De Novo Design of Helical Bundles as Models for Understanding Protein Folding and Function

R. BLAKE HILL,[†] DANIEL P. RALEIGH,[‡] ANGELA LOMBARDI,[§] AND WILLIAM F. DEGRADO*,[†]

Johnson Research Foundation, Department of Biochemistry and Biophysics, University of Pennsylvania, Philadelphia, Pennsylvania 19104-6059, Department of Chemistry, State University of New York, Stony Brook, New York 11794, and Department of Chemistry, University of Naples, Naples 80134, Italy

Received July 7, 2000

ABSTRACT

De novo protein design has proven to be a powerful tool for understanding protein folding, structure, and function. In this Account, we highlight aspects of our research on the design of dimeric, four-helix bundles. Dimeric, four-helix bundles are found throughout nature, and the history of their design in our laboratory illustrates our hierarchic approach to protein design. This approach has been successfully applied to create a completely native-like protein. Structural and mutational analysis allowed us to explore the determinants of native protein structure. These determinants were then applied to the design of a dinuclear metal-binding protein that can now serve as a model for this important class of proteins.

Protein folding is an important and multifaceted scientific problem.[1−8] One facet of this problem involves the prediction of three-dimensional structure from amino acid sequence, an endeavor the importance of which has been highlighted by the recent release of gene sequences of many whole organisms. As the size of the protein structural database expands, it will be increasingly possible to assign amino acid sequences to a given structural family on the basis of the homology of their sequences with proteins of known structure. Another facet of the folding problem is the delineation of the kinetic mechanism by which an unfolded protein chain undergoes the transition from a random coil with an astronomical number of configurations to a uniquely folded structure. A final facet of the protein folding problem is to understand, at the atomic level, the detailed physicochemical features that kinetically and thermodynamically direct the formation of an unique, cooperatively folded, native structure. This latter endeavor is particularly important as it lays the groundwork for the design of novel inhibitors of natural proteins as well as the construction of novel proteins and related biopolymers not found in nature.

De novo protein design is an approach to the protein folding problem that critically tests our understanding of all these facets of this problem.[9,10] This method involves the design of a sequence that is intended to fold into a predetermined three-dimensional structure without the sequence being patterned after any natural protein. Through an iterative process of design and rigorous characterization, the principles governing protein folding and function can be evaluated. Thus, "failures" highlight gaps in our understanding, whereas "successes" confirm the principles used in the design and often provide simple model systems to further refine the relevant parameters. *De novo* protein design has been the subject of several recent reviews, and spatial limitations prevent an exhaustive discussion of the topic here.[11] Instead, we will endeavor to illustrate this approach through a brief discussion of our work on the design of dimeric four-helix bundles.

Conformational Specificity

Conformational specificity refers to the ability of a protein to fold into a unique three-dimensional structure. There are four levels at which conformational specificity is apparent, beginning with the correct oligomeric state. Second, the overall fold and topology of a protein must be correctly specified. For example, the sequence of a three-helix bundle can specify a clockwise-turning topology as in the Z-domain of protein A[12] or a counterclockwise-turning topology, as in a recent *de novo*-designed protein, the success of which was experimentally demonstrated.[13] Third, the relative spatial orientation of secondary structural elements is a consideration for the proper conformational specificity of a native protein. For example, scores of related folds within a structural family are found that differ simply by rigid body shifts of secondary structural elements.[14] Fourth, the packing of the hydrophobic core needs to be favorable. Side chains in protein interiors are generally well packed, and the individual residues tend to populate a single rotameric state of low energy.[15−19] Thus, conformational specificity

R. Blake Hill received his Ph.D. with James H. Prestegard from Yale University (1995); he was a postdoctoral fellow with William F. DeGrado at the University of Pennsylvania. He is now an Assistant Professor of Biology at the Johns Hopkins University.

Daniel P. Raleigh received his Ph.D. with Robert G. Griffin at the Massachusetts Institute of Technology; he was a postdoctoral fellow with Christopher Dobson at the University of Oxford and with William F. DeGrado at the Du Pont-Merck Pharmaceutical Co., and is now an Associate Professor of Chemistry at the State University of New York at Stony Brook.

Angela Lombardi received her Ph.D. in chemistry from the University "Federico II" of Naples, Italy, in 1986. She was a postdoctoral fellow and an associate researcher in the group of Vincenzo Pavone at the same university. She then spent one year with William F. DeGrado at the DuPont-Merck Pharmaceutical Co. and, since 1998, has been an Associate Professor of Chemistry at the University of Naples.

William F. DeGrado received his Ph.D. in organic chemistry with E. Thomas Kaiser from the University of Chicago (1981), was a member of the scientific staff of DuPont and Du Pont-Merck Pharmaceutical Co. from 1981 to 1996, and is currently a professor in the department of Biochemistry and Biophysics, University of Pennsylvania.

* To whom correspondence should be addressed. E-mail: degrado@mail.med.upenn.edu. Telephone: (215) 898-4590. Fax: (215) 573-7229.
[†] University of Pennsylvania.
[‡] State University of New York, Stony Brook.
[§] University of Naples.

10.1021/ar970004h CCC: $19.00
Published on Web 10/24/2000

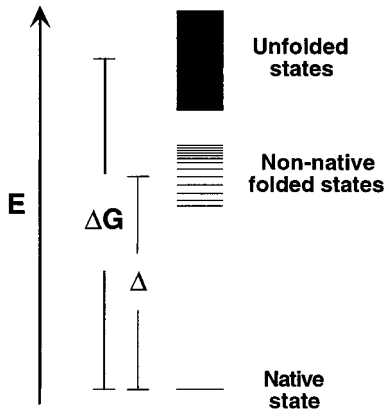

FIGURE 1. Hypothetical free energy diagram for a protein. Each line or bar represents a distinct conformational state, with the native state as the lowest energy state and the unfolded states a densely populated, nearly isoenergetic ensemble. Between these extremes are non-native folded states, oftentimes referred to as the molten globule ensemble. These non-native folded states exhibit extensive secondary structure but lack well-defined tertiary structure. The population of each state is dictated by the Boltzmann distribution. For native protein structure the free energy gap, Δ, must be large enough to significantly populate a distinct native state. For natural proteins Δ has evolved to be much larger than in *de novo*-designed proteins. Thus, a single mutation in designed proteins offers the advantage of being able to access the non-native states in a manner not often observed with natural proteins.

reflects the tendency of a protein to achieve a unique state (or set of closely related states) at each of these different levels.

The conformational specificity of a protein can also be viewed in thermodynamic terms. A unique structure requires a large free energy gap between the native state and the ensemble of non-native folded states and partially folded states,[20] often denoted molten globules.[21−23] Without a large enough free energy gap, a sequence can fail to display one or more of the four levels of conformational specificity. A consensus concerning the nomenclature for this free energy gap has not appeared in the literature;[24−26] here, we refer to it as Δ. Thus, successful protein design requires sufficient thermodynamic stability to favor the folded state over the unfolded states, as well as a large Δ (relative to thermal energy, kT) to ensure conformational specificity. A hypothetical free energy diagram for a protein is shown in Figure 1.

In natural proteins, a lack of conformational specificity can have severe pathophysiological consequences. A striking failure to attain the first level of conformational specificity occurs in sickle cell anemia, where a Glu-to-Val mutation on the surface of hemoglobin results in a loss of this protein's ability to specify the correct oligomeric state.[27,28] More recently, Booth et al. have described mutants of human lysozyme that cause amyloid formation in which the native state has been destabilized.[29] Destabilization of the native state relative to partially folded structures leads to a transient population of molten-globule-like states that lack global cooperativity. Although the proteins are enzymatically active, the increased population of non-native folded states leads to protein aggregation with concomitant formation of plaques *in vivo*. The important role of alternative or partially folded structures in amyloid formation is nicely illustrated by the demonstration that some proteins which do not form amyloid *in vivo* can be induced to form amyloid *in vitro* by incubation under mildly denaturing conditions.[30] The elegant studies of Kelly and co-workers on the formation of amyloids by transthyretin have further highlighted the important role of conformational specificity of the formation of amyloids important to many disease states. In this example, conditions or mutations that destabilize the native tetramer of transthyretin may lead to amyloid formation from a monomeric intermediate, while mutations that stabilize the native state inhibit amyloid formation, as do small molecules which bind to and stabilize the native tetramer.[31,32] Prion proteins and the recently discovered yeast prions offer yet another example of the extreme effects that can accompany a lack of conformational specificity (for recent reviews, see Mihara et al., 1998;[33] Lindquist, 1997;[34] and Prusiner, 1982,[35] 1998[36]).

