De Novo Design of Proteins—What Are the Rules?

Lars Baltzer,* Helena Nilsson, and Jonas Nilsson

Department of Chemistry, Linköping University, 581 83 Linköping, Sweden

Received March 1, 2001

Contents

I. Introduction

The understanding at the atomic level of how and why proteins fold is crucial to the understanding of the life processes and to our ability of designing novel proteins with nativelike properties. Not only is the elucidation of the underlying principles of great fundamental interest, but so is the possibility of introducing tailor-made functions into tailor-made proteins; the perspectives in biocatalysis and in biomolecular recognition are truly exciting. Understanding the relationship between sequence and properties is also central to the dramatic development that has followed upon the mapping of the human genome. Many diseases depend on the misfolding of proteins, and the actual function of the large majority of proteins that are currently known only by their amino acid sequences is the focus of attention of medicinal chemists and pharmaceutical industries. The prospect of major advances in drug discovery and drug development appears to drive tremendous efforts that go into solving the complex problem of translating amino acid sequences into protein structure and function.

The problem can be and is approached in many ways, but ultimately our understanding of protein structure must be put to the litmus test of designing from scratch unnatural sequences that fold according to prediction. In designing a protein from first principles, we rely on much knowledge that has been generated from the analysis of naturally occurring proteins in terms of the structures of the secondary structure elements, i.e., the helix, the β-sheet, loops, and turns. Much work on propensities of amino acids for secondary structure formation[1] also originates in the analysis of proteins for which the three-dimensional structures have been obtained from crystallography or NMR spectroscopy, although context-free propensity scales have been generated in designed peptide scaffolds.[2−4] While the study of native proteins has provided a wealth of information about protein folding, structure, and function in proteins that are known to fold, an understanding of the basic principles that will eventually allow us to design new proteins from scratch has not emerged. The study of fundamental principles of biomolecular structure in naturally occurring enzymes, messengers, and receptors is complicated by the evolutionary baggage that includes memories of primeval functions. Each amino acid residue may have many functions, some of which are no longer obvious, and mutations are likely to have given rise to functional changes that are mediated through indirect and complex pathways that are difficult to deconvolute. De novo designed proteins are virgin with respect to structure and function and therefore possible to analyze in fundamental detail. The approach to understanding biomolecular structure and function by designing proteins de novo is therefore the only one that can provide unequivocally the answer to the question of whether we in fact know how to construct new proteins with tailored structures and properties. The advances in this field over the last 10 years have been the subject of several excellent reviews.[4−18]

It is fair to say that by now we understand fairly well how to make a helix and have done so for more than a decade. It is at the next level of design that the most recent advances have been focused, the joining together of secondary structures in three-dimensional space in unique conformations that show the hallmarks of native proteins. It is a highly complex problem to encode into the amino acid sequence of a polypeptide chain information that controls not only the formation of secondary structures but also the subsequent formation of supersecondary structures, the latter controlled mainly through long-range interactions. Nevertheless, it is the purpose of this review to attempt to demonstrate that we now feel rather confident in claiming that it is

* To whom correspondence should be addressed. Fax: 46-13-122587. E-mail: Lars.Baltzer@ifm.liu.se.

10.1021/cr0000473 CCC: $36.00
Published on Web 09/20/2001

Lars Baltzer is Professor of Organic Chemistry at Linköping University. He obtained his Ph.D. degree with Professor Lars Melander at Göteborg University in 1981 in the field of physical organic chemistry. Lars Baltzer was a Swedish Natural Science Research Council (SNSRC) Postdoctoral Fellow and did postdoctoral work in NMR spectroscopy with Professor E. D. Becker at NIH, Bethesda, MD, 1981–1983. He was an Assistant Professor in Physical Organic Chemistry at Göteborg in 1983–1988, a SNSRC Research Fellow in Bioorganic Chemistry in 1988–1994, and an Associate Professor of Organic Chemistry at Göteborg University in 1994–1999. His research interests include protein design, biomolecular recognition, and mechanistic studies of catalysis.

Helena Nilsson was born in Arvika, Sweden, in 1974. She received her M.Sc. degree in Chemistry from Göteborg University and is currently working toward her Ph.D. degree with Professor Baltzer at Linköping University. Her research is focused on peptide design and catalysis.

Jonas Nilsson was born in Göteborg, Sweden, in 1970. He received his M.Sc. degree in Chemistry at Göteborg University. He is a Ph.D. student working with Professor Baltzer at Linköping University in the field of peptide and polypeptide catalyst design.

possible to design new proteins with nativelike properties that approach or even surpass 100 resi-

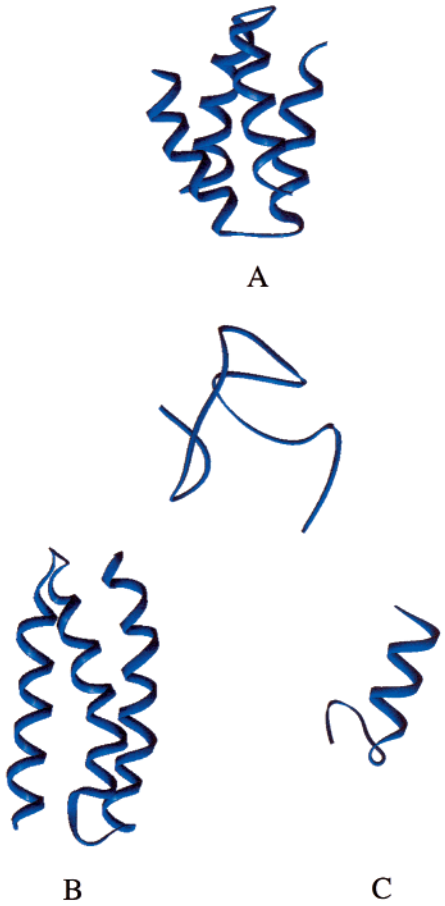

Figure 1. De novo design amounts to designing sequences from scratch that fold into predetermined three-dimensional structures. Encoded into the sequence must be not only the propensity for forming secondary structures, but also the driving force for the formation of shape- and charge-complementary interfaces that exclude alternative global folds as well as alternative but similar conformations. Several examples now exist in the literature of successful designs that show the hallmarks of native proteins and where NMR structures have been determined that allow assessments of the degrees of success of the design principles. Important examples from the RCSB PDB[94] shown here are (A) the helix−loop−helix dimer, α_2D (PDB ID, 1QP6), (B) the three-helix bundle α_3D (PDB ID, 2A3D), and (C) the $\beta\beta\alpha$ motif BBA1 (PDB ID, 1HCW).

dues in size. Helical bundles, β-sheets, and a $\beta\beta\alpha$ motif have been engineered and reengineered in systematic fashions that make it clear that at this level of design the "rules" are understood to an extent where the probability of success in designing a new protein is high.

The determination of several NMR structures,[19−23] Figure 1, represents the crowning achievement in the endeavors to learn how to construct natural folds from unnatural sequences. A high-resolution NMR structure can only be obtained if the three-dimensional structure is well defined, which is in itself proof of a successful design strategy. In addition, the analysis of high-resolution structures of designed proteins has provided the opportunity to identify at the atomic level to what extent the design principles were in fact successful in providing the necessary structural constraints. In the following it will be presented what we have learned about how to design

proteins de novo with an emphasis on the understanding of the underlying principles. The scope is that of "pure" de novo design, i.e., the design of entirely unnatural sequences rather than the reengineering of natural proteins. Neither the functionalization of de novo designed proteins nor the combinatorial approach to protein design is treated here, and theoretical algorithms for protein design have also been omitted; the reader is referred to leading references.[10,16,24−27] A heavy emphasis on helical bundles is unavoidable, as many of the fundamental principles of understanding "structural uniqueness" have been developed through their design. Selected attempts at attacking the "entropy" problem in folding by the use of unnatural scaffolds[28] that induce directionality and proximity is included because of their fundamental value. Significant advances have been made in the area of designing monomeric β-sheets, and the research that has led to the understanding of what factors inhibit aggregation and induce cooperativity in folding will also be discussed. The thrust of the review focuses on developments during the last five years.

