Structures of a Series of 6-kDa Trypsin Inhibitors Isolated from the Stigma of Nicotiana alata

ARTICLE *in* BIOCHEMISTRY · NOVEMBER 1995

Impact Factor: 3.02 · DOI: 10.1021/bi00044a007 · Source: PubMed

4 AUTHORS, INCLUDING:

Marilyn Anne Anderson
La Trobe University
131 PUBLICATIONS **6,286** CITATIONS

SEE PROFILE

Structures of a Series of 6-kDa Trypsin Inhibitors Isolated from the Stigma of *Nicotiana alata*[†,‡]

Katherine J. Nielsen,[§] Robyn L. Heath,[∥] Marilyn A. Anderson,[∥] and David J. Craik*[,§]

Centre for Drug Design and Development, University of Queensland, St. Lucia, Queensland, 4072 Australia, and Plant Cell Biology Research Centre, School of Botany, University of Melbourne, Parkville, Victoria, 3052 Australia

Received June 20, 1995; Revised Manuscript Received August 9, 1995[⊗]

ABSTRACT: The three-dimensional structures of a series of 6-kDa trypsin inhibitors isolated from the stigma of the ornamental tobacco *Nicotiana alata* have been determined by ^{1}H NMR spectroscopy combined with simulated annealing calculations. The proteins, T1−T4, are proteolytically cleaved from a 40.3-kDa precursor protein, NA-proPI, together with a chymotrypsin inhibitor, C1, the structure of which was reported recently [Nielsen, K. J., Heath, R. L., Anderson, M. A., & Craik, D. J. (1994) *J. Mol. Biol. 242,* 231−243]. Each of the proteinase inhibitors comprises 53 amino acids, including 8 cysteine residues which are linked to form 4 disulfide bridges. The proteins have a high degree of sequence identity and differ mainly in residues around the putative reactive sites. The structure of T1 was determined using a set of 533 interproton distance restraints derived from NOESY spectra, combined with 33 dihedral restraints derived from $^3J_{NH-H\alpha}$ coupling constants and 16 hydrogen bonds. The structures of the remaining inhibitors (T2−T4) were deduced to be almost identical to T1, on the basis of their similar chemical shifts and 2D spectra. The current study demonstrates that the structures of the trypsin inhibitors (T1−T4) are similar to that previously found for the chymotrypsin inhibitor, C1. Despite differences in sequence, there is conservation in backbone geometry between the reactive site loops of the two classes of inhibitors. From this, it is clear that the nature of the side chain on the primary binding residue, rather than the backbone fold, is the main determinant of the enzyme specificities of these proteinase inhibitors.

Small proteins which function as proteinase inhibitors (PIs)[1] are found in a wide variety of plants. They accumulate in fruits, storage organs, and plant leaves in response to wounding (Richardson, 1977; Brown & Ryan, 1984) and are believed to be involved in protection against predators and disease, since they inhibit a wide range of proteinases of animal and microbial, but not plant, origin. Recently, high levels of serine PIs have been found in the stigmas of the ornamental tobacco *Nicotiana alata* (Atkinson *et al.*, 1993), consistent with a role in preventing or limiting damage to the reproductive tissue by pathogenic microorganisms and insects.

The serine PIs in *N. alata* stigmas are produced from a 40.3-kDa protein with six potential PI sites (Atkinson *et al.*, 1993). This protein, NA-proPI, is processed proteolytically *in vivo* to produce five homologous PIs of about 6 kDa. In addition to the PI domains, NA-proPI contains two predicted peptides from the N- and C-termini of 2.7 and 5.9 kDa, respectively (Atkinson *et al.*, 1993; Heath *et al.*, 1995), which have not been isolated. The N-terminal fragment contains the sixth potential PI reactive site.

The 40.3-kDa precursor protein has been detected in stigmas at the early stages of bud development, but it is processed to the 6-kDa PIs as the flower matures. It is likely that processing occurs after the precursor protein has passed through the endomembrane system into the vacuoles (Atkinson *et al.*, 1993). Cleavage occurs at a specific site, adjacent to or within a short linker sequence (EEKKN) between the PI domains, with subsequent trimming by nonspecific endogenous peptidases (Heath *et al.*, 1995). The involvement of nonspecific peptidases for trimming is supported by the detection of small quantities of PIs containing either one more, or one fewer, residues at their termini, giving rise to the terminology "ragged ends".

Protein-based serine proteinase inhibitors are characterized by a reactive site region which mimics the scissile bond in the normal substrate. The PIs bind but do not diffuse from the enzyme, thereby blocking binding of the substrate. These inhibitor complexes have been described as distorted Michaelis complexes (Bode & Huber, 1991). Most PIs previously found in plants have just a single or two kinetically independent reactive sites [see reviews by Richardson (1991) and Garcia-Olmedo *et al.* (1987)]. Recently, an 85-kDa cysteine PI composed of eight closely related domains has been isolated from potato (Walsh & Strickland, 1993); however, unlike the multidomain PI from *N. alata*, the 85-kDa protein is unprocessed in plant tissues. The large number of PI reactive sites found in NA-proPI represents an efficient molecular design since only one set of transcription, translation, and targeting machinery is required to

[†] This work was supported from a grant by the Australian Research Council and the Clive and Vera Ramaciotti Foundation.

[‡] Coordinates of the structure have been deposited in the Brookhaven Protein Data Bank under the file name 1TIH.

* Author to whom correspondence should be addressed.

[§] University of Queensland.

[∥] University of Melbourne.

[⊗] Abstract published in *Advance ACS Abstracts,* October 1, 1995.

[1] Abbreviations: PI, proteinase inhibitor; NA-proPI, *Nicotiana alata* precursor protein; NMR, nuclear magnetic resonance; SA, simulated annealing; DQF-COSY, double-quantum-filtered *J*-correlated spectroscopy; NOE, nuclear Overhauser effect; NOESY, two-dimensional NOE spectroscopy; TOCSY, total correlated spectroscopy; E-COSY, exclusive *J*-correlated spectroscopy; DIPSI, decoupling in the presence of scalar interactions; DSS, 4,4-dimethyl-4-silapentane-1-sulfonate; 2D, two dimensional; 3D, three dimensional; RMSD, root mean square deviation; H-bond, hydrogen bond; SGPB, *Streptomyces griseus* proteinase B.

