Acc. Chem. Res. **1997**, *30,* 153−161

ARTICLES

Construction and Design of β-Sheets

CATHERINE K. SMITH AND LYNNE REGAN*

Department of Molecular Biophysics and Biochemistry, Yale University, 266 Whitney Avenue, P.O. Box 208114, New Haven, Connecticut 06520-8114

Received November 14, 1996

Introduction

De novo protein design and architecture both focus on the construction and design of three-dimensional structures. Although these disciplines work on vastly different scales, they nevertheless share two requirements: a structural design and an understanding of the physical properties which govern the stability of that structure. Our knowledge of physics and engineering has allowed architects to devise magnificent buildings that can be stably constructed to serve the intended purpose. Protein designers, on the other hand, have a greater challenge realizing their intended structures because accurately predicting a protein's stability is not yet possible. Although it is well-documented that a protein's folded three-dimensional structure is encoded by its amino acid sequence, currently that folded structure cannot be predicted from sequence information alone. Therefore, the studies of protein stability, protein secondary structure, and *de novo* protein design are intimately interconnected. Stability studies provide insight for the design of proteins that will fold into predetermined structures and perform specified functions. Protein design, on the other hand, provides an opportunity to test our grasp of the rules that underlie protein structure and stability.

Understanding β-sheet formation is the key to a host of problems and applications involving protein folding and design. For example, the formation of a β-hairpin has a profound effect on reducing the conformational space and defining the long-range interactions for a folding protein. Although the characterization and *de novo* design of α-helical structures have dominated the field in the past, interest in β-sheet stability and design has intensified for several reasons. Recent studies have emphasized that there are many proteins in which β-sheets play functionally important roles. β-Sheets can provide the key element in protein−DNA,[1] protein−RNA,[2] and protein−protein recognition.[3] Several of these interactions are based upon direct, edge-on β-sheet contacts, which can often be mimicked by peptides, for example, the dimerization of HIV protease[4] and P pilin binding to the PapD chaperone.[5] Even the behavior of the hormone erythropoetin can be mimicked by disulfide-linked β-hairpin peptides.[6] Aggregated protein fibrils exhibiting predominantly β-structure have been implicated in amyloid diseases.[7] Recently, several groups have begun to quantify the energetics of the interactions that stabilize β-structure in simple model systems and to formulate guidelines which will allow the structure and stability of β-sheets to be manipulated in a rational fashion.

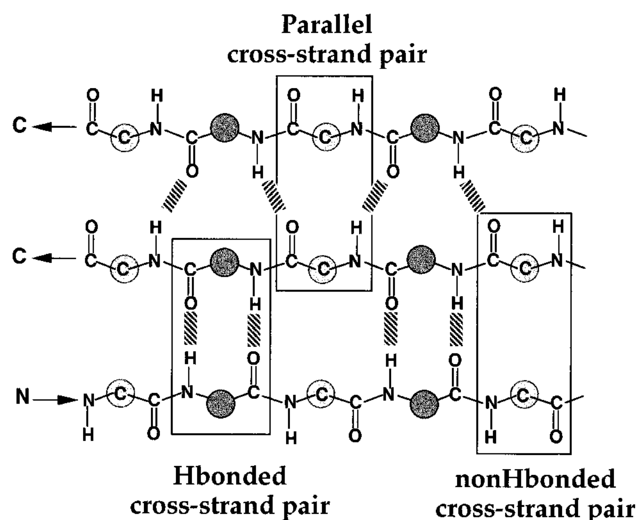

FIGURE 1. Hydrogen-bonding pattern for parallel and antiparallel β-strands and illustration of cross-strand side-chain pairs. Hydrogen bonds are represented by hatched blocks and side chains by gray filled circles. Arrows show the amide (N) to carbonyl (C) direction of the strand.

This review focuses on recent advances in the experimental study of factors influencing the stability and design of β-sheet structures. Although there has been much related work on the design of αβ structure, template-assisted protein folding,[8] and β-turn mimics,[9] the scope of this Account is limited to the study of water-soluble β-sheets that employ mainly natural L-amino acids.

Catherine K. Smith was born on May 18, 1967, in New Jersey. She received a B.A. in chemistry from Princeton University in 1989 and was subsequently employed at Merck Research Laboratories. She is currently finishing her doctoral studies at Yale University in the Department of Molecular Biophysics and Biochemistry and will soon start postdoctoral research at MIT.

Lynne Regan was born on January 16, 1959, in Leeds, England. She received a B.A. in biochemistry from Oxford University in 1981 and a Ph.D. from MIT in 1987. After postdoctoral studies at DuPont and the MRC in Cambridge, U.K., she joined the Yale faculty in 1990. Her principal research interests include protein folding and design and protein−ligand recognition.

* To whom correspondence should be addressed. Phone: (203) 432-9843. Fax: (203) 432-5175. E-mail: regan%hhvms8@venus.cis.yale.edu.

(1) Somers, W. S.; Phillips, S. E. V. *Nature* **1992**, *359*, 387−393.
(2) Puglisi, J. D.; Chen, L.; Blanchard, S.; Frankel, A. D. *Science* **1995**, *270*, 1200−1203.
(3) Derrick, J. P.; Wigley, D. B. *Nature* **1992**, *359*, 752−754.
(4) Babé, L. M.; Rosé, J.; Craik, C. S. *Protein Sci.* **1992**, *1*, 1244−1253.
(5) Kuehn, M. J.; Ogg, D. J.; Kihlberg, J.; Slonim, L. N.; Flemmer, K.; Bergfors, T.; Hultgren, S. J. *Science* **1993**, *262*, 1234−1241.
(6) Livnah, O.; Stura, E. A.; Johnson, D. L.; Middleton, S. A.; Mulcahy, L. S.; Wrighton, N. C.; Dower, W. J.; Jolliffe, L. K.; Wilson, I. A. *Science* **1996**, *273*, 464−471.
(7) Tan, S. Y.; Pepys, M. B. *Histopathology* **1994**, *25*, 403−414.
(8) Tuchsherer, G.; Mutter, M. *J. Biotechnol.* **1995**, *41*, 197−210.
(9) Ball, J. B.; Alewood, P. F. *J. Mol. Recognit.* **1990**, *3*, 55−64.

S0001-4842(96)00104-5 CCC: $14.00 © 1997 American Chemical Society

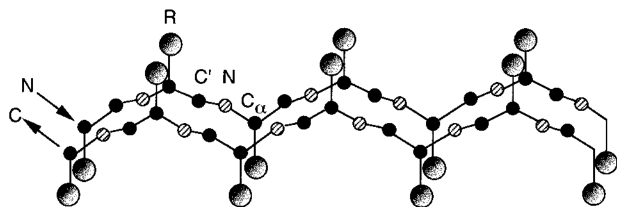

FIGURE 2. Illustration of antiparallel β-strands. Side chains are represented by gray circles, the amide moieties by hatched circles, and both the carbonyl carbons and α-carbons by filled black circles. Arrows indicate the amide (N) to carbonyl (C) direction of the strand.

