Articles

^{1}H 2D NMR and Distance Geometry Study of the Folding of *Ecballium elaterium* Trypsin Inhibitor, a Member of the Squash Inhibitors Family†

Annie Heitz, Laurent Chiche, Dung Le-Nguyen, and Bertrand Castro*
Centre CNRS-INSERM de Pharmacologie-Endocrinologie, Rue de la Cardonille, 34094 Montpellier Cedex 2, France
Received July 29, 1988; Revised Manuscript Received October 13, 1988

ABSTRACT: The solution conformation of synthetic *Ecballium elaterium* trypsin inhibitor II, a 28-residue peptide with 3 disulfide bridges, has been studied by ^{1}H 2D NMR measurements. Secondary structure elements were determined: a miniantiparallel β-sheet Met 7–Cys 9 and Gly 25–Cys 27, a β-hairpin 20–28 with β-turn 22–25, and two tight turns Asp 12–Cys 15 and Leu 16–Cys 19. A set of interproton distance restraints deduced from two-dimensional nuclear Overhauser enhancement spectra and 13 ϕ backbone torsion angles restraints were used as the basis of three-dimensional structure computations including disulfide bridges arrangement by using distance geometry calculations. Computations for the 15 possible S–S linkage combinations lead to the proposal of the array 2–19, 9–21, 15–27 as the most probable structure for EETI II.

The Cucurbitacaea family has recently provided an interesting series of powerful trypsin inhibitors, the squash family, characterized by their very short length, about 30 residues (Hojima et al., 1982; Leluk et al., 1983; Pham et al., 1985). The sequences of a dozen of these inhibitors have been published recently (Polonowski et al., 1980; Wilusz et al., 1983; Joubert, 1984; Wieczorek et al., 1985). One of them has been synthesized (Kupryszewski & Rolka, 1986). Some structural indications concerning the secondary structure and the three disulfide bridges arrangement have been published (Siemion et al., 1984; Hider et al., 1987), based either on the structural analogy with wheat germ agglutinin or on CD[1] spectroscopic data. However, no complete NMR study was presented to support the structural assignments presented.

EETI II (Favel et al., 1988), a trypsin inhibitor extracted from *Ecballium elaterium*, a member of the cucurbitacaea family widespread around the Mediterranean area, has been recently isolated and the primary sequence determined. EETI II has a large homology with the other members of the squash family, as seen in the comparison presented in Figure 1. It is peculiarly characterized by the very large amount of glycine residues and the replacement of the very constant tyrosine by a phenylalanine. A 28-residue peptide bearing the sequence of EETI II has been synthesized, the air cyclization of which gave a product having the same affinity constant to trypsin (5×10^{-11}) as the natural product. The synthesis yielded the rather large quantities needed for structural studies by NMR.

Material and Methods

Synthesis. Synthetic EETI II was obtained at a 50-mg scale by solid-phase synthesis; the details of the synthesis, cyclization, and purification will be published elsewhere.

NMR Measurements. For the NMR experiments about 20 mg of peptide was dissolved in 0.4 mL of a mixture of $H_2O/^2H_2O$ (95/5) and 20 or 10 mg in 100% ^2H_2O at pH 2. The pH meter reading was used without correction for isotope effect. In ^2H_2O two sets of experiments were done. In the first set the acquisition was started just after the dissolution of the peptide; in the second set acquisition was started after several lyophilizations of the peptide in ^2H_2O to remove all exchangeable protons.

All 2D ^{1}H NMR spectra were recorded at 32 °C on a Bruker WM 360 WB spectrometer equipped with an Aspect 3000 computer. COSY (Aue et al., 1976; Nagayama et al., 1980), RELAYED-COSY (Wagner, 1983), DQF-COSY (Rance et al., 1984), and NOESY (Bodenhausen et al., 1984; Otting et al., 1986) were recorded as described previously, using quadrature detection in both dimensions. The spectra were obtained with 2048 points for each value of t_1. There were 512 t_1 measurements, except in DQF-COSY, where 800 t_1 values were used. NOESY spectra were obtained with 100-, 200-, and 300-ms mixing time. For experiments in H_2O the solvent was eliminated by continuous low-power irradiation applied all along the sequence and stopped prior to the acquisition. The carrier frequency was centered on the water signal, and a recycle delay of 1 s was used in all spectra. Prior to Fourier transformation, the time domain data were multiplied with phase-shifted sine bell window ($\pi/8$ in ω_2 and $\pi/4$ in ω_1) for phase-sensitive experiments and with no phase-shifted sine bell for magnitude spectra. The digital resolution for NOESY spectra was 3.5 Hz/point in t_2 dimension and 7.0 Hz/point in t_1 dimension. For DQF-COSY spectra the digital resolution was 3.5 Hz/point in both dimensions.

NH–CαH coupling constants were determined on a 1D spectrum with a digital resolution of 0.2 Hz/point.

Computing Procedure. The DISGEO distance geometry program (Havel, 1985; Havel & Wüthrich, 1984; Havel et al., 1983) was used for the calculation of 3D molecular structures

† This work was partially supported by the French Ministère de la Recherche et de l'Enseignement Supérieur (Project No. 87 C 0573).
* Correspondence should be addressed to this author.

[1] Abbreviations: 2D, two-dimensional; CD, circular dichroism; CMTI, *Cucurbita maxima* trypsin inhibitor; COSY, two-dimensional correlated spectroscopy; CPTI, *Cucurbita pepo* trypsin inhibitor; CSTI, *Cucumis sativa* trypsin inhibitor; DQF-COSY, double quantum filtered homonuclear correlation spectroscopy; EETI, *Ecballium elaterium* trypsin inhibitor; MRTI, *Momordica repens* trypsin inhibitor; NMR, nuclear magnetic resonance; NOE, nuclear Overhauser effect; NOESY, two-dimensional nuclear Overhauser enhancement spectroscopy, RELAYED-COSY, two-dimensional relayed coherence transfer spectroscopy; TSP-d_4, sodium 3-(trimethylsilyl)[2,2,3,3-$^2H_4$]propionate.

