ELSEVIER

Biophysical Chemistry 101–102 (2002) 67–79

Biophysical Chemistry

www.elsevier.com/locate/bpc

Hydrogen exchange, core modules, and new designed proteins

Natàlia Carulla[a,1], George Barany[a], Clare Woodward[b,*]

[a]*Department of Chemistry, University of Minnesota, 207 Pleasant St. S.E., Minneapolis, MN 55455, USA*
[b]*Department of Biochemistry, Molecular Biology and Biophysics, University of Minnesota, 1479 Gortner Ave., St. Paul, MN 55108, USA*

Received 1 February 2002; accepted 27 March 2002

Abstract

A strategy for design of new proteins that mimic folding properties of native proteins is based on peptides modeled on the slow exchange cores of natural proteins. We have synthesized peptides, called core modules, that correspond to the elements of secondary structure that carry the very slowest exchanging amides in a protein. The expectation is that, if soluble in water, core modules will form conformational ensembles that favor native-like structure. Core modules modeled on natural bovine pancreatic trypsin inhibitor have been shown by NMR studies to meet this expectation. The next step toward production of a native state mimic is to further shift the conformational bias of a core module toward more ordered structure by promoting module–module interactions that are mutually stabilizing. For this, two core modules were incorporated into a single molecule by means of a long cross-link. From a panel of several two-module peptides, one very promising lead emerged; it is called BetaCore. BetaCore is monomeric in water and forms a new fold composed of a four-stranded, antiparallel β-sheet. The single, dominant conformation of BetaCore is characterized by various NMR experiments. Here we compare the individual core module to the two-module BetaCore and discuss the progressive stabilization of intramodule structure and the formation of new intermodule interactions.

Keywords: Protein design; H/D exchange; β-sheet protein

1. Introduction

The goal of protein design is to produce, for use in biological and non-biological systems, new molecules that mimic the properties of natural proteins [1–7]. Several hallmarks of natural proteins distinguish them from isomeric peptides with the same number and ratio of natural amino acids but in randomized sequences. These include: folding on the millisecond time scale at moderate pH and temperature to an ensemble of compact, low energy conformations of similar structure that are at once globally stable and internally mobile; adaptation of the irregular folded surface to specific binding of biological molecules and unique arrangements of catalytically active side chains;

*Corresponding author. Tel.: +1-435-946-2949; fax: +1-435-946-8252.
E-mail address: clare@biosci.cbs.umn.edu (C. Woodward).
[1] Present address: Department of Chemistry, University of Cambridge, Lensfield Road, Cambridge CB2 1EW, UK

0301-4622/02/$ - see front matter © 2002 Elsevier Science B.V. All rights reserved.
PII: S0301-4622(02)00149-7

and maintenance of sufficient instability to permit clearance from the cell after functional use. In view of their selection and refinement by billions of years of biological evolution, it is not surprising that these properties are generally elusive to chemists seeking to incorporate them into new proteins. Just how rare natural proteins are in the universe of polypeptides with the same amino acid composition is readily appreciated by anyone who attempts to synthesize a new sequence, only to be confronted by insolubility and other intractable molecular behaviors.

The design strategy we pursue is aimed at producing new proteins that mimic one of the above mentioned hallmarks of natural proteins, namely the spontaneous folding in aqueous solution to a family of closely related conformations that are more stable than any other accessible conformations. We refer to this design target as a native state mimic, after the common designation of N for native state in the N-to-D reaction used with reference to global unfolding and denaturation. In an analysis of amide hydrogen isotope exchange in proteins, a strong tendency was noted for the same secondary structural elements to contain the amide groups that are slowest exchanging in the native protein, the amide groups that are first protected during folding and the amide groups that are most protected in partially folded analogs [8]. Based on this, we proposed that the slow exchange core is the folding core and that 'core elements' (secondary structure containing the slowest exchanging amide protons) are the most likely to be native-like in the earliest stages of folding. Of several inferences drawn from this generalization [8–11], one is that peptides corresponding to slow exchange core elements of a natural protein will favor native-like structure and can form the basis of a strategy for design of new proteins that mimic the native state. Favored native-like structure of core peptides refers to the collection of random and non-random conformations, in which the non-random conformations are biased toward native-like.

Following this line of reasoning, 'core modules' (CMs), peptides containing the secondary structural elements in the slow exchange core of the bovine pancreatic trypsin inhibitor (BPTI), were constructed. BPTI CMs in aqueous solution are ensembles of ordered and disordered molecules among which ordered subpopulations are detected by NMR spectroscopy. The major ordered conformations are native-like 4:4 β-hairpins, in equilibrium with a smaller population of non-native 3:5 β-hairpins [12]. The next step was to further stabilize packed native-like structure within the modules and to organize them into a larger molecule. For this, two modules were joined by a covalent cross-link designed to promote unimolecular self-association of modules that is both specific and mutually stabilizing. A series of two-module analogs were constructed, with cross-links of varying position and length, and several proved to be promising [13]. One of these, called BetaCore, shows significant progress toward the design goal [14]. While not precisely a native state mimic, it is monomeric and soluble in water and has one very dominant conformation composed of a four-stranded antiparallel β-sheet that is amenable to characterization by NMR. In folded BetaCore, the component modules are clearly not three-dimensionally identical, although they are essentially covalent repeats. However, as BetaCore undergoes thermal denaturation, the unfolded species show the symmetry expected for identical disordered domains. Here, we compare the structure of the BPTI core module to the two-module molecule, BetaCore, and discuss the implications for future design of native state mimics based on CMs.

2. Experimental

The core module (CM) and BetaCore were chemically synthesized, either unlabeled or selectively ^{15}N-labeled, by automated Fmoc solid-phase peptide synthesis according to methods previously described [12,13]. CM peptide and BetaCore were confirmed to be monomeric in water by size-exclusion chromatography [12] and sedimentation equilibrium analysis [14], respectively.

NMR samples for CM were 0.5–0.8 mM in 50 mM sodium acetate buffer, pH 4.5, and for BetaCore were 0.4 mM at pH 3. Samples in D_2O were dissolved under argon. Spectra of CM and BetaCore were obtained in 90:10 H_2O/D_2O and

99.9% D_2O at temperatures ranging from 1 to 55 °C on a Varian 600-MHz or 800-MHz Inova instrument. Spectral acquisition parameters for homo- and heteronuclear TOCSY, NOESY, ^{15}N HSQC, ^{15}N-^{1}H HSQC-TOCSY, ^{15}N-^{1}H HSQC-NOESY and HNHA experiments are given in [12,14]. Suppression of intense solvent resonances was achieved by presaturation or use of the WATERGATE sequence [15]. Data were processed and analyzed using the programs NMRPipe [16] and NMRView [17]. Data processing details are given in [12,14]. An ensemble of conformations was calculated, using methods previously described, for the CM [12] and for BetaCore [14]. The coordinates and constraint files for the ensemble of 20 structures calculated for BetaCore have been deposited in the Protein Data Bank (accession code: 1K09). Chemical shifts and coupling constants have been deposited in the BioMagResBank (accession code: 5183).