β-Sheet Anatomy

β-Sheet structure is much more complex than a simple ribbon diagram would imply: the different hydrogen-bonding patterns of antiparallel and parallel sheets produce definite structural differences within the sheet itself. More globally, β-sheets may twist, curl, and even fold back on themselves to varying degrees. Because the details of β-sheet structure have been previously reviewed,[10−13] we present here a simple overview of the relevant structural features to provide a basis for the experiments discussed.

A distinct feature of a β-sheet is the pattern of hydrogen bonds formed between the amide and carbonyl groups of the protein backbone (Figure 1). In parallel β-sheets, where the β-strands run in the same amide-to-carbonyl direction, the backbone hydrogen bonds are evenly spaced and angle across to the adjacent main chain. In contrast, the hydrogen bonds formed in antiparallel sheets are approximately perpendicular to the main chain. Because β-strands necessarily interact with one another to form these hydrogen bonds, the β-sheet can bring together amino acids which are very distant in sequence.

Unlike the side chains of α-helices which extend from the outer face of the helix, β-sheet side chains alternate above and below the plane of the sheet along each strand (Figure 2). On any one face, however, the direction of two adjacent "cross-strand" side chains is in register. The characteristics of such a cross-strand pair depend on the hydrogen-bonding pattern between them (Figure 1). A parallel β-sheet arrangement produces an asymmetric cross-strand pair of amino acids which do not share hydrogen bonds with each other. In antiparallel β-sheets, there are two types of symmetrical cross-strand pairs: a narrow hydrogen-bond pair and a wide pair[10] also referred to as H-bonded and non-H-bonded pairs.[14] The amino acids of H-bonded pairs hydrogen-bond to each other directly, whereas those of non-H-bonded pairs hydrogen bond either to solvent or to another adjacent strand.

β-Strands can alter the direction of the main chain dramatically by 180° through a β-turn[15−17] or more subtly

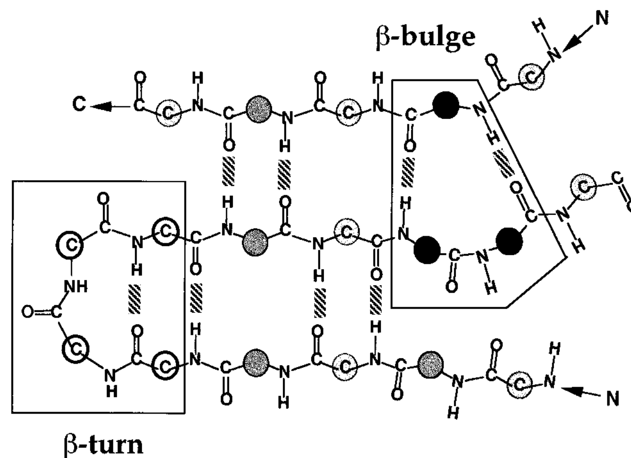

FIGURE 3. Schematic diagram of a β-turn and β-bulge region in a set of hypothetical antiparallel β-sheets. Side chains are represented by gray circles (above the plane of the sheet in dark gray and below the plane of the sheet in light gray) and hydrogen bonds by hatched blocks. Arrows indicate the amide (N) to carbonyl (C) direction of the strand.

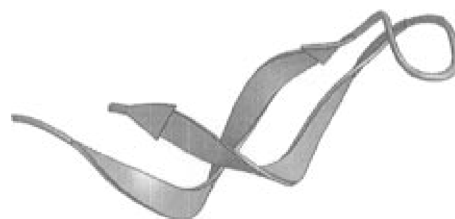

FIGURE 4. Illustration of β-sheet twist. Ribbon diagram of a β-hairpin derived from lactate dehydrogenase residues 265−293 (PDB accession code 9ldh).

through a β-bulge[18,19] (Figure 3). β-sheets also usually exhibit a right-handed twist (Figure 4) which is favored by intrastrand nonbonded interactions and interstrand geometric constraints.[20,21] At the tertiary level, layers of β-sheets are usually oriented relative to one another either at a small angle (−30°) in aligned β-sheet packing or close to 90° in orthogonal β-sheet packing[22,23] (Figure 5).

β-Sheet Construction

For a detailed understanding of protein stability, it is important to delineate experimentally the energetic contributions of each amino acid to both intrinsic β-sheet stability and the higher order interactions described above. The availability of nonaggregated β-sheet-model systems (Figures 6−8) has allowed researchers to begin to measure properties specific to β-sheets and to compare experimental results with statistical surveys.

Intrinsic β-Sheet-Forming Propensities. Statistical studies of proteins of known structure reveal that β-branched and aromatic amino acids (Tyr, Phe, Ile, Thr, Trp, and Val) occur most frequently in β-sheets.[24] The

(10) Salemme, F. R. *Prog. Biophys. Mol. Biol.* **1983**, *42*, 95−133.
(11) Richardson, J. S. *Adv. Protein Chem.* **1981**, *34*, 167−339.
(12) Chothia, C. *Annu. Rev. Biochem.* **1984**, *53*, 537−572.
(13) Rose, G. D.; Gierasch, L. M.; Smith, J. A. *Adv. Protein Chem.* **1985**, *37*, 1−109.
(14) Wouters, M. A.; Curmi, P. M. G. *Proteins: Struct., Funct., Genet.* **1995**, *22*, 119−131.
(15) Dyson, H. J.; Rance, M.; Houghten, R. A.; Lerner, R. A.; Wright, P. E. *J. Mol. Biol.* **1988**, *201*, 161−200.
(16) Hutchison, E. G.; Thornton, J. M. *Protein Sci.* **1994**, *3*, 2207−2216.
(17) Sibanda, B. L.; Blundell, T. L.; Thornton, J. M. *J. Mol. Biol.* **1989**, *206*, 759−777.
(18) Richardson, J. S.; Getzoff, E. D.; Richardson, D. C. *Proc. Natl. Acad. Sci. U.S.A.* **1978**, *75*, 2574−2578.
(19) Chan, A. W. E.; Hutchinson, E. G.; Harris, D.; Thornton, J. M. *Protein Sci.* **1993**, *2*, 1574−1590.
(20) Chothia, C. *J. Mol. Biol.* **1973**, *75*, 295−302.
(21) Chou, K.-C.; Némethy, G.; Scheraga, H. A. *J. Mol. Biol.* **1983a**, *168*, 389−407.
(22) Chothia, C.; Levitt, M.; Richardson, D. *Proc. Natl. Acad. Sci. U.S.A.* **1977**, *74*, 4130−4134.
(23) Chou, K.-C.; Némethy, G.; Rumsey, S.; Tuttle, R. W.; Scheraga, H. A. *J. Mol. Biol.* **1986**, *188*, 641−649.
(24) Chou, P. Y.; Fasman, G. D. *Biochemistry* **1974**, *13*, 211−222.