II. Helical Bundles

A. Four-Helix Bundles

The four-helix bundle occupies a unique position in the hierarchy of de novo designed proteins because of the systematic way that its folding has been pursued and the way in which it can be used for detailed analyses. The systematic buildup of the helix−loop−helix dimer α_2D,[20,29−35] starting from single self-aggregating helices, has been most elegantly achieved by DeGrado and co-workers over a time period that spans more than a decade. An excellent review describing the developments in the DeGrado laboratory was published recently.[17] The thermodynamic and structural rules that govern helical bundle folding appear to be quite general and apply also to two-, three-, four-, as well as *n*-helix bundle design. A brief description is presented below.

1. Binding Energy versus Specificity

The unfavorable entropy contribution that results from limiting the degrees of freedom of a random coil polypeptide upon folding is enormous but more than offset, apparently, by noncovalent interactions such as hydrogen-bonding, charge−charge, and hydrophobic interactions. Helical bundles are based on the concept of amphiphilic helices that aggregate because of the hydrophobic interactions that in the folded state provide enough binding energy to drive the folding equilibrium to completion. The binding energy obtained from the burial of a hydrophobic residue has been measured several times and is now generally accepted to be 50 cal/Å^2.[36] In practice, the side chain of a Leu residue provides approximately 5 kcal/mol at room temperature to the free energy of folding.

In thermodynamic terms, for a polypeptide chain to fold, the free energy of the folded state must be several kcal/mol less than that of the unfolded state, i.e., the free energy of unfolding must be positive and in practice approximately equal to or larger than 5 kcal/mol to provide a fully folded and functional protein. The unfavorable entropy contribution is readily overcome in practice by the incorporation of a "sufficient" number of hydrophobic residues. As a rule of thumb, the binding energy needed to drive the folding process of a four-helix bundle will be amply provided for if four hydrophobic residues are included in helices of approximately 20 residues length, building a core of the side chains of, say, 16 residues. The reports of a multitude of helix−loop−helix dimers that follow approximately this general design principle have demonstrated the validity of this concept. In contrast, charge−charge interactions alone cannot provide binding energies sufficient to cause a sequence of this size to fold. Essentially all of the reported early designs showed disordered hydrophobic cores but substantial secondary structure content, i.e., they showed properties that resemble those of folding intermediates, the so-called molten globules, that differ in properties from those of native proteins.

The identification of nativelikeness typically includes the observation of ^{1}H NMR spectra that show narrow line widths and large chemical shift dispersions, narrow temperature intervals for thermal denaturation, and large values of ΔC_p. In particular, the temperature dependence of the ^{1}H NMR spectrum provides a qualitative assessment of the unfolding transition and of whether the unfolding is cooperative, in which case all residues should be affected simultaneously by temperature, Figure 2. The concept of nativelikeness is also closely linked to well-defined secondary structure contents as measured by CD spectroscopy as well as to well-defined states of aggregation. Very recently, the study of relaxation rates of amino acid side chains in a folded designed protein has added to our understanding of internal packing.[35] An excellent analysis of the nativelikeness of a β motif was presented by Searle.[23]

2. Molten Globule-like Folds

NMR structures of molten globules have been reported, but the structural details of such folds are not available at high resolution since the NOEs that are observable in many cases cannot be shown to derive from a single conformation but may be sampled from a number of very different structures that are in rapid equilibrium on the NMR time scale. Crystal structures of molten globule-like folds may not be reliable as the crystal packing forces may dominate the selection of the dominant conformer in the crystal from those populated in solution. Nevertheless, in solution, NMR structural information at the level of secondary structure elements and "dominant" folds have been obtained that demonstrate clearly that the global folds are not very far from the designed ones.[37] The designed polypeptide $\alpha\tau\alpha$ was shown to fold into a monomeric helix−loop−helix motif,[38] although the RMS deviation suggested that the structure was a mixture of rapidly equilibrating conformers. It was designed to serve as a model for an early stage of folding.

In a study of SA-42, a designed helix−loop−helix dimer,[37,39] the secondary structure was elucidated

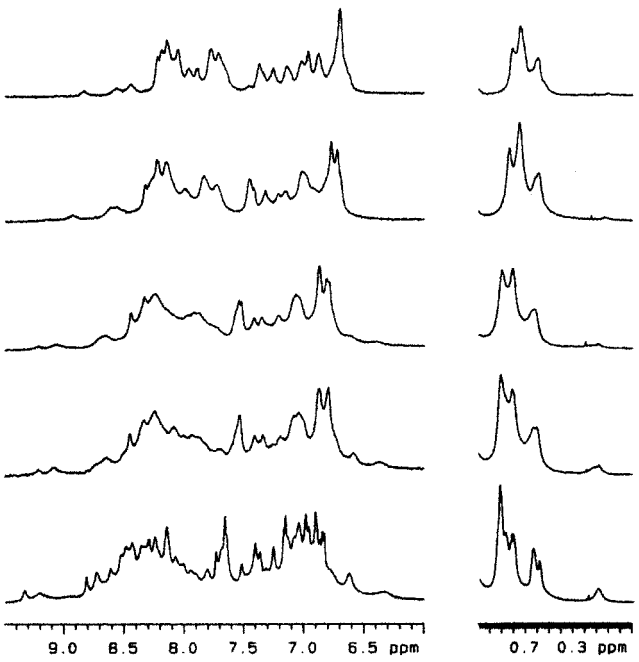

Figure 2. Temperature dependence of parts of the 500 MHz ^{1}H NMR spectrum of GTD-C in 90% H_2O−10% D_2O at pH 4.4. From bottom to top, the temperatures are 283, 298, 303, 313, and 323 K. At 283 K the resolution is dramatically better, showing that at this temperature the conformational exchange processes are considerably slower than at the higher temperatures. At low temperature GTD-C approaches the slow-exchange limit typical of a well-folded protein. The unfolding is reasonably cooperative as the line broadening and the loss of chemical shift dispersion is relatively uniform. (Reprinted with permission from ref 50. Copyright 1997 Elsevier Science.)

from the observation of αH chemical shift deviations from those of random coils and from medium-range NOEs. The observation of long-range NOE connectivities between residues almost 30 residues apart in the sequence were found to be indicative of hairpin formation, and the observation of NOE connectivities between residues at opposite ends of a helix demonstrated that at least part of the time dimers were formed in the antiparallel orientation. Corroborating evidence for the fold came from studies of catalysis where the reactivity of His residues in helix II was increased by the introduction of residues capable of transition-state stabilization in helix I.[40−42] Although the levels of population of each fold cannot be reliably estimated, there is nevertheless a solid foundation for the understanding of the supersecondary structure of the four-helix bundle.