0006-2960/95/0434-14304$09.00/0

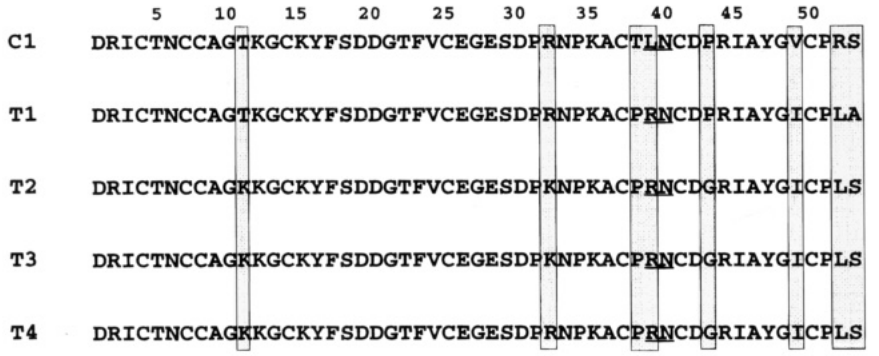

FIGURE 1: Sequence alignment for C1 and T1−T4 (Atkinson *et al.*, 1993). Regions of sequence variability are boxed, and the reactive site is underlined.

ultimately produce several PIs. It is therefore important to gain a knowledge of the tertiary structure of the individual PIs and their relative organization within NA-proPI.

Figure 1 shows the primary structures of the five PIs which share 85−100% sequence identity. One is a chymotrypsin inhibitor, referred to as C1, and the other four are trypsin inhibitors (T1−T4). The sequence differences between the individual PIs occur mainly in the C-terminal region, including substitutions near the putative reactive sites. The enzymes trypsin and chymotrypsin have similar 3D structures apart from the primary binding pocket, and specificity for the cognate enzyme is conferred by the nature of the P_1 [using the nomenclature of Schecter and Berger (1967)] residue side chain on the substrate [for review, see Fersht (1985)]. In the case of C1, this residue is Leu39 while for the trypsin inhibitors, T1−T4, it is Arg39. From this, and the high sequence identity between C1 and T1−T4, it may be anticipated that the trypsin inhibitors adopt similar 3D structures to C1.

In a recent study (Nielsen *et al.*, 1994b) we determined the 3D structure in solution of C1 using a combination of 2D ^{1}H NMR spectroscopy and simulated annealing calculations. This work showed that the protein contains a triple-stranded β-sheet as the dominant secondary structural element, with several turns and a short region of 3_{10} helix also being present. The putative chymotrypsin-reactive site is located on an exposed loop which is less defined than the rest of the protein. The overall shape of C1 is disk-like, and the N- and C-termini are exposed, consistent with the proposal that this protein results from posttranslational processing of the 40.3-kDa precursor protein. The structure is similar to that derived by X-ray methods for PCI-1 (Greenblatt *et al.*, 1989), the chymotrypsin-active domain of a PI found in potato tubers which shares 70% sequence identity with C1. This finding is of significant interest, since although the chymotrypsin-active domains C1 and PCI-1 are nearly identical, the nature of their respective precursor proteins is different. The PCI-1 precursor protein is only double headed and lacks the hydrophilic linker sequence found in NA-proPI.

The profile of enzyme inhibition of the trypsin inhibitors isolated from *N. alata* stigmas is different from that of C1, and it is of interest to determine if there is a structural basis for this difference in activity. The inhibitor C1 binds to chymotrypsin with 1:1 stoichiometry and has no inhibitory activity against trypsin (Heath *et al.*, 1995). The trypsin inhibitors, T1−T4, on the other hand, have the highest affinity for trypsin but also bind to and inhibit chymotrypsin, despite the presence of arginine in the P_1 position at the single reactive site. The interesting property of inhibition of both trypsin and chymotrypsin has also been observed in other small, single-domain trypsin inhibitors from the Solanaceae family (Yamada *et al.*, 1977; Pearce *et al.*, 1982, 1993). The structure of the reactive site of the trypsin inhibitors is of further interest in view of observations that trypsin inhibitors retard insect growth more effectively than chymotrypsin inhibitors when expressed in transgenic plants [reviewed by Ryan (1990)] and thus have potential applications in crop protection against insect damage.

In the current paper we report the structures of the trypsin inhibitors, T1−T4, as derived by 2D ^{1}H NMR spectroscopy and simulated annealing calculations. A full 3D structure determination was made of T1 for comparison with C1, and the structures of T2−T4 were deduced from their similar patterns of chemical shifts and 2D spectra. The NMR measurements were made using gradient methods for suppression of the water signal, resulting in improved spectra relative to those obtained for C1 using presaturation/binomial suppression methods. The larger number of NOEs and better quality spectra have allowed us to produce a highly refined structure for T1. The greater precision in defining the backbone and side-chain atoms increases the utility of the information in structure−function studies. The structural information now available on the individual domains from this study of T1−T4, and from the previous study (Nielsen *et al.*, 1994b) of C1, provides the basis for deducing the arrangement of the domains in the precursor protein, NA-proPI. In turn, this should lead to a greater understanding of the processing mechanisms responsible for the production of the individual PIs.

MATERIALS AND METHODS

Materials. PI proteins were extracted from 21 000 stigmas excised from mature flowers of *N. alata* and purified as described by Heath *et al.* (1995). Samples for NMR spectroscopy contained 1−2 mM protein in either 100% D_2O or 90% H_2O/10% D_2O v/v at pH 3.5.

NMR Experiments. Spectra were recorded on a Bruker ARX 500 spectrometer at temperatures in the range 288−313 K. 2D experiments included NOESY (Jeener *et al.*, 1979; Kumar *et al.*, 1980) with mixing times of 100 and 200 ms and TOCSY (Braünschweiler & Ernst, 1983; Bax & Davis, 1985; Davis & Bax, 1985) with a MLEV-17 spin-lock mixing sequence of 65 ms. All 2D spectra were recorded over 5050 Hz with 4K data points, 400−600 FIDs, and a recycle delay of 1.0 s. $^3J_{NH-H\alpha}$ coupling constants

were derived from a DQF-COSY (Rance *et al.*, 1983) spectrum, and $^3J_{H\alpha-H\beta}$ coupling constants were measured from an E-COSY (Greisinger *et al.*, 1987) spectrum. Slow-exchange protons were determined from TOCSY spectra in D_2O recorded 1 h after dissolution. Water suppression for the NOESY and TOCSY spectra was achieved by the use of two sine-shaped gradient pulses of 6 G cm^{-1} gradient strength on either side of a binomial 3-9-19-19-9-3 pulse of 10 kHz field strength. Other experiments used continuous, low-powered irradiation during the recycle delay.

