Sequential NMR Resonance Assignment and Structure Determination of the Kunitz-Type Inhibitor Domain of the Alzheimer's β-Amyloid Precursor Protein

Sarah L. Heald,*,‡ Robert F. Tilton Jr.,‡ Lisa J. Hammond,§ Alice Lee,§ Richard M. Bayney,§ Michael E. Kamarck,§ Triprayar V. Ramabhadran,§ Robert N. Dreyer,§ Gary Davis,§ Axel Unterbeck,§,∥ and Paul P. Tamburini§

Miles Research Center, Miles, Incorporated, and Molecular Therapeutics, Incorporated, 400 Morgan Lane, West Haven, Connecticut 06516

Received March 5, 1991; Revised Manuscript Received August 9, 1991

ABSTRACT: Certain precursor proteins (APP_{751} and APP_{770}) of the amyloid β-protein (AP) present in Alzheimer's disease contain a Kunitz-type serine protease inhibitor domain (APPI). In this study, the domain is obtained as a functional inhibitor through both recombinant ($APPI_r$) and synthetic ($APPI_s$) methodologies, and the solution structure of APPI is determined by ^1H 2D NMR techniques. Complete sequence-specific resonance assignments (except for P13 and G37 NH) for both $APPI_r$ and $APPI_s$ are achieved using standard procedures. Ambiguities arising from degeneracies in the NMR resonances are resolved by varying sample conditions. Qualitative interpretation of short- and long-range NOEs reveals secondary structural features similar to those extensively documented by NMR for bovine pancreatic trypsin inhibitor (BPTI). A more rigorous interpretation of the NOESY spectra yields NOE-derived interresidue distance restraints which are used in conjunction with dynamic simulated annealing to generate a family of APPI structures. Within this family, the β-sheet and helical regions are in good agreement with the crystal structure of BPTI, whereas portions of the protease-binding loops deviate from those in BPTI. These deviations are consistent with those recently described in the crystal structure of APPI (Hynes et al., 1990). Also supported in the NMR study is the hydrophobic patch in the protease-binding domain created by side chain–side chain NOE contacts between M17 and F34. In addition, the NMR spectra indicate that the rotation of the W21 ring in APPI is hindered, unlike Y21 in BPTI, showing a >90% preference for one orientation in the hydrophobic groove.

Research in Alzheimer's disease has recently focused on the mechanism of amyloid β-protein (AP) deposition in neuritic plaques and cerebrovasculature, a process believed to involve altered proteolysis [Glenner and Wong, 1984; Esch et al., 1990; for a review, see Muller-Hill and Beyreuther (1989)]. Targeted in these efforts are two membrane-bound AP precursor proteins, APP_{751}[1] and APP_{770} (110–130 kDa); each contains a 56 amino acid region (APPI) of unknown function but possesses a high degree of sequence homology with the superfamily of Kunitz-type serine protease inhibitors (Ponte et al., 1988; Tanaka et al., 1988; Tanzi et al., 1988). Elevated levels of APP_{751} have been correlated with plaque density in individuals with Alzheimer's disease (Johnson et al., 1990). The secreted form of APP_{751} is, in fact, protease nexin II (PNII), a protein involved in normal regulation of extracellular serine proteases (Oltersdorf et al., 1989; Van Norstand et al., 1989). APP_{751} inhibits a spectrum of serine proteases (Kitaguchi et al., 1988; Van Norstand et al., 1990), a property conferred on the precursor apparently through the APPI domain since APP_{751} and APPI exhibit indistinguishable protease inhibitory profiles (Dovey et al., 1990). Research presently suggests that, during amyloid formation, APP undergoes an altered process pathway from normal PNII maturation (Esch et al., 1990; Sisodia et al., 1990).

Besides its possible role in Alzheimer's disease, investigation into the structure of human APPI provides an opportunity to further understand the structural basis for the biological diversity found in the Kunitz inhibitor superfamily, as well as the functional role of Kunitz domains found in larger macromolecules. The Kunitz superfamily includes among its over 30 members [for a review, see Creighton and Charles (1987)] bovine pancreatic trypsin inhibitor (BPTI), dendrotoxins (neurotoxin and K^+ channel blocker; Lucchesi & Moczydlowski, 1990), inter-α-trypsin inhibitor (118 kDa; two Kunitz domains, Kaumeyer et al., 1986), and lipoprotein-associated coagulation inhibitor (38 kDa; three Kunitz domains; Wun et al., 1988). APPI and BPTI yield significantly different inhibitory profiles toward known proteases (Castro et al., 1990; Kido et al., 1990). Computational studies on APPI suggest that this difference in specificity pattern arises from structural differences occurring in the P2′ (M17) site in the protease-binding loops (Toma et al., 1989). Surprisingly, no Kunitz domain, identified in a large macromolecule through cDNA screening, has been isolated and analyzed by NMR to date.

This paper presents results on the generation of APPI and its structure elucidation by the integrated use of 2D NMR and dynamic simulated annealing. The Kunitz domain ($APPI_r$), as produced through the bacterial (*Escherichia coli*) expression of residues 289–345 of APP_{751} (Ponte et al., 1988), is obtained as a 65 amino acid functional inhibitor of trypsin activity. Alternatively, $APPI_s$ (56 residues) is synthetically generated and yields an inhibitory profile identical to that of $APPI_r$. Complete sequence-specific assignments of the ^1H NMR resonances (except for G37 NH and P13) are obtained for both

* Author to whom correspondence should be addressed.
‡ Miles Research Center, Miles, Inc.
§ Molecular Therapeutics, Inc.
∥ Present address: Bayer AG, Institut for Biochemie, Aprath 18A, D5600 Wuppertal, Germany.

[1] Abbreviations: APP, β-amyloid precursor protein; $APPI_r$, recombinant APP inhibitor domain; $APPI_s$, synthetic APP inhibitor domain; BPTI, bovine pancreatic trypsin inhibitor; RP-HPLC, reverse-phase high-performance liquid chromatography; FAB-MS, fast atom bombardment mass spectroscopy; NMR, nuclear magnetic resonance; COSY, 2D correlated spectroscopy; NOE, nuclear Overhauser effect; NOESY, 2D NOE spectroscopy; TOCSY, total correlated spectroscopy; RMS, root mean square.

0006-2960/91/0430-10467$02.50/0

$APPI_r$ and $APPI_s$ through standard procedures (Wuthrich, 1986). The secondary structural elements and tertiary fold in APPI are shown to be similar to those described in detail for BPTI (Wagner et al., 1981; Wagner & Wuthrich, 1982a). These assignments are supported by a temperature-based amide exchange study at pH 6.5 (Wagner & Wuthrich, 1982b). The NMR findings are summarized in a family of APPI structures refined using NOE-defined distance restraints with dynamic simulated annealing. The refined model of APPI is compared to BPTI, as well as to the recently reported crystal structure of APPI (Hynes et al., 1990).

Materials and Methods

Recombinant $APPI_r$. The portion of APP_{751} cDNA corresponding to the 56 amino acid Kunitz-type protease inhibitor domain (amino acid residues 289–345) was cloned by polymerase chain reaction into the bacterial expression vector pIN-III-ompA1 (Duffaud et al., 1987). Insertion of this minigene in-frame with the ompA1 secretion signal facilitates transport of the recombinantly expressed protein into the periplasmic space of *E. coli*. In our large scale fermentation experiments, we observed that the majority of the expressed protein was secreted into the media as an active protease inhibitor. A 200-L fermentation culture (M-9 media, containing 100 μg/mL ampicillin) was inoculated with a 4 L/16 h culture of *E. coli*, and the cells were grown at 37 °C (pH 7.4; sparge rate = 120 L/min; agitation = 200 rpm) to an A = 0.72 at 540 nm over 4.5 h, whereupon the culture was induced with 2 mM final filter-sterilized IPTG (isopropyl β-thiogalactopyranoside).[2] Three hours after induction, the cells were collected by centrifugation, and the culture media was concentrated to 1.75 L using a 10-kDa filtration membrane.

The concentrated media was purified on a trypsin Sepharose column preequilibrated with 100 mM TEA (triethylamine)/300 mM NaCl (pH 7.8). The loaded column was subsequently washed with equilibration buffer, then eluted with 250 mM KCl/10 mM HCl at 2.3 mL/min. Protein-containing fractions were combined, concentrated 2.3-fold on a YM-5 membrane (Amicon), and then applied directly to a 300-Å 21 × 250 mm Dynamax C_{18} column (RP-HPLC, Rainin). Separation was achieved at 25 °C using a linear gradient of 0–30% AcCN (acetonitrile) in 0.1% TFA (trifluoroacetic acid) at 10 mL/min over 150 min. A single symmetrical peak region corresponding to active $APPI_r$ was collected, lyophilized to dryness, and stored at −20 °C. The final preparation (28 mg) was homogeneous by electrophoresis; however, 20 cycles of automated N-terminal sequencing revealed the presence of a subpopulation (5% of total) in which the N-terminal dipeptide alanylasparagine was absent. The N-terminal amino acid sequencing was performed on an Applied Biosystems model 477A protein sequencer with an on-line model 120A PTH analyzer and Nelson Analytical model 2600 chromatography software. The following amino acid sequence[3] was deduced by nucleotide sequencing of the cDNA clone:

A-N^0-S-M-E-V-C-S-E-Q-A-E^{10}-T-G-P-C-R-A-M-I-S-R^{20}-

W-Y-F-D-V-T-E-G-K-C^{30}-A-P-F-F-Y-G-G-C-G-G^{40}-N-R-

N-N-F-D-T-E-E-Y^{50}-C-M-A-V-C-G-S-A-K-L^{60}-G-S-G

This sequence was consistent with FAB-MS analysis (MH^+ = 7000.3; calcd = 7000.3), amino acid composition (93.1% accuracy), and amino acid squencing of the peptide mixture generated by a cyanogen bromide digest.