While molten globule-like structures are not useful in providing an understanding of what factors control the uniqueness of a fold, the focus of attention in de novo design, they are likely to become very interesting model proteins for the study of what factors will be important in cooperatively forming larger protein complexes, where the function of the folding and structure of the whole complex is the issue rather than the structure of the subdomains, as is the case in many biologically important contexts. Thus the molten globule-like proteins are likely to be put to good use in the study of context-dependent folding multiprotein complexes as well as providing highly useful scaffolds in the design of functional proteins, such as catalysts.[12] They have also been proposed to be good model proteins for the study of what factors, for example, glycosylation, control maturation of folding intermediates that occur on the folding pathway.[43]

Interestingly the designed helix−loop−helix dimer αSSα exhibits a well-defined monomer subunit, structurally well characterized by NMR spectroscopy,[21,44] but no NOE contacts between the hairpin subunits were observable. Hydrogen exchange protection factors, well-defined melting points, shape complementarity in the hydrophobic core, and the general appearance of the 1D ^{1}H NMR spectrum all suggested that the structure was nativelike. The most probable explanation for the lack of intersubunit NOEs is that while the recognition of helical interfaces within the helix−loop−helix motif is highly specific, that between subunits is not, leading to a disordered interface. It demonstrates that specificity at all levels is required to form a nativelike structure.

3. Unique Folds

a. Negative Design. For a polypeptide chain to fold into a "unique" conformation, a free energy gap of several kcal/mol between that of the correctly folded protein and those of alternate folds is required. This has proven to be a more complex problem, and it is only recently that solutions have been provided to the problem of introducing by design a discriminating element of molecular recognition into the folding process. In structural terms, the introduction of sufficient binding energy at the interfaces between the secondary structural elements is straightforward, but to introduce residues that make this binding highly selective or even specific so that only one conformation is populated requires that alternative folds are destabilized "on purpose". From the dependence of binding energies on buried hydrophobic surface area it can be deduced with certainty that shape complementarity is a key factor in providing the necessary folding free energy. However, shape complementarity alone may not be a sufficient condition for the formation of well-defined tertiary structures. With the exception of highly symmetric and repetitive sequences, there is considerable evidence to suggest that the destabilization of alternative folds by what is known as "negative design" is a necessary requirement for the formation of structural uniqueness.

b. Heptad Repeat and Conformational Constraints. The structural features of negative design are conveniently discussed based on the regularity of the heptad repeat, Figure 3, applied to the design of helix−loop−helix dimers. The requirements for forming a unique fold are formidable as a helix−loop−helix motif can fold into six different global folds, five of which will have to be excluded by design. In addition, within the framework of the selected fold, the interactions between the helices in each subunit, as well as between helices in different subunits, must be highly specific. Well-defined molecular interactions must be included that severely limit the number of helical rotamers in the bundle and that exclude the possibility of register-shifted helix−helix

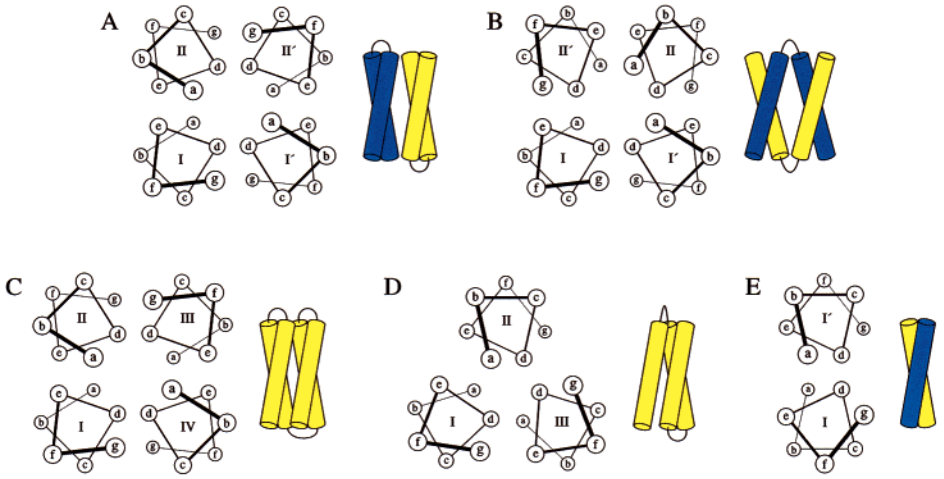

Figure 3. Schematic representation of helical bundles illustrating helix−helix interactions between amphipathic helices based on the heptad repeat. (A) The helix−loop−helix dimer in the antiparallel mode where the hydrophobic core is formed from the a and d positions and the interface between subunits is also controlled by residues in the c and g positions. A parallel mode of dimerization is also possible. The parallel and the antiparallel dimers can each be formed in two ways, where the alternative dimers arise if the "blue" and "yellow" peptides shift position and the helices are rotated 90° to form the four-helix bundle hydrophobic core. In those conformations the b and e positions are those that in addition to the a and d positions control subunit interactions. (B) The intersecting U motif which can also be formed in two different conformations, giving a total of six possible global folds for the helix−loop−helix dimer. (C) The single-chain four-helix bundle which can be formed in clockwise or anticlockwise conformations and, depending on the length of the loops, can fold in a motif which is analogous to that of the intersecting U. The three-helix bundle (D) has two possible folds, clockwise and counterclockwise, and the coiled coil (E) can form a parallel or antiparallel dimer.

interactions. Reported conformational constraints include those that have been incorporated in the hydrophobic core as well as those that function in positions that flank those in the core.

The residues along the sequence that form the hydrophobic core are those in the a and d positions. They provide most of the binding energy, although electrostatic contributions can in special cases be critical. The incorporation of exclusively aliphatic side chains (Leu, Ile, Val) has led to well-folded four-helix bundle motifs in sequences of high symmetry.[45,46] In sequences where the amino acid composition has been extensively varied, it appears that the incorporation of constraints into the hydrophobic core is a promising strategy. For example, α_2, designed by DeGrado et al.,[29] was designed to have an all-Leu core, and although the binding energy was sufficient to induce folding, it had the properties of a molten globule-like structure.[47] After redesign of the hydrophobic core of α_2 to form the sequence α_2C, the introduction of aromatic residues appeared to provide substantial improvements in conformational stability,[31] suggesting that aromatic clusters could function as elements of negative design, although the factors responsible for the specificity were not elucidated in detail. The aromatic core was formed also in α_2D, that forms a nativelike structure according to the NMR spectroscopic analysis. GTD-43, designed by Dolphin et al.,[48] also showed the characteristics of a well-defined tertiary structure and it, too, was based on the formation of an aromatic cluster; two Phe and one Trp in each monomer were shown by NMR to interact in the folded state.[49] GTD-43 also had an internal salt bridge[50] that according to several pH-dependent properties provided significant structural uniqueness at the pH where the salt bridge was the most likely to form. Two principles of negative design within the hydrophobic core have thus been demonstrated experimentally, aromatic clusters and internal salt bridges.

Positions that flank the hydrophobic core are the b and e positions as well as the g and c positions. In the folded antiparallel helix−loop−helix dimer[34] α_2D, Glu-7 is in a b position and replacing Glu-7 by an uncharged residue results in partial collapse of the ordered structure. It was proposed that Glu-7 in its ionized state provided an example of another principle of negative design in that a small positional change might force the anionically charged residue into the hydrophobic core and raise the free energy by several kcal/mol. One way of controlling the conformation would therefore be to introduce charged residues next to the a and d positions to decrease the probability of alternative similar folds. These effects are apparently very site specific, as in the sequence of α_2D there are several positions where the hydrophobic residues are not flanked on both sides by charged ones, and so is the case in GTD-43.[48] Why it is critical in some positions and not in others still eludes a detailed explanation.

