Pergamon

Bioorganic & Medicinal Chemistry 7 (1999) 93–103

BIOORGANIC & MEDICINAL CHEMISTRY

β-Hairpin and β-Sheet Formation in Designed Linear Peptides

Marina Ramírez-Alvarado,[a] Tanja Kortemme,[a] Francisco J. Blanco[b] and Luis Serrano[a,*]

[a]*European Molecular Biology Laboratory (EMBL), Meyerhofstrasse 1, 69117, Heidelberg, Germany*
[b]*National Institutes of Health, Laboratory of Chemical Physics, NIDDKD, 9000 Rockville Pike, Building 5, Room 406, Bethesda, MD 20892-0520, USA*

Received 7 May 1998; accepted 25 September 1998

Abstract—Recent knowledge about the determinants of β-sheet formation and stability has notably been improved by the structural analysis of model peptides with β-hairpin structure in aqueous solution. Several experimental studies have shown that the turn region residues can not only determine the stability, but also the conformation of the β-hairpin. Specific interstrand side-chain interactions, hydrophobic and polar, have been found to be important stabilizing interactions. The knowledge acquired in the recent years from peptide systems, together with the information gathered from mutants in proteins, and the analysis of known protein structures, has led to successful design of a folded three-stranded monomeric β-sheet structure.

Introduction

Elucidating the protein folding and stability problem involves unraveling the mechanisms by which natural proteins attain their special physical properties, as compared to random heteropolymers. To explain these properties, it has been theoretically proposed that there is a sufficiently large energy gap between the native and any of the other possible conformations of a protein.[1,2] This gap requires the native set of noncovalent interactions between the amino-acid residues of a protein to be more stabilizing than those in all other conformations. There are two types of noncovalent interactions in proteins: local and nonlocal, defined in terms of the distance in the sequence between the interacting residues.[3] Importantly, the noncovalent interactions are geometrically different; local interactions participate in defining secondary structure while nonlocal interactions are involved in defining the tertiary structure. Both types of interactions need to be understood in order to have a rational solution to the protein-folding problem. In recent years, significant advances have been made in understanding the energetic contribution of local interactions to protein stability in a quantitative way (see following). This, by itself is enough not only to modify protein stability in a rational way, but also to design small proteins in a hierarchical fashion.

Site-directed mutagenesis in proteins has been one of the methods of choice to characterize and quantify the free energy contribution of noncovalent interactions. However, there are some drawbacks in the use of proteins for such a purpose. First, it is not always possible to carry out mutations that only monitor a single interaction. Second, the effect of mutations on the denatured state of the protein can be determined only with some difficulty. The latter problem becomes more severe in the study of local interactions, because they could fix some degrees of freedom in the polypeptide chain. Changes in local interactions are likely candidates to contribute to the stability of both the native state and the denatured ensemble. An alternative system to analyze local interactions is the use of synthetic model peptides. Peptides could be short and monomeric, and present less context problems than proteins. Such systems have been successfully used to dissect the contribution of local interactions to α-helix stability.[4–6]

A major incentive for using simple peptide model systems to study secondary structure formation and stability is to develop guidelines for protein engineering and design. This approach comprises certain advantages: As discussed above, peptide systems can be devised to quantify the energetic factors contributing to secondary structure stability in a context-free or context-controlled environment. In contrast, context-effects can play a significant role in secondary structure formation in proteins. In a particular dramatic case, the same 11 amino acid peptide sequence has been shown to adopt different secondary structures when implanted into the same protein at two different positions.[7] Therefore, values for the energetic contributions of certain interactions derived from peptide models might be more generally applicable to protein design.

Key words: β-hairpin; β-sheet; peptides; protein folding; NMR.
*Corresponding author. Tel.: +49-6221-387-402; fax: +49-6221-387-306.

0968-0896/99/$ - see front matter
PII: S0968-0896(98)00215-6

In this review, we will briefly describe the extent of our knowledge regarding α-helices and then concentrate on a much less understood secondary structure element, the β-sheet, and its smallest unit, the β-hairpin. We will first analyze the structural characteristics of β-sheets. Then we will provide a description of model peptide systems used to analyze the energy contributions that stabilize this secondary structure element. Finally, we will explain in detail the design and structural characterization of a monomeric three-stranded β-sheet.

α-Helices

In general, peptides corresponding to helical regions in proteins have shown marginal stability in aqueous solution.[8] The first protein fragment that was found to populate helical conformation in aqueous solution was the native 13-residue C-peptide from RNase A that showed &25% helical population at low temperature.[9] Designed de novo peptides with high alanine content have shown high helical content (up to 80%).[10,11] By using polyalanine-based designed peptides with some polar residues intercalated, it has been possible to dissect many of the interactions responsible for α-helix stability.[5,6] The model most often used for that purpose, the helix–coil transition, is based on statistical mechanics. Its simplest version[12] only considered two parameters, an elongation factor and a nucleation factor corresponding to equilibrium constants and characteristics of each amino acid, responsible for the helical tendency of a particular sequence. Later, helix–coil transition theory has been modified to include other parameters: capping interactions, side chain–side chain interactions, long-range electrostatic effects, interaction of charged groups with the helix dipole, etc.[4,6] These modifications have been introduced using experimental data in different helix–coil transition algorithms.[4,5,13–15] In addition, another algorithm using a statistical analysis of the protein database has been described.[16]

Helix–coil transition algorithms can predict the helical tendency of protein fragments and can, therefore, be used to design mutations in order to modify helical tendencies in a quantitative manner. In the case of proteins, it has been found that an increase in helix propensity in different proteins showed chemical and thermal stabilization of the entire protein.[17–19]

β-Hairpins

Structural description and classification. The basic unit of a β-sheet is a β-strand, with the polypeptide backbone almost fully extended. If the amino acids in the interacting β-strands are aligned in the same direction (amino terminal to carboxyl terminal), the β-sheet is described as parallel. If the amino acids in successive strands have alternating directions, the β-sheet is called antiparallel. Each arrangement of strands has a distinctive pattern of hydrogen bonding. Almost all β-sheets (parallel, antiparallel, and mixed) show a right-handed twist in protein structures. This right-handed twist was suggested to represent the conformation of lower free energy when compared with straight β-sheets, or left-handed twist, based on theoretical calculations and empirical observations.[20]

Proteins have been used to derive experimental scales of the propensity of β-sheet formation for the 20 amino acids,[21–23] as well as to study the effect of different residue pairs in an antiparallel β-sheet.[24] The intrinsic propensities scales correlate well among them and with a scale derived from ϕ, Ψ analysis of the protein structure database.[25] In addition, nonlocal context effects in β-sheets are important for β-sheet stability.[23]