All spectra were processed using the spectrometer software package UXNMR. The FIDs were generally multiplied by a polynomial function and apodized using a 60° shifted sine-bell function in both dimensions prior to Fourier transformation. Baseline correction, using a fifth-order polynomial function, was applied in F_2, and chemical shifts were referenced externally to DSS at 0.00 ppm. Coupling constant determinations required high digital resolution, and therefore the E-COSY and DQF-COSY spectra were strip transformed to 8K × 1K over the regions of interest. A deconvolution program which fitted Lorentzian line shapes to 1D rows of these spectra was used to measure coupling constants. This procedure minimizes errors due to finite line widths of antiphase peak components.

Distance Restraints and Structure Calculations. Peak volumes were obtained from a 200 ms NOESY spectrum and classified as strong, medium, or weak, corresponding to interproton distance restraints of 1.8−2.7, 1.8−3.5, and 1.8−5.0 Å, respectively (Williamson *et al.*, 1985; Clore *et al.*, 1986b). Peaks observed at 200 ms but not at 100 ms mixing times were classified as very weak and were assigned a distance restraint of 1.8−6.0 Å. Upper limits for nonstereospecifically assigned methylene and methyl protons were corrected appropriately (Wüthrich *et al.*, 1983), and distances of 1.5 and 2.0 Å were added to the upper limits of restraints involving methyl protons and phenyl rings, respectively. $^3J_{NH-H\alpha}$ coupling constants were used to determine ϕ angle restraints (Pardi *et al.*, 1984), and $^3J_{H\alpha-H\beta}$ coupling constants in conjunction with relevant NOESY strengths were used to determine χ_1 angle restraints (Wagner *et al.*, 1987). All peptide bonds were constrained to be in the *trans* configuration.

Simulated annealing and energy minimization calculations were done using X-PLOR (Brünger *et al.*, 1986; Brünger, 1992). Structures were calculated using a two-stage simulated annealing protocol starting from a series of random template structures (Nilges *et al.*, 1988). The simulated annealing calculations used the standard force-field parameter set and topology file in X-PLOR version 3.1. The first stage consisted of a total of 30 ps of molecular dynamics at high temperature (1000 K) followed by cooling (0 K). During this stage, the S−S connectivities were assigned as pseudo-NOE restraints. Disulfide bonds were formally included during the second stage, which consisted of 10 ps of molecular dynamics at 1000 K and cooling to 0 K. Final structures were subjected to 2000 cycles of restrained energy minimization using the conjugate gradient Powell algorithm (Clore *et al.*, 1986a) under the influence of the CHARMm force field (Brooks *et al.*, 1983).

RESULTS

Protein Purification. The five 6-kDa PIs derived from the *N. alata* PI precursor were isolated as described by Heath *et al.* (1995). The reversed-phase HPLC used as the final purification step resulted in clear separation of C1 (20 mg) and T1 (8 mg) from T2−T4 (40 mg, combined). T2−T4 could not be separated (T2 and T3 are identical in any case) and hence were examined as a mixture of components.

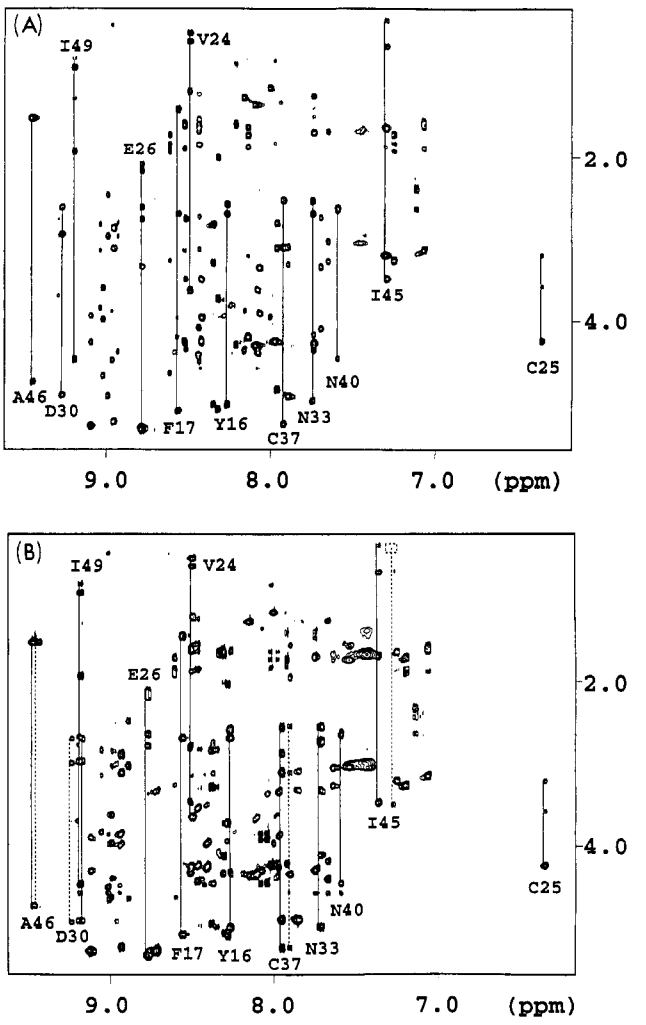

FIGURE 2: TOCSY spectra (mixing time 65 ms) of the NH−aliphatic region showing several of the spin system connectivities for (A) T1 and (B) T2−T4. In (B) the dashed line indicates T4 connectivities while the solid line represents T2/T3 connectivities.