The formation of hydrogen bonds between residues in b and e positions in flanking helices would be expected to provide conformational specificity in a similar way as a change in the relative positions of each helix would break the hydrogen bonds. The net result of such movement would be the loss of the free energy that corresponds to that of the hydrogen bond formation. This is nicely demonstrated by hydrogen bond formation between His residues in the structure of α_2D. Hydrogen bonds between uncharged species in aqueous solution are known to be weak, and those between charged residues typically provide less than 0.5 kcal/mol of binding energy, but apparently the preorganization of the hydrogen bond donors and

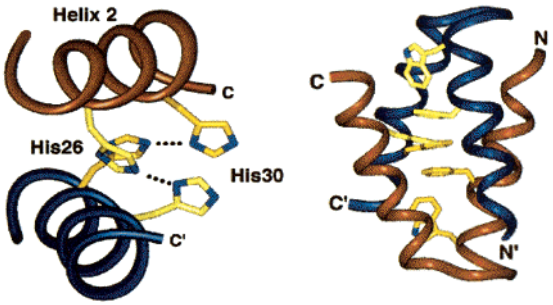

Figure 4. Details from the high-resolution NMR structure of α_2D, showing the aromatic cluster (right) and the His-His hydrogen bonds, two conformational constraints that appear to be crucial in forming the nativelike structure of the helix−loop−helix dimer. (Adapted from ref 17.)

acceptors enhances the efficiency of the interaction in α_2D.

Clearly, the introduction of conformational constraints in the folded motif serves as a successful example of negative design. The high-resolution NMR structure of α_2D,[20] Figure 4, provided the first example of a de novo designed four-helix bundle protein to be reported at that level of accuracy and demonstrates the function of virtually all of the negative design principles discussed here. It represents a landmark achievement in de novo design.

c. Charge−Charge Interactions. The role of charge−charge interactions in protein design has so far not been pursued to any great extent but is expected to be important mainly in controlling specificity in folding as well as in heterooligomeric constructs as demonstrated by the internal salt bridge of GTD-43 and by the detrimental effect of removing Glu-7 in α_2D. Even more far-reaching effects on specificity in dimerization are suggested by effects observed on the folding of helix−loop−helix dimers due to changes of charged residues at the monomer−monomer interface, where essentially complete unfolding was observed upon replacement of two lysine residues by two glutamate residues, although it is not clear yet whether the breakdown of the dimer is due to the effects of total charge rather than specific repulsive interactions.[51] The complexity of design is illustrated by the fact that a dimeric helix−loop−helix motif can fold in six different ways, Figure 3, where all but one will have to be excluded. Repulsions between residues of like charge are likely to be key determinants in selecting one fold from several alternatives. This was most elegantly demonstrated in the design of three-helix bundles discussed below.

4. Role of the Loop

The single-chain four-helix bundle DHP1 by Stroud and co-workers[46] exhibited the hallmarks of a well-packed structure, a highly cooperative unfolding transition, large amide proton exchange protection factors, and a well-dispersed ^{1}H NMR spectrum. The structure was obtained by crystallography, and the sequence was based on that of the 24-residue peptide PD1,[52] an amphiphilic helix that aggregates in solution to form helical bundles and based on only seven amino acid types. The sequence of DPH1 is highly symmetric and does not appear to contain elements of negative design, suggesting that a high degree of symmetry may be an alternative design strategy or that the influence of connecting the helices by suitable loops may provide substantial selectivity in folding. A highly symmetric sequence would be used with the expectation that the hydrophobic core would pack in a highly symmetrical and therefore well-defined arrangement. If the goal is to functionalize the protein, then the symmetry will to some extent have to be abandoned. High symmetry may make the assignment of the NMR spectra difficult due to spectral overlap.

The 33-residue coiled coil RH4 by Kim et al.[53] forms a four-helix bundle in solution that is composed of four single helices. It was designed to fold into a coiled-coil structure and based on an undeca repeat pattern. Its unfolding transition was cooperative, demonstrating the viability of design principles for more complex folds.

Dolphin et al. investigated the possible role of the loop in four-helix bundle design starting from the sequence of GTD-C.[50] The loop sequence of the 43-residue peptide GTD is GTGP, and the sequence was divided into two segments, split between GT and GP, corresponding to helix I and helix II. In the next step two helix I sequences were connected via a GTGP loop to form helixI−GTGP−helixI′, and the same was carried out with helix II to form helixII−GTGP−helixII′. If the helix packing was the dominant determinant of structure, then the equimolar mixture of these peptides would show the same behavior as GTD-C, but instead the chemical shift dispersion was that of a highly molten globule. However, linking the two peptides by chemical ligation to form a 90-residue single chain four-helix bundle, helixI−GTGP−helixI′−GCGP−helixII−GTGP−helixII′, improved the quality of the 1D ^{1}H NMR spectrum to approximately the level of that of GTD-C, suggesting that loops may influence the folding properties of helical bundles.[54]

The role of loops have also been explored by Nagi and Regan,[55] by increasing the length of the loops in the ROP protein, resulting in decreased stability toward thermal and chemical denaturation. Folding rates and lifetimes of folding intermediates were also affected showing that the role of loops in helix bundle design may be decisive. This was indeed found to be the case in the ROP protein where a single amino acid mutation in the loop changed the overall topology from the left-handed antiparallel bundle to a right-handed mixed parallel and antiparallel bundle.[56]

An interesting result was obtained relevant to the understanding of helical propensity, when a sequence with high overall helix propensity designed to fold into a four-helix bundle protein upon heating folded into a β-sheet structure and formed fibrils.[57] The perhaps surprising result suggests that the protein context is the most important determinant of structure. To explore how and why proteins switch from helix to sheet, a number of sequences were designed by Mihara that were shown to reproducibly undergo helix to sheet transitions.[58−61] Apart from the fundamental value of understanding conformational changes, these peptides may find use as molecular switches.

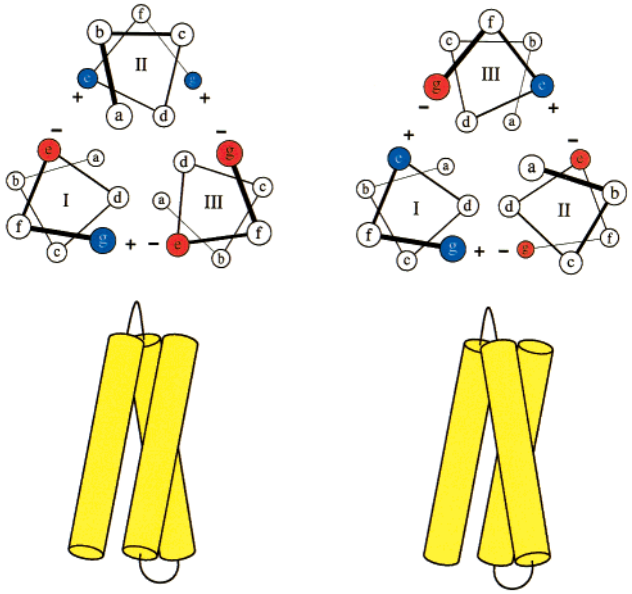

Figure 5. Schematic representation of the clockwise and counterclockwise mode of folding of the single-chain three-helix bundle motif, exploited by Johansson[63] et al. and Walsh et al.,[22] respectively. The fold is controlled by charge−charge interactions of residues in the g and e positions of the heptad repeat. Negatively charged Glu residues are represented by red circles, and positively charged Lys and Arg residues are shown in blue.