Antiparallel β-sheets in proteins show an alternating pattern of side-chain interactions in either one or the other face of the β-sheet. A β-hairpin is the simplest form of an antiparallel β-sheet, defined by a turn, or loop, region flanked by two antiparallel strands with a defined backbone hydrogen-bonding pattern and has been proposed as the nucleation event of a β-sheet.[26] According to Sibanda & Thornton,[27] the residues corresponding to the β-hairpin strands are called..., −B3, −B2, −B1 for the N-terminal strand and +B1, +B2 and +B3... for the C-terminal strand. Turn, or loop, residues are called L1, L2, ..., L*n*. Thus, a two-residue turn β-hairpin is denoted by (...−B3, −B2, −B1, L1, L2, +B1, +B2, +B3,...). β-Hairpins can be further classified with respect to the hydrogen-bonding pattern, using a X:Y nomenclature.[27] The X:Y denotes the number of residues in the loop or turn (regions of the polypeptide chain where the Cα(i)−Cα(i+3) distance is less than 7 Å), using two different IUPAC conventions. X is defined as the number of residues in the connecting segment between the two strands, when the residues belonging to the strands are defined as those that have at least one of their NH or CO main chain groups involved in the typical hydrogen bond pattern of β-sheets. Y is defined as the number of residues in the connecting segment between the two strands, when the residues belonging to the strands are defined as those that have both their NH and CO main chain groups involved in the typical hydrogen bond pattern of β-sheets. If both hydrogen bonds are formed, then X = Y, and the classification just depends on the number of residues in the loop or turn, (i.e. a hairpin with a two-residue turn will be defined as 2:2). If only a single hydrogen bond is formed between the terminal NH of the first strand and the first CO of the second strand, then Y = X + 2, forming hairpins of type 2:4, type 3:5, and so on. In protein structures, 2:2 β-hairpins are most abundant followed by 3:5 β-hairpins and 4:4 β-hairpins[28] (Fig. 1).

β-Turns are the most abundant regular elements connecting two β-strands and can be classified by their backbone dihedral angles[29] (Table 1). In a study of all turns in proteins of known structure, it was found that 60% were of type I + III and 15% were of type II. For β-hairpins with a two-residue turn (2:2), however, type I′ and II′ turns predominated and have been analyzed and both theoretically and experimentally characterized.[30–35] The total absence of type II turns in hairpins and relative lack of type I turns in two-residue turns strongly implies that these turns, which are by far the

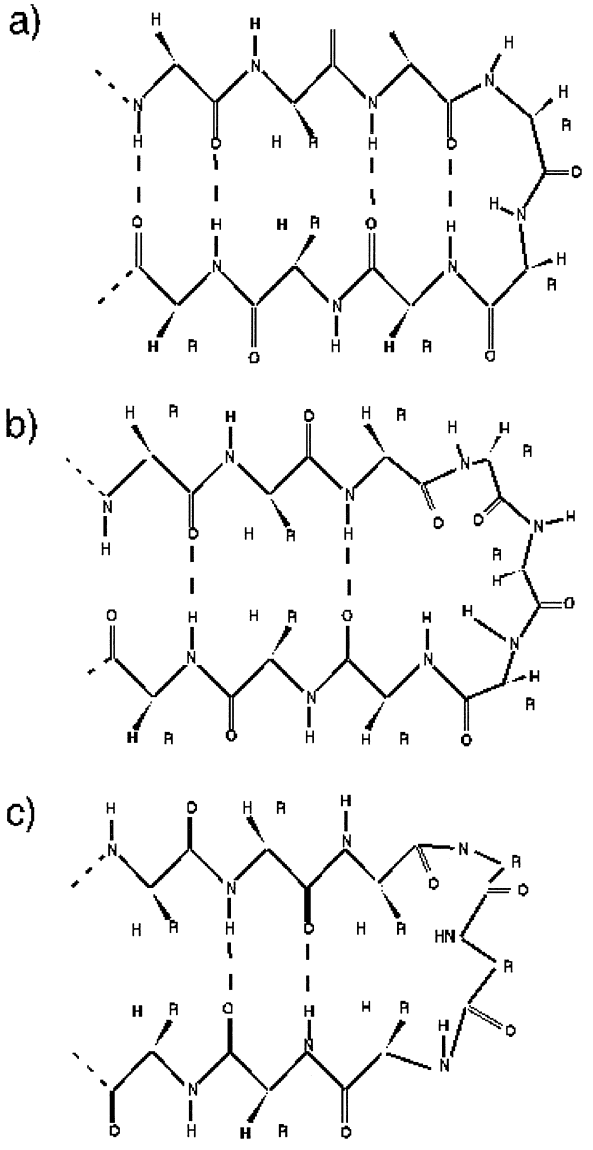

Figure 1. Schematic representation of the most abundant β-hairpin structures in proteins: (a) 2:2 β-hairpin, (b) 3:5 β-hairpin, and (c) 4:4 β-hairpin. R = amino acid side chain.

Table 1. Most common hydrogen-bonded β-turns in β-hairpins: dihedral angles

	L1		L2	
Turn type	ϕ	Ψ	ϕ	Ψ
Type I β	−60	−30	−90	0
Type I′ β	60	30	90	0
Type II′ β	60	−120	−90	0

Table adapted from ref 29.

most commonly observed throughout proteins, are not compatible with a 2:2 antiparallel β-hairpin. The reason for it is purely geometric, as these turns cannot direct the polypeptide chain to match the right-handed twist which is always observed in β-sheets. More recently, it was found that β-hairpins having two-residue type I β-turns tend to have longer strands than those formed by type I′ β-turns.[36] This observation suggests that the additional interstrand hydrogen bonding in longer strands may indeed compensate for any unfavorable interaction in adjusting the strand conformation.

Peptide systems to dissect the energy contributions to β-hairpin stability

Natural β-hairpin fragments. Most natural peptides encompassing β-hairpins in protein structures studied to date are mainly devoid of native structure in water or form aggregates.[37–48] An exception is the 16-residue fragment at the C-terminus of protein G B1 domain, corresponding to the second β-hairpin of this domain. This fragment adopts the native hairpin structure (a 4:4 hairpin) with a population of ∼40% in aqueous solution.[49] This highly populated β-hairpin is stabilized by a complex hydrogen-bond network involving residues at the β-turn, and by the high number of Thr and aromatic residues in the β-strands, that have been found to be good β-sheet formers.[24,50]

The first reported case of a linear peptide with a modified natural sequence forming a β-hairpin was the 2:2 β-hairpin from the α-amylase inhibitor Tendamistat.[51] The native sequence formed transiently populated turns in the central region, but no β-hairpin structure. Replacement of the turn residues by a sequence with a high frequency in the protein database to form a type I β-turn (Asn-Pro-Asp-Gly) resulted in a highly populated 3:5 non-native hairpin with a three-residue turn (type I β-turn plus a β-bulge), in equilibrium with other alternative hairpin conformations populated to a minor extent. Similar results were found with a 16-residue peptide derived from the first β-hairpin of ubiquitin, when the authors introduced the same Asn-Pro-Asp-Gly sequence.[52]

Both the modified natural β-hairpin fragments and the β-hairpin of protein G B1 domain have complex interactions that make it difficult to deduce general principles of β-hairpin formation. Moreover, in some cases a mixture of different hairpin structures with alternative strand pairing is found. Therefore, it has been necessary to find different alternative simpler model systems populating only one type of structure and having simple side-chain interactions involving only a small subset of residues, as it is discussed below.

