Journal of Biomolecular Structure & Dynamics, ISSN 0739-1102
Volume 20, Issue Number 1, (2002)

The ^{1}H-NMR Solution Structure of the Antitryptic Core Peptide of Bowman-Birk Inhibitor Proteins: A Minimal 'Canonical Loop'

http://www.jbsdonline.com

Arnd B.E. Brauer[1,2]
Geoff Kelly[2‡]
Stephen J. Matthews[2]
Robin J. Leatherbarrow[1*]

Departments of [1]Chemistry and
[2]Biological Sciences
Imperial College of Science
Technology and Medicine
South Kensington, London
SW7 2AY, U.K.

‡Current address:
Biomedical NMR Centre
National Institute for Medical Research
The Ridgeway, Mill Hill, London
NW7 1AA, UK.

Telephone: +44 207 594 5752
Fax: +44 1342 841 939
Email: R.Leatherbarrow@ic.ac.uk

Abstract

Bowman-Birk inhibitor (BBI) proteins contain an inhibitory motif comprising a disulfide-bonded sequence that interacts with serine proteinases. Recently, a small 14-residue peptide from sunflowers (SFTI-1), which has potent anti-trypsin activity, has been found to have the same motif. However, this peptide also has an unusual head-to-tail cyclisation. To address the role of the core inhibitory sequence itself, we have solved the ^{1}H-NMR solution structure of an antitryptic 11-residue cyclic peptide that corresponds to the core reactive site loops of both SFTI-1 and Bowman-Birk inhibitor proteins. A comparison is made between the secondary chemical shifts found in this family and the canonical regions of several other inhibitors, giving some insight into relative flexibility and hydrogen bonding patterns in these inhibitors. The solution structure of the core peptide in isolation is found to retain essentially the same three-dimensional arrangement of both backbone and side chains as observed in larger antitryptic BBI and SFTI-1 fragments as well as in the complete proteins. The retention of the canonical conformation in the core peptide explains the peptide's inhibitory potency. It therefore represents a minimization of both the BBI and SFTI-1 sequences. We conclude that the core peptide is a conformationally defined, canonical scaffold, which can serve as a minimal platform for the engineering of biological activity.

Introduction

Inhibition by serine proteinase inhibitor proteins is often mediated by an exposed reactive site loop that is fixed in a characteristic "canonical" conformation, which is complementary to the proteinase active site (1, 2). Bowman-Birk inhibitors (BBIs), which are in clinical trials (3) for their anti-carcinogenic effects (reviewed in 4), are small plant proteins of typically 60 to 90 residues stabilised by seven disulfide bridges. Typically, they have a symmetrical structure of two tricyclic domains, each containing an independent reactive site loop, one of which usually inhibits trypsin (5, and references therein). An isolated tricyclic BBI domain maintains a functional reactive site loop (6) and is, therefore, amongst the smallest canonical inhibitors, comparable in size to members of the squash seed family (2). Only very recently, a significantly smaller canonical inhibitor, the sunflower trypsin inhibitor 1 (SFTI-1), was isolated from sunflower seeds (7). SFTI-1 encompasses a disulfide-linked ring of 9 amino acids, which is homologous in sequence to the antitryptic reactive site loops of BBI proteins. The other five SFTI-1 residues form a second, backbone-cyclised loop not present in BBI proteins. SFTI-1 combines a minimal size with potent inhibition of serine proteinases such as trypsin, cathepsin G, and martriptase (7, 8). SFTI-1 has been chemically synthesised by several groups (8, 9). Both NMR solution and crystal structures (7, 9) reveal a canonical inhibitor and demonstrate that the sequence homology with BBI reactive site loops results in an equivalent three-dimensional arrangement. BBI-derived synthetic peptides, even shorter than SFTI-1, have previously been demon-

strated to retain much of the inhibitory potential of the complete protein (reviewed in 5). This has lead to the proposal that the BBI reactive site loop sequence represents an independent structural motif (10). In this study we present the NMR solution structure of an 11-residue antitryptic peptide, which represents the biologically active "core structure" of both BBI and SFTI-1 reactive site loops.

Materials and Methods

The peptide of this study was generated by established solid-phase peptide synthesis methods as previously described (11). The NMR analysis was performed in aqueous solution (90% H_2O / 10% D_2O and 100% D_2O, with 3-(trimethylsilyl)-1-propane sulfonic acid as internal reference) at 298 K and pH* of 3.8. DQF-COSY (12), TOCSY (mixing time 80 ms; 13), and NOESY (mixing times of 50, 100, 200, and 300 ms; 14) and ROESY (mixing times of 50, 100, 200, and 300 ms; 15) experiments were recorded on a Bruker AMX 600 spectrometer, and processed and analysed with X-WinNMR and Aurelia software packages on Silicon Graphics work stations.

Peptide structures were calculated with the program X-PLOR (version 3.851; 16) by employing a modified version (10) of a standard simulated annealing and molecular dynamics protocol (17). The most important modification was the addition of direct refinement against $^3J_{HNH\alpha}$ coupling constants for values > 8 Hz during the final conjugate gradient minimisation. In the peptide structure calculations, direct refinement against $^3J_{HNH\alpha}$ coupling constants was not used during the simulated annealing phase in order to avoid an early bias for one of two possible ϕ values. No hydrogen bond restraints were used during the calculations. The structures were visualised and analysed with INSIGHT II (Molecular Simulations Inc.), Swiss PDB Viewer (18), WebLab (Molecular Simulations Inc.), and WHAT IF (19) programs.

Results and Discussion

Design and Trypsin Inhibition of the Core Peptide

The sequence of the peptide of this study, here referred to as the core peptide, is shown in Table I (no. 1). The sequence closely resembles those of both BBI and SFTI-1 proteins. The disulfide-linked loop sequence (P3 to P6', Schechter and Berger nomenclature; 20) is identical to the one found in BBI proteins from at least nine different sources (*Macrotyloma axillare* seeds (21), *Torresea cearensis* seeds (22), adzuki bean (23), cow pea (24), kidney bean (25), lima bean (26), mung bean (27), horse gram (28), alfalfa leaves (29)). It differs from that of SFTI-1 (no. 3) only in a single substitution at P5'. The two terminal residues, which are exocyclic extentions of the reactive site loop, were chosen as suggested by Maeder *et al*. (30): The P7' position, which is occupied by a Phe in SFTI-1, was modified to a Tyr to serve as a spectrophotometric marker; a small residue, Ser, was placed at the P4 position in order to minimize the potential for steric clashes with trypsin as indicated by molecular modelling.