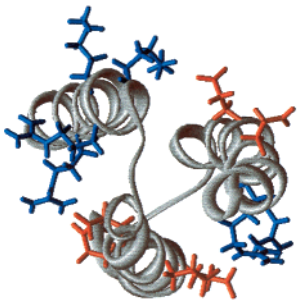

Figure 6. NMR structure of α_3D showing the spatial proximity and organization of the charged residues in the g and e positions that are responsible for the formation of the counterclockwise fold. Glu residues are shown in red, and Lys and Arg residues are shown in blue.

B. Three-Helix Bundles

The principles discussed in conjunction with the four-helix bundle motifs are applicable also in the design of three-helix bundles. Again, the a and d positions form the hydrophobic core with complimentary hydrophobic residues from the three helices. The first two helices are antiparallel, and the third helix is folded back on either side of the first two, generating a clockwise or counterclockwise configuration. The DeGrado group successfully designed a counterclockwise three-helix bundle,[62] Figure 5, and also published its NMR structure,[22] Figure 6. In early design attempts the two topologies interconverted and the structure was more molten globule-like. To improve the structure, the loops and the hydrophobic core were redesigned and the charges in the g and e positions were changed in a clever way to promote the formation of the counterclockwise fold while the opposite fold was disfavored. In the first attempts, negatively and positively charged residues were placed in an alternating fashion, making it possible for favorable interhelical interactions to develop in both topologies. In the final design this was changed to exclusively positive charges in the first helix, exclusively negative charges in the second helix, and alternating positive and negative charges in the third helix. Favorable interhelical interactions could then only arise in the counterclockwise mode. This design principle is certainly an element of negative design, where favorable interhelical interactions are only possible in one of the two topologies. The opposite approach was used by Johansson et al.[63] in the design of a clockwise three-helix bundle, where the charges were reversed and the topology changed "according to plan".

C. Two- and Four-Stranded α-Helical Coiled Coils

Coiled coils are assemblies of two to four α-helices that pack in parallel or antiparallel orientations. Their design is based on the heptad repeat, and the assembled coiled coils make a left-handed twist along the helical axes.[11] The twist gives rise to a larger hydrophobic contact area over extended helical lengths, while the helices of a bundle diverge at the ends as they lack significant supercoiling. One of the major differences between designing coiled coils, especially two-stranded coils, and helical bundles is that a coiled coil is limited to a more narrow hydrophobic face with less buried surface area from hydrophobic residues mainly in the a and d positions of the heptad repeat. Helical bundles may as well have hydrophobic residues in the e and g positions, leading to a wider hydrophobic face and higher oligomerization states. To increase the hydrophobic interactions and make the helices of a coiled coil associate more readily, they are usually somewhat longer than the helices of a bundle. So far the diversity of amino acid residues in the hydrophobic core and on the surface of designed coiled coils has been limited and based on a repeating pattern of only a few amino acid residues, leading to difficulties in assigning the NMR spectra. A greater diversity is needed to deduce the solution structure of designed coiled coils as well as to mimic the amino acid diversity found in natural proteins.

D. Unnatural Conformational Constraints and the Entropy Problem

Chemists are enthusiastically pursuing the mimicking of nature's concepts in man-made systems. The opportunities of noncovalent chemistry are apparent, and the introduction of protein function into minimal scaffolds is an important goal in bioorganic chemistry. Replacing the "useless" part of a protein with a smaller and readily available construct is attractive for reasons of design economy and for the reason that if the same function as that of a protein can be achieved in an unnatural context it implies that the function is understood. These ideas have driven research in particular in peptide mimicking model compounds. In designing a linear protein that folds,

it is an established rule of thumb that approximately 20 residues are needed to provide enough binding energy to make the peptide fold,[19,64] although in the case of helical bundles the number of residues is considerably larger.[14] The main barrier to folding for a smaller polypeptide is the entropic cost of limiting the degrees of freedom of the rotable bonds of backbone and side chains. If, rather than using a linear polypeptide chain, a scaffold is used to which peptides are covalently attached, then a substantial degree of order has been introduced that limits the degrees of freedom of the unfolded state. The negative entropy of folding is therefore reduced considerably, and the number of amino acids needed to ensure folding is therefore reduced. The concept of template-assembled synthetic proteins (TASP) was introduced by Mutter in 1985, and the progress to date has been reviewed recently.[65,66]

The introduction of a covalently linked scaffold amounts to introducing a "hard" conformational constraint. It should, in principle, introduce similar structural uniqueness as the conformational constraints mentioned above. However, covalent linkages are less tolerant toward mismatches than noncovalent ones because bond lengths and bond angles will not deviate much from the average and the noncovalent bonds must therefore adapt to the constraints imposed by the covalent linkage. At the current level of understanding this is a considerable challenge to overcome in design, and strategies have been developed where flexible linkers have been introduced between the scaffold and the peptides. The TASP concept has not yet produced well-folded proteins, as demonstrated by the small number of reported NMR investigations of TASP molecules, the exception being collagen mimics,[67] but its usefulness in designing multifunctional polypeptides is well established.[68] The recent report of a four-helix bundle TASP shows promise toward structural study. The TASP concept has considerable potential in generating multifunctional constructs due to orthogonal synthesis strategies that are by now well developed. The difficulties in obtaining well-folded TASP molecules highlight the problem of introducing several constraints into the same construct. While peptidic folds are to some extent able to adapt by populating alternative rotamers and suboptimal hydrogen bond distances and angles, hard constraints represent a more severe optimization problem. The use of disulfide bridges in design has not met with success in terms of generating well-defined tertiary structures, probably for this reason. The most impressive results to date in this regard are therefore probably those involving metal-binding proteins.[69−71]

III. β-Sheet Design

While designed α-helical proteins have been extensively studied as described above, the design of β-sheet structures remained a challenge for many years. Considerable difficulties were experienced due to aggregation, and the main focus of attention centered on turn sequence design, folding cooperativity, and overall stability. Sequences with a high propensity for β-sheet formation are inherently prone to aggregate, and the successful designs of monomeric triple-stranded β-sheets were only reported within the past few years.[23,72−75] The tendency to aggregate is manifested most clearly in nature in the formation of fibrils that follow upon extensive β-sheet formation of proteins and is believed to be the cause of several diseases.[60,76] The problem of aggregation is, however, quite readily resolved by the introduction of a number of residues of the same charge into the amino acid sequence so that the formation of higher order aggregates generate highly charged and therefore destabilized species.

The principles that determine the efficiency of turn sequences in initiating and stabilizing β-sheet formation were studied by several groups.[77−79] While the formation of α-helical bundles do not depend strongly on the structure of the connecting loops, in β-hairpin formation the function of the turn is crucial. The two-residue sequence Asn-Gly was shown to initiate folding of simple β-hairpins in water, although the folded state was only partly populated.[80−83] Quantitative measurements of the degree of structure formation must be treated with some caution, so a more detailed analysis must await more well-defined folds; nevertheless, valuable information was obtained that was relevant to the understanding of folding and stability. The experimental evidence came from the observation of interstrand NOEs by NMR spectroscopy, observation of β-sheet signatures in CD spectra, and analytical ultracentrifugation. It remains unclear why the Asn-Gly sequence favors hairpin formation since no direct interactions between the Asn side chain and the peptide backbone were observed, but it may be due to indirect interactions that involve water molecules from the solvent.[82] A turn based on the unnatural enantiomer of Pro, D-Pro, was used by Gellman et al. in driving β-sheet formation.[78,84−86] The two-residue sequence D-Pro-X, where X is in principle any amino acid, although Gly was commonly used, was able to provide partially populated structures in water. The motif was highly sensitive to the configuration of the Pro residue, and structure formation depends critically on which enantiomer of Pro was used; L-Pro did not give rise to any observable element of ordered structure.