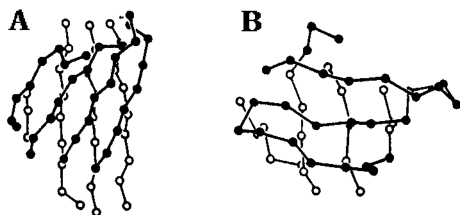

FIGURE 5. Relative orientation of packed β-sheets in the (A) aligned and (B) orthogonal packing arrangements. Circles, representing the Cα atoms, are filled for the β-sheet above the plane of the page and open for the β-sheet below the plane. Reprinted with permission from ref 12. Copyright 1984 Annual Reviews, Inc.

implication is that the β-sheet-forming propensities of the amino acids are important for β-sheet stability. However, the observed statistical distribution is also consistent with the frequent occurrence of β-sheets in the hydrophobic core of proteins. Therefore, the distribution may reflect a hydrophobic requirement rather than a β-sheet-forming propensity. This, combined with the more stringent conformational requirements of α-helices, could govern the statistical distribution. The β-sheet may be a default structure into which any amino acid could substitute. Although early studies of β-sheet formation suggested that there *are* qualitative differences in the β-sheet-forming propensities of selected amino acids,[25,26] there was a clear need for detailed thermodynamic measurements in water-soluble, unaggregated model systems to extend these initial results.

Experimental measurements of intrinsic conformational propensities of the amino acids are based on a "host−guest" method first described for measuring α-helical propensities.[27,28] In host−guest studies, the stability of a standard protein or peptide is compared with those of mutants in which the other 19 amino acids are individually substituted into the guest site. The results of the experimentally measured α-helical propensities show correlations with each other and with statistical preferences.[28]

The first comprehensive experimental measurement of the β-sheet-forming propensities employed a guest site located in a solvent-exposed, non-H-bonded position on the antiparallel β-strand of a consensus zinc-finger peptide[29,30] (Figure 6). Two subsequent host−guest studies used different variants of the B1 domain of streptococcal protein G (hereafter called the B1 domain) as the model system[31−33] (Figure 7). The guest sites in both B1 domain studies were positioned along a central strand in a solvent-exposed, H-bonded, antiparallel β-sheet.

All three studies showed that there are measurable differences between the β-sheet-forming propensities of the amino acids (Table 1). β-Branched and aromatic amino acids tend to be the best β-sheet-forming residues,

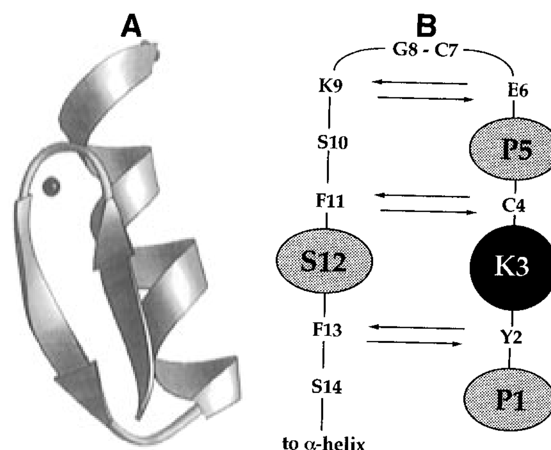

FIGURE 6. (A) Ribbon diagram[78] of the consensus zinc-finger peptide (PDB accession code 1mey). Bound zinc is illustrated by the black sphere. (B) Hydrogen-bonding diagram of the β-sheet region. Hydrogen bonds are indicated by arrows. The guest site (position 3) is highlighted in black. The nearest neighbors to the guest residue on the same face of the β-sheet are highlighted in gray.

while Gly and Pro tend to be the poorest. As might be expected, there was not an exact, one-to-one correspondence between the experimental studies, in which host-specific interactions can modulate the specific ranking, and the statistical surveys, which average over different environments and the various types of β-strands. Nevertheless, the results of all three experimental studies show a strong overall correlation with statistical and theoretical analyses[24,34−36] and indicate that the amino acids have different intrinsic propensities to adopt the β-sheet conformation.

The energetic range between the best and poorest β-sheet-forming amino acids measured in the zinc-finger study (0.5 kcal·mol^{-1}) is less than that of the B1 domain system (∼2.8 kcal·mol^{-1}). This difference may result from the different locations of the two guest sites, in a non-H-bonded edge strand and a H-bonded central strand, respectively. Both the edge location, in which there are fewer interactions with the β-sheet than in an internal strand, and the lack of β-sheet hydrogen bonds in a non-H-bonded site can increase the conformational freedom of the guest site. These factors could attenuate the magnitude of the differences in the β-sheet-forming propensities at this edge-strand position.

Edge-strand propensities have also been measured in the B1 domain on a solvent-exposed H-bonded position on an antiparallel β-sheet[37] (position 44, Figure 7B). The β-branched and aromatic residues were still among the better β-sheet-forming amino acids, and Gly and Pro were among the poorest. Greater deviations from the Chou−Fasman statistical preferences were observed in the edge-strand B1 domain study than in the edge-strand zinc-finger study. Perhaps related to the H-bonded position of the guest site, the range in the B1 domain edge-strand study was not significantly attenuated relative to the central-strand study. These results demonstrate that

(25) Kemp, D. S. *Tibtech* **1990**, *8*, 249−255.
(26) Mutter, M.; Altmann, K.-H. *Int. J. Pept. Protein Res.* **1985**, *26*, 373−380.
(27) Scheraga, H. A. *Pure Appl. Chem.* **1978**, *50*, 315−324.
(28) Chakrabartty, A.; Baldwin, R. L. *Adv. Protein Chem.* **1995**, *46*, 141−176.
(29) Kim, C. A.; Berg, J. M. *Nature* **1993**, *362*, 267−270.
(30) Kim, C. A.; Berg, J. M. *Nat. Struct. Biol.* **1996**, *3*, 940−945.
(31) Smith, C. K.; Withka, J. M.; Regan, L. *Biochemistry* **1994**, *33*, 5510−5517.
(32) Minor, D. L., Jr.; Kim, P. S. *Nature* **1994**, *367*, 660−663.
(33) Regan, L. *Curr. Biol.* **1994**, *4*, 656−658.
(34) Muñoz, V.; Serrano, L. *Proteins: Struct., Funct., Genet.* **1994**, *20*, 301−311.
(35) Finkelstein, A. V. *Protein Eng.* **1995**, *8*, 207−209.
(36) Avbelj, F.; Moult, J. *Biochemistry* **1995**, *34*, 755−764.
(37) Minor, D. L., Jr.; Kim, P. S. *Nature* **1994**, *371*, 264−267.

