Structure and dynamics of de novo proteins from a designed superfamily of 4-helix bundles

ABIGAIL GO,[1] SEHO KIM,[2] JEAN BAUM,[2,3] AND MICHAEL H. HECHT[1]

[1]Department of Chemistry, Princeton University, Princeton, New Jersey 08544, USA
[2]Department of Chemistry and Chemical Biology, Rutgers University, Piscataway, New Jersey 08854, USA
[3]BioMaPS Institute for Quantitative Biology, Rutgers University, Piscataway, New Jersey 08854, USA

(RECEIVED November 30, 2007; FINAL REVISION February 29, 2008; ACCEPTED March 3, 2008)

Abstract

Libraries of de novo proteins provide an opportunity to explore the structural and functional potential of biological molecules that have not been biased by billions of years of evolutionary selection. Given the enormity of sequence space, a rational approach to library design is likely to yield a higher fraction of folded and functional proteins than a stochastic sampling of random sequences. We previously investigated the potential of library design by binary patterning of hydrophobic and hydrophilic amino acids. The structure of the most stable protein from a binary patterned library of de novo 4-helix bundles was solved previously and shown to be consistent with the design. One structure, however, cannot fully assess the potential of the design strategy, nor can it account for differences in the stabilities of individual proteins. To more fully probe the quality of the library, we now report the NMR structure of a second protein, S-836. Protein S-836 proved to be a 4-helix bundle, consistent with design. The similarity between the two solved structures reinforces previous evidence that binary patterning can encode stable, 4-helix bundles. Despite their global similarities, the two proteins have cores that are packed at different degrees of tightness. The relationship between packing and dynamics was probed using the Modelfree approach, which showed that regions containing a high frequency of chemical exchange coincide with less well-packed side chains. These studies show (1) that binary patterning can drive folding into a particular topology without the explicit design of residue-by-residue packing, and (2) that within a superfamily of binary patterned proteins, the structures and dynamics of individual proteins are modulated by the identity and packing of residues in the hydrophobic core.

Keywords: binary patterning; NMR spectroscopy; heteronuclear NMR; 4-helix bundle; protein design; de novo

Supplemental material: see www.proteinscience.org

The structures and dynamics of proteins are dictated by the physical chemistry of the polypeptide sequence interacting with itself and with the surrounding solvent (Scheraga 1970; Anfinsen 1972; Willis et al. 2000). For natural proteins, the structures and dynamics that are ''allowed'' are also constrained by the biological requirements of the host organism. Proteins in present day organisms also reflect the biological and environmental factors that influenced the selection of ancestral sequences through millions of years of evolutionary history. Thus, the properties observed in modern proteomes are biased both by current biology and past history.

Understanding the true potential of protein sequence space would benefit from studies of proteins that are neither required to sustain living organisms nor biased by ''artifacts'' associated with evolutionary history. In

Reprint requests to: Michael Hecht, Department of Chemistry, Princeton University, Princeton, NJ 08544, USA; e-mail: hecht@princeton.edu; fax: (609) 258-6746; or Jean Baum, Department of Chemistry and Chemical Biology, Rutgers University, Piscataway, NJ 08854, USA; e-mail: baum@rutchem.rutgers.edu; fax: (609) 258-6746.

Article and publication are at http://www.proteinscience.org/cgi/doi/10.1110/ps.073377908.

principle, an ideal bias-free collection of proteins would be a stochastic combinatorial collection of sequences constructed at random. However, the vast majority of random sequences do not fold into protein-like structures. Since most random sequences form insoluble aggregates (Mandecki 1990; Keefe and Szostak 2001; Watters and Baker 2004; Chiarabelli et al. 2006), a stochastic collection of sequences is *not* an appealing sample for assessing the structural and dynamic properties of an unevolved proteome. A more appropriate collection of sequences would be combinatorially diverse but would focus on those regions of sequence space that are consistent with folding into protein-like three-dimensional structures.

Building a collection of folded, but unselected sequences requires the use of a single, overarching approach. The patterning of polar and nonpolar amino acids has proven to be a powerful method to design protein structures (Kamtekar et al. 1993; Hecht et al. 2004) and may be used to build macromolecules comparable to early proteins (Lopez de la Osa et al. 2007). In previous work, we described a method that focuses combinatorial libraries into productive regions of sequence space and thereby facilitates the design and construction of vast collections of folded proteins (Hecht et al. 2004; Bradley et al. 2006). Our binary patterning method samples enormous sequence diversity; yet it favors proper folding by rigorously defining which positions in a sequence must be polar (and exposed to solvent), and which must be nonpolar (and buried in the interior). Binary patterning of polar and nonpolar residues specifies the target topology by directing the hydrophobic collapse of a sequence into the desired shape. For example, to specify an α-helical fold, the sequence periodicity of polar and nonpolar residues is designed to match the structural repeat of 3.6 residues per α-helical turn. A sequence of polar (○) and nonpolar (●) residues with the pattern ○●○○●●○ has a nonpolar amino acid every three or four positions and is consistent with the formation of an amphiphilic α-helix. Such an α-helix would contain a hydrophobic face, which would be buried in the final tertiary structure. Conversely, to specify a β-sheet fold, polar and nonpolar residues are designed to alternate every other residue. Thus, a designed sequence with the pattern ○●○●○●○ has a sequence periodicity that matches the structural repeat of amphiphilic β-strands. Such strands would bury their hydrophobic faces upon folding. The designed segments of α-helical and β-sheet secondary structure may then be connected with glycine-rich turns.

Implementation of the binary code strategy is enabled by the organization of the genetic code, with the degenerate codon VAN (V = A, G, or C; N = A, G, C, or T) encoding a mixture of polar residues (Lys, His, Glu, Gln, Asp, and Asn), and the degenerate codon NTN encoding a mixture of nonpolar residues (Met, Leu, Ile, Val, and Phe). By constructing a library of synthetic genes in which these two degenerate codons are used at defined locations in the sequence, the polarity of amino acids can be specified without explicit design of unique side chains at each site.

We previously reported the successful construction of several binary patterned libraries including all-α and all-β structures (Kamtekar et al. 1993; West et al. 1999; Xu et al. 2001; Wang and Hecht 2002; Wei et al. 2003b; Hecht et al. 2004; Bradley et al. 2007). The α-helical designs focused on the 4-helix bundle topology. Our first-generation library used a 74-residue template. All proteins purified from this library were soluble and α-helical, and several displayed cooperative folding (Kamtekar et al. 1993; Roy et al. 1997; Roy and Hecht 2000). Nonetheless, these first-generation proteins were not sufficiently ordered for structure determination by X-ray crystallography or NMR. We hypothesized that to favor well-ordered structures, it would be important to elongate the helices, thereby generating a larger number of hydrophobic contacts. This hypothesis was confirmed by constructing a second-generation library. We constructed this library by choosing one 74-residue molten globule sequence from the first-generation library and elongating the structure by adding two turns to each of its four α-helices. The strategy succeeded, and the second-generation library produced a majority of stable, monomeric, α-helical proteins with well-ordered hydrophobic cores (Wei et al. 2003b).

This second-generation library of 4-helix bundles provides an opportunity to assess the range of structural, dynamic, and functional properties that can be found in a superfamily of proteins that has not been constrained by biological evolution. Initial studies probing the functional capabilities of these de novo proteins demonstrated that some of them bind cofactors and exhibit low levels of enzymatic activity (Wei and Hecht 2004; Das et al. 2006; Das and Hecht 2007). The structural and dynamic properties of these de novo proteins are the focus of the current study.

Here we report the solution structure of the second-generation protein, S-836, and compare it with the structure of its sibling, S-824, which was determined previously (Wei et al. 2003a). We also determine the dynamic behavior of both proteins. The dynamics of de novo designed proteins is a relatively unexplored area of study with limited published research (Walsh et al. 1999). Comparison of the structure and dynamics of S-836 and S-824 would provide a window into the range of structural and dynamic behaviors that can be expected from a library of proteins that was designed ''from scratch'' and not subjected to the constraints of biological selection. In addition to their implications for the design of proteins de novo, these studies may contribute to understanding of the properties of pre-evolved ancestral proteins.