Synthetic $APPI_s$. $APPI_s$ was synthesized on an Applied Biosystems model 420A peptide synthesizer using NMP-HOBt (*N*-methylpyrrolidone-1-hydroxybenzotriazole) Fmoc (9-fluorenylmethoxycarbonyl) chemistry (Fields & Noble, 1989). The peptide was cleaved and deprotected in 90% TFA, 4% thioanisole, 2% ethanedithiol, and 4% liquified phenol (2 h at room temperature). Refolding was performed as follows: A portion of the final crude peptide (260 mg) was dissolved in 50 mM Tris-HCl (pH 8.0)/8 M urea/5 mM DTT (dithiothreitol)/1 mM EDTA (ethylenediaminetetraacetic acid). After 1 h at 25 °C, the sample was dialyzed (2 times over 6 h) against 100 mM Tris-HCl/200 mM KCl/10 mM GSH (reduced glutathione)/1 mM GSSG (oxidized glutathione)/1 mM EDTA (pH 8.7) and then dialyzed overnight against 100 mM Tris-HCl/200 mM KCl/1 mM EDTA (pH 8.7). The refolded sample was further purified to electrophoretic homogeneity using the affinity and RP-HPLC chromatography procedures described for $APPI_r$ above. Unlike $APPI_r$, the RP-HPLC chromatogram showed the following elution peaks: 134.4 min (2% total), 146.5 min (6%), 150.0 min (72%), 164.2 min (13%), 167.9 min (4%), and 172.0 min (3%). The most abundant peak corresponded to full length $APPI_s$. The final preparation (6.5 mg) yielded the correct amino acid composition and molecular ion by FAB-MS analysis (MH^+ = 6154.3; calcd = 6154.4) for the predicted sequence:[3]

E-V-C-S-E-Q-A-E^{10}-T-G-P-C-R-A-M-I-S-R^{20}-W-Y-F-D-V-T-E-G-K-C^{30}-A-P-F-F-Y-G-G-C-G-G^{40}-N-R-N-N-F-D-T-E-E-Y^{50}-C-M-A-V-C-G-S-A

Biological Assays. APPI was routinely quantified according to its capacity to inhibit trypsin-catalyzed hydrolysis of N^{α}-benzoyl-L-arginine-*p*-nitroanilide (BAPNA). Following preincubation (20 min, 22 °C) of inhibitor with trypsin (bovine pancrease) buffered to pH 7.8, reactions were initiated by the addition of stock BAPNA (17 mg/mL in DMSO, dimethyl sulfoxide), thereby achieving the following initial component concentrations: APPI solution (diluted serially), trypsin (2 μg/mL), and BAPNA (0.37 mM) in 200 mM TEA buffer (pH 7.8). Formation of *p*-nitroaniline was followed spectrophotometrically at λ = 405 nm. A unit of inhibitor is defined as the amount of APPI necessary to produce a 50% inhibition of trypsin activity under the specified conditions.

APPI inhibition of both human leukocyte and porcine pancreatic elastase were similarly determined using *N*-t-boc-L-alanine-*p*-nitrophenyl ester (NBA) as substrate. Final reaction conditions contained APPI, human (93 nM) or porcine (260 nM) elastase, and NBA (0.17 mM) in 50 mM sodium phosphate buffer (pH 6.5).

Sample Preparation. The recombinant sample contained 4 mM $APPI_r$ in 90% H_2O/10% D_2O at pH 5.5, 36 °C, and at pH 6.5, 25 °C. The synthetic sample contained 2 mM $APPI_s$ in 90% H_2O/10% D_2O at pH 6.5, 25 °C. In both samples, the desired pH was reached by adding microliter amounts of HCl or NaOH. Both samples were internally referenced to sodium 3-(trimethylsilyl)propionate-2,2,3,3-d_4.

NMR Spectroscopy. NMR spectra were recorded at 500 MHz on a Bruker AM 500 spectrometer. All 2D NMR experiments employed time-proportional phase incrementation (TPPI; Marion & Wuthrich, 1983) in the pure-phase absorption mode. Except where noted, 2K t_2 × 512 t_1 experiments were collected using presaturation for water suppression. Double-quantum filtered COSY (DQF-COSY) spectra were recorded as described (Rance et al., 1983; Shaka & Freeman,

[2] Except where noted, all reagents were obtained from Sigma, St. Louis, MO.

[3] Numbering has been adjusted to match the Cys residues in BPTI: R-P-D-F-C-L-E-P-P-Y^{10}-T-G-P-C-K-A-R-I-I-R^{20}-Y-F-Y-N-A-K-A-G-L-C^{30}-Q-T-F-V-Y-G-G-C-R-A^{40}-K-R-N-N-F-K-S-A-E-D^{50}-C-M-R-T-C-G-G-A.

1983) with 700 t_1 records. In addition, three abbreviated DQF-COSY spectra were acquired sequentially at 25, 45, and 65 °C. 1D spectra were collected at 5 °C increments between each DQF-COSY experiment. 2D total correlation spectra (TOCSY; Braunschweiler & Ernst, 1983) utilized a MLEV-17 sequence (Davis & Bax, 1985) and z-filters (Rance, 1987) before and after the spin lock period. Hard pulses (19 μs) were used throughout the TOCSY sequence with a recovery delay (2.4 times the 90° pulse) inserted during the spin lock period (Griesinger et al., 1988) to minimize rotating-frame NOE effects. Isotopic mixing periods ranged from 32 to 46 ms. NOESY spectra (Jeener et al., 1979; Kumar et al., 1980) were recorded with mixing times of 50, 100, 150, and 200 ms. Mixing periods were randomly varied by 10% (Macura et al., 1981) to remove coherent transfer effects. One NOESY spectrum (100-ms mixing period) was acquired with a semi-selective jump–return read pulse (Plateau & Gueron, 1982) for water suppression. All data were processed on a SUN 4/110 or a VAX 11/750 equipped with an array processor using the FTNMR software from D. Hare (Hare Research, Woodinville, WA).

Molecular Modeling. Using the program XPLOR (Brunger, 1988; topology file topall6x.pro; parameter file parmall3x.pro), 15 independent initial structures were generated with random backbone torsion angles ϕ and ψ. These initial structures were subjected to simulated annealing refinement in five stages. First, to provide an energetically feasible starting structure, 2500 steps of Powell energy minimization were performed. Next, 15 ps of high-temperature molecular dynamics (1000 K) were run using 0.002-ps time steps, fixed bond lengths (SHAKE), a greatly reduced repulsive weight in the van der Waals term (0.002) to allow atoms to move freely, and a "soft square well" potential (r_{SW} = 0.5 Å; asymptote slope = 0.1) as the effective NOE energy term. A force constant (k_{NOE}) for the experimental NOE distance restraints of 50.0 kcal/(mol·Å^2) and geometric center averaging were used for ambiguous assignments, e.g., methylene or methyl protons. In the next stage, the slope of the asymptote was increased to 1.0, and the repulsive weight in the van der Waals term was increased to 0.1 in 10 sequential steps of 1.0-ps simulations using the old velocities from the end of the previous 1000 K simulation. The system was then slowly cooled (annealed) to a temperature of 300 K during the next 2.8 ps with the incorporation of slightly smaller atomic van der Waal parameters. In the final stage, 200 steps of Powell minimization were performed. Details of the simulated annealing have been published (Brunger, 1990). All of the refinement simulations were carried out in the absence of the internally bound waters or surrounding solvent molecules, and the electrostatic contributions from Arg, Lys, Glu, and Asp residues were not included. Information regarding the disulfide pairing of the six cysteines within APPI were introduced into the simulated annealing refinement in different ways. According to the homology with BPTI, the six cysteines were modeled as three disulfides (14–38, 5–55, and 30–51) and were constrained by covalent bonds from the initial starting structure or were refined with the S–S distance restraint of 2.02 ± 0.1 Å. In other refinements, no a priori knowledge of the disulfide pairing was used, and only the NMR-derived NOE distance restraints were included. These NOE distance restraints were divided into three categories: (1) strong, 1.8–2.3 Å (strong NOESY cross peak at a mixing time of 50 ms); (2) medium, 1.8–3.3 Å (medium or weak NOE connectivity at a mixing time of 50 ms); and (3) weak, 1.8–5.0 Å (very weak NOE at 50 ms or weak NOE at 100 ms). Other NOESY cross peaks appearing at longer mixing times (150 and 200 ms) were not included in the refinement. Each structure was analyzed for NOE violations. Ten of the structures was assessed as to their RMS geometric differences.

Table I

purification step	tot vol (mL)	protein conc[a] (mg/mL)	tot IC_{50}[b] (Units)	spec act. (Units/mg)	act. recovered (%)
(A) Purification of Secreted $APPI_r$					
concentrated media	1747	2.57	88 121	17.37	100
trypsin affinity eluate	54	0.63	113 684	3334	100
RP-HPLC eluate	110.5	0.26	80 364	2850	91
(B) Purification of Synthetic $APPI_s$					
refolded peptide	75	3.47	75 000[c]	258	100
trypsin affinity eluate	41.5	0.41	46 696	2746	62
RP-HPLC eluate	14.5	0.45	21 970	3390	29

[a] Protein concentration values obtained using the BCA assay (Pierce). [b] Determined at an assay concentration of 2 μg/mL trypsin. [c] Value estimated from a nonlinear IC_{50} plot.

Results and Discussion

Biochemistry. Two protein samples, $APPI_r$ and $APPI_s$, were produced in the course of our investigation of the Kunitz-type protease inhibitor domain. $APPI_r$ was the result of recombinant methodology in which a minigene was cloned by PCR amplification and expressed in *E. coli* cells. $APPI_r$ was found to contain an additional four and five amino acid extensions at the N- and C-termini, respectively. Attempts at propagating the secretion vector containing only the desired 56 amino acids were unsuccessful due to the apparent instability of the vector in *E. coli*. Despite this, both periplasmic extracts and media isolated from the bacterial cultures showed the same trypsin inhibitory activity. To be absolutely certain that the additional amino acids did not adversely affect the structure of APPI, an alternate source of the inhibitor, $APPI_s$ (56 residues), was obtained via solid-phase peptide synthesis followed by an oxidative folding procedure.

Both recombinant $APPI_r$ and synthetic $APPI_s$ were effectively purified using a two-step procedure (Table IA,B). In the case of recombinant $APPI_r$, the trypsin affinity chromatography selectively isolated 0.8% of the protein and 100% of the trypsin inhibitory activity present in the media so that subsequent chromatography was needed only to remove trace contaminants. For $APPI_s$, the starting synthetic mixture contained many truncated and quasifolded peptides. This mixture was further complicated by degradation during the cleavage process and by nonspecific cross-linking during the folding procedure. While the trypsin affinity chromatography effectively removed many of these byproducts, a lengthy RP-HPLC procedure proved important in obtaining $APPI_s$ purified to homogeneity.