Hydrogen isotope exchange rates for CM were obtained by measuring peak volumes vs. time in a series of two-dimensional TOCSY spectra at pH 4.15 and 6 °C and for BetaCore were obtained by measuring peak volumes vs. time in a series of two-dimensional ^{15}N HSQC spectra at pH 3 and 5 °C. Pseudo-first order rate constants were determined from non-linear least-squares fit of an exponential rate equation to experimental data. Thermal unfolding curves for BetaCore were obtained by measuring peak volumes vs. temperature in a series of two-dimensional ^{15}N HSQC spectra. A relaxation delay time of 3 s was used to permit quantitative volume integration. Reversibility of thermal denaturation was verified by comparison of low temperature spectra acquired before and after unfolding.

3. Results

3.1. Synthesis of peptides based on core elements of BPTI

The BPTI core elements are two strands of antiparallel β-sheet (residues 18–24 and 29–35) and a very small β-bridge (residues 43–44) (Fig. 1). A CM, designed to correspond to the antiparallel β-sheet, is comprised of 25 residues spanning

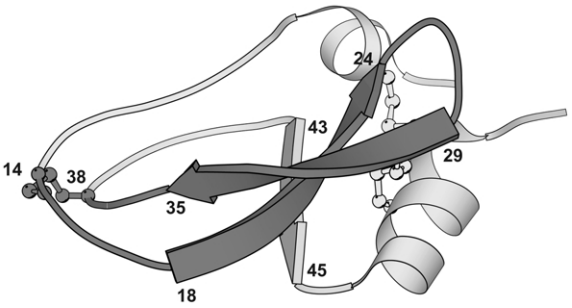

Fig. 1. The BPTI parent structure, showing in dark gray the segments on which the CM is modeled. The slow exchange core elements are the two long strands of antiparallel β-sheet (residues 18–24 and 29–35) and the small β-bridge (residues 43–44). Cystine atoms in the three disulfide bonds (14–38, 5–55 and 30–51) are ball-and-stick models. The diagram was produced using the program MOLSCRIPT [35].

positions 14–38 of the natural protein (dark gray in Fig. 1, and Fig. 2b). Native BPTI residue numbering is retained in describing the CM which corresponds to β-sheet strands 18–24 and 29–35 and the critical 25–28 reverse turn; this hairpin is thought to be the folding initiation site in BPTI [18,19]. Since in the native protein the core peptide is flanked by the rest of the sequence, neutral end groups are incorporated into the CM; residue 14 is N-terminal β-mercaptopropionate (B in Fig. 2) and residue 38 is C-terminal cysteinamide. Isosteric replacement of C30 by α-amino-*n*-butyric acid (X in Fig. 2) retains atom homology but eliminates formation of extraneous intra- and intermolecular disulfide bonds. The CM retains the natural 14–38 disulfide cross-link to exclude from the ensemble of conformations those that are more stable because they are highly extended and flexible (of high entropy). Compared to the same module with no disulfide bond, the cross-linked ensemble is on average more collapsed and the population of the favored native-like conformations is higher [12]. Finally, to stabilize a type I β-turn, the reverse turn was modified by replacement of Ala 27 by aspartic acid [20].

In the next step toward production of a native state mimic, a series of two-module analogs were constructed, using oxime-forming ligation chemistry. Six candidates of ~50 residues each were produced with variations in cross-link position

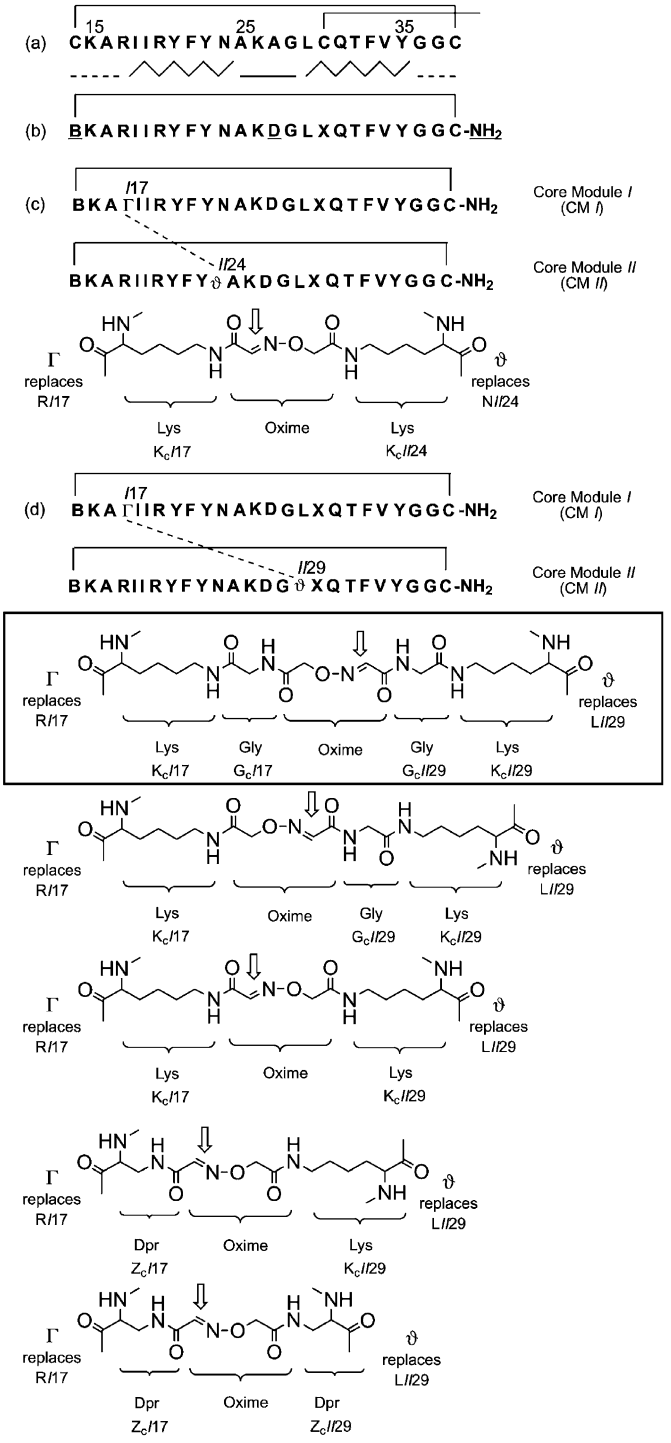
(a) CKARIIRYFYNAKAGLCQTFVYGGC
(b) BKARIIRYFYNAKDGLXQTFVYGGC-NH2
(c) BKARIIRYFYNAKDGLXQTFVYGGC-NH2
Core Module I (CM I)
BKARIIRYFYAKDGLXQTFVYGGC-NH2
Core Module II (CM II)
Γ replaces R/17
ϑ replaces N/24
Lys Kc/17
Oxime
Lys Kc/24
(d) BKARIIRYFYNAKDGLXQTFVYGGC-NH2
Core Module I (CM I)
BKARIIRYFYNAKDGϑXQTFVYGGC-NH2
Core Module II (CM II)
Γ replaces R/17
ϑ replaces L/29
Lys Kc/17
Gly Gc/17
Oxime
Gly Gc/29
Lys Kc/29
Γ replaces R/17
ϑ replaces L/29
Lys Kc/17
Oxime
Gly Gc/29
Lys Kc/29
Γ replaces R/17
ϑ replaces L/29
Lys Kc/17
Oxime
Lys Kc/29
Γ replaces R/17
ϑ replaces L/29
Dpr Zc/17
Oxime
Lys Kc/29
Γ replaces R/17
ϑ replaces L/29
Dpr Zc/17
Oxime
Dpr Zc/29

Fig. 2.