Designed de novo peptide systems

A peptide model system required certain characteristics: It should be as small as possible, while retaining all the features of the structural element studied, but with the minimum number of interactions necessary for structure formation. The peptide should also be soluble and monomeric in the millimolar concentration range, necessary for nuclear magnetic resonance (NMR) studies. Currently, there are two main model systems used (Table 2): One based on a complete de novo design (BH system[53–55]) and the other produced after an iterative de novo design process starting with the modified Tendamistat fragment mentioned above, but resulting in a completely changed sequence (Rico and co-workers system).[56–58] In the following section we will explain in

Table 2. β-hairpin sequences discussed in this review

Name	Sequence	Designed	Hairpin Formed
Peptide 1t	YQNPDGSQA	type II	type I + G1 bulge
Peptide 2t	IYSNPDGTWT	type II	3:5, 4:4
Peptide 3t	IYSNSDGTWT	type II	3:5, 4:4
Peptide 4t	IYSNSDGVWT	des 3:5	des 3:5
Peptide 5t	IYSAPDGTWT	stab. 4:4	dec. pop. 3:5
Peptide 6t	IYSAKAGTWT	BPTI 4:4 turn	4:4 turn
Peptide 7t	IYSYNGKTWT	I′ β-turn	I′,II′ turns
Peptide 1s	(3T) IYSNSDGTWT	4:4	3:5(30%),4:4(30%)
Peptide 2s	SYINSDGTWT	4:4	3:5(n.d),4:4(50%)
Peptide 3s	YISNSDGTWT	4:4	3:5(n.d),4:4(20%)
Peptide 4s	SYSNSDGTWT	4:4	3:5(30%),4:4(40%)
Peptide 5s	YITNSDGTWT	4:4	3:5(n.d),4:4(80%)
Peptide 6s	YYTNSDGTWT	4:4	3:5(10%), 4:4(40%)
Peptide 7s	TISNSDGTWT	4:4	3:5(n.d), 4:4(10%)
BH8	ITVNGKTY	I′β-turn 2:2	I′β-turn 2:2
BH1	ATANGATA	no hairpin	no hairpin
BH9	ITVDGKTY	2:2	2:2
BH11	ITVAGKTY	2:2	2:2
BH12	ITVGGKTY	2:2	2:2
BH13	ITVSGKTY	2:2	2:2
BHEK	EITVNGKTYK	2:2	2:2
BHAK	AITVNGKTYK	2:2	2:2
BHKE	KITVNGKTYE	2:2	2:2
BHAE	AITVNGKTYE	2:2	2:2
BHKA	KITVNGKTYA	2:2	2:2
BHAA	AITVNGKTYA	2:2	2:2

des = destabilized structure, dec. pop.: decrease population.

detail the different problems encountered when designing a model peptide system, as well as the principles and strategies used in the design process. Later, we will dwell on the requirements to obtain energetic contributions from model peptide systems and finally discuss the results obtained so far by different groups.

Avoiding aggregation in β-hairpin peptides

Aggregation constitutes a major problem in β-hairpin peptides. Even small amounts of aggregated material can result in a wrong interpretation of the NMR data, especially of nuclear Overhauser effects (NOEs).[59] It is therefore essential to design a system that will not aggregate and to test experimentally that the designed molecule is monomeric. There are several strategies that can be employed to prevent aggregation in a peptide system: (1) to have an overall net charge under the pH conditions used to analyze the peptide structure; or (2) to avoid having an amphiphilic secondary structure with one hydrophobic face; or (3) to block the formation of hydrogen bonds on the groups which are not involved in intrachain hydrogen bonds in the target structure.

The first two strategies have been applied in the design of the BH peptide system.[53] In this system, the typical pattern of alternate hydrophobic and hydrophilic residues of β-sheets in proteins was not followed.[60] This cannot, however, prevent lateral oligomerization of the extended polypeptide chains. To avoid this problem, the authors decided as an additional strategy to place Arg residues at the N- and C-termini of the designed peptides. This has the disadvantage that upon formation of the β-hairpin, an electrostatic repulsion could take place. Gly is the most conformationally unrestrained amino acid and generally breaks secondary structure. Therefore, placing Gly residues as spacers between the Arg and the hairpin residues should result in the first and last two residues not forming part of the hairpin and should consequently diminish this putative electrostatic effect. The third strategy was applied using *N*-methyl amino acids.[61]

Sequence design for a β-hairpin system

Several factors need to be taken into consideration when designing the sequence for a β-hairpin structure. They can be roughly classified as those regarding the experimental analysis, and those important for obtaining the desired structure. The first group includes the sequence requirements important to avoid aggregation, mentioned above, and also the need to have well dispersed NMR signals. Since the structured population of a designed model peptide system will not necessarily be high, it is important to restrict the number of the same amino acids of the same type and to have an aromatic residue in the sequence. In this way overlap of NMR signals that could prevent identification of critical NOEs required to determine the existence of a hairpin structure can be avoided. As a result, the designed sequence will be a compromise between the experimental requirements to avoid aggregation and a good experimental system, and the need to have sufficiently stabilizing interactions. The following section will explain in some detail the strategy used to design completely de novo the BH model peptide system used to study β-hairpin formation.

In the protein structure database, the most abundant regular hairpins are those with two-residue turns,[27] and

three-residue strands (The positions are defined as (−B3, −B2, −B1, L1, L2, B+1, B+2, B+3). The BH system was designed to have that structural composition. Positions −B2 and +B2 were fixed to be occupied by threonine residues since they have a high intrinsic preferences for β-sheet formation,[21,22,62] and their simultaneous presence in the nonhydrogen bonded site of an antiparallel β-sheet is statistically highly favorable in proteins.[50] At the central positions of the turn (L1 and L2), the pair Asn-Gly was placed, since it is the most frequent in type I′ β-turns in β-hairpins. At other positions, the selection was based on individual and residue-pair preferences in the database, trying to place an aromatic residue at one of the four positions. Based on all the above criteria, a peptide sequence (peptide BH8; RG-ITVNGKTY-GR), was chosen as a good candidate for folding as a β-hairpin (Table 2). NMR and CD analysis of this peptide revealed that it was monomeric and adopted the expected conformation in aqueous solution with an estimated population of ∼35% (Fig. 2).[53,54]

Regarding the system of Rico and co-workers,[56] it was originally based on a modified sequence from the α-amylase inhibitor, Tendamistat.[51] This sequence was modified by placing residues with higher β-sheet

