Experimental support for the foldability-function tradeoff hypothesis: Segregation of the folding nucleus and functional regions in FGF-1.

ARTICLE *in* PROTEIN SCIENCE · DECEMBER 2012

Impact Factor: 2.85 · DOI: 10.1002/pro.2175 · Source: PubMed

3 AUTHORS:

Liam Michael Longo
Weizmann Institute of Science
18 PUBLICATIONS **91** CITATIONS
SEE PROFILE

Jihun Lee
Celltrion
27 PUBLICATIONS **395** CITATIONS
SEE PROFILE

Michael Blaber
Florida State University
127 PUBLICATIONS **4,529** CITATIONS
SEE PROFILE

THE PROTEIN SOCIETY

Experimental support for the foldability–function tradeoff hypothesis: Segregation of the folding nucleus and functional regions in fibroblast growth factor-1

Liam Longo, Jihun Lee, and Michael Blaber*

Department of Biomedical Sciences, College of Medicine, Florida State University, Tallahassee, Florida 32306-4300

Received 21 August 2012; Revised 28 September 2012; Accepted 1 October 2012
DOI: 10.1002/pro.2175
Published online 9 October 2012 proteinscience.org

Abstract: The acquisition of function is often associated with destabilizing mutations, giving rise to the stability–function tradeoff hypothesis. To test whether function is also accommodated at the expense of foldability, fibroblast growth factor-1 (FGF-1) was subjected to a comprehensive φ-value analysis at each of the 11 turn regions. FGF-1, a β-trefoil fold, represents an excellent model system with which to evaluate the influence of function on foldability: because of its threefold symmetric structure, analysis of FGF-1 allows for direct comparisons between symmetry-related regions of the protein that are associated with function to those that are not; thus, a structural basis for regions of foldability can potentially be identified. The resulting φ-value distribution of FGF-1 is highly polarized, with the majority of positions described as either folded-like or denatured-like in the folding transition state. Regions important for folding are shown to be asymmetrically distributed within the protein architecture; furthermore, regions associated with function (i.e., heparin-binding affinity and receptor-binding affinity) are localized to regions of the protein that fold after barrier crossing (late in the folding pathway). These results provide experimental support for the foldability–function tradeoff hypothesis in the evolution of FGF-1. Notably, the results identify the potential for folding redundancy in symmetric protein architecture with important implications for protein evolution and design.

Keywords: φ-value analysis; β-trefoil; protein folding; stability–function tradeoff; protein evolution; protein symmetry; protein design

Introduction

Protein function typically relies on the precise alignment of specific main-chain or side-chain groups within the folded structure. Active conformations often require that functional residues be constrained by the global fold to adopt energetically suboptimal arrangements, such as solvent-exposed hydrophobic patches,[1] regions of high-charge density, strained conformations,[2] and buried polar/charged groups.[3,4] Numerous researchers have reported successful enzyme stabilization by active-site redesign at the expense of function.[5–8] Careful analysis of active site mutagenesis data has led to the hypothesis that optimization of functional activity is often accommodated at the expense of thermostability; that is, there exists a fundamental "stability–function tradeoff."[6,7] In this viewpoint, enzymes preorganize a small subset of residues for efficient function by enforcing structurally strained conformations within the active site, and this is offset by favorable interactions distributed over the remaining majority of the protein structure.[3,4,9–12]

Computational studies have provided strong support for an additional but distinct *foldability–function*

Jihun Lee's current address is Celltrion Inc., 13-1 Songdo-dong, Yeonsu-gu 406-840, Incheon City, Korea.

Grant sponsor: Florida State University College of Medicine.

*Correspondence to: Michael Blaber, 3350G BMS 1115 West Call St., Tallahassee, FL 32306-4300.
E-mail: michael.blaber@med.fsu.edu

tradeoff hypothesis, which suggests that native state strain is only one manifestation of the burden imposed by the acquisition of function. In this viewpoint, the requirements of function enforce structural features that are unlikely to be optimal to nucleate protein folding (i.e., they will fold after barrier crossing and be denatured-like in the folding transition state) or are more likely to form intermediates before folding properly (requiring folding "back-tracking"). For example, Gō-type simulations paired with molecular dynamics simulations revealed that the functional β-bulge of interleukin-1β (IL-1β) undergoes significant backtracking (proper folding preceded by a misfolding/unfolding event) before folding correctly—a computational result that was subsequently confirmed using real-time refolding NMR experiments.[13–15] Thus, the foldability–function tradeoff hypothesis posits that regions contributing to specific function in a protein are likely to be segregated from regions contributing to efficient folding. Folding and stability are entangled properties, in that stability defines thermodynamics (but not kinetics of folding), while folding defines kinetics (but from which thermodynamic stability can be derived).

Experimental support for the stability–function tradeoff has been well-established,[1–9] in part owing to the relative simplicity of characterizing protein stability and functional activity in response to active site mutation. Similar experimental support for a foldability–function tradeoff, however, remains scarce, reflecting a more recent postulate as well as the more significant demands of characterizing a protein's folding pathway. φ-Value analysis,[16–18] although labor-intensive, remains one of the few direct experimental techniques with which to identify key positions contributing to the formation of the folding transition state; thus, experimental evaluation of the foldability–function tradeoff hypothesis can be directly tested using φ-value analysis in combination with functional data. Furthermore, by selecting a protein fold having internal (i.e., rotational) symmetry, regions of function can be compared directly to symmetry-related positions of the protein that are not associated with function but adopt the same general structure to ask whether they make equivalent contributions to the folding transition state. In essence, the protein provides an internal standard that explicitly demonstrates the folding potential (or lack thereof) of a given structural element. If symmetry-related positions do not fold concurrently, with one folding early and another folding late, it suggests that sequence divergence between symmetry-related subdomains (either by evolution of function or neutral drift) has diminished the foldability at specific sites. Furthermore, probing the foldability of symmetry-related regions in symmetric protein architecture can elucidate the potential for folding redundancy. Thus, although φ-value studies have been performed on a variety of proteins, analysis of symmetric protein folds provides unique and valuable information about protein folding, symmetric redundancy thereof, and with important implications for protein evolution and design.

Figure 1. Left panel: Ribbon diagram view of FGF-1 (PDB ID: 1RG8) oriented parallel to the threefold axis of rotational symmetry. The N-terminal β-trefoil subdomain is indicated by black shading. Right panel: a side view of FGF-1 with identical shading to indicate the N-terminal β-trefoil subdomain.