(Fig. 2c,d) and length (Fig. 2d) [13]. To screen for ordered structure in the conformational ensembles of the two-module analogs, one dimensional proton (1D ^{1}H) NMR spectra were acquired at pH 2. The most promising candidate, with the narrowest line widths and highest degree of chemical shift dispersion, was BetaCore (Fig. 2d).

3.2. CM favors native-like β-hairpin structures in rapid equilibrium with non-native β-hairpins

The CM peptide has considerably more chemical shift dispersion than a random coil with the same amino acid composition (Fig. 3b,c). The deviation of observed chemical shift (δ) of Cα protons from random coil values, expressed by the relationship $\Delta\delta H\alpha = \delta H\alpha(\text{obs}) - \delta H\alpha(\text{random coil})$ [21–23], is an indicator of secondary structure. Relative to random coil values of the same amino acid, δHα for β-sheet protons are shifted downfield while δHα for α-helix and turn protons are shifted upfield [24]. For residues 25–28, which in native BPTI are a turn, ΔδHα is negative (upfield shift); for residues 18–24 and 29–35, which in native BPTI are strands, ΔδHα is positive (downfield shift) (Fig. 4, bar graphs). Thus, CM has CαH proton chemical shift deviations similar to native BPTI both in the strands and in the turn. In addition to chemical shift data, multiple long-range NOEs observed in spectra of the CM indicate organized β-hairpin structure (Fig. 5a,b). These NOEs report subpopulations of preferred structures, in contrast to NOEs of native proteins which arise from a dominant population of closely related conformations. The CM is an equilibrium ensemble of conformations among which a major population is similar to the native-like 4:4 β-hairpin (Fig. 5a) and a minor population approximates 3:5 β-hairpins (Fig. 5b). Native-like backbone CαH-CαH NOEs have approximately twice the intensity of non-native. Likewise, for sidechain–sidechain NOEs, the native-like are more numerous and more intense than non-native. Simulated annealing calculations do not support a single family of structures but are consistent with an equilibrium between major 4:4 β-hairpin and minor 3:5 β-hairpin populations. Equilibria between 4:4 and 3:5 β-hairpins are also reported for other peptides [25–27].

3.3. Compared to the CM, modules in BetaCore are much more native-like and are organized into a new fold, a stable four-stranded β-sheet

BetaCore, constructed with a 22-atom cross-link between two CMs (Fig. 2d), folds to a dominant conformation amenable to structural characterization by multidimensional NMR methods [14]. BetaCore behaves as a monomer in sedimentation experiments, under conditions adopted for NMR studies. At low temperature and pH 3, NMR spectra of BetaCore have far more chemical shift dispersion than the CM (Fig. 3a vs. Fig. 3b). Structural analysis of assigned spectra indicates a dominant, folded conformation composed of a four-stranded β-sheet and non-identical packing of equivalent residues in the constituent modules, referred to as CM *I* and CM *II* [14].

The increased native-like behavior of BetaCore, compared to CM, is evident in their NOE patterns. The sequential backbone NOEs of BetaCore imply a strong preference of β-sheet structure; interresidue $d_{NN}(i, i+1)$ NOEs are very weak or entirely absent and $d_{\alpha N}(i, i+1)$ are intense. This is in contrast to the CM, for which sequential backbone NOEs of the same residues (18–24 and 29–35, designed to be strands) indicate conformational averaging in interresidue $d_{NN}(i, i+1)$ as well as

Fig. 2. Sequences and cross-links for CM and BetaCore. Amino acid sequences for (a) residues 14–38 in native BPTI over a line diagram of native secondary structure; (b) CM with differences from the native sequence underlined; (c) CM dimer with a cross-link between positions 17 and 24; and (d) CM dimers with a cross-link between positions 17 and 29 and varying lengths. In c and d, the dotted line between Γ and ϑ depicts the cross-link and solid lines indicate spanning disulfide bridges in CM. The chemical structure of the covalent cross-links that join the two CMs is shown below the amino acid sequences; a thick arrow indicates where the cross-link is formed. The subscript c indicates that the residue is part of the cross-link. BetaCore cross-link is indicated by a box. A '*I*' or '*II*' before the residue number indicates which of the two CM units contains the residue. B is an abbreviation for β-mercaptopropionic acid and X is for α-amino-*n*-butyric acid.