Results and Discussion

Protein S-836 forms a well-defined 4-helix bundle

The solution structure of protein S-836 was solved by NMR spectroscopy. The structure is an up-down-up-down 4-helix bundle connected by relatively short turns (Fig. 1A,B). The adjacent helices are not perfectly antiparallel. The slight tilt of these helices (~20°) relative to one another is typical of 4-helix bundles with ''knobs-in-holes'' packing (Crick 1953; Chothia et al. 1977; Harris et al. 1994).

The final, calculated structure of S-836 is well-resolved, with the 15 lowest energy structures showing a backbone root mean square deviation (RMSD) of 0.39 ± 0.05 Å relative to the mean (Table 1). When limited to the helical regions, this RMSD decreases to 0.32 ± 0.06 Å. Overall, low RMSD indicates that the protein tertiary structure is well-defined and well-ordered. The helical regions are better defined than the turn regions, with the latter showing more variability.

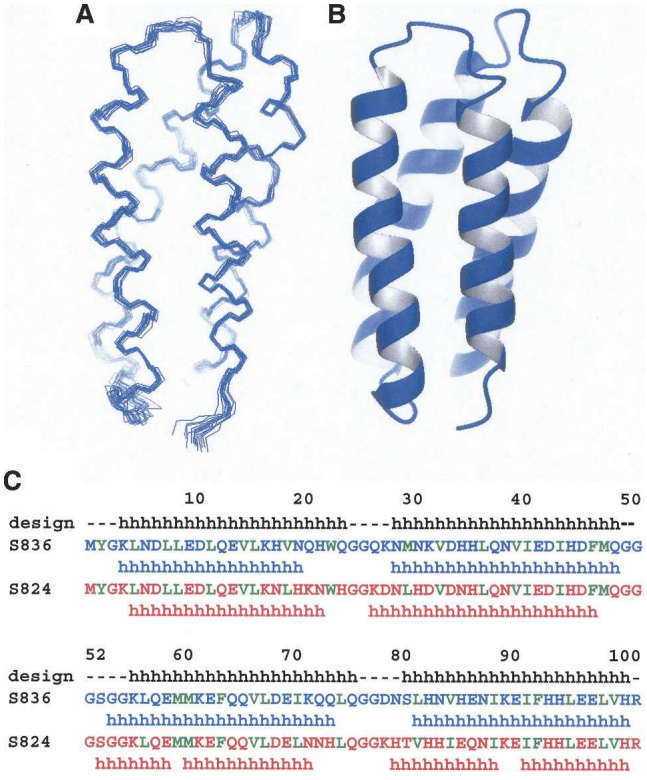

Figure 1. Helical backbone of protein S-836. (*A*) Line rendering of the 15 lowest energy structures. (*B*) Ribbon diagram of one representative structure. Both renderings show S-836 in the same orientation, with the N terminus in the foreground. (*C*) Sequence and secondary structure of protein S-836 compared with protein S-824 and with the original binary patterned design. ''h'' Indicates helical secondary structure. There is high sequence identity between S-836 and S-824. Differences in their primary structure occur at positions 18–35 and at positions 71–87. Helices were identified from solved structures using MOLMOL software (Koradi et al. 1996) and vary slightly from the design template. Residues that are nonpolar by binary patterning design are shown in green.

The overall topology of the bundle is left-turning (viewed from the outside, the chain turns left to traverse from helix 1 to helix 2). Right-turning and left-turning 4-helix bundles both occur frequently among natural proteins (Presnell and Cohen 1989), and the binary code strategy does not explicitly design for one topology versus the other. Thus, it is noteworthy that both S-824 (Wei et al. 2003a) and S-836 form left-turning bundles.

The helices are highly consistent with—but not identical to—those specified by the design template (Fig. 1C). The first and fourth helices are slightly shorter than expected from the design, and the third helix begins two residues earlier in the sequence, incorporating Gly54 and Gly55 into its N-terminal end. We surmise that inclusion of these glycines into helix 3 lengthens the inter-helical core, thereby providing space to accommodate the large hydrophobic side chains on neighboring helices (e.g., Phe47).

The hydrophobic core of protein S-836

The main premise of protein design by binary patterning is that hydrophobic collapse of strategically placed nonpolar residues—irrespective of their exact side-chain identities—is sufficient to drive the polypeptide chain to fold into a desired structure. Because the identities of the side chains are not defined a priori, the design strategy cannot specify the residue-by-residue packing of nonpolar residues in the hydrophobic core. Therefore a diversity of hydrophobic packing is expected in the different proteins in a binary code library. In the structure of S-836, most nonpolar residues are indeed buried in the core, and all polar residues are exposed to solvent (Fig. 2A,B). Heavy atoms in this core deviate from the mean structure by an average of 0.41 ± 0.10 angstroms, indicating that the tertiary structure is well-defined (Fig. 2C).

Although most of the nonpolar residues are fully buried, there are a number of notable exceptions. Several nonpolar side chains are only partially buried, and surprisingly, three of the four methionine side chains are completely exposed to solvent (Fig. 3A). The protein/solvent contact surface areas (PyMOL; DeLano Scientific) of Met30, Met48, and Met61 are comparable to polar α-helical amino acids of similar size. The exposure of these three methionine side chains may be attributed to the proximity of large aromatic side chains: Each of these exposed methionines shares a cross-sectional packing layer with either a phenylalanine or a tryptophan side chain (Fig. 3B). Furthermore, methionine is less hydrophobic than the other nonpolar residues utilized by the binary patterning design strategy (Wolfenden et al. 1981; Kyte and

Table 1. *Structural statistics for the 20 lowest energy structures*

NOE distance restraints	
Intraresidue (/i − j/ = 0)	828
Medium range (0</i − j/ < 5)	998
Long range (/i − j/ ≥ 5)	256
Φ angle restraints	84
Total	2166
Mean RMSD from restraints	
Distance restraints (Å)	0.010 ± 0.0005
Dihedral angles (°)	0.196 ± 0.066
Mean RMSD from ideal geometry	
Bonds (Å)	0.002
Angles (°)	0.250 ± 0.008
Impropers (°)	0.213 ± 0.007
RMSD from mean structure (Å)[a]	
All backbone	0.39 ± 0.05
All heavy atoms	0.91 ± 0.06
Backbone in helical region	0.32 ± 0.06
Heavy atoms in helical region	0.86 ± 0.07
Heavy atoms in hydrophobic core	0.41 ± 0.10
Ramachandran statistics from ProCheck (%)[b]	
Residues in most favored regions	83.1
Residues in additional allowed regions	16.9
Residues in generously allowed regions	0.0
Residues in disallowed regions	0.0
Ramachandran statistics from MolProbity (%)[c]	
Residues in favored regions	77.9 ± 3.2
Residues in allowed regions (includes favored regions)	97.2 ± 1.4

[a] Calculated using MOLMOL (Koradi et al. 1996).
[b] Calculated using Procheck (Laskowski et al. 1996).
[c] Calculated using Moprobity (Davis et al. 2007).

Doolittle 1982; Fauchère and Pliska 1983). Another factor that may favor exposure of methionine is the entropic advantage that is gained when this straight chain residue is freed from packing interactions in the interior. Exposing any of the other nonpolar side chains utilized in binary patterning—valine, leucine, isoleucine, and phenylalanine—would produce less of an entropic benefit. The observed ejection of the methionine side chains from the hydrophobic core indicates that steric hindrance takes precedence over the hydrophobicity of methionine, and suggests that the designed hydrophobic core of S-836 is overly packed: Indeed, the methionines may have been "pushed away" by the nearby packing of Phe47, Phe64, and Phe93 into a pi-stacking arrangement (Fig. 3C).

Similarities with the structure of protein S-824

Protein S-836 is one of five proteins that were biophysically characterized from a second-generation combinatorial library designed to produce 4-helix bundles (Wei et al. 2003b). All five were shown to be monomeric and α-helical. Thermodynamic and NMR characterization determined that four of the five proteins were stable and well-ordered. (The fifth, protein S-23, proved to be a molten globule.) Of these four well-folded proteins, the solution structures of two have now been solved: S-824 was reported previously (Wei et al. 2003a), and S-836 is presented here.

