning calorimetry endotherms of FGF-1 (solid line) and Phifoil (dashed line). Fitted parameters for isothermal equilibrium denaturation and differential scanning calorimetry are provided in Tables S1 and S2, respectively.

β-trefoil architecture with the major structural differences compared to FGF-1 localized to regions of relative insertions/deletions (root-mean-square deviation $[rmsd]_{C\alpha}$ for conserved regions = 1.0 Å). Structural analysis of the β-trefoil fold highlights 15 key positions that form a solvent-excluded central hydrophobic packing group (core-packing data provided in Table S3). In Phifoil, these positions are completely excluded from solvent (as determined with a 1.2 Å radius probe), whereas in FGF-1, this set demonstrates a partial solvent accessible surface area of 15.9 $Å^2$ (the principle contributors being Leu14, Val109, and Cys117). Overall core volumes (calculated from the side chain volumes and ignoring partial accessibilities) are 2,497 $Å^3$ for FGF-1 and 2,387 $Å^3$ for Phifoil. Thus, the Phifoil core buries ~110 $Å^3$ (~4%) less hydrophobic volume than FGF-1, a modest reduction in hydrophobicity. Cavity calculations using a 1.2 Å radius probe identify an ~30 $Å^3$ central cavity in the FGF-1 structure that is absent in Phifoil. Although cavities are identified in the Phifoil structure, they are distributed and do not form a large central cavity and have a combined volume of ~18 $Å^3$; that is, 12 $Å^3$ less than the central cavity of FGF-1. Thus, despite a 4% loss in hydrophobic volume, Phifoil packs its core more efficiently than FGF-1, as evidenced by the reduction in cavity size and the lack of exposed hydrophobic surface area. The preceding observation is explained principally by the fact that Phifoil lacks two functional insertions (positions 104–106 and 120–122) that are present in the third trefoil-fold subdomain of FGF-1. These insertions provide for heparin-binding functionality in FGF-1 but distort the three-fold symmetry of the central barrel (Brych et al., 2004).

"Trefoil-Fold" Structural Subdomain Symmetric Expansion

The regions comprising residue positions 11–52, 53–93, and 94–140 define the three repeating "trefoil-fold" subdomains in FGF-1 (Figure 3). As a control experiment, three different β-trefoil constructs were prepared by symmetric expansion of these domains (i.e., utilizing a structure-based motif, not a folding nucleus-based motif). This structure-based motif approach reflects a common strategy in attempts to design symmetric protein architectures such as β-propeller (Yadid and Tawfik, 2007, 2011) and $(\beta/\alpha)_8$-barrel (i.e., TIM barrel; Höcker et al., 2004; Richter et al., 2010) proteins—although such studies have met with limited success. Expression and purification of these three β-trefoil proteins (i.e., symmetric expansion of regions of FGF-1 not containing an intact folding nucleus) was attempted but failed in each case (data not shown). The constructs derived from the first β-trefoil subdomain and the third β-trefoil subdomain precipitated completely upon cell lysis; the construct derived from the second β-trefoil subdomain failed to express.

DISCUSSION

The prevalence of symmetric protein folds is something of a conundrum because their evolutionary emergence would appear to involve two consequences for protein folding that, individually, are typically considered lethal. The first is that gene duplication and fusion represents a major replication error—producing gross alterations to protein structure. Proteins have a delicate thermodynamic balance in favor of the native structure, and while conservative point mutations can often be accommodated, major alterations invariably tip this balance in favor of unfolding. The second is that gene duplication and fusion yields extensive regions of exact repeating primary structure—a situation postulated to frustrate folding pathways, resulting in kinetically trapped misfolded forms (Borgia et al., 2011; Wright et al., 2005). The present results, however, suggest robust biophysical properties of folding for the Phifoil protein produced by symmetric expansion of the folding nucleus.