The observation that hairpins in aqueous solution were of only marginal stability prompted the question whether a third strand might stabilize the motif and therefore provide cooperativity in folding. Peptide sequences that contained the two-residue turn NG in two positions were shown to fold and form monomeric triple-stranded β-sheets that were more stable than the corresponding hairpin.[87,88] In addition to providing a folded model system of higher stability, the principle of cooperativity was also demonstrated experimentally. A designed peptide Betanova[74] was based on similar principles, and it was demonstrated by NMR spectroscopy that the Asn-Gly segments form turns in water. The unfolding transition showed signs of cooperativity, although it took place over a wide temperature interval. Gellman compared the Asn-Gly turn with an unnatural D-Pro-Gly sequence

in β-hairpins[78,85] and demonstrated that the stability improved significantly in the latter case, whereas the stabilization of L-Pro-Gly was only marginal. There may therefore be room for further development of turn sequences based on the natural amino acids that can stabilize such structures to a larger degree.

The triple-stranded monomeric β-sheet Betanova showed cooperative unfolding in aqueous solution, and the reported NMR structure demonstrated the medium- and long-range interactions diagnostic of β-sheet formation. In the design process the incorporation of a Trp residue was critical in providing stability, and although no compact hydrophobic core was observable, transient contacts between other residues and the Trp appears to have contributed to the conformational stability, although it may be that Trp serves to destabilize alternative folds rather than stabilizing the prevailing one. An analysis of the relative importance of hydrogen-bonded, side-chain charge−charge and hydrophobic interactions will greatly improve our understanding of β-sheet design. Searle investigated in detail the role of the turn in β-hairpin formation of the two antiparallel strands from the met repressor dimer. The Asn-Gly two-residue sequence forms a turn I′ turn that induces some structure formation in water based on NMR spectroscopic analysis.[79−81] It was also shown that interstrand interactions between a Lys residue and the C-terminal were important for folding.[82,89] A 24-residue triple-stranded β-sheet was designed that included the Asn-Gly turn sequence which was included twice to enforce the formation of the monomeric triple-stranded motif.[23] An aromatic cluster was found to be important for structure formation as demonstrated by the determined NMR structure, Figure 7. While the observed NOEs were sufficient in number and intensity to ensure that a three-dimensional structure could be calculated, the stability of the motif is not as developed as those of, e.g., the four-helix bundles. This may be an intrinsic property of the shorter β strands due to the relatively small number of conformational constraints rather than a sign of imperfect design.

So far it has been demonstrated that important determinants of β-sheet formation are well-defined turns, residues with the same charge that inhibit aggregation and increase solubility, amino acid residues with high β propensity, and inter- and intrastrand side-chain interactions between hydrophobic residues. These latter determinants provide interesting possibilities in the design of mixed motifs.

IV. Miniprotein $\beta\beta\alpha$ Motif

The first mixed motif to be designed was a short 23-residue peptide, BBA1, that forms a short β-sheet−β-turn segment connected to a third strand that forms an α helix.[19] Apart from the fact that it is a mixed motif, the design is remarkable in that it forms a unique compact fold from only 23 residues.[90] The initial $\beta\beta\alpha$ miniprotein was designed based on the structure of the naturally occurring zinc-finger motif. A short type II′ turn was introduced by using a D-Pro-Ser sequence, and a second unnatural amino acid,

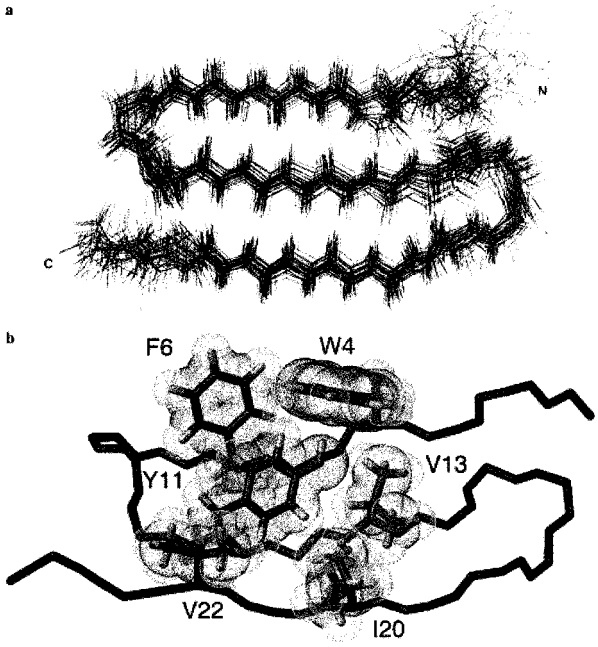

Figure 7. NMR structure of the designed monomeric triple-stranded β-sheet of Griffiths-Jones and Searle. The formation of an aromatic cluster is clearly demonstrated (bottom), and the suggestion of dynamics is illustrated by the relatively large differences between the minimized structures (top). (Reprinted with permission from ref 23. Copyright 2000 American Chemical Society.)

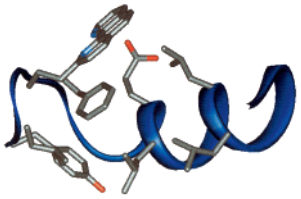

Figure 8. NMR structure of the miniprotein BBA1, hightlighting the residues of the hydrophopic core. A key feature of the first design was the artificial amino acid Fen, which was later replaced by Tyr or Trp residues, without loss of its nativelike properties.

3-(1,10-phenantrol-2-yl)-1-alanine (Fen), was used for its potential in binding metals and as a reporter group. BBA1 was found to fold into a well-defined structure, without the constraining capacity of zinc complexation. The NMR structure, Figure 8, was solved and showed that a hydrophobic cluster was formed from the Fen residue and additional aromatic residues from both the helix and the sheet. The aromatic cluster was probably instrumental in restricting the conformational freedom enough to form the nativelike structure. In subsequent designs the Fen residue was replaced by a Tyr (or Trp) residue resulting in the formation of molten globule-like structures.[91] This was compensated for by the introduction of an interstrand Arg-Asp pair that regenerated the nativelike properties of the structure (BBA5). These charged residues were not found to form the intended salt bridge but rather destabilized alternate folds where the Arg and Asp residues would be forced to move into the hydrophobic core. This is again an example where the concept of negative design has been successfully used.

Very recently the BBA miniprotein was used to induce peptide oligomerization by truncating the loop between the sheet and the helix to deliberately expose the hydrophobic surface.[92] Analytical ultracentrifugation experiments showed that the resulting peptide formed trimers, and NMR data indicated that the structure was well defined.