0006-2960/89/0428-2392$01.50/0

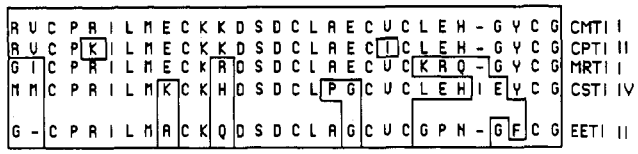

FIGURE 1: Alignment of the primary sequences of EETI II, CMTI I (Wilusz et al., 1983), CPTI II, CSTI IV (Wieczorek et al., 1985), and MRTI I (Joubert, 1984). The boxed residues are those that differ from the sequence of CMTI I.

compatible with the constraints deduced from NMR data. Substructures containing the atoms N, Hα, C, Cβ, and Cγ and the pseudoatom QR in place of the aromatic ring were first computed, subsequent calculations leading to complete structures. By use of different initial random number seed, different possible conformations were generated.

The computed structures were visualized and compared by using the MANOSK molecular graphics package (Vaney et al., 1985) running on a Microvax-hosted PS330 Evans and Sutherland graphic display.

For the best structures, hydrogen atoms were fixed in theoretical positions, and listing of distances between hydrogen atoms of less than a threshold value were generated. This allows careful comparison between computed structures and NMR data.

^1H–^1H Distance Constraints Derived from the NMR Measurements for Use in the DISGEO *Program.* Distance constraints were derived from the NOESY spectra recorded with 100-, 200-, and 300-ms mixing times. Among the 105 assigned cross peaks, 15 corresponding to sequential NOEs between Cβ_i and NH$_{i+1}$ were not used because, in the absence of diastereotopic assignment, the derived distance constraints were looser than the ones fixed by the covalent structure.

The calibration between NOESY cross-peak heights and ^{1}H–^{1}H distance constraints was established in the manner already described (Williamson et al., 1985) on a 100-ms mixing time NOESY spectra. NOEs that are not observed on a 100-ms time scale but are detected on 200- or 300-ms mixing time NOESY spectra are considered weak ones.

For medium- and long-range distances, we empirically assign upper bounds of 2.5 Å to strong NOEs, 3.5 Å to medium NOEs, and 4.0 Å to weak NOEs. If the connectivity involved side-chain protons, the upper bounds were raised respectively to 3.0, 4.0, and 5.0 Å, to accommodate with the higher assumed internal mobility (Williamson et al., 1985); however, these corrections were not applied to cysteines that are part of disulfide bridges.

For sequential connectivity, we used upper limits of 2.5, 3.0, and 3.5 Å for $d_{\alpha N}$ and 3.0 (strong and medium NOEs) and 4.0 (weak NOEs) for d_{NN}. All the distance constraints thus defined were corrected to account for the use of pseudoatoms when necessary (Wüthrich et al., 1983).

Additional Distance Constraints. Disulfide bridges were imposed by adding constraints of 2.0–2.1 Å on the S_i–S_j bond and of 3.0–3.1 Å on the S_i–Cβ_j and S_j–Cβ_i distances across the bridge. Moreover, a value of about ±80–100° was imposed to the χ_3 dihedral angle (Richardson, 1981) by a distance constraint of 3.75–3.95 Å between the two Cβ_i–Cβ_j.

Hydrogen bonds between amide protons and carbonyl oxygen that could be inferred from slow amide proton exchange rates and at least two observed NOEs were imposed by adding constraints of 1.8–2.0 Å on the NH–O distances and of 2.7–3.0 Å on the N–O distances.

Dihedral Angle Constraints. Trans planar geometry for all the peptide links was assumed and imposed, and the planarity of the aromatic ring of the Phe 26 was achieved by forcing all the corresponding dihedral angles to zero value.

Some ϕ angles were constrained in the range −90° to −40° ($^3J_{NH-C\alpha H} < 4$ Hz) or −160° to −80° ($^3J_{NH-C\alpha H} > 8.5$ Hz). In addition, eight χ_1 angles (6 Cys, Phe, and Asp 14) for which a preferential rotamer was determined by examination of intraresidue coupling constants and NOEs strength were constrained in the corresponding range (Wagner et al., 1987); all these angle constraints were ensured by fixing three dihedral angles around the rotating bond (Havel & Wüthrich, 1985).

Results and Discussion

Assignment of the ^1H NMR Spectrum. The elucidation of a peptide or protein structure by NMR needs the sequential assignment of resonances. This was accomplished by well-established methods (Wagner & Wüthrich, 1982; Billeter et al., 1982). The identification of amino acid spin systems was first established by the mean of direct and relayed through-bond connectivities (COSY and RELAYED-COSY) followed by sequential resonance assignments using through-space NOE connectivities (NOESY) essentially between NH, CαH, and CβH protons.

Spin Systems Identification. In H_2O, on 1D spectrum, we observed 25 resonances (in the range 9.6–7.3 ppm) corresponding to all expected amide protons. The 2D COSY spectra showed 26 NH–CαH cross peaks (8 corresponding to 4 Gly residues). This led to the identification of 22 NH–CαH connectivities (3 CαH being obscured by H_2O residual signal; no NH–CαH cross peaks were detected for 3 residues). In the same way 18 NH–CβH connectivities were observed in the 2D RELAYED-COSY.

In ^2H_2O the CαH–CβH cross peaks of the DQF-COSY could be divided into three regions on the 2D contour plot map (Neuhaus et al., 1985): the first one corresponds to Gly and Ser residues (Figure 2A), the second one to Cys, Asp, Asn, and Phe residues with typical AMX patterns (Figure 2B), and the third one to Ala, Val, Leu, Ile, Met, Gln, Pro, Arg, and Lys residues (Figure 2C). Gly resonances were easily recognized due to their typical AX spectrum with a large doublet splitting arising from their large geminal coupling constant. Ser residue was identified as an AMX spin system with all chemical shifts greater than 3.8 ppm. The other AMX systems were identified by the fact that the chemical shifts of the β protons were at higher field than those of serine and at lower field than those of aliphatic side chains: generally between 3.7 and 2.5 ppm. Phe and Asn residues were distinguished from the other AMX spin systems by the NOEs observed between their CβH and aromatic protons for Phe and NδH$_2$ protons for Asn. The sequential resonance assignment was necessary to differentiate the six Cys and two Asp residues. In the third contour plot region Ala residue was identified by a cross peak with a methyl at about 1.4 ppm. For Met and Gln residues two spin systems were attributed, but these two residues were only distinguished by the sequential assignment. The Ile and Val spin systems were identified by the fact that they possess only one β proton that is coupled with one γCH$_3$ for Ile and two γCH$_3$ for Val (see Figure 2C).