FNSE provides a straightforward approach by which knowledge of the folding nucleus—determined either experimentally or computationally—can be used to simply and rapidly generate an intact symmetric protein fold. Motivated by hypotheses regarding symmetric protein evolution, FNSE should, in principle, be applicable to any common protein architecture characterized by internal symmetry (i.e., approximately one-third of known protein architectures). FNSE highlights a fundamental relationship between protein evolution, folding, and design by rebranding the critical folding nucleus as the key evolutionary building block and de novo design element. Proteins generated by FNSE will exhibit exact symmetry; that is, both the primary and tertiary structures of the resulting architecture will be symmetric. The fitness for folding of such exact symmetry has previously been contested (Borgia et al., 2011; Wright et al., 2005; Yadid and Tawfik, 2011), although recent reports have unambiguously demonstrated that exact symmetry is compatible with efficient foldability (Alsenaidy et al., 2012; Höcker et al., 2009; Lee and Blaber, 2011) and that symmetric sequences can maintain foldability in the face of major structural rearrangement (Longo et al., 2013). Indeed, symmetric sequences built from FNSE would make ideal protein "scaffolds" because they

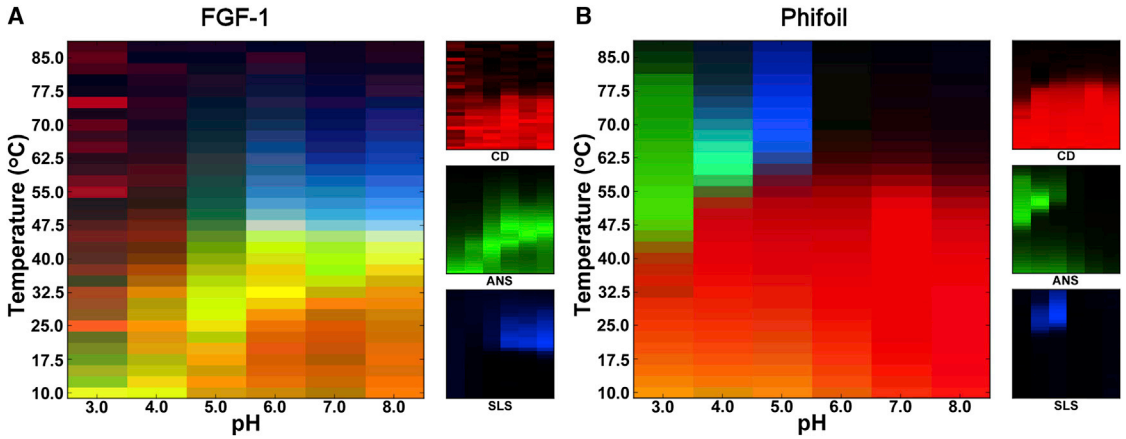

Figure 5. pH versus Temperature Empirical Phase Diagrams of FGF-1 and Phifoil

EPDs (Kim et al., 2012) were constructed using three structural probes: circular dichroism (CD), in which the red color indicates native-like secondary structure; ANS binding, in which green color indicates partially folded or molten-globule-like states; and static light scattering (SLS), in which blue color indicates protein aggregation. Individual contributions from each probe are shown on the right; with the large images showing the additive overlay of all probe data (A, FGF-1; B, Phifoil). Source data used to generate the Phifoil EPD are shown in Figure S1.

potentially contain built-in folding pathway redundancy, robust to diverse subsequent functional mutation.

The success of FNSE helps to elucidate the failure of structural motif-based symmetric protein design, as well as provides a possible explanation as to why proteins sharing a common symmetric fold do not necessarily share a related folding nucleus (Jennings et al., 1998; Liu et al., 2002; Longo et al., 2012). A number of experimental attempts have been made to leverage structural symmetry to simplify the protein design process, because symmetry provides a means to efficiently parameterize the protein design problem. In the majority of these studies, the repeating structural subdomain (i.e., typically defined beginning with the N terminus) is utilized as a building block and expressed as a fragment, or expanded as a multimer according the symmetry of the fold, with the goal of generating a foldable symmetric target architecture. Such approaches have proven largely unsuccessful, and were similarly unsuccessful in this study when using the different trefoil-fold repeat definitions in FGF-1 as building blocks. The key shortcoming of this structural motif-based approach is that it tacitly assumes that the critical folding nucleus is neatly contained within one of the repeating structural motifs. In the ϕ-value analysis of FGF-1 this assumption is clearly false; the folding nucleus is not neatly encapsulated within any of the three trefoil-fold repeating motifs. Furthermore, the folding nucleus is not neatly contained within any of the three exons of FGF-1. Thus, the folding nucleus of FGF-1 is a cryptic region using either structure-based or exon-based definitions as a probe (Figure 1).