De novo protein design has led to a variety of proteins showing a gradation of conformational specificity. Early attempts to design proteins led to structures that showed extremely high thermodynamic stability, but whose interior-facing side chains were dynamically disordered as in some forms of molten globules.[21,37] Several groups followed this work with the successful design of peptides that adopt partially structured states. For example, Mayo and co-workers have designed a tetrameric β-sheet protein, based on the sequence of cytokines PF4 and IL8.[38,39] The monomers adopt two distinct antiparallel three-stranded sheet conformations with different registries of the strands. The monomers assemble into a dimer of six-stranded sheets that is mediated through hydrophobic surfaces that appear to interact in a number of dynamically interconverting conformations, as assessed from the lack of intersheet NOEs. In a similar manner, Wand, Dutton, and co-workers described tetrameric four-helix bundles,[40] composed of partially structured dimers that assemble into tetramers via a structurally mobile hydrophobic interface.[41] These proteins appear to have a marginal energetic difference between parallel and antiparallel topologies, a fact that was recently exploited to create a "topological switch" in response to heme ligation.[42] Also, Baltzer and co-workers described a family of partially native-like, dimeric, four-helix bundles with 4-nitrophenylesterase activity.[11,43,44]

Natural proteins that form stable molten globule states are typically larger than *de novo*-designed proteins, and it has been suggested that proteins need to exceed a certain minimum size to be able to form a stable molten globule, or partially folded, state. The early work on *de novo*-designed proteins coupled with recent efforts to prepare "peptide models" of folding intermediates[45,46] demonstrates that moderate sized polypeptides form molten globule-like structures rather easily. These results suggest that nature has worked quite hard to avoid forming molten globules. Many research groups have also worked quite hard to avoid designing molten globules,

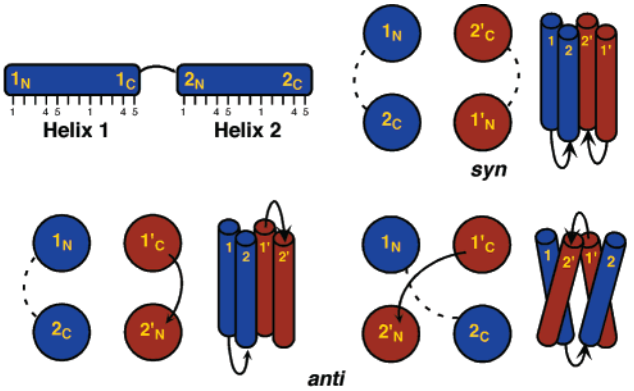

FIGURE 2. A dimeric, four-helix bundle can adopt six distinct topologies. The topology diagrams shown are for the three possible clockwise-turning helical bundles. Counterclockwise-turning helical bundles are also possible and have interfacial interactions different from their clockwise-turning correlates.

and, recently, the design of proteins that adopt a unique conformation has been accomplished. Examples include peptides of approximately 25 residues patterned after the zinc finger motif.[47−51] Further, approximately 20-residue peptides that adopt marginally stable antiparallel β-sheets have been described,[52−56] and native-like two-helix,[57] three-helix,[13] and four-helix[58,59] bundles have been designed and structurally characterized. Here we describe a family of dimeric four-helix bundles with properties ranging from highly mobile molten globules to fully native-like folds. An analysis of these proteins has led to a deeper understanding of conformational specificity as well as the design of a new class of metalloproteins that serve as models of larger, functional proteins.

Determinants of the Free Energy Gap

Dimeric four-helix bundles occur in nature in the form of noncovalently self-assembled proteins such as Rop.[60] These proteins adopt bundles with the helices lying in either parallel or antiparallel orientations. Dimeric helical bundles have a large number of possible topologies and hence are ideally suited for studies of the determinants of conformational specificity. While a single-chain four-helix bundle with short connecting loops has only two topologies—clockwise-turning or counterclockwise-turning—a homodimeric four-helix bundle has six possible topologies (Figure 2): Four topologies have clockwise- or counterclockwise-turning loops that connect neighboring helices on either the same or opposite sides of the bundle.[61] Alternatively, the loops can cross diagonally over the tops of the bundle in a motif denoted the bisecting U, which was first described in the *de novo*-designed protein α_2D.[58] Thus, the successful design of dimeric helical bundles requires stabilization of one topology while destabilizing the other five possible topologies.

The residues that comprise the loops between the helices may play an important role in determining conformational specificity. Loop residues between helices can adopt a number of different geometries (ϕ,ψ angles), analogous to the well-known turn motifs that connect

FIGURE 3. Amino acid sequences of the α_2 family, reflecting the hierarchical approach to protein design. Each peptide is comprised of 35 residues with the N-terminus acetylated and the C-terminus amidated. α_2B is comprised solely of leucine residues in the hydrophobic core positions. α_2C is comprised of a more diverse set of nonpolar and aromatic residues in the hydrophobic core positions. α_2D has three additional changes at positions 7, 26, and 30.

antiparallel β-hairpins.[62] Two frequently occurring three-residue loops, helix−(α_L-β-β)−helix and helix−(γ-α_L-β)−helix, are comprised of two common helical motifs.[10,62] First, the C-capping Schellman motif (helix-α_L-β) caps the end of the first helix.[63] Second, the N-capping motif (β-helix) or hydrophobic staple (β-β-helix) caps the start of the second helix.[64] The two motifs are defined by conformational preferences, which are matched with sequence preferences that help define the turn site. Thus, it is not surprising that the conformational specificity, and the thermodynamic stability, of a helix−loop−helix motif depends on both the composition and the length of its loop sequence.

For example, the thermodynamic stability of Rop is modulated over 3.6 kcal mol^{-1}, depending on the nature of a single residue in its helix−(α_L-β)−helix loop.[65] Also, insertion of Gly residues into this loop have been shown to be entropically destabilizing relative to the natural sequence by 0.5−1.0 kcal mol^{-1} for each Gly residue inserted.[66] Thus, a well-designed loop can stabilize a protein by several kilocalories per mole relative to a randomly chosen or poorly designed sequence.[67] In some cases, loop residues can also play an important role in conformational specificity. In one particularly extreme example, a single-residue replacement of Pro for Ala31 in the loop of the dimeric helical bundle, Rop, switches this protein's topology from an anti counterclockwise-turning bundle to a bisecting U![68]

The α_2 family of dimeric four-helix bundle proteins (Figure 3) provides an excellent opportunity to examine other determinants of conformational specificity. An early member of this family, α_2B,[69] consisted of a pair of interconnected, identical helices, whose sequences consisted of only Leu, Glu, and Lys. Although this peptide dimerized to form an exceedingly stable helical bundle, it adopted molten globule-like conformations.[20] The dynamic character of α_2B most probably arose from rapid motions of side chains in the bundle and the formation of multiple, interconverting topologies. With identical sequences for helices 1 and 2 (Figure 3), and a flexible loop, the six different topologies illustrated in Figure 2 should be nearly energetically degenerate and therefore roughly equally populated. In the next generation design, α_2C, the sequence degeneracy of the helices was reduced, and the diversity of the side chains within the core was increased.[70] Half of the Leu residues were replaced by side chains with aromatic and more conformationally restricted β-branched apolar side chains, resulting in a peptide, α_2C, that showed an increase in its ability to adopt

