Structure and multiple conformations of the Kunitz-type domain from human type VI collagen α3(VI) chain in solution

ARTICLE *in* STRUCTURE · MARCH 1996

Impact Factor: 5.62 · DOI: 10.1016/S0969-2126(96)00022-6 · Source: PubMed

Structure and multiple conformations of the Kunitz-type domain from human type VI collagen α3(VI) chain in solution

Markus Zweckstetter[1], Michael Czisch[1], Ulrike Mayer[1], Mon-Li Chu[2], Wolfgang Zinth[3], Rupert Timpl[1] and Tad A Holak[1*]

Background: The Kunitz-type inhibitor motif is found at the C terminus of the human collagen α3(VI) chain. This 76-residue module (domain C5) was prepared in recombinant form and showed high stability against proteases; however, it lacked any inhibitory activity against trypsin, thrombin, kallikrein and several other proteases. We have undertaken the determination of the three-dimensional (3D) structure of domain C5 in solution, by nuclear magnetic resonance (NMR), in order to establish the structural basis for the properties of this protein.

Results: The 7 N-terminal and 12 C-terminal residues of domain C5 are disordered in the solution structure. The 55-residue core, which shows high homology to bovine pancreatic trypsin inhibitor, retains the characteristic fold of all members of the Kunitz-type inhibitor family. 24 residues of this main structural body show more than one resonance, symptomatic of multiple conformations slowly exchanging on the NMR time scale. In addition, significant proton chemical exchange line broadening is observed for residues in the vicinity of the disulfide bridge between residues 20 and 44: this indicates interconversion, on the micro- to millisecond time scale, between multiple conformations.

Conclusion: The NMR study demonstrates that domain C5 is a highly dynamic molecule at temperatures studied (between 10 and 30°C). Indeed, some 44% of the main body structure of C5 showed multiple conformations. The existence of multiple conformations was not necessarily expected in view of the conformational constraints imposed by the 3D structure of proteins as rigid as C5; it should therefore be considered in the interpretation of its structural and dynamical properties. The accessibility of the inhibitory binding loop (Gly18 [P4] to Leu25 [P4′]) should be relatively unaffected by this conformational exchange and thus would not explain the unusual specificity of C5. Most serine proteinase inhibitors that, like C5, have an arginine at the P1 position inhibit trypsin; the lack of trypsin inhibition of C5 must therefore arise from the amino-acid side-chain composition of the adjoining positions in the binding loop.

Addresses: [1]Max-Planck-Institut für Biochemie, D-82152 Martinsried, Germany, [2]Department of Biochemistry and Molecular Biology and Department of Dermatology, Thomas Jefferson University, Philadelphia, USA, [3]Fakultät für Physik, Ludwig-Maximilians-Universität, Institut für Medizinische Optik, D-80797 München, Germany.

*Corresponding author.

Key words: collagen, Kunitz-type domain, NMR, tertiary structure

Received: **1 Nov 1995**
Revisions requested: **20 Nov 1995**
Revisions received: **20 Dec 1995**
Accepted: **5 Jan 1996**

Structure 15 February 1996, **4**:195–209

Introduction

Collagen α1 (VII) and α3 (VI) chains were shown to possess a sequence of about 60-amino-acid residues that is highly similar to a sequence found in serine protease inhibitors of the Kunitz type [1]. The corresponding module of α3 (VI) chain (domain C5) was recently prepared in recombinant form but it was shown to lack inhibitory activity for serine proteases [2]. This module is also found in other large proteins, for example, a human Alzheimer amyloid precursor protein [3–5]. The amino-acid sequence of these modules shows 42–50% identity to bovine pancreatic trypsin inhibitor (BPTI), a representative member of the Kunitz-type serine protease inhibitors. BPTI is not only among the best characterized of all Kunitz-type inhibitors but is one of the most thoroughly investigated of all proteins. Three high-resolution crystal structures of BPTI are available [6–8], one with neutron diffraction data [9], and a high quality structure was determined by NMR in solution [10]. Extensive structural studies have been carried out on other members of the Kunitz family, including the X-ray structures of α-dendrotoxin [11], Alzheimer's protein precursor inhibitor [12] and the smaller subunit of β-bungarotoxin [13], as well as NMR structures of dendrotoxin K [14] and dendrotoxin I [15]. Characteristic features of the structures are the presence of six cysteine residues, two antiparallel β strands and a short α helix; an exposed binding loop projects away from the supporting scaffold. All six cysteines are oxidized and form disulfide bridges, which contribute to the overall stability of the protein. The stability against denaturation by heat or chemical denaturants is very high; the melting temperature for

BPTI is 95°C [16,17]. The X-ray diffraction studies of Kunitz-type domains seemed to highlight this feature with all modules having a compact shape and a rigid structure with a well defined hydrophobic core; however, even the earliest NMR studies indicated increased internal mobility, at least in a part of the BPTI molecule [10,18–20]. Recent NMR studies showed that BPTI in solution exists in two exchanging conformational isomers that most probably arise because of the different chirality of disulfide bridge Cys14–Cys38 [21].

In this paper, we describe the three-dimensional (3D) structure of the Kunitz-type domain of human type VI collagen α3(VI) chain (domain C5) in solution, determined by the 2D NMR technique. We demonstrate the presence of the exchanging multiple conformations in domain C5, which occurs to an extent even greater than for BPTI. This conformational mobility has important consequences for the interpretation of both its structural and functional properties. Recently, the results of an independently determined X-ray structure of a fragment of domain C5 (residues 7–64; see below) were published [22]. Superposition of the well structured main core of the major conformation of the nuclear magnetic resonance (NMR) structures with the X-ray structure showed that the overall folds obtained by the X-ray and NMR methods are very similar. The crystallographic data, however, showed no evidence for the presence of multiple conformations in C5.

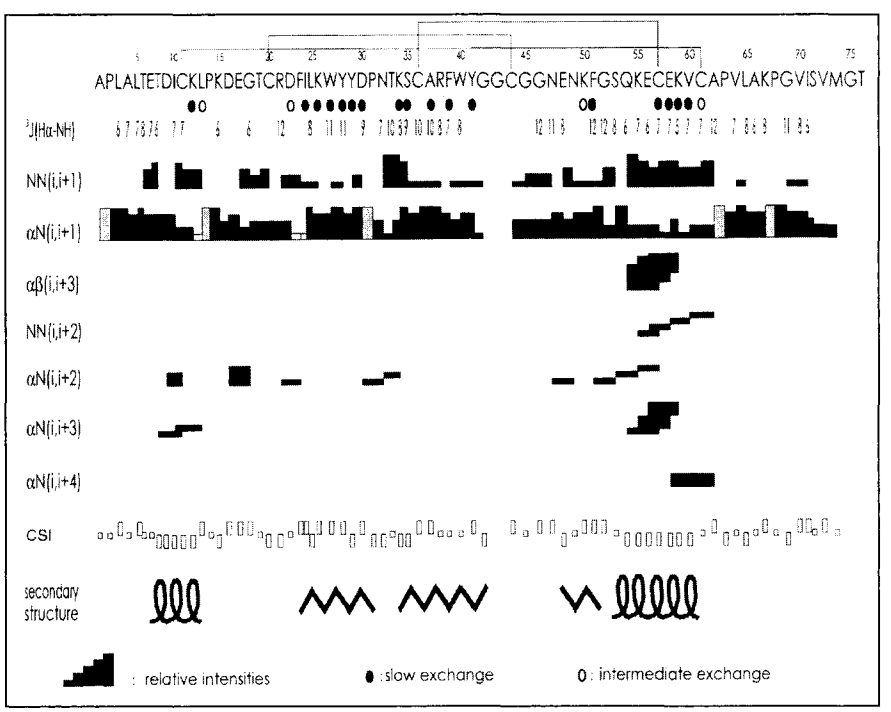
Results

Domain C5 corresponds to positions 2869 to 2943 of the human type VI collagen α3(VI) chain [1]. This sequence, plus an additional N-terminal alanine, was obtained as a recombinant fragment from mammalian cell clones [2] and will, for the sake of simplicity, be referred to as residues 1 to 76 (Fig. 1). Sequence alignment of domain C5 with other members of the Kunitz family is provided in [1] and [22].

Sequence-specific assignment of ^{1}H NMR spectra

The NMR spectra of domain C5 were assigned with the 2D ^{1}H homonuclear NMR methods [23] based on the following spectra: 2D nuclear Overhauser effect (NOE) spectroscopy (NOESY) [24,25], total correlation spectroscopy (TOCSY) [26,27], correlated spectroscopy (COSY) [28], double quantum filtered homonuclear correlated spectroscopy (DQF COSY) [29] and rotating frame Overhauser spectroscopy (ROESY) [30,31]. Because of a slow dynamic equilibrium present in C5, the ROESY spectra were essential for assigning the resonances of the major and minor conformations of the protein. Proton resonances of all backbone and side-chain atoms were assigned with the exception of the spin systems of Gly43, Gly75 and Thr76 and several side-chain protons (H^{ϵ} of Lys15, H^{β} of Thr19, H4 of Trp40 and ϵ-CH_3 of Met74). A section of a typical water NOESY spectrum is shown in Figure 2 and a complete list of assignments is given in Table 1. The signals of Thr19 and Cys20 (two residues in the binding loop) and Cys44 (the residue across the disulfide bridge 20–44) were only

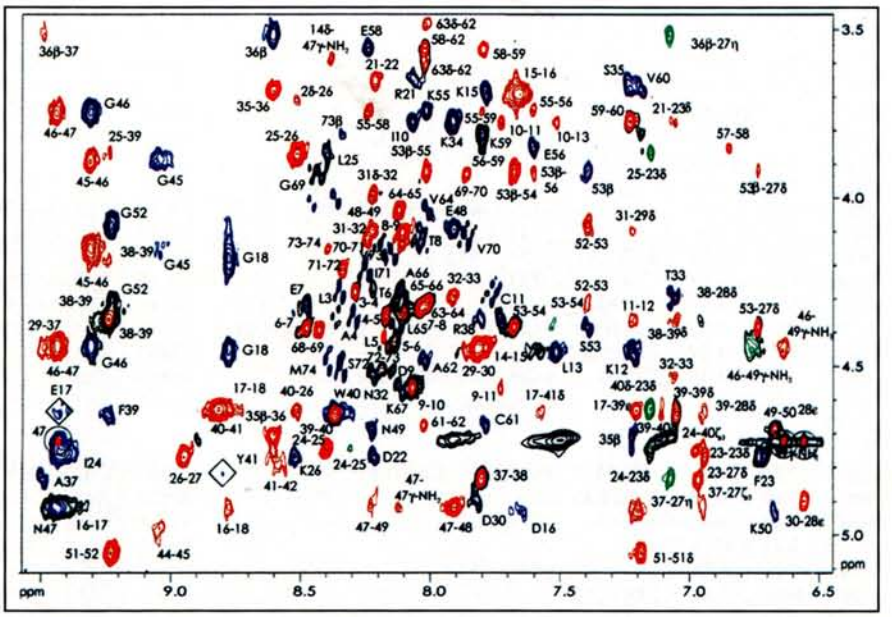
Figure 1



The amino-acid sequence of recombinant domain C5 and a survey of the NMR data used for locating secondary structure elements. The NOEs ($|i-j|<5$), classified as weak, weak-medium, medium, strong, and very strong, are represented by the heights of the bars and were extracted from the NOESY spectra with mixing times of 46 ms. The $H^{\alpha}(i)$–$H^{\delta}(i+1)$ (Pro) NOEs are shown in dark grey along the same line as the $H^{\alpha}(i)$–$H^{N}(i+1)$ connectivities. $H^{\alpha}(i)$–$H^{N}(i+1)$ connectivities which could not unambiguously be identified due to overlap are marked by white bars. Filled circles indicate H^{N}s that did not exchange against D_2O after 6 h; open circles indicate H^{N}s that did exchange after 3 months at pH 4.3. The values of the $J_{C\alpha H\text{-}NH}$ coupling constants (in Hz) are indicated below the sequence and the disulfide bridges above. The chemical shift indices (CSI) of H^{α} protons are defined in [35].