Spectral Assignments. The peaks in the fingerprint (NH−Hα) region of the 2D spectra were well dispersed, as illustrated in Figure 2A, which shows the TOCSY spectrum of T1. Sequential resonance assignments were made using well-established methodology (Wüthrich, 1986) in which scalar connectivities were identified from DQF-COSY and TOCSY spectra and through-space connectivities from NOESY spectra. In general, the spin-system patterns were similar to those previously encountered for C1, with the main differences occurring for the regions around the sites of amino acid substitutions, that is, those near the reactive site, including residues 37−40 and the C-terminus residues 51−53. The five proline residues in T1 were assigned as *trans* from the observation of strong sequential Hα−Hδ_{i+1} NOE cross peaks and the absence of Hα−Hα_{i+1} connectivities. The complete chemical shift assignments for T1 are provided as supporting information.

A similar protocol was used to assign peaks in the mixture containing T2−T4. In many cases corresponding residues in the different proteins were close in chemical shift and

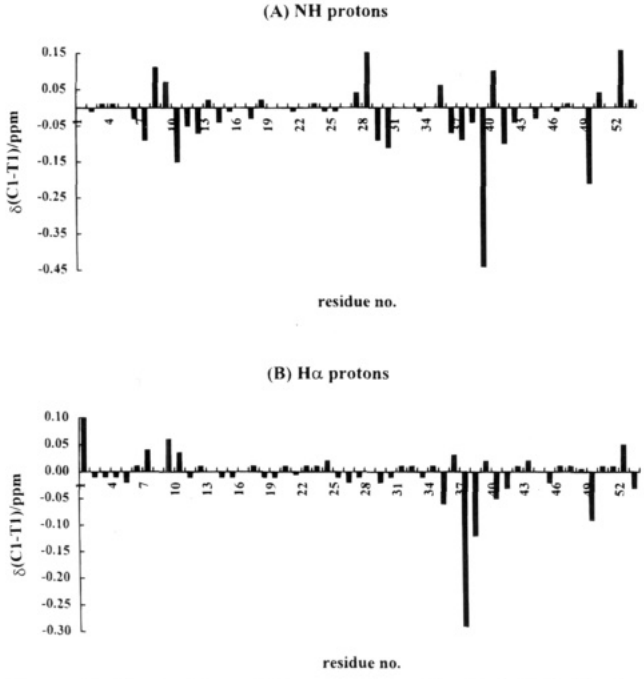

FIGURE 3: Comparison of the secondary chemical shifts for the Hα and NH protons of T1 and C1.

could be readily identified as "brothered" peaks in the spectrum. The peaks corresponding to T2/T3 were approximately double the intensity of those corresponding to T4, and therefore it was straightforward to distinguish T2/T3 from T4 in cases where there were two sets of peaks. For most resonances, the offset between the T2/T3 and T4 peaks was in the NH dimension only, with the aliphatic shifts being identical. This is illustrated in Figure 2B, which shows part of the TOCSY spectrum of the T2−T4 mixture. The similarity of the 2D spectra for T1 and T2−T4 (Figure 2) suggests that all of the trypsin inhibitors from NA-proPI adopt the same 3D structure.

Chemical Shifts. Changes in chemical shifts of backbone protons relative to random coil values provide information on protein secondary structure (Bundi & Wüthrich; 1979; Wishart *et al.*, 1991). For T1, these secondary shifts are large (up to 1 ppm), indicating that well-defined regions of secondary structure are present. For C1, these regions included β-sheet, 3_{10} helix, and several turns. Figure 3 shows the differences in secondary shifts between C1 and T1 for the Hα and NH protons. While these values are generally small, indicating that their secondary structures are very similar, there are regions with significant differences. These deviations correspond to residues around the reactive site and near the C-terminus which are in regions with the highest sequence variability between the proteins. The region encompassing the 3_{10} helix also has differences in secondary shifts for both NH and Hα protons which, in this case, cannot be attributed to sequence variability but may be due to a subtle difference in structure. For T2−T4, the Hα chemical shifts are almost identical to those of T1 and C1 (data not shown), indicating that the secondary structure and global fold for all of these inhibitors are similar.

Secondary Structure. Measurements of coupling constants and the identification of slowly exchanging NH protons, together with analysis of sequential, medium-range, and long-range NOE intensities, were used to characterize the secondary structure of T1 and T2−T4. Due to the similarity in spectra for T1 and T2−T4, only the data for T1 are shown to represent the structure of the trypsin inhibitors (Figure 4). The cross peaks in the NOESY spectra for T1−T4 had strengths similar to those of C1, and the long- and medium-range connectivities were also similar over most residues, confirming that their structures are alike. This is further supported by the slowly exchanging NH protons and coupling constants which were the same in all of the proteins.

The secondary structure in C1 and T1−T4 can be summarized as follows: a triple-stranded β-sheet incorporating residues 14−18 as the central β-strand and residues 21−26 and 45−50 as the two peripheral β-strands, a β-turn (18−21) linking β-strands 14−18 and 21−26 forming a β-hairpin structure, three more β-turns at residues 11−14, 30−33, and 42−45, and a short stretch of 3_{10} helix (residues 7−10). Irregularities in the secondary structure include two classic type β-bulges at residues 24−25 and 46−47 and a G1-type β-bulge at residues 22−23 (Nielsen *et al.*, 1994b).

Reactive Site Region. The region encompassing residues 34−41 contains no medium-range NOEs representative of secondary structure in any of the PIs from NA-proPI. The lack of secondary structure is supported by an absence of slow-exchange NH protons in the region. However, the presence of several strong NH−NH_{i+1} connectivities and $^3J_{NH-H\alpha}$ coupling constants either ≤6 Hz or ≥9 Hz in this region is indicative of some local structural definition, particularly near the scissile bond where residue 39 has a large $^3J_{NH-H\alpha}$ coupling constant and a strong NH−NH_{i+1} connectivity to residue 40. This applies for C1 and for T1−T4. In addition, there are several long-range NOEs between the residues of the reactive site loop and the 3_{10} helix, further indicating that this region is not entirely disordered.