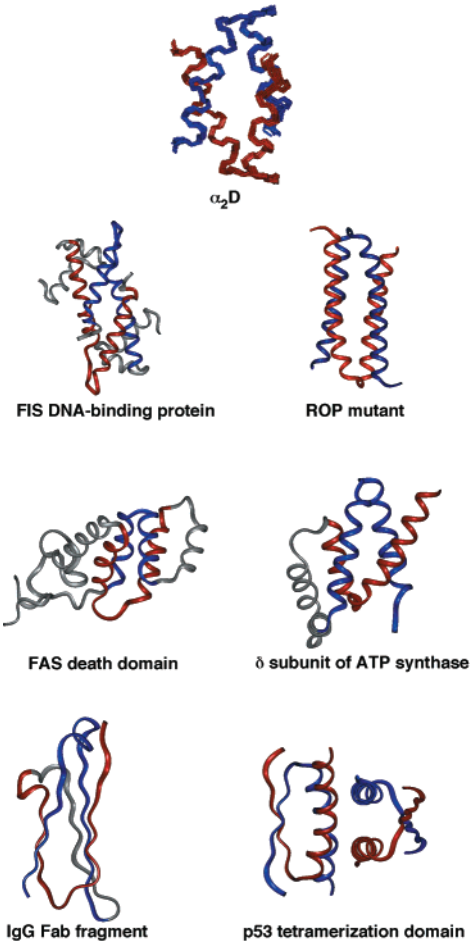

FIGURE 4. The bisecting U motif is a common structural motif. α_2D was the first protein in which the bisecting U motif was recognized.[58] (A backbone overlay of the 10 lowest energy NMR structures is shown.) Subsequently, this motif has been identified in a wide variety of proteins, including other dimeric helical bundles (FIS DNA-binding protein[116] and Rop mutant[68]), intramolecularly folded helical bundles (FAS death domain[117] and N-terminal domain of the δ subunit of the ATPsynthase[118]), and β-sheet proteins (the IgG fold[119]), and multimeric proteins with both α/β structures (p53 tetramerization domain[120]).

a unique conformation.[70] However, a fully native structure was obtained only after changing three more residues. To induce conformational specificity, two apolar residues were changed to polar residues (Leu 7 to Glu and Phe 26 to His),[71] and an interfacial Lys residue at position 30 was switched to His. These substitutions were designed to introduce a metal-ion-binding site, and this protein, α_2D, was indeed found to bind Zn(II). However, even in the absence of metal ions, the resulting dimer behaved like a native protein, although it was thermodynamically less stable than α_2B. Thus, the increase in conformational specificity occurred at the price of decreased thermodynamic stability.

The three-dimensional structure of α_2D was determined at high resolution (0.28 Å rms deviation for backbone atoms) by NMR (Figure 4).[58] To date, this is the most highly resolved solution structure of a designed protein, providing an excellent starting place for the study of conformational specificity. The topology of α_2D was unanticipated and provided the first clear example of the bisecting U motif. Interestingly, examination of the protein database showed that the bisecting U is a ubiquitous folding motif found throughout nature in both monomeric and dimeric α-helical and β-sheet proteins (Figure 4).

α_2D contains three distinct helix/helix interfaces that arise from the symmetry of its structure (Figure 5). The first interface is quite similar to the four equivalent interfaces originally designed in α_2B and resembles an antiparallel Leu zipper with interdigitated Leu residues shielded by electrostatically interacting Glu and Lys residues (Figure 5a). The substitutions required for the native fold are found at the other two helix/helix interfaces. The second interface consists of a well-packed collection of diverse hydrophobes (Figure 5b) and also includes Glu 7, which is essential for adopting a unique conformation. The third interface is also quite rich in interacting aromatic side chains (Figure 5c). This interface also includes the pair of critical His residues, which interact in an interlocking set of intermonomer hydrogen-bonded interactions.

Given that the sequence of α_2D differs from its partially molten globule-like precursor by changes at only three residues, it provides an outstanding opportunity to explore the determinants of conformational specificity. The location of these residues at the surface of the protein was particularly interesting, because many earlier studies into conformational specificity had focused on the role of hydrophobic residues in the interior of the protein. For example, the apolar cores of several natural proteins have been replaced with a single type, or a random collection, of different sized apolar residues with retention of folding and/or activity.[72−77] The changes often result in a loss in stability, and also in an increase in the dynamic properties of the side chains. However, these proteins retained folded structures, suggesting that hydrophobic interactions within the protein core are not the sole contributors to native structure.

A Polar and Solvent-Exposed Residue Is Essential for the Native Structure of α_2D. One of the residues in α_2D that differs from its precursor is at position 7, where an apolar Leu was replaced with a hydrophilic Glu (Figure 3). In the structure of α_2D, Glu 7 is exposed to solvent and adopts multiple rotamers in solution (Figure 6), suggesting that it might not strongly influence the overall thermodynamic stability of the native dimer relative to the unfolded monomers. Thus, its role must be to increase the free energy gap Δ by destabilizing alternatively folded conformations. To test this idea, Glu7 was changed to a variety of hydrophobic and polar residues.[78] As expected, the free energies of folding of these proteins were largely independent of the substitutions and could be explained almost entirely from the expected variations in helix stability. However, there was a very strong correlation between the hydrophobicity of the side chain and the dynamic behavior of the dimers, with aliphatic residues such as Val giving rise to a molten globule-like conformations, as assessed by examining the NMR spectrum and ΔC_p for folding.

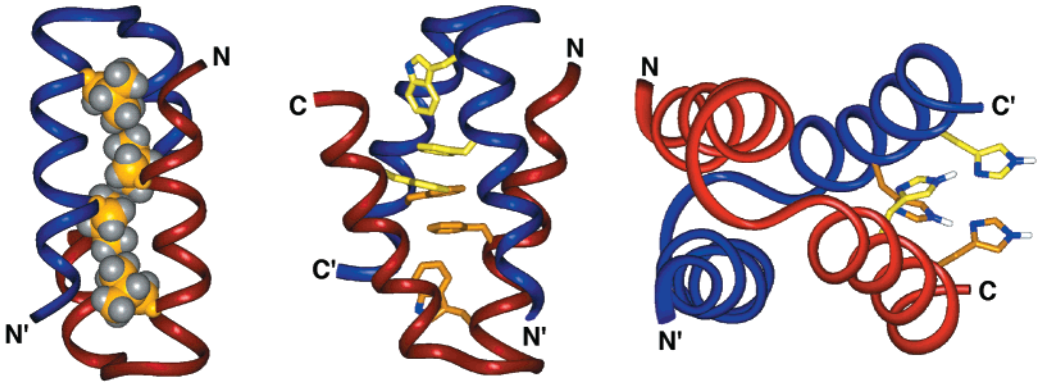

FIGURE 5. α_2D has three nonequivalent interfaces: (a) Interdigitated Leu residues stabilize the interface between helices 1 and 1′. (b) A diverse collection of aromatic and hydrophobic residues stabilize the interface between helices 1 and 2′. (c) Hydrogen-bonded clusters of His residues stabilize the interface between helices 2 and 2′.

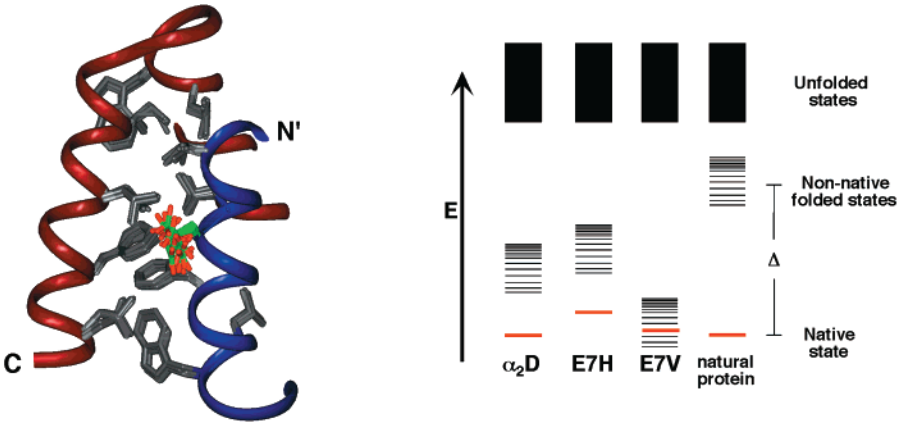

FIGURE 6. Glu7 is not involved in stabilizing the native state of α_2D. Superposition of the 10 lowest energy NMR-derived structures (left panel) shows that Glu7 adopts multiple conformations typical of solvent-exposed residues (hydrophobic core residues in gray). Replacing Glu7 with other polar or nonpolar residues has little effect on the overall stability (ΔG) of the protein but has a large effect on the conformational specificity (Δ) of the protein (right panel). For example, replacing Glu7 with a polar residue (as in E7H) results in a protein with a Δ large enough to adopt a native structure. However, replacing Glu7 with a hydrophobic residue (as in E7V) results in a protein that has lost its ability to adopt a native structure. Note that the number of conformations that E7V adopts cannot be too large because this would result in a large entropic stabilization of this ensemble, which is not observed. In conclusion, Glu7 is involved in "negative design" by increasing Δ through the destabilization of the non-native folded states relative to the native state. We assume that the energy level of the unfolded ensemble is not significantly affected by these substitutions.