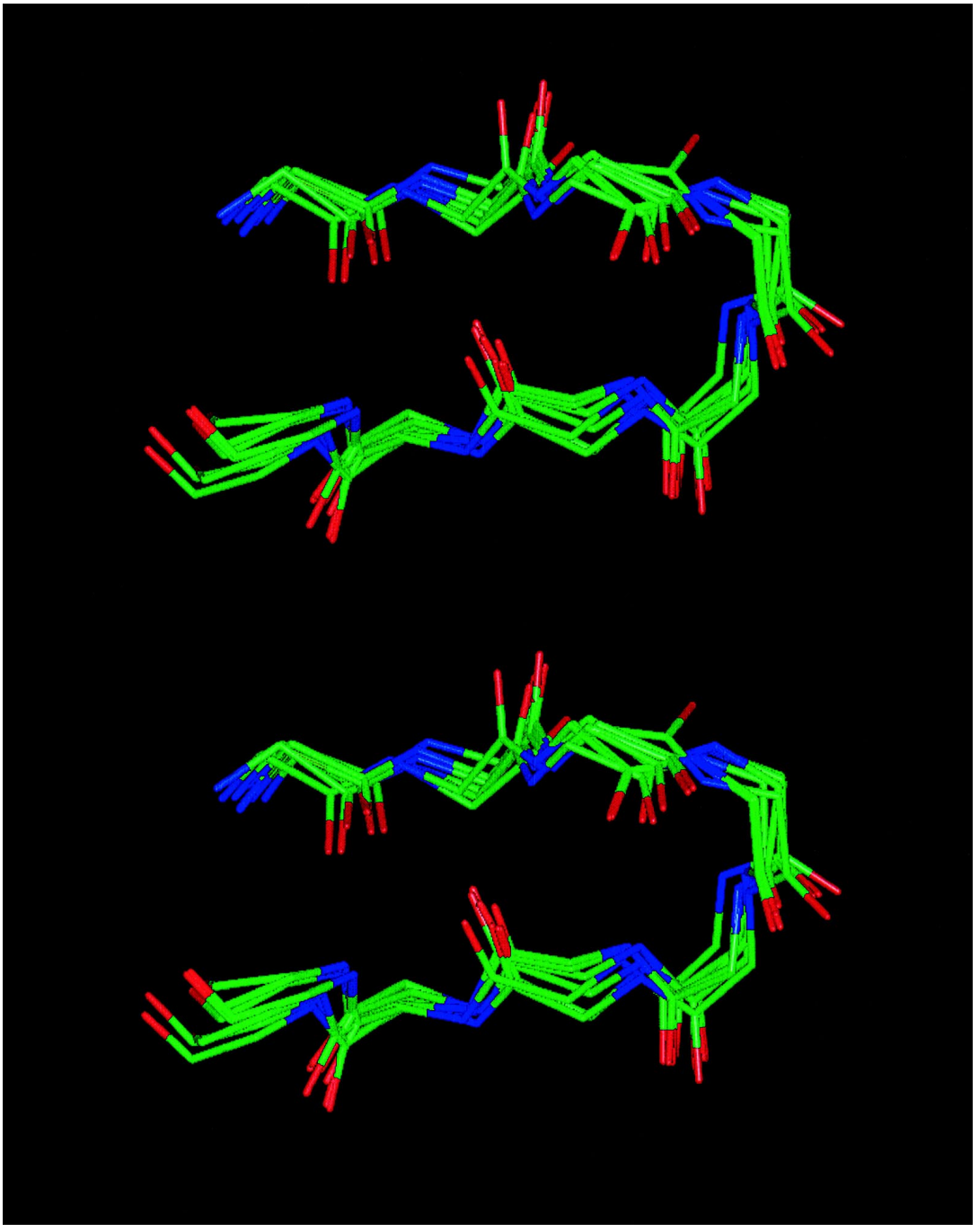

Figure 2. Stereo view of the superimposed ten best calculated structures of peptide BH8 from residue 3–10, showing backbone atoms.[53]

propensities on the β-strands to improve β-hairpin formation. The resulting sequence adopted a β-hairpin conformation in solution with two hairpin populations (a 3:5 and a 2:2 hairpin), providing a good system to explore the factors determining the pairing of β-strands. This indicates that a design strategy based on the statistical analysis of the protein database, in combination with experimental propensities can be successful.

Quantification of the structured population

A method for the quantification of the structured population is essential for the use of designed peptides as appropriate model systems for the formation and stability of secondary structure elements. In contrast to what happens with α-helix structure, the far-UV circular dichroism (CD) spectrum is not the most appropriate method for estimation of the hairpin population,[63] due to the much smaller CD signal and the strong influence of the β-turn conformation. As a result, in some cases, β-hairpins could show similar structured populations with different CD spectra. The advantage of NMR over CD is that there are several parameters that can be used to provide an internal consistency of the calculation. In addition, information at a residue level can be obtained, and therefore the population at the β-strands can be measured separately without interference of the turn region. Currently, and in the absence of the appropriate theoretical models for analyzing the experimental data (see below), the basis to determine the population of a β-hairpin peptide lies in the assumption of a two-state unfolding transition. This means, that the peptide is either folded or unfolded and there are no intermediate conformations.

The most commonly used NMR parameters for structure quantification are: (a) the NOE intensity is the most sensitive method to detect a transiently populated folded conformation due to its inverse dependency with the sixth power of the inter-proton distance. This emphasizes the short distances typical of folded conformers. However, the number of NOEs that are distinct to the hairpin structure are in general small and in many cases overlapped with other NOEs, and the NOE intensity measurement could contain effects arising from experimental limitations; and (b) the chemical shifts of Cα protons in proteins are largely dependent on the dihedral angle ϕ; the ^{13}Cα carbon in proteins are dependent on the ϕ, Ψ dihedral angles[64–67] although both chemical shifts mentioned before are also affected by the chemical environment (in the case of ^{13}Cα to a smaller extent).

The differences between the values obtained in random coil model pentapeptides,[68,69] and in proteins (conformational shifts) is related to the secondary structure conformation. In β-sheets, the exclusive use of the Cα proton or ^{13}Cα carbon conformational shifts for a population estimate can lead to errors since the backbone dihedral angles of the β-strands are not as regular as in α-helices, and vary from position to position. They are less sensitive to detect folded structures than NOEs since they are linear averages over all the conformers. However, they can be easily measured, with high accuracy, at any position of the peptide. (c) The $^{3}J_{NH_\alpha}$ coupling constants are also linear averages over all conformers. They are independent of the particular chemical environment and measure only the ϕ dihedral angle population at each residue position of the main chain. However, the precision with which the $^{3}J_{NH_\alpha}$ coupling constants can be measured is much lower than for the chemical shifts, and signal overlapping can prevent their measurement at some positions.

At the moment, the most appropriate strategy is to use all of the possibilities and try to integrate the different values in a single picture. In addition, the choice of the references for the two conformational states (0% and 100% populations) can influence the population values, especially when these are small. The same reference for all the methods should be used whenever possible, to minimize this problem. In the case of the intensity between Cα–Cα proton NOEs, the reference for 100% is taken from the expected value assuming a fixed distance between two strands in β-hairpins obtained from the protein structure database. For the other two parameters, the reference for 100% population cannot be obtained from average values in the protein structure database, since as explained above, the dihedral angles for each position in the hairpin vary from position to position. One way to obtain the correct reference values is to enhance the structured population of the peptide by adding organic solvents (methanol or 2,2,2-trifluoroethanol) and to check that the intensity of the Cα–Cα proton NOEs is consistent with a population close to 100%. For 0% population, the reference values can be obtained from random coil peptides,[68,69] or from the protein structure database.[70,71]

Having the adequate reference for 100% population (obtained either from proteins or from the peptide with an enhanced population) and 0%, the conformational shifts and the $^{3}J_{NH_\alpha}$ coupling constant values can be used for population estimates applying the following equation:

$$\%\ \text{Structure} = 100 \times \frac{(\text{NMR par. peptide} - \text{NMR par.0\% ref})}{(\text{NMR par.100\% ref.} - \text{NMR par.0\% ref})}$$

where:

NMR par.100% ref = NMR parameter value used as a 100% reference

NMR par. 0% ref = NMR parameter value used as a 0% reference.