In spite of the difference in protein preparation, $APPI_r$ and $APPI_s$ showed comparable specific activities (Table IA,B). Apparently the presence of the additional N- and C-terminal amino acids in $APPI_r$ did not affect the inhibitory function of the molecule. Indeed, the potency of trypsin inhibition for $APPI_r$ was tight enough to be stoichiometric even at enzyme concentrations <10 nM, which was consistent with the K_i of 2–3 nM reported for APPI (Dovey et al., 1990). Also in

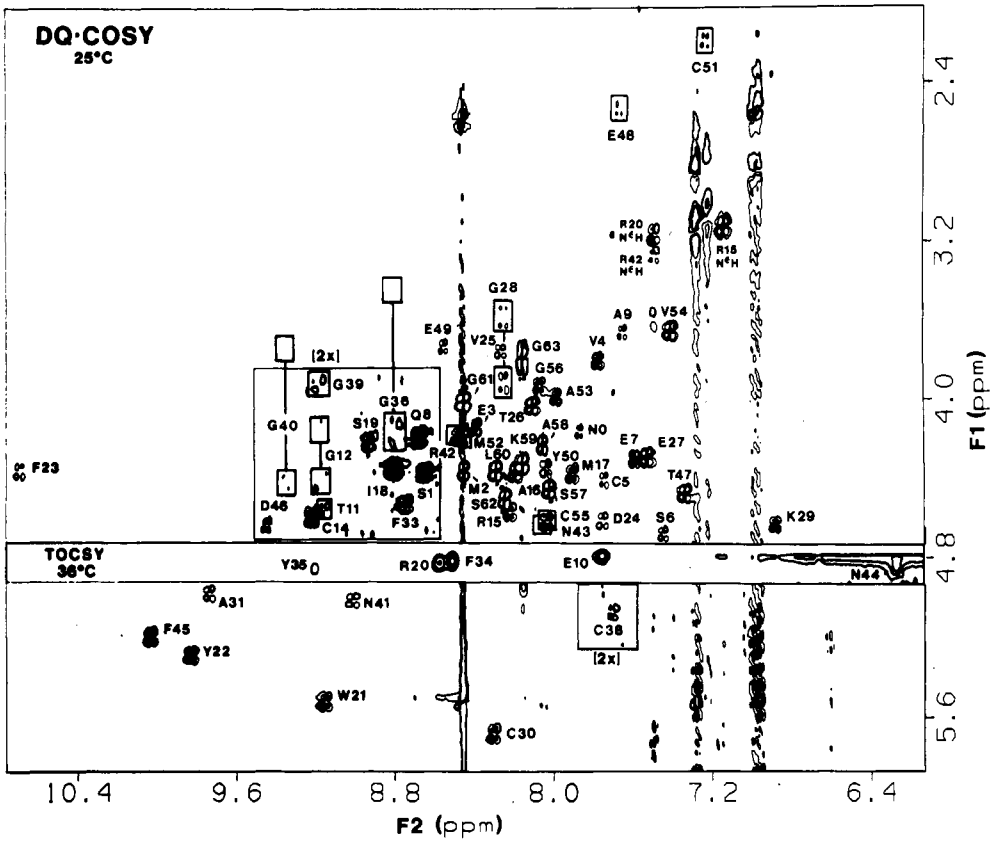

FIGURE 1: Sequential assignments for $APPI_r$ (pH 6.5) in the DQF-COSY fingerprint region at 25 °C. Boxed areas, marked 2×, are plotted at a lower contour level. The F_2 water stripe has been overlaid with the same region in the (pH 5.5, 36 °C) TOCSY spectrum.

agreement with the reported inhibitory profile for APPI (Kitaguchi et al., 1990), $APPI_r$ was only a weak inhibitor of human leukocyte elastase (K_i approximately 0.83 μM) and did not inhibit porcine pancreatic elastase at concentrations >15 μM.

NMR and Modeling Strategy. The ^1H NMR spectra of several Kunitz-type trypsin inhibitors, BPTI [for a review, see Wuthrich (1986)], inhibitor E (Arseniev et al., 1982), and inhibitor K (Keller et al., 1983), have been assigned by 2D NMR techniques. APPI and BPTI have >40% sequence homology, and, more importantly, the six cysteines, as well as several other residues located in the trypsin binding domain, are conserved. Conversely, $APPI_r$ has nine additional residues, a proline substitution in what should be β-sheet and an acidic instead of a basic pI. Our strategy, therefore, was to empirically complete the sequence-specific assignments of the ^1H NMR resonances and identification of the secondary structural elements, then to use NOE-defined distance restraints and dynamic simulated annealing to generate a family of APPI structures. This approach to modeling NMR data proved quite successful in the recent study of thymosin β_4 (Zarbock et al., 1990).

Our investigation began with $APPI_r$, the 65 amino acid inhibitor. Protein resonance assignments were obtained in two steps: (1) assignment of scalar coupled spin systems according to amino acid type and (2) utilization of backbone NOEs, primarily NH(i)–NH(i+1) and C^αH–NH(i+1), to place the residue in the sequence. $APPI_r$ is composed of 23 unique spin systems (9 Gly, 6 Ala, 3 Thr, 3 Val, 1 Ile, and 1 Leu), 25 AMX[4] spin systems (2 Asp, 4 Asn, 6 Cys, 5 Ser, 4 Phe, 3 Tyr, and 1 Trp), and 17 long-chain spin networks (3 Arg, 6 Glu, 1 Gln, 2 Lys, 3 Met, and 2 Pro). This protein contains a diverse set of amino acid types which maintain a fairly stable globular structure giving rise to well-dispersed NMR resonances (see Figure 1). This meant that standard 2D NMR techniques, such as COSY, TOCSY, and NOESY, proved sufficient for our purposes. Ambiguities in the resonance assignments were, for the most part, resolved by obtaining spectra under two unique sample conditions (pH 5.5, 36 °C, and pH 6.5, 25 °C). Once the assignment process neared completion, regular elements of secondary structure were readily deduced from qualitative interpretation of the NOESY spectra (Wuthrich, 1986).

$APPI_r$ and Spin System Assignments. The acidic pI (4.5) of APPI precluded the use of NMR sample conditions reported for other trypsin inhibitors, i.e., pH 3.2–4.6 (Arseniev et al., 1982; Wagner & Wuthrich, 1982a; Keller et al., 1983). Spectra acquired at low pH (≤3.2) contained a wide range of line widths (5–50 Hz). Raising the pH one unit above the pI gave well-resolved ^1H NMR spectra with fairly uniform line widths (<12 Hz). A standard starting set of 2D NMR spectra (DQF-COSY, TOCSY, and NOESY) on $APPI_r$ was collected at pH 5.5 (36 °C). The resulting scalar coupling spectra appeared quite reasonable. However, the NOESY spectra contained fewer than the expected number of cross peaks. Comparison of the NOESY and TOCSY spectra revealed that many of the intraresidue NOEs were weak or absent, but, more importantly, the fingerprint region apparently contained only short-range NOEs. Spectral effects such as these are typically attributed to unfavorable protein dynamics and could be caused by either the temperature being too high and/or the pH being too close to the pI of the protein. Sample conditions were adjusted to pH 6.5 (24 °C), and the experiments were repeated (Table II).

[4] AMX spin systems include amino acids which have a functional substituent at the β-carbon causing the $C^{\beta1,\beta2}H_2$ resonances to be shifted to a unique region in the ^1H NMR spectrum (Wuthrich, 1986).

Table II: ^{1}H NMR Chemical Shifts (in ppm) of APPI$_r$ at 25 °C and pH 6.5[a]

residue	NH	C$^\alpha$H	C$^\beta$H	others
A(-1)		4.14	1.52	
N0	7.87	4.17	3.83	N$^\delta$H$_2$ 7.67, 7.00
S1	8.65	4.39	3.97, 3.95	
M2	8.46	4.37	2.11, 2.09	C$^\gamma$H 2.63, 2.58; C$^\epsilon$H$_3$ 2.07
E3	8.40	4.15	2.10, 2.08	C$^\gamma$H 2.33
V4	7.77	3.81	2.11	C$^\gamma$H$_3$ 0.99, 0.91
C5	7.75	4.42	2.96, 2.86	
S6	7.45	4.69	4.06, 3.95	
E7	7.59	4.31	2.31, 2.17	C$^\gamma$H 2.65
Q8	8.68	4.21	2.09, 1.96	C$^\gamma$H 2.47, 2.43; N$^\epsilon$H$_2$ 7.70, 6.89
A9	7.65	3.67	−0.39	
E10	7.78	4.80	1.96, 1.80	C$^\gamma$H 2.13
T11	9.16	4.59	4.33	C$^\gamma$H$_3$ 1.54
G12	9.19	4.44, 4.18		
P13				(C$^\gamma$H 1.98; C$^{\delta1,\delta2}$H 3.72, 3.62)
C14	9.22	4.61	3.31, 2.88	
R15	8.23	4.58	1.57, 1.53	C$^\gamma$H 1.60, 1.17; C$^\delta$H 3.13; N$^\epsilon$H 7.14
A16	8.20	4.38	1.25	
M17	7.91	4.39	1.77, 1.50	C$^\gamma$H 2.27, 2.15; C$^\epsilon$H$_3$ 1.62
I18	8.81	4.37	1.94	C$^\gamma$H 1.54, 1.12; C$^\gamma$H$_3$ 1.03; C$^\delta$H$_3$ 0.86
S19	8.94	4.24	3.88, 3.75	
R20	8.57	4.85	1.95, 0.84	C$^\gamma$H 1.53, 1.44; C$^\delta$H 3.41, 3.18; N$^\epsilon$H 7.49
W21	9.16	5.54	3.10, 2.82	C$^{\delta1}$H 6.81; C$^{\zeta2}$H 7.50; C$^\eta$H 7.17; C$^{\zeta3}$H 6.92; C$^{\epsilon3}$H 7.05; N$^{\epsilon1}$H 10.03
Y22	9.84	5.30	2.72	C$^\delta$H 6.97; C$^\epsilon$H 6.61
F23	10.70	4.36	3.54, 2.88	C$^\delta$H 7.40; C$^\epsilon$H 7.01; C$^\zeta$H 7.08
D24	7.76	4.63	2.79, 2.02	
V25	8.27	3.77	2.33	C$^\gamma$H$_3$ 1.20, 1.16
T26	8.11	4.04	4.35	C$^\gamma$H$_3$ 1.27
E27	7.53	4.30	2.17, 1.56	C$^\gamma$H 2.27, 2.04
G28	8.25	3.93, 3.59		
K29	6.88	4.65	1.91, 1.79	C$^\gamma$H 1.68; C$^\delta$H 1.15; C$^\epsilon$H 3.00
C30	8.31	5.70	3.74, 2.95	
A31	9.75	4.99	1.20	
P32		4.48	1.28, 0.81	C$^\gamma$H 1.68; C$^\delta$H 3.39, 3.53
F33	8.75	4.55	3.12, 2.88	C$^\delta$H 6.99; C$^\epsilon$H 7.15; C$^\zeta$H 6.87
F34	8.55	4.84	2.88, 2.72	C$^\delta$H 7.00; C$^\epsilon$H 7.29; C$^\zeta$H 7.22
Y35	9.21	4.86	2.59, 2.53	C$^\delta$H (6.83, 7.80); C$^\epsilon$H 6.70
G36	8.81	4.19, 3.44		
G37		4.32, 3.13		
C38	7.71	5.07	3.83, 2.94	
G39	9.19	3.95, 3.95		
G40	9.36	4.43, 3.80		
N41	9.01	5.03	3.42, 2.76	N$^\delta$H$_2$ 8.36, 7.84
R42	8.49	4.20	2.03, 1.83	C$^\gamma$H 1.75; C$^\delta$H 3.28, 3.18; N$^\epsilon$H 7.50
N43	8.06	4.71	3.18, 2.92	N$^\delta$H$_2$ 8.15, 7.95
N44	6.19	4.82	2.56, 2.40	N$^\delta$H$_2$ 7.08, 6.39
F45	10.04	5.21	3.35, 2.84	C$^\delta$H 7.23; C$^\epsilon$H 7.67; C$^\zeta$H 7.20
D46	9.46	4.65	2.89, 2.19	
T47	7.34	4.49	4.41	C$^\gamma$H$_3$ 1.22
E48	7.67	2.54	1.01, −0.69	C$^\gamma$H 1.76, 1.47
E49	8.56	3.74	1.93, 1.79	C$^\gamma$H 2.27, 2.18
Y50	8.04	4.36	3.16, 2.96	C$^\delta$H 7.29; C$^\epsilon$H 6.97
C51	7.23	2.20	3.29, 3.04	
M52	8.43	4.16	2.00, 1.91	C$^\gamma$H 2.71, 2.63; C$^\epsilon$H$_3$ 2.21
A53	7.99	4.00	1.43	
V54	7.41	3.66	1.51	C$^\gamma$H$_3$ 0.86, 0.68
C55	8.04	4.63	2.14, 1.77	
G56	8.08	3.95, 3.93		
S57	8.03	4.47	3.89, 3.87	
A58	8.06	4.24	1.31	
K59	8.16	4.34	1.87, 1.77	C$^\gamma$H 1.70; C$^\delta$H 1.45; C$^\epsilon$H 3.02
L60	8.29	4.37	1.71, 1.62	C$^\gamma$H 1.66; C$^\delta$H$_3$ 0.95, 0.89
G61	8.46	4.03, 4.01		
S62	8.25	4.52	3.93, 3.88	
G63	8.16	3.83, 3.75		