These results suggest that Glu 7 of α_2D is a "negative" element of protein design, which promotes native structure in α_2D not by stabilizing the native state, but by destabilizing alternatively folded conformations. This concept is pictured in Figure 6, showing the effect of mutating Glu 7 to either His or Val. A mutation to a polar residue such as His affects the overall stability via helix propensity effects (in this case destabilizing the protein), without markedly changing the size of the free energy gap. By contrast, introduction of Val has little effect on the overall thermodynamic stability of the protein but stabilizes the misfolded conformations, which now are also populated at room temperature. Presumably, these alternatively folded states are dimeric, four-helix bundles with alternative topologies (Figure 3) or more subtle alterations of the bisecting U motif. Clearly, however, position 7 in these alternative folds is less exposed than in the native state, allowing hydrophobic stabilization of their structures.

A Hydrogen-Bonded Cluster of His Residues Is Essential for the Native, Dimeric Structure of α_2D. Substitutional studies of His 26 and His 30 of α_2D have demonstrated that the buried residue, His 26, is also a "negative" element of design, while the interfacial His 30 plays a more "positive" role by specifically stabilizing the native state (Figure 7). When the buried His 26 was changed to the nearly isosteric but apolar residue, Phe, the protein was strongly stabilized to thermal or guanidine-induced denaturation.[79] However, this change also induced a switch from a dimeric to a trimeric oligomeric state. Thus, His 26 serves as a buried hydrogen-bonded group that contributes to the free energy gap between the dimeric and trimeric states of α_2D. The importance of hydrogen-bonding was illustrated through the synthesis of a variant in which His26 was replaced by Asn, which appeared to be able to form an analogous hydrogen bond in models. This variant formed a native, dimeric bundle. Together, these findings are analogous to earlier results from studies on two-stranded coiled coils in which a buried, hydrogen-bonded Asn side chain was changed to Val.[80] This substitution resulted in the formation of

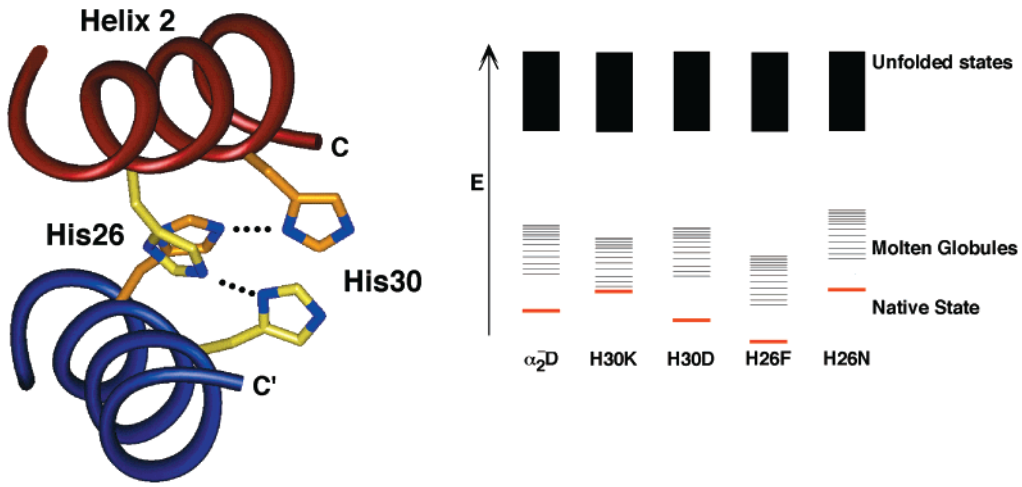

FIGURE 7. Hydrogen-bonding within a His cluster is essential for the native state of α_2D. An interior His (H26) forms a hydrogen bond across the interface with an exterior His (H30′) on the other monomer (left panel). A symmetrical interaction also occurs between H26′ and H30. Substitution of residues that are unable to form hydrogen bonds (as in H30K) results in a protein that has lost all ability to adopt the native structure. Substitution of residues that are capable of forming hydrogen bonds (as in H30D) results in a protein that maintains the native structure (right panel). Note that the non-native states of H30K are depicted slightly lower in energy than those of α_2D because Lys is known have a higher helical propensity than His. In conclusion, the His cluster is involved in "positive design" by increasing Δ through the direct stabilization of the native state relative to the non-native folded states.

trimers, both in an analogue of the coiled coil from GCN4 and in designed peptides. In each of these cases, buried polar residues promote the formation of a native, dimeric state, although this specificity is energetically costly relative to the burial of a hydrophobic side chain.

In α_2D, the more interfacial His 30 plays a more positive role in stabilizing the native structure of the dimer.[79] When this residue was replaced by Lys, which cannot participate in a hydrogen bond with His 26′ of a neighboring helix, the protein was thermodynamically destabilized but still adopted a dimeric, four-helix bundle. However, this protein lost its conformational specificity and adopted molten globule-like conformations. Thus, the introduction of this mutation appeared to destabilize the protein, making its energy level similar to those of alternatively folded states. This finding suggests that the hydrogen-bonded interaction between the His side chains is present only in the native conformation. To demonstrate that a hydrogen bond is, indeed, important for this native structure, His 30 was replaced by Asp, which also can serve as a hydrogen bond acceptor from His 26′. The resulting variant was found to be even more stable than α_2D and retained the ability to adopt a unique conformation.

Hierarchic Principles of Protein Design. Together, these studies on variants of α_2D illustrate the hierarchic principles of protein structure formation. Helix propensity and hydrophobicity play an early and thermodynamically important role in defining the folding of a protein.[81] However, these features are somewhat geometrically non-specific, and more specific interactions are often important for obtaining a uniquely folded structure. Geometric complementarity in the packing of hydrophobic side chains must be important at some level, as the replacement of small, hydrophobic, buried side chains with larger hydrophobes strongly destabilizes the native states of natural proteins.[74] Similarly, cavity-forming mutations also destabilize native proteins.[82] The studies on variants of α_2D illustrate that solvent-exposed residues at interfaces can play an important, even decisive, role in defining structure. In summary, it is important to stress that many interactions are critical for maintaining a large free energy gap, and that these determinants of conformational specificity are typically distributed throughout the protein. Thus, conformational specificity oftentimes reflects the sum of small contributions throughout the chain.

This view of conformational specificity—which stresses the multiplicity of factors required for native structure—extends and is somewhat more detailed than models that stress the importance of the pattern of hydrophobic and hydrophilic residues in a sequence for defining the unique tertiary structure of a protein. Experiments in which the cores of natural proteins are repacked with a designed or randomly selected collection of hydrophobic side chains have shown that such proteins often retain a native-like structure, indicating that a variety of combinations of side chains can give rise to native folds.[72−77] It is tempting to conclude from these studies that the packing of hydrophobic side chains is not an important determinant of conformational specificity, and that only the pattern of hydrophobic and hydrophilic residues actually matters. However, only a fraction of the sequence of a protein is normally varied in these studies, and stabilizing interfacial interactions remain constant. Further, a critical feature in these studies is that the sum of the volumes of the side chains should be roughly equivalent to that of the native sequence. Too large a summed volume gives rise to molten globule-like structures, while too small volumes lead to marginal stability.[83] Also, if the side chains are unable to achieve a good geometric fit, the interior side chains can be less well ordered than in natural proteins.[84]

A second series of studies appear to support the hypothesis that binary patterning is the sole contributor