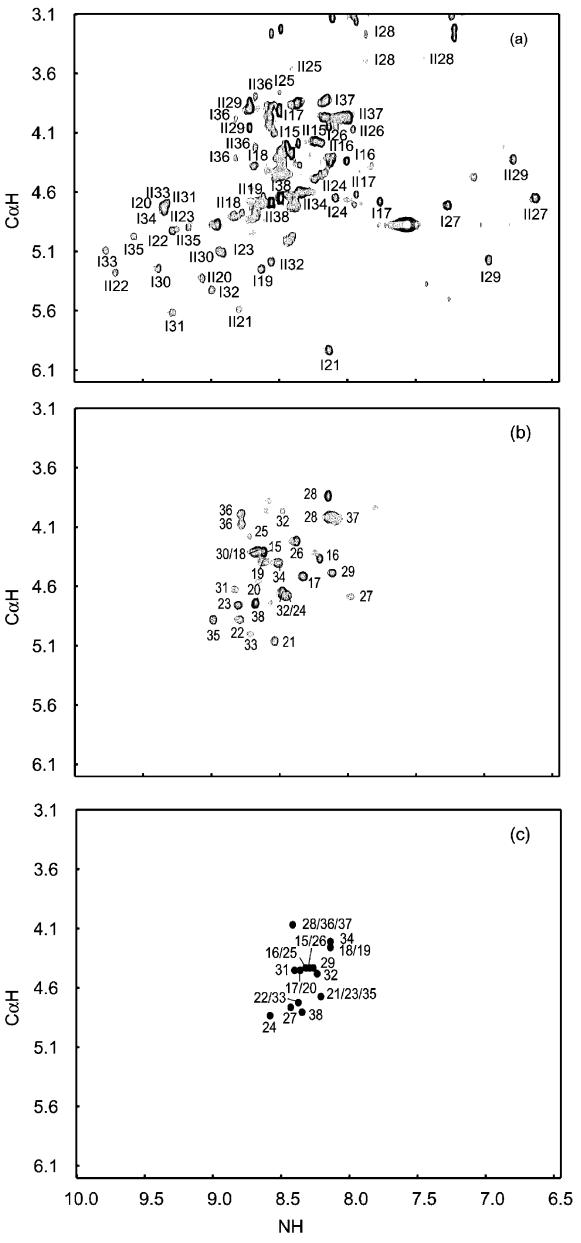

Fig. 3. The NH–CαH region of the TOCSY spectrum taken at 600 MHz. (a) BetaCore in $^{1}H_2O$: $^{2}H_2O$ (9:1 by vol) at 15 °C and pH 3; a '*I*' or '*II*' before the residue number indicates which of the two CM units contains the residue. (b) Oxidized CM in $^{1}H_2O$: $^{2}H_2O$ (9:1 by vol) at 3 °C and pH 4.5. (c) Plot of random coil NH–CαH chemical shifts [21–23] of amino acids in the CMs, at 3 °C. Panels b and c are similar to spectra published in [12].

$d_{\alpha N}(i, i+1)$ NOEs. Information about strand alignment and tertiary structure comes from long-range NOEs. Antiparallel β-sheet is indicated by long-range NOE cross-peaks between CαH*i*–CαH*j* protons, where *i* and *j* are residues that face each other in adjacent β-strands [28]. In BetaCore, CαH–CαH NOEs between residues in the same CM (Fig. 5c, thick arrows) are the same as in native BPTI [29], providing strong evidence that each CM unit samples 4:4 β-hairpin structure. BetaCore spectra do not contain any NOEs consistent with non-native β-hairpins and there are no indications of an equilibrium between 4:4 and 3:5 hairpins, as observed in the CM alone [12].

In BetaCore, there are long-range NOEs within the same module, as well as CαH–CαH NOEs between residues in CM *I* and CM *II*. The cross-link extends from a position near the N-terminal of the first strand of CM *I* to a position past the turn on the second strand of CM *II*. All inter-module NOEs (Fig. 5c, thick arrows) are between the second strand (C-terminal half) of CM *I* and the first strand (N-terminal half) of CM *II*. Importantly, no NOEs between the second strand of CM *II* and the first strand of CM *I* are observed. The data are entirely consistent with the two CMs in a four-stranded antiparallel β-sheet (Fig. 5c).

Comparison of ΔδHα values for equivalent residues in BetaCore and in a single CM unit is shown in Fig. 4. Overall shapes of the bar graphs follow patterns expected from native-like BPTI β-sheet, i.e. downfield shifts for residues 18–24 and 29–35 and upfield shifts for residues 25–28. The absolute values of the deviations from random coil shifts are much higher for residues in BetaCore (e.g. Y*I*21 $\Delta\delta H\alpha = 1.38$) than for corresponding residues in CM units (e.g. Y21 $\Delta\delta H\alpha = 0.38$), indicating considerable stabilization of structure associated with cross-linking two CMs in BetaCore. For comparison, the maximum ΔδHα for a residue in the β-sheet core of native BPTI is ~1.15 at 68 °C [30]; at this temperature BPTI is fully folded.

The fine structure of chemical shift patterns in Fig. 4 shows a notable mirror symmetry between the two CMs in BetaCore. An *i*, *i*+1 periodicity in the first strand of CM *I* and the second strand of CM *II* reflect the *i*, *i*+1 hydrogen-bonding

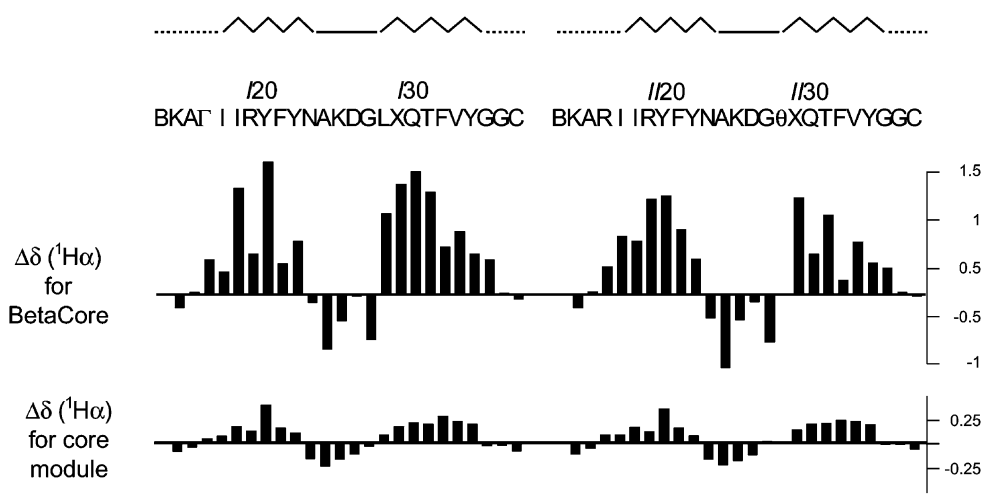

Fig. 4. The change in chemical shift, relative to random coil, of CαH protons for BetaCore and for core module alone, at pH 3 and 15 °C. The secondary structure and identity of residues in CM *I* and CM *II* are indicated above the plots of BetaCore. Alignment of plots for the core module (bottom) show chemical shift deviations for corresponding residues in CM; the same CM plot is repeated twice to show a comparison to each CM in BetaCore. Data were published in [14] as part of another figure and are reprinted here with permission.

pattern in simple β-hairpins [24]. Alternating residues involved in cross-strand contacts (e.g. residues I*I*19, Y*I*21, Y*I*23, X*II*30, T*II*32 and V*II*34 in Fig. 5c) have a more positive δHα proton shift than residues having no interactions with a neighboring strand (e.g. residues I*I*18, R*I*20, F*I*22, Q*II*31, F*II*33 and Y*II*35 in Fig. 5c). This type of two-fold *i*, *i*+1 periodicity is commonly observed in native proteins. Smoother profiles arising from loss of *i*, *i*+1 periodicity are observed for the second strand of CM *I* and the first strand of CM *II*, consistent with strands in the middle of a β-sheet where each residue is involved in contacts with a neighboring strand. The periodicity in ΔδHα plots for the outer two strands and its absence for the inner strands (Fig. 4) are entirely consistent with the four-stranded antiparallel β-sheet conformation determined as the dominant conformation of BetaCore (Fig. 5c).