Top-down symmetric deconstruction (TDSD; Lee et al., 2011) has been used to successfully generate a purely symmetric β-trefoil architecture (the Symfoil protein) starting from the highly asymmetric FGF-1 β-trefoil protein. The TDSD approach relies on the iterative experimental identification of mutations that can increase the symmetry of the protein and incorporates a thermostability and folding cooperativity screen; as such, TDSD is labor intensive, difficult to implement computationally, and requires significant user expertise. In contrast, FNSE is straightforward to implement once knowledge of the folding nucleus is obtained; thus, computational formulation of FNSE is feasible (and can potentially be exclusively computational if folding nuclei can be computationally identified). Critically, the success of TDSD appears to be due to the conceptual framework identified by FNSE: Symfoil-4T, the final product of TDSD of FGF-1 to generate a symmetric β-trefoil protein, shares 77% sequence identity with Phifoil, indicating that TDSD was successful because it effectively isolated the cryptic folding nucleus of FGF-1 (which, at the time, was unknown).

The FNSE results, furthermore, suggest that the earliest symmetric proteins (i.e., the immediate product of duplication and fusion events) may have enjoyed built-in folding redundancy by virtue of having multiple—potentially overlapping—folding nuclei present within a single sequence. This condition is intrinsic to a gene duplication and fusion evolutionary process and comes with clear evolutionary advantages, because it may permit foldability in the face of major structural rearrangement (Longo et al., 2013), and subsequent acquisition of mutations that may disrupt folding of one nucleus while still maintaining overall protein foldability through the presence of one or more redundant folding nuclei. This latter postulate explains why symmetric proteins distantly related by evolution need not exhibit the same folding nucleus location within the sequence. Related to this point is the observation that the folding nucleus of FGF-1 is not localized to any individual trefoil-fold subdomain, and is thus an apparent circular permutation of the presumed archetypical trefoil-fold subdomain. Circular permutation can result from gene duplication and fusion events also involving truncations—such that the apparent structural repeat domain no longer defines the folding nucleus. Thus, it is unclear whether the original β-trefoil building block may have used an alternate termini definition; however, it seems more likely that multiple nucleating sequences present in a single structure may be eroded by evolution to preserve only what is necessary to maintain folding—with no selective pressure to observe repeating motif or exon demarcations. The failure of repeat domain-based symmetric expansion highlights

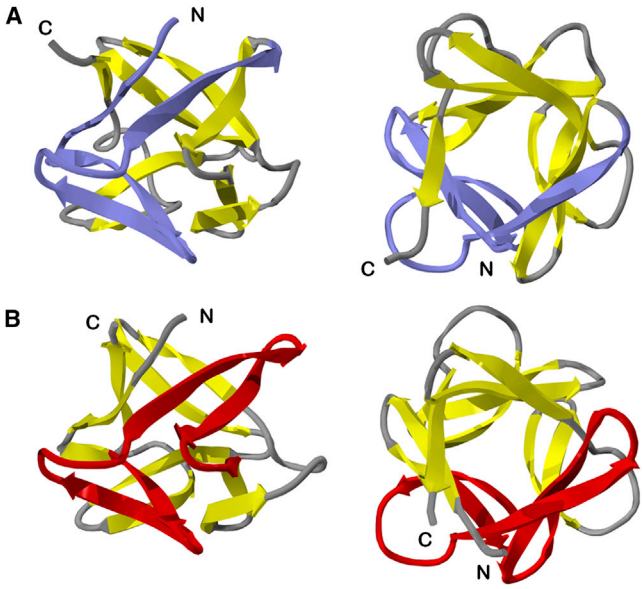

Figure 6. Main Chain Ribbon Diagram of FGF-1 and Phifoil Crystal Structures

(A) Ribbon diagram of FGF-1 (side view, left; top view [i.e., parallel to the three-fold axis of rotational symmetry], right) and with the blue region identifying the three-fold repeating trefoil-fold structural subdomain (PDB accession 2AFG). (B) A ribbon representation for Phifoil (same orientations as the FGF-1 ribbon diagrams) and with the red region identifying residue positions 23–64; the folding nucleus extracted from FGF-1 and expanded by the intrinsic three-fold symmetry to generate the purely symmetric Phifoil primary structure. The N- and C-termini positions are indicated.