Initially, we chose to solve the structure of S-824 because it produced NMR spectra of extremely high quality: The $^{1}H,^{15}N$-HSQC spectrum of S-824 was highly dispersed, and the two-dimensional nuclear Overhauser enhancement spectra (2D-NOESY) showed many side-chain interactions indicative of a native-like tertiary structure. In contrast, NMR and thermodynamic studies of the other three proteins (S-213, S-285, and S-836) showed less dispersion, fewer NOE peaks, and less cooperative folding (Wei et al. 2003b). Therefore, for the current study, we chose to solve the structure of S-836 because it is more representative of the type of structure

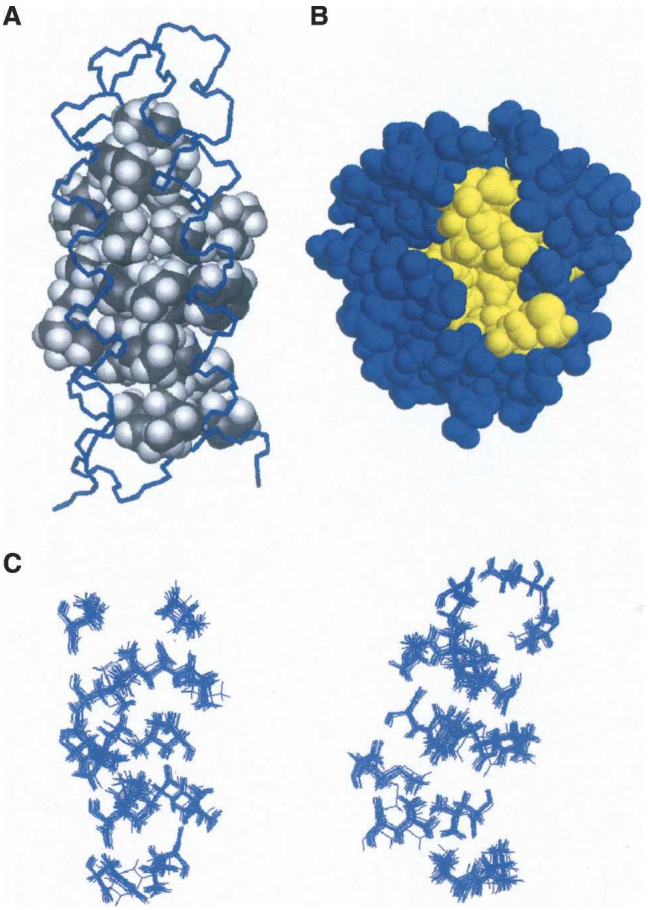

Figure 2. (*A*) The core of the S-836 consists of side chains of nonpolar amino acids. (*B*) In general, nonpolar amino acids (yellow) are buried and polar amino acids (blue) are exposed to solvent. (*C*) Nonpolar amino acids comprising the core exhibit well-defined structure. The figure shows nonpolar amino acids in the 15 lowest energy structures, with residues in helices 1 and 2 on the *left* and residues in helices 3 and 4 on the *right*. This well-ordered hydrophobic core indicates that S-836 forms a native-like structure.

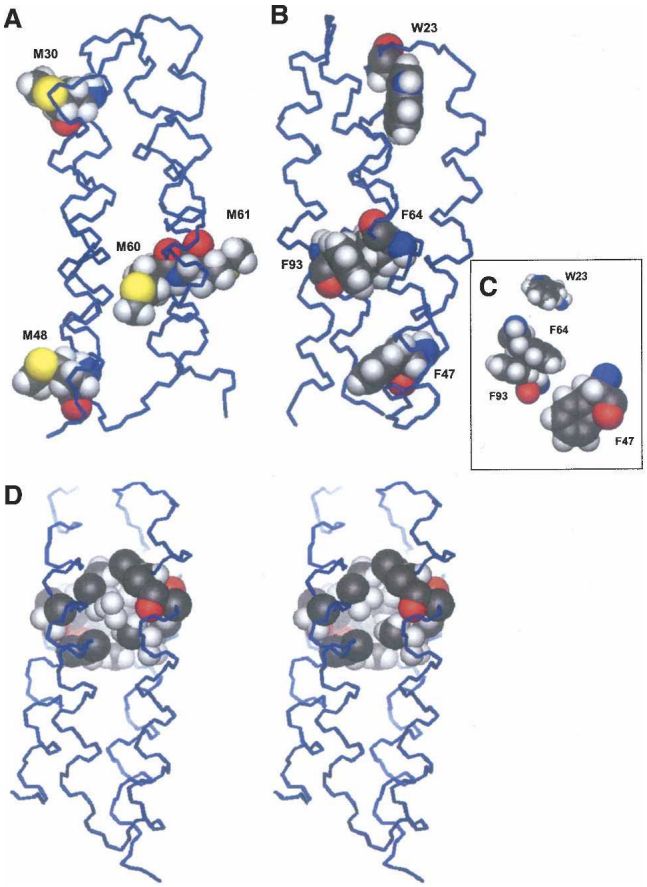

Figure 3. Methionines, aromatic residues, and the hydrophobic core of S-836. (*A*) All four methionines occur in helical regions of the protein. However, with the exception of M60, they do not contribute to the hydrophobic core, and their side chains are exposed to solvent. Their exposure might relieve overpacking of the core. (*B*) Bulky, highly hydrophobic aromatics near the methionines may sterically hinder packing of methionine side chains and thereby push the Met side chains out of the core. Panels *A* and *B* were rotated to better show the relevant amino acids. (*C*) Phe64 and Phe93 are also involved in pi-stacking interactions. (*D*) Protein S-836 has an interconnected network of small cavities in the hydrophobic core. This network is only marginally exposed to solvent. To enhance clarity, the cavity is rendered on the *right* without the methyl from Val67. The atoms that border this cavity are depicted. They are atoms from the following amino acids: V15, L16, V19, N20, W23, V33, D34, L37, V67, L68, I71, L82, V85, H86, and I89.

and dynamics that are likely to occur frequently in our library of de novo proteins.

The solution structures of S-836 and S-824 show a number of general similarities: Both proteins are up-down-up-down 4-helix bundles, consistent with their design. Both also exhibit the same left-turning topology, despite the lack of explicit design of this feature, and despite the more-or-less equal propensity in nature for left-turning and right-turning bundles (Presnell and Cohen 1989). The similarity in topology also extends to similar interhelical angles; therefore, both display knobs-in-holes packing in their hydrophobic cores. Moreover, as specified by the binary patterning design strategy, both structures partition their polar and nonpolar side chains to the surface and core, respectively. These similarities in structure are not unexpected, as S-836 and S-824 have very similar sequences. Based on these two structures, we expect that a large fraction (probably the majority) of proteins in this binary-patterned library fold into up-down-up-down 4-helix bundles with knobs-in-holes packing.

As described above, several methionine side chains in protein S-836 are exposed to solvent. Similar observations were made in the structure of S-824 (Wei et al. 2003a). This deviation from the binary code design, which is based on the burial of nonpolar residues, may indicate that the template used to encode the binary patterning of the current library encodes slightly too many nonpolar positions and could be improved in future libraries.

Cavities in the structure of protein S-836

The structures of S-824 and S-836 also differ in that the latter contains larger and more numerous cavities in its hydrophobic core. Solvent accessible cavities in both proteins were probed using CASTP (Dundas et al. 2006). With a probe radius of 1.4 Å, CASTP identified 11 cavities in S-836 and only six in S-824. Indeed, S-836 has a network of four interconnected cavities, ranging from 12 to 24 $Å^3$. At a probe radius of 1 Å, these four small cavities converge into one large cavity with a volume of 142 $Å^3$ (Fig. 3D). No such network was identified in protein S-824.

Of the 15 residues that border this large cavity, five are nonidentical between S-836 and S-824. That is, the network of cavities on S-836 includes 10 spatially adjacent amino acids that also occur in the S-824 sequence and five spatially adjacent ones that are unique to S-836. This spans the region in S-836 around the boundary between amino acid identity and nonidentity with S-824. The size and location of this network of cavities demonstrate an important sequence-related perturbation in hydrophobic packing. Hydrophobic packing differs between S-836 and S-824 at this interface, and this difference propagates itself into shifted packing layers throughout the structures of the proteins

Hydrophobic core packing in natural proteins is often imperfect, and small cavities of the size seen in S-836 do not preclude folding into well-ordered structures. Nonetheless, such an extensive network of cavities decreases the local packing density in the hydrophobic core. This provides room for the movement of nearby side chains and presumably accounts for the reduced unfolding cooperativity of protein S-836 relative to S-824. Moreover, the

correlation between the reduced stability of S-836 and the presence of this cavity suggests that the stability of this protein could be enhanced by the judicious substitution of one or two residues bordering this cavity.