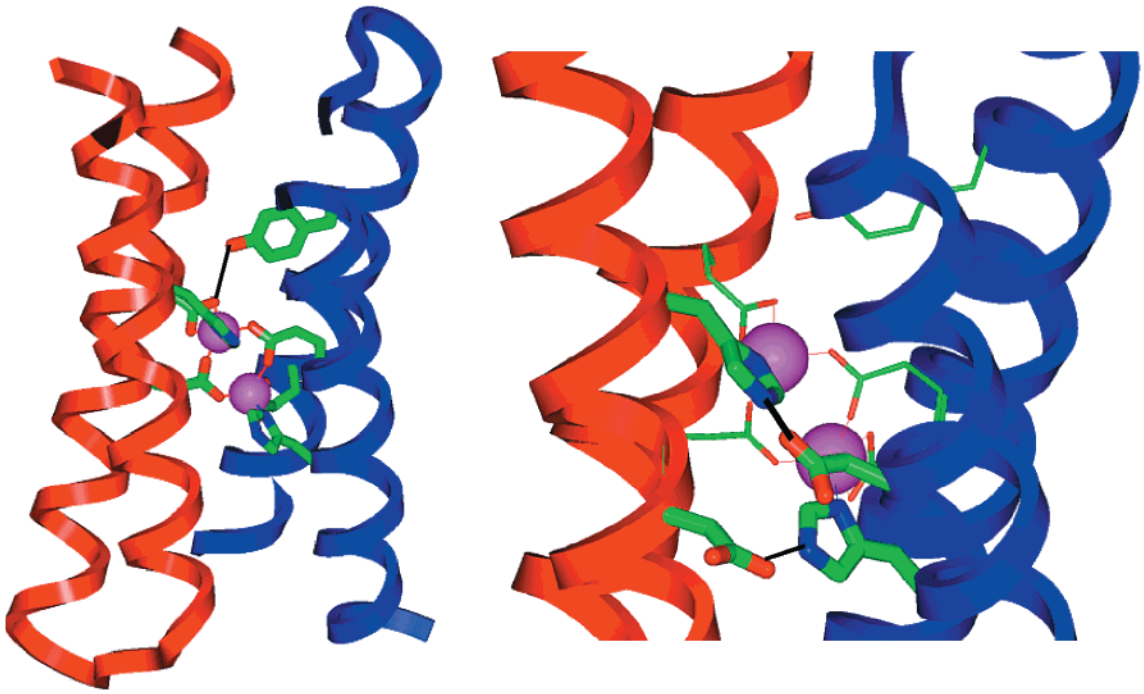

FIGURE 8. X-ray structure of the di-Zn(II) form of DF1 (2.5 Å resolution), which is nearly identical to the intended design. The backbone of the structure plus the ligands are shown in the two views. At left, a Tyr phenolic group hydrogen bonds to a Glu carboxylate (a second, symmetry-related Tyr-Glu interaction is not shown for clarity). At right, hydrogen bonds between Asp carboxylates and His side chains are shown.

of a fold.[85−87] In this work, protein sequences were prepared that were random within the constraints of a hydrophobic/hydrophilic pattern. For example, a series of single-chain four-helix bundles with the same hydrophobic pattern as α_2B has been shown to adopt compact structures with significant helical content. Individual members of the library have various degrees of native-like behavior, as assessed from their thermodynamic properties and the degree of dispersion of their NMR spectra. However, structures have not been determined for these and other random-sequence proteins, so it is not clear what fraction of the members of these libraries have a uniquely folded structure.

Thus, the hydrophobic/hydrophilic pattern is an important driving force that strongly predisposes a protein toward a family of related folds (e.g., clockwise- and counterclockwise-turning four-helix bundles). For some folds, such as coiled coils, a hydrophobic period may be sufficient to specify an elongated helical bundle. However, our studies with α_2D suggest that more fine-tuned interactions may frequently contribute to a uniquely folded state.

The process of protein design has certain parallels with the folding of natural proteins. Forming collapsed molten globule-like states is relatively rapid, indicating that there is a small free energy barrier to this step in folding, and this is often the easy part of the design process. The slow step in folding and the difficult step in design is usually the transition from a relatively noncooperative condensed state to a fully native state.

From Structure to Function: Design of Metalloproteins. The above studies indicate that *de novo* protein design has increased our understanding of how amino acid sequence specifies native protein structure. The rules for designing native-like four-helix bundle proteins are now well in hand, and it is possible to use this technology to address how an amino acid sequence specifies catalytic function. Toward this goal, many groups are pursuing the design of metalloproteins (reviewed recently[10,88]). There are basically three approaches to designing mimics of metal-binding sites: In the first, one uses automated methods to graft a metal-binding site into the structure of a natural protein of known structure.[89−92] A second approach has been to design small, flexible peptides that, although intrinsically flexible, are nevertheless able to fold around a metal ion. Examples of such peptides include Cu(II)-binding motifs related to Gly-Gly-His,[93−96] as well as small peptides that assemble into Fe_4S_4 clusters.[97,98]

However, complete control of a cofactor's environment might be best effected through the *de novo* design[99−102] of proteins whose active sites are defined by the favorable free energy of folding of the polypeptide chain. There has also been much progress made in the *de novo* design of proteins that bind metalloporphyrins[92−95,103−105] and Zn(II).[20,106] The structures of some simple porphyrin peptide complexes have been solved by NMR.[107,108] However, the structures of larger designed proteins with bound cofactors have not been solved, possibly because they have dynamically averaging structures.[41,104,109,110]

As the principles and methods for *de novo* protein design have matured, it has recently become possible to design structurally defined models for metalloproteins. The diiron class of proteins, which are capable of performing a diverse range of functions yet contain a four-

helix bundle at the heart of the protein, provided an attractive target for these studies. Individual members of this group of proteins either bind oxygen reversibly or catalyze a diverse set of oxygen-dependent reactions, including epoxidation, desaturation, monohydroxylation, and radical formation. We therefore sought to design minimal models for these large, highly complex proteins to address how the amino acid sequence modulates the type of chemistry performed. Subsequent studies of the spectroscopic properties and chemical reactivities of these designed models (with variations within their active sites) will allow one to systematically address how changes in polarity, electrostatics, and solvent accessibility of the active site define the activities of the natural proteins. Furthermore, by introducing non-natural metal ions, it might be possible to design catalysts for a variety of chemical transformations.

A geometric analysis of six natural diiron proteins suggested that they all may have evolved from a dimer of helix−loop−helix motifs.[100,111] Each of the helices donates a single Glu residue, and the second helix of each helix−loop−helix donates a single His residue to the diiron site. In the diferrous form, the ligands often show approximate two-fold symmetry with two bridging Glu residues, two Glu residues that bind a single metal ion, and two His residues. The ligands are arranged in an approximate square pyramidal array, with the His side chains trans to the empty sites (which presumably bind oxygen). On the basis of this analysis, we designed a model diiron protein, DF1, with four Glu residues and two His residues projecting toward the center of the bundle.[111] The backbone structure was dictated by a mathematical parametrization of the structures of natural diiron proteins, thus allowing the design process to proceed in an objective manner that is consistent with fully automated approaches to protein design.[50,112,113] Hydrogen-bonded interactions were introduced to stabilize the ligands in the proper orientation for binding to iron. In particular, an Asp residue was introduced to serve as a hydrogen bond acceptor for the buried His, in precisely the same manner as a His 30-to-Asp variant of α_2D.[79] A hydrogen bond was also modeled between a buried Tyr and the nonbridging Glu residue in the core of the structure. The remaining positions in the core were chosen to stabilize the geometry of the active site, and the interfacial and exposed side chains were chosen using the principles described previously.[60,114]

The crystal structures of the di-Zn(II)[115] and di-Mn(II) forms of the resulting protein conform precisely to the design, with an RMSD of approximately 1.0 Å between the model and the structure (Figure 8). DF1 has a well-defined dimetal-binding site with two bridging Glu carboxylates, two chelating Glu carboxylates, and two His ligands. Further, the intended second-shell ligands form hydrogen bonds to the first-shell ligands. The protein also folds in solution in the absence of metal ions, and preliminary NMR investigations of the solution structure dimer suggest that the native fold and metal-binding site are largely preorganized in the apo-protein. Clearly, this highly simplified minimal protein will serve as a robust template for examining how a protein is able to bind and tune the reactivity of transition metal ions. Further, the mathematical approach used in the design of DF1 can be extended to the design of a variety of metalloproteins.

Conclusion

De novo protein design has contributed significantly to our understanding of the principles of protein folding and stabilization. As this understanding has matured and tools for design have become available, it has become possible to design proteins of increasing complexity. Our early studies in the design of diiron proteins suggest that it should now be possible to approach the *de novo* design of highly functionalized metalloproteins with well-defined three-dimensional structures. Thus, in the future, the attention of protein design may increasingly shift from the problem of protein folding to the design of proteins with specific biological and catalytic functions.