Temperature Coefficients. From a series of TOCSY spectra recorded at five temperatures, the dependence of NH chemical shifts on temperature was determined. These temperature coefficients provide a method of assessing

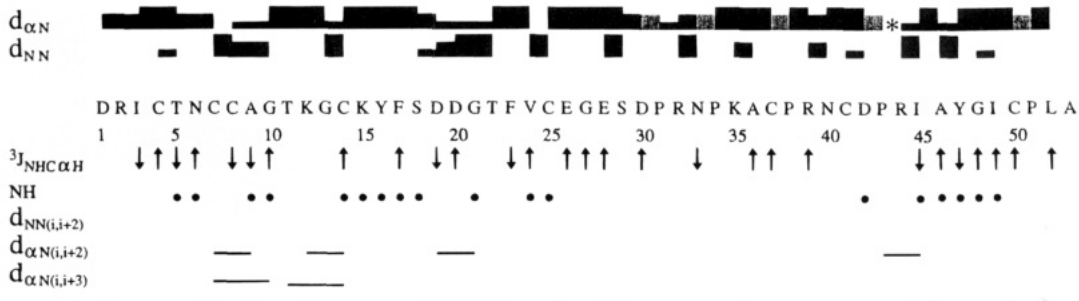

FIGURE 4: Summary of sequential and medium-range NOE information, $^3J_{NH-H\alpha}$ coupling constant, and slow-exchange NH data. The height of the bars represents relative NOE strength. ↑ and ↓ symbols represent $^3J_{NH-H\alpha}$ coupling constants of ≥8.5 and ≤ 6 Hz, respectively. Filled circles indicate slow-exchange NH protons.

Table 1: Comparison of NH Temperature Coefficients and Slow-Exchange Data Obtained for T1

residue	temp coeff (ppb/K)[a]	slow-exchange NH[b]	residue	temp coeff (ppb/K)[a]	slow-exchange NH[b]
Arg2	6.38		Glu26	7.26	
Ile3	4.89		Gly27	2.86	
Cys4	7.81		Glu28	0.06	
Thr5	4.22	•	Ser29	5.81	
Asn6	0.98	•	Asp30	10.81	
Cys7	3.00		Arg32	1.95	
Cys8	1.90		Asn33	2.46	
Ala9	3.20	•	Lys35	4.96	
Gly10	4.70	•	Ala36	7.94	
Thr11	8.97		Cys37	7.79	
Lys12	1.89		Arg39	7.87	
Gly13	4.09		Asn40	6.09	
Cys14	2.08	•	Cys41	8.72	
Lys15	0.03	•	Asp42	6.07	•
Tyr16	1.34	•	Arg44	2.34	
Phe17	0.53	•	Ile45	0.71	•
Ser18	2.11	•	Ala46	3.84	•
Asp19	4.40		Tyr47	1.31	•
Asp20	2.76		Gly48	7.74	•
Gly21	4.05	•	Ile49	4.04	•
Thr22	3.73		Cys50	9.85	
Phe23	7.92		Leu52	9.65	
Val24	3.97	•	Ala53	4.46	
Cys25	0.53	•			

[a] Coefficient determined from a plot of chemical shift versus temperature. [b] Detected as slowly exchanging in the D_2O TOCSY spectra.

H-bonding which is complementary to that derived from the D_2O exchange experiments, although other factors can also contribute to the temperature coefficients. The results shown in Table 1 support the conclusion of a recent study (Nielsen *et al.*, 1994a) which found that while, in general, low-temperature coefficients correlate with slowly exchanging protons, some slowly exchanging NH protons have relatively high temperature coefficients. For example, while most slow-exchange NH protons for T1 have temperature coefficients less than 5 ppb/K, Asp42 and Gly48 have NH protons that exchange slowly with D_2O and have high temperature coefficients, i.e., close to random coil values for these residue types (Merutka *et al.*, 1995). By contrast, there are other NH protons that appeared to exchange rapidly with the solvent yet had low temperature coefficients. The very low temperature coefficients for the NH protons of Lys12 and Asn33 suggest that these protons are involved in H-bond interactions, yet they were not observed in the 2D slow-exchange experiment. Further comments on these observations are given in the Discussion section, in relation to the derived 3D structures.

Three-Dimensional Structure. The NMR data obtained from T1 and from T2−T4 are consistent with a conserved secondary structure and a similar tertiary fold in each of the component PIs derived from NA-proPI. The structure previously determined for the chymotrysin inhibitor C1 (Nielsen *et al.*, 1994b) is therefore a reasonable model for the trypsin-active inhibitors. However, some regions of the C1 structure were not well defined, and the new data for T1 were used to refine the structural model for the inhibitors and to investigate any subtle differences between the reactive loop conformations of C1 and T1 that may contribute to the differences in their enzyme specificity. The additional information was obtained by performing the NOESY and TOCSY experiments at several temperatures, resulting in more long- and medium-range NOEs and the unambiguous assignment of previously unidentified NOEs. Gradient-enhanced NMR methods for water suppression were also used to improve definition of cross peaks close to the water resonance.

A total of 533 distance restraints derived from 273 intraresidual, 128 sequential, and 132 medium- and long-range NOEs was used to generate an initial set of 25 structures which were checked for errors and violations prior to the successive addition of 26 ϕ and 17 χ_1 dihedral restraints and 36 H-bond restraints (defining 18 H-bonds) for the generation of further structures. H-Bond restraints were placed on the basis of the observation of a slow-exchange NH proton and/or the measurement of a very low temperature coefficient (<3 ppb) and agreement with the distance and orientational criteria (Kabsch & Sander, 1983) as defined in the program PROCHECK (Laskowski *et al.*, 1993). The disulfide pairings for T1 were assumed to be identical to those deduced previously for C1 (Nielsen *et al.*, 1994b).

A refined set of 50 structures was generated with all available NOE, dihedral, H-bond, and disulfide restraints, and the 25 structures with the lowest total energies were chosen to represent the solution conformation of T1. A backbone superimposition of these structures (Figure 5) showing a front view and two side views illustrates clearly the global fold of T1, which, like C1, is disk-shaped with a slightly concave face and exposed ends. The main difference between the C1 and T1 structures is in the region preceding

FIGURE 5: Backbone superimposition of the 25 lowest energy refined structures of T1 showing a front view and two side views.