Packing interactions

The hydrophobic core of S-836 consists of five stacked packing layers, with each layer perpendicular to the long axis of the protein (Fig. 4A). Protein S-824 also displays five stacked packing layers (Wei 2003). In S-824, side chains from all four helices contribute to every layer; however, in protein S-836, only four of the five layers contain side chains from all four helices. The first packing layer in S-836 has no side chain from helix 2 (see below).

In protein S-836, helices 2 and 4 are vertically translated relative to their positions in the structure of S-824. As shown in Figure 4B, this translation shifts the components of all the packing layers. It is noteworthy that the differences in packing occur not only in regions where the sequences differ: The translation of helices 2 and 4 causes packing differences to propagate throughout the 4-helix bundle—even into regions where the sequences are identical. This finding demonstrates that a novel superfamily of proteins can encompass a range of structural details—even in cases where the amino acid sequences are very similar.

The first packing layer in S-836 (Fig. 4A) is relatively sparse and is made up of only three nonpolar side chains—Trp23, Leu75, and Leu82. The intended fourth residue, Met30, is exposed to solvent rather than involved in hydrophobic packing. The solvent exposure of this methionine and the existence of a relatively less compact core are presumably driven by a highly stabilizing structural feature, most likely the partial burial of Trp23.

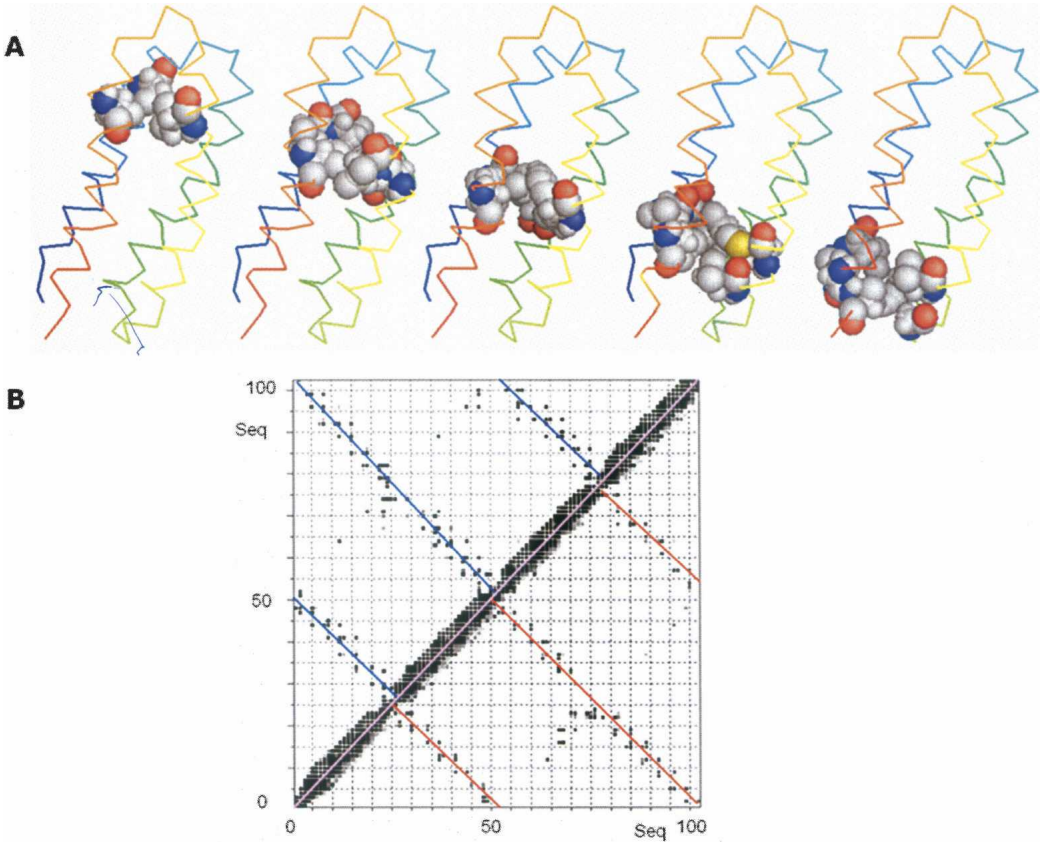

Figure 4. (*A*) Hydrophobic packing layers. S-836 contains five stacked packing layers in its hydrophobic core. Each layer typically contains four to six contributing amino acids, with the exception of the *topmost* layer (far *left*) which consists of only three amino acids due to an exposed methionine. Corresponding layers in S-824 consist of different hydrophobic amino acids (data not shown). (*B*) Contact maps. Map of intraresidue contacts for protein S-836 (upper *left*) and protein S-824 (lower *right*). All black squares represent residues that are 3 Å or less in distance. Shades of gray represent contacts greater than three angstroms but less than five angstroms apart. Contact networks perpendicular to the diagonal show interhelical contacts. Note that the networks are slightly shifted *between* the two proteins, demonstrating altered packing between the two structures.

Tryptophans are known to stabilize local hydrophobic packing (MacDonald et al. 1994; Black et al. 2001; Klein-Seetharaman et al. 2002), and the location of Trp23 suggests that it influences nearby packing interactions. The Trp23 side chain is mostly buried; and its indole secondary amine is located within 5 Å of several nearby non-hydrogen bonded carbonyls. However, the carbonyls are not directly aligned with the secondary amine, hinting at the possibility of dipole–dipole interactions rather than outright hydrogen bonding. The Trp23 indole ring also forms part of the border of the network of cavities described above. Restrictions from packing of the relatively rigid Trp23 side chain probably prevent other nearby nonpolar side chains from packing in a more efficient manner, thereby producing the observed cavities.

We had previously hypothesized that Trp23, which is also present in protein S-824, might function as a nonpolar cap, sealing the hydrophobic core of the bundle from solvent (Wei et al. 2003a). The four well-folded proteins (S-213, S-285, S-824, and S-836) all have a tryptophan at position 23, whereas the molten globule, S-23, has a leucine at this position (Wei et al. 2003b). The existence of an abridged first packing layer on S-836 and the burial of the equivalent tryptophan on S-824 reinforce this hypothesis.

Protein dynamics

To further probe the range of behaviors in our library of de novo proteins, we performed NMR dynamics studies to measure the mobility and rigidity of the backbones of proteins S-836 and S-824. We measured ^{15}N longitudinal (R1) and transverse (R2) relaxation rates and steady-state ^{1}H-^{15}N NOEs for both proteins. These data were analyzed using the Modelfree program (Palmer III et al. 1991; Mandel et al. 1995) to identify mobility at various timescales, and to obtain values for the residue-by-residue generalized order parameters (s^2).

Our Modelfree analysis of proteins S-836 and S-824 used an axial symmetric model, which corresponds with the elongated structure of 4-helix bundles. This analysis yielded values for overall correlation times (τ_m) and ratios of longitudinal versus latitudinal diffusion ($D_{\|}/D_{\perp}$). Protein S-836 yielded a correlation time of 6.728 ± 0.112 ns and a $D_{\|}/D_{\perp}$ ratio of 1.437 ± 0.072, while protein S-824 yielded a correlation time of 6.523 ± 0.056 ns and a $D_{\|}/D_{\perp}$ ratio of 1.585 ± 0.037. These data indicate that proteins S-836 and S-824 tumble at about the same rates as other proteins of similar size and shape (Gibney et al. 1997). Moreover, consistent with the results from previous work (Wei et al. 2003b), the dynamics results indicate that both proteins are compact and monomeric even at the high concentrations (1.3 mM) at which the spectra were collected. S-836 exhibits a slightly higher τ_m and slightly lower $D_{\|}/D_{\perp}$ than S-824. These differences are minor, but greater than the error, and most likely stem from differences in the dimensions of the two proteins: The structure of protein S-836 is slightly longer and fatter than that of S-824 and should have a slightly longer tumbling time and a more isotropic diffusion. Hydrodynamics calculations using the NMR structures (Garcıa de la Torre et al. 2000) show similar differences in anisotropy to those determined using the Modelfree analysis (data not shown).