The core peptide is exceptional amongst the antitryptic BBI-derived peptides, of which at least 30 variants of different length and sequences have been described in the literature, in that it combines minimal size with the highest inhibitory potency. Several inhibition constants (K_i) with bovine pancreatic trypsin have been reported for the core peptide. These range between 10 nM (11) and 100 nM (30). A detailed kinetic study of the inhibition properties of the core peptide indicates that this discrepancy is in part a result of a slow hydrolytic turnover of the peptide by trypsin. This study estimates a K_i value of 19 nM for the intact, bioactive peptide (31).

While ten structures of complete BBI proteins have been solved to date (see 10, 32), only two structures of antitryptic BBI-derived peptides have been determined.

Table I

Comparison of the sequences and antitryptic K_i values of the peptide of this study (no. 1) with BBI and SFTI-1 proteins and fragments. Inhibitor residues identified in the crystallographic analyses are highlighted in bold. The positions are labelled according to the Schechter and Berger nomenclature (20). Solid bars show positions of disulfide bonds; dotted bars indicate head-to-tail cyclisation in SFTI-1.

No.	*Position* (*P4 P3 P2 P1 P1′ P2′ P3′ P4′ P5′ P6′ P7′*)	*Name*	*AA*	*K_i in nM*
	NMR structures in isolation			
1	S C T K S I P P Q C Y	Core Peptide*	11	19
2	G R C T K S I P P I C F P D	Monocyclic SFTI-1*	14	12
3	G R C T K S I P P I C F P D	SFTI-1*	14	0.5
4	~E S P K C P E Y C F D T I~ ~S S K P C C D Q C A C T K S N P P Q C R C S D M R L~	BBI (SOY)	71	0.6 [a]
	Crystal structures in complex with bovine pancreatic trypsin			
5	**G R C T K S I P P I C F P D**	SFTI-1	14	0.1
6	Q P C C D S C R **C T K S I P P Q C** H C A N I	BBI (M.B.) Fragment*	22	120
7	M S E C P K Y C F S S E P C C D S **C R C T K S I P P Q C H C** A N I R L	BBI (M.B.) Fragment*	35	5 [b]
8	~E S P **K C P E Y C F D T I**~ ~S M **K P C C D Q C A C T K S N P P Q C R C S D M R L**~	BBI (SOY)	71	0.13

*Synthetic material; AA, number of amino acid residues; SOY, soy bean; M.B., mung bean. PDB codes and references: 1, 1gm2; 2, 1jbn (9); 3, 1 jbl (9); 4, 1bbi (33), [a] (40); 5, 1sfi (7); 6, 1sfm (64); 7, 1sbw (27), [b] (64); 8, 1d6r (65). Please note that the sequences are written from the N- to C-terminus, except for the attached third loop in no. 4, 7, and 8, which is written from the C-terminus to the N-terminus.

These relatively large synthetic fragments of mung bean BBI in complex with trypsin have been investigated by X-ray crystallography (no. 6 and 7, Table I). In both structures only those residues that constitute the reactive site loop, which includes the contact region with the enzyme surface, could be identified in the electron density maps. These residues are highlighted in bold. This finding raises the question whether the antitryptic core region retains its structural integrity in isolation. We therefore investigated the structural properties of the uncomplexed core peptide in aqueous solution by NMR and compared these with those of larger fragments and complete proteins.

NMR parameters of the fully assigned core peptide are summarised in Table II. Coupling constants, chemical shifts and line widths do not change significantly in spectra over a concentration range of 0.3 mM to 5 mM. This, in combination with narrow resonance lines throughout the spectrum, suggests that the peptide does not aggregate at NMR concentrations. Several factors indicate that this peptide is comparatively rigid in solution. These include a wide chemical shift dispersion for HN and Hα protons; large $^3J_{HNH\alpha}$ coupling constants typically >8.3Hz, diagnostic $^3J_{H\alpha H\beta}$ coupling constants (< 4 Hz or >10 Hz for all but two non-proline residues); HN temperature coefficients which suggest occurrence of hydrogen bonding; almost complete stereospecific assignment; defined side-chain conformations; and several long-range NOE contacts (Table III).

Table II

^{1}H-NMR assignment table of the core peptide in aqueous solution at pH 3.8* and 298 K. Stereospecifically assigned protons are numbered in brackets according to (66); [a] assigned in the process of structure calculation. Coupling constants were extracted from one-dimensional spectra recorded at 298 K ([b] at 285 K). χ^1-Values are to indicate the dominant low-energy rotamer (67, 68). Amide temperature coefficients ($\Delta\delta_{HN}/\Delta T$) are derived from one-dimensional spectra recorded at temperatures ranging from 285 K to 298 K.

		HN	*Hα*	*Hβ*	*Hγ*	*Hδ*	***other H***	$\Delta\delta_{HN}/\Delta T$ ***in ppb/K***	$^3J_{HNH\alpha}$ ***in Hz***	$^3J_{H\alpha H\beta}$ ***in Hz***	χ_1 ***in °***
P4	***Ser***	-	4.22	3.99 3.88				-	-	4.1 5.8	
P3	***Cys***	8.90	5.52	3.10 (2) 2.96 (3)				– 6.7	9.0[b]	11.0 4.0	– 60
P2	***Thr***	8.89	4.47	4.58	1.47			– 5.3	7.2[b]	3.5	+ 60
P1	***Lys***	8.52	4.48	2.03 1.71	1.52 1.44	1.71	ε 3.02 ζ_{NH} 7.54	– 7.0 – 6.4	7.6	10.2 4.7	
P1'	***Ser***	7.43	4.39	3.90 (2)[a] 3.80 (3)[a]				– 0.7	7.2[b]	4.1 4.1	+ 60
P2'	***Ile***	8.23	4.30	1.80	γ^1 1.47 1.07 γ^2 0.85	0.85		– 9.7	9.4	10.5	– 60
P3'	***Pro***	-	5.11	2.06 (2) 2.47 (3)	1.86 (2) 1.98 (3)	3.61 (2) 3.55 (3)		-	-	3.2 8.8	+ 26
P4'	***Pro***	-	4.26	1.85 (2) 2.44 (3)	2.09 (2) 2.11 (3)	3.70 (2) 3.83 (3)		-	-	5.0 8.8	
P5'	***Gln***	7.92	4.56	2.03	2.27		ε^{21}_{NH} 7.42 ε^{22}_{NH} 6.86	– 1.7 – 5.3 – 5.8	9.5		
P6'	***Cys***	8.68	5.21	2.97 (2) 3.06 (3)				– 8.8	9.3	11.7 3.6	– 60
P7'	***Tyr***	8.18	4.47	3.12 (2) 2.90 (3)		7.10	ε 6.79	– 4.9	8.3	4.7 10.1	– 60

In order to reflect the solvent and the temperature of biological activity assays, the NMR spectra were recorded in aqueous solution at 298 K. On the basis of these spectra, the core peptide could be fully assigned. This is in contrast to both BBI protein and SFTI-1, where the completion of the assignments required the presence of approximately 20% organic co-solvent to enhance dispersion of the chemical shifts. The co-solvent acetonitrile appeared to be necessary to suppress self-association of BBI protein (33), whereas TFE as a co-solvent and low temperature served to increase the population of the dominant conformer of SFTI-1 at the expense of minor conformers (9). The population of minor SFTI-1 conformers in aqueous solution was reported to be approximately 40%. Although we also observed minor conformers in the spectra of the core peptide, these were significantly less populated and did not interfere with the assignment process.