(the random coil values that are usually reported).

NMR par. peptide = NMR parameter value measured in the peptide studied

In the case of the NOE intensities, the method described by Bradley,[72] can be used. It assumes that the overall correlation time of the molecule is the same in both states (folded and random coil). The average over the different conformations is nonlinear due to the r^{-6} dependence of the NOE intensity on the inter-proton

distance. The contribution of the random-coil state to the intensity of the diagnostic NOE (Cα–Cα proton across β-strands), is considered to be zero since this distance is very large, and the distance for the NOE diagnostic of the secondary structure is taken from the average value found in proteins. This method requires an internal calibration for translating NOE cross-peak intensities to inter-proton distances to estimate the hairpin population. The reference is usually taken from the NOE intensity of protons with constant distances in the folded and random coil states (i.e. CαH of a glycine residue).

The equation used for the intensity of diagnostic NOE is the following:

$$\%\ \text{structure} = 100 \times \left(I_{\beta-\text{hairpin}}/I_{\text{ref}}\right)/\left(r_{\text{ref}}/r_{\text{hairpin}}\right)^6$$

where:

r_{ref}	=	distance between the reference protons
$r_{\beta-\text{hairpin}}$	=	distance between the diagnostic protons involved in the diagnostic NOE
$I_{\beta-\text{hairpin}}$	=	intensity of diagnostic NOE
I_{ref}	=	intensity of reference NOE

All these methods to estimate β-hairpin population have been tested on different peptides derived from the BH system and found to be in good agreement.[54]

Obtaining an accurate estimation of the structured population in a peptide system is enough to determine the free energy of the system, or to quantify the effect of point mutations, as long as the system follows a two-state transition (β-hairpin and random coil). However, it is likely that partly folded conformations exist, which would contribute to NMR parameters, as found in α-helices. In this case statistical mechanics should be used to obtain free energy contributions for β-hairpin formation.

Discussion

Factors that determine β-hairpin stability

Turn preferences. The importance of the turn sequence in β-hairpin formation and stability has been analyzed by the structural characterization of a series of model peptides by various groups.[54,57]

Using the BH β-hairpin system,[53,54] it was found that the sequence Asn-Gly is sufficient to promote the formation of a β-turn, but not enough to promote β-hairpin formation without the side-chain interactions in both strands of the β-hairpin. An analysis of the dynamics on the model BH8 peptide by ^{13}C relaxation measurements, showed that the Glycine residue in the turn region displays restricted motions at high temperature, while NOE conformational information is lost and CαH chemical shifts reached a plateau close to random coil values.[55] This finding supports the idea that the β-turn by itself can be partly folded. Rico and co-workers have analyzed different peptides with several different sequences at the turn region[57] (Table 2, peptides 1t–7t). Remarkably, the sequence at the turn position in this β-hairpin system determines not only the stability of the hairpin, but also the register of the β-strands. Furthermore, it is possible to rationally modify the sequences to attain some of the alternative pairings of the β-strands. Especially interesting is the peptide t7 (IYSY-**NG-KTWT**) which is very similar to the BH8 peptide (GVTI-**NG-KTY**GT). This peptide shows a mixture of two populations, one with a type I′ β-turn at the Asn-Gly positions, and the other with a type II′ turn at the Gly-Lys positions. Statistical analysis of anti-parallel β-sheets in proteins[50] reveals that the residue pairing is quite favorable in both conformations. In the case of the BH8 peptide, the alternative β-hairpin structure would put Gly11 at the end of the second strand, which will not produce any favorable side chain–side chain interaction. Therefore, it is likely that the conformational mixture found in the peptides analyzed by Rico and co-workers,[57] is due to the possibility of establishing good side chain–side chain interactions in both two different alternative conformations.

Mutational analysis at position L1 of the BH8 peptide, in which Asn6 has been substituted by Asp, Ser, Ala and Gly (peptides BH9, BH11, BH12, BH13, see Table 2), has shown that all the turn variants formed regular two-residue turn β-hairpins but with different populations.[54] The different stabilizing properties of these residues correlate very well with the statistical abundance in the protein database at position L1 of 2:2 β-hairpins with type I′ β-turns. The relative positions of Asn, Ser and Ala within this hierarchy can be explained by their intrinsic residue preferences in populating the backbone dihedral angles (ϕ, Ψ),[70,71] corresponding to the first position of a type I′ β-turn. Asp and Gly were found to be more and less stabilizing, respectively, than expected from their ϕ, Ψ propensities. In the case of Asp, this can be explained by a specific side chain χ_1 rotamer preference that creates a favorable electrostatic interaction with its own amide group only when it is preceded by a residue in an extended conformation.

Therefore, it seems clear that β-turns can contribute significantly to the stability of a β-hairpin. In some cases very strong turn tendencies[57] can overcome poor side chain–side chain interactions and force the formation of a β-hairpin, or shift the register between β-strands.

Side-chain–side-chain interactions

Mutational analysis in the B1 domain of protein G, has shown the importance of side chain–side chain interactions across β-strands in protein stability.[24] These results are corroborated by similar studies in the β-hairpin model peptide systems. In the case of the BH system it has been found that the packing of voluminous hydrophobic side chains provides part of the stabilization energy required for β-hairpin formation[53] as has been proposed by theoretical calculations.[73] These results suggest that hydrophobic interactions between the strands stabilize and determine the conformation of

the β-hairpin. Also, electrostatic interactions across strands have been found to significantly contribute to the stability of the hairpin.[74] The fact that these interactions are conformationally specific in nature has been shown in the BH system by mutating residues with similar stereospecificity and different hydrophobicities (i.e. Ile into Leu), or the opposite (i.e. Tyr into Phe), as well as by changing the order of ion pairs across β-strands (Glu-Lys and Lys-Glu)(peptides BHEK, BHKE, and controls BHAA, BHAK, BHKA, BHAE see Table 2), which produced significant changes in the stability of the peptide. The group of Rico and coworkers, have also recently shown that changes in the residues of the strands can produce changes in the relative populations of the different hairpin structures coexisting in their system (Table 2).[57] However, the authors proposed that the intrinsic conformational preferences of the turn are more important in β-hairpin conformation than particular patterns of cross-strand side chain–side chain interactions. It is interesting to point out that these authors found in their system and in the protein structure database, that the twist of the β-hairpin is more pronounced in the 3:5 class than in the 4:4 class.[58] This suggests that the turn could influence the twist of the β-sheet and consequently the packing of side chains.