Figure 2

The H^N–H^α region of the 2D NOESY spectrum of domain C5, recorded at pH 6.2, 30°C and with a mixing time of 150 ms. Intraresidual peaks are marked in blue, inter-residual NOEs in red, and connectivities involving minor conformations in green. NOEs between non-exchangeable protein protons and hydration water molecules are indicated by the pink dot and circle. The H^α–H^N cross peaks of Glu17 and Tyr41, which are very weak in this spectrum, are highlighted by a square. Intraresidual H^N–H^α NOEs are labeled with one-letter abbreviations and inter-residual cross peaks are indicated by two numbers.



observed at 10°C. In addition, low-temperature spectra were crucial for the assignment of Glu17 and Gly18. For these residues, the sequential connectivites were missing in the spectra at higher temperatures and, in addition, the amide proton of Glu17 was obscured by spectral overlap. At 10°C, the amide proton of Glu17 was resolved, and both Glu17 and Gly18 showed sequential contacts. There are five proline residues in the sequence: Pro2, Pro14, Pro31, Pro63 and Pro68. All of them had *trans* peptide bond conformations, as inferred from the presence of very strong H^α(i)-H^δ(i+1) cross peaks in the NOESY D_2O spectra [32].

Stereospecific assignments were obtained for the β protons of 15 residues, the methyl groups of Val60 and the γ-CH_2 protons of Arg38 (Table 1), by using $^3J_{\alpha\beta}$ coupling constants and the NOE intensity patterns [19,33]. The methyl groups of Leu13 were assigned stereospecifically during the refinement stage of the structure determination.

Secondary structure

Figure 1 summarizes the NMR data used to determine the secondary structure of C5. Helical conformations are present between residue 8 and 13 and between 53 and 61 (Figs 1,3). The helix from Ser53 to Cys61 could be unambiguously identified as an α helix. Several properties are consistent with this determination: strong sequential H^N(i)–H^N(i+1) cross-peaks; H^N(i)–H^N(i+2), H^α(i)–H^β(i+3) connectivities; some H^α(i)–H^N(i+2) and H^α(i)–H^N(i+3,4) NOEs [32]; medium or slow amide exchange rates [23]; $^3J_{HN\alpha}$ coupling constants smaller than 7 Hz [34]; and chemical shift H^α indices of –1 [35]. Characterization of the N-terminal helix (residues 8–13) is not as clear because of the paucity of the NMR data. For example, no H^α(i)–H^N(i+4) contacts were observed. However, the pattern of slowly exchanging amide protons, in combination with the H^α(i)–H^N(i+3) connectivities, indicated that this helix is also best described as an α helix. A three-stranded antiparallel β sheet (residues 22–31, 34–42 and 49–51) was identified based on the presence of strong sequential H^α(i)–H^N(i+1) cross peaks, weak intraresidual H^α(i)–H^N(i) cross peaks and H^N(i)–H^N(j) interstrand NOEs. Six interstrand H^α(i)–H^α(j) connectivities and five interstrand H^N(i)–H^N(j) cross peaks were observed between strands comprising residues 22–31 and 34–42, allowing the identification of 11 hydrogen bonds. Although the third β strand, from Asn49 to Phe51, is much shorter than the two other strands, it unambiguously participates in the β sheet, as shown by the strong HN-HN contact between Trp27 and Phe51 and the H^α–H^α NOE between Tyr28 and Lys50 seen in the NOESY spectra. The fragment comprising residues 31–34 forms a type I turn. Weak H^α(i)–H^N(i+1) and strong H^N(i)–H^N(i+1) cross peaks from Asn32 to Thr33 are observed and the $^3J_{HN\alpha}$ coupling constant is 10 Hz for Thr33 [23].

In addition to the hydrogen bonds involved in the formation of the triple stranded β sheet, two hydrogen bonds, between Glu48 and Lys50 and between the side chains of Asn49 and Tyr29, were identified from the 3D structures (Fig. 3). After changing the pH to 2.7, the amide protons of Phe23, Arg38, Trp40, Gly42, Cys44 and Cys61 also exhibited slow amide exchange. However, these data were not used in the structure calculations as distances from the H^N proton to the possible backbone oxygen

Table 1

^{1}H chemical shifts (ppm) of domain C5 at 30°C and pH 6.2.

Residue	NH	CαH	CβH	Others	Residue	NH	CαH	CβH	Others
Ala1	–	4.38	1.54		Phe39	9.24	4.66	3.08,2.96	C2,6H 7.06; C3,5H 7.22; C4H 6.44
Pro2	–	4.50	2.35	γ-CH_2 1.91,2.01; δ-CH_2 3.73,3.62	Trp40	8.37	4.64	2.96*,2.75	N1H 9.91; C2H 7.12; C5H 7.00; C6H 7.20; C7H 7.41
Leu3	8.35	4.29	1.66,1.64	γ-CH 1.59; δ-CH_3 0.95,0.90					
Ala4	8.29	4.36	1.37		Tyr41	8.82	4.81	2.66*,2.40	C2,6H 7.59; C3,5H 6.74
Leu5	8.17	4.44	1.67	γ-CH 1.62; δ-CH_3 0.94,0.88	Gly42	8.57	3.38,3.01		
Thr6	8.14	4.4	4.32	γ-CH_3 1.23	Gly43				
Glu7	8.48	4.34	2.15,2.47*	γ-CH_2 2.51,2.47	Cys44	7.40†	4.77†	2.77†,1.49†	
Thr8	8.04	4.15	4.11	γ-CH_3 1.21	Gly45	9.04	4.17,3.90		
Asp9	8.11	4.58	2.83		Gly46	9.31	4.45,3.76		
Ile10	8.07	3.79	1.86	γ-CH_2 1.62; γ-CH_3 1.16; δ-CH_3 0.93	Asn47	9.44	4.93	3.18*,2.91	γ-NH_2 8.39,8.13
Cys11	7.73	4.37	2.96,2.90*		Glu48	7.91	4.1	1.72	γ-CH_2 2.51,2.32
Lys12	7.22	4.48	2.17	γ-CH_2 1.50,1.43; δ-CH_2 1.79; ε-CH_2 3.02	Asn49	8.22	4.7	3.18,2.90*	γ-NH_2 7.59, 6.65
					Lys50	6.67	4.94	1.68*,1.54	γ-CH_2 0.49; δ-CH_2 1.06; ε-CH_2 2.89, 2.56
Leu13	7.52	4.47	2.01	γ-CH 1.52; δ-CH_3 1.04,0.96					
Pro14	–	4.46	2.28	γ-CH_2 2.08, 1.89; δ-CH_2 3.85,3.61	Phe51	9.93	5.06	3.28,2.70*	C2,6H 7.20; C3,5H 7,67; C4H 7,22
Lys15	7.79	3.7	0.56	γ-CH_2 0.2; δ-CH_2 0.47	Gly52	9.23	4.32, 4.09		
Asp16	7.66	4.94	2.76,2.42		Ser53	7.40	4.4	3.93,3.66*	
Glu17	9.44	4.65	2.57	γ-CH_2 3.03	Gln54	7.68	2.51	1.11,–0.74	γ-CH_2 1.72,1.49; ε-NH_2 7.29,6.83
Gly18	8.78	4.47, 4.19			Lys55	8.02	3.75	1.69	γ-CH_2 1.37,1.29; δ-CH_2 1.59; ε-CH_2 2.91
Thr19	7.35†	4.26†		γ-CH_3 0.97†					
Cys20	8.82†	4.34†	2.88†,2.47†		Glu56	7.61	3.86	2.07,2.02	γ-CH_2 2.44,2.41
Arg21	8.06	3.66	1.92	γ-CH_2 1.38, 1.33; δ-CH_2 3.19,3.13; ε-NH 7.25	Cys57	6.85	2.08	3.11,2.91*	
					Glu58	8.24	3.57	1.94	γ-CH_2 2.54,2.27
Asp22	8.21	4.78	2.72,2.49		Lys59	7.80	3.82	1.82	γ-CH_2 1.34; δ-CH_2 1.46; ε-CH_2 2.91
Phe23	6.72	4.78	2.92,2.61*	C2,6H 6.95; C3,5H 7.26; C4H 7.23	Val60	7.19	3.69	1.67	γ-CH_3 1.06,0.99
Ile24	9.41	4.77	1.98	γ-CH_2 1.66,1.31; γ-CH_3 1.09; δ-CH_3 1.04	Cys61	7.79	4.69	2.02*,1.73	
					Ala62	8.03	4.5	1.32	
Leu25	8.40	3.78	1.18,0.99	γ-CH 0.66; δ-CH_3 0.05,0.02	Pro63	–	4.33	2.27	γ-CH_2 1.96, 1.83; δ-CH_2 3.62,3.50
Lys26	8.50	4.78	1.82	γ-CH_2 1.07,0.85; δ-CH_2 1.26,1.09; ε-CH_2 2.73	Val64	8.02	4.05	2.04	γ-CH_3 0.88
					Leu65	8.13	4.36	1.62	γ-CH 1.56; δ-CH_3 0.93,0.88
Trp27	8.95	5.73	3.04, 2.76*	N1H 10.13; C2H 6.75; C4H 6.86; C5H 6.99; C6H 7.12; C7H 7.44	Ala66	8.10	4.29	1.34	
					Lys67	8.12	4.58	1.81	γ-CH_2 1.50, 1.46; δ-CH_2 1.70; ε-CH_2 3.02
Tyr28	10.01	5.28	2.89,2.79	C2,6H 6.97; C3,5H 6.57					
Tyr29	10.30	4.46	3.36,2.86*	C2,6H 7.23; C3,5H 6.39	Pro68	–	4.41	2.3	γ-CH_2 2.01,1.93; δ-CH_2 3.81,3.64
Asp30	7.82	4.91	3.14,2.05		Gly69	8.43	3.94,3.68		
Pro31	–	4.12	2.49	γ-CH_2 2.17, 2.11; δ-CH_2 4.30,4.01	Val70	7.87	4.13	2.06	γ-CH_3 0.91
Asn32	8.22	4.54	3.04*,2.86	γ-NH_2 7.86,7.10	Ile71	8.25	4.23	1.86	γ-CH_2 1.49,1.22; γ-CH_3 0.91; δ-CH_3 0.86
Thr33	7.07	4.33	4.3	γ-CH_3 1.07					
Lys34	7.91	3.78	2.04	γ-CH_2 1.35; δ-CH_2 1.96; ε-CH_2 3.05	Ser72	8.34	4.52 3.83		
Ser35	7.23	3.69	4.72		Val73	8.19	4.17	2.1	γ-CH_3 0.93
Cys36	8.61	5.58	3.53*,2.89		Met74	8.40	4.51	2.12,2.03	γ-CH_2 2.63,2.55
Ala37	9.49	4.84	1.24		Gly75				
Arg38	7.81	4.38	1.05*,0.82	γ-CH_2 0.25, 0.00; δ-CH_2 2.64,2.59; ε-NH 6.80	Thr 76				

The chemical shift of water was 4.66 ppm. *The chemical shift corresponds to either Hα1, Hβ2 or Cγ1H_3 (IUPAC notation). †The chemical shift was measured in a H_2O NOESY recorded at 10°C and pH 4.3.