Table 1. Crystallographic Data Collection and Refinement Statistics

	Phifoil[a]
Space group	$P2_12_12_1$
Cell constants (Å)	a = 52.9
	b = 68.6
	c = 76.0
	$\alpha = 90^\circ$
	$\beta = 90^\circ$
	$\gamma = 90^\circ$
Max Resolution (Å)	2.15
Mosaicity (°)	0.48
Redundancy	2.2
Mol/ASU	2
Matthews coeff. ($Å^3$/Da)	2.50
Total reflections	27,752
Unique reflections	12,829
I/σ (overall)	18.8
I/σ (highest shell[b])	4.1
Completion overall (%)	81.7
Completion highest shell[b] (%)	86.9
R_{merge} overall (%)	8.0
R_{merge} highest shell[b] (%)	19.6
Nonhydrogen protein atoms	1,939
Solvent molecules/ion	132/1
R_{cryst} (%)	22.8
R_{free} (%)	26.7
Rmsd bond length (Å)	0.005
Rmsd bond angle (°)	0.94
Ramachandran plot	
Most favored (%)	97.5
Additional allowed (%)	2.5
Generously allowed (%)	0
Disallowed region (%)	0
PDB accession	4OW4

[a]800 mM $(NH_4)_2SO_4$, 100 mM citric acid, pH 4.0.
[b]Highest shell 2.21–2.15 Å.

the essentially irreversible evolutionary process associated with such mutational "erosion," i.e., that the folding nucleus in a symmetric protein is no longer constrained to reside neatly within the apparent repeating structural motif. A previously published circular permutation study of a related symmetric β-trefoil protein (with permutations at all possible surface turn positions) showed that alternative definitions for the N and C termini, in comparison to the native termini definitions, are destabilizing in each case (Longo et al., 2013). Fragmentation studies of a related symmetric β-trefoil solution also showed that the native termini definitions constitute a foldable solution for a putative ancient 42-mer peptide motif (Lee and Blaber, 2011). Furthermore, alternative termini definitions that occur at position 23 and 64 do not residue within an exposed surface turn, but rather, within the central region of a β strand—with an expectation of major destabilization resulting for termini location. Thus, the folding nucleus of FGF-1 is a centrally located structural element, but beyond this appears to have minimal association with current boundary definitions of the structural repeat, exons, or putative ancient progenitor motif.

The improved biophysical properties of Phifoil compared to FGF-1 are an additional demonstration of the burden of function on foldability, stability, and solubility. It has been reported that functional regions of FGF-1 are highly segregated from the region that comprises the folding nucleus (Longo et al., 2012); thus, the construction of Phifoil, which was built from a region of FGF-1 devoid of any known function, may help disentangle which biophysical properties of FGF-1 are the consequence of functional mutation. Perhaps the most striking difference between Phifoil and FGF-1 is the fact that Phifoil is dramatically less prone to aggregation (only aggregating near the pI), suggesting that functional acquisition has negatively affected the folding and aggregation properties of FGF-1. In this regard, whereas the construction of Phifoil using the folding nucleus of FGF-1 has captured the essential stability of FGF-1, the precipitation of FGF-1 during unfolding is due to the (functional) regions outside the essential folding nucleus.

FNSE is an approach to rapidly design a stable protein scaffold with favorable biophysical properties useful for subsequent mutational engineering. The success of this approach underpins the evolutionary emergence of symmetric protein folds, and highlights symmetry as a structural property linking the evolution, folding, and efficient design of complex protein architecture with the critical folding nucleus.