Fibroblast growth factor-1 (FGF-1) is a heparin-binding protein that adopts the β-trefoil fold, a common protein fold that exhibits a threefold rotational symmetry at the tertiary structure level.[19,20] This architecture is composed of three repeating "trefoil-fold" subdomains (40–50 amino acids in length) each composed of a pair of antiparallel β-hairpin structures. Thus, within the structure are a total of 12 β-strands (numbered #1–12) and 11 reverse turns (numbered #1–11; Fig. 1). The primary structure of FGF-1, however, does not directly reflect the higher-order (i.e., tertiary structure) symmetry of the β-trefoil architecture: only a single amino acid position is conserved across the three repeating trefoil-fold subdomains. Such sequence asymmetry may reflect not only genetic drift but also functional and folding regions distributed asymmetrically in the primary structure.

Turns are the principle secondary structure element that permit 180° changes in the polypeptide direction and are thus a requirement for globular protein architecture (i.e., α-helix and β-strand secondary structure are linear structural elements). For efficient folding, approximate turn structure is likely formed early in the folding pathway to allow for structural collapse, and several studies have characterized the importance of turns to both folding (as folding nuclei)[21–23] and stability.[24] We report a comprehensive φ-value analysis for each of the 11 turns of FGF-1, involving a total of 44 residue positions within the 140 amino acid FGF-1 protein. The folding transition state of FGF-1 is shown to be highly polarized, with the majority of turns adopting either native-like or denatured-like structure in the folding transition state. However, symmetry-related turns do not fold concurrently, indicating that

structures important for foldability are asymmetrically distributed over the protein structure. Turns associated with heparin-binding functionality and receptor-binding sites appear largely unstructured in the folding transition state. Functionally important residues are therefore highly segregated from the regions of the protein optimized for foldability, in support of the foldability–function tradeoff hypothesis.

Results

The isothermal equilibrium denaturation data for all turn mutations used in the φ-value analysis of FGF-1 are given in Table I. FGF-1 is a weak mesophile as regards thermostability ($\Delta G_{\text{unfolding}}$ = 21.1 kJ/mol),[25,28] and certain turn mutations can destabilize the protein by a magnitude approaching the overall $\Delta G_{\text{unfolding}}$; thus, stabilizing background mutations were necessary for a subset of mutations. Of the 44 positions evaluated, 28 yielded $|\Delta\Delta G| \geq 2.0$ kJ/mol (i.e., $>3\sigma$ for $\Delta\Delta G$ measurements) and were therefore of sufficient magnitude to permit accurate φ-value analysis.[16] Complete folding and unfolding kinetic data were collected for this set of mutations (Table II). A comparison of $\Delta\Delta G$ values determined from isothermal equilibrium denaturation ($\Delta\Delta G_{\text{iso}}$) and folding kinetic data ($\Delta\Delta G_{\text{kin}}$) is in good agreement (Fig. 2, upper panel) supporting the two-state denaturation assumption.[25] A plot of the derived φ-values (Fig. 2, lower panel) indicates that FGF-1 has a highly polarized transition state, with the vast majority of evaluated positions having φ-values clustered around 1.0 or 0.0 (and are therefore either fully native-like, or denatured-like, in the folding transition state, respectively). Thus, a specific subset of turn positions in FGF-1 makes a critical contribution to the folding pathway, while others make little contribution. Despite the threefold internal symmetry characteristic of the β-trefoil architecture (involving three repeated trefoil-fold subdomains of 42–45 amino acids having a r.m.s.d. for main chain atoms of $\leq$1.3 Å[29]), the key regions identified for folding are *asymmetrically* distributed in the FGF-1 structure (Fig. 3, upper panel). For example, turns #1, #5, and #9 are related by the threefold internal symmetry; however, turn #1 is unstructured in the folding transition state, while turn #5 is natively structured, and turn #9 is partially structured. Notably, *there is no set of symmetry-related positions that fold concurrently.* The general regions that contribute to the folding transition state appear essentially contained within the second half of the first trefoil-fold subdomain and most of the second trefoil-fold subdomain (i.e., comprising ~50% of the protein structure). In contrast, the other regions (including essentially all the third trefoil-fold subdomain) contribute little to the folding transition state (and therefore fold late in the folding pathway).

Discussion

Heparin and membrane-bound heparan (heparan sulfate proteoglycan; HSPG), with a large negative charge density, are known to bind FGF-1 with high affinity.[34,35] The role of HSPG binding appears crucial for the proper biological function of FGF-1. Increased tissue levels of HSPG restrict the distribution of signaling molecules, such as FGF-1, can regulate concentration gradients of such signaling molecules and may play a role in pattern formation in embryogenesis; conversely, decreased levels of HSPG in vasculature promote long range transport of such signaling molecules.[36] Thus, HSPG binding is postulated to be a key determinant of the pharmacokinetic properties of FGF-1.[37–39] Furthermore, the competent signal transduction complex of FGF-1 involves a ternary interaction between FGF-1, FGF receptor, and HSPG.[30–33] The addition of soluble heparin to FGF-1 confers resistance to thermal denaturation, chemical denaturation, and proteolysis,[28,40,41] and inclusion of heparin in the formulation of FGF-1 greatly improves its potency, stability, storage, and reconstitution properties.[28] Thus, heparin binding represents a key functionality that regulates the tissue distribution, pharmacokinetics, and receptor signaling of FGF-1.

A large body of published work, involving numerous investigators and wide-ranging methodologies, has unambiguously identified amino acid positions associated with heparin-binding and receptor-binding functionality in FGF-1. For example, there are nine different molecular structures (seven X-ray and two NMR) of FGF-1 in complex with receptor and/or heparin analogues (PDB accession 1EVT, 1DJS, 1E0O, 1RY7, 2ERM, 1HKN, 1RML, 1AFC, and 1AXM) that serve to identify structural details of the regions of FGF-1 associated with both receptor and heparin binding. A large number of functional studies (involving chemical modification, point mutations, deletion mutations, homologous substitution mutations, and peptide-binding competition studies) in combination with analytical ultracentrifugation, surface plasmon resonance, and affinity chromatography validate the above structural data.[42–46] Heparin binds a specific cluster of basic residues with positive charge density and comprised mostly from the first β-hairpin and the last two-thirds of the third trefoil-fold subdomain (Fig. 3, lower panel). Furthermore, within such regions, heparin-binding and receptor-binding functionalities are associated with local structural deviations from ideal threefold symmetry. For example, positions 120–122 in turn #11 (which contribute significantly to heparin-binding functionality[46]) represent a structural insertion in comparison with the symmetry-related turns #3 and #7 (which are not involved in heparin-binding function). Additionally, residue positions 104–106 in turn #9 (which contribute to the low-affinity