[a] Chemical shifts are referenced to sodium 3-trimethylsilylpropionate-2,2,3,3-d_4.

In the adjusted sample conditions, the ^{1}H NMR line width was narrower by 10–15% contrary to the known relationship between line width and temperature. Nearly all of the intraresidue NOEs were significantly stronger due presumably to decreased side-chain motion at the lower temperature. Qualitative comparison of the (100 ms) NOESY spectra, after sequence-specific assignments were complete, showed that all of the medium-range NOEs important in defining the helical and turn regions of the protein and a number of the weak long-range NOEs were only visible at pH 6.5 (24 °C). This may be a consequence of increased motion associated with exchange reactions on the surface of the protein resulting from the pH being too close to the pI.

The fingerprint region from the DQF-COSY spectrum at APPI$_r$ collected at 25 °C is labeled in Figure 1 with the sequence-specific assignments. Within the unique spin systems,

the assignment of the nine glycines was marginally problematic. Only G28 showed connectivity to all three spins in the COSY and TOCSY spectra. The complete spin networks for G12, G36, and G40 were visible only in the TOCSY spectrum. Other glycines (G39, G56, G61, and G63) were seen in the COSY spectrum, but their $C^{\alpha 1}H$ and $C^{\alpha 2}H$ resonances were determined to be degenerate. The G37 NH resonance has been assigned in the spectrum of BPTI (Tuchsen & Woodward, 1987) at the unusual chemical shift of 4.3 ppm. This assignment was based, in part, on an observed NOE from G36 NH. In the spectra of APPI, G36 NH was degenerate with I18 NH, which, in turn, showed a strong NOE to M17 $C^{\alpha}H$ at 4.3 ppm. This double degeneracy prevented the unambiguous assignment of G37 NH.

Among the other unique spin systems, the six alanines were readily identified at both 25 and 36 °C. The assignment of both T11 and T47 involved recognizing degeneracies which were resolved at different temperatures. In both V4 and V25, magnetization from the amide resonance was not relayed beyond $C^{\alpha}H$. Fortunately, in both cases the $C^{\alpha}H$ resonances occurred in a well-resolved region of the spectrum so that the complete side-chain assignment was derived from the $C^{\alpha}H$ resonance. The remaining spin networks for the unique residues were readily assigned from relays observed from their respective NH resonances in the TOCSY spectra.

In 17 of 25 AMX[4] spin systems, both NH–$C^{\beta 1}H$ and NH–$C^{\beta 2}H$ TOCSY cross peaks were seen. Another six showed only one NH–$C^{\beta}H$ cross peak, but the second β-resonance was readily identified in the upfield portions of the spectra. In the last two (both cysteines), no relayed cross peaks were observed from the amide resonance, but both had a well-resolved $C^{\alpha}H$ resonance enabling unambiguous assignment of the β-protons. The assignment of the AMX spin systems was facilitated by subdividing the residues into the following three groups on the basis of unique NMR characteristics: aromatics, serines, and Cys/Asp/Asn.

The aromatic residues were identified from the assignment of the aromatic spin systems. The single tryptophan (W21) ring gave the only aromatic four-spin system. The two other indole resonances, $C^{\delta 1}H$ and $N^{\epsilon 1}H$, were assigned on the basis of their observed coupling in the TOCSY and NOESY spectra. The ring was then connected by NOEs to an AMX spin system, completing the assignment of W21. The four phenylalanine rings were assigned to the aromatic three-spin systems. For both F23 and F45, a phenyl ring was then matched using the NOESY spectra to an AMX spin system to complete their assignment. The $C^{\delta}H$ ring resonances in residues F33 and F34 were degenerate with each other and with Y22 $C^{\delta}H$, making their complete assignment somewhat tenuous at this stage. The two-spin tyrosine ring protons were differentiated from the two-spin side-chain $N^{\delta}H_2$ resonances by comparison of the data sets collected at different pH units. In addition, the hindered rotation of the aromatic rings in Y35 and W21 (see *Unique Spectral Features*) resulted in two sets of resonances with associated exchange NOEs. The slow rotation of the Y35 ring was also observed in the other assigned BPTI homologoues (Wuthrich, 1986).

All five serines in $APPI_r$ showed connectivities to complete spin networks in the TOCSY spectra, and their β-protons were all located characteristically between 4.05–3.70 ppm. Of the remaining AMX spin systems, three asparagines (N41, N43, and N44) showed weak NOEs between the side-chain $N^{\delta}H_2$ and β-proton resonances, making their complete assignment unambiguous. N0 was assigned, in part, by the absence of its resonances in the spectra of $APPI_s$. The final classification of the two aspartic acids and six cysteines was carried out during the sequential assignment process. The six cysteines are conserved in all members of the Kunitz family. Excellent agreement was found between our assignments and those reported for BPTI and its two homologues (Heald et al., 1990).

Only one of the 17 long-chain spin systems (P13) could not be completely assigned. Tentative assignments have been made for P13 $C^{\delta 1,\delta 2}H$ and $C^{\gamma}H$ on the basis of weak cross peaks to G12 NH and $C^{\alpha}H$ and C14 NH. The remainder of the spin system could not be identified due to multiple degeneracies. The other proline, P32, was readily assigned from the observed TOCSY cross peak pattern. The β-proton resonances were shifted by >1 ppm upfield and showed good relayed cross peaks to the $C^{\alpha}H$ and $C^{\delta 1,\delta 2}H$ resonances. Two of the glutamic acids, E48 and E49, did not show any relayed cross peaks from the amide resonance. Fortunately, both residues had well-resolved $C^{\alpha}H$ resonances which provided complete side-chain connectivity patterns. The other spin systems (Met, Glu, and Gln) were assigned from reliable NH–$C^{\gamma 1,\gamma 2}H$ TOCSY cross peaks and were differentiated during the sequential assignment process. The $-SCH_3$ resonances were assigned on the basis of the BPTI literature (Wagner & Wuthrich, 1982a) and comparison of the $APPI_r$ and $APPI_s$ spectra. The Arg and Lys residue were differentiated, in part, by the presence or absence of the side-chain amine resonance. The $N^{\epsilon}H$ resonance in arginine was seen at pH 6.5, whereas the $N^{\epsilon}H_3^+$ resonance in lysine was not visible. Each arginine showed NH–$C^{\beta 1,\beta 2}H$ relayed cross peaks and $N^{\epsilon}H$–$C^{\delta 1,\delta 2}H$ cross peaks. These were connected by cross-peak patterns between the α- and δ-proton resonances in the TOCSY spectra. Both lysines gave discernible NH–$C^{\gamma}H_2$ TOCSY cross peaks and $C^{\beta 1,\beta 2}H$–$C^{\epsilon 1,\epsilon 2}H$ cross peaks, allowing the two ends of the spin system to be correctly aligned.

Sequential Assignments. There is no clear point at which amino acid typecasting ends and sequential assignment begins or where sequential assignment ends and the identification of secondary structural elements begins. The sequential assignment process (Wagner et al., 1981; Wuthrich, 1986) relies upon strong NOE interactions between backbone NH and $C^{\alpha}H$ resonances on adjacent residues. Typically, the most important NOESY cross peaks to be assigned are NH(i)–NH(i+1) ($=d_{NN}$), $C^{\alpha}H(i)$–NH(i+1) ($=d_{\alpha N}$), and $C^{\beta}H(i)$–NH(i+1) ($=d_{\beta N}$). Local secondary structure dictates which of these short-range NOEs will dominate. In β-sheet, a series of strong $d_{\alpha N}$ and $d_{\beta N}$ connectivities are found, whereas helical and turn regions give strong d_{NN} connectivities. Loop regions are less predictable. Strategically approaching sequential assignment in $APPI_r$, there are four marker residues (Q8, I18, W21, and L60), amino acids which occur only once in the primary sequence. Every individual residue in $APPI_r$ occurs within a unique tripeptide segment. In fact, there are only five pairs of redundant dipeptide segments.