Intrinsic propensities

It is likely that the intrinsic secondary structure propensities of the different amino acids[22,25,50] contribute to the stability of β-hairpins. However, it is difficult to determine their relative contribution, without affecting interstrand side-chain–side-chain interactions in the case of the strand variants.

What can be learned from design approaches based on simple models for β-hairpin formation? Different interactions have been found to be important for the stability of β-hairpins, such as intrinsic propensities,[22,24,53] hydrogen bonding,[75] electrostatic side-chain,[74] hydrophobic interactions,[44,47,53,57] and conformational preferences in β-turns.[54,57,76] All these results indicate that the stability of hairpin structures is the result of a delicate balance involving different specific interactions ranging from hydrophobic interactions to hydrogen bonds, as also found in α-helices. Depending on the system, any of the interactions mentioned above could have a more pronounced contribution. Therefore, no specific force can be considered as the main contributor of hairpin stability.

The application of knowledge about β-sheet formation: from β-hairpins to β-sheets

As we have seen above, the β-hairpin systems can offer important information to understand the formation of β-sheet structures in proteins. However, in proteins β-sheets normally consist of more than two strands packing together. This means that the residues in the central strand will have a different environment than any of the two edge strands in a β-hairpin system. Therefore, the next logical step to understand β-sheet formation was to extend the studies on β-hairpin formation in peptide model systems to the design of a higher-order structure, a three-stranded β-sheet.

Several attempts have been made to the design of all β-sheet proteins.[77–81] A number of de novo design studies were based on statistical rules derived from known protein structures.[77–79] Other approaches used modifications of chimerical sequences taken from natural β-sheet proteins.[80,81] CD studies indicated the formation of β-sheet secondary structure in all of these cases, but aggregation presented a major obstacle to detailed structural characterization of the designed protein. No high-resolution structure has been obtained of a well-folded, monomeric designed β-sheet protein, although an NMR model of β-sheet formation coupled to oligomerization has been reported.[81]

Our approach for β-sheet design was based on the simple BH peptide model system for β-structure and relied on a twofold strategy: In an initial step, a target backbone framework was chosen, a three-stranded antiparallel β-sheet with four residues per strand and two-residue turns (Fig. 3). Subsequently, the selection of a suitable sequence was performed using experimental results from the de novo BH β-hairpin system described above,[53–56] statistical information derived from the protein structure database, information from structurally stabilizing motifs and an evaluation of packing interactions through modeling of side-chain rotamers compatible with the target structure using the program ICM.[82]

Several sequences were designed by placing a longer version of BH8 (peptide BHKE, Table 2), at the positions of β-strands 2 and 3 of the target backbone, and choosing a suitable sequence for β-strand 1 by rotamer modeling. However, all peptides designed using this strategy failed to show incorporation of the β-strand 1 into the β-sheet while β-strands 2 and 3 formed a stable β-hairpin in aqueous solution. The expected three-stranded β-sheet structure was, however, formed in the presence of co-solvents such as 2,2,2-trifluoroethanol (TFE) and methanol, known to enhance secondary structure formation in peptides.[83] Similar results were obtained in a different de novo peptide system designed to form a three-stranded antiparallel β-sheet.[84] Whereas there was no evidence for significant structured population in water, the peptide appeared to adopt β-structure in 50% (v/v) methanol–water mixtures.

The origin of the cosolvent effect on secondary structure formation is not fully understood.[83] Nevertheless, these findings indicated that the energetic contribution of the interactions in our system was not sufficient to drive the formation of the three-stranded β-sheet in water, although the side-chain packing was compatible with the target structure, as it could be adopted in a different solvent. The reason could be insufficient burial of hydrophobic surface. In the case of the BH system it was found that the Tyr side chain at position B+3 establishes contacts with the side chains of residues −B3, −B1 and +B1, thus creating a small hydrophobic cluster. Based on this and on the negative results obtained

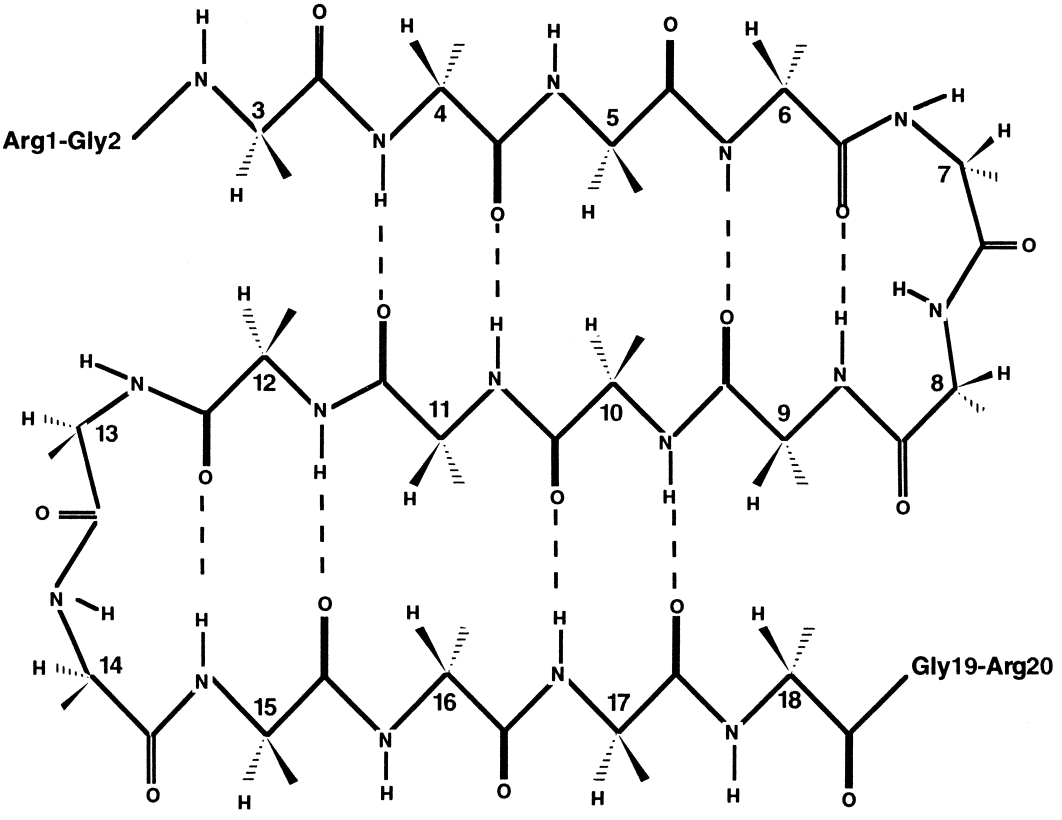

Figure 3. Schematic representation of a three-stranded antiparallel β-sheet, with the additional first and last two residues required for the solubility in the BH system.[53–56]

by placing aliphatic side chains on the first β-strand, we reasoned that an aromatic side chain in the first β-strand could bury more hydrophobic surface, incorporating this strand into the three-stranded β-sheet. As a consequence, a redesign of the sequence was performed using rotamer modeling as before. We found that an aromatic group in the first β-strand can only interact with the second strand in the absence of β-branched residues on the same face of the second β-strand. A similar local motif is found in the WW domain.[85] Therefore, we introduced a cluster of four residues of the WW domain containing a Trp residue in the first β-strand, into our design. This sequence (RG-WSVQNGKYTNNGKTTE-GR) was termed Betanova. A molecular model of the Betanova molecule calculated from NOE distance and coupling constant restraints showed the presence of the expected β-sheet structure, stabilized by specific tertiary interactions of the tryptophane residue packing onto one sheet surface and making contacts with residues on β-strands 2 and 3, as expected from the design, furthermore, thermal and chemical denaturation of the Betanova molecule showed a cooperative unfolding transition, which is a hallmark of natural protein.[88]