Table I. *Isothermal Equilibrium Denaturation Data for FGF-1 Turn Mutants*

Protein	Turn no.	ΔG (kJ/mol)	m-Value (kJ/mol M)	C_m (M)	$\Delta\Delta G$ (kJ/mol)
FGF-1[a]		21.1 ± 0.6	18.9 ± 0.6	1.11 ± 0.01	
FGF-1/K12V/C117V/P134V[b]		37.6 ± 0.6	18.5 ± 0.3	2.03 ± 0.01	
FGF-1/H93G[c]		28.6 ± 0.1	19.8 ± 0.1	1.45 ± 0.01	
S17A	1	26.6 ± 0.1	19.8 ± 0.1	1.34 ± 0.01	−4.5
N18A	1	16.7 ± 0.1	18.0 ± 0.2	0.93 ± 0.00	3.5
G19A	1	16.8 ± 0.2	20.0 ± 0.4	0.84 ± 0.01	5.5
G20A	1	12.9 ± 0.2	17.8 ± 0.3	0.72 ± 0.01	7.3
L26A	2	15.7 ± 0.1	18.4 ± 0.1	0.85 ± 0.01	5.0
P27A	2	16.8 ± 0.4	19.9 ± 0.5	0.84 ± 0.02	5.3
D28A	2	12.5 ± 0.4	18.0 ± 0.4	0.69 ± 0.01	7.8
G29A[d]	2	29.8 ± 0.6	19.9 ± 0.7	1.50 ± 0.02	10.3
D36A	3	14.2 ± 0.1	18.3 ± 0.1	0.78 ± 0.00	6.3
R37A	3	21.2 ± 0.4	19.4 ± 0.2	1.09 ± 0.02	0.5
S38A	3	19.1 ± 0.1	18.3 ± 0.2	1.04 ± 0.01	1.5
D39A	3	15.3 ± 0.2	17.3 ± 0.3	0.88 ± 0.01	4.2
E49A	4	22.0 ± 0.3	19.9 ± 0.4	1.11 ± 0.01	0.2
S50A	4	21.6 ± 0.5	18.9 ± 0.4	1.14 ± 0.01	−0.3
V51A	4	21.0 ± 0.5	18.8 ± 0.6	1.12 ± 0.01	0.0
G52A	4	15.3 ± 0.4	18.7 ± 0.4	0.82 ± 0.01	5.6
T59A	5	21.7 ± 0.3	19.3 ± 0.3	1.13 ± 0.01	−0.1
E60A	5	26.3 ± 0.1	20.5 ± 0.2	1.28 ± 0.01	−3.3
T61A	5	14.6 ± 0.1	20.9 ± 0.3	0.70 ± 0.01	8.3
G62A[e]	5	15.1 ± 0.7	19.2 ± 0.9	0.79 ± 0.01	6.1
D68A	6	15.7 ± 0.4	18.1 ± 0.1	0.87 ± 0.02	4.6
T69A	6	22.8 ± 0.4	17.9 ± 0.2	1.27 ± 0.02	−2.8
D70A[f]	6	28.0 ± 0.8	19.3 ± 0.3	1.45 ± 0.02	−0.1
G71A[d,g]	6	16.9 ± 0.7	17.7 ± 0.8	0.96 ± 0.02	19.4
N80A	7	16.5 ± 0.2	19.7 ± 0.3	0.84 ± 0.01	5.4
E81A	7	18.0 ± 0.6	17.1 ± 0.6	1.06 ± 0.01	1.1
E82A	7	19.2 ± 0.1	18.5 ± 0.2	1.04 ± 0.01	1.6
C83A	7	18.1 ± 0.5	21.5 ± 0.6	0.84 ± 0.00	5.5
E90A	8	22.2 ± 0.1	20.2 ± 0.1	1.10 ± 0.01	0.4
E91A	8	22.7 ± 0.3	19.4 ± 0.3	1.17 ± 0.01	−0.9
N92A	8	17.5 ± 0.3	18.9 ± 0.1	0.93 ± 0.01	3.6
H93A[e]	8	19.3 ± 0.1	20.4 ± 0.3	0.95 ± 0.02	3.1
K100A	9	21.0 ± 0.3	19.0 ± 0.3	1.11 ± 0.01	0.3
K101A	9	17.7 ± 0.2	15.9 ± 0.2	1.11 ± 0.01	0.0
H102A	9	12.6 ± 0.4	17.1 ± 0.4	0.74 ± 0.01	6.8
A103G[e]	9	19.4 ± 0.6	19.4 ± 0.3	1.00 ± 0.01	2.1
K112A[f]	10	30.0 ± 0.9	19.9 ± 0.6	1.50 ± 0.02	−1.3
K113A[f]	10	28.1 ± 0.4	18.5 ± 0.4	1.52 ± 0.01	−1.3
N114A[f]	10	17.4 ± 0.6	17.0 ± 0.4	1.02 ± 0.01	7.7
G115A[d]	10	21.7 ± 0.8	16.6 ± 0.5	1.31 ± 0.02	12.7
H124A	11	21.9 ± 0.1	18.5 ± 0.1	1.19 ± 0.00	−1.2
Y125A	11	12.5 ± 0.4	17.9 ± 0.3	0.70 ± 0.02	7.7
G126A	11	16.3 ± 0.2	20.1 ± 0.3	0.81 ± 0.00	6.0
Q127A	11	15.0 ± 0.3	20.2 ± 0.3	0.74 ± 0.00	7.3

[a] Refs. 25–27.
[b] Ref. 27.
[c] Ref. 26.
[d] Reference protein is FGF-1/K12V/C117V/P134V.
[e] Ref. 26.
[f] Reference protein is FGF-1/H93G.[24]
[g] Thermodynamic parameters measured by CD due to atypical fluorescence signal.

receptor-binding site) are another apparent structural insertion in comparison with the symmetry-related turns #1 and #5.[31,43] Previous studies on forms of FGF-1 mutated to have increased primary sequence symmetry have observed a significant stability–function tradeoff in this region: deletion of the insertions at positions 104–106 and 120–122 (part of turns #9 and #11, respectively) increases protein stability by a substantial 16 kJ/mol (i.e., increasing by 50% the $\Delta G_{unfolding}$ of the protein) but diminish heparin-binding affinity (i.e., K_D for sucrose octasulfate) by an order of magnitude.[46] Notably, deletion of residues in these heparin-binding regions increases the folding rate constant by a factor of 20, while the