V. Outlook and Conclusion

Intense efforts at elucidating how and why proteins fold have improved confidence in protein design to a level where the design of helical bundle proteins, monomeric β-sheets, and mixed motifs at least for sequences that approach 100 residues or so in size represent realistic research targets. The reported successful designs are not shots in the dark but carefully analyzed and characterized molecules that have been developed based on concepts of molecular recognition, conformational preferences, and analysis of structures of native proteins. More has probably been learned from the failures than from the successes, in the iterative process, but today there appears to be a consensus about what makes a polypeptide chain fold and what is needed to preferentially populate a small group of very similar folds. The high-resolution NMR structures of $\beta\beta\alpha$, α_2D, and β-sheets clearly show that the design principles work according to plan. The next challenges then clearly lie in the design of polypeptides that fold into structures that are preorganized to form specific protein−protein interactions that control further assembly into even larger superstructures. The beginning of such a development was demonstrated recently.[93] The rules that control four-helix bundle and three-helix bundle formation are very likely to control also *n*-helix bundle formation and for that matter also mixed motifs. The construction of cavities is a necessary next step in the design of efficient catalysts and selective receptors, and the interplay between binding energies that cause folding into superstructures that do not collapse into compact proteins but leave cavities for the binding of ligands, substrates, and transition states (but expel reaction products) is now within conceptual reach.

Considerable progress has been made in the design of membrane binding proteins, where an understanding of what drives helix formation has emerged and where an understanding of what governs helix−helix interactions is underway. Thus, the further development of functional membrane-bound proteins remains an intense area of research. The energetics of protein−protein interactions in membranes are perhaps more subtle than those in aqueous solution, but the structural properties are very similar, linking the design of water-soluble proteins to those of membranes.

The historical perspective on protein design tells us that we have moved from studying natural folds with natural sequences to making natural folds from unnatural sequences. The foundation has now been laid for the design of unnatural folds from unnatural sequences. Thus, the fascinating perspective of biocatalysis and biomolecular recognition may not remain a perspective for long.

VI. References

(1) Chou, P. Y.; Fasman, G. D. *Biochemistry* **1974**, *13*, 211.
(2) Padmanabhan, S.; Marqusee, S.; Ridgeway, T.; Laue, T. M.; Baldwin, R. *Nature* **1990**, *344*, 268.
(3) O'Neil, K. T.; DeGrado, W. F. *Science* **1990**, *250*, 646.
(4) Bryson, J. W.; Betz, S. F.; Lu, H. S.; Suich, D. J.; Zhou, H. X.; O'Neil, K. T.; DeGrado, W. F. *Science* **1995**, *270*, 935.
(5) Kamtekar, S.; Hecht, M. H. *FASEB J.* **1995**, *9*, 1013.
(6) Beasly, J. R.; Hecht, M. H. *J. Biol. Chem.* **1997**, *272*, 2031.
(7) Blanco, F.; Ramirez-Alvarado, M.; Serrano, L. *Curr. Opin. Struct. Biol.* **1998**, *8*, 107.
(8) Gellman, S. H. *Curr. Opin. Chem. Biol.* **1998**, *2*, 717.
(9) Schafmeister, C. E.; Stroud, R. M. *Curr. Opin. Biotechnol.* **1998**, *9*, 350.
(10) Baltzer, L. *Curr. Opin. Struct. Biol.* **1998**, *8*, 466.
(11) Kohn, W. D.; Hodges, R. S. *TIBTECH* **1998**, *16*, 379.
(12) Baltzer, L. *Top. Curr. Chem.* **1999**, *202*, 39.
(13) Imperiali, B.; Ottesen, J. J. *J. Pept. Res.* **1999**, *54*, 177.
(14) DeGrado, W. F.; Summa, C. M.; Pavone, V.; Nastri, F.; Lombardi, A. *Annu. Rev. Biochem.* **1999**, *68*, 779.
(15) Lacroix, E.; Kortemme, T.; Lopez de la Paz, M.; Serrano, L. *Curr. Opin. Struct. Biol.* **1999**, *9*, 487.
(16) Micklatcher, C.; Chmielewski, J. *Curr. Opin. Chem. Biol.* **1999**, *3*, 724.
(17) Hill, R. B.; Raleigh, D. P.; Lombardi, A.; DeGrado, W. F. *Acc. Chem. Res.* **2000**, *33*, 745.
(18) Searle, M. S. *J. Chem. Soc., Perkin Trans. 2* **2001**, in press.
(19) Struthers, M. D.; Cheng, R. P.; Imperiali, B. *Science* **1996**, *271*, 342.
(20) Hill, B. R.; DeGrado, W. F. *J. Am. Chem. Soc.* **1998**, *120*, 1138.
(21) Skalicky, J. J.; Gibney, B. R.; Rabanal, F.; Bieber Urbauer, R. J.; Dutton, P. L.; Wand, A. J. *J. Am. Chem. Soc.* **1999**, *121*, 4941.
(22) Walsh, S. T. R.; Cheng, H.; Bryson, J. W.; Roder, H.; DeGrado, W. F. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 5486.
(23) Griffiths-Jones, S. R.; Searle, M. S. *J. Am. Chem. Soc* **2000**, *122*, 8350.
(24) Kamtekar, S.; Schiffer, J. M.; Xiong, H.; Babik, J. M.; Hecht, M. H. *Science* **1993**, *262*, 1680.
(25) Roy, S.; Ratnaswamy, G.; Boice, J. A.; Fairman, R.; McLendon, G.; Hecht, M. H. *J. Am. Chem. Soc.* **1997**, *119*, 5302.
(26) Dahiyat, B. I.; Mayo, S. L. *Science* **1997**, *278*, 82.
(27) Gordon, D. B.; Marshall, S. A.; Mayo, S. L. *Curr. Opin. Struct. Biol.* **1999**, *9*, 509.
(28) Mutter, M.; Vuilleumier, S. *Angew. Chem., Int. Ed. Engl.* **1989**, *28*, 535.
(29) Ho, S. P.; DeGrado, W. F. *J. Am. Chem. Soc.* **1987**, *109*, 6751.
(30) Regan, L.; DeGrado, W. F. *Science* **1988**, *241*, 976.
(31) Raleigh, D. P.; DeGrado, W. F. *J. Am. Chem. Soc.* **1992**, *114*, 10079.
(32) Raleigh, D. P.; Betz, S. F.; DeGrado, W. F. *J. Am. Chem. Soc.* **1995**, *117*, 7558.
(33) Hill, R. B.; Hong, J.; DeGrado, W. F. *J. Am. Chem. Soc.* **2000**, *122*, 746.
(34) Hill, R. B.; DeGrado, W. F. *Struct. Folding Des.* **2000**, *8*, 471.
(35) Hill, R. B.; Bracken, C.; DeGrado, W. F.; Palmer, A. G. *J. Am. Chem. Soc* **2000**, *122*, 11610.
(36) Fersht, A. *Structure and Mechanism in Protein Science*; W. H. Freeman and Company: New York, 1999.
(37) Olofsson, S.; Baltzer, L. *Folding Des.* **1996**, *1*, 347.
(38) Fezoui, Y.; Connolly, P. J.; Osterhout, J. J. *Protein Sci.* **1997**, *6*, 1869.
(39) Olofsson, S.; Johansson, G.; Baltzer, L. *J. Chem. Soc., Perkin Trans. 2* **1995**, 2047.
(40) Broo, K. S.; Nilsson, H.; Nilsson, J.; Flodberg, A.; Baltzer, L. *J. Am. Chem. Soc.* **1998**, *120*, 4063.
(41) Broo, K. S.; Nilsson, H.; Nilsson, J.; Baltzer, L. *J. Am. Chem. Soc.* **1998**, *120*, 10287.
(42) Baltzer, L.; Broo, K. S.; Nilsson, H.; Nilsson, J. *Bioorg. Med. Chem.* **1999**, *7*, 83.
(43) Andersson, L. K.; Dolphin, G. T.; Kihlberg, J.; Baltzer, L. *J. Chem. Soc., Perkin Trans. 2* **2000**, 459.
(44) Gibney, B. R.; Rabanal, F.; Skalicky, J. J.; Wand, A. J.; Dutton, P. L. *J. Am. Chem. Soc.* **1999**, *121*, 4952.
(45) Jiang, X.; Bishop, E. J.; Farid, R. S. *J. Am. Chem. Soc.* **1997**, *119*, 838.
(46) Schafmeister, C. E.; LaPorte, S. L.; Miercke, L. J. W.; Stroud, R. M. *Nat. Struct. Biol.* **1997**, *4*, 1039.
(47) Handel, T. M.; Williams, S. A.; DeGrado, W. F. *Science* **1993**, *261*, 879.
(48) Dolphin, G. T.; Brive, L.; Johansson, G.; Baltzer, L. *J. Am. Chem. Soc.* **1996**, *118*, 11297.
(49) Brive, L.; Dolphin, G. T.; Baltzer, L. *J. Am. Chem. Soc.* **1997**, *119*, 8598.
(50) Dolphin, G. T.; Baltzer, L. *Folding Des.* **1997**, *2*, 319.
(51) Vijayalekshmi, S.; Baltzer, L. Unpublished results.
(52) Schafmeister, C. E.; Miercke, L. J. W.; Stroud, R. M. *Science* **1993**, *262*, 734.