Backbone secondary chemical shifts of the core peptide and the corresponding

regions of the complete proteins are compared in Figure 1. The similarity of the patterns is striking and indicates a close structural similarity. The common pattern is dominated by positive deviations, which characterise extended or β-type conformations (34). The magnitude of the deviation from random coil values is marginally attenuated in the core peptide, which may reflect the absence of organic co-solvents. It may, however, also indicate some increased mobility in the smaller structure. The most pronounced increase of mobility would be expected for the points of disconnection from the complete protein. This is indeed the case for the core peptide's terminal residues, the backbone protons of which have chemical shifts close to random coil values and show no non-sequential NOEs. This, however, does not result in a corruption of the peptide's structural integrity. The pattern of sequential NOE correlations is essentially identical to those reported for the corresponding regions of SFTI-1 and BBI proteins: All non-Pro residues are correlated by strong $H\alpha_i$-HN_{i+1} NOEs in the presence of only one intense HN_i-HN_{i+1} NOE across the scissile peptide bond between P1 Lys and P1' Ser. A strong $H\alpha_i$-$H\alpha_{i+1}$ NOE between P3' Ile and P4' Pro is diagnostic of a *cis* peptide bond, while strong $H\alpha_i$-$H\delta_{i+1}$ NOEs between P4' Pro and P5' Pro define a *trans* peptide bond (35). This suggests a β-hairpin structure in which stretches of extended conformation are connected by a type VI β-turn centred on the *cis* peptide bond. This is in agreement with the backbone chemical shifts and $^3J_{HNH\alpha}$ coupling constants > 8 Hz. As indicated by the only prominent HN_i-HN_{i+1} NOE, a particular arrangement is found at the scissile peptide bond. While most $^3J_{HNH\alpha}$ coupling constants > 8 Hz reported for SFTI-1 are reproduced in the core peptide, the values for P1 Lys and P1' Ser are marginally attenuated. It cannot be directly distinguished whether this reflects a small local rearrangement or increased mobility. The high level of structural integrity of the core peptide, however, is further confirmed by the presence of the interstrand NOEs and χ^1 conformations reported for the corresponding region of SFTI-1 (9). The reduction in size from SFTI-1 to the core peptide is accompanied by a reduced number of non-sequential NOE correlations. The pattern of interstrand NOEs, however, is retained. There are only three changes compared with the interstrand NOEs explicitly reported for SFTI-1, all of which involve cysteine protons and appear to be related to the discontinuation of the β-sheet structure in the shorter core peptide. The preferred χ^1 side chain conformations of most residues of the core peptide could be determined unambiguously (Table II). These are fully consistent with the χ^1 conformations reported for SFTI-1. In addition, the pyrrolidine ring pucker of P3' Pro in the core peptide could be determined as DOWN, whereas the pyrrolidine ring of P4' Pro appears to be in rapid exchange. The co-existence of a *cis* peptide bond and DOWN ring pucker at P3' Pro is fully consistent with the finding that the majority of *cis*-Pro residues in proteins (91%) and polypeptides (70%) adopt a DOWN ring pucker (36).

Secondary Chemical Shifts of Canonical Loops

As canonical inhibitor proteins share the geometry of the P3 to P3' polypeptide backbone within narrow limits (1, 2), we investigated whether this is reflected by the backbone secondary proton chemical shifts in the more general context of canonical inhibitors. Examples from five different families were identified from the BMRB data bank and are compared with the core peptide in Figure 2. As positive deviations are generally associated with extended or β-strand conformations (34), this is broadly consistent with the canonical conformation, which is dominated by extended β-type backbone geometries (1). The HN proton of the P1' residue exhibits a prominent negative deviation of at least –0.6 ppm in all examples but one. The dominant determinants of HN chemical shifts are backbone dihedral angles and hydrogen bonds (37). It would appear from the ϕ–ψ–chemical shift hypersurface presented by Beger and Bolton (38) that the P1' backbone dihedral angles alone are insufficient to explain this negative shift. A significant influence from the small value of the ϕ dihedral angle of the preceding bulged P1 residue (9° to 50°; 1) is likely (37). It is this dihedral angle which contributes to hyperexpos-

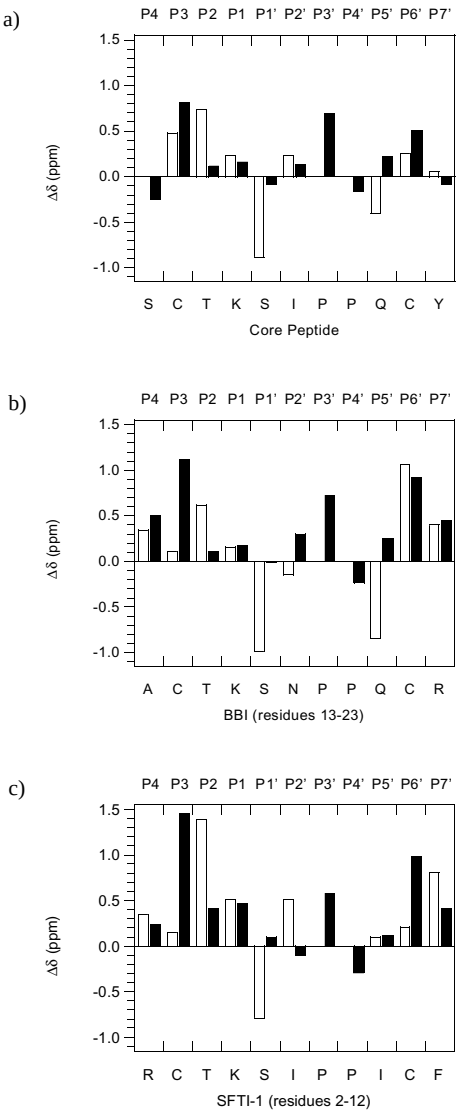

Figure 1: Secondary chemical shifts Δδ (deviation from random coil values; 37) for backbone HN (□) and Hα (■) protons.