Table II. *Folding and Unfolding Kinetic Data for FGF-1 Turn Mutantsx*

Protein	Turn no.	k_f (s^{-1})	m_f (kJ/mol M)	k_u (1×10^{-6} s^{-1})	m_u (kJ/mol M)	m-Value (kJ/mol M)	C_m (M)	$\Delta\Delta G$ (kJ/mol)	$\Delta\Delta G_u$ (kJ/mol)	$\Delta\Delta G_f$ (kJ/mol)	φ_f	φ_u
FGF-1		3.74	−16.4	808	1.1	17.5	1.19					
FGF-1/K12V/ C117V/P134V		9.01	−13.0	58	2.2	15.2	1.57					
FGF1/H93G		55.1	−18.4	440	1.2	19.6	1.49					
S17A	1	2.65	−15.9	162	1.4	17.2	1.40	−3.5	3.7	0.2	−0.04	0.83
N18A	1	4.49	−17.5	2640	1.4	18.9	0.98	3.9	−3.2	0.7	0.20	0.93
G19A	1	0.96	−13.5	4370	1.5	15.0	0.89	4.9	−4.5	0.4	0.08	0.83
G20A	1	1.06	−16.9	3980	1.1	17.9	0.77	7.5	−3.9	3.6	0.49	0.54
L26A	2	1.10	−18.8	732	1.1	19.9	0.91	5.3	0.3	5.6	1.12	−0.06
P27A	2	0.57	−16.1	1170	1.1	17.2	0.89	5.2	−0.8	4.4	0.82	0.16
D28A	2	0.06	−14.8	767	1.1	15.9	0.68	8.6	0.2	8.8	1.12	−0.02
G29A[a]	2	0.90	−13.0	58	2.2	15.2	1.57	7.7	0.9	8.6	0.83	−0.08
D36A	3	0.24	−15.2	992	1.1	16.2	0.84	6.0	−0.4	5.5	0.88	0.07
D39A	3	0.42	−15.1	1034	1.1	16.2	0.92	4.6	−0.6	4.0	0.95	0.14
G52A	4	0.57	−17.6	895	1.0	18.6	0.86	6.0	−0.2	5.9	1.05	0.03
E60A	5	5.26	−15.4	837	1.1	16.6	1.31	−2.0	−0.1	−2.1	0.63	−0.03
T61A	5	0.21	−16.9	865	1.2	18.1	0.75	7.8	−0.2	7.6	0.92	0.02
G62A[b]	5	0.96	−18.7	820	1.2	14.0	0.88	6.1	−0.1	5.8	0.92	0.01
D68A	6	2.24	−18.6	1937	1.3	20.0	0.88	6.0	−2.4	3.6	0.77	0.52
T69A[c]	6	6.20	−16.1	1414	1.2	17.3	1.20	−1.5	−1.5	−1.6	0.56	−0.53
G71A	6	0.12	−7.7	109	3.4	11.1	1.05	14.0	−8.0	5.9	0.30	0.42
N80A	7	0.17	−15.1	678	1.3	16.3	0.84	6.0	0.3	6.3	1.17	−0.06
C83A	7	0.32	−15.7	808	1.2	16.9	0.88	5.4	−0.1	5.3	0.97	0.01
N92A	8	0.90	−16.4	1507	1.1	17.4	0.91	5.0	−1.5	3.5	0.97	0.42
H93A[b]	8	2.43	−17.8	117	1.0	18.9	1.00	3.1	−0.8	2.6	0.79	0.24
H102A	9	0.57	−17.0	186	1.1	18.1	0.79	7.2	−2.0	5.2	0.76	0.30
A103G[b]	9	2.21	−14.4	172	1.1	15.6	1.14	2.1	−1.9	−1.0	−0.44	0.84
N114A[d]	10	6.23	−12.2	172	2.2	14.4	1.12	7.7	−8.9	−2.7	−0.35	1.15
G115A[a]	10	23.4	−12.2	915	4.9	17.0	1.48	9.7	−10.5	−0.9	−0.07	0.83
Y125A	11	0.57	−19.2	1457	1.0	20.2	0.73	8.8	−1.4	7.4	0.96	0.18
G126A	11	0.48	−11.7	5928	1.1	12.8	0.85	5.2	−4.9	0.3	0.04	0.82
Q127A	11	2.21	−15.3	7891	1.5	16.8	0.83	6.2	−6.0	0.3	0.03	0.82

[a] Background protein is FGF-1/K12V/C117V/P134V.
[b] Ref. 26.
[c] Excluded from φ-value analysis due to non-two-state behavior.
[d] Background protein is FGF-1/H93G.[24]

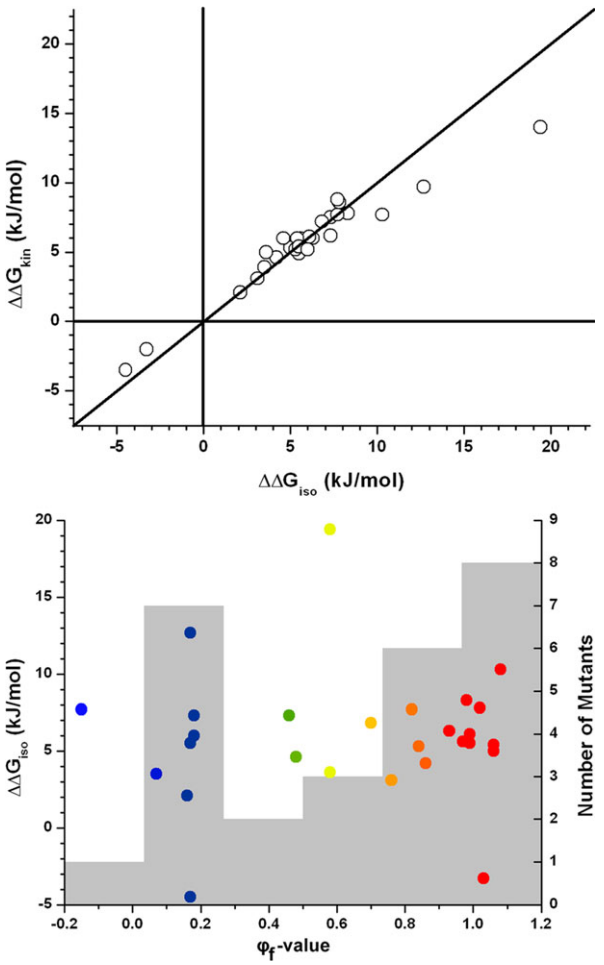

Figure 2. Top panel: a plot of mutant $\Delta\Delta G_{iso}$ (Table I) versus $\Delta\Delta G_{kin}$ (Table II) values. The good agreement between these $\Delta\Delta G$ values supports the two-state denaturation model.[25] Bottom panel, foreground: $\Delta\Delta G_{iso}$ values plotted against φ_f-values for turn mutations. φ_f-values were calculated according to the equation $\varphi_f = 1 - \varphi_u$, due to the greater accuracy with which unfolding kinetic parameters can be measured for FGF-1. The φ_f-values are colored using a heat map ranging from blue to red (see Fig. 3) and indicate whether the position has denatured-like or native-like structure in the folding transition state, respectively. Bottom panel, background shading: the histogram of φ_f-value distribution is bimodal with a peak around $\varphi_f = 0.15$ (unfolded-like) and $\varphi_f = 1$ (folded-like). These data are consistent with a polarized transition state.

unfolding rate constant is largely unaffected.[46] Thus, heparin-binding functionality is accommodated at the expense of thermostability, and the kinetic basis of such thermostability is principally upon foldability (i.e., folding kinetics).