References

(1) Creighton, T. E. Protein folding. *Biochem. J.* **1990**, *270*, 1−16.
(2) Cordes, M. H. J.; Davidson, A. R.; Sauer, R. T. Sequence space, folding and protein design. *Curr. Opin. Struct. Biol.* **1996**, *6*, 3−10.
(3) Dill, K. A. Dominant forces in protein folding. *Biochemistry* **1990**, *31*, 7133−7155.
(4) Dobson, C. M. Characterization of protein folding intermediates. *Curr. Opin. Struct. Biol.* **1991**, *1*, 22−27.
(5) Eaton, W. A.; Thompson, P. A.; Chan, C.-K.; Hagen, S. J.; Hofrichter, J. Fast events in protein folding. *Structure* **1996**, *4*, 1133−1139.
(6) Fersht, A. R. Optimization of rates of protein folding: The nucleation−condensation mechanism and its implications. *Proc. Natl. Acad. Sci. U.S.A.* **1995**, *92*, 10869−10873.
(7) Honig, B.; Yang, A.-S. Free energy balance in protein folding. *Adv. Prot. Chem.* **1995**, *46*, 27−58.
(8) Shortle, D. Probing the determinants of protein folding and stability with amino acid substitutions. *J. Biol. Chem.* **1989**, *264*, 5315−5318.
(9) Bryson, J. W.; Betz, S. F.; Lu, H. S.; Suich, D. J.; Zhou, H. X.; O'Neil, K. T.; DeGrado, W. F. Protein design: a hierarchic approach. *Science* **1995**, *270*, 935−941.
(10) DeGrado, W. F.; Summa, C. M.; Pavone, V.; Nastri, F.; Lombardi, A. De novo design and structural characterization of proteins and metalloproteins. *Annu. Rev. Biochem.* **1999**, *68*, 779−819.
(11) Baltzer, L.; Broo, K. S. De novo designed polypeptide catalysts with adopted folded structures. *Biopolymers* **1998**, *47*, 31−40.
(12) Gouda, H.; Torigoe, H.; Saito, A.; Arata, Y.; Shimada, I. Three-dimensional solution structure of the B domain of *staphylococcal* protein A: Comparisons of the solution and crystal structures. *Biochemistry* **1992**, *31*, 9665−9672.
(13) Walsh, S. T. R.; Cheng, H.; Bryson, J. W.; Roder, H.; DeGrado, W. F. Solution structure and dynamics of a de novo designed three helix bundle protein. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *96*, 5486−5491.
(14) Lesk, A. M.; Chothia, C. How different amino acid sequences determine similar protein structures: the structure and evolutionary dynamics of the globins. *J. Mol. Biol.* **1980**, *136*, 225−270.
(15) Dunbrack, R. L., Jr.; Cohen, F. E. Bayesian statistical analysis of protein side-chain rotamer preferences. *Protein Sci.* **1997**, *6*, 1661−81.
(16) Dunbrack, R. L., Jr.; Karplus, M. Conformational analysis of the backbone-dependent rotamer preferences of protein sidechains. *Nature Struct. Biol.* **1994**, *1*, 334−340.
(17) Dunbrack, R. L., Jr.; Karplus, M. Backbone-dependent rotamer library for proteins. Application to side-chain prediction. *J. Mol. Biol.* **1993**, *230*, 543−574.
(18) McGregor, M. J.; Islam, S. A.; Sternberg, M. J. Analysis of the relationship between side-chain conformation and secondary structure in globular proteins. *J. Mol. Biol.* **1987**, *198*, 295−310.
(19) Schrauber, H.; Eisenhaber, F.; Argos, P. Rotamers: To be or not to be? *J. Mol. Biol.* **1993**, *230*, 592−612.

(20) Handel, T. M.; Williams, S. A.; DeGrado, W. F. Metal Ion-Dependent Modulation of the Dynamics of a Designed Protein. *Science* **1993**, *261*, 879−885.

(21) Betz, S. F.; Raleigh, D. P.; DeGrado, W. F. De novo protein design: from molten globules to native-like states. *Curr. Opin. Struct. Biol.* **1993**, *3*, 601−610.

(22) Kuwajima, K. The molten globule state as a clue for understanding the folding and cooperativity of globular-protein structure. *Proteins* **1989**, *6*, 87−103.

(23) Ptitsyn, O. B.; Pain, R. H.; Semisotnov, G. V.; Zerovnik, E.; Razgulyaer, O. L. Evidence for a molten globule state as a general intermediates in protein folding. *FEBS Lett.* **1990**, *262*, 20−24.

(24) Lazaridis, T.; Tobias, D. J.; C. L. Brooks, I.; Paulaitis, M. E. Reaction paths and free energy profiles for conformational transitions: An internal coordinate approach. *J. Chem. Phys.* **1991**, *95*, 7612−7625.

(25) Mirny, L. A.; Abkevich, V.; Shakhnovich, E. I. Universality and diversity of the protein folding scenarios: a comprehensive analysis with the aid of a lattice model. *Folding Design* **1996**, *1*, 103−116.

(26) Onuchic, J. N.; Wolynes, P. G.; Luthey-Schulten, Z.; Socci, N. D. Towards an outline of the topography of a realistic protein folding funnel. *Proc. Natl. Acad. Sci. U.S.A.* **1995**, *92*, 3626−3630.

(27) Ingram, V. Gene mutation in human hemoglobin: the chemical difference between normal and sickle cell hemoglobin. *Nature* **1957**, *180*, 326−328.

(28) Pauling, L. Sickle cell anemia: a molecular disease. *Science* **1949**, *110*, 543−548.

(29) Booth, D.; Sunde, M.; Bellotti, V.; Robinson, C.; Hutchinson, W.; Fraser, P.; Hawkins, P.; Dobson, C.; Radford, S.; Blake, C.; Pepys, M. Instability, unfolding and aggregation of human lysozyme variants underlying amyloid fibrillogenesis. *Nature* **1997**, *385*, 787−793.

(30) Dobson, C. Protein misfolding, Evolution and Disease. *Trends Biochem. Sci.* **1999**, *24*, 329−332.

(31) Kelly, J. The alternative conformations of amyloidogenic proteins and their multistep assembly patterns. *Curr. Opin. Struct. Biol.* **1998**, *8*, 101−106.

(32) Klabunde, T.; Petrassi, H.; Oza, V.; Raman, P.; Kelly, J.; Sacchettini, J. Rational design of potent human transthyretic amyloid disease inhibitors. *Nat. Struct. Biol.* **2000**, *7*, 312−321.

(33) Mihara, H.; Takahashi, Y.; Ueno, A. Design of peptides undergoing self-catalytic α-to-β transition and amyloidogenesis. *Biopolymers* **1998**, *47*, 83−92.

(34) Lindquist, S. Mad cows meet psi-chotic yeast: the expansion of the prion hypotheses. *Cell* **1997**, *89*, 495−498.

(35) Prusiner, S. Novel proteinaceous infectious particles cause scrapie. *Science* **1982**, *216*, 136−144.

(36) Prusiner, S. Prions *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 13363−13368.

(37) Richardson, J. S.; Richardson, D. C. The de novo design of protein structures. *Trends Biochem. Sci.* **1989**, *14*, 304−309.

(38) Ilyina, E.; Roongta, V.; Mayo, K. H. NMR structure of a de novo designed, peptide 33mer with two distinct, compact β-sheet folds. *Biochemistry* **1997**, *36*, 5245−5250.

(39) Mayo, K. H.; Ilyina, E.; Park, H. A recipe for designing water-soluble, β-sheet-forming peptides. *Protein Sci.* **1996**, *5*, 1301−1315.

(40) Gibney, B. R.; Rabanal, F.; Skalicky, J. J.; Wand, A. J.; Dutton, P. L. Design of a unique protein scaffold for maquettes. *J. Am. Chem. Soc.* **1997**, *119*, 2323−2324.

(41) Skalicky, J.; Gibney, B.; Rabanal, F.; Urbauer, R.; Dutton, P.; Wand, A. Solution structure of a designed four α-helix bundle maquette scaffold. *J. Am. Chem. Soc.* **1999**, *121*, 4941−4951.

(42) Gibney, B. R.; Rabanal, F.; Reddy, K. S.; Dutton, P. L. Effect of four helix bundle topology on heme binding and redox properties. *Biochemistry* **1998**, *37*, 4635−4643.

(43) Baltzer, L. Functionalization of designed folded polypeptides. *Curr. Opin. Struct. Biol.* **1998**, *8*, 466−470.