As a prerequisite to sequential assignment and secondary structure determination, a series of time-dependent NOESY spectra were collected at both 25 and 36 °C. A summary of the short- and medium-range NOEs which have been assigned in the 50- and 100-ms NOESY spectra of $APPI_r$ and subsequently used to define distance restraints are schematically represented in Figure 2. The fingerprint region from the 100-ms NOESY spectrum (25 °C) is shown in Figure 3. Three stretches of the sequence (residues 17–25, 28–[P32]–36, and 44–46), well defined by $d_{\alpha N}$ and $d_{\beta N}$ connectivities, are identified. Other segments (not labeled in Figure 3) which gave similar well-resolved $d_{\alpha N}$ and $d_{\beta N}$ NOE patterns included residues 7–10, 14/15, 39–42, and 43/44. The d_{NN} connec-

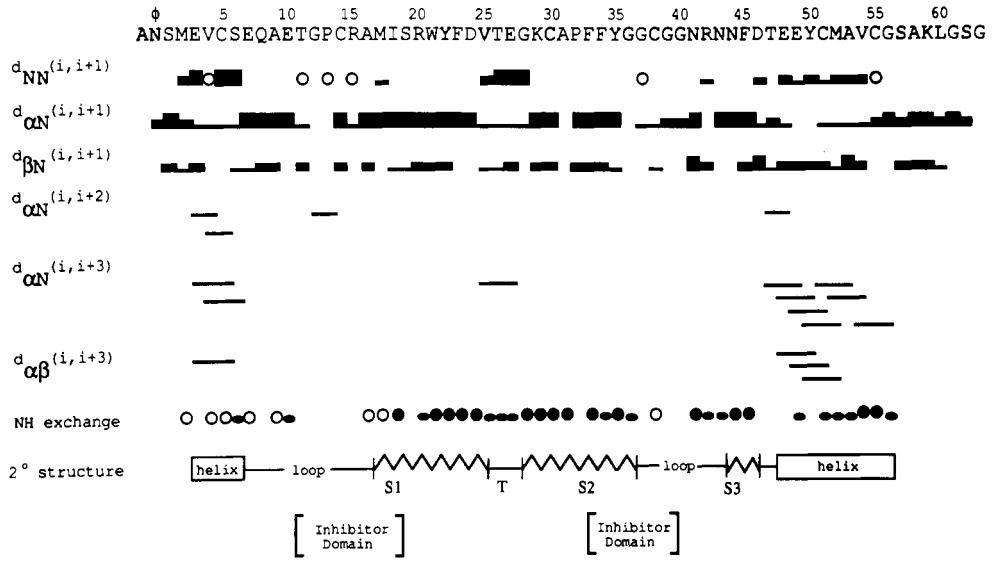

FIGURE 2: Summary of the short- and medium-range NOEs assigned in the 50- and 100-ms NOESY spectra (pH 6.5, 25 °C). The thickness of the line corresponds to a strong, medium, or weak NOE intensity. Open circles and dashed lines indicate ambiguities in the data (degeneracies or incomplete assignments). Amide resonances, residual at 65 °C, are shown at the bottom (NH exchange) as filled circles; half circles represent weak cross peaks. Regions of secondary structure and the protease-binding domain are delineated figuratively.

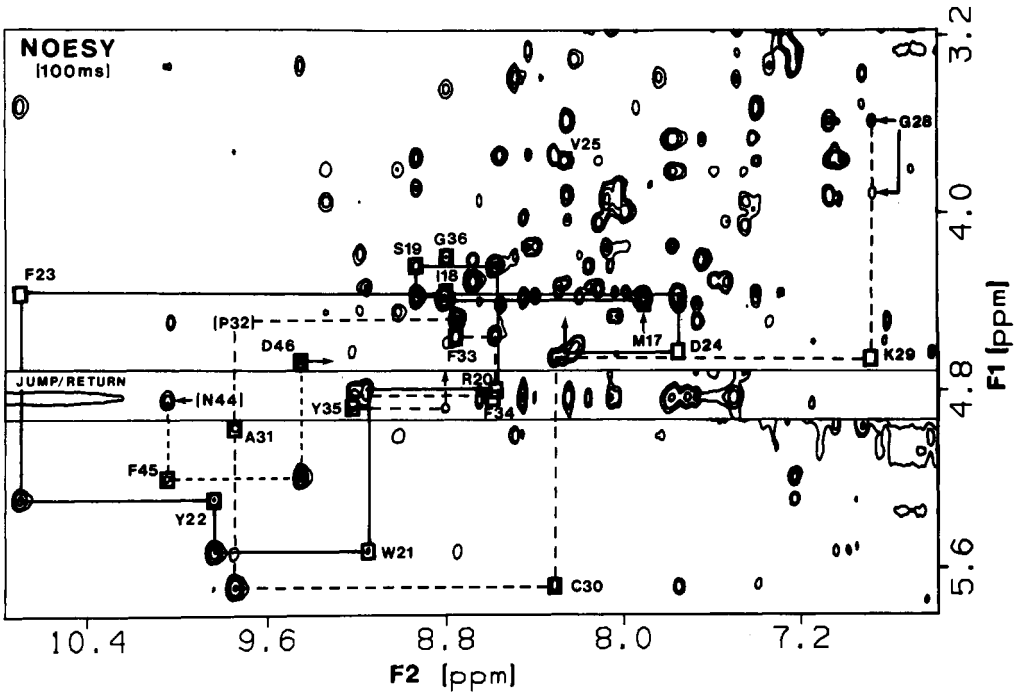

FIGURE 3: $d_{\alpha N}$ connectivities for residues M17–V25, G28–[P32]–G36, and N44–D46 observed for $APPI_r$ in the 100-ms NOESY (pH 6.5, 25 °C) spectrum. The F_2 water stripe has been covered with the same region from the jump–return NOESY (pH 6.5, 25 °C) spectrum.

tivities for residues 2–7, 25–29, and 48–56 have been labeled in the NOESY spectrum in Figure 4.

Connecting the $d_{\alpha N}$ and $d_{\beta N}$ NOE maps listed above and shown in Figure 3 was relatively straightforward due to the distribution of amino acid types and the dispersion in NH and $C^{\alpha}H$ resonances. Three of the four marker amino acids, as well as seven of the eight aromatic residues, were located in these segments. The one completely assigned proline spin system was readily identified as P32 from the $d_{\alpha N}$ and $d_{\beta N}$ connectivities to F34 as well as from NOEs between P32 $C^{\delta 1,\delta 2}H$ and A31. Extending the assignments from I18 back to C14 was somewhat obscured by the functional degeneracy (chemical shift difference <0.05 ppm) of the I18, M17, and A16 $C^{\alpha}H$ resonances and the R15 and A16 NH resonances. These assignments were resolved by interresidue side-chain and medium-range NOEs and, interestingly enough, by the occurrence of this same pattern of $C^{\alpha}H$ degeneracy in BPTI. These two long segments (C14–V25 and G28–G36) were then connected by a sequence of d_{NN} NOEs observed for V25–K29 (see Figure 4).

The next long segments (M2–E10 and E48–G56) were identified mainly from d_{NN} connectivities (Figure 4) and contained only one functional NH degeneracy each (V4/C5 and C55/G56). The sequential assignments were then extended from G56 to the marker residue, L60, via $d_{\alpha N}$ and $d_{\beta N}$ connectivities. Assignments were verified during the analysis of $APPI_s$ since residues A(−1)–M2 and A58–G63 were absent in this protein. This left three tripeptide segments (A(−1)–S1, T11–P13, and G61–G63) to be identified. T11–P13 was not apparent at 25 °C due to the functional degeneracy of the T11, G12, W21, and G39 NH resonances. These were, however, resolved at 36 °C. P13 could only be assigned tentatively on

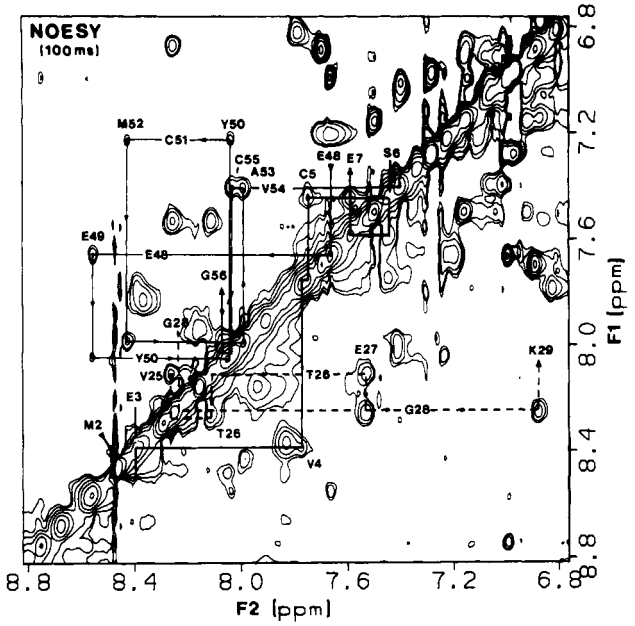

FIGURE 4: d_{NN} connectivities for residues M2–E7, V25–K29, and E48–G56 as seen in the 100-ms NOESY (pH 6.5, 25 °C) spectrum of $APPI_r$.

the basis of weak NOEs from G12 and C14. The final two segments occur in the N- and C-termini of the sequence. These residues were sequenced, for the most part, by $d_{\alpha N}$ NOEs and were consistent with the subsequent assignment of $APPI_s$.

Secondary Structure and Tertiary Folds. The unique NOESY cross-peak patterns which signify regions of β-sheet, helix, and turns have been described in detail elsewhere (Wuthrich, 1986). These patterns were readily recognized in the spectra of $APPI_r$ and can be seen in schematic representation of the unambiguous interresidue NOESY cross peaks which have been catalogued (Figure 2). The presence of continuous d_{NN} and $d_{\alpha N}(i,i+3)$ connectivities indicated that residues 3–7 and 48–56 formed helical structures. The short segment of helix (E3–E7) also showed regular $d_{\alpha N}(i,i+2)$ NOEs. The NOESY spectra were further analyzed for $d_{\alpha N}(i,i+4)$ connectivities within the longer stretch of helix (E48–G56) in order to distinguish α-helix from 3_{10}-helix. Only one could be unambiguously assigned due to degeneracy.