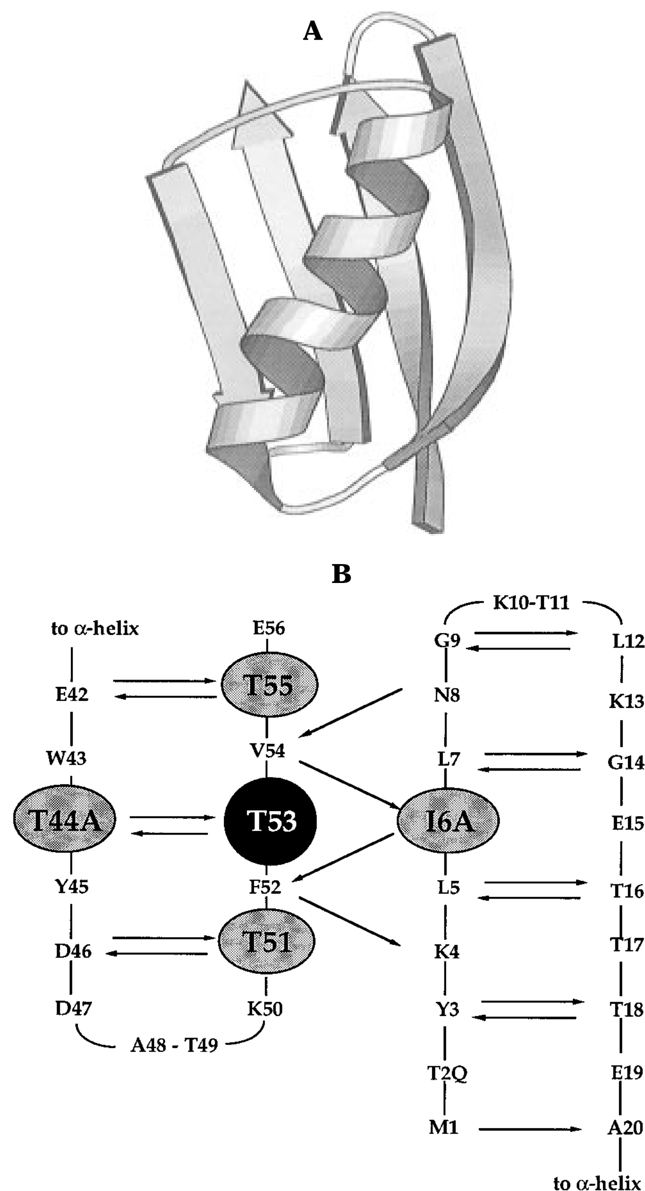

FIGURE 7. (A) Ribbon diagram of the B1 domain of streptococcal protein G (PDB accession code 2gb1). (B) Hydrogen-bonding diagram of the *β*-sheet region. Hydrogen bonds are indicated by arrows. The guest site (position 53) is highlighted in black. The nearest neighbors to the guest residue on the same face of the *β*-sheet are highlighted in gray.

positioning the guest site in an edge versus central strand can modulate the precise rankings of the measured β-sheet propensities. Statistical and theoretical studies also reveal differences between the ranking of the residues preferred in a β-sheet when edge and central strands are considered separately.[14,38,39]

Long-Range Interactions. β-Strands necessarily interact to form a β-sheet, and it seemed likely that side-chain interactions could make a substantial contribution to β-sheet stability. Statistical surveys reveal a nonrandom *pairwise* distribution of amino acids in cross-strand positions in antiparallel β-sheets.[14,40,41] The specific pairings of amino acids depend on their positions in H-bonded and non-H-bonded sites.[14] Furthermore, theoretical analyses suggest that specific interactions between side chains play an important role in determining β-sheet stability,[42] and such pairwise interactions have been included in protein structure prediction.[43] However, β-sheets are often amphipathic, with one face solvent-exposed and the other contributing to the hydrophobic core. It is therefore likely that two cross-strand amino acids would be of similar hydrophobicity. The statistical distribution is consistent with this possibility and as a result may not reflect any particular interaction between cross-strand side chains.

Table 1. Intrinsic β-Sheet-Forming Propensities[a]

Amino acid	Statistics[b]	Zinc Finger[c]	B1[d]	B1[e]
Val	1.87	-0.53	-0.94	-0.82
Ile	1.67	-0.56	-1.25	-1.0
Tyr	1.45	-0.50	-1.63	-0.96
Cys	1.40	-0.47	-0.78	-0.52
Trp	1.35	-0.48	-1.04	-0.54
Phe	1.33	-0.55	-1.08	-0.82
Leu	1.22	-0.48	-0.45	-0.51
Thr	1.17	-0.48	-1.36	-1.1
Met	1.14	-0.46	-0.90	-0.72
Gln	0.98	-0.40	-0.38	-0.23
Ser	0.96	-0.39	-0.87	-0.70
Arg	0.84	-0.44	-0.40	-0.45
His	0.80	-0.46	-0.37	0.02
Ala	0.72	-0.35	0	0
Lys	0.69	-0.41	-0.35	-0.27
Gly	0.58	0	1.21	1.2
Glu	0.52	-0.41	-0.23	-0.01
Asn	0.48	-0.38	-0.52	0.08
Asp	0.39	-0.41	0.85	0.94
Pro	0.31	-0.23	ND	>3

[a] The six best β-sheet-forming amino acids are highlighted in dark gray. The four poorest are highlighted in light gray. [b] Normalized frequencies for each amino acid in a β-sheet conformation in proteins of known structure calculated from the fraction of each amino acid that occurred in that conformation divided by this fraction for all residues.[79] [c] $\Delta\Delta G$ (kcal·mol^{-1}) relative to the variant containing Gly at the guest site calculated at room temperature.[29] [d] $\Delta\Delta G_{333\ K}$ (kcal·mol^{-1}) relative to the variant containing Ala at the guest site. The $\Delta\Delta G$ for Pro was not determined (ND) because it was unfolded at 10 °C.[31] [e] $\Delta\Delta G_{321\ K}$ (kcal·mol^{-1}) relative to the variant containing Ala at the guest site.[32]

With these considerations in mind, we set out to measure the energetic contribution of side-chain interactions to β-sheet stability using the B1 domain.[44] Of the different types of cross-strand pairs possible, we chose to measure interactions between residues which occupied a H-bonded site in an antiparallel β-sheet environment. The most useful information for the design and engineering of stable and soluble proteins can be gained from studying the interaction between pairs of amino acids with high intrinsic β-sheet-forming propensities and pairs of complementary charge. Such pairs were substituted into a solvent-exposed "double" guest site (positions 44 and 53, Figure 7B), and the thermal stabilities of these proteins were measured. The side-chain interaction energy was calculated as the difference in free energy beyond that which the simple additive sum of the propensities would predict (Table 2). For example, side chains interact favorably if the stability of the double mutant is greater than that predicted by the sum of the intrinsic propensities.

(38) Garratt, R. C.; Thornton, J. M.; Taylor, W. R. *FEBS* **1991**, *280*, 141–146.
(39) Finkelstein, A. V.; Reva, B. A. *Nature* **1991**, *351*, 497–499.
(40) von Heijne, G.; Blomberg, C. *J. Mol. Biol.* **1977**, *117*, 821–824.
(41) Lifson, S.; Sander, C. *J. Mol. Biol.* **1980**, *139*, 627–639.
(42) Yang, A.; Honig, B. *J. Mol. Biol.* **1995**, *252*, 366–376.
(43) Hubbard, T. J. *Proceedings of the biotechnology computing track, protein structure prediction minitrack of the 27th HICSS*; IEEE Computer Society Press: Los Alamitos, CA, 1994; pp 336–354.
(44) Smith, C. K.; Regan, L. *Science* **1995**, *270*, 980–982.