(44) Broo, K. S.; Nilsson, H.; Nilsson, J.; Flodberg, A.; Baltzer, L. Cooperative nucleophilic and general-acid catalysis by the HisH-(+)-His pair and arginine transition state binding in catalysis of ester hydrolysis reactions by designed helix−loop−helix motifs. *J. Am. Chem. Soc.* **1998**, *120*, 4063−4068.

(45) Wu, L.; Laub, P.; Elove, G.; Carey, J.; Roder, H. A noncovalent peptide complex as a model for an early folding intermediate of cytochrome c. *Biochemistry* **1993**, *32*, 10271−10276.

(46) Demarest, S.; Fairman, R.; Boice, J.; Raleigh, D. Defining the core structure of the α-lactalbumin molten globule. *J. Mol. Biol.* **1999**, *1999*, 213−221.

(47) Struthers, M.; Ottesen, J. J.; Imperiali, B. Design and NMR analyses of compact, independently folded BBA motifs. *Folding Design* **1998**, *3*, 95−103.

(48) Struthers, M. D.; Cheng, R. P.; Imperiali, B. Economy in protein design: Evolution of a metal-independent $\beta\,\beta\,\alpha$ motif based on the zinc finger domains. *J. Am. Chem. Soc.* **1996**, *118*, 3073−3081.

(49) Struthers, M. D.; Cheng, R. P.; Imperiali, B. Design of a monomeric 23-residue polypeptide with defined tertiary structure. *Science* **1996**, *271*, 342−345.

(50) Dahiyat, B. I.; Mayo, S. L. De novo protein design: fully automated sequence selection. *Science* **1997**, *278*, 82−87.

(51) Dahiyat, B. I.; Sarisky, C. A.; Mayo, S. L. De novo protein design: towards fully automated sequence selection. *J. Mol. Biol.* **1997**, *273*, 789−96 [see comments].

(52) Schenck, H. L.; Gellman, S. H. Use of a designed triple-stranded antiparallel β-sheet to probe β-sheet cooperativity in aqueous solution. *J. Am. Chem. Soc.* **1998**, *120*, 4869−4870.

(53) Kortemme, T.; Ramirez-Alvarado, M.; Serrano, L. Design of a 20-amino acid, three-stranded β-sheet protein. *Science* **1998**, *281*, 253−256.

(54) Sharman, G. J.; Searle, M. S. Cooperative interaction between the three strands of a designed antiparallel β-sheet. *J. Am. Chem. Soc.* **1998**, *120*, 5291−5300.

(55) Das, C.; Raghothama, S.; Balaram, P. A designed three-stranded β-sheet peptide as a multiple β-hairpin model. *J. Am. Chem. Soc.* **1998**, *120*, 5812−5813.

(56) Das, C.; Raghothama, S.; Balaram, P. A four stranded β-sheet in a designed, synthetic polypeptide. *Chem. Commun.* **1999**, 967−968.

(57) Fezoui, Y.; Connolly, P. J.; Osterhout, J. J. Solution structure of α t α, a helical hairpin peptide of de novo design. *Protein Sci.* **1997**, *6*, 1869−1877.

(58) Hill, R. B.; DeGrado, W. F. Solution structure of α_2D, a native-like de novo designed protein. *J. Am. Chem. Soc.* **1998**, *120*, 1138−1145.

(59) Schafmeister, C. E.; LaPorte, S. L.; Mierke, L. J.; Stroud, R. M. A designed four helix bundle with native-like structure. *Nature Struct. Biol.* **1997**, *4*, 1039−1046.

(60) Banner, D. W.; Kokkinidis, M.; Tsernoglou, D. Structure of ColE1 rop protein at 1.7 Å resolution. *J. Mol. Biol.* **1987**, 657−675.

(61) Betz, S. F.; DeGrado, W. F. Controlling topology and native-like behavior of de novo-designed peptides: design and characterization of antiparallel four-stranded coiled coils. *Biochemistry* **1996**, *35*, 6955−62.

(62) Efimov, A. V. Patterns of loop regions in proteins. *Curr. Opin. Struct. Biol.* **1993**, *3*, 379−384.

(63) Schellman, C. In *The α_L-conformation at the ends of helices*; Schellman, C., Ed.; Elsevier: North-Holland, New York, 1980; pp 53−61.

(64) Muñoz, V.; Blanco, F. J.; Serrano, L. The hydrophobic-staple motif and a role for loop-residues in α-helix stability and protein folding. *Nature Struct. Biol.* **1995**, *2*, 380−385.

(65) Predki, P. F.; Agrawal, V.; Brunger, A. T.; Regan, L. Amino-acid substitutions in a surface turn modulate protein stability. *Nature Struct. Biol.* **1996**, *3*, 54−58.

(66) Nagi, A. D.; Anderson, K. S.; Regan, L. Using loop length variants to dissect the folding pathway of a four-helix-bundle protein. *J. Mol. Biol.* **1999**, *286*, 257−265.

(67) Zhou, H. X.; Hoess, R. H.; DeGrado, W. F. *In vitro* evolution of thermodynamically stable turns. *Nature Struct. Biol.* **1996**, *3*, 446−451.

(68) Glykos, N.; Cesareni, G.; Kokkinidis, M. Protein plasticity to the extreme: changing the topology of a four α-helical bundle with a single amino acid substitution. *Structure* **1999**, *7*, 597−603.

(69) Ho, S. P.; DeGrado, W. F. Design of a 4-helix bundle protein: synthesis of peptides which self-associate into a helical protein. *J. Am. Chem. Soc.* **1987**, *109*, 6751−6758.

(70) Raleigh, D. P.; DeGrado, W. F. A *de novo* designed protein shows a thermally induced transition from a native to a molten globule-like state. *J. Am. Chem. Soc.* **1992**, *114*, 10079−10081.

(71) Raleigh, D. P.; Betz, S. F.; DeGrado, W. F. A *de novo* designed protein mimics the native state of natural proteins. *J. Am. Chem. Soc.* **1995**, *117*, 7558−7559.

(72) Axe, D. D.; Foster, N. W.; Fersht, A. Active barnase variants with completely random hydrophobic cores. *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 5590−5594.

(73) Gassner, N. C.; Baase, W. A.; Matthews, B. W. A test of the "jigsaw puzzle" model for protein folding by multiple methionine substitutions within the core of T4 lysozyme. *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 12155−12158.

(74) Matthews, B. W. Studies on protein stability with T4 lysozyme. *Adv. Protein Chem.* **1995**, *46*, 249−278.

(75) Lim, W. A.; Hodel, A.; Sauer, R. T.; Richards, F. M. The crystal structure of a mutant protein with altered but improved hydorphobic core packing. *Proc. Natl. Acad. Sci. U.S.A.* **1994**, *91*, 423−427.