Generalized order parameter

The residue-by-residue generalized order parameters (s^2) obtained from Modelfree analysis of NMR relaxation data describe the mobilities about specific bonds, in this case the H-N bonds on the protein backbone, on the picosecond timescale. The order parameter obtained from ^{1}H,^{15}N-HSQC relaxation experiments is essentially a measure of the flexibility of a protein at a specific point on its backbone. Order parameter values range from 0 to 1, with freedom of motion at 0, and restricted motion at 1. Proteins S-836 and S-824 both exhibit order parameters greater than 0.9 throughout most of their α-helices, and low order parameters, spanning 0.5–0.9, in their turn regions (Fig. 5). These values correspond to the behavior of secondary structures and their connecting turns, respectively (Redfield et al. 1992; Barchi Jr. et al. 1994; Buck et al. 1995). Low order parameters are also typical of highly flexible glycines, contributing to relatively lower order parameter values in the turn regions where they have been placed (Figs. 1C, 5). The average order parameters of both proteins are identical, within error, at around 0.92. Of particular interest is the behavior of turn 3 in both proteins. The order parameter indicates that, at the picosecond timescale, turn 3 is more rigid than the other two turns. The solved structures do not hint at such variability in turn mobility. In both structures, turn 3 is sequentially and structurally similar to turn 1; the two turns have two glycines each and are in van der Waals contact. The relative rigidity of turn 3, as determined by order parameter, may be a consequence of proximity to Trp23 on both S-836 and S-824. Van der Waals contacts between the large and rigid Trp side chain and residues in turns 3 on both proteins may reduce the mobility of nearby segments of structure.

Conformational exchange (R_{ex})

Modelfree analysis shows that protein S-836 displays higher frequencies of millisecond timescale motion than protein S-824. This is reflected in the rates of conformational exchange (R_{ex}) that were generated for S-836 and S-824 (Fig. 6; Kim and Baum 2004). In particular, S-836 has 11 amino acids with R_{ex} greater than 2.0 s^{-1} while S-824 has only eight. Sites of such exchange sample

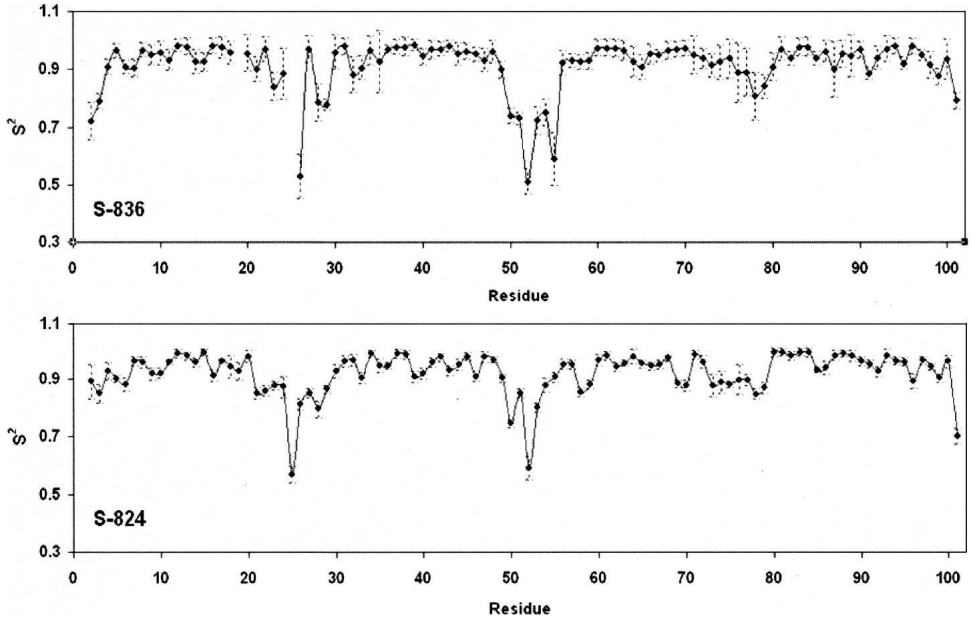

Figure 5. Residue-by-residue generalized order parameters (s^2) for protein S-836 and protein S-824. Both proteins show dips corresponding with the turn regions. The order parameters for the turn regions in S-836 are lower than for the corresponding regions in S-824. In both proteins, turn three exhibits significantly higher order parameters than the other two turns.

multiple conformational states. With both proteins, some of the millisecond timescale motion occurs at turn 3 and, to a lesser degree, at turn 1. There are also the occasional single occurrences of significant exchange at the N or C termini and along the helices. Overall, the difference between R_{ex} on S-836 and S-824 is one of degree. S-836 and S-824 display a number of common, albeit shifted, sites of conformational exchange; S-836 just has larger and more frequent R_{ex}. The shift is probably a consequence of the differences in their hydrophobic core packing.

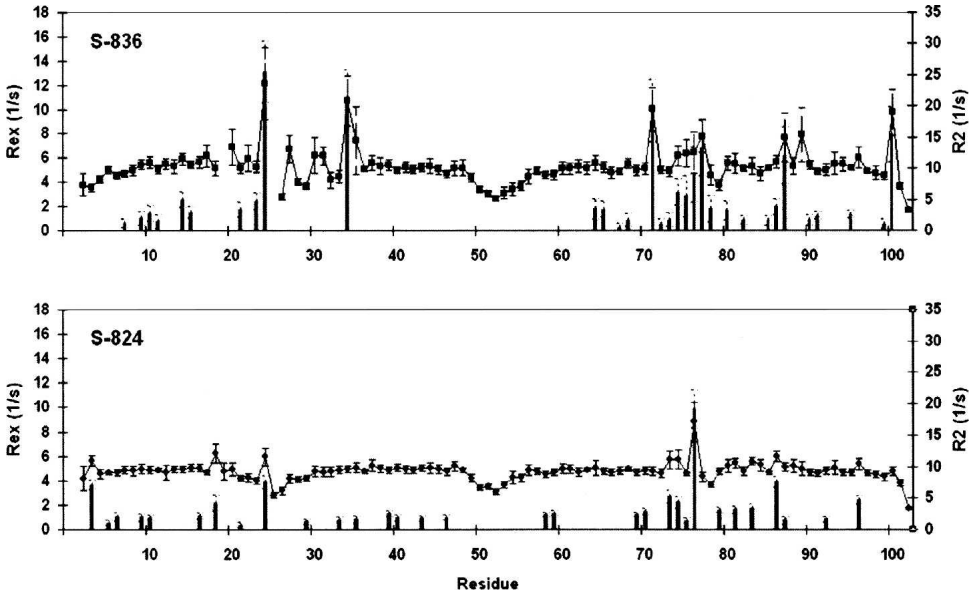

Figure 6. R2 relaxation constants (squares) and conformational exchange rate constants (R_{ex}, bars), both at 500 MHz, for proteins S-836 and S-824. Conformational exchange is more pronounced and occurs for more residues in S-836 than S-824. For both proteins, higher R_{ex} values occur at and near turns one and three, indicating motion at the millisecond timescale at these turns, rather than the fast picosecond timescale local motion typical of turns, and exhibited by turn two.

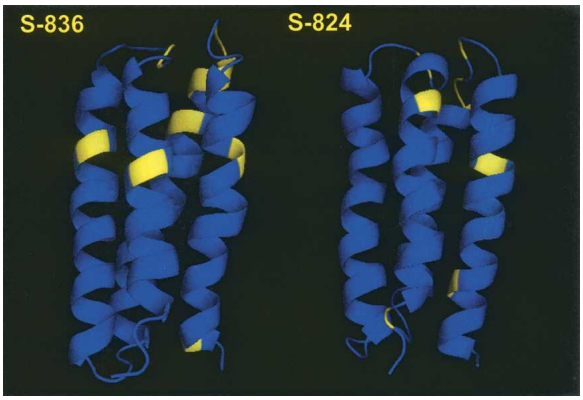

Figure 7. Backbone conformational exchange in S-836 and S-824. Residues with R_{ex} values greater than 2.0 s^{-1}, are shown in yellow. Highlighted are residues 14, 23, 24, 34, 71, 74, 75, 76, 77, 87, and 100 on S-836 and 3, 18, 24, 73, 74, 76, 86, and 96 on S-824.