The hydrogen exchange protection factors of BetaCore amides are at least an order of magnitude higher than those of corresponding residues in the CM [12] (Table 1) and about the same order of magnitude observed for a partially folded protein, e.g. $[14–36]_{Abu}$ [31]. The hydrogen exchange pattern in BetaCore is in good agreement with the four-stranded antiparallel β-sheet model calculated from NOE (Fig. 5c) and chemical shift data (Fig. 4). The most protected residues are in centrally positioned hydrogen bonds between donors and acceptors in different strands of the same module (F*I*22, F*II*22, F*I*33 and F*II*33; $k_{rc}/k_{obs}=30–75$), or in the hydrogen bond of the native-like BPTI type I β-turn (G*I*28 and G*II*28; k_{rc}/k_{obs} 100 and 190, respectively). Moderately protected residues engage in hydrogen bonds between donors and acceptors in different strands of the same module (L*I*29; $k_{rc}/k_{obs}=17$) or between donors and acceptors in strands of different modules (V*I*34 and G*I*36; $k_{rc}/k_{obs}=17–20$). Residues that exchange relatively rapidly (A*I*16, A*II*16, G_cI17, A*I*25, A*II*25, G_cII29, V*II*34, G*II*36, G*I*37 and G*II*37; $k_{rc}/k_{obs}=3–10$) include amides in loop and turn regions, the cross-link, and a solvent-exposed side of β-strand, none of which are expected to be intramolecularly hydrogen bonded.

^{15}N HSQC spectra acquired at increasing temperature between 1 and 55 °C show clear evidence of a temperature-induced global transition from one dominant folded conformation to disordered, unfolded conformations (Fig. 6). At 1 °C, the two modules in BetaCore are clearly not identical and

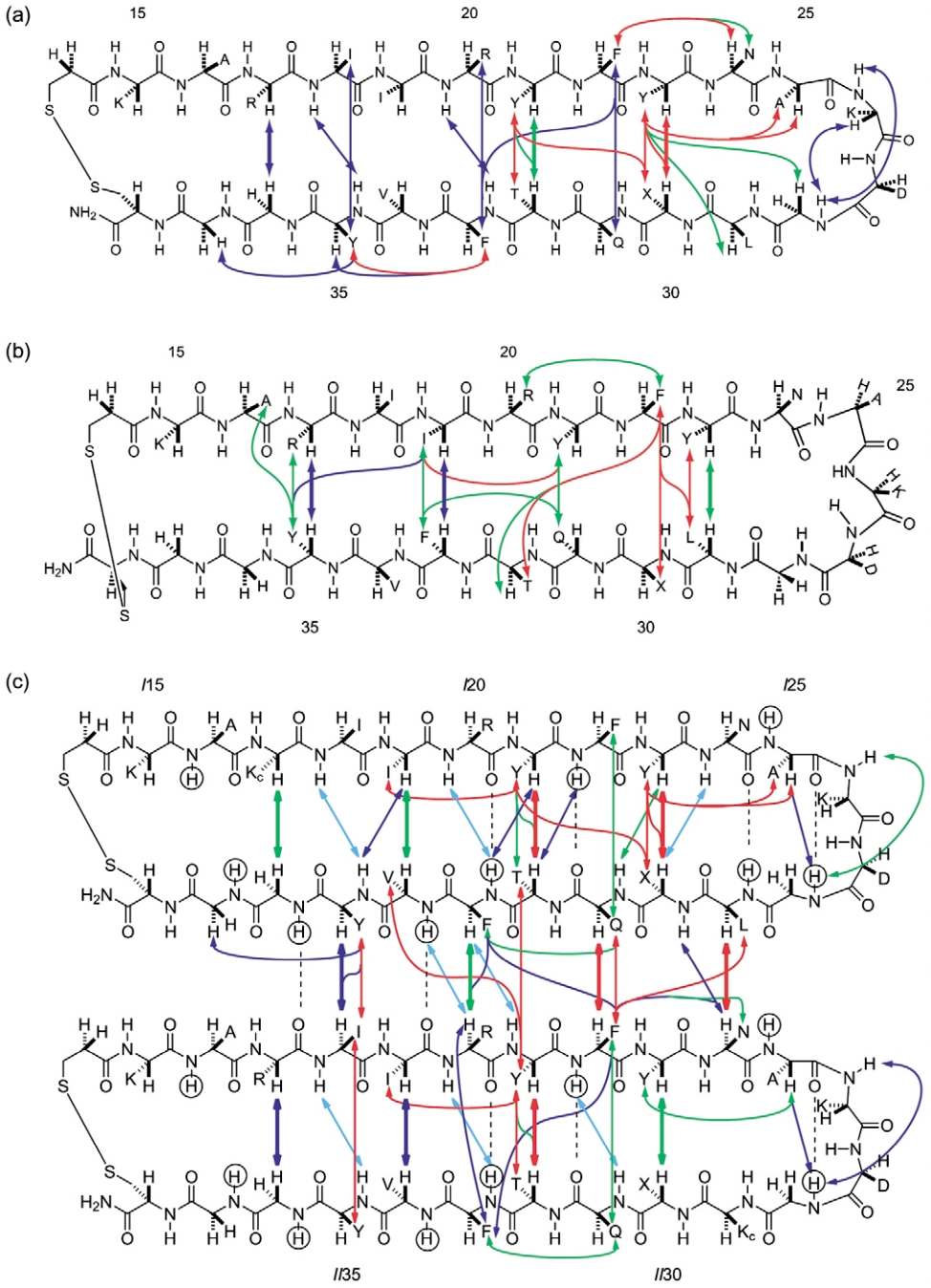

Fig. 5. Schematic representation of long-range CαH–CαH, NH–CαH and aromatic–aliphatic NOEs for (a) the CM consistent with a native 4:4 β-hairpin conformation, (b) the CM consistent with a non-native 3:5 β-hairpin, and (c) BetaCore, which is drawn to indicate the proposed four-stranded antiparallel β-sheet. Amide protons for which hydrogen exchange is evaluated (Table 1) are circled; for those with significant protection, the corresponding deduced hydrogen bonds are indicated by dashed lines. One-letter abbreviations indicate amino acid residue side chains. K_c at positions 17 of CM *I* and 29 of CM *II* are lysines in the oxime cross-link. A '*I*' or '*II*' before the residue number indicates which CM unit contains the residue. In calculated structures of BetaCore, the cross-link is primarily but not exclusively beneath the plane of the present Figure. In panels a, b and c, the CαH–CαH NOEs are shown as thicker arrows. The color of the arrow indicates NOE strength: strong (red), medium (green), weak (purple) or very weak (light blue). Panels a and b are published in [12] and panel c in [14], both are reprinted here with permission.

Table 1
Comparison of protection factors[a] for the same ^{15}N labeled residues in BetaCore at 5 °C and pH 3 and in CM at 6 °C and pH 4.15.

Residue	CM	BetaCore
A*I*16	1.0	3.5
F*I*22	3.1	57.2
A*I*25	[b]	6.7
G*I*28	3.2	102.5
L*I*29	2.7	16.7
F*I*33	3.2	61.2
V*I*34	2.2	18.0
G*I*36	1.7	19.8
G*I*37	[b]	8.3
A*II*16	1.0	4.6
F*II*22	3.1	29.1
A*II*25	[b]	9.5
G*II*28	3.2	191.5
F*II*33	3.2	73.5
V*II*34	2.2	4.4
G*II*36	1.7	4.6
G*II*37	[b]	5.8
G_c*I*17	[c]	[d]
G_c*II*29	[c]	[b]

[a] Protection factors are computed as k_{rc}/k_{obs}, where k_{rc}, the rate constant expected in a random coil model at the same pH and temperature, is calculated as in [36] using an E_a of 17 Kcal/mol.