Segments showing continuous $d_{\alpha N}$ and $d_{\beta N}$ connectivities typically occur in rigid β-sheet or in irregular loop regions. The hydrogen-bonding network unique to the β-sheet structure produces a pattern of slowly exchanging amide resonances and stretches of long-range backbone–backbone NOEs which align these segments into parallel or antiparallel configurations. The NOE pattern associated with the β-sheet region in BPTI has been described in detail (Wagner et al., 1981). A similar pattern was identified in the NOESY spectra of $APPI_r$. The amide exchange rates in $APPI_r$ were categorized by using temperature rather than the more commonly used deuterium method (see *Hydrogen Exchange* and Figure 2). The β-sheet, seen in $APPI_r$, was composed of two long antiparallel strands, M17–V25 and G28–G36, and a third shorter strand, N44–D46, which aligned with R20–Y22 in an antiparallel configuration. The alignment of the β-strands was fairly straightforward with the exception that three of the five possible $C^{\alpha}H$–$C^{\alpha}H$ cross peaks (signatures of antiparallel β-sheet) could not be seen due to degeneracy with the water resonance.

A tight turn centered at T26 and E27, typical of type I or type I′, was characterized by strong d_{NN} NOEs and weak $d_{\alpha N}$ connectivities (Figure 2 and 4). This pattern was also found for the two residues (V25 and G28) at the junction between the β-sheet and the tight turn, suggesting an irregularity in the β-sheet. The second and third β-strands were interrupted by an irregular loop region (G37–N43) as characterized by an NOE pattern inconsistent with regular secondary structure. Another such region (Q8–A16) was found between the first short helix and the β-sheet. Portions of both loop regions contain the stretches of sequence (T11–I18 and F34–G39) which are homologous with the protease-binding domain in the Kunitz family. The first four residues of the N-terminus and the last seven residues of the C-terminus did not give NOE pattern indicative of well-ordered structure. In addition, no NOEs were found between these residues and the remainder of the globular protein structure.

Extensive patterns of long-range NOEs indicated the tertiary folding of the regular elements of secondary structure. The short helix was aligned so that S6/E7 was in contact with both the β-sheet (F23/V25) and the second loop region (N41–N43). The short helix was connected to the first loop region which crossed over the β-sheet (NOEs: A9–Y22/F33) and was aligned with the second loop region. These two loop stretches formed an isolated domain as seen from the numerous cross-strand connectivities (NOEs: E10–N41, T11–G36/G37, G12–G39, C14–C38, and A16–G36). The alignment of the loop regions was interrupted by the β-sheet which made up the hydrophobic core of the protein. This core was tightly defined by both backbone–backbone and side chain–side chain NOE interactions. The long helix was then aligned with the β-sheet so that E48 was in close contact with the side chain of W21 and residues C55/G56 were near F23.

Hydrogen Exchange. As a final test of the sequence-specific assignments and the proposed secondary structure in $APPI_r$, amide exchange rates were qualitatively measured as a function of temperature. The rationale behind this approach was quite simple. Amide exchange rates generally increase with increasing pH and temperature. In order for an NH resonance to be seen by ^1H NMR, it must exchange at a rate slower than 10^3 min^{-1} (Englander et al., 1972; Molday et al., 1972). Therefore, since the $APPI_r$ sample was only slightly acidic (pH 6.5), increases in temperature would cause amide resonances not involved in hydrogen bonding to be *bleached out* by rapid exchange. Important factors which were considered before undertaking this approach were that BPTI was known to tolerate high temperatures and that thorough deuterium studies on BPTI have been documented (Wagner & Wuthrich, 1982b; Roder et al., 1985).

The study was initiated by collecting 1D spectra at 5 °C increments from 25 to 65 °C. The amide and aromatic regions progressively simplified until at 65 °C the spectrum resembled those typical of deuterium-exchanged samples. The spectrum returned to its original appearance when the temperature was lowered back to 25 °C, verifying that the temperature effects were readily reversible (data not shown). This was further investigated by collecting abbreviated DQF-COSY experiments at 45 and 65 °C. At 45 °C, the amide resonances assigned to the first four residues at the N-terminus and the last six residues at the C-terminus, as well as to the arginine side-chain $N^{\epsilon}Hs$, had disappeared. At 65 °C, only 31 of the 64 original NH–$C^{\alpha}H$ cross peaks were observed in the fingerprint region. The assignments of these residual amides and their relative intensities are summarized in Figure 2. Amides involving in strong hydrogen bonds within the β-sheet were clearly visible, as well as several amides in the helical region. These data are completely consistent with the early stages of deuterium exchange in BPTI (Wagner & Wuthrich, 1982b).

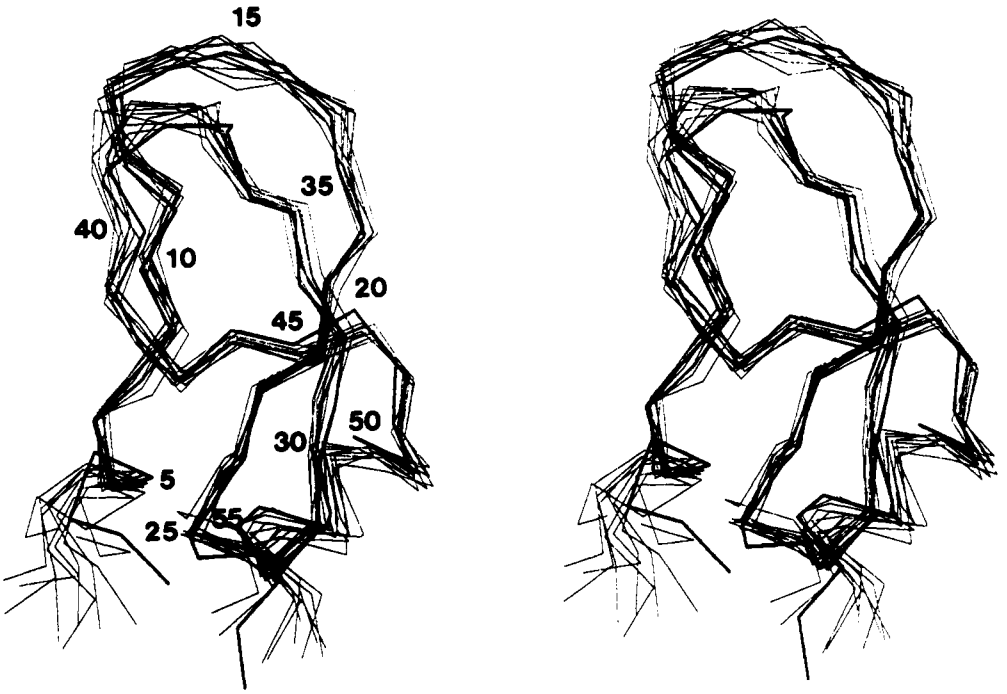

FIGURE 5: Stereoview of the family of 10 NOE-refined structures using simulated annealing (XPLOR 2.1) and starting random initial conformations. The thick line is the X-ray crystal structure of BPTI (4PTI; Wlodawer et al., 1987). Structures were overlaid by a least-squares fit procedure using only the backbone atoms (C, C^{α}, N, O) of residues 3–55. The orientation shown is similar to that used for APPI (Hynes et al., 1990).

Molecular Model. Although initial models of APPI can be generated by homology building with the BPTI X-ray crystal structure (Wlodawer et al., 1987; Toma et al., 1989) for subsequent restrained molecular dynamics refinement (Clore et al., 1985), a de novo approach to the structural determination was pursued. Fifteen independent structures were initially generated with random backbone torsion angles ϕ and ψ. Each was then subjected to simulated annealing refinement using 272 unambiguous interresidue NOE distance restraints (123 short-range, 60 medium-range, and 89 long-range NOEs) and 28 intraresidue NOE restraints, the latter retaining to residues V4, T11, C14, N41, N42, and N44. All of the resulting structures satisfied the experimental data (NOE RMS difference = 0.06 ± 0.01 Å). Ten of these structures (NOE RMS difference = 0.054 ± 0.004 Å) were fitted using a least-squares procedure (Figure 5) and analyzed as to their atomic RMS differences (Figure 6). Representative structures from refinement using covalent S–S bonds, distance restrained S–S pairing, and no S–S pairing information (see Materials and Methods) were retained in this family of 10 structures, showing that the NOE restraints correctly aligned the cysteine pairs. In Figure 5, highlighted as a thick line, the X-ray crystal structure of BPTI (4PTI; Wlodawer et al., 1987) has been overlaid with the family of APPI structures. As is evident in the structures shown in Figure 5, the helical and β-sheet regions of the structure were in good agreement with BPTI. Even the irregularity in the β-sheet, i.e., the matched bend at V25/G28, was found in all the individual structures. The loop regions, however, showed significant deviations. This is apparent in the atomic RMS differences graphically represented in Figure 6 (A, backbone atoms only; B, side-chain atoms). The regions of regular secondary structure show very little variation in the backbone atoms, even the side chains of the buried residues are fairly consistent. However, the loop regions, the surface residue side chains, and the N- and C-termini residues show elevated RMS differences. For the most part, these findings are a direct consequence of the number of NOEs, or lack thereof, used to

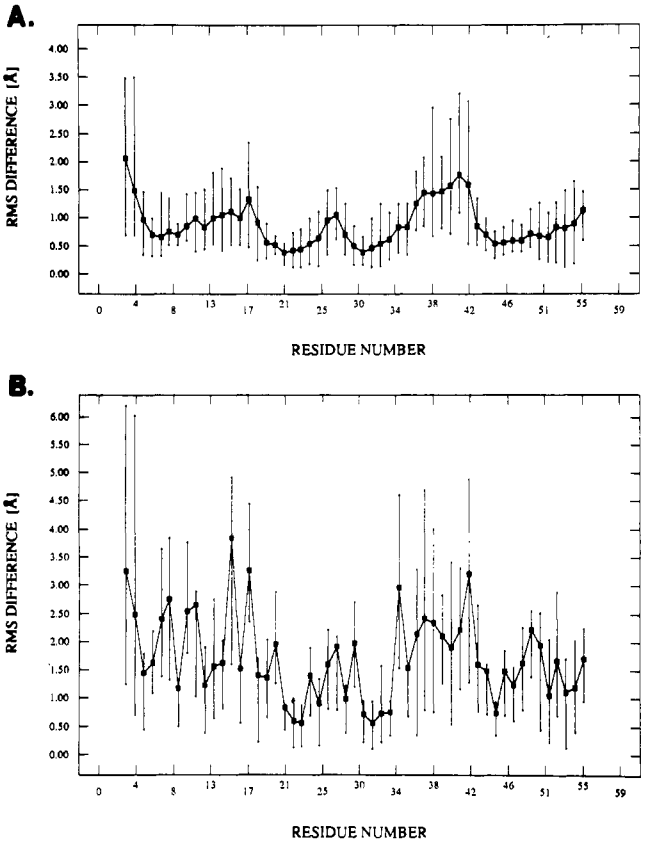

FIGURE 6: (A) RMS differences for backbone atoms (C^{α}, C, N) within the family as a function of residue number. Individual structures were fitted by a least-squares procedure using only the backbone C, C^{α}, and N atoms of residues 3–55. The average RMS difference between the 10 structures is 0.88 Å. The mean RMS difference between individual APPI structures and the X-ray crystal structure of BPTI (4PTI; Wlodawer et al., 1987) is 1.49 ± 0.37 Å. RMS differences for each residue are shown with the average (■) and the minimum and maximum difference within the family of structures. (B) RMS difference for side-chain atoms between individual structures within the family as a function of residue number.