Significant exchange that is unique to S-836 occurs at residues E14 and D34 which occur two-thirds of the way up on the first and second helices, respectively. Together with E87, these helical, conformationally exchanging amino acids border the network of packing cavities described above (Figs. 3D, 7). These findings provide strong evidence that the network of small cavities influences the backbone dynamics of protein S-836. Although the differences in sequence between S-836 and S-824 are localized to residues 18–35 and 71–87, the impact of the sequence differences propagates throughout the structure. Thus, it becomes more probable that increased dynamics of S-836 relative to S-824 indeed stems from inefficient hydrophobic core packing, which leads to a local network of cavities bordered by a wobbly polypeptide backbone.

Conclusions

The finding that protein S-836, like protein S-824, folds into the expected 4-helix bundle demonstrates that the desired topology can be encoded by binary patterning without the need for explicit design of residue-by-residue packing. Moreover, these results suggest that our second-generation library of 102-residue sequences contains a vast number of stable 4-helix bundles.

The similar peak dispersion in the $^{1}H,^{15}N$-HSQC spectra of proteins S-285, S-213, and S-836 (Wei et al. 2003b) suggests that S-285 and S-213 also form stable well-ordered tertiary structures with well-defined hydrophobic cores. We believe that S-836 represents a ''typical'' protein from this library. With the determination of the structure and dynamics of S-836, we have now sampled a range of behaviors from this library: S-836 represents an average protein; S-824 represents proteins that are slightly ''better'' than average (more ordered, less dynamic); and S-23 represents proteins that are slightly ''worse'' than average (more dynamic to the point of resembling a molten globule).

This second-generation binary patterned library can be seen as a de novo protein superfamily of 4-helix bundles. Because this collection was neither designed a priori by computational methods, nor subjected to evolutionary selection for biological function, this superfamily provides an unbiased estimate of the range of structures and dynamics that are possible for 4-helix bundles. Thus, this collection provides a model for estimating the properties of ancestral proteins.

The structural and dynamical studies of proteins S-836 and S-824 indicate that although perfectly packed tertiary structures are not easily obtained, highly stable and reasonably well-ordered structures do occur frequently in an unselected library. In the case of protein S-836—and presumably in many other proteins in this superfamily—imperfect packing produces cavities and favors structures that are somewhat more dynamic. A correlation between packing and protein stability and/or dynamics has also been observed for several other proteins—both natural and designed (Lim and Sauer 1989; Eriksson et al. 1992; Gassner et al. 1996; Munson et al. 1996; Dahiyat and Mayo 1997; Walsh et al. 1999, 2001; Willis et al. 2000). Wobbly cavities and dynamic structures may be disadvantageous for highly evolved protein activities requiring ''lock and key'' binding. However, these properties would be advantageous for ancestral proteins requiring malleable structures capable of broad specificity toward a range of substrates. Such pluripotency would provide catalytic versatility in an ancestral cell functioning with limited repertoire of enzymes (Jensen 1976; Nagano et al. 2002; Orengo and Thornton 2005; Glasner et al. 2006). Similarly, initial studies on our binary patterned libraries of 4-helix bundles demonstrate that many of these de novo proteins also possess low levels of enzymatic activity with a broad range of substrate specificity (Rojas et al. 1997; Moffet et al. 2000, 2001, 2003; Wei and Hecht 2004; Das et al. 2006; Das and Hecht 2007).

Materials and Methods

Protein expression and purification

Protein S-836 was expressed in *Escherichia coli* strain BL21 (DE3). ^{15}N-labeled proteins were expressed by growing for 14–16 h in autoinducing minimal media (Studier 2005) with 100 mg/mL ampicillin, and $^{15}NH_4Cl$ was used as nitrogen source. $^{15}N,^{13}C$-labeled S-836 was prepared using a procedure designed to increase yield during isotopic labeling (Marley et al. 2001) with minor modifications. To obtain pure protein, *E. coli* cells were lysed using the freeze-thaw method (Johnson and Hecht 1994), and impurities were removed by acid precipitation at pH 4 with sodium acetate buffer. Protein was purified using cation-exchange HPLC (HS, Millipore POROS). Details of the

expression and purification are described elsewhere (Go et al. 2007).

NMR structure determination

NMR spectra for structure determination were collected with a Varian Inova 600-MHz spectrometer. All spectra were collected at 298°K. The typical protein sample consisted of 2 mM ^{15}N- or ^{13}C,^{15}N- S-836 in 10% D_2O (v/v) at a volume of 250 μL in a thin-walled Shigemi tube. 2,2-dimethyl-2-silapentane-5-sulfonate (DSS) was used as a chemical shift reference (Wishart et al. 1995).

The structure of S-836 was calculated using distance restraints obtained from ^{15}N-edited NOESY (with ^{15}N-HSQC-TOCSY) and ^{13}C-edited NOESY (with HCCH-TOCSY), and *phi* dihedral angles from a HNHA J-coupling experiment (Vuister and Bax 1993). All spectra were processed using NMRPipe (Delaglio et al. 1995) and analyzed using both NMRPipe (with NMRDraw), and Sparky (T.D. Goddard and D.G. Kneller, University of California, San Francisco).

Distance restraints were obtained by classifying peak intensities from the two NOESY experiments into strong (1.8–2.9 Å), medium (1.8–3.3 Å), and weak intensity peaks (1.8–5.0 Å). The contents of these groups were calibrated using known proton–proton distances in helical secondary structures. The upper limits of these restraints were then corrected to adjust for lack of stereo-specific assignments (Wüthrich 1986). Phi-dihedral angles were obtained from the HNHA experiment. The ratios of amide and alpha hydrogen intensities were converted to dihedral angles using the Karplus relationship between coupling constants and dihedral angles (Karplus 1963).

Three-dimensional structures were calculated using XPLOR-NIH (Schwieters et al. 2003). The initial structures were calculated by simulated annealing starting from a temperature of 1000°K, cooling in 6000 steps, to a final temperature of 100°K. Refinement was performed iteratively, starting at a temperature of 1500°K, cooling down in 12,000 steps to 100°K. The 20 lowest energy structures were aligned using MOLMOL (Koradi et al. 1996) and validated by comparing structures against experiments restraints. Additional validation was performed using AQUA/PROCHECK (Laskowski et al. 1996) and MolPRrobity (Davis et al. 2007). RMSD values were calculated using MOLMOL and cavity surface areas, using CASTp (Dundas et al. 2006).

Relaxation and dynamics

Experiments to measure R1 and R2 relaxation rates and NOEs were performed at field strengths of 500 MHz and 600 MHz. The 600-MHz data were obtained with a cryoprobe. Samples of ^{15}N-labeled S-836 and S-824 were prepared at pH 4.0 at a concentration of 1.3 mM in a volume of 0.6 mL with 10% D_2O.

R2 relaxation experiments for both S-836 and S-824 were collected at eight different relaxation delays (10, 30, 50, 70, 90, 130, 170, and 190 ms) at 500 MHz, and at nine different delays (10, 30, 50, 70, 90, 110, 130, 170, and 190 ms) at 600 MHz. We also used nine R1 relaxation delays at 500 MHz (30, 100, 150, 250, 400, 600, 750, 900, and 1100 ms) and 600 MHz (30, 100, 150, 250, 400, 600, 750, 900, and 1200 ms). Measuring NOEs required two experiments with and without the ^{1}H saturation. The spectra were processed using NMRPipe and analyzed using both NMRPipe and NMRDraw.

R1 and R2 relaxation rate constants were obtained by plotting the exponential decay of peak intensities against their respective relaxation delays. To confirm that the proteins existed entirely in monomeric form, a second R2 relaxation experiment was collected at 500 MHz on a sample of S-836 at a concentration of 0.6 mM. R1 and R2 experiments for S-824 were also performed twice at 600 MHz. Relaxation rates were identical within error for both duplications.