Conclusions

Implications for β-sheet protein design and engineering

The success of the Betanova design demonstrates that we begin to understand the principles behind β-sheet stabilization. For Betanova, the main factor allowing the formation of a unique β-sheet conformation appears to be the introduction of long-range side-chain contacts between β-strands 1 and 3. These tertiary interactions form the underlying basis of the observed cooperativity in the unfolding transition of Betanova, since the two hairpins making up the peptide (comprising β-strands 1 and 2, or β-strands 2 and 3) are expected not be stable in isolation. This inherent instability of isolated elements of secondary structure, which are stabilized by context-dependent tertiary interactions in Betanova, is also one of the key features of natural proteins.

In the case of α-helices, an empirical description of the α-helix coil transition taking into account several different individual, or pairwise, interactions, has been very successful for the prediction of helix stability in the absence of tertiary interactions,[5,6] and for the rational modification of protein stability.[17–19] Due to the complexity of the interactions present in β-sheets such an empirical description will probably not be feasible. For example, statistical analysis suggests that the energetic contribution of a pairwise inter-strand side-chain interaction depends on the face of the β-sheet on which they are located,[50] intrinsic propensities are likely to display a significant position-dependence along the strands as well as in the different turn positions, and will vary for the different turn types. Furthermore, parallel β-sheets will vary in their properties from the antiparallel β-sheets considered here. Taken together, these different contributions would yield an astronomical number of parameters to be experimentally determined for any

empirical model of β-sheet formation. Most likely an all-atomic model will be needed for a rational approach to β-sheet design and stability. Simple peptide model systems adopting β-hairpin and β-sheet conformations will be essential for the testing and refinement of such an algorithm.