Sequential Assignments. The short-range NOEs involving the NH, CαH, and CβH protons as well as CδH protons of proline are summarized in Figure 3, and the resulting resonance assignments are listed in Table I. A part of the sequential assignment is presented in Figure 4. Since the peptide contains unique Ala, Val, Ile, Ser, and Phe residues, any of these could be used as a starting point.

Starting from Phe 26, the assignment proceeded via $d_{\alpha N}$ connectivities until Gly 28 in one direction and Gly 25 in the reverse direction. Similarly, starting from Val 20 and Ile 5

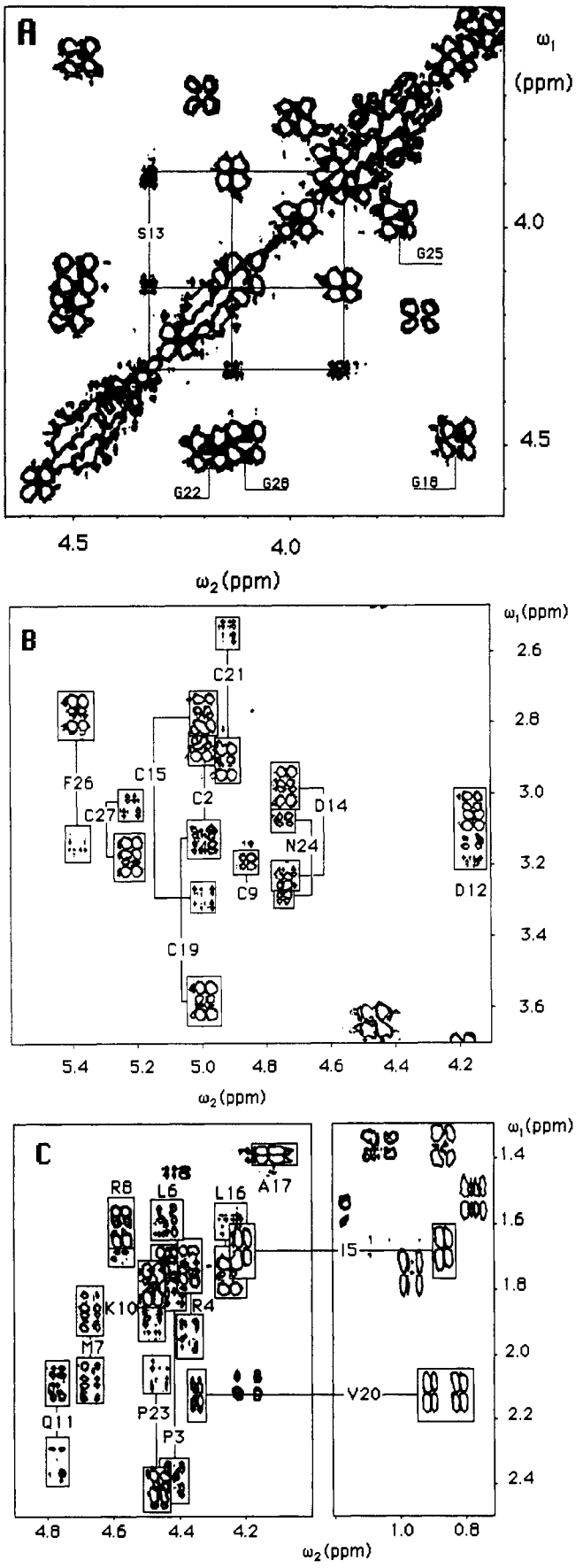

FIGURE 2: Expansion of the region containing the Hα–Hα and Hα–Hβ cross peaks in a phase-sensitive DQF-COSY spectrum of a 8 mM solution of EETI II in ^2H_2O at 32 °C at pH 2. Identification of the spin systems of glycines and serine (A), CαH–CβH_2 AMX systems other than serine (B), Cα–CβH_2 spin systems of prolines, methionine, glutamine, lysine, arginines, and leucines and CαH–CβH–CγH_3 connectivities of valine and isoleucine (C).

Table I: Resonance Assignments for EETI II Protons[a]

position	residue	NH	Hα	Hβ	others
1	Gly	–	3.91–3.82	–	–
2	Cys	8.69	5.02	3.12–2.84	–
3	Pro	–	4.42	2.38–1.79	γCH_2 2.09, δCH_2 4.20–3.71
4	Arg	8.58	4.37	1.93–1.72	γCH_2 1.69, δCH_2 3.25, εNH 7.18
5	Ile	7.54	4.22	1.68	γCH_2 1.38–1.05, γCH_3 0.86, δCH_3 0.87
6	Leu	8.71	4.45	1.72–1.57	γCH 1.53, δCH_3 0.79–0.76
7	Met	9.12	4.68	2.07–1.87	γCH_2 2.68–2.52, SCH_3 2.19
8	Arg	8.71	4.58	1.70–1.62	γCH_2 1.58–1.18, δCH_2 2.96–2.90, εNH 7.05
9	Cys	8.26	4.85	3.17	–
10	Lys	8.34	4.48	1.89–1.79	γCH_2 1.44, δCH_2 1.69, εCH_2 3.00, ζNH_2 7.55
11	Gln	7.84	4.75	2.35–2.08	γCH_2 2.23
12	Asp	9.55	4.18	3.17–3.06	–
13	Ser	8.46	4.33	4.13–3.89	–
14	Asp	7.72	4.80	3.31–2.95	–
15	Cys	7.79	5.02	3.28–2.75	–
16	Leu	8.69	4.27	1.76–1.60	γCH 1.76, δCH_3 0.97–0.94
17	Ala	8.10	4.13	1.40	–
18	Gly	8.61	4.48–3.63	–	–
19	Cys	8.23	4.99	3.56–3.11	–
20	Val	8.80	4.34	2.11	γCH_3 0.90–0.82
21	Cys	8.49	4.93	2.90–2.53	–
22	Gly	8.37	4.50–4.19	–	–
23	Pro	–	4.48	2.40–2.05	γCH_2 2.10, δCH_2 3.84–3.74
24	Asn	8.12	4.75	3.26–3.03	δNH_2 7.70–7.00
25	Gly	8.33	3.99–3.76	–	–
26	Phe	7.33	5.39	3.13–2.76	2,6H 7.06, 3,4,5H 7.29
27	Cys	8.87	5.22	3.17–3.02	–
28	Gly	9.38	4.55–4.15	–	–