Table 2. Ranking of Side-Chain Interaction Energy, Pairwise Statistical Correlations, and Double Mutant Stability in the B1 Domain[44]

cross-strand pair	$\Delta(\Delta\Delta G)_{346\ K}$ [a] (kcal·mol^{-1})	cross-strand pair	$\Delta\Delta G_{346\ K}$ [b] (kcal·mol^{-1})
Glu-Arg	−0.96 (3.4)	Phe-Tyr	−2.37
Glu-Lys	−0.95 (3.4)	Phe-Phe	−2.26
Phe-Phe	−0.91 (2.4)	Ile-Tyr	−2.16
Ile-Tyr	−0.61 (1.4)	Phe-Ile	−1.93
Phe-Tyr	−0.61	Thr-Thr	−1.91
Ile-Tyr	−0.59	Thr-Tyr	−1.91
Phe-Ile	−0.44	Thr-Trp	−1.80
Ile-Ile	−0.36	Ile-Phe	−1.75
Ile-Trp	−0.34	Thr-Phe	−1.71
Phe-Trp	−0.31	Ile-Ile	−1.67
Ile-Val	−0.27	Phe-Trp	−1.63
Phe-Val	−0.20 (1.4)	Phe-Thr	−1.59
Phe-Thr	−0.19	Ile-Trp	−1.47
Ile-Thr	−0.19	Thr-Val	−1.47
Thr-Thr	0.21 (1.6)	Thr-Ile	−1.47
Thr-Trp	0.24 (0.3)	Phe-Val	−1.38
Thr-Phe	0.36	Ile-Thr	−1.37
Thr-Val	0.47 (0.7)	Ile-Val	−1.30
Thr-Tyr	0.54	Glu-Arg	−1.20
Thr-Ile	0.75	Glu-Lys	−1.09

[a] Side-chain interaction energy between residues at positions 44 and 53 in the B1 domain. Shown in brackets are the statistical pair correlations which are known at the ≥90% confidence level.[14] The correlation value is the ratio between the number of times a pair of residues is found together in the data set and the expected number of times that pair would occur randomly. [b] The B1 domain double mutant (positions 44 and 53) stability.

We found both stabilizing and destabilizing interaction energies with a range of nearly 2 kcal·mol^{-1}, which is comparable to that of the measured propensities. The pairs found to be most interactive experimentally (Phe-Phe, Phe-Tyr, Glu-Arg, and Glu-Lys) are indeed the pairs found together most often statistically. Conversely, the least interactive pairs experimentally (Thr-Val and Thr-Trp) are found together with low frequency.

An important point to note is that absolute effects on stability and side-chain interaction energy are not necessarily correlated (Table 1). The intrinsic β-sheet-forming propensities and side-chain interaction energies work in conjunction to make a stable protein. Clearly, overall protein stability is a balance between local and long-range interactions, and for protein design, it is important to optimize both β-sheet-forming propensities and cross-strand side-chain interactions. For example, the charged pairs did not form the most stable proteins despite a high interaction energy. Here the local effects of the low intrinsic propensity of Glu, Arg, and Lys dominate. In the case of the Thr-Thr pair, the high β-sheet-forming propensity of Thr results in a high overall protein stability in spite of a low side-chain interaction energy. By contrast, long-range interactions may dominate if local stabilization is low. Identical 5-amino acid sequences have been found as part of a β-sheet in one protein or as part of an α-helix in another.[45] Similarly, it has been shown that an 11-amino-acid sequence could assume either α-helical or β-sheet conformation depending on its position within the B1 domain (Figure 7A).[46] Modulation of protein stability by both the intrinsic β-sheet-forming propensities and interactions with neighboring amino acids has also been observed in a host−guest study of β-sheet stability using chymotrypsin inhibitor 2.[47]

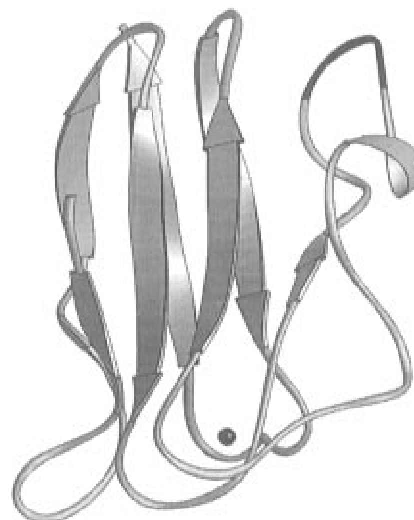

FIGURE 8. Ribbon diagram[78] of poplar plastocyanin (PDB accession code 5pcy). The central β-turn (Pro47-Ser48-Gly49-Val50) is highlighted in black. The bound copper is shown by the black sphere.

β-Turns. Statistical studies reveal a preference for certain amino acids (Asn, Gly, Pro, Asp, Ser) to occur in β-turn structure[24,48] in a positionally and structurally dependent fashion.[15−17,49] Because β-turns have been reviewed extensively elsewhere,[13] we describe here only two recent examples.

Arbitrary substitution of the highly-conserved central turn of a β-barrel protein, poplar plastocyanin[50] (Figure 8), was not tolerated. A comparison with statistical turn potentials[16] suggests that the wild-type β-turn sequence is already optimized for stability. In a related study, random substitution of the native turn sequence in B1 domain variants (positions 45−49, Figure 7B) revealed that the wild-type turn and sequences matching established turn preferences produced the most thermostable proteins.[51] These results imply that the proper turn sequence is critical for a successful design and demonstrate that the turns found most often statistically best describe those that produce stable proteins experimentally.

Protein and Peptide Design

The success of a protein or peptide design both tests and furthers our understanding of the basic principles underlying protein structure and stability. For a number of reasons, the design of β-sheet proteins has proven more difficult than that of α-helical proteins.[52] The β-sheet is composed of many structurally distinct regions with different conformational requirements—all of which must be incorporated into a single design. To form hydrogen-bonded β-sheets, β-strands necessarily interact with one another and therefore also have a strong tendency to aggregate. Design not only must include favorable interactions between several β-strands but also must disfavor

(45) Kabsch, W.; Sander, C. *Proc. Natl. Acad. Sci. U.S.A.* **1984**, *81*, 1075−1078.
(46) Minor, D. L., Jr.; Kim, P. S. *Nature* **1996**, *380*, 730−734.
(47) Otzen, D. E.; Fersht, A. R. *Biochemistry* **1995**, *34*, 5718−5724.
(48) Chou, P. Y.; Fasman, G. D. *J. Mol. Biol.* **1977**, *115*, 135−175.
(49) Wilmot, C. M.; Thornton, J. M. *J. Mol. Biol.* **1988**, *203*, 221−232.
(50) Ybe, J. A.; Hecht, M. H. *Protein Sci.* **1996**, *5*, 814−824.
(51) Zhou, H. X.; Hoess, R. H.; DeGrado, W. F. *Nat. Struct. Biol.* **1996**, *3*, 446−451.
(52) Hecht, M. H. *Proc. Natl. Acad. Sci. U.S.A.* **1994**, *91*, 8729−8730.