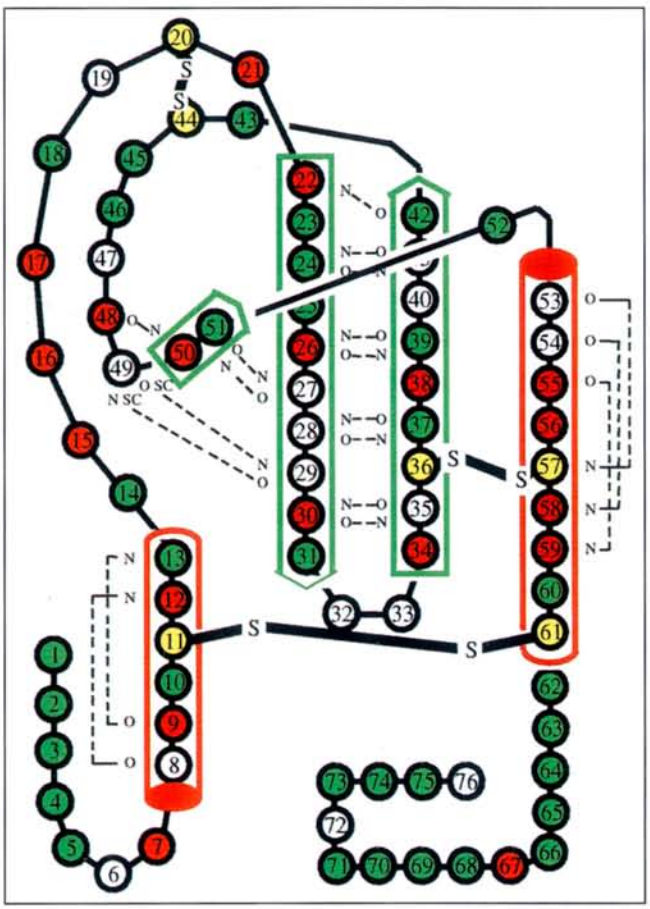
receptors for hydrogen bonds were longer than 2.5 Å in the final 3D structures.

On the basis of the high homology between domain C5, BPTI and other Kunitz-type inhibitors, we predicted disulfide pairing between Cys11–Cys61, Cys20–Cys44, and Cys36–Cys57. The disulfide bridge 20–44 was confirmed by the characteristic NOE contacts between the H^{β} protons of Cys20 and the H^{N} proton of Cys44. For the other two S–S pairings, no direct contacts could be observed between the corresponding cysteines. For these cases, either NOEs between residues that are in the vicinity of cysteines, for example Cys36 and Glu54, or long-range NOEs between residues close to the cysteines, were used to confirm the predicted S–S pairings.

Tertiary structure

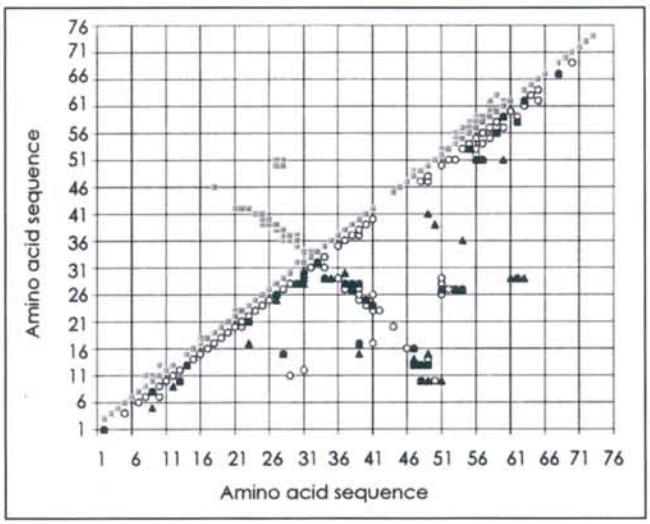
The 3D structure of C5 was calculated from 533 approximate interproton-distance constraints, including 36 hydrogen-bond constraints and 24 non-NOEs. No intraresidue distance constraints were included. The distance-constraint data were supplemented with 39 backbone ϕ, 10 ψ and 13 $\chi 1$ torsion-angle constraints. Figure 4 shows a diagonal plot of the NOEs identified in the NOESY spectra of C5. From Figures 4 and 5, it can be seen that the global folding of the polypeptide chain in the core is uniquely

Figure 3



Schematic diagram of the secondary structure elements in domain C5. Numbered circles represent the amino acids in the sequence. The colour of the circles indicates the type of amino acid: green for hydrophobic, light grey for polar, red for charged residues and yellow for cysteines. Red cylinders indicate the position of α helices; the green arrows represent the observed NOE connectivities that define the β sheet. Hydrogen bonds observed at pH 4.3 and 30°C are indicated by dashed lines.

Figure 4

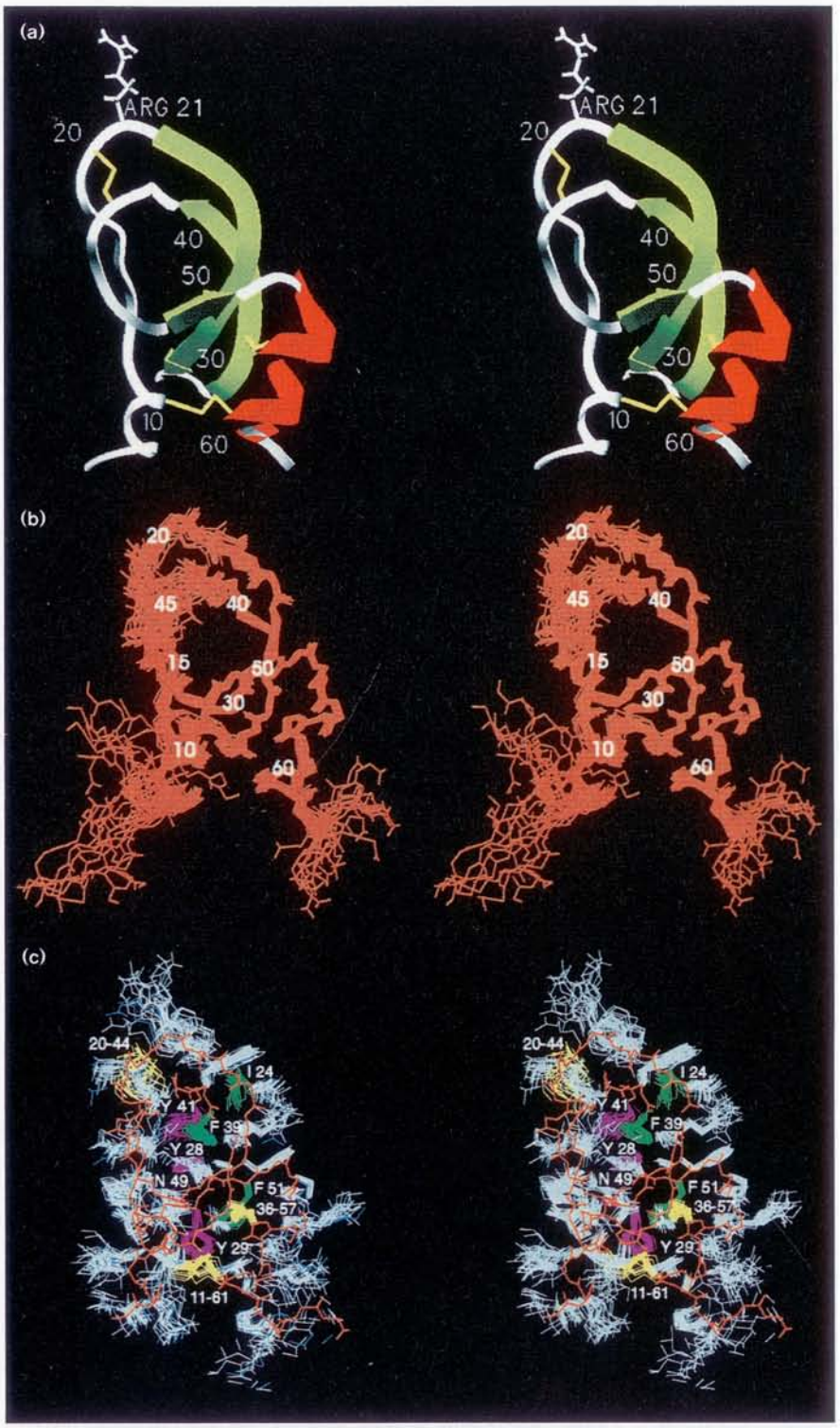


Diagonal plot of the interproton-distance constraints used as input in the structure calculations. Backbone–backbone interactions (squares) are plotted above the diagonal; backbone–side-chain (circles) and side-chain–side-chain (triangles) constraints are shown below the diagonal.

defined due to the large number of NOEs. The exception is the region in the vicinity of the disulfide bridge Cys20–Cys44. The structural statistics of the final structures are given in Table 2 (<SA> represents the final 2D structures and the minimized mean structure is represented by (SA)m). It is evident that all simulated annealing (SA) structures, as well as the (SA)m structure, satisfy the experimental constraints and exhibit only small deviations from the idealized covalent geometry. The average atomic rms difference, among the 20 calculated structures, for heavy atoms in secondary structure elements of domain C5 (residues 8–13, 22–31, 34–42, 50–51, 53–61) (Fig. 6) was 0.4±0.1 Å for the backbone atoms and 0.9±0.3 Å for all atoms. The rms deviations among the SA structures for both the atomic coordinates and the backbone torsion angles are larger for residues 15–23 and 42–46; this was especially evident around the disulfide bridge Cys20–Cys44. In addition, the structures exhibited very large rms deviations for the N and C terminus between residues 1–6 and 64–76, respectively (7.5±1.5 Å for heavy backbone atoms and 7.7±1.8 Å for all heavy atoms). These two regions of the protein seem to be completely unstructured. There were no long-range NOEs and only three medium-range NOEs were observed for these segments (Fig. 4). Also, the TOCSY peaks were stronger with noticeable reduction in the line widths compared with the peaks of residues in well-structured regions.