[a] ^1H chemical shifts of EETI II in H_2O at pH 2, 32 °C. Values were measured ±0.02 ppm relative to TSP-d_4 as internal reference. Note that there were slight differences in the chemical shifts of one CαH proton of Gly 28 and one CβH proton of Asp 14 compared to EETI II in ^2H_2O shown in Figure 2. This is because the pHs of the two samples were not exactly the same.

residues led to sequential assignment of the segments 15–22 and 3–11, respectively. A strong NOE observed between CδH protons of Pro 3 and an CαH proton at 5.02 ppm led to the attribution of Cys 2 and Gly 1. The CδH protons of the second proline gave a NOE with two CαH protons and allowed the assignment of Pro 23 and the confirmation of Gly 22 and indicated that the peptide bond 22–23 is trans as the 2–3 one.

In H_2O, the CαH protons of Gln 11, Asp 14, and Asn 24 being obscured by the residual water signal, the α-N cross peaks 11–12, 14–15, and 24–25 were not observed in the NOESY spectra; however, we detected NOEs between CαH Pro 23 and NH Asn 24 and between CαH Ser 13 and NH Asp 14. No αN 12–13 NOE was detected, but βN 12–13 and NN 12–13 NOEs led to the unambiguous assignment of Asp 12 (the sequence containing only one Ser residue in position 13). In addition, on a NOESY spectra made on a freshly prepared ^2H_2O solution, the slow exchanging rates of Cys 15 and Gly 25 amide protons led to the observation of αN cross peaks 14–15 and 24–25.

Secondary Structure. In addition to the short-range NOEs, Figure 3 provides a summary of the slowly exchanging amide protons and the $^3J_{NH-C\alpha H}$ coupling constants measured on the 1D spectrum in H_2O. All NOEs observed in the NOESY

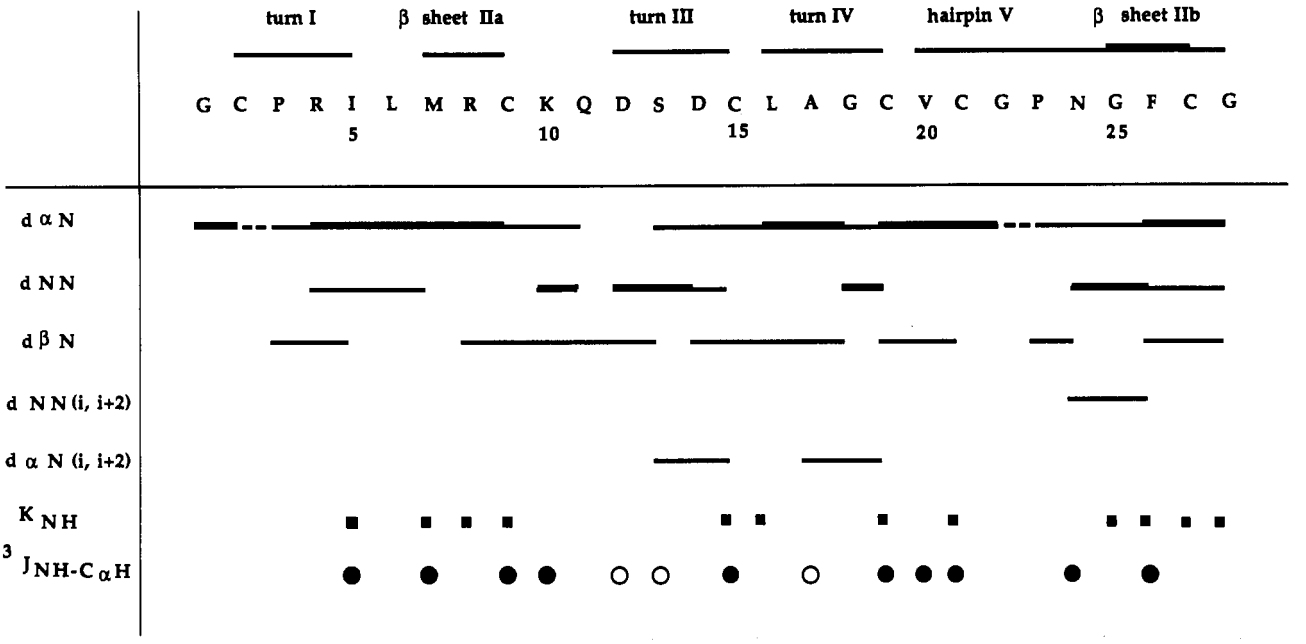

FIGURE 3: Sequence of EETI II together with a summary of the observed short-range NOEs involving the NH, CαH, and CβH protons as well as the CδH protons of proline residues (dotted lines). NH protons that were still present after 24 h of dissolving the peptide in ^2H_2O are indicated by solid squares (■). $^3J_{NH-C\alpha H}$ >8.5 Hz and <4 Hz are represented by solid circles (●) and open circles (○), respectively. No spin–spin coupling constants were given for glycines. The secondary structure elements are indicated at the top of the figure.

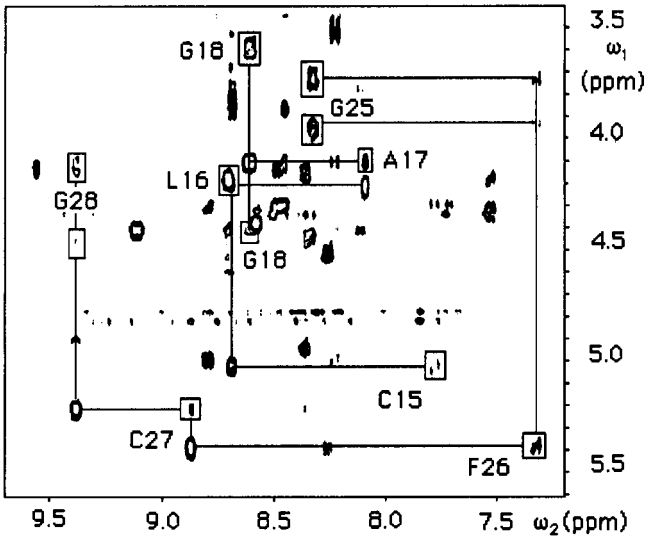

FIGURE 4: Contour plot of the NH–CαH region of the 300-ms ^1H NOESY spectrum of EETI II in H_2O. Connectivities from Cys 15 to Gly 18 and from Gly 25 to Gly 28 were followed via $d_{\alpha N}$ NOEs.