(76) Lim, W. A.; Farruggio, D. C.; Sauer, R. T. Structural and energetic consequences of disruptive mutations in a protein core. *Biochemistry* **1992**, *31*, 4324−4333.
(77) Lim, W. A.; Sauer, R. T. The role of internal packing interaction in determining the structure and stability of a protein. *J. Mol. Biol.* **1991**, *219*, 359−376.
(78) Hill, R. B.; DeGrado, W. F. A polar, solvent-exposed residue can be essential for native protein structure. *Structure* **2000**, *8*, 471−479.
(79) Hill, R. B.; Hong, J.-K.; DeGrado, W. F. Hydrogen bonded cluster can specify the native state of a protein. *J. Am. Chem. Soc.* **2000**, *122*, 746−747.
(80) Harbury, P. B.; Zhang, T.; Kim, P. S.; Alber, T. A switch between two-, three-, and four-stranded coiled coils. *Science* **1993**, *262*, 1401−1407.
(81) Sauer, R. T. Protein folding from a combinatorial perspective. *Folding Design* **1996**, *1*, R27−R30.
(82) Xu, J.; Baase, W. A.; Baldwin, E.; Matthews, B. W. The response of T4 lysozyme to large-to-small substitutions within the core and its relation to the hydrophobic effect. *Protein Sci.* **1998**, *7*, 158−177.
(83) Dahiyat, B. I.; Mayo, S. L. Probing the role of packing specificity in protein design. *Proc. Natl. Acad. Sci. U.S.A.* **1997**, *94*, 10172−10177.
(84) Johnson, E.; Lazar, G.; Desjarlais, J.; Handel, T. Solution structure and dynamics of a designed hydrophobic core variant of ubiquitin. *Structure* **1999**, *7*, 967−976.
(85) Kamtekar, S.; Schiffer, J. M.; Xiong, H.; Babik, J. M.; Hecht, M. H. Protein design by binary patterning of polar and nonpolar amino acids. *Science* **1993**, *262*, 1680−5168.
(86) Roy, S.; Helmer, K. J.; Hecht, M. H. Detecting native-like properties in combinatorial libraries of de novo proteins. *Folding Design* **1997**, *2*, 89−92.
(87) Roy, S.; Hecht, M. Cooperative thermal denaturation of proteins designed by binary patterning of polar and nonpolar amino acids. *Biochemistry* **2000**, *39*, 4603−4607.
(88) Cheng, R. P. In *Design and synthesis of metallopeptides: Incorporation of unnatural amino acids and construction of a structural template*; Cheng, R. P., Ed.; California Institute of Technology: Pasadena, 1998; p 170.
(89) Hellinga, H. W. The construction of metal centers in proteins by rational design. *Folding Design* **1998**, *3*, R1−R8.
(90) Desjarlais, J. R.; Clarke, N. D. Computer search algorithms in protein modification and design. *Curr. Opin. Struct. Biol.* **1998**, *8*, 471−475.
(91) Sigman, J. A.; Kwok, B. C.; Gegenbach, A.; Lu, Y. Design and creation of a Cu(II)-binding site in cytochrome c peroxidase that mimics the Cu_B-heme center in terminal oxidases. *J. Am. Chem. Soc.* **1999**, *121*, 8949−8950.
(92) Brittain, I. J.; Huang, X.; Long, E. C. Selective recognition and cleavage of RNA loop structures by Ni(II).Xaa-Gly-His metallopeptides. *Biochemistry* **1998**, *37*, 12113−12120.
(93) Hay, R. W.; Hassan, M. M.; You-Quan, C. Kinetic and thermodynamic studies of the copper (II) and nickel (II) complexes of glycylglycyl-L-histidine. *J. Inorg. Biochem.* **1993**, *52*, 17−25.
(94) Stemmler, A. J.; Burrows, C. J. The sal-XH motif for metal-mediated oxidative DNA-peptide cross-linking. *J. Am. Chem. Soc.* **1999**, *121*, 6956−6957.
(95) Torrado, A.; Walkup, G. K.; Imperiali, B. Exploiting polypeptide motifs for the design of selective Cu(II) ion chemosensors. *J. Am. Chem. Soc.* **1998**, *120*, 609−610.
(96) Gibney, B. R.; Mulholland, S. E.; Rabanal, F.; Dutton, P. L. Ferredoxin and ferredoxin-heme maquettes. *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 15041−15046.
(97) Coldren, C.; Hellinga, H. W.; Caradonna, J. P. The rational design and construction of a cuboidal iron−sulfur protein. *Proc. Natl. Acad. Sci. U.S.A.* **1997**, *94*, 6635−6640.
(98) Lombardi, A.; Marasco, D.; Maglio, O.; Di Constanzo, L.; Nastri, F.; Pavone, V. Miniaturized metalloproteins: Application to iron−sulfur proteins. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, in press.
(99) Richardson, J.; Richardson, D. C.; Tweedy, N. B.; Gernert, K. M.; Quinn, T. P.; Hecht, M. H.; Erickson, B. W.; Yan, Y.; McClain, D. M. E.; Surles, M. C. Looking at proteins: representations, folding, packing, and design. Biophysical Society National Lecture. *Biophys. J.* **1992**, *63*, 1186−1209.
(100) Summa, C. M.; Lombardi, A.; Lewis, M.; DeGrado, W. F. Tertiary templates for the design of diiron proteins. *Curr. Opin. Struct. Biol.* **1999**, *9*, 500−508.
(101) Ghadiri, M.; Case, M. De-novo design of a novel heterodinuclear 3-helix bundle metalloprotein. *Angew. Chem., Int. Ed. Engl.* **1993**, *32*, 1594−1597.
(102) Mutz, M.; Case, M.; Wishart, J.; Ghadiri, M.; McLendon, G. De novo design of protein function: Predictable structure-function relationships in synthetic redox proteins. *J. Am. Chem. Soc.* **1999**, *121*, 858−859.
(103) Rojas, N. R.; Kamtekar, S.; Simons, C. T.; McLean, J. E.; Vogel, K. M.; Spiro, T. G.; Farid, R. S.; Hecht, M. H. De novo heme proteins from designed combinatorial libraries. *Protein Sci.* **1997**, *6*, 2512−2524.
(104) Choma, C. T.; Lear, J. D.; Nelson, M. J.; Dutton, P. L.; Robertson, D. E.; DeGrado, W. F. Design of a heme-binding four-helix bundle. *J. Am. Chem. Soc.* **1994**, *116*, 856−865.
(105) Gibney, B. R.; Dutton, P. L. Histidine placement in de novo-designed heme proteins [In Process Citation]. *Protein Sci.* **1999**, *8*, 1888−1898.
(106) Regan, L.; Clarke, N. D. A tetrahedal zinc(II)-binding site introduced into a designed protein. *Biochemistry* **1990**, *29*, 10878−10883.
(107) D'Auria, G.; Magio, O.; Nastri, F.; Lombardi, A.; Mazzeo, M.; Morelli, G.; Paolillo, L.; Pedone, C.; Pavone, V. Hemoprotein models based on a covalent helix-heme-helix sandwich. 2. Structural characterization of Co−III mimochrome I Delta and Delta isomers. *Chem.−Eur. J.* **1997**, *3*, 350−362.
(108) Nastri, F.; Lombardi, A.; D'Andrea, L. D.; Sanseverino, M.; Maglio, O.; Pavone, V. Miniaturized hemoproteins. *Biopolymers* **1998**, *47*, 5−22.
(109) Gibney, B. R.; Rabanal, F.; Dutton, P. L. Synthesis of novel proteins. *Curr. Opin. Chem. Biol.* **1997**, *1*, 537−542.
(110) Robertson, D. E.; Farid, R. S.; Moser, C. C.; Urbauer, J. L.; Mulholland, S. E.; Pidikiti, R.; Lear, J. D.; Wand, A. J.; DeGrado, W. F.; Dutton, P. L. Design and synthesis of multi-haem proteins. *Nature* **1994**, *368*, 425−432.
(111) Lombardi, A.; Summa, C.; Geremia, S.; Randaccio, L.; Pavone, V.; DeGrado, W. F. Retrostructural analysis of metalloproteins: Application to the design of a minimal model for diiron proteins. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *97*, 6298−6305.
(112) Lazar, G.; Handel, T. Hydrophobic core packing and protein design. *Curr. Opin. Chem. Biol.* **1998**, *2*, 675−679.
(113) Street, A.; Mayo, S. Computational protein design. *Structure* **1999**, *7*, R105−R109.
(114) Betz, S. F.; Liebman, P. A.; DeGrado, W. F. De novo design of native proteins: characterization of proteins intended to fold into antiparallel, rop-like, four-helix bundles. *Biochemistry* **1997**, *36*, 2450−2458.
(115) Lombardi, A.; DeGrado, W., manuscript in preparation, 2000.
(116) Kostrewa, D.; Granzin, J.; Koch, C.; Choe, H.-W.; Raghunathan, S.; Wolf, W.; Labahn, J.; Kahmann, R.; Saenger, W. Three-dimensional structure of the *E. coli* DNA-binding protein FIS. *Nature* **1991**, *349*, 178−180.
(117) Huang, B.; Eberstadt, M.; Olejniczak, E.; Meadows, R.; Fesik, S. NMR structure and mutagenesis of the Fas (APO-1/CD95) death domain. *Nature* **1996**, *384*, 638−641.
(118) Wilkens, S.; Dunn, S.; Chandler, J.; Dahlquist, F.; Capaldi, R. Solution structure of the N-terminal domain of the delta subunit of the E. coli ATP synthase. *Nature Struct. Biol.* **1997**, *4*, 198−201.
(119) Rini, J.; Schulze-Gahmen, U.; Wilson, I. Structural evidence for induced fit as a mechanism for antigen−antibody recognition. *Science* **1992**, *255*, 959−963.
(120) Lee, W.; Harvey, T.; Yin, Y.; Yau, P.; Litchfield, D.; Arrowsmith, C. Solution structure of the tetrameric minimum transforming domain of p53. *Nature Struct. Biol.* **1994**, *1*, 877−890.

AR970004H