(a) Core peptide (cyclized *via* the cysteines) in 90% H_2O / 10% D_2O, pH* 3.8, 298 K.

(b) Soybean BBI protein in aqueous solution with 18% co-solvent acetonitrile, pH 5.7, 308 K (69).

(c) SFTI-1 protein in aqueous solution with 20% co-solvent TFE, pH 4.5, 273 K (9).

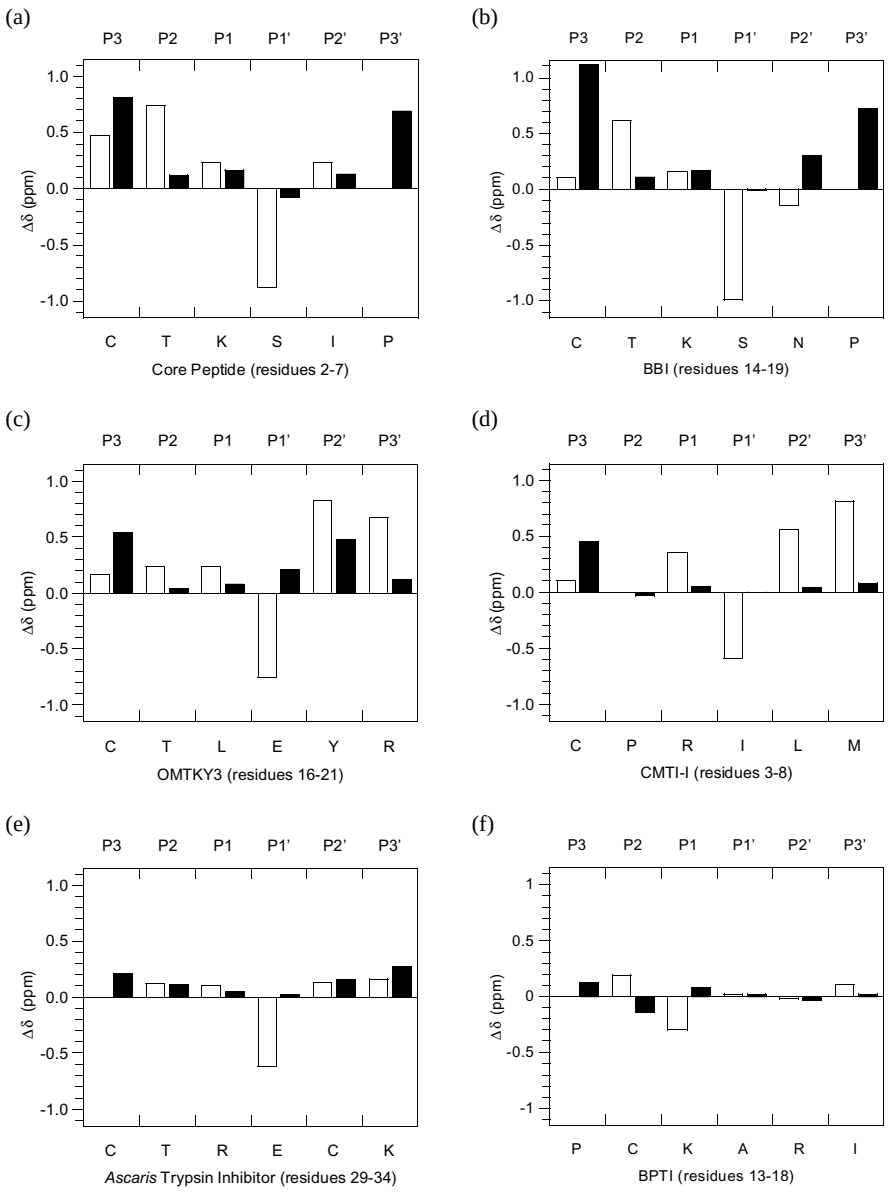

Figure 2: Secondary chemical shifts Δδ (37) for the backbone HN (□) and Hα (■) protons of the P3 to P3' residues of canonical inhibitors belonging to different families.

(a) Core peptide.

(b) BBI family: Soybean BBI (69).

(c) Kazal family: Turkey ovomucoid third domain (OMTKY3) (70).

(d) Squash seed family: *Curcurbita maxima* trypsin inhibitor I (CMTI-1) (71).

(e) Ascaris family: *Ascaris* trypsin inhibitor (43).

(f) BPTI-Kunitz family: Bovine (basic) pancreatic trypsin inhibitor (BPTI) (67).

ing the P1 residue by changing the direction of the backbone, as can be seen in Figure 3. Negatively shifted HN protons have been noted in partially formed turns, helices, and cyclic peptides and have been associated with weak intramolecular hydrogen bonding (39). Indeed, the bulge in the canonical backbone arrangement appears to be stabilised by a hydrogen bond originating from the P1' HN, as in the crystal structures of the proteins in Figure 2 (soybean BBI: 1d6r, (40); OMTKY3: 1cho, 1ppf; CMTI-1: 1ppe; BPTI: 6pti, 1bpi). In the examples in which the P1' HN exhibits a negative deviation, the hydrogen bond closes a ring of 7 atoms, in which the acceptor is the P2 carbonyl, or (in the case of OMTKY3) the side chain of the P1' Glu. This P2-P1' backbone hydrogen bond is conserved in the core peptide. Since HN chemical shifts depend on the hydrogen bond energy (34), it can be speculated that the requirements of the 7-membered rings impose a geometry on the hydrogen bond, which contributes to the observed negative shift. A different hydrogen bond arrangement in BPTI, in which the acceptor is located on another loop, is notably not associated with a negative shift. The involvement of the P1' HN in hydrogen bonding is reflected by the temperature coefficients, which are –0.7 ppb/K for the core peptide, and –2.3 ppb/K for BPTI (41). However, the P1' HN proton exchange rate was not classified as slow for BBI (33), SFTI-1 and its acyclic variant (9), OMTKY3 (42), *Ascaris* trypsin inhibitor (43), and BPTI (44). These seemingly contradictory findings provide a further example to support the

view of Baxter and Williamson (41) that surface exposed hydrogen bonded amide protons are likely to exhibit temperature coefficients more positive than –4.5 ppb/K but not necessarily slow HN proton exchange.