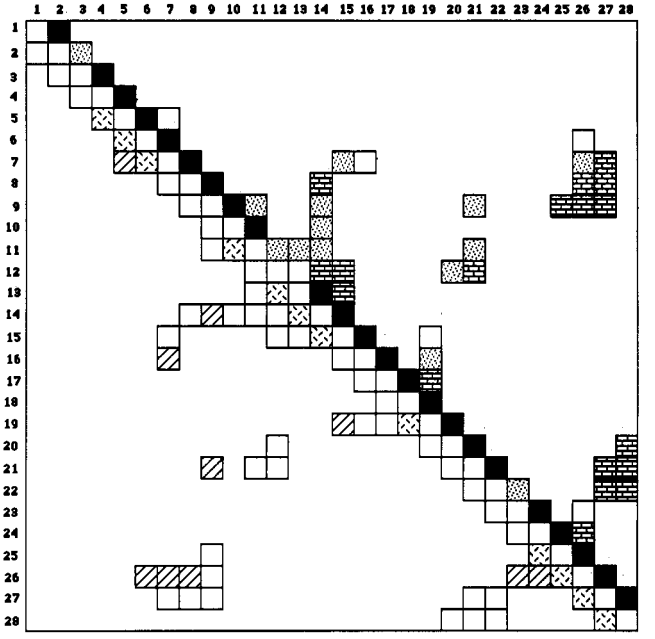

FIGURE 5: Diagonal plot of short-, medium-, and long-range NOEs in EETI II derived from 300-ms NOESY spectra obtained in H_2O and ^2H_2O at pH 2 and 32 °C. Sequential NOEs corresponding to short $d_{\alpha N}$ (solid box) and d_{NN} (latticework box) distances are plotted above and below the diagonal, respectively. Nonsequential interresidues NOEs, backbone–backbone (brickwork box) and side chain–side chain (hatched box), are reported above the diagonal, and backbone–side chain NOEs (stippled box) are plotted below the diagonal.

spectra in H_2O and ^2H_2O were summarized in a diagonal plot in Figure 5. The presence of CαH–CαH NOE between Phe 26 and Arg 8 and the backbone proton–backbone proton NOEs presented on the diagonal plot in Figure 5, where several NOEs form a continuous line perpendicular to the diagonal, indicated that the sequences Met 7–Cys 9 and Gly 25–Cys 27 adopt an antiparallel β structure (Wüthrich et al., 1984). This is in good agreement with the large $^3J_{NH-C\alpha H}$ coupling constants of Met 7, Cys 9, and Phe 26 and the slow exchanging rate of Met 7, Cys 9, and Cys 27 amide protons.

The observation of several spin–spin coupling constants $^3J_{NH-C\alpha H}$ > 8.5 Hz (Pardi et al., 1984), the CαH–CαH NOE between Cys 21 and Cys 27, and another set of NOEs forming a line perpendicular to the diagonal proved that the sequence 20–28 adopts an antiparallel β structure with a reversal in the region 22–26. This folding presented a $NN_{i,i+2}$ NOE between Asn 24 and Phe 26, suggesting that we have located a 23–26 turn with a hydrogen bond between NH of Phe 26 and CO of Pro 23 and probably a β bulge (Richardson et al., 1978) in position Gly 22. In fact, the presence of the sequence Gly-Pro-Asn-Gly, highly favorable to form a β turn according to Chou and Fasman (1978), led us to propose a turn 22–25 with an hydrogen bond between the NH of Gly 25 and CO of Gly 22. In this situation, the proximity of the amide protons of Asn 24 and Phe 26, giving a weak NOE, was observed on molecular models. The two β sheets are shown schematically in Figure 6.

Two $\alpha N_{i,i+2}$ NOEs observed between Ser 13 and Cys 15 and between Ala 17 and Cys 19 led us to propose two tight turns:

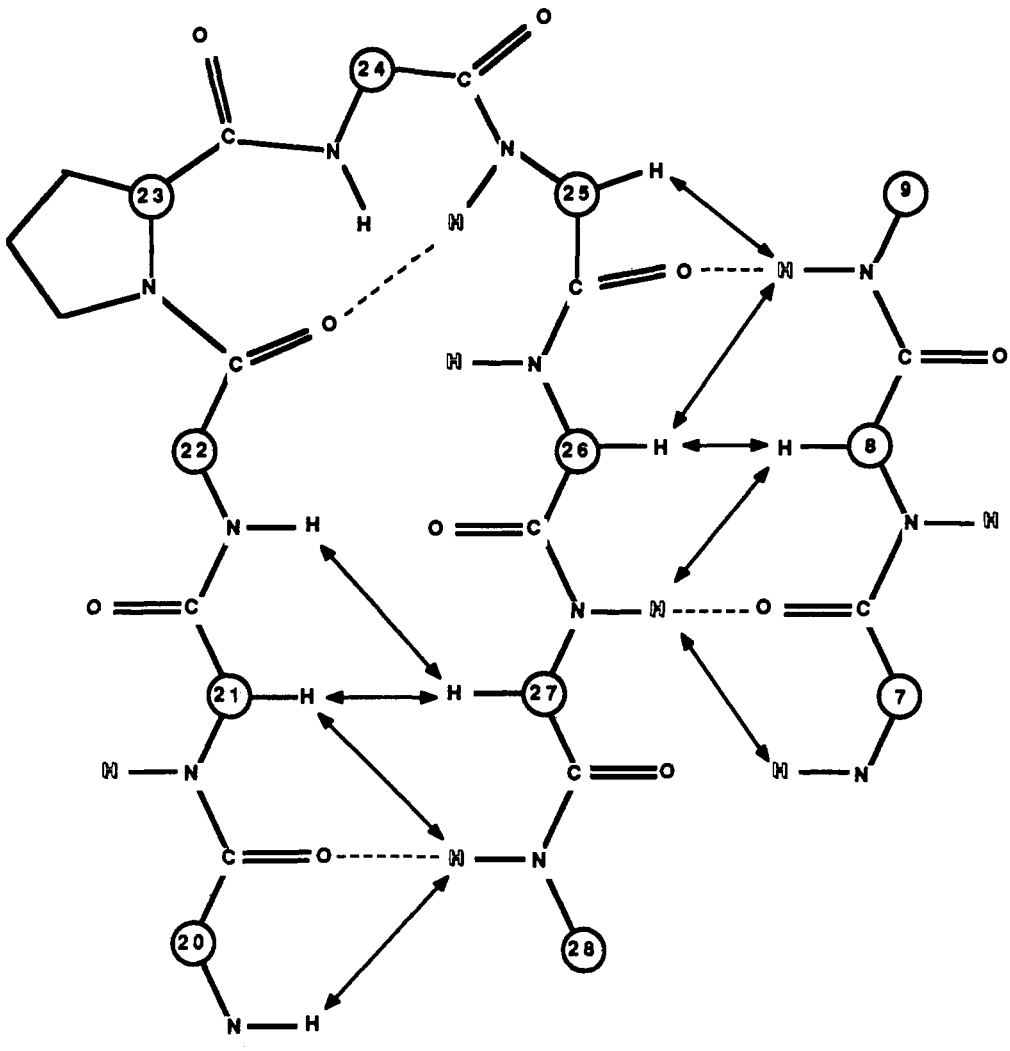

FIGURE 6: Schematic diagram of the two antiparallel β sheets of EETI II. The interstrand backbone NOEs and hydrogen bonds are indicated by arrows and dashed lines, respectively. The slowly exchanging protons are represented by open H's.

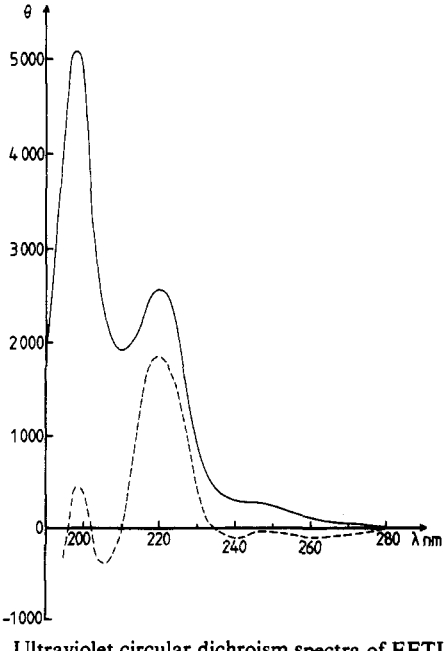

FIGURE 7: Ultraviolet circular dichroism spectra of EETI II in H_2O: (---) pH 2; (—) pH 7.9. Peptide concentration was 1.1×10^{-5} M and cell path length 2 mm.

Asp 12–Cys 15 and Leu 16–Cys 19 with hydrogen bonds CO Asp 12···NH Cys 15 and CO Leu 16···NH Cys 19, respectively. The type of these turns were not defined due to lack of information and the reasons discussed by Wüthrich (1986).

No more experimental structural elements were available, but the fact that the amide proton of Ile 5 has a slow exchanging rate and that Cys 2 is involved in a disulfide bridge led us to make a reversal in the Cys 2–Ile 5 segment.

In EETI II no interaction of the type $i,i+3$ typical of a helical structure was detected in the NOESY spectra, so we concluded that the peptide contains no α helix, as corroborated by CD spectra shown in Figure 7 (two positive extrema at 197 and 220 nm instead of a positive extremum at 191 nm and two negative extrema at 207 and 222 nm for α helix). The conformation of CMTI I, which possesses 20 common amino acids compared to EETI II, was recently studied by Hider et al. (1987) on the basis of CD and Chou–Fasman analysis. These authors concluded that their peptide contains 40–45% α helix with two segments corresponding to helical conformation. The comparison of the Chou–Fasman secondary structure analysis for CMTI I and EETI II (supplementary material) shows that the two peptides are indeed different: CMTI I exhibits a higher α helix tendency than EETI II, especially in the 7–10 and 16–25 regions. Strikingly, EETI II shows a very high β turn propensity at Gly 22 (turn Gly-Pro-Asn-Gly) that is completely absent in CMTI I.

The Disulfide Bridges Problem. In hirudin, an inhibitor of coagulation, Sukumaran et al. (1987) associated the CβH–CβH NOEs between two pairs of cysteines to two disulfide

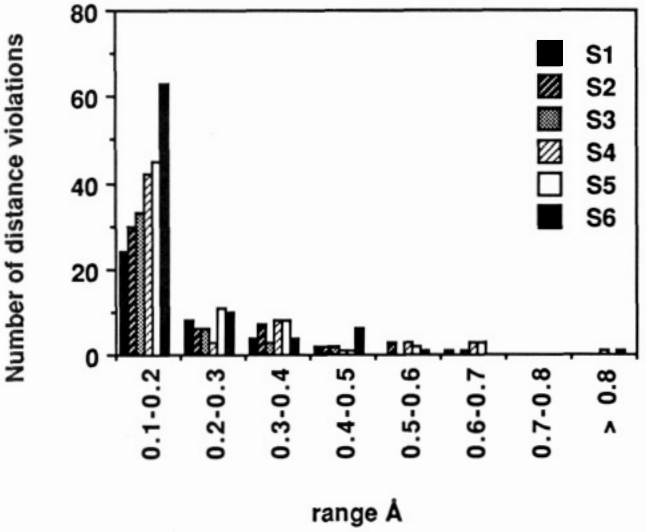

FIGURE 8: Histogram of the residual violations of input distance constraints for the six best computed DISGEO structures for EETI II: S1 (2–19, 9–21, 15–27); S2 (2–19, 9–15, 21–27); S3 (2–9, 15–19, 21–27); S4 (2–15, 9–19, 21–27); S5 (2–19, 9–27, 15–21); S6 (2–27, 9–15, 19–21).

bridges. In our peptide, we observed strong NOEs between CβH of Cys 15 and Cys 19, on one hand, and Cys 9 and Cys 21, on the other hand. From these observations we could conclude that we had located three disulfide bridges, 2–27, 9–21, and 15–19, the same as proposed by Hider et al. for CMTI I. On the other hand, the presence of a NOE between CαH of Cys 21 and CαH of Cys 27 did not permit us to exclude a disulfide bridge 21–27, this NOE being observed in several peptides possessing only one disulfide bridge (Sugg et al., 1988; Liu et al., 1988; Fehrentz et al., 1988). In fact, we could observe on a mechanical molecular model that disulfide bridges are very close to one another and that it was possible to interchange them without any drastic changes in the secondary structure. A similar situation has been described by Gariépy et al. (1986) for the toxic domain of the *Escherichia coli* heat-stable enterotoxin ST I.