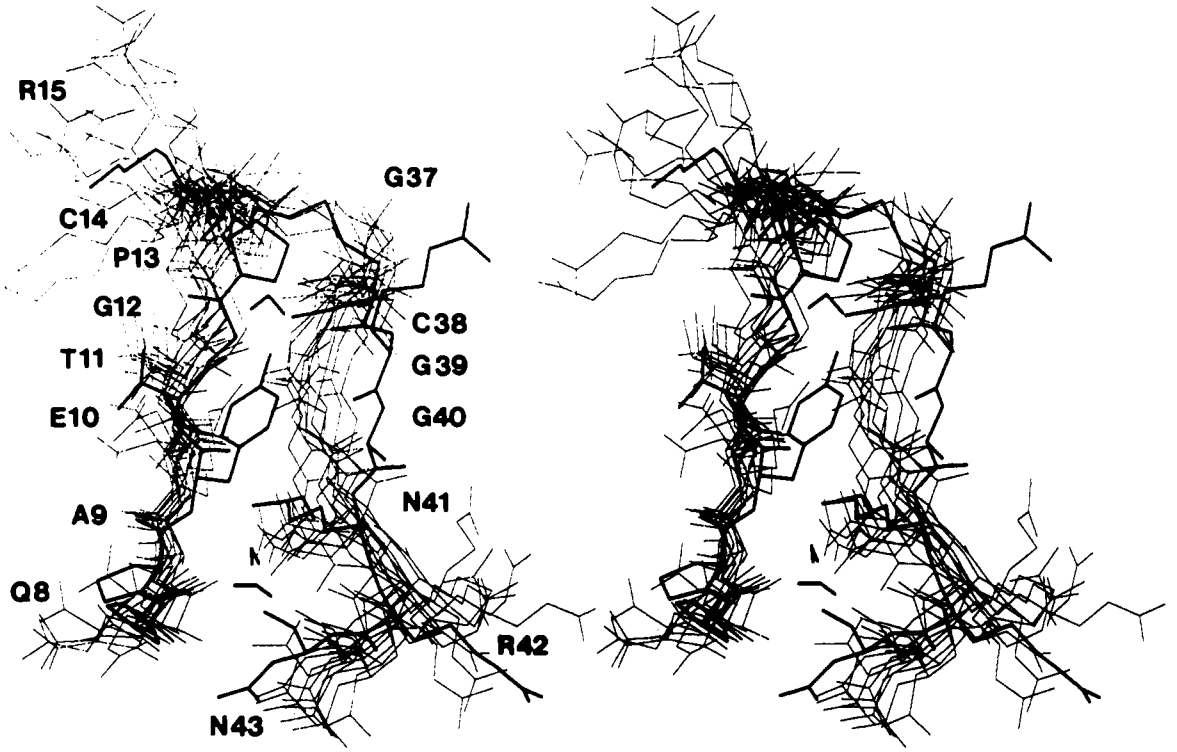

FIGURE 7: Expansion of Figure 5 [model rotated by 90°, similar to Hynes et al. (1990)] showing residues 8–15 and residues 37–43. All atoms with the exception of hydrogen are shown. As in Figure 5, the BPII crystal structure (4PTI) with its four internal water molecules are highlighted as a thick line.

define the exact spatial position of each residue.

Recently, the X-ray crystal structure of APPI refined with data to 1.5-Å resolution was reported by Hynes et al. (1990). This work provided a very precise picture of the packed protein. These authors found excellent agreement between all regions of BPTI and APPI with the exception of the protease-binding loops. Especially noted was the alteration of residues 39–41, as well as the hydrogen-bond formation between the side-chain amide of N41 and the backbone oxygen of Q8 and the backbone nitrogen of N43 with subsequent displacement of one of the four bound water molecules (no. 111 in 4PTI) in BPTI. Our NOE data and, subsequently, the converged structures were consistent with this finding. Figure 7 shows this portion of the protease-binding loops. BPTI and its internally bound water molecules (4PTI; Wlodawer et al., 1987; highlighted as a thick line) have been overlaid with the family of APPI structures. As is also reflected in the RMS differences shown in Figure 6, this region in the NOE-defined APPI structures contained a higher degree of variability. Even so, in all cases, the backbone positions of residues 39–41 were shifted closer to the neighboring strand (Q8–G12) of the protease-binding loops as a consequence of cross-strand NOEs. The backbone atoms of Q8–G12 converged tightly, while the surface side chains of Q8 and E10 appear disordered due to a lack of NOEs to restrain the terminal protons. On the other hand, multiple NOEs observed between E7–N41/N43 and E10–N41 resulted in the side chains of N41 and N43 to be well ordered. In each case, the sidechain $N^{\delta}H_2$ of N41 and the backbone oxygen of Q8 were oriented in positions consistent with hydrogen-bond formation and inconsistent with a location of the bound water molecule. The NMR model did not favor the alignment of the side-chain oxygen of N41 and the backbone nitrogen of N43 due to a critical lack of NOE restraints between the side-chain β-resonances of N41 and N43. These possible cross-peak assignments would have been obscured by degeneracy. Supportive NMR evidence that the structure of the protease-binding loops differed in APPI and BPTI was found in the observed chemical shifts of the N44 side-chain $N^{\delta}H_2$ resonances. Tuchsen and Woodward (1987) showed that anisotropic shielding of protons G37 NH and N44 $N^{\delta}H$[5] by the Y35 ring caused these protons to be unusually shifted upfield. In $APPI_r$, the N44 $N^{\delta}H_2$ were found to resonate at 6.39 and 7.08 ppm. Not shown in Figure 7, the N44 side-chain $N^{\delta}H_2$ was positioned approximately 6 Å, instead of 4 Å (Tuchsen & Woodward, 1987), away from the Y35 ring in the individual APPI structures. This change in position was coupled to the shift of residues G39–N41. Also noted in the crystal structure is the hydrophobic face created by M17 and F34 in the protease-binding domain. Differences between protease inhibitory profiles obtained for APPI and BPTI have been attributed to this unique structural feature in APPI by both the crystallography investigation as well as modeling studies (Toma et al., 1989). This hydrophobic face is also appeared in solution as seen by the numerous side chain–side chain NOEs between M17 and F34.

Unique Spectral Features. Clear spectral evidence existed in the 1D spectrum of $APPI_r$ to indicate that this protein maintained a similar globular structure to BPTI. The most upfield-shifted resonances in BPTI arise from anisotropic shielding of the side-chain resonances of P9. In $APPI_r$, these signals were missing, but a new methyl resonance was found at −0.39 ppm. $APPI_r$ has an alanine in the 9 position. Early analysis of the NOESY spectra revealed that the A9 methyl resonance was shifted upfield by close contacts with the aromatic rings in Y22 and F33, as was also the case in BPTI. This line of reasoning did not, however, account for the occurrence of the E48 $C^{\beta 2}H$ resonance at −0.71 ppm, a 2.7 ppm upfield shift. The replacement residue, A48 in BPTI, was reported at nearly random coil chemical shifts. Moreover, the

[5] Reported chemical shifts: N44 H_Z at 3.4 ppm and N44 H_E at 7.8 ppm.

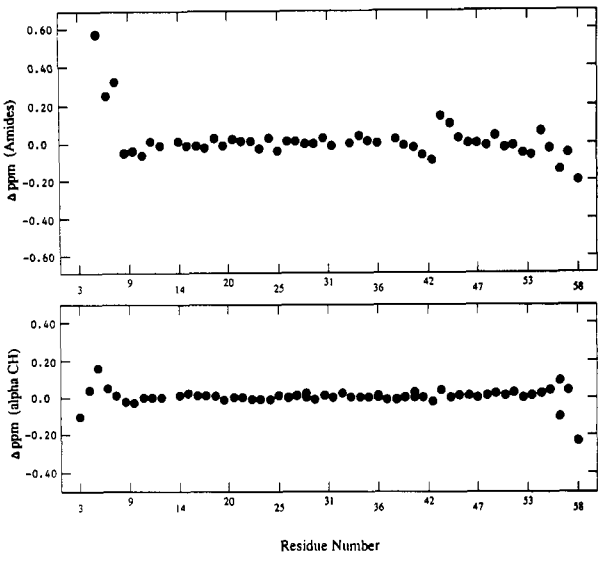

FIGURE 8: Chemical shift difference ($APPI_r$ minus $APPI_s$) of the NH (above) and $C^{\alpha}H$ (below) resonances recorded at pH 6.5, 25 °C.

NOESY spectrum of $APPI_r$ did not indicate that the E48 $C^{\beta 2}H$ resonance made close contact with any aromatic ring. Only one very weak NOE was observed to W21 $C^{\zeta 2}H$ (data not shown). This anomaly in the spectral data was resolved in the converged structures. The E48 $C^{\beta 2}H$ is apparently sandwiched against the center of the tryptophan ring, thereby experiencing the strongest anisotropic field and yet located >3 Å away from the ring protons.