Strikingly, the regions associated with heparin affinity are, without exception, observed to be denatured-like in the folding transition state (Fig. 3). Sites known to support receptor binding and positions identified as being denatured in the transition state are largely coincident, with limited exceptions. Notably, turns related by symmetry to the heparin-binding site, but do not participate in heparin-binding or receptor-binding functionality, are observed to be folded in the transition state (e.g., turn 5 compared to turns #1 or #9; turn #2 compared to turn #10 and turn #7 compared to turn #11), critically demonstrating that the inherent architecture of these positions is compatible with foldability. On the basis of the above-mentioned data, we conclude that formation of the heparin-binding site, and the majority of the receptor-binding site, necessitates specific physicochemical properties (i.e., positive charge density repulsion or strain associated with structural insertions that provide for molecular recognition), which preclude such regions from also efficiently participating in protein folding nucleation. The folding nucleus of FGF-1 appears largely localized to the second trefoil-fold subdomain; a region of the protein with a role that appears principally structural rather than functional. Taken together, our results show a general asymmetric segregation (despite the

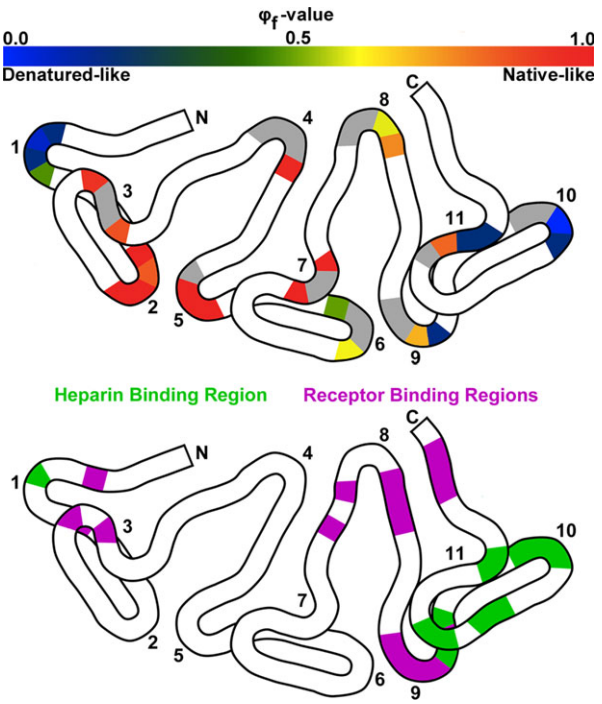

Figure 3. Upper panel: φ_f-value heat map for reverse turn regions in FGF-1 plotted upon a flattened ribbon diagram of the β-trefoil architecture. φ_f-value colors range from blue to red and indicate whether the position has denatured-like or native-like structure in the folding transition state, respectively. Light gray indicates positions where $|\Delta\Delta G| <$ 2.0 kJ/mol, and therefore φ_f cannot be accurately determined. As mentioned earlier, φ_f-values were calculated using φ_u according to the equation $\varphi_f = 1 - \varphi_u$. Lower panel: similar representation of FGF-1 where light green corresponds to those positions involved in heparin binding and magenta denotes positions critical for receptor binding and determined from X-ray structure data of the FGF-1/FGF receptor/heparin fragment ternary complex.[30–33]

threefold symmetric tertiary structure) of critical folding and functional elements in FGF-1, thus supporting the foldability–function tradeoff hypothesis.

Seven of 11 total turn regions in FGF-1 are observed by φ-value analysis to possess native structure in the folding transition state, a result that underpins the importance of turn formation early in the folding of FGF-1. As highlighted by Meiering and coworkers,[47] the folding of FGF-1 was thought to differ from the folding pathways of both hisactophilin and IL-1β. The basis for this conclusion was a report that probed the folding of FGF-1 using NMR H/D exchange studies and identified the association of the N- and C-termini as the first step in the folding pathway.[48] The φ-value analysis presented here is inconsistent with this proposed folding pathway: both turns #1 and #11 are observed to fold after barrier crossing, and neither turn is part of the folding nucleus (which approximately spans turns #2–#7). In the absence of significant residual structure, termini closure as the first step in folding seems unlikely from an entropic standpoint, as it would require association of the two most distal structural elements of the protein. Indeed, the success of relative contact order,[49] a measure of topological complexity, to predict folding rates is based on the preference of nearest-neighbor interactions to form first, which then support the formation of more distal interactions (so called, "zipping and assembly"[50]). Finally, refinement of anisotropic displacement parameters from a 1.10 Å X-ray diffraction dataset suggests that the N- and C-terminal regions of FGF-1 do not form a rigid body; instead, the N- and C-termini appear to be sliding past one another, suggestive of a tenuous interaction (even when stabilized by the folded structure).[51] Thus, our analysis of the folding of FGF-1 is inconsistent with the published report of Yu and coworkers,[48] but is in general agreement with the folding studies of other β-trefoil proteins in which the central strands of the protein are observed to fold faster than those on the periphery.[47,52,53]