Another aspect highlighted by Figure 2 is that the magnitude of the deviation from random coil values varies greatly. While the magnitude is particularly large in BBI-related loops, it is not significant for most of the BPTI loop. This may indicate that the BBI-related loops are particularly rigid structures, whereas the BPTI loop is more flexible. The latter is supported by a recent study of the effect of thermal motion on the secondary chemical shifts of BPTI (45), in which the authors state that only a mixture of different conformations of this flexible loop could explain the observed chemical shifts. The different degree of loop flexibility may help to explain why even the best peptides based on the BPTI reactive site loop showed a reduction of the inhibitory potency by seven orders of magnitude (46, 47). An 11-residue bicyclic BPTI-derived peptide with a K_i value of 33 μM, for example, exhibited even in complex with trypsin in the crystallographic analysis only poorly defined electron density and large spatial differences to the corresponding region of the complete BPTI protein structure (46). This contrasts with the high inhibitory potency and structural rigidity of the BBI-derived core peptide of the same size. In this study we demonstrate that the retention of biological activity results from the retention of the three-dimensional structure of the complete proteins.

Table III

NMR restraints and statistics for the family of the 30 lowest-energy simulated annealing structures of a set of 300 structures calculated for the core peptide.

Distance restraints [a]	
intraresidue	56
sequential	21
$i + 2$	3
$i + 3$	1
$i + 6$	5
$i + 8$	3
Total	89
Stereospecific methylenes	9
Dihedral restraints [b]	
ϕ	8
χ_1	7
Total	15
$^3J_{HNH\alpha}$ restraints [c]	5
Hydrogen bond restraints	0
Violations (average r.m.s.d.)	
Distance (Å)	0.0033 ± 0.0012
Dihedral. (°)	0 ± 0
$^3J_{HNH\alpha}$ (Hz)	0.22 ± 0.11
Deviation from ideal geometry (average r.m.s.d.)	
Bond length (Å)	0.00187 ± 0.00003
Bond angles (°)	0.4914 ± 0.0026
Improper angles (°)	0.1014 ± 0.0025
Pairwise r.m.s. deviations (Å)	
all residues over backbone atoms (N, Cα, C, O)	0.61 ± 0.19
all residues over heavy atoms	1.50 ± 0.30
residues 2-10 over backbone atoms (N, Cα, C, O)	0.43 ± 0.14
residues 2-10 over heavy atoms	1.49 ± 0.28

[a] Interproton distance restraint limits were set to ± 0.35 Å for the shortest distances and relaxed up to 1 Å for longer distances and methyl groups.

[b] ϕ-Dihedral restraints were set to – 120° with limits of ± 60°. χ^1-Dihedral restraints limits were set to ± 30°, and to ± 15° for P3' Pro that adopts the DOWN pyrrolidine ring pucker.

[c] $^3J_{HNH\alpha}$ restraints were only implemented during final gradient minimisation.

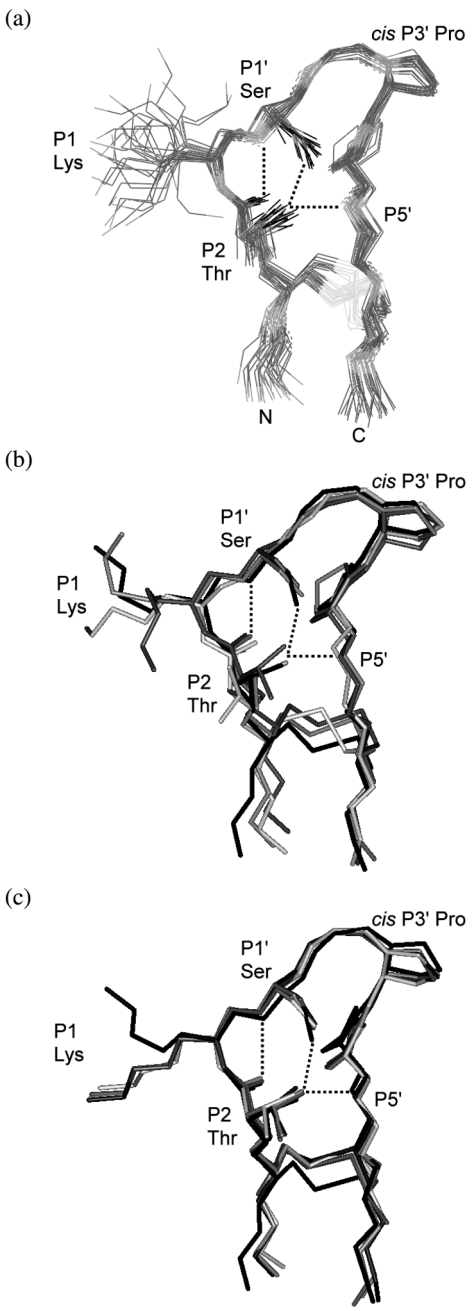

Figure 3: Structures (P4 to P7') superimposed over the backbones of the P3 to P6' residues. Only selected side chains and potential hydrogen bonds are shown for clarity.

(a) Family of 30 lowest-energy simulated annealing structures of the core peptide (O, C, N, and S atoms in increasingly lighter shades of grey).

(b) Core peptide (black) and NMR structures no. 2, 3 and 4 (increasingly lighter shades of grey).

(c) Core peptide (black) and crystal structures no. 5, 6, 7 and 8 (increasingly lighter shades of grey) in complex with trypsin.

For the structure calculations of the core peptide we found that additional direct refinement against $^3J_{HNH\alpha}$ coupling constants for values > 8 Hz during the final conjugate gradient minimisation significantly improved the convergence of the structures and also their consistency with all experimental restraints. This confirms the improvement of the accuracy reported for protein NMR structures (48). The statistics of the family of 30 lowest-energy simulated annealing structures are summarised in Table III. All residues fall into the allowed areas of the Ramachandran plot. The final structures were checked for consistency with the experimental restraints. All the NOEs that could be expected from the interproton distances in the final structures were present in the spectra within the detection limit.