Hence, we checked the 15 disulfide bridge possible combinations using the distance geometry program DISGEO.

DISGEO *Modelization*. For each of the 15 possible combinations of cysteines—S1–S15—we computed three different conformations that led to 45 total structures. Angle constraints as well as hydrogen bonds were not included in the first calculations, but subsequently added.

Only 5 of the 12 slowly exchanging amide protons could be unambiguously attributed to H-bonds with known donor atom, particularly among the elements of secondary structure determined above; the following hydrogen bonds were thus imposed: 9 Cys NH···O=C Gly 25, 15 Cys NH···O=C Asp 12, 19 Cys NH···O=C Leu 16, 27 Cys NH···O=C Met 7, 28 Gly NH···O=C Val 20.

Then the constraints on ϕ dihedral angles were first introduced, and subsequently those on χ_1 angles.

Table II: RMS Deviations[a] between Pairs for the Six Best-Computed Structures for EETI II

	S1	S2	S3	S4	S5	S6
S1	–	1.38[b]	1.03	1.35	1.05	2.00
S2	1.84	–	1.57	1.55	1.55	2.10
S3	1.40	2.08	–	0.96	0.88	1.46
S4	1.81	2.12	1.40	–	1.34	1.51
S5	1.49	2.00	1.30	2.00	–	1.59
S6	2.37	2.64	1.92	2.09	1.94	–

[a] The 1–4 segment was not included in the calculations due to its undetermined folding. [b] The above-diagonal entries are between α carbons; the below-diagonal entries are between all heavy atoms.

At last, in consideration of the high probability of β-turn folding for the segment Gly-Pro-Asn-Gly, as shown in Figure 6 (Chou & Fasman, 1978), we also imposed the hydrogen bond between the carbonyl of Gly 22 and the amide proton of Gly 25, although it could not be unambiguously inferred from NOE data.

The use of all these additional constraints did not increase significantly the number or the height of violations of the imposed constraints, which seems indicative of the quality of the data. Finally, with all the constraints described above, only one conformer giving rather low residual violations was obtained for each set of disulfide bridges, which probably comes from a rather severe set of constraints.

Figure 8 shows the residual violations of input distance constraints for the six best computed structures. The results obtained for these structures show that there are no severe inconsistencies in the experimental data. No violations exceeding 0.7 Å or few degrees were found for the four best structures.

Visual inspection of the six best 3D structures (S1–S6) reveals that the overall folding is quite similar in all of these, except for the 1–4 segment. The folding of these residues is only determined by the imposed disulfide bridge for cysteine 2, since we have no NMR constraints on the four first residues. The conservation of the global folding between the different structures has been quantified by the root mean square deviation between fitted structures as given in Table II.

The global folding and the elements of secondary structure determined above can be seen from the stereoview of structure S1 (Figure 9), although slightly distorted: antiparallel β sheet 7–9 and 25–27, β hairpin 20–28, and turns 12–15 and 16–19.

It can be seen from Figure 8 that disulfide links between cysteine 2 and respectively cysteines 21 and 27 are unlikely and that the preferred linkage of cysteine 2 is with cysteine 19 and, to a lesser extent, with cysteines 9 or 15.

CONCLUSION

The NMR study and the distance geometry calculations presented here led to the determination of the overall folding and approximate spatial structure of EETI II. We have shown that at least 6 of the 15 possible disulfide bridge arrangements could be accommodated with only minor changes in the backbone folding.

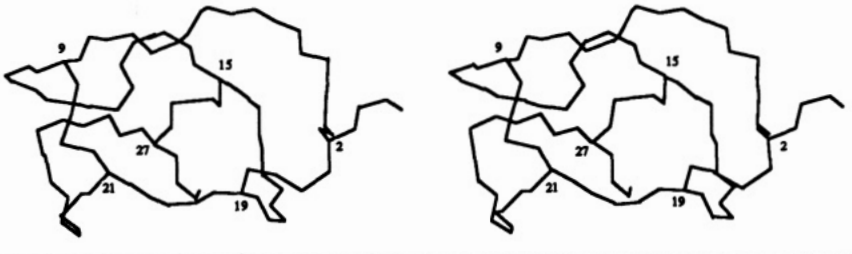

FIGURE 9: Stereoview of the main chain and disulfide bridges of the most probable DISGEO structure S1 for EETI II.

However, given the results shown in Table II, we can propose the structure S1 (disulfide bridges 2–19, 9–21, 15–27) as the most likely. The structures S2 (2–19, 9–15, 21–27) and S4 (2–15, 9–19, 21–27) that cannot be excluded on the extent of distance violations are in fact rejected because of the values obtained for the χ_3 dihedral angles for the disulfide bridges (S2, $\chi_3(21-27) = +4°$; S4, $\chi_3(9-19) = -173°$, $\chi_3(21-27) = -4°$). The structure S3 (2–9, 15–19, 21–27), which also gives rather good results, seems unlikely in a biochemical sense because the three disulfide bridges are independent, with no overlapping able to ensure the overall rigidity of the molecule.

This model of EETI II will be very useful in facilitating the refinement of the X-ray structure of the EETI II/trypsin complex that is now in progress (in collaboration with J. P. Mornon and C. Gaboriau). It provides a better understanding of the structural features of the new squash inhibitors family, and it stands as an interesting model for folding processes.