[b] Exchange is too rapid to be measured.

[c] As the subscript 'c' indicates, this residue is in the cross-link and, therefore, is not present in CM.

[d] Residue attached to a side chain N^{ε} of Lys. A protection factor cannot be computed since this configuration is not included in the rules for estimation of k_{rc} [36]. Its exchange rate is measurable and is about the same as G*I*37.

different cross-peaks are observed for the same residue in different modules (Fig. 6a). As the temperature is increased, extensive overlap is observed for cross-peaks of the same residue in different modules, as expected for identical disordered domains (Fig. 6c). The fact that throughout the unfolding transition each resonance is reported by two peaks (e.g. Fig. 6b), one with a chemical shift well resolved and away from the random coil envelope and the other with a chemical shift in the random coil envelope, implies that BetaCore is in slow conformational exchange on the NMR time scale between folded and unfolded forms [32].

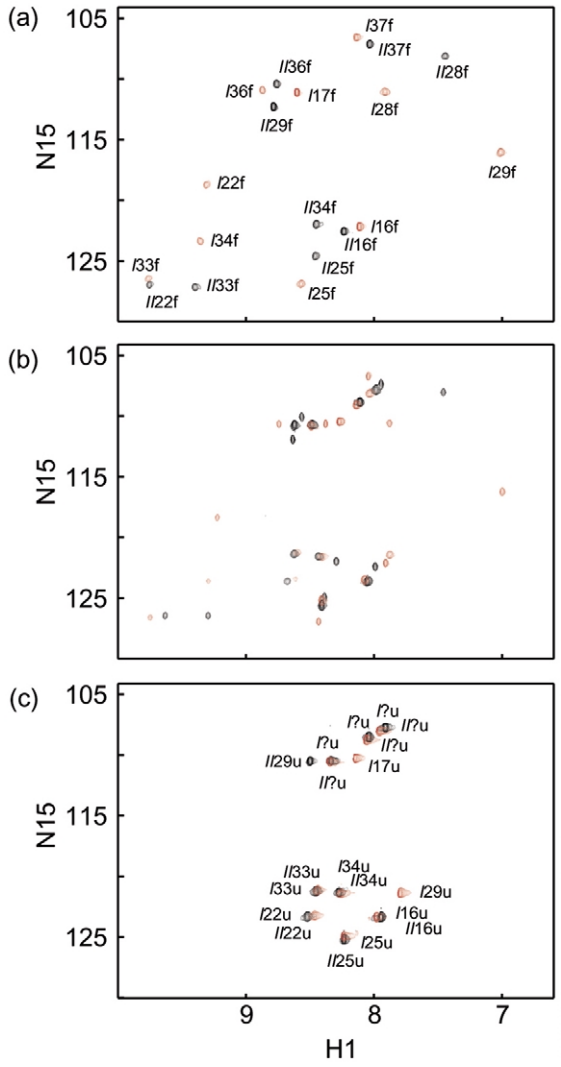

Fig. 6. BetaCore ^{15}N-^{1}H HSQC spectra taken at 800 MHz in H_2O:D_2O (9:1 by vol) at pH 3 and (a) 1 °C, (b) 35 °C and (c) 55 °C. Each frame is composed of an overlay of two spectra, one obtained with a sample in which 10 positions of CM *I* were ^{15}N-labeled (red cross-peaks) and a second obtained under identical conditions with a sample in which 9 positions of CM *II* were ^{15}N-labeled (black cross-peaks). Each cross-peak in (a) and in (c) is designated by CM (*I* or *II*), residue number (14–38) and conformation ('f' or 'u' for folded or unfolded), except for 6 glycine residues in the 'u' conformation for which the sequence position cannot be determined unambiguously. At 35 °C, both f and u cross-peaks are observed for each residue. Cross-peak positions in frame (b), observed at 35 °C, are duplicated if one overlays frames (a) and (c). Published in [14] and reprinted here with permission.

4. Discussion

Our general strategy for design of a native state mimic is to incorporate cross-links that shift the conformational equilibria of CMs toward more native-like structure. The strategy is predicated on a bias toward compact, well packed species of the native core elements on which the modules are modeled. Intramodule cross-links are not expected to organize compact structure around themselves; rather they are designed to eliminate more extended, higher entropy conformations from the ensemble and thereby increase the population of more collapsed conformations, among which the native-like are favored.

An intermodule cross-link incorporates two modules in one molecule and enhances the probability of interactions between CM *I* and CM *II*. We hypothesize that intermodule interactions that are mutually stabilizing will involve specific contacts between hydrophobic surfaces of compact conformations of individual modules. By specific, we mean that a set of contact surface interactions are, collectively, significantly more stable than other sets that are possible. In the panel of new two-module peptides with cross-links at varying positions and lengths, the presence of increased native-like structure, easily screened in 1D NMR spectra, was taken as a strong suggestion of specific intermodule interactions. These expectations are in large part fulfilled in BetaCore, in which each constituent module is far more native-like than an individual CMs and specific intermodule interactions are evident in the new fold.

BetaCore is the first designed protein which folds to a four-stranded antiparallel β-sheet conformation that is soluble in water. Its dominant conformation at low temperature is indicated by well-dispersed chemical shifts, i, $i+1$ periodicity, numerous long-range NOEs sufficient to permit calculation of low resolution NMR structures and slowed amide hydrogen isotope exchange patterns. This conformation contains features of native BPTI structure, but has very different tertiary structure. Each constituent module retains a native-like 4:4 β-hairpin, but not the distinctive twist of the native hairpin. Altogether new structural features are the contacts between antiparallel strands of different modules.

Folded BetaCore undergoes reversible, global, moderately cooperative, non-two state thermal transitions to an equilibrium ensemble of unfolded conformations [14]. Folded and unfolded ensembles interconvert slowly on the NMR time scale, indicating a significant energy barrier between them. BetaCore has properties that compare favorably to some of the more successful designed proteins in the literature. However, it is not a perfect mimic of a native state, as evidenced by its lack of very slow hydrogen exchange in the most organized and well packed elements of secondary structure, its limited number of long-range NOEs and its deviation from two-state global folding/unfolding. Nevertheless, it is an encouraging lead toward the next generation of two-module proteins designed to mimic a native state