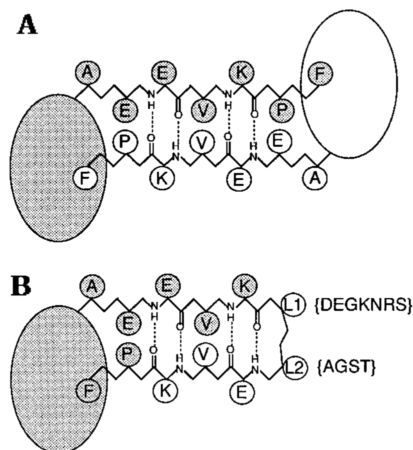

FIGURE 9. (A) Schematic diagram of the antiparallel β-sheet dimeric interface of wild-type λ-Cro. The globular portion of each subunit is represented as a large oval. Hydrogen bonds are indicated by the dashed lines. (B) Monomer design: The five-residue insertion (open circles) into the wild-type sequence (shaded circles) consists of two loop positions (L1 and L2) and a duplication of residues 54−56. Letters in brackets indicate the residues at each turn position found by selection. Reprinted with permission from ref 53. Copyright 1990 American Association for the Advancement of Science.

unplanned, competing interactions with other β-strands. Despite these diverse considerations, designs for β-sheet proteins and β-hairpin peptides have met with some success.

Designed β-Hairpin Peptides. The design of β-hairpin peptides, the most basic unit of β-sheet structure, attempts to reproduce fully folded secondary structure in the absence of any tertiary context. To allow detailed thermodynamic and structural analyses of such systems, it is critical that the peptide exhibit a significant degree of secondary structure, a discrete oligomeric state, and high solubility in aqueous solution.

One of the first successful β-hairpin designs was the conversion of a homodimeric protein, λ-Cro, to a folded and functional monomeric structure.[53] The intermolecular dimer interface, formed by two antiparallel β-strands, was replaced with an intramolecular β-hairpin in which the β-turn sequence was optimized by selection (Figure 9). The protein containing the best turn (Asp-Gly) behaved as the predicted monomer in solution, and its crystal structure confirmed that the hairpin design was successful.[54]

Despite this early progress, β-hairpin structures were thought to be unstable as isolated peptides in aqueous solution. Recently, however, some degree of native-like β-hairpin structure was observed in aqueous solution for short, monomeric, linear peptides derived from natural proteins.[55−58] These exciting results spurred much interest in the design and conformational analysis of β-hairpin systems (Figure 10) as described in the following section.

β-Hairpin Peptide Design Considerations. One of the first designed β-hairpin peptides was based on the sequence of the cyclic peptide gramicidin S[59] (Figure 10A). The ornithine residues in gramicidin S were replaced with structurally similar Lys residues to enhance solubility and reduce aggregation. A distinguishing feature of this design was the replacement of the wild-type D-Phe-Pro turn with 4-(2-aminoethyl)-6-dibenzofuranpropionic acid. This dibenzofuran moiety can nucleate a β-sheet hydrogen-bonding pattern in the attached amino acid strands by creating a hydrophobic cluster with neighboring amino acid side chains.[60]

The sequence for the eight-residue β-hairpin peptide BH8[61] was based on experimental intrinsic β-sheet-forming propensities,[29,31,32] pairwise correlations,[14] statistical analyses of eight-residue β-hairpins in proteins of known structure, and type I′ turn sequences favored statistically[16] (Figure 10B). To minimize aggregation, Arg residues were placed at the N- and C-termini, and the typical pattern of alternating hydrophilic and hydrophobic residues along the strands was avoided. As a control, variations on the BH8 peptide sequence included a poor β-sheet-forming residue, Ala, along the strands.

Design considerations for the 12-residue peptide model BB[62] included intrinsic β-sheet- and β-turn-forming propensities[63] (Figure 10C). Of the β-branched residues used, four Thr were included because of their potential for cross-strand hydrophobic and hydrogen-bonding interactions. To facilitate polar interactions, Asp, Ser, and His residues were located on the same face of the sheet. The potential for disulfide bridge formation across the β-strands was included by positioning D-cysteine residues at the N- and C-termini.

Two short linear peptides (designated peptides 2 and 3) were designed to improve upon the β-hairpin structural properties of a peptide derived from tendamistat[64] (Figure 10D). Experimental β-sheet-forming propensities were used to optimize β-strand-promoting sequences[29,31,32] in the context of a potentially stabilizing Thr-X-Thr motif observed in natural β-hairpin-forming peptides. Statistical probabilities for type I β-turn formation determined the turn sequence.[16]

β-Hairpin Peptide Characterization. The four peptide systems were shown to be monomeric and highly water-soluble. CD suggested the presence of β-sheet structure for the dibenzofuran-based peptide BH8 and the disulfide-bonded model BB (BB-O), but not for the Ala-substituted BH8 control peptide, the reduced model BB (BB-R), or peptides 2 and 3. All four peptide systems showed several sequential and cross-strand NOEs characteristic of antiparallel β-strands (Figure 10). Although these model systems clearly form some β-structure, accurately quantifying this folded population is difficult. It has been

(53) Mossing, M. C.; Sauer, R. T. *Science* **1990**, *250*, 1712−1715.
(54) Albright, R. A.; Mossing, M. C.; Matthews, B. W. *Biochemistry* **1996**, *35*, 735−742.
(55) Blanco, F. J.; Rivas, G.; Serrano, L. *Nat. Struct. Biol.* **1994**, *1*, 584−590.
(56) Neira, J. L.; Fersht, A. R. *Folding Des.* **1996**, *1*, 231−241.
(57) Viguera, A. R.; Jiménez, M. A.; Rico, M.; Serrano, L. *J. Mol. Biol.* **1996**, *255*, 507−521.
(58) Searle, M. S.; Zerella, R.; Williams, D. H.; Packman, L. C. *Protein Eng.* **1996**, *9*, 559−565.
(59) Díaz, H.; Tsang, K. Y.; Choo, D.; Espina, J.; Kelly, J. W. *J. Am. Chem. Soc.* **1993**, *115*, 3790−3791.
(60) Díaz, H.; Espina, J. R.; Kelly, J. W. *J. Am. Chem. Soc.* **1992**, *114*, 8316−8318.
(61) Ramírez-Alvarado, M.; Blanco, F. J.; Serrano, L. *Nat. Struct. Biol.* **1996**, *3*, 604−611.
(62) Sieber, V.; Moe, G. R. *Biochemistry* **1996**, *35*, 181−188.
(63) Chou, P. Y.; Fasman, G. D. *Annu. Rev. Biochem.* **1978**, *47*, 251−276.
(64) de Alba, E.; Jiménez, M. A.; Rico, M.; Nieto, J. L. *Folding Des.* **1996**, *1*, 133−144.