The family of 30 lowest-energy structures is shown in Figure 3a. The core peptide adopts a well-defined β-hairpin structure with a slight right-handed twist and a type VIb β turn centred around the *cis* peptide bond at the P3' Pro. The terminal residues are the least well-ordered region of the molecule, which is consistent with their secondary chemical shifts. In the family of structures both chiralities of the disulfide bridge are populated. This is possibly a consequence of spectral overlap between the Hβ resonances of the two Cys residues, which limited the unambiguous identification of NOE contacts between these two residues to a single, but particularly strong Hα - Hα NOE. This short Hα - Hα distance and the observed χ^1 rotamers of −60° in both Cys residues are consistent with the short right-handed hook conformation (49, 50) observed in the BBI protein crystal structures. Moreover, the population of this native-like disulfide bridge conformation increases sharply in families of structures calculated using tighter limits on the NMR-derived restraints. One strand of the hairpin has a bulge which hyperexposes the P1 Lys. This is an important characteristic of the canonical conformation of small serine proteinase inhibitor proteins (1), which the core peptide adopts between the P3 and P2' residues. The P3' *cis*-Pro causes, as in all BBI protein structures, a deviation from the canonical conformation at that position. Differences at that position from other families of canonical inhibitors are also evident in the secondary chemical shifts shown in Figure 2.

The covalently closed backbone of the core peptide shows only small root-mean-square deviations (r.m.s.d.) ranging between 0.4 Å to 0.6 Å with the structures of complete SFTI-1 and BBI proteins as well as larger protein fragments (Table IV). Figure 3b and 3c show that the retention of the native structure extends to important side chains, two of which are involved in a cross-strand hydrogen bond network. At the central position is the side-chain hydroxyl of P2 Thr. This is the hydrogen bond acceptor with the most favourable geometry for the HN of P5' Gln, which was identified as a donor on the basis of its temperature coefficient of –1.7 ppb/K. Another hydrogen bond of optimal geometry (51) is possible between the side chains of P2 Thr and P1' Ser. This bifurcated hydrogen bond at the P2 Thr side chain is present in all available BBI and SFTI-1 protein crystal structures. In addi-

Table IV

Pairwise r.m.s.d. values in Å over the backbone of the disulfide-linked reactive site loop (P3 to P6'). The numbering of the structures is as in Table I.

		NMR structures				*Crystal structures*			
No.		***1***	***2***	***3***	***4***	***5***	***6***	***7***	***8***
	PDB code	1gm2	1jbn	1jbl	1bbi	1sfi	1smf	1sbw	1d6r
1	1gm2	-	0.58	0.41	0.60	0.44	0.48	0.42	0.59
2	1jbn	0.58	-	0.34	0.80	0.45	0.45	0.44	0.49
3	1jbl	0.41	0.34	-	0.67	0.34	0.37	0.33	0.41
4	1bbi	0.60	0.80	0.67	-	0.73	0.78	0.75	0.86
5	1sfi	0.44	0.45	0.34	0.73	-	0.22	0.15	0.28
6	1sfm	0.48	0.45	0.37	0.78	0.22	-	0.19	0.28
7	1sbw	0.42	0.44	0.33	0.75	0.15	0.19	-	0.30
8	1d6r	0.59	0.49	0.41	0.86	0.28	0.28	0.30	-

tion, two native, regular β-sheet type hydrogen bonds are geometrically possible on either side of the disulfide bridge. However, the intermediate temperature coefficients of about –5 ppb/K of the potential donors, HN of P2 Thr and P7' Tyr, may indicate that these hydrogen bonds are of a more transient nature than they are in the complete proteins. This would be consistent with the increased flexibility at the termini resulting from the separation of the rest of the protein. The same effect of weakened hydrogen bonds in the terminal region ascribed to changed dynamics but without major structural rearrangements has been observed in SFTI-1 upon the opening of the smaller loop. This modification, the creation of free exocyclic termini, is analogous to the truncation of the core peptide, and results in almost the same increase of the K_i value by two orders of magnitude. This suggests a link between the internal dynamics of the inhibitor and its biological activity.

Sequence-Inherent Stability of the Core Peptide

The core peptide is currently the smallest member in the series of structures solved for antitryptic BBI fragments (Table I). Although its size is minimal, it equally retains the three-dimensional structure observed for the covalently closed reactive site loop in complete native proteins (Table IV and Figure 3). The core peptide is, therefore, an independent structural β-hairpin motif and presents itself as a context-free minimal model system for the detailed analysis of the nature of this sequence-inherent stability. Upon the reduction of the disulfide bridge, no significant inhibition was detected (11), highlighting a critical role as a conformational restraint. The disulfide bridge adopts a conformation which is typically found to bridge antiparallel β strands (50), suggesting that strand alignment is its primary function. This is confirmed by the finding that this disulfide bridge is dispensable in the context of the complete BBI protein (52), presumably because the β-strands are sufficiently aligned by the extended β-sheet structure. The aligning function of the disulfide bridge appears to be supported by regular main-chain cross-strand hydrogen bonds. Additional irregular hydrogen bonds, which are shown in Figure 3 and were discussed above, appear to be instrumental in shaping and stabilising the central segment of the canonical conformation (53). The role of the two Pro residues in the core peptide has been analysed by sequential Ala substitutions (31). A P3' Pro in the *cis* conformation was found to be essential for the retention of a native-like biologically active three-dimensional structure. Substitution of the P4' Pro resulted in increased *cis-trans* isomerisation at the preceding P3' Pro. Thus the restraining effect of the P3'-P4' Pro-Pro motif is instrumental in stabilising the native *cis* conformation at the P3' position.

Role of Exocyclic Extensions

The sequence-inherent stability of the covalently closed reactive site loop raises questions about the role of the exocyclic extensions in the parent proteins. The core peptide described in the present study has a single amino acid extension at both N- and C-termini. Only two BBI-derived loops without any extension have been structurally investigated. These were targeted at chymotrypsin, though only one of these showed any inhibition, which was of a moderate level (55). The NMR analyses in DMSO revealed conformational heterogeneity and were hampered by a lack of non-sequential NOEs and defined side-chain conformations. However, NMR parameters such as chemical shifts, $^3J_{HNH\alpha}$ coupling constants, and temperature coefficients of the moderately active peptide bear a significant element of resemblance to those of the core peptide. This indicates that two exocylic residues in a paired arrangement, as in the core peptide, provide substantial additional structural stabilisation. This stabilisation may result from electrostatic interactions between the termini and/or from a favourable extension of the β-sheet structure. Both effects have been shown to contribute to the stability of linear peptides that highly populated β-hairpin conformations (56-58). Additionally, in the complete proteins these extensions may serve other purposes. One obvious role is to provide the scope for

additional interactions with the enzyme, which may be aimed not only at tight binding but also at specificity. The P4 position of the core peptide, for example, was found useful in generating potent inhibitors of different specificities (59, 60). Larger extensions may also be required to achieve the inhibition mechanism usually associated with the canonical inhibitor proteins, the so-called standard mechanism (1, 2). The standard mechanism describes a dynamic equilibrium in which the scissile bond of the inhibitor is both cleaved and re-synthesised by the cognate proteinase. The NMR parameters of the core peptide show significant changes upon cleavage of the scissile peptide bond (31). Although intermediate $^3J_{HNH\alpha}$ coupling constants indicate substantially increased flexibility, the HN resonances are still well-dispersed, and the secondary chemical shift pattern of the cleaved core peptide is still dominated by positive deviations. This suggests that the cleaved core peptide retains structural elements of extended or β-type conformations. Further extensions, possibly as small as the non-binding second SFTI-1 loop, may restrain the dynamics sufficiently to promote adherence to the standard mechanism.