In general, conformations of the side chains were also well defined. Figure 5c shows that only side chains of residues 12, 14, 21, 44 and 50 (excluding the unstructured terminal regions of C5) had greater rms differences in the calculated structures. The side chains of these residues, with the exception of Lys50, were not assigned stereospecifically; Lys50 was assigned stereospecifically but had few long-range NOEs. The side chains of Phe39, Phe51, Ile24, Tyr28, Tyr29 and Tyr41, in addition to the three disulfide bridges, were involved in the formation of the core of domain C5. Asn49 also plays an important role. Its side chain amide proton showed very slow exchange in D_2O. The side chain amide proton and the oxygen formed hydrogen bonds with the backbone oxygen and amide proton of Tyr29 (Fig. 3). Most of the backbone torsion angles of non-glycine residues fell within the allowed regions of the Ramachandran plot (data not shown). Only Thr19 and Lys67 showed energetically unfavorable positive ϕ torsion angles of 76.2° and 39.0°, respectively.

Figure 5



3D structure of domain C5. **(a)** Ribbon drawing of domain C5. Helices are marked in red, β strands in green and disulfide bridges in yellow. (The figure was generated by the program SETOR [59].) **(b)** Stereoview of the backbone atoms (N, Cα, C, O) of domain C5 structures best fitted to the N, Cα, and C atoms of all residues located in secondary structure elements (8–13, 22–31, 34–42, 50–51, 53–61). (The figure was generated by the program WHAT IF [60].) **(c)** Stereo picture of the Cα tracing of one of the C5 structures with the side chains of all the final structures shown. Side chains important for the protein core formation have been marked in green (hydrophobic residues), magenta (polar residues) and yellow (disulfide bridges).

The topology of the global fold of C5 is shown in Figures 3 and 5a. The core of domain C5 consists of a triple stranded antiparallel β sheet comprising a strongly twisted, short third strand. This β sheet is close to the two α helices, which are connected by the disulfide bridge Cys11–Cys61. The C-terminal helix is also connected to the β sheet by the disulfide bridge Cys36–Cys57. The portion of the protein between the second and the third

Table 2

Average deviations from ideality, energies of the structures, and rms differences between experimental and calculated distance constraints.

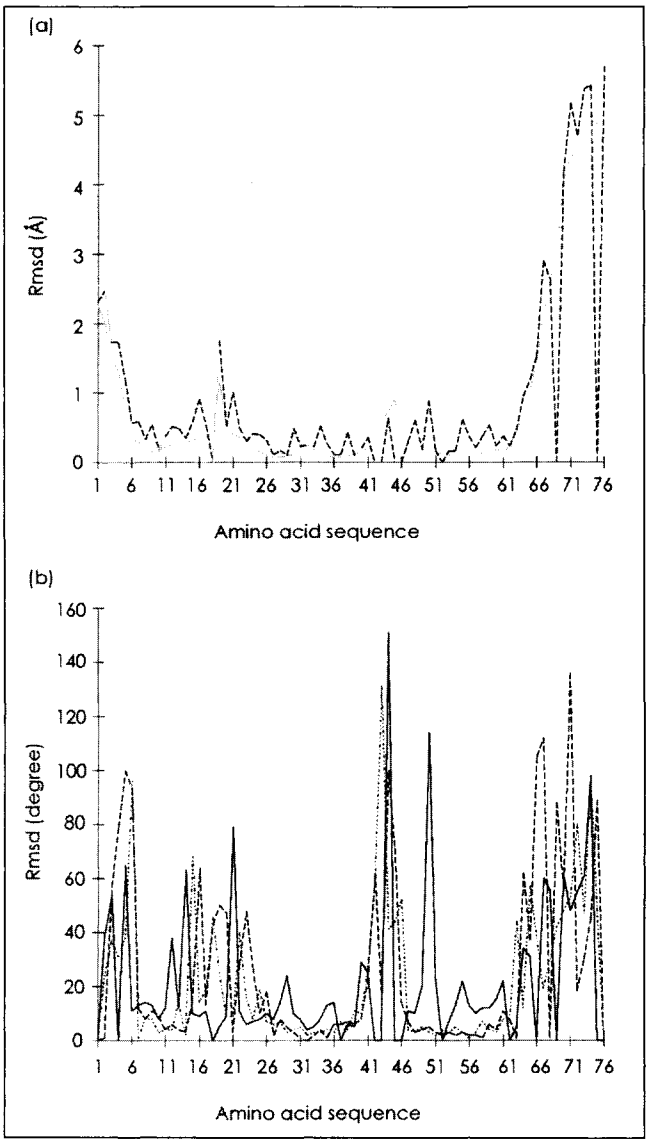
Parameter	<SA>*	(SA)m†
Deviations from idealized geometry		
bonds (Å)(1169)‡	0.004±0.000	0.006
Angles (°)(2124)‡	0.744±0.016	0.872
Impropers (°)(610)‡	0.430±0.015	0.452
Energies (kcal mol^{-1})§		
E_{NOE}	85.7±4.8	124.2
E_{tor}	2.5±1.3	1.7
E_{vdw}	16.8±3.3	154.2
E_{all}	316.5±13.3	594.4
Number of residual constraint violations for all 533 constraints		
0.3 Å≤NOE violation	4.3±1.0	10
0.4 Å≤NOE violation	0	2
0.5 Å≤NOE violation	0	0
Rms differences from experimental constraints (Å)#		
All (533)	0.057±0.002	0.076
Sequential (|i–j|=1) (232)	0.073±0.003	0.098
Medium range (1<|i–j|<5) (81)	0.032±0.005	0.045
Long range (|i–j|≥5) (160)	0.040±0.004	0.050
Non NOEs (24)	0.025±0.013	0.017
Hydrogen bonds (36)	0.060±0.006	0.092

*<SA> represents the 20 final structures. †(SA)m is the structure obtained by constrained minimization of the mean structure. The mean structure was obtained by averaging the coordinates of the 20 final <SA> structures best fitted to N, Cα and C atoms of all residues located in secondary structure elements (8–13, 22–31, 34–42, 50–51, 53–61). ‡The number of bond, angle, and improper terms is given in parentheses. §Force constants used to calculate energy terms are the same as those in [39]. #The rms deviations (in Å) from the interproton distance constraints were calculated as described in [39]. Number of distance constraints is given in parentheses.

strand of the β sheet, residues 45–48, is oriented antiparallel to residues 16–19 of the binding loop, which comprises residues Gly18 (P4) to Leu25 (P4′).

Figure 5b shows the final structures that had the smallest number of distance violations and lowest energies. Another distinct set of structures was obtained during the initial stage of structure determination of C5. This set differed substantially in the conformation of the fragment between residues 16 and 19 showing nearly the mirror image of the final structures in this region (Fig. 7). This resulted in a configuration in which the N-terminal part of the protein passed through the large loop built by residues 40 to 50. These structures also fulfilled all experimental constraints, with allowed deviations of 0.5 Å for distance constraints, 5° for torsion angle constraints, 0.1 Å for bonds, 10° for bond angles and 15° for impropers. Also, the backbone torsion angles were in the allowed regions of the Ramachandran plot. The introduction of specific non-NOEs in this region of the protein, obtained by comparison of the two different sets of structures, did not result in the selection of one of the two sets. However, the

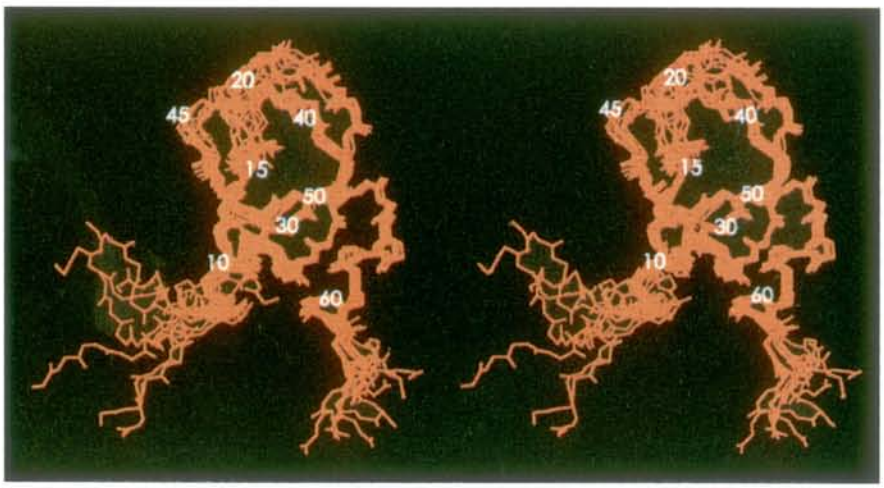
Figure 6



Rms deviations by residue in C5. **(a)** Residue-based rms deviations of the atomic coordinates among the 20 SA structures for the backbone atoms (solid line) and the non-hydrogen side chain atoms (dashed line). **(b)** Residue-based rms deviations of the backbone ϕ (dashed line), ψ (dotted line) and χ (solid line) torsion angles for the 20 SA structures.

structures with the different loop conformation had significantly higher improper (~+11 kcal mol^{-1}), bond (~+7 kcal mol^{-1}), angle (~+55 kcal mol^{-1}), van der Waals (~+28 kcal mol^{-1}), NOE (~+32 kcal mol^{-1}) and torsion angle energies (~+3 kcal mol^{-1}). (Values for the final structures are given in Table 2.) These higher energies could be traced to small, but distinct, violations of experimental constraints. For example, the calculated distance constraint between H^{α} of Thr19 and H^{N} of Cys20 was always 3.6 Å in these structures whereas the experimental value

Figure 7



Stereoview of the backbone atoms of the domain C5 structures with the misfolded chain for residues 17–19 best fitted to N, Cα, and C atoms of residues 8–13, 22–31, 34–42, 50–51 and 53–61.

was 3.1 Å. Also, the ϕ angle of Lys15 had a value of –36° for the observed $^3J_{HN\alpha}$ coupling constant of 6 Hz. Various bond angle constraints of Cys20, Gly43 and Cys44, were violated by more than 3°; for example, the S–S–Cγ angle of the disulfide bridge Cys20–Cys44 was 108.2°, compared with the ideal value of 104.2°, and the improper angle for 19Cα–19C–20N–20Cα was violated by 8°. As discussed below, this set of the structures was designated to be an artifact of the NMR structure determination, the origin of which could be traced to an insufficient amount of NOEs for residues 19 and 20.