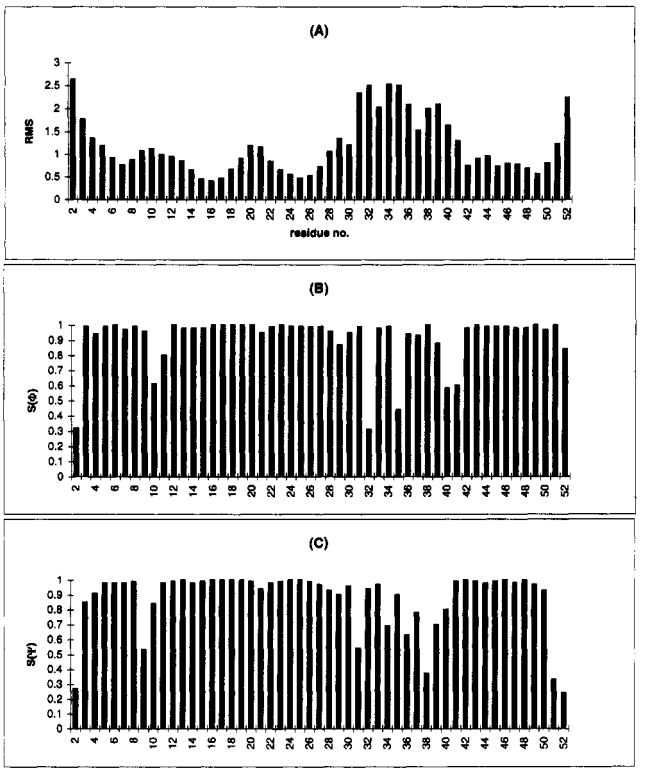

FIGURE 6: RMSD (Å) of the backbone atoms from the averaged structure (A) and the S values for the dihedral angles ϕ (B) and ψ (C) versus residue number.

the reactive site loop (residues 27−37), which is better defined in T1, due to the larger number of medium- and long-range experimental restraints used in the structure determination.

All regions containing secondary structure show excellent agreement among the structures, while the termini and reactive site loop are not well-defined. This is supported by the atomic RMSD by residue for the backbone atoms (Figure 6A), which is low for the residues involved in secondary structure but high for the reactive site and terminal regions. The angular order parameters, S (Pallaghy *et al.*, 1993; Hyberts *et al.*, 1992), for the backbone dihedral angles ϕ and ψ (Figure 6B,C) are close to unity for most residues, indicating high precision in the structures; however, some isolated residues, particularly in the reactive site loop (residues 37−41), have low S values, representative of disorder.

DISCUSSION

The most ordered region of the structures of C1, T1, and, by analogy, T2−T4 is the triple-stranded β-sheet. A short 3_{10} helix lies above one surface of this sheet anchoring, via a disulfide bridge network, the more disordered reactive site loop to the stable β-sheet scaffold. Analysis of the Ramachandran plot for the average structure shows that all well-defined residues reside in low-energy regions of ϕ,ψ space. A comparison of this plot to that for C1 shows that the local conformations are nearly identical for the two proteins (Figure 7).

For T1, the degree of spread in ϕ,ψ space for the residues of the reactive site loop is greater than that of the β-sheet and turn regions; nevertheless, in the majority of structures, these residues also occupy low-energy regions and adopt the

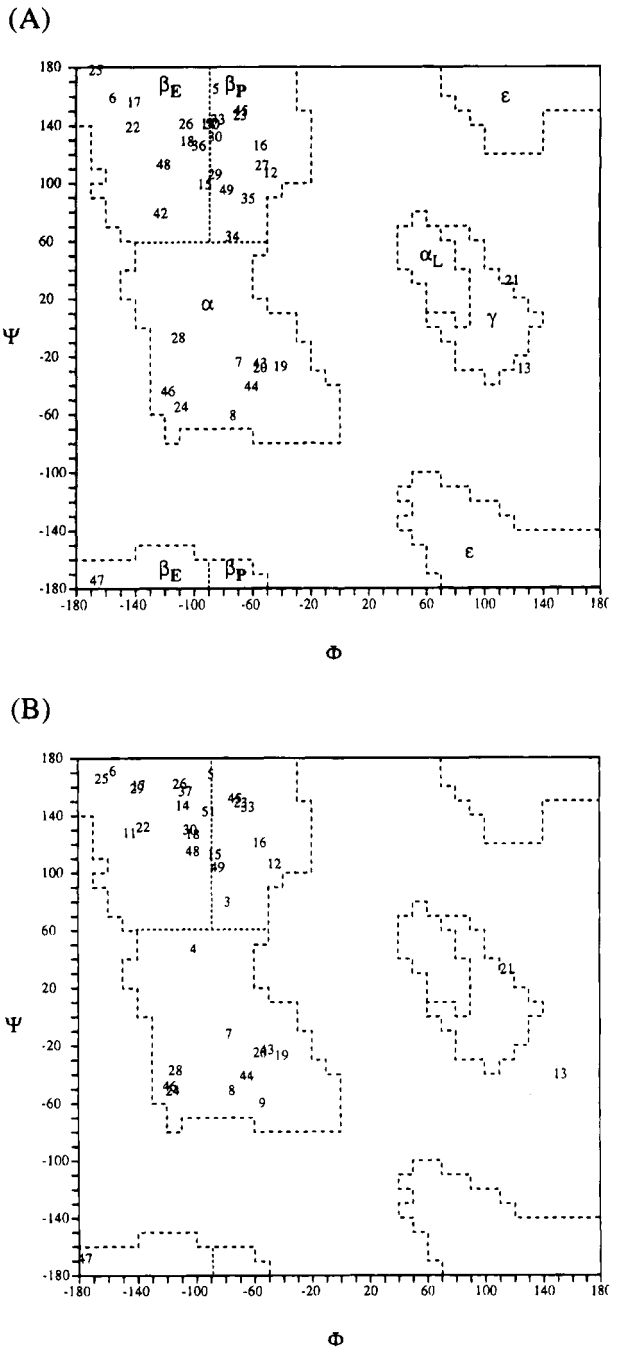

FIGURE 7: (A) Ramachandran plot for the averaged structure of T1 showing the ϕ and ψ dihedral angles of well-defined residues [i.e., where both $S(\phi)$ and $S(\psi)$ are ≥ 0.8 in the ensemble of 25 calculated structures]. The corresponding plot for C1 (Nielsen *et al.*, 1994b) is shown in (B). Conformations in regions marked on the Ramachandran plot in (A) are denoted as follows: β_E, antiparallel β-sheet; β_P, parallel β-sheet; α, right-handed α-helix; α_L, left-handed α-helix; γ and ϵ, usually accessible to glycines and smaller residues, i.e., Asp and Asn [adapted from Wilmot and Thornton (1990)].

conformations shown in Table 2. These are compared to the conformations of the residues of the reactive site loop in C1, which have previously been shown (Nielsen *et al.*, 1994b) to adopt conformations similar to those of the corresponding residues in the related PI, PCI-1, complexed with the chymotrypsin-like enzyme, SGPB (Greenblatt *et al.*, 1989). Unlike the remainder of the molecule, the conformation of the T1 P_2 residue is slightly less defined than its C1 counterpart. This is not likely to be due to greater motion

FIGURE 8: Front and side CPK views of the lowest energy T1 structure showing the residues that have both low-temperature coefficients and slow-exchange NH protons (orange), those with low-temperature coefficients and fast-exchange NH protons (green), and those with high-temperature coefficients and fast-exchange NH protons (white).