Conclusions

The structure of the core peptide confirms our proposal that the reactive site loop sequence of BBI and SFTI-1 proteins is an independent structural β-hairpin motif. The retention of the canonical conformation of serine proteinase inhibitor proteins is identified as the basis of the potent biological activity. The core peptide, therefore, presents itself as a mini-protein and thus as a minimal, structurally defined "molecular scaffold" (for a review of other molecular scaffolds see 61) for the engineering of biological activity. For example, the minimal size of this scaffold was recently demonstrated to be essential for the successful inhibition of human mast cell tryptase, a tetrameric serine proteinase that is not inhibited by the complete BBI protein or macromolecular endogenous inhibitors (62). The core peptide's high level of integrity as a β hairpin also makes it a suitable model system for structural studies (63).

Co-ordinates

The molecular co-ordinates have been deposited with the RCSB Protein Data Bank (code 1gm2).

Acknowledgements

The authors wish to thank Dr. H. Toms and P. Haycock of the University of London 600 MHz NMR facility at Queen Mary and Westfield College for their expert advice and assistance, and Dr. J. McBride and Dr. R. Cooke for helpful discussions. We gratefully acknowledge the financial support of GlaxoWellcome and the BBSRC.

References and Footnotes

1. W. Bode and R. Huber, *Eur. J. Biochem. 204*, 433-451 (1992).
2. M. Laskowski and M. A. Qasim, *Biochim. Biophys. Acta 1477*, 324-337 (2000).
3. S. B. Malkowicz, W. G. McKenna, D. J. Vaughn, X. S. Wan, K. J. Propert, K. Rockwell, S. H. Marks, A. J. Wein, and A. R. Kennedy, *Prostate 48*, 16-28. (2001).
4. A. R. Kennedy, *Am. J. Clin. Nutr. 68*, S1406-S1412 (1998).
5. J. D. McBride and R. J. Leatherbarrow, *Curr. Med. Chem. 8*, 909-917. (2001).
6. S. Odani and T. Ikenaka, *J. Biochem. 83*, 747-753 (1978).
7. S. Luckett, R. S. Garcia, J. J. Barker, A. V. Konarev, P. R. Shewry, A. R. Clarke, and R. L. Brady, *J. Mol. Biol. 290*, 525-533 (1999).
8. Y. Long, S. Lee, C. Lin, I. J. Enyedy, S. Wang, P. Li, R. B. Dickson, and P. P. Roller, *Bioorg. Med. Chem. Lett. 11*, 2515-2519. (2001).
9. M. L. Korsinczky, H. J. Schirra, K. J. Rosengren, J. West, B. A. Condie, L. Otvos, M. A. Anderson, and D. J. Craik, *J. Mol. Biol. 311*, 579-591. (2001).
10. A. B. E. Brauer, G. Kelly, J. D. McBride, R. M. Cooke, S. J. Matthews, and R. J. Leatherbarrow, *J. Mol. Biol. 306*, 799-807 (2001).