Multiple conformations

The NMR data indicated that, at temperatures between 10°C and 30°C, C5 existed in at least two or three slowly exchanging conformations. The residues that showed two exchanging conformations and the chemical shifts of their minor conformations are listed in Table 3 and are shown schematically in Figure 8. In addition, C7H of Trp27 exhibited three exchanging conformations. Most of the residues in the core of C5 showed the presence of a second conformation. In most cases, the second conformation could be unambiguously assigned from the ROE exchange peaks between amide resonances of the two conformations in the H^N–H^N region of the ROESY spectra taken at 30°C (Fig. 9). In cases where no such ROE connectivities could be observed, for example Leu13, the minor conformation could still be identified from the NOESY spectra as the NOESY cross peak patterns of minor conformations were similar to that of the major conformation. Except for the NOE peak between Trp27 H^α and Arg38 H^α at 30°C, no inter-residue connectivities within the minor conformation were observed. Some weak NOEs were found between amide resonances of the major conformation and H^α resonances of the minor conformation, and vice versa, for example between Ala37 and Cys36. These peaks corresponded to the exchange-relayed ROE cross peaks [21]. Also, a few NOEs between a resonance that showed only one conformation and a resonance of a minor conformation were observed. At 15°C,

Table 3

^{1}H chemical shifts (ppm) of the second conformation at 30°C and pH 6.2.

Residue	NH	CαH	CβH	Others
Cys11	7.50*	4.12*		
Lys12	7.05*	4.22*	1.85*	γ-CH_2 1.21*; δ-CH_2 1.49*
Leu13	7.16*	4.22*	1.69*	
Asp22	8.17†			
Phe23				C2,6H 7.16
Ile24	9.38†			
Leu25	8.31			γ-CH 0.52‡; δ-CH_3 –0.14‡
Lys26	8.21			
Trp27	9.07	5.83‡		C2H 6.57; C6H 7.01; C7H 7.52 (7.36)
Tyr28	9.79	5.22‡		
Tyr29	10.29†			
Asp30		4.93‡		
Cys36	8.64	5.44‡		
Ala37	9.20			
Arg38	8.09	4.19‡		ε-NH 7.06
Phe39	9.29			C2,6H 6.97
Trp40				C6H 7.07§
Tyr41	8.84			
Gly45	8.67*			
Glu48	7.67*			
Asn49				γ-NH_2 6.78
Phe51	9.90			
Gln54	7.53			δ-NH_2 7.52
Cys61	7.65*	4.45*		

The chemical shift of water was 4.66 ppm. *Chemical shift measured in an H_2O NOESY at 10°C, pH 4.3. †Chemical shift measured in a D_2O ROESY at 30°C, pH 4.3. ‡Chemical shift measured in a D_2O NOESY at 30°C, pH 6.2. §Chemical shift measured in a H_2O TOCSY at 30°C, pH 6.2.

Figure 8

Ribbon plot of domain C5 showing the parts of the protein which exhibit multiple conformations (yellow) or in which the spin systems of the residues were only observed at and below 15°C (lilac). The unstructured N- and C-terminal residues are colored in red.

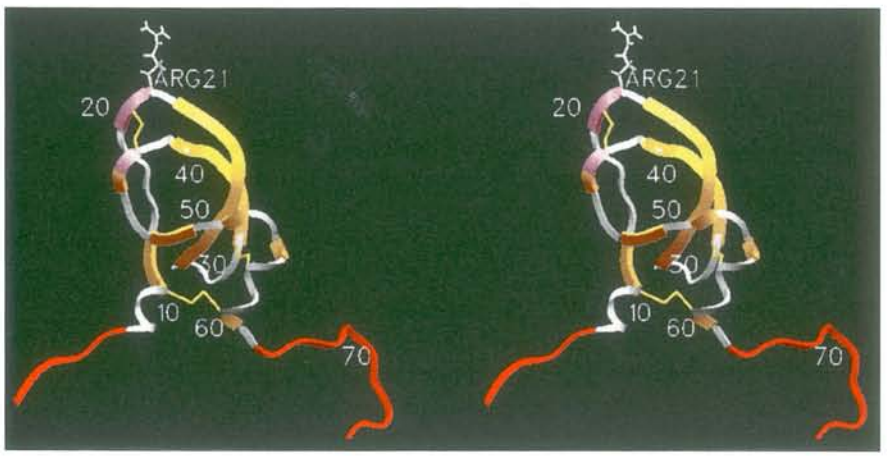
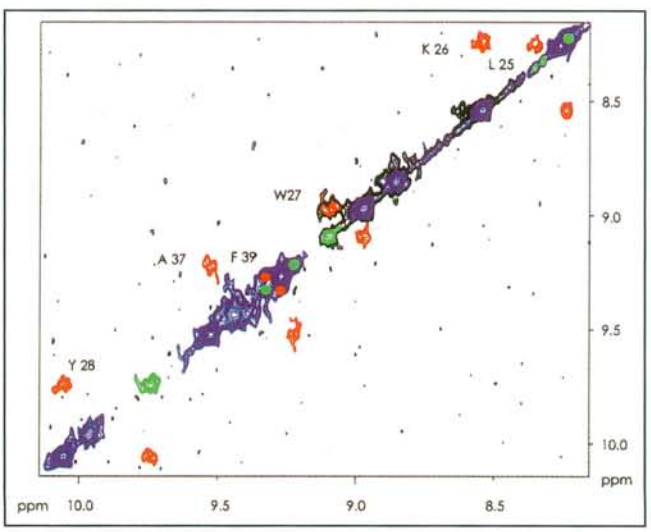


Figure 9



H^N–H^N region of the 2D ROESY spectrum of domain C5 recorded at pH 4.3, 30°C and with a mixing time of 45 ms. Intraresidual H^N–H^N diagonal peaks of the major conformation are marked in blue; intraresidual H^N–H^N peaks of the minor conformations are displayed in green; exchange peaks between the major and minor conformations are indicated in red.

most of the intermolecular exchange connectivities between the H^N–H^N resonances of the minor and major conformations disappeared both in the ROESY and NOESY spectra. The only exchange cross peaks present were those from the side chain of the arginine and lysine residues. However, at this temperature, additional intramolecular cross peaks to H^N resonances of the minor conformation were observed for Cys11, Gly45, Glu48 and Cys61. These residues exhibited only single broadened cross peaks at 30°C, indicating that the coalescence temperature for the exchange process is slightly below 30°C.

Ratios of the major to minor conformations are shown in Table 4 for temperatures 10°C and 30°C. The most reliable determination of the ratios was obtained for H^N of Tyr28 and H^α of Trp27; for these residues, the cross peaks from the major and minor conformations were well separated throughout the temperature and pH ranges studied. The data for these two resonances, as well as for H^N of Trp27 and H^N of Phe39, indicated that the amount of the minor conformation for the residues located in the β sheet decreased with increasing temperature. For H^N of Lys12, the ratio of the major to the minor conformations slightly decreased with the increasing temperature (Table 4); however, the ratios for the residues located in vicinity of the disulfide bridge 11–61 were all approximately 1 (or even <1 in the case of H^N of Lys12 at 10°C). Variation in the pH, between 2.7 and 4.3, also changed the ratios. For example, for H^N of Tyr28 the ratio of the major to the minor conformation at 30°C increased from 23.69±1.22 at pH 2.7 to 38.12±1.55 at pH 4.3. However, the temperature dependence of the ratios was identical in both pHs.

Table 4

Ratios of the major (M) to the minor (m) conformations for HN and HA resonances at pH 2.7

Resonance	$\frac{[M]}{[m]}$ (10°C)	$\frac{[M]}{[m]}$ (30°C)
Lys12HN	0.97±0.03	
Leu13HN	1.44±0.03	1.24±0.04
Trp27HN	5.45±0.28	19.39±0.58
Trp27HA	8.47±0.30	13.76±0.34
Tyr28HN	6.47±0.18	23.69±1.22
Cys36HA		15.96±0.56
Phe39HN	6.55±0.33	12.58±0.47
Cys61HN	1.06±0.06	

Errors are standard deviations.

Attempts to determine the exchange rates and activation energies (assuming a two-state equilibrium) were unsuccessful as it was not possible to determine the line widths for resonances of the minor conformation either due to overlap or because peak intensities were too low.

Hydration water

The amide protons of Gly18, Cys20, Tyr29, Gly42, Gly45, Gly46, Asn47, Asn49 and Phe51, the H^{ϵ} of Tyr28 and H^{δ} of Asn49 showed cross peaks at the water frequency in the NOESY and ROESY at pH 4.3 and 15°C. At 30°C, cross peaks from the amide protons of Asn47, Phe51, the H^{ϵ} of Tyr28, and the H^{δ} of Asn49 were still present in the spectra. All of these cross peaks were checked carefully for their origin. None of them could arise from the intraprotein NOEs (Table 1 and Table 3). There are several other possible sources of these magnetization transfers [36]: a direct intermolecular NOE between the protein proton and the protons of hydration water molecules that slowly exchange with the bulk water; an intramolecular NOE between a protein proton and a labile protein proton which is in rapid exchange with the bulk water and therefore appears at the water chemical shift; and chemical exchange of a labile protein proton with bulk water protons. With the exception of the amide proton of Phe51, the cross peaks of all other residues had an opposite sign relative to the diagonal of the ROESY spectrum identifying them as NOEs originating from either of the first two possibilities described. To distinguish between these two mechanisms, the distances were extracted from the 3D structures for protons that exhibited the water cross peaks and the labile amino, hydroxyl or carboxylate protons. A direct NOE between the hydration water molecule and the protein proton can only be positively established if the nearest rapidly exchanging proton is at least 4.0 Å away from the protein proton [36]. Most of the cross peaks at the chemical shift of water, except for H^{N} of Phe51, H^{N} of Gly46 and H^{ϵ}s of Tyr28, could be unambiguously identified as direct NOEs.

Water molecules were introduced into the calculations in the later stages of the structure determination. First, only one water molecule was introduced; however, this lead to the violations of several distance constraints involving the water molecule that were larger than 1 Å. (For example, the shortest distance between H^{δ}s of Asn49 and the water protons was 5.5 Å, whereas the experimental NOE intensity corresponded to a distance constraint of 2.7 Å.) Introduction of the second water molecule resulted in an excellent agreement between the experimental and calculated distance constraints involving water molecules (the H^{δ} of Asn49 to the second water molecule was 2.9 Å). However, the introduction of the two bound internal water molecules did not induce any significant conformational changes into the structures calculated without water molecules. Bound water was also not helpful for differentiating between the final 'correct' structures and the structures that had an incorrect chain folding between residues 16 and 19.