References and Notes

1. Bryngelson, J. D.; Onuchic, J. N.; Socci, N. D.; Wolynes, P. G. *Proteins: Struct., Funct., Genet*. **1995**, *21*, 167.
2. Karplus, M.; Sali, A. *Curr. Opin. Struct. Biol*. **1995**, *5*, 58.
3. Dill, K. A. *Biochemistry* **1990**, *29*, 7133.
4. Scholtz, J. M.; Baldwin, R. L. *Annu. Rev. Biophys. Biomol. Struct*. **1992**, *21*, 95.
5. Chakrabartty, A.; Baldwin, R. L. *Adv. Protein Chem*. **1995**, *46*, 141.
6. Muñoz, V.; Serrano, L. *Curr. Opin. Biotech*. **1995**, *6*, 382.
7. Minor, Jr. D. L.; Kim, P. S. *Nature* **1996**, *380*, 730.
8. Muñoz, V.; Serrano, L. *Fold. Des*. **1996**, *1*, 167.
9. Brown, J. E.; Klee, W. A. *Biochemistry* **1971**, *10*, 470.
10. Marqusee, S.; Baldwin, R. L. *Proc. Natl. Acad. Sci. USA* **1987**, *84*, 8898.
11. Marqusee, S.; Robbins, V. H.; Baldwin, R. L. *Proc. Natl. Acad. Sci. USA* **1989**, *86*, 5286.
12. Zimm, B. H.; Bragg, J. K. *J. Chem. Phys*. **1959**, *31*, 526.
13. Muñoz, V.; Serrano, L. *Biopolymers* **1997**, *41*, 495.
14. Lomize, A. L.; Mosberg, H. I. *Biopolymers* **1997**, *42*, 239.
15. Andersen, N. H.; Tong. H. *Protein Sci*. **1998**, *6*, 1920.
16. Misra, G. P.; Wong, C. F. *Proteins: Struct., Funct., Genet*. **1997**, *28*, 344.
17. Villegas, V.; Viguera, A. R.; Aviles, F. X.; Serrano, L. *Fold. Des*. **1996**, *1*, 29.
18. Muñoz, V.; Cronet, P.; Lopez-Hernandez, E.; Serrano, L. *Fold. Des*. **1996**, *1*, 167.
19. Viguera, A. R.; Villegas, V.; Aviles, F. X.; Serrano, L. *Fol. Des*. **1996**, *2*, 23.
20. Chothia, C. *J. Mol. Biol*. **1973**, *75*, 295.
21. Kim, C. A.; Berg, J. M. *Nature* **1993**, *362*, 267.
22. Minor, Jr. D. L.; Kim, P. S. *Nature* **1994**, *367*, 660.
23. Minor, Jr. D. L.; Kim, P. S. *Nature* **1994**, *371*, 264.
24. Smith, C. K.; Regan, L. *Science* **1995**, *270*, 980.
25. Muñoz, V.; Serrano, L. *Proteins: Struct., Funct., Genet*. **1994**, *20*, 301.
26. Ptitsyn, O. B. *FEBS Lett*. **1991**, *131*, 197.
27. Sibanda, B. L.; Thornton, J. M. *Nature* **1985**, *316*, 170.
28. Sibanda, B. L.; Blundell, T. L.; Thornton, J. M. *J. Mol. Biol*. **1989**, *206*, 759.
29. Rose, G. D.; Gierasch, L. M.; Smith, J. A. *Adv. Protein Chem*. **1985**, 37.
30. Chou, P. Y.; Fasman, G. D. *Biochemistry* **1977**, *13*, 211.
31. Mattos, C.; Petsko, G. A.; Karplus, M. *J. Mol. Biol*. **1994**, *238*, 660.
32. Wilmot, C. M.; Thornton, J. M. *J. Mol. Biol*. **1988**, *203*, 221.
33. Hutchinson, E. G.; Thornton, J. M. *Protein Sci*. **1994**, *3*, 2207.
34. Yang, A.-S.; Hitz, B.; Honig, B. *J. Mol. Biol*. **1996**, *259*, 873.
35. Ohage, E. C.; Graml, W.; Walter, M. M.; Steinbacher, S.; Steipe, B. *Protein Sci*. **1997**, *6*, 233.
36. Gunasekaran, K.; Ramakrishnan, C.; Balaram, P. *Protein Eng*. **1997**, *10*, 1131.
37. Blanco, F. J.; Jiménez, M. A.; Rico, M.; Santoro, J.; Herranz, J.; Nieto, J. L. *Eur. J. Biochem*. **1991**, *200*, 345.
38. Dyson, H. J.; Wright, P. E. *Annu. Rev. Biophys. Biophys. Chem*. **1991**, *20*, 519.
39. Cox, J. P. L.; Evans, P. A.; Packman, L. C.; Williams, D. H.; Woolfson, D. N. *J. Mol. Biol*. **1993**, *234*, 483.
40. Kemmink, J.; Creighton, T. E. *J. Mol. Biol*. **1983**, *234*, 861.
41. Blanco, F. J.; Jiménez, M. A.; Pineda, A.; Rico, M.; Santoro, J.; Nieto, J. L. *Biochemistry* **1994**, *33*, 6004.
42. Viguera, A. R.; Jiménez, M. A.; Rico M.; Serrano, L. *J. Mol. Biol*. **1996**, *255*, 507.
43. Blanco, F. J.; Serrano, L. *Eur. J. Biochem*. **1995**, *230*, 634.
44. Searle, M. S.; Zerella, R.; Williams, D. H.; Packman, L. C. *Protein Eng*. **1996**, *9*, 559.
45. Neira, J. L.; Fersht, A. R. *Fold. Des*. **1996**, *1*, 231.
46. Ramírez-Alvarado, M.; Serrano, L.; Blanco, F. J. *Protein Sci*. **1997**, *6*, 162.
47. Maynard, A. J.; Searle, M. S. *Chem. Commun*. **1997**, 1297.
48. Maynard, A. J.; Sharman, G. J.; Searle, M. S. *J. Amer. Chem. Soc*. **1998**, *120*, 1996.
49. Blanco, F. J.; Rivas, G.; Serrano, L. *Nat. Struct. Biol*. **1994**, *1*, 584.
50. Wouters, M. A.; Curmi, P. M. G. *Proteins: Struct., Funct., Genet*. **1995**, *22*, 119.
51. Blanco, F. J.; Jiménez, M. A.; Herranz, J.; Rico, M.; Santoro, J.; Nieto, J. L. *J. Am. Chem. Soc*. **1993**, *115*, 5887.
52. Searle, M. S.; Williams, D. H.; Packman, L. C. *Nat. Struct. Biol*. **1995**, *2*, 999.
53. Ramírez-Alvarado, M.; Blanco, F. J.; Serrano, L. *Nat. Struct. Biol*. **1996**, *3*, 604.
54. Ramírez-Alvarado, M.; Blanco, F. J.; Niemann, H.; Serrano, L. *J. Mol. Biol*. **1997**, *273*, 898.
55. Ramírez-Alvarado, M.; Daragan, V.; Serrano, L.; Mayo, K. *Protein Sci*. **1998**, *7*, 720.
56. de Alba, E.; Jiménez, M. A.; Rico, M.; Nieto, J. L. *Folding & Design* **1996**, *1*, 122.
57. de Alba, E.; Jiménez, M. A.; Rico, M. *J. Am. Chem. Soc*. **1997**, *119*, 175.
58. de Alba, E.; Rico, M.; Jiménez, M. A. *Protein Sci*. **1997**, *6*, 2548.
59. Wright, P. E. *Annu. Rev. Biophys. Biophys. Chem*. **1991**, *20*, 519.
60. Richardson, J. S. *Adv. Protein Chem*. **1981**, *34*, 167.
61. Doig, A. J. *Chem. Comm*. **1997**, 2153.
62. Smith, C. K.; Withka, J. M.; Regan, L. *Biochemistry* **1994**, *33*, 5510.
63. Johnson, Jr. W. J. *Annu. Rev. Biophys. Biophys. Chem*. **1988**, *17*, 145.
64. Williamson, M. P. *Biopolymers* **1990**, *29*, 1423.
65. Osapay, K.; Case, D. *J. Amer. Chem. Soc*. **1991**, *113*, 9436.
66. Herranz, J.; González, C.; Rico, M.; Nieto, J. L.; Santoro, J.; Jiménez, M. A.; Bruix, M.; Neira, J. L.; Blanco, F. *J. Magn. Res. Chem*. **1992**, *30*, 1012.
67. Oldfield, E. *J. Biomol. NMR* **1995**, *5*, 217.
68. Merutka, G.; Dyson, H. J.; Wright, P. E. *J. Biomol. NMR* **1995**, *5*, 14.
69. Wishart, D. S.; Bigam, C. G.; Holm, A.; Hodges, R. S.; Sykes, B. D. *J. Biomol. NMR* **1995**, *5*, 67.
70. Serrano, L. *J. Mol. Biol*. **1995**, *254*, 322.
71. Smith, L. J.; Bolin, K. A.; Schwalbe, H.; MacArthur, M. W.; Thornton, J. M.; Dobson, C. M. *J. Mol. Biol*. **1996**, *255*, 494.
72. Bradley, E. K.; Thomason, J. F.; Cohen, F. E.; Kosen, P. A.; Kuntz, I. D. *J. Mol. Biol*. **1990**, *215*, 607.
73. Yang, A.-S.; Honig, B. *J. Mol. Biol*. **1995**, *252*, 366.
74. de Alba, E.; Blanco, F. J.; Jiménez, M. A.; Rico, M.; Nieto, J. L. *Eur. J. Biochem*. **1995**, *233*, 283.
75. Constantine, K. L.; Mueller, L.; Andersen, N. H.; Tong, H.; Wandler, C. F.; Friedrichs, M. S.; Bruccoleri, R. E. *J. Am. Chem. Soc*. **1995**, *117*, 10841.
76. Haque, T. S.; Gellman, S. H. *J. Am. Chem. Soc*. **1997**, *119*, 2303.
77. Richardson, J. S.; Richardson, D. C. *Trends Biochem*. **1989**, *14*, 304.

78. Quinn, T. P.; Tweedy, N. B.; Williams, R. W.; Richardson, J. S.; Richardson, D. C. *Proc. Natl. Acad. Sci. USA* **1994**, *91*, 8747.
79. Yan, Y.; Erickson, B. W. *Protein Sci.* **1994**, *3*, 1069.
80. Pessi, A.; Bianchi, E.; Crameri, A.; Venturini, S.; Tramontano, A.; Sollazzo, M. *Nature* **1993**, *362*, 367.
81. Ilyina, E.; Roongta, V.; Mayo, K. H. *Biochemistry* **1997**, *36*, 5245.
82. Abagyan, R.; Totrov, M.; Kuznetsov, D. *J. Comp. Chem.* **1994**, *15*, 488.
83. Luo, P.; Baldwin, R. L. *Biochemistry* **1997**, *36*, 8413.
84. Sharman, G. J.; Searle, M. S. *Chem. Comm.* **1997**, 1955.
85. Macias, M. J.; Hyvönen, M.; Baraldi, E.; Schultz, J.; Sudol, M.; Saraste, M.; Oschkinat, H. *Nature* **1996**, *382*, 646.
86. Kabsch, W.; Sander, C. *Biopolymers* **1983**, *27*, 2577.
87. Vriend, G. *J. Mol. Graph.* **1990**, *8*, 52.
88. Kortemme, T.; Ramírez-Alvarado, M.; Serrano, L. *Science* **1998**, *281*, 253.