EXPERIMENTAL PROCEDURES

Protein Design by Folding Nucleus Symmetric Expansion

Residue positions identified by ϕ-value analysis as contributing to the folding nucleus of FGF-1 span the region from turn 2 (position 26) through turn 7 (position 83)—approximately 46% of the basic β-trefoil architecture (58 amino acids out of 132 amino acid positions) of FGF-1 (Longo et al., 2012). However, turn 6 (positions 68–71) within this region exhibits ϕ-values indicative of folding after the transition state (i.e., ϕ-values 0.3–0.77). Additionally, symmetric expansion requires a region comprising precisely one-third of the basic β-trefoil architecture. To achieve this length requirement, and to capture the major folding nucleus region, the region comprising turns 2–5 was selected as the folding nucleus. Several amino acids before turn 2 and after turn 5 were incorporated in the final construct, because these positions are also likely key contributors to the folding nucleus. Thus, the design element used to create Phifoil spans residues 23–64 or, equivalently, residues 24–65 (an indistinguishable solution; Figure 3). Although this 42-residue region is smaller than the entire folding nucleus identified by ϕ-value analysis, the symmetric expansion of this sequence to structurally equivalent positions means that the "missing" parts of the experimentally determined folding nucleus are regenerated by structurally equivalent residues from positions residing within the folding nucleus. Alternative definitions involving minor variation in the precise start point may serve equally well as efficient folding nuclei. The final Phifoil construct was generated by symmetric expansion of the region 23–64 (24–65) folding nucleus but retaining the wild-type β-trefoil N and C terminus definitions (which is probably the most stable termini configuration (Longo et al., 2013; Figures 1 and 3). Thus, the Phifoil protein contains two intact folding nuclei from FGF-1 as well as one interrupted folding nucleus (partial regions of which are located at the N and C termini). Alternatively, if the original longer 58 amino acid definition of the FGF-1 folding nucleus is effectively regenerated by the symmetric expansion, there are two intact and overlapping larger folding nuclei.

Phifoil Expression and Purification

Details of protein mutagenesis, expression, and purification are provided in the Supplemental Experimental Procedures.

X-ray Crystallography

Purified Phifoil in phosphate buffer was concentrated to ~12 mg/mL and crystal conditions were screened using the hanging-drop vapor diffusion method at 25°C. Diffraction quality crystals grew in ~1 month from vapor diffusion against 800 mM $(NH_4)_2SO_4$, 100 mM citric acid, pH 4.0. A crystal was mounted using a Hampton Research nylon cryo-loop and cooled in a stream of gaseous nitrogen to 100 K. Diffraction data were collected using an in-house Rigaku RU-H3R rotating anode X-ray source (Rigaku, Tokyo, Japan) equipped with Osmic confocal mirrors (Osmic) and a MarCCD165 (Rayonix) detector. The data were indexed, integrated, and scaled using the HKL2000 software package (Z. Otwinowski, unpublished data, 1993; Otwinowski and Minor, 1997). Molecular replacement and model building used the PHENIX software package (Zwart et al., 2008), with 5% of the data in the reflection files set aside for R_{free} calculations (Brünger, 1992). Symfoil-1 (Protein Data Bank [PDB] accession 3O49) was used as the search model in molecular replacement. Model building and visualization utilized the COOT molecular graphics software package (Emsley and Cowtan, 2004).

Isothermal Equilibrium Denaturation

The 10 μM samples of Phifoil in ADA Buffer were incubated for ~20 hr at 25°C in the presence of 0.0–2.6 M GuHCl (i.e., 2 × C_m) in 0.1 M increments. The folding of Phifoil was monitored by fluorescence on a Cary Eclipse fluorospectrophotometer equipped with a Peltier temperature control unit (Agilent). Samples were loaded into a 1.0 cm path length quartz cuvette and were incubated for 4 min prior to collecting spectra. Tyr fluorescence was excited at 277 nm, emission was monitored from 284 to 410 nm, and slit-widths were set to 5 nm. Each sample was scanned in triplicate and the resulting spectra were averaged, buffer subtracted, and integrated to generate an unfolding curve. The resulting unfolding curve was fit to a six-parameter, two-state model of protein unfolding (Eftink, 1994) using the nonlinear, least-squares fitting program, DataFit (Oakdale Engineering). Reported errors are the SDs of three independent experiments.