Discussion

The tertiary structure of C5 is very similar to all known structures of other members of the Kunitz family [37] (to avoid any nomenclature confusion, the Kunitz-type family is referred to the classification used by Bode and Huber [37]). Of these family members, we made the most detailed comparison between the structure of C5 and BPTI. The average rms deviation between C5 (residues 8–13, 22–31, 34–42, 50–51, 53–61) and BPTI (NMR structure) for the α carbons was 0.9 Å. Comparison of the structure of C5 with those of BPTI, APPI [13], dendrotoxin K [14], and dendrotoxin I [15] shows that the core hydrophobic residues of C5, 22–42 and 53–61 can be superimposed with an rms difference of 0.7 Å. It is also obvious that the minimum length of the Kunitz-like type domain of C5 is from residues 8–63 as the fragments 1–7 and 64–76 were completely unstructured in solution.

At the same time as the NMR structure determination of C5 was completed, the results were published of an independently determined X-ray structure of a fragment from residue 7 to 64 [22]. Superposition of the well-structured main core of the average NMR structure (residues 8–13, 22–31, 34–42, 50–51 and 53–61) with the X-ray structure showed that the overall folds obtained by the X-ray and NMR methods (the major conformation) are very similar. However, the crystallographic data showed no evidence for the presence of multiple conformations in C5. Structures of dendrotoxin K [14] and dendrotoxin I [15] were determined in solution by NMR, and although no conformational inhomogeneities were reported for these toxins, some conformational variability might still be present in dendrotoxin I, as its structure showed that the loop between the second β strand and the second helix was disordered [15].

Conformational variability

C5 exists in solution, at temperatures between 10°C and 30°C, in at least two or three slowly exchanging conformations. 24 residues from the total of 55 residues in the well-defined main part of C5 showed the presence of multiple conformations. There are approximately three patches of residues that have multiple conformations (Fig. 8). The first patch is located in the central β sheet; for these residues, the population of the minor conformation, which was about 13% at 10°C, decreased with the increasing temperature to about 5% at 30°C (Table 4). The second group of conformationally inhomogeneous residues consists of the loop residues around the disulfide bridge 20–44. These residues showed no cross peaks in the TOCSY and (R)NOESY spectra at 30°C and broad or multiple peaks at 10°C. The third group of residues is located

in the vicinity of the disulfide bridge 11–61. For these residues, the population of the minor conformation was high (50%) with a slight trend toward increasing this population at higher temperatures (Table 4).

Otting *et al.* [21] observed two conformational isomers in the ^{1}H NMR spectra of both BPTI and the BPTI mutant in which Gly36 was replaced by serine. In BPTI, the population of the minor conformer increased from about 1.5%, at 4°C, to 8% at 68°C. In the serine mutant, the population of the minor form was larger (about 15%, at 4°C, of the total protein) and there was also a trend towards increasing the population of the minor conformer at higher temperatures. The trend observed in BPTI is therefore opposite to that observed in the central β sheet of C5. It is also evident from our data that the conformational inhomogeneity of C5 is more extensive than that of BPTI and its mutant. For BPTI, it was postulated that the two conformational isomers differed from each other by different chiralities of the Cys14–Cys38 disulfide bridge. It is difficult to establish the origins of conformational isomerism in C5 due to its abundance; several coupled molecular processes can contribute to conformation variability. The inhomogeneity might, to some extent, be associated with the isomerization of the disulfide bonds as multiple conformational exchanges were observed in the 3D structures for residues in the vicinity of the disulfide bridge 11–61 (Table 4). The residues neighboring the disulfide bridge 20–44, which gave no NMR spectra above 15°C, also exhibited similar features. However, it can be seen from Tables 1 and 3 that the largest differences in chemical shifts between the minor and major conformations were exhibited by the residues in the β sheet. Different conformational exchange processes must be responsible for structure fluctuation in this part of domain C5 and they may also influence the whole protein.

A set of structures of C5 obtained during the initial stage of structure determination differed substantially from that of the final structures in the conformation of the fragment between residues 16 and 19; it showed a near mirror image of the final structures in this region (Figs 5,7). The N-terminal part in these structures passed through the large loop built by residues 40–50. Although none of these structures exhibited violations of the NOEs larger than 0.5 Å, all of the NOE and dihedral constraints were better fulfilled in the final structures. This is reminiscent of the situation found in conformations of the binding loop of serine proteinase inhibitors for which NMR structures have been determined: CMTI-I [38,39], LDTI [40], eglin c [41], a low resolution structure of BPTI [19], and dendrotoxin I which exhibits no trypsin activity [15]. However, in each of these cases, the conformational variability in the structures was manifested by a large distribution in conformations of the binding loop only. In the case of C5, the lack of NOEs for residues located at the hinges of the segment 17–20 resulted in a dramatic change in the global tracing of a large fragment of the protein (Figs 5,7). It is also worth mentioning that introduction of two bound internal water molecules did not induce any significant conformational changes in the either set of structures and was of no help in differentiating between the correctly and incorrectly folded structures. This is in contrast to results for a small trypsin inhibitor, EETI-II, of the squash family of trypsin inhibitors, for which CMTI-I is a representative member. For EETI-II, the incorporation of water in the NMR structure determination was successful for indicating the correct versus misfolded models of the inhibitor [42].

Exchange-relayed ROE cross peaks were observed between amide resonances of the major conformation and Hα resonances of the minor conformation in the NOESY and ROESY spectra of C5 (see the Results section). Also, a few NOEs were observed between a resonance that showed only one conformation and a resonance of the minor conformation. Including these NOEs in the complete list of the experimental constraints during the structure calculation resulted in large violations (≥1 Å) of previously fulfilled experimental constraints. As a consequence no structures could be obtained that fulfilled all experimental constraints (data not shown). The implications of the presence of conformational equilibria for the determination of structures of proteins by NMR were discussed by Jardetzky [43] and more recently by Kim and Prestegard [44]. The experimental results for C5 clearly show that the commonly used assumption in the determination of solution structures of proteins, that is, that the data can be interpreted by a single rigid conformation, should be carefully checked before a straightforward conversion of NOE intensities to distances is assumed.

Lack of protease inhibitor activity

The segment comprising residues 19–24 in C5 (Thr19 [P3], Cys20 [P2], Arg21 [P1], Asp22 [P1′], Phe23 [P2′] and Ile24 [P3′]) corresponds to the protease binding loop in Kunitz-type inhibitors. Arginine, at position P1, is the primary determinant of inhibitor specificity and usually results in trypsin inhibition. The lack of inhibitory activity by domain C5 can be now explained at the structural level, from information available from the X-ray [22] and present NMR analyses. The NMR solution structure of C5 shows that the loop exhibits conformational variability. Though disordered, the loop is exposed and poised for insertion into the specificity pocket of the target protease. Such properties of the binding loop have been found in other free serine proteinase inhibitors for which the NMR structure has been determined: CMTI-I [38], LDTI [40], and eglin c [41]. We therefore suggest that the lack of inhibitory activity of C5 cannot be caused by conformational instability of the binding loop but has to be primarily associated with the type of amino-acid side-chains present in the loop. This prediction has now been confirmed by

site-directed mutagenesis of domain C5 (Kohfeldt *et al.*, personal communication). For example, substitution of Asp22 (P1′) for alanine produced inhibitory activity of domain C5 for trypsin. Another example of the importance of the amino-acid side-chain composition in the binding loop is provided by Phe23 in C5. In the C5 structures, the Phe23 ring is fixed in a hydrophobic pocket that consists of Phe23, Trp40 and Gly42. The presence of a hydrophobic residue (phenylalanine) at P2′ in C5 removes the important stabilizing effect of hydrogen bonding that is present in BPTI, where imino protons of the arginine (P2′) side chain make hydrogen bonds to oxygen donors of trypsin. As already observed by Arnoux *et al.* [22], in a hypothetical complex modeled on the BPTI–trypsin interaction, the P2′ phenylalanine in C5 should also result in severe steric hindrance between the aromatic ring of phenylalanine and Gly192 and Gly193 of trypsin.

Biological implications

Collagen α1 (VII) and α3 (VI) chains possess a sequence of about 60 amino-acid residues that is highly similar to a sequence found in serine protease inhibitors of the Kunitz type. The corresponding 76-residue module of α3 (VI) chain (domain C5) was recently prepared in recombinant form and showed high stability against proteases; however, it lacked any inhibitory activity against trypsin, thrombin, kallikrein and several other proteases [2]. This module is also found in other large proteins, for example, a human Alzheimer amyloid precursor protein. The amino-acid sequence of the module shows 42–50% identity with that of bovine pancreatic trypsin inhibitor (BPTI), a representative member of the Kunitz-type serine protease inhibitors. BPTI has been thoroughly investigated by both X-ray diffraction and NMR methods.

All X-ray diffraction studies of Kunitz-type domains indicate that these modules form compact and rigid structures with well-defined hydrophobic cores. Six cysteine residues, two antiparallel β strands and a short α helix are characteristic features of these structures and an exposed binding loop projects away from the supporting scaffold. All six cysteines are oxidized and form disulfide bridges, which contribute to the overall stability of the protein. The stability against denaturation by heat or chemical denaturants is very high.

We report the structure of domain C5 in solution determined by the 2D NMR technique. Recently, the crystal structure of a fragment, residues 7–64 of domain C5 of the collagen α3(VI) chain, has also been determined [22]. The present NMR study demonstrates that domain C5 is a highly dynamic molecule in solution, with exchanging multiple conformations. This is in contrast to the crystallographic data which showed no evidence for the presence of multiple conformations in C5 [22]. Nevertheless, superposition with the X-ray structure of the well-structured main core of the major conformation of the NMR structures showed that the overall folds obtained by the X-ray and NMR methods are very similar.

Previous NMR studies on BPTI showed that BPTI in solution exists in two exchanging conformational isomers. This isomerism was proposed to arise from the isomerization of a single disulfide bridge, which then affected about 10 residues from the total of 58. In domain C5, the origin of multiple conformations is difficult to pinpoint due to the number of residues affected; multiple conformations were observed for 24 residues (44%) of the main body structure (55 residues). The observation of multiple conformations in these proteins may be surprising in view of the conformational constraints imposed by the 3D structure of proteins as rigid as C5 or BPTI. These properties may occur in many proteins with this fold, and should be considered in the interpretation of their structural and dynamical properties. In the case of C5, the accessibility of the inhibitory binding loop should be relatively unaffected by conformational exchange. An arginine at position P1 is the primary determinant of specificity of Kunitz inhibitors and usually results in trypsin inhibition. As arginine is present at P1 position in C5, it seems likely that the lack of trypsin inhibition arises from the amino-acid side-chain composition of the binding loop.

Materials and methods

Sample Preparation

Recombinant domain C5 of human type VI collagen was purified from mammalian cell clones, as described by Mayer *et al.* [2]. The protein was exhaustively dialyzed against $(NH_4)_2CO_3$ and lyophilized before the use in NMR studies. The samples for NMR contained 1.2 mM of C5 in 50 mM NaH_2PO_4/CDC_3OONa buffer in either 90% H_2O/10% D_2O or 100% D_2O. C5 was studied at three different pHs starting with pH 6.2 and consecutively lowering pH to 4.3 and 2.7 by the addition of phosphoric acid.