4.1. Future directions, based on BetaCore results, for design of a native state mimic

It is fairly straightforward to design a core module from hydrogen exchange data, synthesize the module and determine whether it favors native-like structure. The creative challenge comes in the next step, the choice of a cross-link that promotes module self-association that is mutually stabilizing and specific. In these first attempts, we based the positioning of the intermodule cross-link on a dimeric form of BPTI the existence of which is surmised from hydrodynamic [33] and molecular modeling studies [34]. In the molecular model of the putative BPTI dimer, the residues at the hydrophobic monomer-monomer interface are primarily in the β-hairpins (Fig. 7a). Two core modules arranged in head-to-tail fashion retain the dimer interface of the molecular model and Arg^{17} in one module should be close to Leu^{29} in the other (Fig. 7a). However, in a similar head-to-tail arrangement that does not retain the native 'twist' of the β-sheets, Arg^{17} would be close to Asn^{24} (Fig. 7b). Following up on these two possibilities, two-module analogs were constructed with cross-links of length 16 atoms connecting 17–24 or 17–29 (Fig. 2c,d). Evaluation of 1D ^1H NMR spectra indicates that the best positioning is a link between

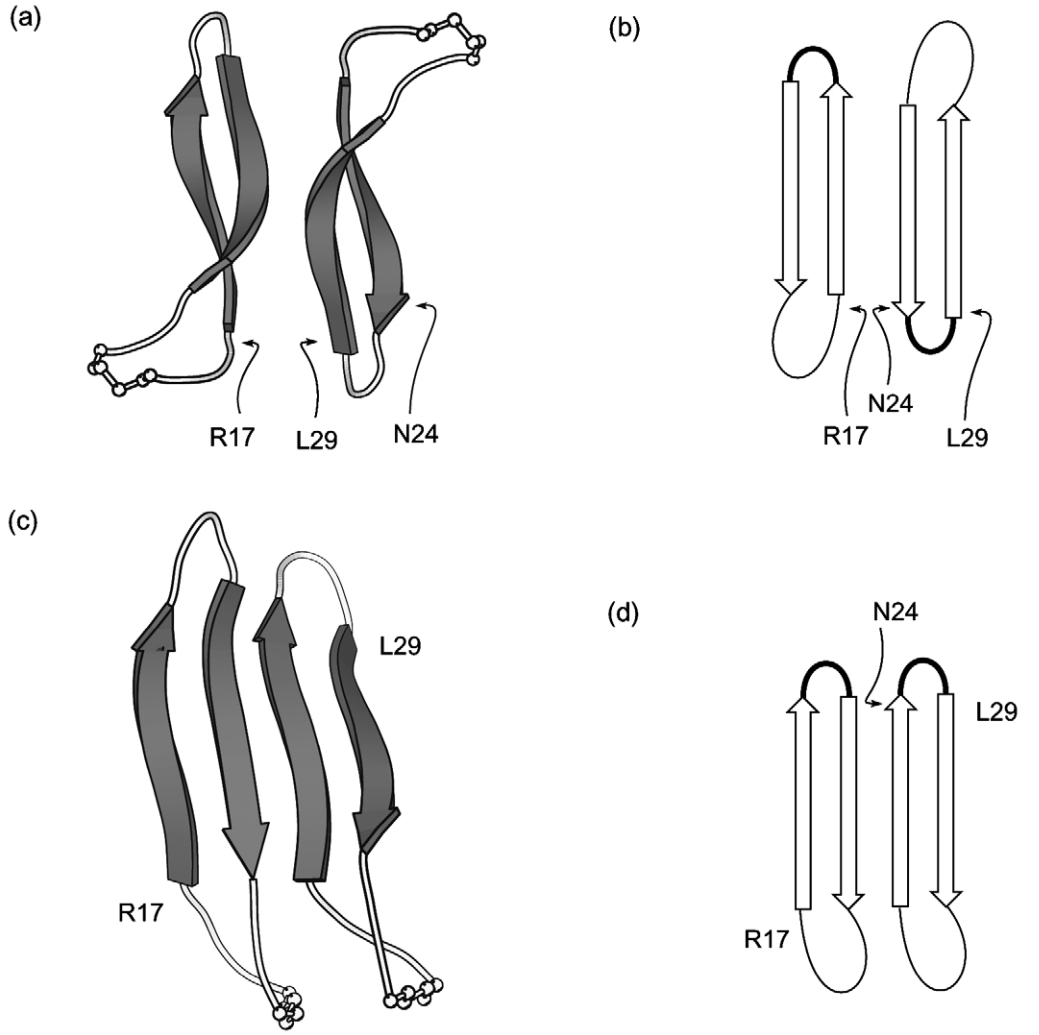

Fig. 7. Intermodule orientation in a calculated BPTI dimer and in BetaCore. (a) The head-to-tail antiparallel β-sheet model, orienting CM *I* and CM *II* as in the dimer of native BPTI proposed in [34]. Residue Arg^{17} in CM *I* (left) is nearer to Leu^{29} than to Asn^{24} in CM *II*. (b) Flattened diagram of (a). Omitting the native-like twist of β-sheet strands suggests that Arg^{17} in CM *I* may be closer to Asn^{24} in CM *II*. (c) The head-to-head antiparallel β-sheet calculated for BetaCore from NMR parameters [14]. Coordinates of an NMR structure close to the average were used for the drawing. (d) Flattened diagram of (c). Both Leu^{29} and Asn^{24} are far from Arg^{17}. Drawings (a) and (c) were generated with the program MOLSCRIPT [35]. In (b) and (d), block arrows are β-sheet strands, bold lines are β-turns and thin lines are the more flexible regions which include the disulfide cross-link.

Arg^{17} and Leu^{29}. Keeping the 17–29 cross-link position, additional two-module analogs were constructed with cross-links of length 10 to 22 atoms (Fig. 2d). The two-module analog with the longest cross-link, BetaCore, was the most promising candidate. However, after complete NMR characterization [14], it became clear that the intermodule self-association of BetaCore is not head-to-tail as in Fig. 7a,b, but head-to-head in an antiparallel β-sheet as in Fig. 7c,d.

What do we learn from the fact that the complementary associating surfaces are not the ones predicted? Besides noting the outcome as a nice example of serendipity in science, we draw two

inferences. It is probably important to use longer intermodule cross-links for the first round of variants, as they permit more alternative arrangements of intermodule contacts. In longer cross-links, the polarity (and related water solubility) is an important concern. A second inference stems from the fact that the BetaCore cross-links are located on the outside of the packed structure, whereas the links were designed to be between strands that were expected to be adjacent. Future iterations of this molecule may position a long cross-link on the outside of the hypothetical head-to-tail dimer, to test whether this permits head-to-tail association.

Further optimization may also include cross-link packing. Alternatively, the first experiments with a long cross-link may be used to identify sites where a main chain peptide linker can effectively replace the cross-link. In this system, for example, that would be simply to connect through peptide linkages the C-terminus of CM *I* and the N-terminus of CM *II*. What we learn from BetaCore iterations will be applied to systems based on different slow exchange cores to produce other native state mimics.

Acknowledgments

This work is supported by NIH grant GM51628.