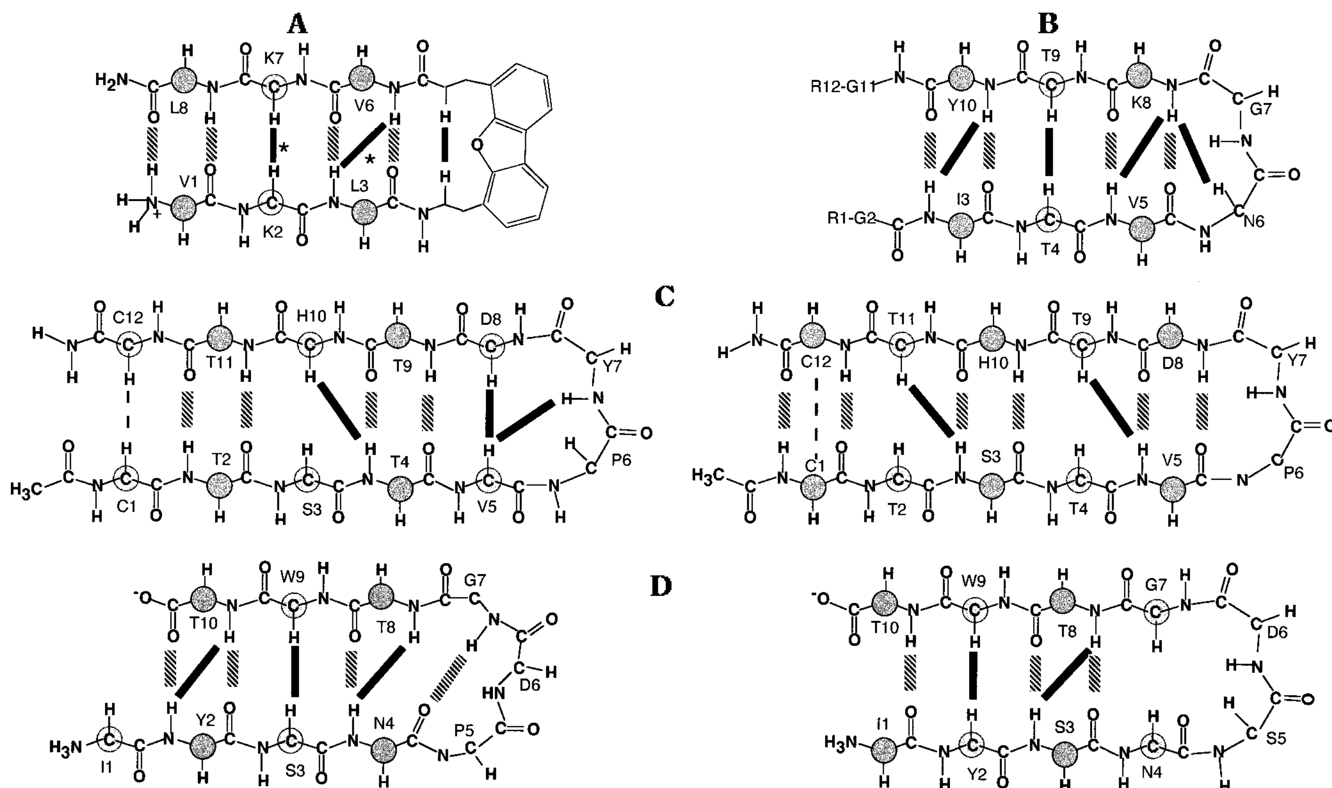

FIGURE 10. Schematic diagrams of the expected structures of the designed peptides (A) dibenzofuran-based peptide,[59] (B) BH8,[61] (C) BB-O (left) and BB-R (right),[62] and (D) peptides 2 (left) and 3 (right).[64] The observed αN, αα, and NN long-range NOEs are shown as thick black lines (sequential NOEs are not shown). An asterisk indicates that the NOE was not unambiguously determined. Hatched blocks correspond to the expected hydrogen bonds. A dashed line corresponds to a disulfide bridge.

estimated that ~30% of the peptide population is folded in these systems. It is therefore not surprising that the complete hydrogen-bonding network typical of fully folded β-sheets is not observed in the NMR analysis of these designed peptides.

What are the interactions stabilizing these structures? In the dibenzofuran-based peptide, the strong, long-range NOEs observed between the dibenzofuran moiety and the neighboring side chains suggest that a hydrophobic cluster supports β-hairpin formation.[59] The lack of structure observed in the BH8 control peptides implies that interstrand side-chain interactions and β-sheet propensities make important stabilizing contributions.[61] Hydrophobic interactions between the Thr γ-methyl groups may be the structure-determining interaction in the model BB peptides.[62] Peptides 2 and 3, which differ by a single residue in the β-turn, exhibit different conformations, indicating the importance of the turn in defining the details of the β-hairpin structure[64] (Figure 10D). These first, encouraging steps toward the design of a water-soluble, β-hairpin peptide highlight the variety of considerations required for a successful design and provide information critical for the future design of fully folded peptide systems.

Designed β-Sheet Proteins. Protein design takes peptide design one step further, attempting to capture in an artificial sequence the properties of natural proteins: a compact globular structure with a discrete oligomeric state, a well-packed hydrophobic interior with a polar exterior, and an abundance of secondary structure. It is important to note that several β-sheet protein designs were attempted prior to both the experimental measurement of β-sheet stability properties and the design of simple β-hairpins.

Betabellin Progeny: Betadoublet and Betabellin 14D. The artificial protein betabellin was intended to adopt a homodimeric, antiparallel β-barrel structure, similar to that of an immunoglobulin V_L domain.[65] Although there have been a number of redesigns, we discuss here the two latest and most successful versions, betabellin 14D[66] and betadoublet[67] (Figure 11). Design considerations in these two proteins included residues of high β-sheet-forming propensity in an alternating hydrophilic/hydrophobic pattern, a nonrepetitive sequence similar to that of native proteins, a disulfide bond connecting the subunits, and statistically and theoretically favored type I′ β-turns.

Both proteins were water-soluble and appeared to adopt compact, unaggregated β-sheet structures. Thermal denaturation of the proteins resembles the cooperative and reversible transitions of native proteins with moderately stable melting temperatures. Although the ^{1}H-NMR spectrum of both proteins showed native-like spectral dispersion, the binding of the hydrophobic dye ANS indicated that the hydrophobic core was not entirely well-packed. To date, these proteins represent the most iterative attempts at the design of a native-like β-sheet protein.

Minibody. The design for the minibody is based on a portion of the immunoglobulin V_H domain[68] (Figure 12).