Only a key subset of structural elements (turns #2–#5 and turn #7, spanning ~50% of the overall protein) appears necessary to confer efficient foldability to FGF-1. The entire protein is not (and apparently does not need to be) optimized for foldability; the regions not contributing to formation of the folding transition state instead segregate to regions of HSPG and receptor-binding functionalities. In the case of FGF-1, the structural regions that are observed to fold late (i.e., have decreased foldability) do so not because of a structurally intrinsic inability to efficiently fold: in every case, at least one symmetry-related position is observed to be part of the folding nucleus (and folds early). Therefore, the inability of regions to efficiently fold appears due to differences in the primary structure or the presence of a symmetry-breaking structural insertion. Thus, if the requirement for function is relaxed, regions associated with functionality could instead be optimized for foldability (e.g., by substitution of the primary structure of the symmetry-related regions that comprise the folding nucleus). The resulting protein would exhibit exact primary sequence symmetry and contain a foldable structural element at each of the symmetry-related positions. It has been postulated that such a protein might exhibit inefficient folding (due to folding frustration involving regions of identical primary structure),[54,55] or conversely, might possess *highly redundant*, overlapping folding nuclei, thereby promoting folding co-operativity.[56] We note that a protein with folding pathway redundancy might, in principle, retain efficient folding despite a localized deleterious mutation due to compensation by the folding-competent symmetry-related positions. Recent reports of *de novo* designed purely symmetric β-trefoil proteins that efficiently fold and are hyperthermophile in stability[57–59] demonstrate that pure primary structure symmetry does not necessarily result in folding frustration and therefore supports the hypothesis of redundant foldability. If folding redundancy is a property of purely symmetric proteins (i.e., as would result from gene duplication and fusion replication errors[60–62]), it highlights a critical advantage of symmetric protein architecture over asymmetric architecture in protein evolution and *de novo* design; namely, an intrinsic ability to tolerate diverse functional mutation and retain efficient foldability.

Materials and Methods

Mutant construction and protein purification

Ala point mutations in FGF-1 for φ-value analysis were constructed using the Quikchange® site-directed mutagenesis method and wild-type FGF-1 as a template. Kinetic parameters and φ-values of several Ala mutations were previously reported using wild-type FGF-1, H93G, or K12V/C117V/P134V as a template depending on the requirement of a stabilizing background. Mutant protein expression and purification procedures were reported previously.[46] Purified mutant protein was exchanged into 20 m*M* *N*-(2-acetamido)iminodiacetic acid (ADA), 0.1*M* NaCl, and pH 6.6 ("ADA buffer") with the addition of 2 m*M* DTT. An extinction coefficient of $E_{280\ \mathrm{nm}}$ (0.1%, 1 cm) = 1.26[63] was used for FGF-1 and mutants thereof.

Isothermal equilibrium denaturation

Isothermal equilibrium denaturation by guanidine HCl (GuHCl) for Ala point mutations was performed as previously described[64] using fluorescence as the spectroscopic probe. Briefly, fluorescence data were collected on a Cary Eclipse fluorescence

spectrophotometer (Agilent Technology, Santa Clara, CA) equipped with a Pelletier controlled-temperature regulator at 298 K and using a 1.0-cm path length cuvette. About 5.0 μ*M* protein samples were equilibrated overnight in ADA buffer at 298 K in 0.1*M* increments of GuHCl. Samples were excited at 295 nm, and emission was measured from 304 to 500 nm. Scans were collected in triplicate, averaged, and buffer-subtracted. Isothermal equilibrium denaturation of the G71A mutation was measured by circular dichroism (CD) due to its abnormal fluorescence profile. CD data of 25 μ*M* samples were collected on a Jasco model 810 CD spectrophotometer (Jasco, Easton, MD) equipped with a Pelletier controlled-temperature regulator at 298 K and using a 1-mm path length cuvette. The unfolding process was monitored by quantifying the change in CD signal at 227 nm with increasing GuHCl. Data were analyzed using the general purpose nonlinear least-squares fitting program DataFit (Oakdale Engineering, Oakdale, PA) implementing a six-parameter, two-state model.[65] The effect of a given mutation upon the stability of the protein ($\Delta\Delta G$) was calculated by taking the difference between the midpoint of denaturation (C_m value) for reference and mutant proteins and multiplying by the average of the m values, as described by Pace and Scholtz[66] and where a negative value indicates that the mutation is stabilizing in relationship to the reference protein.

Folding/unfolding kinetic analysis

Folding and unfolding kinetic measurements of FGF-1 Ala point mutations followed previously described methods.[26] Briefly, denatured protein samples for folding kinetics measurements were prepared by adding GuHCl to 2.0*M* followed by overnight incubation to permit equilibration. All folding kinetic data were collected using an Applied Photophysics SX20 stopped-flow system (Applied Photophysics, Surrey, United Kingdom) at 298 K with excitation wavelength at 295 nm and emission at 350 nm. Folding was initiated by a 1:10 dilution of 40 μ*M* denatured protein into ADA buffer with denaturant concentrations increasing in increments of 0.05*M* up to the midpoint of denaturation as determined by isothermal equilibrium denaturation measurements. The data collection strategy was designed to span approximately five half-lives or >97% of the expected fluorescence signal change between the fully denatured and native states. Because of the comparatively slower kinetics, unfolding kinetics measurements for Ala point mutations were performed using manual mixing. Protein samples (~30 μ*M*) were dialyzed against ADA buffer overnight at 298 K. Unfolding was initiated by a 1:10 dilution into ADA buffer with a final GuHCl concentration of 1.5–5.5*M* in 0.5*M* increments. All unfolding data were collected using a Cary Eclipse fluorescence spectrophotometer (Agilent Technology) equipped with a Pelletier-controlled temperature unit at 298 K. Data collection times for each protein were designed so as to quantify the fluorescence signal over three to four half-lives or >93% of the total expected amplitude.

The kinetic rates and amplitudes versus denaturant concentration were calculated from the time-dependent change in fluorescence using a single exponential model (or double exponential model under low-GuHCl concentrations). Folding and unfolding rate constant data were fit to a global function describing the contribution of both rate constants to the observed kinetics as a function of denaturant (chevron plot) as described by Fersht[67]:

$$\ln(k_{obs}) = \ln\left(k_{f0}\ \exp(m_{kf}[D]) + k_{u0}\ \exp(m_{ku}[D])\right) \quad (1)$$

where k_{f0} and k_{u0} are the folding and unfolding rate constants, respectively, extrapolated to 0*M* denaturant and m_{kf} and m_{ku} are the slopes of the linear refolding and unfolding arms, respectively, of the chevron plot. Changes in the free energy barrier to folding, $\Delta\Delta G_f$, and unfolding, $\Delta\Delta G_u$, were calculated from the global fit of the kinetic data:

$$\Delta\Delta G_f = -RT\ \ln(k_f\ \text{ref}/k_f\ \text{mutant}) \quad (2)$$

$$\Delta\Delta G_u = -RT\ \ln(k_u\ \text{ref}/k_u\ \text{mutant}) \quad (3)$$

where k_f and k_u are calculated at the average midpoint of denaturation for the reference and mutant proteins.