In addition to R1 and R2 relaxation rates and the steady-state NOEs, Modelfree analysis of the two proteins required initial estimates for the overall correlation time, obtained from an estimated rotational diffusion tensor, and the ratio between longitudinal and latitudinal diffusion ($D_{\|}/D_{\perp}$). Estimates for the rotational diffusion tensor and $D_{\|}/D_{\perp}$ were obtained using two programs—*pdbinertia* to rotate and align the solution structures to the coordinate axes and *r2r1diffusion* (Tjandra et al. 1995) to calculate diffusion tensor and $D_{\|}/D_{\perp}$ and realign the translated structure. Both programs are provided on A.G. Palmer's Web site (http://cpmcnet.columbia.edu/dept/gsas/biochem/labs/palmer/software.html). *R2r1diffusion* required modifications in the recommended input. Fluctuations in R2 relaxation rates within the S-836 sequence required very drastic reduction in the number of residues used in calculating the diffusion tensor and $D_{\|}/D_{\perp}$. Whereas data for at least 45 amino acids were used to calculate diffusion tensor and $D_{\|}/D_{\perp}$ for S-824, a combination of less than 20 residues produced initial estimates that yielded reasonable results after cycles of Modelfree analysis. Analysis of S-836 and S-824 dynamics utilized the software Modelfree version 4 and the iterative cycle of residue-by-residue model selection and parameter optimization described by Mandel et al. (1995). Consistent with their protein topology, Modelfree analysis of S-836 and S-824 utilized the axial symmetric model.

References

Anfinsen, C.B. 1972. The formation and stabilization of protein structure. *Biochem. J.* **128:** 737–749.

Barchi Jr., J.J., Grasberger, B., Gronenborn, A.M., and Clore, G.M. 1994. Investigation of the backbone dynamics of the IgG-binding domain of streptococcal protein G by heteronuclear two-dimensional ^{1}H-^{15}N nuclear magnetic resonance spectroscopy. *Protein Sci.* **3:** 15–21.

Black, K.M., Clark-Lewis, I., and Wallace, C.J. 2001. Conserved tryptophan in cytochrome *c*: Importance of the unique side-chain features of the indole moiety. *Biochem. J.* **359:** 715–720.

Bradley, L.H., Thumfort, P.P., and Hecht, M.H. 2006. De novo proteins from binary-patterned combinatorial libraries. *Methods Mol. Biol.* **340:** 53–69.

Bradley, L.H., Wei, Y., Thumfort, P., Wurth, C., and Hecht, M.H. 2007. Protein design by binary patterning of polar and nonpolar amino acids. *Methods Mol. Biol.* **352:** 155–166.

Buck, M., Boyd, J., Redfield, C., MacKenzie, D.A., Jeenes, D.J., Archer, D.B., and Dobson, C.M. 1995. Structural determinants of protein dynamics: Analysis of ^{15}N NMR relaxation measurements for main-chain and side-chain nuclei of hen egg white lysozyme. *Biochemistry* **34:** 4041–4055.

Chiarabelli, C., Vrijbloed, J.W., De Lucrezia, D., Thomas, R.M., Stano, P., Polticelli, F., Ottone, T., Papa, E., and Luisi, P.L. 2006. Investigation of de novo totally random biosequences. Part II: On the folding frequency in a totally random library of de novo proteins obtained by phage display. *Chem. Biodivers.* **3:** 840–859.

Chothia, C., Levitt, M., and Richardson, D. 1977. Structure of proteins: Packing of α-helices and pleated sheets. *Proc. Natl. Acad. Sci.* **74:** 4130–4134.

Crick, F.H. 1953. The packing of α-helices: Simple coiled coils. *Acta Crystallogr.* **6:** 689–697.

Dahiyat, B.I. and Mayo, S.L. 1997. Probing the role of packing specificity in protein design. *Proc. Natl. Acad. Sci.* **94:** 10172–10177.

Das, A. and Hecht, M.H. 2007. Peroxidase activity of de novo heme proteins immobilized on electrodes. *J. Inorg. Biochem.* **101:** 1820–1826.

Das, A., Trammell, S.A., and Hecht, M.H. 2006. Electrochemical and ligand binding studies of a de novo heme protein. *Biophys. Chem.* **123:** 102–112.

Davis, I.W., Leaver-Fay, A., Chen, V.B., Block, J.N., Kapral, G.J., Wang, X., Murray, L.W., Arendall III., W.B., Snoeyink, J., Richardson, J.S., et al. 2007. MolProbity: All-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Res.* **35:** W375–W383. doi: 10.1093/nar/gkm216.

Delaglio, F., Grzesiek, S., Vuister, G.W., Zhu, G., Pfeifer, J., and Bax, A. 1995. NMRPipe: A multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR* **6:** 277–293.

Dundas, J., Ouyang, Z., Tseng, J., Binkowski, A., Turpaz, Y., and Liang, J. 2006. CASTp: Computed atlas of surface topography of proteins with structural and topographical mapping of functionally annotated residues. *Nucleic Acids Res.* **34:** W116–W118. doi: 10.1093/nar/gkl282.

Eriksson, A.E., Baase, W.A., Zhang, X.J., Heinz, D.W., Blaber, M., Baldwin, E.P., and Matthews, B.W. 1992. Response of a protein structure to cavity-creating mutations and its relation to the hydrophobic effect. *Science* **255:** 178–183.

Fauchère, J.L. and Pliska, V. 1983. Hydrophobic parameters pi of amino-acid side chains from the partitioning of N-acetyl-amino-acid amides. *Eur. J. Med. Chem.* **18:** 369–375.

Garcıa de la Torre, J., Huertas, M.L., and Carrasco, B. 2000. HYDRONMR: Prediction of NMR relaxation of globular proteins from atomic-level structures and hydrodynamic calculations. *J. Magn. Reson.* **147:** 138–146.

Gassner, N.C., Baase, W.A., and Matthews, B.W. 1996. A test of the ''jigsaw puzzle'' model for protein folding by multiple methionine substitutions within the core of T4 lysozyme. *Proc. Natl. Acad. Sci.* **93:** 12155–12158.

Gibney, B.R., Johansson, J.S., Rabanal, F., Skalicky, J.J., Wand, A.J., and Dutton, P.L. 1997. Global topology and stability and local structure and dynamics in a synthetic spin-labeled four-helix bundle protein. *Biochemistry* **36:** 2798–2806.

Glasner, M.E., Gerlt, J.A., and Babbitt, P.C. 2006. Evolution of enzyme superfamilies. *Curr. Opin. Chem. Biol.* **10:** 492–497.

Go, A., Kim, S., Hecht, M.H., and Baum, J. 2007. NMR assignment of S836: A de novo protein from a designed superfamily. *J. Biomol. NMR* **1:** 213–215.

Harris, N.L., Presnell, S.R., and Cohen, F.E. 1994. Four-helix bundle diversity in globular proteins. *J. Mol. Biol.* **236:** 1356–1368.

Hecht, M.H., Das, A., Go, A., Bradley, L.H., and Wei, Y. 2004. De novo proteins from designed combinatorial libraries. *Protein Sci.* **13:** 1711–1723.

Jensen, R.A. 1976. Enzyme recruitment in evolution of new function. *Annu. Rev. Microbiol.* **30:** 409–425.

Johnson, B.H. and Hecht, M.H. 1994. Recombinant proteins can be isolated from *E. coli* cells by repeated cycles of freezing and thawing. *Biotechnology (N.Y.)* **12:** 1357–1360.

Kamtekar, S., Schiffer, J.M., Xiong, H., Babik, J.M., and Hecht, M.H. 1993. Protein design by binary patterning of polar and nonpolar amino acids. *Science* **262:** 1680–1685.

Karplus, M. 1963. Vicinal proton coupling in nuclear magnetic resonance. *J. Am. Chem. Soc.* **85:** 2870–2871.

Keefe, A.D. and Szostak, J.W. 2001. Functional proteins from a random-sequence library. *Nature* **410:** 715–718.

Kim, S. and Baum, J. 2004. An on/off resonance rotating frame relaxation experiment to monitor millisecond to microsecond timescale dynamics. *J. Biomol. NMR* **30:** 195–204.

Klein-Seetharaman, J., Oikawa, M., Grimshaw, S.B., Wirmer, J., Duchardt, E., Ueda, T., Imoto, T., Smith, L.J., Dobson, C.M., and Schwalbe, H. 2002. Long-range interactions within a nonnative protein. *Science* **295:** 1719–1722.

Koradi, R., Billeter, M., and Wüthrich, K. 1996. MOLMOL: A program for display and analysis of macromolecular structures. *J. Mol. Graph.* **14:** 29–32, 51–55.

Kyte, J. and Doolittle, R.F. 1982. A simple method for displaying the hydropathic character of a protein. *J. Mol. Biol.* **157:** 105–132.