(53) Harbury, P. B.; Plecs, J. J.; Tidor, B.; Alber, T.; Kim, P. S. *Science* **1998**, *282*, 1462.
(54) Dolphin, G. T.; Baltzer, L. Unpublished results.
(55) Nagi, A. D.; Regan, L. *Folding Des.* **1997**, *2*, 67.
(56) Glykos, N. M.; Cesareni, G.; Kokkinidis, M. *Structure* **1999**, *7*, 597.
(57) Fezoui, Y.; Hartley, D. M.; Walsh, D. M.; Selkoe, D. J.; Osterhout, J. J.; Teplow, D. B. *Nat. Struct. Biol.* **2000**, *7*, 1095.
(58) Takahashi, Y.; Ueno, A.; Mihara, H. *Chem. Eur. J.* **1998**, *4*, 2475.
(59) Takahashi, H.; Yamashita, S.; Ueno, A.; Mihara, H. *Tetrahedron* **2000**, *56*, 7011.
(60) Takahashi, H.; Ueno, A.; Mihara, H. *Structure* **2000**, *8*, 915.
(61) Takahashi, H.; Ueno, A.; Mihara, H. *CHEMBIOCHEM* **2001**, 75.
(62) Bryson, J. W.; Desjarlais, J. R.; Handel, T. M.; DeGrado, W. F. *Protein Sci.* **1998**, *7*, 1404.
(63) Johansson, J. S.; Gibney, B. R.; Skalicky, J. J.; Wand, A. J.; Dutton, P. L. *J. Am. Chem. Soc.* **1998**, *120*, 3881.
(64) Shoemaker, K. R.; Kim, P. S.; York, E. J.; Stewart, J. M.; Baldwin, R. L. *Nature* **1987**, *326*, 563.
(65) Mutter, M.; Tuchscherer, G. *Cell. Mol. Life Sci.* **1997**, *53*, 851.
(66) Tuchscherer, G.; Grell, D.; Mathieu, M.; Mutter, M. *J. Pept. Res.* **1999**, *54*, 185.
(67) Goodman, M.; Bhumralkar, M.; Jefferson, E. A.; Kwak, J.; Locardi, E. *Biopolymers* **1998**, *47*, 127.
(68) Mutter, M.; Dumy, P.; Garrouste, P.; Lehmann, C.; Mathieu, M.; Peggion, C.; Peluso, S.; Razanamne, A.; Tuchscherer, G. *Angew. Chem., Int. Ed. Engl.* **1996**, *35*, 1483.
(69) Tommos, C.; Skalicky, J. J.; Pilloud, D. L.; Wand, A. J.; Dutton, P. L. *Biochemistry* **1999**, *38*, 9495.
(70) Lombardi, A.; Summa, C. M.; Geremia, S.; Randaccio, L.; Pavone, V.; DeGrado, W. F. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 6298.
(71) Summa, C. M.; Lombardi, A.; Lewis, M.; DeGrado, W. F. *Curr. Opin. Struct. Biol.* **1999**, *9*, 500.
(72) Ilyina, E.; Roongta, V.; Mayo, K. H. *Biochemistry* **1997**, *36*, 5245.
(73) Schenck, H. L.; Gellman, S. H. *J. Am. Chem. Soc.* **1998**, *120*, 4869.
(74) Kortemme, T.; Ramirez-Alvarado, M.; Serrano, L. *Science* **1998**, *281*, 253.
(75) De Alba, E.; Santoro, J.; Rico, M.; Jiménez, M. A. *Protein Sci.* **1999**, *8*, 854.
(76) Prusiner, S. B. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 13363.
(77) De Alba, E.; Jiménez, M.; Rico, M. *J. Am. Chem. Soc.* **1997**, *119*, 175.
(78) Stanger, H. H.; Gellman, S. H. *J. Am. Chem. Soc.* **1998**, *120*, 4236.
(79) Griffiths-Jones, S. R.; Maynard, A. J.; Sharman, G. J.; Searle, M. S. *Chem. Commun.* **1998**, 789.
(80) Maynard, A. J.; Searle, M. S. *Chem. Commun.* **1997**, 1297.
(81) Maynard, A. J.; Sharman, G. J.; Searle, M. S. *J. Am. Chem. Soc.* **1998**, *120*, 1996.
(82) Griffiths-Jones, S. R.; Maynard, A. J.; Searle, M. S. *J. Mol. Biol* **1999**, *292*, 1051.
(83) Colley, C. S.; Griffiths-Jones, S. R.; George, M. W.; Searle, M. S. *Chem. Commun.* **2000**, 593.
(84) Haque, T. S.; Gellman, S. H. *J. Am. Chem. Soc.* **1997**, *119*, 2303.
(85) Espinosa, J. F.; Gellman, S. H. *Angew. Chem., Int. Ed.* **2000**, *39*, 2330.
(86) Espinosa, J. F.; Munoz, V.; Gellman, S. H. *J. Mol. Biol.* **2001**, *306*, 397.
(87) Sharman, G. J.; Searle, M. S. *Chem. Commun.* **1997**, 1955.
(88) Sharman, G. J.; Searle, M. S. *J. Am. Chem. Soc.* **1998**, *120*, 5291.
(89) Searle, M. S.; Griffiths-Jones, S. R.; Skinner-Smith, H. *J. Am. Chem. Soc* **1999**, *121*, 11615.
(90) Struthers, M. D.; Cheng, R. P.; Imperiali, B. *J. Am. Chem. Soc.* **1996**, *118*, 3073.
(91) Struthers, M.; Ottesen, J. J.; Imperiali, B. *Folding Des.* **1998**, *3*, 95.
(92) Mezo, A. R.; Ottesen, J. J.; Imperiali, B. *J. Am. Chem. Soc* **2001**, *123*, 1002.
(93) Ogihara, N. L.; Ghirlanda, G.; Bryson, J. W.; Gingery, M.; DeGrado, W. F.; Eisenberg, D. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 1404.
(94) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. *Nucl. Acids Res.* **2000**, *28*, 235.

CR0000473