φ-*Value analysis*

Alanine scanning was performed at each position of the 11 turns that comprise the β-trefoil architecture of FGF-1. Only mutants with a $|\Delta\Delta G| > 2.0$ kJ/mol were considered suitable for φ-value analysis (i.e., $>3\sigma$ in the error of the $\Delta\Delta G$ measurement). φ-Values were calculated following the procedure established by Fersht et al.[68]:

$$\varphi_f = \Delta\Delta G_f/\Delta\Delta G \quad (4)$$

$$\varphi_u = \Delta\Delta G_u/-\Delta\Delta G \quad (5)$$

where $\Delta\Delta G$ is derived from equilibrium experiments as described earlier. Note that all reported φ-values are at the average midpoint of denaturation for the mutant and reference proteins.[24]

Acknowledgment

The authors declare no competing interests that could undermine the objectivity or integrity of this work.

References

1. Clackson T, Wells JA (1995) A hot spot of binding energy in a hormone-receptor interface. Science 267: 383–386.

2. Herzberg O, Moult J (1991) Analysis of the steric strain in the polypeptide backbone of protein molecules. Proteins 11:223–229.
3. Warshel A (1978) Energetics of enzyme catalysis. Proc Natl Acad Sci USA 75:5250–5254.
4. Warshel A, Sussman F, Hwang JK (1988) Evaluation of catalytic free energies in genetically modified proteins. J Mol Biol 201:139–159.
5. Meiering EM, Serrano L, Fersht AR (1992) Effect of active site residues in barnase on activity and stability. J Mol Biol 225:585–589.
6. Schreiber G, Buckle AM, Fersht AR (1994) Stability and function: two constraints in the evolution of barstar and other proteins. Structure 2:945–951.
7. Shoichet BK, Baase WA, Kuroki R, Matthews BW (1995) A relationship between protein stability and protein function. Proc Natl Acad Sci USA 92:452–456.
8. Beadle BM, Shoichet BK (2002) Structural basis of stability—function tradeoffs in enzymes. J Mol Biol 321: 285–296.
9. Richards FM (1974) The interpretation of protein structures: total volume, group volume distributions and packing density. J Mol Biol 82:1–14.
10. Wang X, Minasov G, Shoichet BK (2002) Evolution of an antibiotic resistance enzyme constrained by stability and activity trade-offs. J Mol Biol 320:85–95.
11. Nagatani RA, Gonzalez A, Shoichet BK, Brinen LS, Babbitt PC (2007) Stability for function trade-offs in the enolase superfamily "catalytic module." Biochemistry 46:6688–6695.
12. Tokuriki N, Tawfik DS (2009) Stability effects of mutations and protein evolvability. Curr Opin Struct Biol 19:596–604.
13. Capraro DT, Roy M, Onuchic JN, Jennings PA (2008) Backtracking on the folding landscape of the beta-trefoil protein interleukin-1beta? Proc Natl Acad Sci USA 105:14844–14848.
14. Gosavi S, Chavez LL, Jennings PA, Onuchic JN (2006) Topological frustration and the folding of interleukin-1β. J Mol Biol 357:986–996.
15. Gosavi S, Whitford PC, Jennings PA, Onuchic JN (2008) Extracting function from a beta-trefoil folding motif. Proc Natl Acad Sci USA 105:10384–10389.
16. Fersht AR, Sato S (2004) ϕ-Value analysis and the nature of protein-folding transition states. Proc Natl Acad Sci USA 101:7976–7981.
17. Serrano L, Matouschek A, Fersht AR (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.
18. Goldenberg DP, Frieden RW, Haack JA, Morrison TB (1989) Mutational analysis of a protein-folding pathway. Nature 338:127–132.
19. Murzin AG, Lesk AM, Chothia C (1992) β-Trefoil fold. Patterns of structure and sequence in the kunitz inhibitors interleukins-1β and 1α and fibroblast growth factors. J Mol Biol 223:531–543.
20. McLachlan AD (1979) Three-fold structural pattern in the soybean trypsin inhibitor (Kunitz). J Mol Biol 133: 557–563.
21. Jager M, Nguyen H, Crane JC, Kelly JW, Gruebele M (2001) The folding mechanism of a beta-sheet: the WW domain. J Mol Biol 311:373–393.
22. Marcelino AM, Gierasch LM (2008) Roles of beta-turns in protein folding: from peptide models to protein engineering. Biopolymers 89:380–391.
23. Petrovich M, Jonsson AL, Ferguson N, Daggett V, Fersht AR (2006) Phi-analysis at the experimental limits: mechanism of beta-hairpin formation. J Mol Biol 360:865–881.
24. Lee J, Dubey VK, Longo LM, Blaber M (2008) A logical OR redundancy with the Asx-Pro-Asx-Gly type I β-turn motif. J Mol Biol 377:1251–1264.
25. Blaber SI, Culajay JF, Khurana A, Blaber M (1999) Reversible thermal denaturation of human FGF-1 induced by low concentrations of guanidine hydrochloride. Biophys J 77:470–477.
26. Kim J, Brych SR, Lee J, Logan TM, Blaber M (2003) Identification of a key structural element for protein folding within β-hairpin turns. J Mol Biol 328:951–961.
27. Dubey VK, Lee J, Somasundaram T, Blaber S, Blaber M (2007) Spackling the crack: stabilizing human fibroblast growth factor-1 by targeting the N and C terminus beta-strand interactions. J Mol Biol 371:256–268.
28. Copeland RA, Ji H, Halfpenny AJ, Williams RW, Thompson KC, Herber WK, Thomas KA, Bruner MW, Ryan JA, Marquis-Omer D, Sanyal G, Sitrin RD, Yamazaki S, Middaugh CR (1991) The structure of human acidic fibroblast growth factor and its interaction with heparin. Arch Biochem Biophys 289:53–61.
29. Blaber M, DiSalvo J, Thomas KA (1996) X-ray crystal structure of human acidic fibroblast growth factor. Biochemistry 35:2086–2094.
30. Pantoliano MW, Horlick RA, Springer BA, Van Dyk DE, Tobery T, Wetmore DR, Lear JD, Nahapetian AT, Bradley JD, Sisk WP (1994) Multivalent ligand-receptor binding interactions in the fibroblast growth factor system produce a cooperative growth factor and heparin mechanism for receptor dimerization. Biochemistry 33:10229–10248.
31. Pellegrini L, Burke DF, von Delft F, Mulloy B, Blundell TL (2000) Crystal structure of fibroblast growth factor receptor ectodomain bound to ligand and heparin. Nature 407:1029–1034.
32. Schlessinger J, Plotnikov AN, Ibrahimi OA, Eliseenkova AV, Yeh BK, Yayon A, Linhardt RJ, Mohammadi M (2000) Crystal structure of a ternary FGF-FGFR-heparin complex reveals a dual role for heparin in FGFR binding and dimerization. Mol Cell Biol 6: 743–750.
33. Hecht HJ, Adar R, Hofmann B, Bogin O, Weich H, Yayon A (2001) Structure of fibroblast growth factor 9 shows a symmetric dimer with unique receptor- and heparin-binding interfaces. Acta Cryst D57:378–384.
34. Gospodarowicz D, Cheng J, Lui G-M, Baird A, Bohlen P (1984) Isolation by heparin-sepharose affinity chromatography of brain fibroblast growth factor: identity with pituitary fibroblast growth factor. Proc Natl Acad Sci USA 81:6963–6967.
35. Lobb RR, Fett JW (1984) Purification of two distinct growth factors from bovine neural tissue by heparin affinity chromatography. Biochemistry 23:6295–6299.
36. Hacker U, Nybakken K, Perrimon N (2005) Heparan sulphate proteoglycans: the sweet side of development. Nat Rev 6:530–541.
37. Rosengart TK, Kuperschmid JP, Maciag T, Clark RE (1989) Pharmacokinetics and distribution of heparin-binding growth factor-1 (endothelial cell growth factor) in the rat. Circul Res 64:227–234.
38. Hondermarck H, Courty J, Boilly B, Thomas D (1990) Distribution of intravenously administered acidic and basic fibroblast growth factors in mouse. Experientia 46:973–974.
39. Zinn KR, Kelpke S, Chaudhuri TR, Sugg T, Mountz JM, Thompson JA (2000) Imaging Tc-99m-labeled FGF-1 targeting in rats. Nucl Med Biol 27:407–414.