Table 2: Comparison of the Conformations of Reactive Site Loop Residues P_4-P_3' for C1 (Nielsen *et al.*, 1994b) and T1 (This Study)

	P_4	P_3	P_2	P_1	P_1'	P_2'	P_3'
T1	A36	C37	P38	R39	N40	C41	D42
ϕ,ψ	β	β	β_P	α	β	β	β_E
N[a]	23	24	15	25	14	19	25
C1	A36	C37	T38	L39	N40	C41	D42
ϕ,ψ	β_E	β	β_P	α	β	β	β_E
N	22	25	19	25	14	21	25

[a] N: number of structures out of a possible 25 that adopt the conformation shown. The symbols refer to regions of the Ramachandran plot marked in Figure 7A.

but may be attributed to the presence of a proline residue at this site in T1 which, in the absence of a backbone NH proton, exhibits fewer NOEs than the threonine residue at the P_2 site in C1. Nevertheless, as is evident in Table 2, C1 and T1 adopt similar conformations at the reactive site loop in most structures, despite the differences in sequence. This indicates that the backbone fold is not the primary determinant of differences in biological activity that are observed between the tryptic and chymotryptic PIs of NA-proPI. Rather, it is the nature of the P_1 side chain that confers specificity of the inhibitor to its cognate enzyme.

It was mentioned previously that the NH and Hα chemical shifts for the residues of the 3_{10} helix differed for C1 and T1−T4, although there is complete sequence identity in this region. Since Cys8 and Cys37 are linked by a disulfide bridge, the chemical shift differences may be attributed to the proximity of these residues to those of the variable reactive site loop. This is supported by the observation of several long-range NOEs between these regions, and furthermore, a comparison of the 3D structures of C1 and T1 revealed that there are no structural differences in the 3_{10} helix region.

The presence of slowly exchanging amide protons in spectra recorded for D_2O solutions of proteins is the primary experimental evidence that the NH protons concerned are involved in H-bonding. As part of a wider interest in other potential measures of H-bonding, we measured the temperature coefficient for all amide protons in T1 and T2−T4. The results of the comparison between NH temperature coefficients and slow-exchange data are in agreement with previous studies (Nielsen *et al.*, 1994a; Skalicky *et al.*, 1994) which show that, although there is a general correlation between low-temperature coefficients (<5 ppb/K) and slow-exchange NH protons, high values for temperature coefficients are sometimes observed for slow-exchange NH protons. Furthermore, some NH protons involved in H-bonds may not be detected in slow-exchange experiments, for a variety of reasons including solvent exposure or low-signal intensity in the TOCSY spectrum. To obtain maximum information on NH protons involved in H-bonding interactions, it is beneficial to measure NH temperature coefficients as well as NH exchange rates.

This is illustrated by the data for the NH protons of Lys12 and Asn33, which are not slowly exchanging, but there is reason to believe that they are involved in H-bonds. For example, in most of the 3D structures calculated for C1 (Nielsen *et al.*, 1994b), there were H-bonds involving these NH protons, i.e., 33NH···OC30 and 12NH···29OC. In C1, the turn involving the NH proton of residue 33 is exposed at the surface, suggesting that the H-bond, 33NH···OC30, is susceptible to increased deuterium exchange. In the case of Lys12, the failure to detect the NH proton in slow-exchange D_2O spectra is likely due to its inherently low intensity. The NH−aliphatic connectivities are weak even in the TOCSY spectra in aqueous solution, and in D_2O, the NH signal intensity would be further reduced by partial exchange. The same situation is also likely to apply to Cys7, which has extremely low peak intensities in the TOCSY spectrum which in this case may be due to a very small $^3J_{NH-H\alpha}$ coupling constant. There are also several other NH protons (i.e., Cys8, Arg32, Arg44, Asp20) which, judging by the low-temperature coefficients, are likely to be involved in H-bonds but were not detected as slowly exchanging. A representation of the surface of T1 (Figure 8) shows that most residues with low-temperature coefficients and slow-exchange NH protons are buried in the protein, while those residues with low-temperature coefficients and fast NH exchange rates are generally exposed at the surface.

As was mentioned previously, the series of PIs from NA-proPI does not exhibit slow exchange along the reactive loop region. This is of significant interest because the activity of serine proteinase inhibitors has been linked to the presence and strength of an intramolecular H-bond involving the NH proton of the P_1' residue (i.e., at Asn40 for these inhibitors) (Fujinaga *et al.*, 1982), and indeed, this has been observed for other trypsin inhibitors, even in the uncomplexed state (Heinz *et al.*, 1992; Nielsen *et al.* 1994a). In contrast, the high rate of proton/deuterium exchange at this site and the relatively high temperature coefficient for the NH of Asn40 in T1 (Figure 8) strongly suggest that this proton is not involved in a H-bond. This implies that, at least in the free state, an intramolecular H-bond involving the backbone NH proton at P_1' is not an absolute requirement for PI activity. However, the large $^3J_{NH-H\alpha}$ coupling constant at P_1 and the strong $NH-NH_{i+1}$ NOE between the P_1 and P_1' residues provide evidence that the motion at the reactive site is somewhat restricted. The absence of the aforementioned H-bond in the free state does not preclude its presence in the bound state, since the NH proton of the P_1' residue in the PCI-1/SGPB complex forms a H-bond with its side-chain oxygen (Greenblatt *et al.*, 1989).

This study has provided a unique opportunity for a direct comparison of PIs belonging to the same class, derived from the same source, but having differences in serine proteinase specifities. In summary, it has been shown that the four trypsin inhibitors (T1–T4) of NA-proPI adopt the same secondary and tertiary structure as the chymotrypsin inhibitor, C1. Furthermore, their reactive site loops, although not highly defined, also adopt similar conformations, supporting the view that the nature of the P_1 residue side chain, rather than the local backbone conformation, determines the specificity of the trypsin and chymotrypsin PIs for their cognate enzymes. Current investigations are aimed at determining the relative orientations of the PI domains and the 3D structure of the precursor protein.