Differential Scanning Calorimetry

DSC was performed on samples of 40 μM Phifoil in ADA Buffer using a VP-DSC microcalorimeter (GE Healthcare). Samples were scanned from 10°C to 95°C under 2.3 bar, with a pre-scan equilibration time of 10 min and a scan rate of 0.25°C/min (Blaber et al., 1999). Prior to protein loading, buffer-buffer scans were collected until thermal history was established. Buffer-subtracted, concentration-normalized endotherms were analyzed using the DSCFit software package (Grek et al., 2001) and SDs result from three consecutive protein loads.

Preparation of Empirical Phase Diagrams

Temperature versus pH EPDs (Fan et al., 2007) were generated using three probes of protein structure: CD, SLS, and extrinsic fluorescence. Temperatures ranged from 10.0°C to 87.5°C; pH values ranged from pH 3.0 to 8.0. Each probe at each pH was measured in triplicate and averaged to yield the final data for EPD determination. Three-index EPDs were constructed as previously described using the MiddaughSuite software package (Kim et al., 2012; Maddux et al., 2011). Additional details are provided in the Supplemental Experimental Procedures.

ACCESSION NUMBERS

The PDB accession number for the refined X-ray structure coordinates and structure factors for the Phifoil protein is 4OW4.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Experimental Procedures, one figure, and three tables and can be found with this article online at http://dx.doi.org/10.1016/j.str.2014.08.008.

ACKNOWLEDGMENTS

We thank Dr. Thayumanasamy Somasundaram for assistance with X-ray diffraction data collection and analysis. L.M.L. gratefully acknowledges The Florida State University College of Medicine for financial support. This work was supported by the FSU Research Foundation.

Received: May 16, 2014
Revised: August 2, 2014
Accepted: August 8, 2014
Published: September 18, 2014

REFERENCES

Alsenaidy, M.A., Wang, T., Kim, J.H., Joshi, S.B., Lee, J., Blaber, M., Volkin, D.B., and Middaugh, C.R. (2012). An empirical phase diagram approach to investigate conformational stability of "second-generation" functional mutants of acidic fibroblast growth factor (FGF-1). Protein Sci. *21*, 418–432.

Bellesia, G., Jewett, A.I., and Shea, J.E. (2010). Sequence periodicity and secondary structure propensity in model proteins. Protein Sci. *19*, 141–154.

Bernett, M.J., Somasundaram, T., and Blaber, M. (2004). An atomic resolution structure for human fibroblast growth factor 1. Proteins *57*, 626–634.

Blaber, M., DiSalvo, J., and Thomas, K.A. (1996). X-ray crystal structure of human acidic fibroblast growth factor. Biochemistry *35*, 2086–2094.

Blaber, S.I., Culajay, J.F., Khurana, A., and Blaber, M. (1999). Reversible thermal denaturation of human FGF-1 induced by low concentrations of guanidine hydrochloride. Biophys. J. *77*, 470–477.

Borgia, M.B., Borgia, A., Best, R.B., Steward, A., Nettels, D., Wunderlich, B., Schuler, B., and Clarke, J. (2011). Single-molecule fluorescence reveals sequence-specific misfolding in multidomain proteins. Nature *474*, 662–665.

Bradley, C.M., and Barrick, D. (2006). The notch ankyrin domain folds via a discrete, centralized pathway. Structure *14*, 1303–1312.

Brünger, A.T. (1992). Free *R* value: a novel statistical quantity for assessing the accuracy of crystal structures. Nature *355*, 472–475.

Brych, S.R., Dubey, V.K., Bienkiewicz, E., Lee, J., Logan, T.M., and Blaber, M. (2004). Symmetric primary and tertiary structure mutations within a symmetric superfold: a solution, not a constraint, to achieve a foldable polypeptide. J. Mol. Biol. *344*, 769–780.