Acknowledgments

We thank Dr. M. A. Previero-Coletti (U 58 INSERM, Montpellier) for stimulating discussions. The Centre de Recherche en Informatique de Montpellier provided to us valuable computational facilities.

Supplementary Material Available

Complete listing of observed NOEs and complete DISGEO violation table for the 15 disulfide bridge structures and the Chou–Fasman analysis of EETI II and CMTI I (7 pages). Ordering information is given on any current masthead page. This information can also be obtained from the authors (B.C.).

Registry No. Trypsin inhibitor, 9035-81-8.

References

Aue, W. P., Bartholdi, E., & Ernst, R. R. (1976) *J. Chem. Phys. 64*, 2229–2246.

Billeter, M., Braun, W., & Wüthrich, K. (1982) *J. Mol. Biol. 155*, 321–346.

Bodenhausen, G., Kogler, H., & Ernst, R. R. (1984) *J. Magn. Reson. 58*, 370–388.

Chou, P. Y., & Fasman, G. D. (1978) *Adv. Enzymol. Relat. Areas Mol. Biol. 47*, 45–148.

Favel, A., Mattras, H., Coletti-Previero, M. A., Zwilling, R., Robinson, E. A., & Castro, B. (1988) *Int. J. Pept. Protein Res.* (in press).

Fehrentz, J. A., Heitz, A., Seyer, R., Fulcrand, P., Devilliers, R., Castro, B., Heitz, F., & Carelli, C. (1988) *Biochemistry 27*, 4071–4078.

Gariépy, J., Lane, A., Frayman, F., Wilder, D., Robien, W., Schoolnik, G. K., & Jardetzky, O. (1986) *Biochemistry 25*, 7854–7866.

Havel, T. (1985) *QCPE 507*.

Havel, T., & Wüthrich, K. (1984) *Bull. Math. Biol. 46*, 673–698.

Havel, T., & Wüthrich, K. (1985) *J. Mol. Biol. 182*, 281–294.

Havel, T., Crippen, G. M., & Kunz, I. D. (1983) *Bull. Math. Biol. 45*, 665–720.

Hider, R. C., Drake, A. F., Morrison, I. E. G., Kupryszewski, G., & Wilusz, T. (1987) *Int. J. Pept. Protein Res. 30*, 397–403.

Hojima, Y., Pierce, J. V., & Pisano, J. J. (1982) *Biochemistry 21*, 3741–3746.

Joubert, F. J. (1984) *Phytochemistry 23*, 1401–1406.

Kupryszewski, G., & Rolka, K. (1986) *Peptides*, 811–813.

Leluk, J., Otlewski, J., Wieczorek, M., Polanowski, A., & Wilusz, T. (1983) *Acta Biochim. Pol. 30*, 127–138.

Liu, C. F., Fehrentz, J. A., Heitz, A., Le Nguyen, D., Castro, B., Heitz, F., Carelli, C., Galen, F. X., & Corvol, P. (1988) *Tetrahedron 44*, 675–683.

Nagayama, K., Kumar, A., Wüthrich, K., & Ernst, R. R. (1980) *J. Magn. Reson. 40*, 321–334.

Neuhaus, D., Wagner, G., Vasak, M., Kagi, J. H. R., & Wüthrich, K. (1985) *Eur. J. Biochem. 151*, 257–273.

Otting, G., Widmer, H., Wagner, G., & Wüthrich, K. (1986) *J. Magn. Reson. 66*, 187–193.

Pardi, A., Billeter, M., & Wüthrich, K. (1984) *J. Mol. Biol. 180*, 741–751.

Pham, T. C., Lelur, J., Polanowski, A., & Wilusz, T. (1985) *Hoppe-Seyler's Z. Biol. Chem. 366*, 939–944.

Polanowski, A., Wilusz, T., Nienartowicz, B., Cieslar, E., Slominska, A., & Nowak, K. (1980) *Acta Biochim. Pol. 27*, 371–381.

Rance, M., Sorensen, O. W., Bodenhausen, G., Wagner, G., Ernst, R. R., & Wüthrich, K. (1984) *Biochem. Biophys. Res. Commun. 117*, 479–485.

Richardson, J. S. (1981) *Adv. Protein Chem. 34*, 167–339.

Richardson, J. S., Getzoff, E. D., & Richardson, D. C. (1978) *Proc. Natl. Acad. Sci. U.S.A. 75*, 2574–2578.

Siemion, I. Z., Wilusz, T., & Polanowski, A. (1984) *Mol. Cell. Biochem. 60*, 159–161.

Sugg, E. E., Tourwe, D., Kazmierski, W., Hruby, V. J., & Van Binst, G. (1988) *Int. J. Pept. Protein Res. 31*, 192–200.

Sukumaran, D. K., Clore, G. M., Preuss, A., Zarbock, J., & Gronenborn, A. M. (1987) *Biochemistry 26*, 333–338.

Vaney, M. C., Surcouf, E., Morize, I., Cherfils, J., & Mornon, J. P. (1985) *J. Mol. Graphics 3*, 123–124.

Wagner, G. (1983) *J. Magn. Reson. 55*, 151–156.

Wagner, G., & Wüthrich, K. (1982) *J. Mol. Biol. 155*, 347–384.

Wagner, G., Braun, W., Havel, T., Schaumann, T., Go, N., & Wüthrich, K. (1987) *J. Mol. Biol. 196*, 611–639.

Wieczorek, T., Otlewski, J., Cook, J., Parks, K., Leluk, J., Wilimowska-Pelc, A., Polanowski, A., Wilusz, T., & Laskowski, M., Jr. (1985) *Biochem. Biophys. Res. Commun. 126*, 646–652.

Williamson, M. P., Havel, T., & Wüthrich, K. (1985) *J. Mol. Biol. 182*, 295–315.

Wilusz, T., Wieczorek, M., Polanowski, A., Denton, A., Cook, J., & Laskowski, M., Jr. (1983) *Hoppe-Seyler's Z. Physiol. Chem. 364*, 93–95.

Wüthrich, K. (1986) *NMR of Proteins and Nucleic Acids*, Wiley, New York.

Wüthrich, K., Billeter, M., & Braun, W. (1983) *J. Mol. Biol. 169*, 949–961.

Wüthrich, K., Billeter, M., & Braun, W. (1984) *J. Mol. Biol. 180*, 715–740.