(65) Richardson, J. S.; Richardson, D. C. *Tutorials in molecular and cell biology: 1. Protein Engineering*; Alan R. Liss, Inc.: New York, 1987; pp 149−163.
(66) Yan, Y.; Erickson, B. W. *Protein Sci.* **1994**, *3*, 1069−1073.
(67) Quinn, T. P.; Tweedy, N. B.; Williams, R. W.; Richardson, J. S.; Richardson, D. C. *Proc. Natl. Acad. Sci. U.S.A.* **1994**, *91*, 8747−8751.

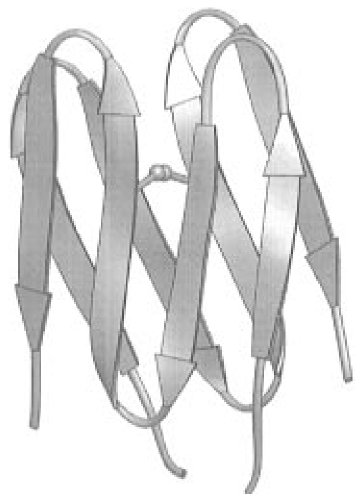

FIGURE 11. Ribbon diagram[78] of the model structure of betadoublet (PDB accession code 1btd).[67]

FIGURE 12. Ribbon diagram illustrating a model structure for the minibody. Reprinted with permission from ref 68. Copyright 1993 Macmillan Magazines Limited. Bound zinc is indicated by the gray circle, and the most probable coordinating side chains are shown explicitly.

Segments corresponding to the exposed hypervariable H1 and H2 loops of the immunoglobulin were also included in the design. The loops show a high tolerance to sequence variability in the natural protein and were therefore a logical site in which to incorporate residues with the potential for metal-binding.

As might be expected for a "first-generation" design, the minibody exhibits low solubility (10 μM). Nevertheless, characterization of the soluble fraction shows promising results. The minibody is monomeric, exhibits β-sheet secondary structure, and shows a native-like denaturation transition and moderate stability. As designed, the minibody interacts with Zn^{2+}, albeit weakly, and with an as yet uncharacterized geometry. The inclusion of a Lys-

(68) Pessi, A.; Bianchi, E.; Crameri, A.; Venturini, S.; Tramontano, A.; Sollazzo, M. *Nature* **1993**, *362*, 367−369.

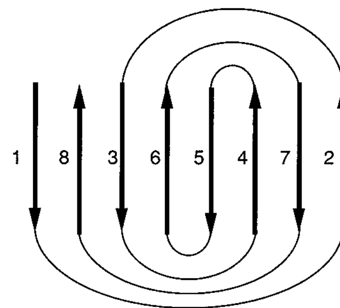

FIGURE 13. Connectivity diagram of the Greek key motif. β-strands, indicated by arrows, are numbered according to their occurrence in the primary sequence. In the three-dimensional structure, β-strands 1 and 2 would share hydrogen bonds.

rich tail and additional polar residues increased the solubility of the minibody to millimolar levels.[69]

Greek Key Motif. A computerized protein design algorithm was developed to generate nonrepetitive amino acid sequences compatible with the geometry and statistical amino acid preferences of the Greek key motif, an eight-stranded barrel-like structure[70] (Figure 13). The first protein designed using this approach tended to precipitate, and the soluble fraction existed as an aggregate.[71] Although CD showed weak evidence of β-sheet structure, the tendency of the protein to bind ANS suggested that the hydrophobic core is not well-packed.

***β*-Peptides 1, 2, 3, and 4.** The four 33-residue peptides, βpep1−4, are designed to self-associate as triple-stranded antiparallel β-sheets.[72] An analysis of the betabellin family, betadoublet, and peptides derived from the α-chemokines, Gro α, IL-8, and PF4,[73] suggested the following design criteria: (1) a net positive charge, (2) a composition of <20% noncharged polar residues (Ser, Thr, Gln, Asn), and (3) a 40−50% content of appropriately paired hydrophobic residues (Ile, Leu, Val, Ala, Met).

βpep1−4 are water-soluble and self-associating, as designed. However, βpep4 seems to exist as a tetramer, and the rest form a variety of aggregates. The NMR spectra showed dispersed αH and NH chemical shifts and limited protection from amide solvent exchange, suggesting the presence of some β-structure which was confirmed by CD.

As might be expected, many of the first-generation β-sheet designs display low solubility, illustrating the delicate balance between designing against unwanted intermolecular aggregation while promoting the necessary interstrand interactions. The encouraging features of betabellin, betadoublet, and the minibody emphasize that an iterative design process can improve solubility while progressing toward the target structure.

Summary

The results presented here illustrate that intrinsic propensities and long-range, side-chain interactions play critical roles in determining β-sheet stability. The infor-

(69) Bianchi, E.; Venturini, S.; Pessi, A.; Tramontano, A.; Sollazzo, M. *J. Mol. Biol.* **1994**, *236*, 649−659.
(70) Henneke, C. M. *Comput. Appl. Biosci.* **1993**, *9*, 709−722.
(71) Smith, D. D. S.; Pratt, K. A.; Sumner, I. G.; Henneke, C. M. *Protein Eng.* **1995**, *8*, 13−20.
(72) Mayo, K. H.; Ilyina, E.; Park, H. *Protein Sci.* **1996**, *5*, 1301−1315.
(73) Ilyina, E.; Mayo, K. H. *Biochem. J.* **1995**, *306*, 407−419.

mation derived from experimental β-sheet studies has and will continue to improve the stability and folded structure of β-sheet designs. Although the current protein designs capture β-sheet secondary structure fairly well, none achieves the tight hydrophobic packing of interior residues observed in natural proteins. Statistical analyses and theoretical calculations of β/β and α/β packing are available.[22,23,74–76] To date, however, there have been no systematic experimental studies of $\beta\beta$ and $\alpha\beta$ packing requirements akin to those dissecting α-helix packing.[77] Such experimental studies should enhance our understanding of secondary structure packing and provide an additional level of guidance for β-sheet design.

AR9601048

(74) Cohen, F. E.; Sternberg, M. J. E.; Taylor, W. R. *J. Mol. Biol.* **1982**, *156*, 821−862.
(75) Chothia, C.; Janin, J. *Biochemistry* **1982**, *21*, 3955−3965.
(76) Lasters, I.; Wodak, S. J.; Alard, P.; van Custem, E. *Proc. Natl. Acad. Sci. U.S.A.* **1988**, *85*, 3338−3342.
(77) Munson, M.; Balasubramanian, S.; Fleming, K. G.; Nagi, A. D.; O'Brien, R.; Sturtevant, J. M.; Regan, L. *Protein Sci.* **1996**, *5*, 1584−1593.
(78) Carson, M. *J. Appl. Crystallogr.* **1990**, *24*, 958−961.
(79) Williams, R. W.; Chang, A.; Juretic, D.; Loughran, S. *Biochim. Biophys. Acta* **1987**, *916*, 200−204.