Courtemanche, N., and Barrick, D. (2008). The leucine-rich repeat domain of Internalin B folds along a polarized N-terminal pathway. Structure *16*, 705–714.

Eftink, M.R. (1994). The use of fluorescence methods to monitor unfolding transitions in proteins. Biophys. J. *66*, 482–501.

Emsley, P., and Cowtan, K. (2004). Coot: model-building tools for molecular graphics. Acta Crystallogr. D Biol. Crystallogr. *60*, 2126–2132.

Fan, H., Li, H., Zhang, M., and Middaugh, C.R. (2007). Effects of solutes on empirical phase diagrams of human fibroblast growth factor 1. J. Pharm. Sci. *96*, 1490–1503.

Farid, T.A., Kodali, G., Solomon, L.A., Lichtenstein, B.R., Sheehan, M.M., Fry, B.A., Bialas, C., Ennist, N.M., Siedlecki, J.A., Zhao, Z., et al. (2013). Elementary tetrahelical protein design for diverse oxidoreductase functions. Nat. Chem. Biol. *9*, 826–833.

Fersht, A.R., and Sato, S. (2004). ϕ-value analysis and the nature of protein-folding transition states. Proc. Natl. Acad. Sci. USA *101*, 7976–7981.

Fortenberry, C., Bowman, E.A., Proffitt, W., Dorr, B., Combs, S., Harp, J., Mizoue, L., and Meiler, J. (2011). Exploring symmetry as an avenue to the computational design of large protein domains. J. Am. Chem. Soc. *133*, 18026–18029.

Fowler, S.B., and Clarke, J. (2001). Mapping the folding pathway of an immunoglobulin domain: structural detail from Phi value analysis and movement of the transition state. Structure *9*, 355–366.

Geierhaas, C.D., Nickson, A.A., Lindorff-Larsen, K., Clarke, J., and Vendruscolo, M. (2007). BPPred: A computational tool to predict biophysical quantities of proteins. Protein Sci. *16*, 125–134.

Goldenberg, D.P., Frieden, R.W., Haack, J.A., and Morrison, T.B. (1989). Mutational analysis of a protein-folding pathway. Nature *338*, 127–132.

Grek, S.B., Davis, J.K., and Blaber, M. (2001). An efficient, flexible-model program for the analyis of differential scanning calorimetry protein denaturation data. Protein Pept. Lett. *8*, 429–436.

Höcker, B., Claren, J., and Sterner, R. (2004). Mimicking enzyme evolution by generating new (betaalpha)8-barrels from (betaalpha)4-half-barrels. Proc. Natl. Acad. Sci. USA *101*, 16448–16453.

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

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

Kim, D.E., Fisher, C., and Baker, D. (2000). A breakdown of symmetry in the folding transition state of protein L. J. Mol. Biol. *298*, 971–984.

Kim, J.H., Iyer, V., Joshi, S.B., Volkin, D.B., and Middaugh, C.R. (2012). Improved data visualization techniques for analyzing macromolecule structural changes. Protein Sci. *21*, 1540–1553.

Kuhlman, B., Dantas, G., Ireton, G.C., Varani, G., Stoddard, B.L., and Baker, D. (2003). Design of a novel globular protein fold with atomic-level accuracy. Science *302*, 1364–1368.

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

Lee, J., and Blaber, M. (2011). Experimental support for the evolution of symmetric protein architecture from a simple peptide motif. Proc. Natl. Acad. Sci. USA *108*, 126–130.

Lee, J., Blaber, S.I., Dubey, V.K., and Blaber, M. (2011). A polypeptide "building block" for the β-trefoil fold identified by "top-down symmetric deconstruction." J. Mol. Biol. *407*, 744–763.

Lehmann, A., and Saven, J.G. (2008). Computational design of four-helix bundle proteins that bind nonbiological cofactors. Biotechnol. Prog. *24*, 74–79.

Lindberg, M.O., Haglund, E., Hubner, I.A., Shakhnovich, E.I., and Oliveberg, M. (2006). Identification of the minimal protein-folding nucleus through loop-entropy perturbations. Proc. Natl. Acad. Sci. USA *103*, 4083–4088.