References

[1] J.R. Beasley, M.H. Hecht, Protein design: the choice of de novo sequences, J. Biol. Chem. 272 (1997) 2031–2034.
[2] C.K. Smith, L. Regan, Construction and design of β-sheets, Acc. Chem. Res. 30 (1997) 153–161.
[3] B. Imperiali, J.J. Ottesen, Design strategies for the construction of independently folded polypeptide motifs, Biopolymer 47 (1998) 23–29.
[4] S.H. Gellman, Minimal model systems for β-sheet structure in proteins, Curr. Opin. Chem. Biol. 2 (1998) 717–725.
[5] R.B. Hill, D.P. Raleigh, A. Lombardi, W.F. DeGrado, De novo design of helical bundles as models for understanding protein folding and function, Acc. Chem. Res. 33 (2000) 745–754.
[6] M.L. Paz, E. Lacroix, M. Ramírez-Alvarado, L. Serrano, Computer-aided design of β-sheet peptides, J. Mol. Biol. 312 (2001) 229–246.
[7] J. Venkatraman, S.C. Shankaramma, P. Balaram, Design of folded peptides, Chem. Rev. 101 (2001) 3131–3152.
[8] R. Li, C. Woodward, The hydrogen exchange core and protein folding, Protein Sci. 8 (1999) 1571–1590.
[9] C. Woodward, Is the slow exchange core the folding core?, Trends Biochem. Sci. 18 (1993) 359–360.
[10] K.S. Kim, J. Fuchs, C. Woodward, Hydrogen exchange identifies native-state motional domains important in protein folding, Biochemistry 32 (1993) 9600–9608.
[11] C. Woodward, E. Barbar, N. Carulla, J. Battiste, G. Barany, Experimental approaches to protein folding based on the concept of a slow hydrogen exchange core, J. Mol. Graphics Modell. 19 (2001) 94–101.
[12] N. Carulla, C. Woodward, G. Barany, Synthesis and characterization of a β-hairpin peptide that represents a 'core module' of Bovine Pancreatic Trypsin Inhibitor (BPTI), Biochemistry 39 (2000) 7927–7937.
[13] N. Carulla, C. Woodward, G. Barany, Towards new designed proteins derived from Bovine Pancreatic Trypsin Inhibitor (BPTI): covalent cross-linking of two 'core modules' by oxime-forming ligation, Bioconjugate Chem. 12 (2001) 726–741.
[14] N. Carulla, C. Woodward, G. Barany, BetaCore, a designed water soluble four-stranded antiparallel β-sheet protein, Protein Sci. 11 (2002) 1539–1551.
[15] M. Piotto, V. Saudek, V. Sklenar, Gradient-tailored excitation for single-quantum NMR spectroscopy of aqueous solution, J. Biomol. NMR 2 (1992) 661–665.
[16] F. Delaglio, S. Grzesiek, G.W. Vuister, G. Zhu, J. Pfeifer, A. Bax, NMRPipe: a multidimensional spectral processing system based on UNIX pipes, J. Biomol. NMR 6 (1995) 277–293.
[17] B.A. Johnson, R.A. Blevins, NMRView: a computer program for the visualization and analysis of NMR data, J. Biomol. NMR 4 (1994) 603–614.
[18] E. Barbar, M. Hare, V. Daragan, G. Barany, C. Woodward, Dynamics of the conformational ensemble of partially folded bovine pancreatic trypsin inhibitor, Biochemistry 37 (1998) 7822–7833.
[19] R. Li, J. Battiste, C. Woodward, Native-like interactions favored in unfolded bovine pancreatic trypsin inhibitor have different roles in folding, Biochemistry 41 (2002) 2246–2253.
[20] E.G. Hutchinson, J.M. Thornton, A revised set of potentials for β-turn formation in proteins, Protein Sci. 3 (1994) 2207–2216.
[21] G. Merutka, J.H. Dyson, P.E. Wright, 'Random coil' ^{1}H chemical shifts obtained as a function of temperature and trifluoroethanol concentration for the peptide series GGXGG, J. Biomol. NMR 5 (1995) 14–24.
[22] D.S. Wishart, C.G. Bigam, A. Holm, R.S. Hodges, B.D. Sykes, ^{1}H, ^{13}C, ^{15}N random coil NMR chemical shifts of the common amino acids. I. Investigations of nearest-neighbor effects, J. Biomol. NMR 5 (1995) 67–81.
[23] N.H. Andersen, J.W. Neidigh, S.M. Harris, G.M. Lee, Z. Liu, H. Tong, Extracting information from the tem-

perature gradients of polypeptide NH chemical shifts. 1. The importance of conformational averaging, J. Am. Chem. Soc. 119 (1997) 8547–8561.
[24] K. Ösapay, D.A. Case, Analysis of proton chemical shifts in regular secondary structure of proteins, J. Biomol. NMR 4 (1994) 215–230.
[25] E. de Alba, M.A. Jimenez, M. Rico, J.L. Nieto, Conformational investigation of designed short linear peptides able to fold into β-hairpin structures in aqueous solution, Folding Des. 1 (1996) 133–144.
[26] E. de Alba, M. Rico, M.A. Jimenez, Cross-strand side-chain interactions versus turn conformation in β-hairpins, Protein Sci. 6 (1997) 2548–2560.
[27] E. de Alba, M.A. Jimenez, M. Rico, Turn residue sequence determines β-hairpin conformation in designed peptides, J. Am. Chem. Soc. 119 (1997) 175–183.
[28] K. Wüthrich, NMR of proteins and nucleic acids, J. Wiley & Sons, New York, 1986.
[29] G. Wagner, W. Braun, T.F. Havel, T. Schaumann, N. Go, K. Wüthrich, Protein structures in solution by nuclear magnetic resonance and distance geometry. The polypeptide fold of the basic pancreatic trypsin inhibitor determined using two different algorithms, DISGEO and DISMAN, J. Mol. Biol. 196 (1987) 611–639.
[30] G. Wagner, K. Wüthrich, Sequential resonance assignments in protein ^{1}H nuclear magnetic resonance spectra of basic pancreatic trypsin inhibitor, J. Mol. Biol. 155 (1982) 347–366.
[31] E. Barbar, G. Barany, C. Woodward, Dynamic structure of a highly ordered β-sheet molten globule: multiple conformations with a stable core, Biochemistry 34 (1995) 11423–11434.
[32] G.C.K. Roberts, Effects of chemical exchange on NMR spectra, in: G.C.K. Roberts (Ed.), NMR of Macromolecules: A Practical Approach, Oxford University Press, New York, 1993, pp. 153–182.
[33] W.H. Gallagher, C.K. Woodward, The concentration dependence of the diffusion coefficient for bovine pancreatic trypsin inhibitor: a dynamic light scattering study of a small protein, Biopolymers 28 (1989) 2001–2024.
[34] P. Zielenkiewicz, Y. Georgalis, W. Saenger, Self-association of bovine pancreatic trypsin inhibitor: specific or nonspecific?, Biopolymers 31 (1991) 1347–1349.
[35] P.J. Kraulis, MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures, J. Appl. Crystallogr. 24 (1991) 946–950.
[36] Y. Bai, J.S. Milne, L. Mayne, S.W. Englander, Primary structure effects on peptide group hydrogen exchange, Proteins Struct Funct Genet 17 (1993) 75–86.