Laskowski, R.A., Rullmannn, J.A., MacArthur, M.W., Kaptein, R., and Thornton, J.M. 1996. AQUA and PROCHECK-NMR: Programs for checking the quality of protein structures solved by NMR. *J. Biomol. NMR* **8:** 477–486.

Lim, W.A. and Sauer, R.T. 1989. Alternative packing arrangements in the hydrophobic core of λ represser. *Nature* **339:** 31–36.

Lopez de la Osa, J., Bateman, D.A., Ho, S., Gonzalez, C., Chakrabartty, A., and Laurents, D.V. 2007. Getting specificity from simplicity in putative proteins from the prebiotic earth. *Proc. Natl. Acad. Sci.* **104:** 14941–14946.

MacDonald, R.I., Musacchio, A., Holmgren, R.A., and Saraste, M. 1994. Invariant tryptophan at a shielded site promotes folding of the conformational unit of spectrin. *Proc. Natl. Acad. Sci.* **91:** 1299–1303.

Mandecki, W. 1990. A method for construction of long randomized open reading frames and polypeptides. *Protein Eng.* **3:** 221–226.

Mandel, A.M., Akke, M., and Palmer III, A.G. 1995. Backbone dynamics of *Escherichia coli* ribonuclease HI: Correlations with structure and function in an active enzyme. *J. Mol. Biol.* **246:** 144–163.

Marley, J., Lu, M., and Bracken, C. 2001. A method for efficient isotopic labeling of recombinant proteins. *J. Biomol. NMR* **20:** 71–75.

Moffet, D.A., Certain, L.K., Smith, A.J., Kessel, A.J., Beckwith, K.A., and Hecht, M.H. 2000. Peroxidase activity in heme proteins derived from a designed combinatorial library. *J. Am. Chem. Soc.* **122:** 7612–7613.

Moffet, D.A., Case, M.A., House, J.C., Vogel, K., Williams, R.D., Spiro, T.G., McLendon, G.L., and Hecht, M.H. 2001. Carbon monoxide binding by de novo heme proteins derived from designed combinatorial libraries. *J. Am. Chem. Soc.* **123:** 2109–2115.

Moffet, D.A., Foley, J., and Hecht, M.H. 2003. Midpoint reduction potentials and heme binding stoichiometries of de novo proteins from designed combinatorial libraries. *Biophys. Chem.* **105:** 231–239.

Munson, M., Balasubramanian, S., Fleming, K.G., Nagi, A.D., O'Brien, R., Sturtevant, J.M., and Regan, L. 1996. What makes a protein a protein? Hydrophobic core designs that specify stability and structural properties. *Protein Sci.* **5:** 1584–1593.

Nagano, N., Orengo, C.A., and Thornton, J.M. 2002. One fold with many functions: The evolutionary relationships between TIM barrel families based on their sequences, structures and functions. *J. Mol. Biol.* **321:** 741–765.

Orengo, C.A. and Thornton, J.M. 2005. Protein families and their evolution—a structural perspective. *Annu. Rev. Biochem.* **74:** 867–900.

Palmer III, A.G., Ranee, M., and Wright, P.E. 1991. Intramolecular motions of a zinc finger DNA-binding domain from Xfin characterized by proton-detected natural abundance ^{13}C heteronuclear NMR spectroscopy. *J. Am. Chem. Soc.* **113:** 4371–4380.

Presnell, S.R. and Cohen, F.E. 1989. Topological distribution of four-a-helix bundles. *Proc. Natl. Acad. Sci.* **86:** 6592–6596.

Redfield, C., Boyd, J., Smith, L.J., Smith, R.A., and Dobson, C.M. 1992. Loop mobility in a four-helix-bundle protein: ^{15}N NMR relaxation measurements on human interleukin-4. *Biochemistry* **31:** 10431–10437.

Rojas, N.R., Kamtekar, S., Simons, C.T., McLean, J.E., Vogel, K.M., Spiro, T.G., Farid, R.S., and Hecht, M.H. 1997. De novo heme proteins from designed combinatorial libraries. *Protein Sci.* **6:** 2512–2524.

Roy, S. and Hecht, M.H. 2000. Cooperative thermal denaturation of proteins designed by binary patterning of polar and nonpolar amino acids. *Biochemistry* **39:** 4603–4607.

Roy, S., Helmer, K.J., and Hecht, M.H. 1997. Detecting native-like properties in combinatorial libraries of de novo proteins. *Fold. Des.* **2:** 89–92.

Scheraga, H.A. 1970. Theoretical and experimental studies of conformations of polypeptides. *Chem. Rev.* **71:** 195–217.

Schwieters, C.D., Kuszewski, J.J., Tjandra, N., and Clore, G.M. 2003. The Xplor-NIH NMR molecular structure determination package. *J. Magn. Reson.* **160:** 65–73.

Studier, F.W. 2005. Protein production by auto-induction in high density shaking cultures. *Protein Expr. Purif.* **41:** 207–234.

Tjandra, N., Feller, S.E., Pastor, R.W., and Bax, A. 1995. Rotational diffusion anisotropy of human ubiquitin from ^{15}N NMR relaxation. *J. Am. Chem. Soc.* **117:** 12562–12566.

Vuister, G.W. and Bax, A. 1993. Quantitative *J* correlation: A new approach for measuring homonuclear three-bond $J(H^NH^\alpha)$ coupling constants in ^{15}N-enriched proteins. *J. Am. Chem. Soc.* **115:** 7772–7777.

Walsh, S.T., Cheng, H., Bryson, J.W., Roder, H., and DeGrado, W.F. 1999. Solution structure and dynamics of a *de novo* designed three-helix bundle protein. *Proc. Natl. Acad. Sci.* **96:** 5486–5491.

Walsh, S.T., Sukharev, V.I., Betz, S.F., Vekshin, N.L., and DeGrado, W.F. 2001. Hydrophobic core malleability of a *de novo* designed three-helix bundle protein. *J. Mol. Biol.* **305:** 361–373.

Wang, W. and Hecht, M.H. 2002. Rationally designed mutations convert de novo amyloid-like fibrils into monomeric β-sheet proteins. *Proc. Natl. Acad. Sci.* **99:** 2760–2765.

Watters, A.L. and Baker, D. 2004. Searching for folded proteins in vitro and in silico. *Eur. J. Biochem.* **271:** 1615–1622.

Wei, Y. 2003. ''Structural and functional studies of well-folded alpha-helical proteins from a designed combinatorial library.'' Ph.D thesis, Princeton University, Princeton, NJ.

Wei, Y. and Hecht, M.H. 2004. Enzyme-like proteins from an unselected library of designed amino acid sequences. *Protein Eng. Des. Sel.* **17:** 67–75.

Wei, Y., Kim, S., Fela, D., Baum, J., and Hecht, M.H. 2003a. Solution structure of a de novo protein from a designed combinatorial library. *Proc. Natl. Acad. Sci.* **100:** 13270–13273.

Wei, Y., Liu, T., Sazinsky, S.L., Moffet, D.A., Pelczer, I., and Hecht, M.H. 2003b. Stably folded de novo proteins from a designed combinatorial library. *Protein Sci.* **12:** 92–102.

West, M.W., Wang, W., Patterson, J., Mancias, J.D., Beasley, J.R., and Hecht, M.H. 1999. De novo amyloid proteins from designed combinatorial libraries. *Proc. Natl. Acad. Sci.* **96:** 11211–11216.

Willis, M.A., Bishop, B., Regan, L., and Brunger, A.T. 2000. Dramatic structural and thermodynamic consequences of repacking a protein's hydrophobic core. *Structure* **8:** 1319–1328.

Wishart, D.S., Bigam, C.G., Yao, J., Abildgaard, F., Dyson, H.J., Oldfield, E., Markley, J.L., and Sykes, B.D. 1995. ^{1}H, ^{13}C and ^{15}N chemical shift referencing in biomolecular NMR. *J. Biomol. NMR* **6:** 135–140.

Wolfenden, R., Andersson, L., Cullis, P.M., and Southgate, C.C. 1981. Affinities of amino acid side chains for solvent water. *Biochemistry* **20:** 849–855.

Wüthrich, K. 1986. *NMR of proteins and nucleic acids*. Wiley, New York.

Xu, G., Wang, W., Groves, J.T., and Hecht, M.H. 2001. Self-assembled monolayers from a designed combinatorial library of de novo β-sheet proteins. *Proc. Natl. Acad. Sci.* **98:** 3652–3657.