Liu, C., Gaspar, J.A., Wong, H.J., and Meiering, E.M. (2002). Conserved and nonconserved features of the folding pathway of hisactophilin, a β-trefoil protein. Protein Sci. *11*, 669–679.

Longo, L., Lee, J., and Blaber, M. (2012). Experimental support for the foldability-function tradeoff hypothesis: segregation of the folding nucleus and functional regions in FGF-1. Protein Sci. *21*, 1911–1920.

Longo, L.M., Lee, J., Tenorio, C.A., and Blaber, M. (2013). Alternative folding nuclei definitions facilitate the evolution of a symmetric protein fold from a smaller peptide motif. Structure *21*, 2042–2050.

MacDonald, J.T., Maksimiak, K., Sadowski, M.I., and Taylor, W.R. (2010). De novo backbone scaffolds for protein design. Proteins *78*, 1311–1325.

Maddux, N.R., Joshi, S.B., Volkin, D.B., Ralston, J.P., and Middaugh, C.R. (2011). Multidimensional methods for the formulation of biopharmaceuticals and vaccines. J. Pharm. Sci. *100*, 4171–4197.

McLachlan, A.D. (1972). Repeating sequences and gene duplication in proteins. J. Mol. Biol. *64*, 417–437.

Ohno, S. (1970). Evolution by Gene Duplication. (New York: Allen and Unwin).

Otwinowski, Z., and Minor, W. (1997). Processing of x-ray diffraction data collected in oscillation mode. Methods Enzymol. *276*, 307–326.

Parmeggiani, F., Pellarin, R., Larsen, A.P., Varadamsetty, G., Stumpp, M.T., Zerbe, O., Caflisch, A., and Plückthun, A. (2008). Designed armadillo repeat proteins as general peptide-binding scaffolds: consensus design and computational optimization of the hydrophobic core. J. Mol. Biol. *376*, 1282–1304.

Ponting, C.P., and Russell, R.B. (2000). Identification of distant homologues of fibroblast growth factors suggests a common ancestor for all beta-trefoil proteins. J. Mol. Biol. *302*, 1041–1047.

Privalov, P.L. (1963). [Study of heat denaturation of egg albumin]. Biofizika *8*, 308–316.

Richter, M., Bosnali, M., Carstensen, L., Seitz, T., Durchschlag, H., Blanquart, S., Merkl, R., and Sterner, R. (2010). Computational and experimental evidence for the evolution of a (β α)8-barrel protein from an ancestral quarter-barrel stabilised by disulfide bonds. J. Mol. Biol. *398*, 763–773.

Saeki, K., Arai, M., Yoda, T., Nakao, M., and Kuwajima, K. (2004). Localized nature of the transition-state structure in goat α-lactalbumin folding. J. Mol. Biol. *341*, 589–604.

Serrano, L., Matouschek, A., and Fersht, A.R. (1992). The folding of an enzyme. III. Structure of the transition state for unfolding of barnase analysed by a protein engineering procedure. J. Mol. Biol. *224*, 805–818.

Went, H.M., and Jackson, S.E. (2005). Ubiquitin folds through a highly polarized transition state. Protein Eng. Des. Sel. *18*, 229–237.

Wright, C.F., Teichmann, S.A., Clarke, J., and Dobson, C.M. (2005). The importance of sequence diversity in the aggregation and evolution of proteins. Nature *438*, 878–881.

Yadid, I., and Tawfik, D.S. (2007). Reconstruction of functional β-propeller lectins via homo-oligomeric assembly of shorter fragments. J. Mol. Biol. *365*, 10–17.

Yadid, I., and Tawfik, D.S. (2011). Functional β-propeller lectins by tandem duplications of repetitive units. Protein Eng. Des. Sel. *24*, 185–195.

Zwart, P.H., Afonine, P.V., Grosse-Kunstleve, R.W., Hung, L.W., Ioerger, T.R., McCoy, A.J., McKee, E., Moriarty, N.W., Read, R.J., Sacchettini, J.C., et al. (2008). Automated structure solution with the PHENIX suite. Methods Mol. Biol. *426*, 419–435.