40. Gospodarowicz D, Cheng J (1986) Heparin protects basic and acidic FGF from inactivation. J Cell Physiol 128:475–484.
41. Rosengart TK, Johnson WV, Friesel R, Clark R, Maciag T (1988) Heparin protects heparin-binding growth factor-1 from proteolytic inactivation *in vitro*. Biochem Biophys Res Commun 152:432–440.
42. Harper JW, Lobb RR (1988) Reductive methylation of lysine residues in acidic fibroblast growth factor: effect on mitogenic acitivity and heparin affinity. Biochemistry 27:671–678.
43. Springer BA, Pantoliano MW, Barbera FA, Gunyuzlu PL, Thompson LD, Herblin WF, Rosenfeld SA, Book GW (1994) Identification and concerted function of two receptor binding surfaces on basic fibroblast growth factor required for mitogenesis. J Biol Chem 269: 26879–26884.
44. Seddon AP, Aviezer D, Li L-Y, Bohlen P, Yayon A (1995) Engineering of fibroblast growth factor: alteration of receptor binding specificity. Biochemistry 34: 731–736.
45. Patrie KM, Botelho MJ, Franklin K, Chiu IM (1999) Site-directed mutagenesis and molecular modeling identify a crucial amino acid in specifying the heparin affinity of FGF-1. Biochemistry 38:9264–9272.
46. Brych SR, Dubey VK, Bienkiewicz E, Lee J, Logan TM, 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.
47. Liu C, Gaspar JA, Wong HJ, Meiering EM (2002) Conserved and nonconserved features of the folding pathway of hisactophilin, a β-trefoil protein. Protein Sci 11: 669–679.
48. Samuel D, Kumar TKS, Balamurugan K, Lin W-Y, Chin D-H, Yu C (2001) Structural events during the refolding of an all β-sheet protein. J Biol Chem 276: 4134–4141.
49. Plaxco KW, Simons KT, Baker D (1998) Contact order, transition state placement and the refolding rates of single domain proteins. J Mol Biol 277:985–994.
50. Dill KA, Ozkan SB, Shell MS, Weikl TR (2008) The protein folding problem. Ann Rev Biophys 37:289–316.
51. Bernett MJ, Somasundaram T, Blaber M (2004) An atomic resolution structure for human fibroblast growth factor 1. Proteins 57:626–634.
52. Varley P, Gronenborn AM, Christensen H, Wingfield PT, Pain RH, Clore GM (1993) Kinetics of folding of the all-β sheet protein interleukin-1β. Science 260: 1110–1113.
53. Heidary DK, Gross LA, Roy M, Jennings PA (1997) Evidence for an obligatory intermediate in the folding of Interleukin-1β. Nat Struct Biol 4:725–731.
54. Wright CF, Teichmann SA, Clarke J, Dobson CM (2005) The importance of sequence diversity in the aggregation and evolution of proteins. Nature 438: 878–881.
55. Borgia MB, Borgia A, Best RB, Steward A, Nettels D, Wunderlich BS, Clarke J (2011) Single-molecule fluorescence reveals sequence-specific misfolding in multidomain proteins. Nature 474:662–665.
56. Haglund E, Lindberg MO, Oliveberg M (2008) Changes in protein folding pathways by circular permutation: overlapping nuclei promote global cooperativity. J Biol Chem 283:27904–27915.
57. Lee J, 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.
58. Lee J, Blaber SI, Dubey VK, Blaber M (2011) A polypeptide "building block" for the β-trefoil fold identified by "top-down symmetric deconstruction." J Mol Biol 407:744–763.
59. Broom A, Doxey AC, Lobsanov YD, Berthin LG, Rose DR, Howell PL, McConkey BJ, Meiering EM (2012) Modular evolution and the origins of symmetry: reconstruction of a three-fold symmetric globular protein. Structure 20:1–11.
60. Ohno S (1970) Evolution by gene duplication. New York: Allen and Unwin.
61. McLachlan AD (1972) Repeating sequences and gene duplication in proteins. J Mol Biol 64:417–437.
62. Tang J, James MN, Hsu IN, Jenkins JA, Blundell TL (1978) Structural evidence for gene duplication in the evolution of the acid proteases. Nature 271:618–621.
63. Brych SR, Blaber SI, Logan TM, Blaber M (2001) Structure and stability effects of mutations designed to increase the primary sequence symmetry within the core region of a β-trefoil. Protein Sci 10:2587–2599.
64. Brych SR, Kim J, Logan TM, Blaber M (2003) Accommodation of a highly symmetric core within a symmetric protein superfold. Protein Sci 12:2704–2718.
65. Eftink MR (1994) The use of fluorescence methods to monitor unfolding transitions in proteins. Biophys J 66:482–501.
66. Pace CN, Scholtz JM (1997) Measuring the conformational stability of a protein. In: Creighton TE, editor. Protein structure: a practical approach. Oxford: Oxford University Press, pp 299–321.
67. Fersht AR (1999) Kinetics of protein folding. New York: W.H. Freeman & Co.
68. Fersht AR, Matouschek A, Serrano L (1992) The folding of an enzyme. I. Theory of protein engineering analysis of stability and pathway of protein folding. J Mol Biol 224:771–782.