11. G. J. Domingo, R. J. Leatherbarrow, N. Freeman, S. Patel, and M. Weir, *Int. J. Pep. Prot. Res. 46*, 79-87 (1995).
12. A. E. Derome and M. P. Williamson, *J. Mag. Res. 88*, 177-185 (1990).
13. A. Bax and D. G. Davis, *J. Mag. Res. 65*, 355-360 (1985).
14. J. Jeener, B. H. Meier, P. Bachman, and R. R. Ernst, *J. Chem. Phys. 71*, 4546-4553 (1979).
15. C. Griesinger and R. R. Ernst, *J. Magn. Reson. 75*, 261-271 (1987).
16. A. T. Brünger, *X-PLOR Version 3.1: A System for X-ray Crystallography and NMR*, Yale University Press, New Haven (1993).
17. M. Hofmann, D. Gondol, G. Bovermann, and M. Nilges, *Eur. J. Biochem. 186*, 95-103 (1989).
18. N. Guex and M. C. Peitsch, *Electrophoresis 18*, 2714-2723 (1997).
19. G. Vriend, *J. Mol. Graph. 8*, 52-56 (1990).
20. I. Schechter and A. Berger, *Biochem. Biophys. Res. Comm. 27*, 157-162 (1967).
21. F. J. Joubert, H. Kruger, G. S. Townshend, and D. P. Botes, *Eur. J. Biochem. 97*, 85-91. (1979).
22. A. S. Tanaka, M. U. Sampaio, S. Marangoni, B. de Oliveira, J. C. Novello, M. L. Oliva, E. Fink, and C. A. Sampaio, *Biol. Chem. 378*, 273-281 (1997).
23. T. Kiyohara, K. Yokota, Y. Masaki, O. Matsui, T. Iwasaki, and M. Yoshikawa, *J. Biochem. 90*, 721-728. (1981).
24. L. Morhy and M. M. Ventura, *An. Acad. Bras. Cienc. 59*, 71-81 (1987).
25. A. Funk, J. K. Weder, and H. D. Belitz, *Z. Lebensm. Unters. Forsch. 196*, 343-350. (1993).
26. F. C. Stevens, S. Wuerz, and J. Krahn, *Proteinase inhibitors (Bayer-Symp. V)*, H. Fritz et al., Eds., Springer-Verlag, Berlin (1974).
27. Y. S. Zhu, Q. C. Huang, and C. W. Chi, *J. Biomol. Struct. Dynam. 16*, 1219-1224 (1999).
28. B. Prakash, S. Selvaraj, M. R. Murthy, Y. N. Sreerama, D. R. Rao, and L. R. Gowda, *J. Mol. Evol. 42*, 560-569 (1996).
29. W. E. Brown, K. Takio, K. Titani, and C. A. Ryan, *24*, 2105-2108. (1985).
30. D. L. Maeder, M. Sunde, and D. P. Botes, *Int. J. Pept. Prot. Res. 40*, 97-102 (1992).
31. A. B. E. Brauer, G. J. Domingo, R. M. Cooke, S. J. Matthews, and R. J. Leatherbarrow, Biochemistry (in the press).
32. H. M. Krishna Murthy, K. Judge, L. DeLucas, and R. Padmanabhan, *J. Mol. Biol. 301*, 759-767 (2000).
33. M. H. Werner and D. E. Wemmer, *Biochemistry 31*, 999-1010 (1992).
34. D. S. Wishart, B. D. Sykes, and F. M. Richards, *J. Mol. Biol. 222*, 311-333 (1991).
35. K. Wüthrich, *NMR of Proteins and Nucleic Acids*, John Wiley & Sons, New York (1986).
36. E. J. Milner-White, L. H. Bell, and P. H. Maccallum, *J. Mol. Biol. 228*, 725-734 (1992).
37. D. S. Wishart and D. A. Case, *Methods. Enzymol. 338*, 3-34 (2001).
38. R. D. Beger and P. H. Bolton, *J. Biomol. NMR 10*, 129-142. (1997).
39. N. H. Andersen, J. W. Neidigh, S. M. Harris, G. M. Lee, Z. H. Liu, and H. Tong, *J. Am. Chem. Soc. 119*, 8547-8561 (1997).
40. R. H. Voss, U. Ermler, L. O. Essen, G. Wenzl, Y. M. Kim, and P. Flecker, *Eur. J. Biochem. 242*, 122-131 (1996).
41. N. J. Baxter and M. P. Williamson, *J. Biomol. NMR. 9*, 359-369. (1997).
42. C. B. Arrington and A. D. Robertson, *296*, 1307-1317. (2000).
43. A. M. Gronenborn, M. Nilges, R. J. Peanasky, and G. M. Clore, *Biochemistry 29*, 183-189. (1990).
44. P. E. Hansen, W. Zhang, C. Lauritzen, S. Bjorn, L. C. Petersen, K. Norris, O. H. Olsen, and C. Betzel, *Biochemistry 37*, 3645-3653. (1998).
45. B. Busetta, P. Picard, and G. Precigoux, *J. Pept. Sci. 7*, 121-127. (2001).
46. R. Kasher, D. A. Oren, Y. Barda, and C. Gilon, *J. Mol. Biol. 292*, 421-429 (1999).
47. J. P. Kitchell and D. F. Dykes, *Biochim. Biophys. Acta 701*, 149-152 (1982).
48. D. S. Garrett, J. Kuszewski, T. J. Hancock, P. J. Lodi, G. W. Vuister, A. M. Gronenborn, and G. M. Clore, *J. Magn. Res. Series B 104*, 99-103 (1994).
49. J. Richardson, *Adv. Prot. Chem. 34*, 167-339 (1981).
50. N. Srinivasan, R. Sowdhamini, C. Ramakrishnan, and P. Balaram, *Int. J. Pep. Prot. Res. 36*, 147-155 (1990).
51. I. K. McDonald and J. M. Thornton, *J. Mol. Biol. 238*, 777-793 (1994).
52 S. Phillip, Y. M. Kim, I. Durr, G. Wenzl, M. Vogt, and P. Flecker, *Eur. J. Biochem. 251*, 854-862 (1998).
53. J. D. McBride, A. B. E. Brauer, M. Nievo, and R. J. Leatherbarrow, *J. Mol. Biol. 282*, 447-457 (1998).
54. A. G. Cochran, R. T. Tong, M. A. Starovasnik, E. J. Park, R. S. McDowell, J. E. Theaker, and N. J. Skelton, *J. Am. Chem. Soc. 123*, 625-632 (2001).
55. V. Pavone, C. Isernia, M. Saviano, L. Falcigno, A. Lombardi, L. Paolillo, C. Pedone, S. Buoen, H. M. Naess, H. Revheim, and J. A. Eriksen, *J. Chem. Soc. - Perkin Trans. 2*, 1047-1053 (1994).
56. S. R. Griffiths-Jones, A. J. Maynard, and M. S. Searle, *J. Mol. Biol. 292*, 1051-1069 (1999).
57. E. de Alba, F. J. Blanco, M. A. Jimenez, M. Rico, and J. L. Nieto, *Eur. J. Biochem. 233*, 283-292 (1995).

58. H. E. Stanger, F. A. Syud, J. F. Espinosa, I. Giriat, T. Muir, and S. H. Gellman, *Proc. Natl. Acad. Sci. USA 98*, 12015-12020 (2001).
59. J. D. McBride, H. N. M. Freeman, and R. J. Leatherbarrow, *Eur. J. Biochem. 266*, 403-412 (1999).
60. J. D. McBride, H. N. M. Freeman, and R. J. Leatherbarrow, *J. Pept. Sci. 6*, 446-452 (2000).
61. A. Skerra, *J. Mol. Recognit. 13*, 167-187 (2000).
62. D. Scarpi, J. D. McBride, and R. J. Leatherbarrow, *J. Pept. Res.* 59, 90-93 (2002).
63. D. Scarpi, E. G. Occhiato, A. Trabocchi, R. J. Leatherbarrow, A. B. E. Brauer, M. Nievo, and A. Guarna, *Bioorg. Med. Chem. 9*, 1625-1632. (2001).
64. Y. L. Li, Q. C. Huang, S. W. Zhang, S. P. Liu, C. W. Chi, and Y. Q. Tang, *J. Biochem. 116*, 18-25 (1994).
65. J. Koepke, U. Ermler, E. Warkentin, G. Wenzl, and P. Flecker, *J. Mol. Biol. 298*, 477-491 (2000).
66. J. L. Markley, A. Bax, Y. Arata, C. W. Hilbers, R. Kaptein, B. D. Sykes, P. E. Wright, and K. Wüthrich, *J. Mol. Biol. 280*, 933-952 (1998).
67. G. Wagner, W. Braun, T. F. Havel, T. Schaumann, N. Go, and K. Wüthrich, *J. Mol. Biol. 196*, 611-639 (1987).
68. M. G. Cai, Y. Huang, J. H. Liu, and R. Krishnamoorthi, *J. Biomol. NMR 6*, 123-128 (1995).
69. M. H. Werner and D. E. Wemmer, *Biochemistry 30*, 3356-3364 (1991).
70. S. Choe, Z. Dzakula, E. S. Kuloglu, and J. L. Markley, *J. Biomol. NMR 12*, 193-195. (1998).
71. T. A. Holak, D. Gondol, J. Otlewski, and T. Wilusz, *J. Mol. Biol. 210*, 635-648 (1989).

Date Received: January 24, 2002

Communicated by the Editor Ramaswamy H Sarma