NMR spectroscopy

Most of the NMR studies of C5 were carried out at pH 4.3 and pH 6.2. The data at pH 4.3 were used for the structure determination of the protein, whereas spectra measured at pH 6.2 were used for the sequence-specific assignments.

NMR measurements were carried out at 500 MHz and 600 MHz on the Bruker AMX-500 and DRX-600 spectrometers, respectively. All 2D spectra [45] were recorded in the pure-phase absorption mode using the time proportional incrementation methods [46]. The following spectra were recorded in H_2O and D_2O: NOESY [24,25] with mixing times of 46 ms, 100 ms, 126 ms and 150 ms, and TOCSY [26,27] with the MLEV-17 mixing sequence [47,48] of 50 ms and 60 ms duration. The NOESY spectra with a mixing time of 126 ms and the TOCSY with a mixing time of 50 ms were recorded for samples at the three pH values used in the study. A COSY [28,45] spectrum was

recorded in H_2O, whereas in D_2O a DQF COSY [29] was acquired. Both were measured at pH 4.3 and 15°C. The ROESY spectrum [30,31] was acquired in D_2O at pH 4.3 and 30°C whereas the water ROESY spectra were recorded at pH 2.7 and 15°C. For the NOESY spectra in H_2O, the water resonance was suppressed by the use of a semi-selective excitation pulse in which the last 90° pulse in the sequence was replaced by the jump-return sequence with the carrier placed at the position of water [49,50]. A homospoil pulse of 8 ms duration was used during the mixing time. Essentially the same pulse sequences were used for the D_2O samples, with the exception that no or only slight presaturation was used for suppression of residual water. For the TOCSY and COSY spectra in H_2O, the water signal was suppressed by presaturation of the water resonance. For ROESY spectra in H_2O, presaturation or a water suppression proposed by Otting and Wüthrich [36] and Sklenar and Bax [51] were used. To study NOEs located at the chemical shift of water, a ROESY spectrum was recorded using a pulse sequence proposed by Otting and Wüthrich [36]. The mixing time was set to 67 ms. For the majority of the spectra, 700 t_1 increments were collected, each with 4K data points and 192 or 288 scans per t_1 value, over a spectral width of 12.65 ppm in both dimensions. The spectra were processed using a Gaussian function in F_2 and a quadratic sinus in F_1 after zero filling to 1K in F_1.

Amide exchange rates were obtained by acquiring two 2D NOESYs of domain C5 at pH 4.3 and 15°C. One NOESY was measured immediately after domain C5 lyophilized from water was dissolved in D_2O, the other was recorded after domain C5 was in D_2O for three months (Fig. 1).

Measurements of the populations of different conformations

Relative concentrations of conformational forms were determined from the intensities in the NOESY spectra with the shortest mixing time. The use of the NOE off-diagonal cross-peaks was problematic, as the intensities of the NOE peaks were to too weak to be reliably measured in the case where the population of the minor conformation was small (Trp27, Tyr28, Cys36, Phe39 in Table 4). Therefore, relative concentrations of the conformations were determined from volumes of well-separated diagonal peaks for the ^{N}H and $^\alpha$H resonances (compare also Fig. 9). The volumes of cross-peaks were estimated using the procedure described in [52]. These volumes are directly proportional to the populations of the multiple forms provided that the forms have similar relaxation properties and coupling constants. This should be the case for domain C5, as it is not expected that the 3D global folds and secondary structures of the forms differ so dramatically that, for example, Tyr27 (Table 4), which is in the β strand (large coupling constant $^3J_{HN\alpha}$) in the major form, would be in the helix in the minor form (the small coupling constant). The manifestation that the two forms have similar properties was provided by the fact that, for each resonance in Table 4, the sum of the measured intensities for the major and minor conformations was equal to the intensity of the cross peaks, which did not show the presence of multiple conformations.

Stereospecific assignments

For stereospecific assignments, $^3J_{\alpha\beta}$ coupling constants were extracted from the DQF COSY spectrum in D_2O. Stereospecific assignments of $C^\beta H$ protons, $C^\gamma H$ protons of arginines, and methyl groups of valines and leucines were obtained either using procedures described by Wagner *et al.* [19] and Hyberts *et al.* [33] or during the refinement of the structures.

Interproton distance constraints

The intensities of 2D NOE peaks derived from the 2D NOESY spectra recorded in H_2O and D_2O at pH 4.3 and 15°C with a mixing time of 46 ms were determined from volume integrals and converted into distance constraints according to the procedures described earlier [38,39]. Additional constraints were obtained for residues Thr19, Cys20 and Cys44 in the 2D NOESY spectra in H_2O and D_2O with a mixing time of 126 ms at pH 2.7 and 10°C. The distance bounds of the distance constraints were set: d −0.2 Å/+0.3 Å for distance constraints of 2.3 Å–2.6 Å, d ±0.4 Å for distance constraints 2.7 Å–3.4 Å, and d −0.6 Å/+0.8 Å for distance constraints of 3.5 Å-4.2 Å. All protons were explicitly defined in the dynamical simulated annealing calculations; however, for ring protons and protons that were not stereospecifically assignable, additional terms were added to the upper bounds that correspond to the pseudoatom corrections introduced by Wüthrich [23].

Torsion angle constraints

The distance constraints were supplemented with backbone ϕ torsion angle constraints derived from the $^3J_{HN\alpha}$ coupling constant data measured from the COSY spectrum (Table 1 and the supporting information). To increase resolution the spectrum was folded in F1 to 6.4 ppm and in F2 to 11.6 ppm. In addition, the rows with maximum intensity for an appropriate cross peak were extracted from the 2D spectrum, after inverse Fourier transformation and zero filling to 32K, they were again Fourier transformed. This resulted in a resolution of 0.18 Hz pt^{-1}. The apparent $^3J_{HN\alpha}$ coupling constants measured from the splittings of the NH–C^αH cross-peaks were corrected for line width using a method of Kim and Prestegard [53]. The α-helical segments, for which a complete set of the $H^N(i)$–$H^N(i+1)$, $H^\alpha(i)$–$H^\beta(i+3)$, $H^\alpha(i)$–$H^N(i+3,4)$ NOE cross-peaks and slow proton amide exchange rates were observed, were further constrained by restricting the backbone torsion angles to −35°±20° [54]. A constraint on the torsion angle in the fragment X-Pro was also added using the procedure described by Arseniev *et al.* [55]. In total 10 such α-helical and X-Pro constraints were used in the calculations.

Hydrogen bonds and disulfide bridge constraints

The initial structures were calculated without the presence of any disulfide bridges; S–S bonds were introduced after these structures confirmed the expected S–S pattern [23,39]. In the final structures the S–S bridges were treated as normal bonds.

Hydrogen bonds established by slowly exchanging amide protons and proximity in preliminary 3D structures were introduced during the structure refinement using an allowed distance range of 1.6 Å–2.5 Å for proton–oxygen distances and 2.2 Å–3.0 Å for amide nitrogen–oxygen distances.

Structure calculations

Structure calculations were carried out with the program X-PLOR 3.1 [56,57] using essentially the basic protocol described previously [39,58]. Starting structures were an extended strand, randomized coordinates, and the structures obtained from the distance geometry method. At the initial stage of the structure calculation, only those torsion constraints were used which corresponded to $^3J_{HN\alpha}$ coupling constants larger than 8 Hz (ϕ=−120°±40°) [23,34]. The rest of the ϕ angle constraints, derived for $^3J_{HN\alpha}$ coupling constants equal to 8 Hz and ≥6 Hz [39] were introduced only during the refinement procedure. These ϕ constraints were applied for residues for which ϕs were close to the mean value (within ±20°) and fulfilled the Karplus relationship [23] in structures calculated without the dihedral constraints. The minimum ranges employed for ϕ and χ1 were ±30° and ±40°, respectively.

Supplementary material available

The supplementary material available from the authors includes the coordinates of the final structures of C5 and the complete list of NMR constraints including the NOEs of the minor conformational form. The atomic coordinates have been deposited in the Brookhaven Protein Data Bank.

Acknowledgements

We thank Richard Engh, Robert Huber and Wolfram Bode for stimulating discussions. This work was supported by the Deutsche Forschungsgemeinschaft (Projects B11 and C4, Sonderforschungbereich 266 of the Technical University of Munich).