SUPPORTING INFORMATION AVAILABLE

A table of ^1H chemical shift assignments for T1 in H_2O at 313 K (3 pages). Ordering information is given on any current masthead page.

REFERENCES

Atkinson, A. H., Heath, R. L., Simpson, R. L., Clarke, A. C., & Anderson, M. A. (1993) *Plant Cell 5*, 203–213.

Bax, A., & Davis, D. G. (1985) *J. Magn. Reson. 65*, 355–360.

Bode, W., & Huber, R. (1991) *Curr. Opin. Struct. Biol. 1*, 45–52.

Braünschweiler, L., & Ernst, R. R. (1983) *J. Magn. Reson. 53*, 521–528.

Brooks, B. R., Bruccoleri, R. E., Olafson, B. D., States, D. J., Swaminathan, S., & Karplus, M. (1983) *J. Comput. Chem. 4*, 187–217.

Brown, W. E., & Ryan, C. A. (1984) *Biochemistry 23*, 3410–3422.

Brünger, A. T. (1992) *X-PLOR version 3.0 Manual*, Yale University, New Haven, CT.

Brünger, A. T., Clore, G. M., Gronenborn, A. M., & Karplus, M. (1986) *Proc. Natl. Acad. Sci. U.S.A. 83*, 3801–3805.

Bundi, A., & Wüthrich, K. (1979) *Biopolymers 18*, 285–298.

Clore, G. M., Brünger, A. T., Karplus, M., & Gronenborn, A. M. (1986a) *J. Mol. Biol. 191*, 523–551.

Clore, G. M., Nilges, M., Sukuraman, D. K., Brünger, A. T., Karplus, M., & Gronenborn, A. M. (1986b) *EMBO J. 5*, 2729–2735.

Davis, D. G., & Bax, A. (1985) *J. Magn. Reson. 64*, 533–535.

Fersht, A. (1985) *Enzyme Structure and Mechanism*, W. H. Freeman and Co., New York.

Fujinaga, M., Read, R. J., Sielecki, A. R., Ardelt, W., Laskowski, M., Jr., & James, M. N. G. (1982) *Proc. Natl. Acad. Sci. U.S.A. 79*, 4868–4872.

Garcia-Olmedo, F., Salcedo, G., Sanchez-Monge, R., Gomez, L., Royo, J., & Carbonero, P. (1987) *Oxford Surv. Plant Mol. Cell Biol. 4*, 275–334.

Greenblatt H. M., Ryan, C. A., & James, M. N. G. (1989) *J. Mol. Biol. 205*, 201–228.

Greisinger, C., Sørenson, O. W., & Ernst, R. R. (1987) *J. Magn. Reson. 75*, 474–492.

Heath, R. J., Barton, P. A., Simpson, R. L., Reid, G. E., Lim, G., & Anderson, M. A. (1995) *Eur. J. Biochem. 230*, 250–257.

Heinz, D. W., Hyberts, S. G., Peng, J. W., Priestle, J. P., Wagner, G., & Grutter, M.G. (1992) *Biochemistry 31*, 8755–8766.

Hyberts, S. G., Goldberg, M. S., Havel, T. S., & Wagner, G. (1992) *Protein Sci. 1*, 736–751.

Jeener, J., Meier, B. H., Bachmann, P., & Ernst, R. R. (1979) *J. Chem. Phys. 71*, 4546–4553.

Kabsch, W., & Sander, C. (1983) *Biopolymers 22*, 2577–2637.

Kumar, A., Ernst, R. R., & Wüthrich, K. (1980) *Biochem. Biophys. Res. Commun. 95*, 1–6.

Laskowski, R. A., MacArthur, M. W., Moss, D. S., & Thornton, J. M. (1993) *J. Appl. Crystallogr. 26*, 283–291.

Merutka, G., Dyson, H. J., & Wright, P. E. (1995) *J. Biomol. NMR 5*, 14–24.

Nielsen K. J., Alewood, D., Andrews, J., Kent, S. B. H., & Craik, D. J. (1994a) *Protein Sci. 3*, 291–302.

Nielsen, K. J., Heath, R. L., Anderson, M. A., & Craik, D. J. (1994b) *J. Mol. Biol. 242*, 231–243.

Nilges, M., Clore, G. M., & Gronenborn, A. M. (1988) *FEBS Lett. 239*, 129–136.

Pallaghy, P. K., Duggan, B. M., Pennington, M. W., & Norton, R. S. (1993) *J. Mol. Biol. 233*, 405–420.

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

Pearce, G., Sy, L., Russell, C., Ryan, C. A., & Hass, G. M. (1982) *Arch. Biochem. Biophys. 213*, 456–462.

Pearce, G., Johnson, S., & Ryan, C. A. (1993) *Plant Physiol. 102*, 639–644.

Rance, M., Sørensen, O. W., Bodenhausen, G., Wagner, G., Ernst, R. R., & Wüthrich, K. (1983) *Biochem. Biophys. Res. Commun. 117*, 479–485.

Richardson, M. (1977) *Phytochemistry 16*, 159–169.

Richardson, M. (1991) *Methods Plant Biochem. 5*, 259–305.

Ryan, C. A. (1990) *Annu. Rev. Phytopathol. 28*, 425–429.

Schecter, I., & Berger, A. (1967) *Biochem. Biophys. Res. Commun. 27*, 157–162.

Skalicky, J. J., Selsted, M. E., & Pardi, A. (1994) *Proteins: Struct., Funct., Genet. 20*, 52–67.

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

Walsh, T. A., & Strickland, J. A. (1993) *Plant Physiol. 103*, 1227–1234.

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

Wilmot, C. M., & Thornton, J. M. (1990) *Protein Eng. 3*, 479–493.

Wishart, D. S., Sykes, B. D., & Richards, F. M. (1991) *J. Mol. Biol. 222*, 311–333.

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

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

Yamada, M., Santo, S., Yamaguchi, H., Tashiro, M., Ibuki, F., & Kanamori, M. (1977) *Agric. Biol. Chem. 41*, 2343–2347.

BI951394N