References

1. Chu, M.-L., *et al.*, & Timpl, R. (1990). Mosaic structure of globular domains in the human type VI collagen α3 chain: similarity to von Willebrand factor, fibronectin, actin, salivary proteins and aprotinin type protease inhibitors. *EMBO J.* **9**, 385–393.
2. Mayer, U., *et al.*, & Timpl, R. (1994). Recombinant expression and properties of Kunitz-type protease inhibitor module from human type VI collagen α3(VI) chain. *Eur. J. Biochem.* **225**, 573–580.
3. Kitaguchi, N. Takahashi, Y., Tokushima, Y., Shiojiri, S. & Ito, H. (1988). Novel precursor of Alzheimer's disease amyloid protein shows protease inhibitory activity. *Nature* **331**, 530–532.
4. Ponte, P., *et al.*, & Cordell, B. (1988). A new A4 amyloid mRNA contains a domain homologous to serine proteinase inhibitors. *Nature* **331**, 525–527.
5. Tanzi, R.E., McClatchey, A.I., Lamperti, E.D., Villa-Komaroff, L., Gusella, J.F. & Neve, R.-L. (1988). Protease inhibitor domain encoded by an amyloid protein precursor mRNA associated with Alzheimer's disease. *Nature* **331**, 528–530.
6. Deisenhofer, J. & Steigemann, W. (1975). Crystallographic refinement of the structure of bovine pancreatic trypsin inhibitor at 1.5 Å resolution. *Acta Cryst. B* **31**, 238–250.
7. Wlodawer, A., Walter, J., Huber, R. & Sjölin, L. (1984). Structure of bovine pancreatic trypsin inhibitor. *J. Mol. Biol.* **180**, 301–329.
8. Wlodawer, A., Nachman, J., Gilliland, G. L., Gallagher, W. & Woodward, C. (1987). Structure of form III crystals of bovine pancreatic trypsin inhibitor. *J. Mol. Biol.* **198**, 469–480.
9. Wlodawer, A., Deisenhofer, J. & Huber, R. (1987). Comparison of two highly refined structures of bovine pancreatic trypsin inhibitor. *J. Mol. Biol.* **193**, 145–156.
10. Berndt, K.D., Güntert, P., Orbons, L.P.M. & Wüthrich, K. (1992). Determination of a high-quality nuclear magnetic resonance solution structure of bovine pancreatic trypsin inhibitor and comparison with three crystal structures. *J. Mol. Biol.* **227**, 757–775.
11. Skarzynski, T. (1992). Crystal structure of α-dendrotoxin from the green mamba venom and its comparison with the structure of bovine pancreatic trypsin inhibitor. *J. Mol. Biol.* **224**, 671–683.
12. Hynes, T.R., Randal, M., Kennedy, L.A., Eigenbrot, C. & Kossiakoff, A.A. (1990). X-ray crystal structure of the protease inhibitor domain of Alzheimer's amyloid β-protein precursor. *Biochemistry* **29**, 10018–10023.
13. Kwong, P.D., McDonald, N.Q., Sigler, P.B. & Hendrickson, W.A. (1995). Structure of β_2-bungarotoxin. Potassium channel binding by Kunitz modules and targeted phospholipase action. *Structure* **3**, 1109–1119.
14. Berndt, K.D., Güntert, P. & Wüthrich, K. (1993). Nuclear magnetic resonance solution structure of dendrotoxin K from venom of *Dendraoapsis polylepis polylepis*. *J. Mol. Biol.* **234**, 735–750.
15. Lancelin, J.-M., Foray, M.-F., Poncin, M., Hollecker, M. & Marion, D. (1994). Proteinase inhibitor homologues as potassium channel blockers. *Nat. Struct. Biol.* **1**, 246–250.
16. Moses, E. & Hinz, H.-J. (1983). Bovine pancreatic trypsin inhibitor has unusual thermodynamic stability parameters. *J. Mol. Biol.* **170**, 765–776.
17. Schwarz, H., Hinz, H.-J., Mehlich, A., Tschesche, H. & Wenzel, H.R. (1987). Stability studies on derivates of the bovine pancreatic trypsin inhibitor. *Biochemistry* **26**, 3544–3551.
18. Wagner, G. & Wüthrich, K. (1982). Sequential resonance assignment in protein ^{1}H nuclear magnetic resonance spectra bovine pancreatic trypsin inhibitor. *J. Mol. Biol.* **155**, 347–366.
19. Wagner, G., Braun, W., Havel, T.F., Schaumann, T., Go, N. & Wüthrich, K. (1988). Protein structures in solution by NMR and distance geometry. The polypeptide fold of the basic trypsin inhibitor determined using two different algorithms, DISGEO and DISMAN. *J. Mol. Biol.* **196**, 611–639.
20. Wagner, G. & Nirmala, N.R. (1989). Studies of protein dynamics by heteronuclear NMR: individual ^{13}C relaxation times and evidence for multiple conformations in the reactive site of BPTI. *Chem. Scripta* **29A**, 27–30.
21. Otting, G., Liepinsh, E. & Wüthrich, K. (1993). Disulfide bond isomerization in BPTI and BPTI(G36S): an NMR study of correlated mobility in proteins. *Biochemistry* **32**, 3571–3582.
22. Arnoux, *et al.*, & Ducruix, A. (1995). The 1.6 Å structure of Kunitz-type domain from the α3 chain of human type VI collagen. *J. Mol. Biol.* **246**, 609–617.
23. Wüthrich, K. (1986). *NMR of Proteins and Nucleic Acids.* Wiley, New York.
24. Jeener, J., Meier, B.H., Bachman, P. & Ernst, R.R. (1979). Investigation of exchange processes by two-dimensional NMR spectroscopy. *J. Chem. Phys.* **71**, 4546–4553.
25. Macura, S. & Ernst R.R. (1980). Elucidation of cross relaxation in liquids by two-dimensional NMR spectroscopy. *Mol. Phys.* **41**, 95–117.
26. Braunschweiler, L. & Ernst, R.R. (1983). Coherence transfer by isotropic mixing: Application to proton correlation spectroscopy. *J. Magn. Reson.* **53**, 521–528.
27. Davis, D.G. & Bax, A. (1985). Assignment of complex ^{1}H NMR spectra via two-dimensional homonuclear Hartmann–Hahn spectroscopy. *J. Am. Chem. Soc.* **107**, 2820–2821.
28. Aue, W.P., Bartholdi, E. & Ernst, R.R. (1976). Two dimensional spectroscopy: application to nuclear magnetic resonance. *J. Chem. Phys.* **64**, 2229–2246.
29. Rance, M., Srensen, O.W., Bodenhausen, G., Wagner, G., Ernst, R.R. & Wüthrich, K. (1983). Improved spectral resolution in COSY ^{1}H NMR spectra of proteins via double quantum filtering. *Biochem. Biophys. Res. Commun.* **117**, 479–485.
30. Bothner-By, A.A., Stephens, R.L., Lee, J., Warren, C.D. & Jeanloz, R.W. (1984). Structure determination of a tetrasaccharide: transient nuclear Overhauser effects in the rotating frame. *J. Am. Chem. Soc.* **106**, 811–813.
31. Bax, A. & Davis, D.G. (1985). Practical aspects of two-dimensional transverse NOE spectroscopy. *J. Magn. Reson. A* **63**, 207–213.
32. Wüthrich, K., Billeter, M. & Braun, W. (1984). Polypeptide secondary structure determination by nuclear magnetic resonance observation of short proton-proton distances. *J. Mol. Biol.* **180**, 715–740.
33. Hyberts, S.G., Mäki, W. & Wagner, G. (1987). Stereospecific assignments of side-chain protons and characterization of torsion angles in Eglin c. *Eur. J. Biochem.* **164**, 625–635.
34. Pardi, A., Billeter, M. & Wüthrich, K. (1984). Calibration of the angular dependence of the amide proton–Cα proton coupling constants, $^3J_{HN\alpha}$, in a globular protein. *J. Mol. Biol.* **180**, 741–751.
35. Wishart, D.S., Sykes, B.D. & F.M. Richards (1992). The chemical shift index: a fast and simple method for the assignment of protein secondary structure through NMR spectroscopy. *Biochemistry* **31**, 1647–1651.
36. Otting, G. & Wüthrich, K. (1989). Studies of protein hydration in aqueous solution by direct NMR observation of individual protein-bound water molecules. *J. Am. Chem. Soc.* **111**, 1871–1875.
37. Bode, W. & Huber, R. (1992). Natural protein proteinase inhibitors and their interaction with proteinases. *Eur. J. Biochem.* **204**, 433–451.
38. Holak, T.A., Bode, W., Huber, R., Otlewski, J. & Wilusz, T. (1989). Nuclear magnetic resonance solution and X-ray structures of squash trypsin inhibitor exhibit the same conformation of the proteinase binding loop. *J. Mol. Biol.* **210**, 649–654.
39. Holak, T.A., Gondol, D., Otlewski, J. & Wilusz, T. (1989). Determination of the complete 3-dimensional structure of the trypsin-inhibitor from squash seeds in aqueous-solution by nuclear magnetic-resonance and a combination of distance geometry and dynamical simulated annealing. *J. Mol. Biol.* **210**, 635–648.
40. Mühlhahn, P., *et al.*, & Holak, T.A. (1994). Structure of leech derived tryptase inhibitor (LDTI-C) in solution. *FEBS Lett.* **355**, 290–296.
41. Hyberts, S.G., Goldberg, M.S., Havel, T.F. & Wagner, G. (1992). The solution structure of eglin c based on measurements of many NOEs and coupling constants and its comparison with X-ray structures. *Protein Sci.* **1**, 736–751.40
42. Chiche., L., Gaboriaud, C., Heitz., A. Mornon, J.-P., Castro, B. & Kollman,P. A. (1989). Use of restrained molecular dynamics in water to determine three-dimensional protein structure: prediction of the three-dimensional structure of *Ecballium elaterium* trypsin inhibitor II. *Proteins* **6**, 405–417.
43. Jardetzky, O. (1980). On the nature of molecular conformations inferred from high-resolution NMR. *Biochem. Biophys. Acta* **621**, 227–232.
44. Kim, Y. & Prestegard, J.H. (1989). A dynamic model for the structure of acyl carrier protein in solution. *Biochemistry* **28**, 8792–8797.
45. Ernst, R.R., Bodenhausen, G. & Wokaun, A. (1987). *Principles of NMR in One and Two Dimension.* Clarendon Press, Oxford.
46. Marion, D. & Wüthrich, K. (1983). Application of phase sensitive two dimensional correlated spectroscopy (COSY) for measurements of ^{1}H-^{1}H spin-spin coupling constants in proteins. *Biochem. Biophys. Res. Commun.* **113**, 967–974.
47. Bax, A. & Davis, D.G. (1985). MLEV-17-based two-dimensional homonuclear magnetization transfer spectroscopy. *J. Magn. Reson.* **65**, 355–360.
48. Rance, M. (1987). Improved techniques for homonuclear rotating-frame and isotropic mixing experiments. *J. Magn. Reson.* **74**, 557–564.

49. Plateau, P. & Gueron, M. (1982). Exchangeable proton NMR without base-line distortion using new strong-pulse sequences. *J. Am. Chem. Soc.* **104**, 7310–7311.
50. Gueron, M., Plateau, P. & Decorps, M. (1991). Solvent signal suppression in NMR. *Prog. NMR Spectr.* **23**, 135–209.
51. Sklenar, V. & Bax, A. (1987). A new water suppression technique for generating pure-phase spectra with equal excitation over a wide bandwidth. *J. Magn. Reson.* **75**, 378–383.
52. Holak, T.A., Scarsdale, N.J., & Prestegard, J.H. (1987). A simple method for quantitative evaluation of cross-peak intensities in two-dimensional NOE spectra. *J. Magn. Reson.* **74**, 546–549.
53. Kim, Y. & Prestegard, J.H. (1989). Measurement of vicinal couplings from cross peaks in COSY spectra. *J. Magn. Reson.* **84**, 9–13.
54. Holak, T.A., Nilges , M. & Oschkinat, H. (1989). Improved strategies for the determination of protein structures from NMR data: the solution structure of acyl carrier protein. *FEBS Lett.* **242**, 218–224.
55. Arseniev, A., *et al.*, & Wüthrich, K. (1988). Three-dimensional structure of rabbit liver [Cd_7] metallothionein-2a in aqueous solution determined by nuclear magnetic resonance. *J. Mol. Biol.* **201**, 637–657.
56. Brünger, A.T. (1993). *X-PLOR Version 3.1 Manual.* Yale University, New Haven, CT.
57. Brünger, A.T. & Nilges, M. (1993). Computational challenges for macromolecular structure determination by X-ray crystallography and solution NMR spectroscopy. *Q. Rev. Biophys.* **26**, 49–125.
58. Holak, T.A., Kearsley, S.K., Kim, Y. & Prestegard, J.H. (1988). 3-dimensional structure of acyl carrier protein determined by NMR pseudoenergy and distance geometry calculations. *Biochemistry* **27**, 6135–6142.
59. Evans, S.V. (1993). *SETOR: hardware-lighted three-dimensional solid model representations of macromolecules Version 4.13.0.* University of British Columbia, Vancouver.
60. Vriend, G. (1991). *WHAT IF.* EMBL, Heidelberg.