The second anomaly in the spectral data involved the presence of exchange NOEs arising from amide resonances (F45/W21 $N^{\epsilon 1}H$, Y22, and A31) which were involved in the stable hydrogen bonds in β-sheet (data not shown). Further assignment of the spectra revealed that W21 $C^{\alpha}H$, $C^{\delta 1}H$, and $C^{\zeta 3}H$ also showed strong exchange NOEs. This suggested that the exchange NOEs might be a consequence of dynamic reorientation of the tryptophan ring. Extensive long-range NOEs between the tryptophan ring protons and the surrounding residues (D46, T47, E48, C30, A31, and P32) clearly placed the indole ring close to the C30–C51 disulfide bond in the hydrophobic groove created between the β-sheet and long segment of helix. Moreover, in this conformation, the aromatic ring was situated over two hydrogen bonds in the β-sheet involving Y22 NH and A31 NH, causing their chemical shifts to be sensitive to local anisotropic effects. Since many of the exchange NOEs occurred in well-resolved regions of the spectrum, it was estimated that there was a >90% preference for the orientation of the tryptophan ring defined by our distance restraints. Since the second conformation made up such a small percentage of the population, its orientation could not be determined. The correct position of the tryptophan ring was not initially predicted from the BPTI crystal structure or the APPI starting model, which had the tryptophan ring oriented by 180° around the χ_1 from the final converged structures.

Synthetic Protein. 1D ^1H NMR spectra collected on $APPI_s$ indicated clearly that the synthetic protein had folded into a structure quite similar to $APPI_r$. Spectra recorded at low pH (3.5) showed the same pattern of diversified line widths as seen for $APPI_r$. Well-resolved spectra were obtained by raising the pH above the pI. By simply overlaying the upfield regions of the 1D 1H NMR spectra (pH 6.5, 25 °C), the methyl resonances common to both proteins were readily recognized (data not shown). In fact, the methyl resonances, unique to $APPI_r$ (A(−1), M2, and L60), were clearly seen due to their significantly sharper line widths, another indication that these tail-end residues were freely rotating. The downfield regions of the spectra were also superimposable, including the irregularities noted in the aromatic region. A set of 2D NMR spectra was collected under exactly the same conditions as for $APPI_r$. Resulting spectra were nearly identical to those obtained from $APPI_r$, such that the assignment process for both sequential assignments and secondary structure was trivial. Cross peaks arising from residues unique to $APPI_r$ were clearly absent. The relative intensities of the interresidue NOEs used in the simulated annealing refinement of the APPI structure were found to be essentially the same in $APPI_s$, indicating that there was no significant differences in the globular structures of these two proteins. The similarity in the NMR data was best reflected in the difference of the observed chemical shifts for the backbone NH and $C^{\alpha}H$ resonances ($APPI_r$ minus $APPI_s$; Figure 8). As expected, the most significant deviations in chemical shifts occurred in the N- and C-termini, where the proteins were structurally different. Overall, considering how different the preparation was for these two proteins, the agreement between the chemical shifts was excellent.

$APPI_s$, we believe, is the first case in which a fully functional 56 amino acid protease inhibitor has been generated entirely by solid-state synthesis. Partial synthetic approaches have involved chemical modification of the P_1 site in BPTI to determine inhibitory specificity (Beckmann et al., 1988). More recently, in an eloquent study by Staley and Kim (1990) on the folding pathways in BPTI, the helical and sheet subdomains were prepared synthetically and then shown to fold and form the correct disulfide linkages. Taking a completely synthetic approach has produced a homogeneous 56 amino acid protein ($APPI_s$), correctly folded by NMR standards, in less than 2 weeks of concentrated wet chemistry. This opens the opportunity of accelerated preparation of human APPI variants in order to screen for protease inhibitors with improved immune-tolerance characteristics and therapeutically useful for disease indications such as emphysema or posttraumatic shock.

ACKNOWLEDGMENTS

We thank Yale University, Department of Chemistry, for the use of their NMR spectrometer and, in particular, Peter DeMou for his many hours of support and involvement. In addition, we thank Thomas Buckholz of Molecular Therapeutics, Inc., for his efforts in amino acid analysis, peptide synthesis, and protein sequencing and Dr. Walter McMurray, Yale University, for providing FAB-MS.

REFERENCES

Arseniev, A. S., Wider, G., Joubert, F. J., & Wuthrich, K. (1982) *J. Mol. Biol. 159*, 323–351.

Beckmann, J., Mehlich, A., Schroder, W., Wenzel, H. R., & Tschesche, H. (1988) *Eur. J. Biochem. 176*, 675–682.

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

Brunger, A. T. (1988) XPLOR *Manual*, version 2.0; (1990) XPLOR *Manual*, version 2.1, Yale University, New Haven, CT.

Castro, M., Marks, C. B., Nilsson, B., & Anderson, S. (1990) *FEBS Lett. 267*, 207–212.

Clore, G. M., Brunger, A. T., Karplus, M., & Gronenborn, A. M. (1985) *J. Mol. Biol. 186*, 435–455.

Creighton, T. E., & Charles, I. G. (1987) *Cold Spring Harbor Symp. Quant. Biol. 52*, 511–519.

Davis, D. G., & Bax, A. (1985) *J. Am. Chem. Soc. 107*, 2820–2821.

Dovey, H. F., Seubert, P., Ward, P. J., Blacher, R. W., Blacher, M., Bradshaw, R. A., Arici, M., Mobley, W. C., Lieberburg, I., & Sinha, S. (1990) *ICOP: 8th Conference on Proteolysis*, abstract P-30.

Duffaud, G. D., March, P. E., & Inouye, M. (1987) *Methods Enzymol. 153*, 492–507.

Englander, S. W., Downer, N. W., & Teitelbaum, H. (1972) *Annu. Rev. Biochem. 41*, 903–924.

Esch, F. S., Keim, P. S., Beattie, E. C., Blacher, R. W., Culwell, A. R., Otlersdorf, T., McClure, D., & Ward, P. J. (1990) *Science 248*, 1122–1124.

Fields, G. B., & Noble, R. L. (1989) *Int. J. Peptide Res. 35*, 161–214.

Glenner, G. C., & Wong, C. W. (1984) *Biochem. Biophys. Res. Commun. 120*, 885–890.

Griesinger, C., Otting, G., Wuthrich, K., & Ernst, R. R. (1988) *J. Am. Chem. Soc. 110*, 7870–7872.

Heald, S. L., Tilton, R. T., Jr., Bayney, R., Hammond, L., Ramabhadran, T. V., & Tamburini, P. (1990) *XIV Int. Conf. Magn. Reson. Biol. Syst.*, abstract P8-30.

Hynes, T. R., Randal, M., Kennedy, L. A., Eigenbrot, C., & Kossiakoff, A. A. (1990) *Biochemistry 29*, 10018–10022.

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

Johnson, S. A., McNeill, T., Cordell, B., & Finch, C. E. (1990) *Science 248*, 854–857.

Kaumeyer, J. F., Polazzi, J. O., & Kotick, M. P. (1986) *Nucleic Acids Res. 14*, 7839–7850.

Keller, R. M., Baumann, R., Hunziker-Kwik, E.-H., Joubert, F. J., & Wuthrich, K. (1983) *J. Mol. Biol. 163*, 623–646.

Kido, H., Fukutomi, A., Schilling, J., Wang, Y., Cordell, B., & Katunuma, N. (1990) *Biochem. Biophys. Res. Commun. 167*, 716–721.

Kitaguchi, N., Takahashi, Y., Tokushima, Y., Shiojiri, S., & Ito, H. (1988) *Nature 331*, 530–532.

Kitaguchi, N., Takahashi, Y., Oishi, K., Shiojiri, S., Tokushima, Y., Utsunomiya, T., & Ito, H. (1990) *Biochim. Biophys. Acta 1038*, 105–113.

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

Lucchesi, K., & Moczydlowski, E. (1990) *Neuron 2*, 141–148.

Macura, C., Huang, Y., Suter, D., & Ernst, R. R. (1981) *J. Magn. Reson. 43*, 259–282.

Marion, D., & Wuthrich, K. (1983) *Biochem. Biophys. Res. Commun. 113*, 967–974.

Molday, R. S., Englander, S. W., & Kallen, R. G. (1972) *Biochemistry 11*, 150–158.

Muller-Hill, B., & Beyreuther, K. (1989) *Annu. Rev. Biochem. 58*, 287–307.

Oltersdorf, T., Fritz, L. C., Schenk, D. B., Leiberburg, I., Johnson-Wood, K. L., Beattie, E. C., Ward, P. J., Blacher, R. W., Dovey, H. F., & Sinha, S. (1989) *Nature 341*, 144–147.

Plateau, P., & Gueron, M. (1982) *J. Am. Chem. Soc. 104*, 7310–7311.

Ponte, P., Gonzalez-DeWhitt, P., Schilling, J., Miller, J., Hsu, D., Greenberg, B., Davis, K., Wallace, W., Leiberburg, I., Fuller, F., & Cordell, B. (1988) *Nature 331*, 525–527.

Rance, M. (1987) *J. Magn. Reson. 74*, 557–564.

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

Roder, H., Wagner, G., & Wuthrich, K. (1985) *Biochemistry 24*, 7407–7411.

Shaka, A. J., & Freeman, R. (1983) *J. Magn. Reson. 51*, 169–173.

Sisodia, S. S., Koo, E. H., Beyreuther, K., Unterbeck, A., & Price, D. L. (1990) *Science 248*, 492–495.

Staley, J. P., & Kim, P. S. (1990) *Nature 344*, 685–688.

Tanaka, S., Nakamura, S., Ueda, K., Kameyama, M., Shiojiri, S., Takahashi, Y., Kitaguchi, N., & Ito, H. (1988) *Biochem. Biophys. Res. Commun. 157*, 472–479.

Tanzi, R. E., McClatchey, A. I., Lamperti, E. D., Villa-Komaroff, L., Gusella, J. F., & Neve, R. L. (1988) *Nature 331*, 528–530.

Toma, K., Kitaguchi, N., & Ito, H. (1989) *J. Mol. Graphics 7*, 202–218.

Tuchsen, E., & Woodward, C. (1987) *Biochemistry 26*, 1918–1925.

Van Nostrand, W. E., Wagner, S. L., Suzuki, M., Choi, B. H., Farrow, J. S., Geddes, J. W., Cotman, C. W., & Cunningham, D. D. (1989) *Nature 341*, 546–549.

Van Nostrand, W. E., Wagner, S. L., Farrow, J. S., & Cunningham, D. D. (1990) *J. Biol. Chem. 265*, 9591–9594.

Wagner, G., & Wuthrich, K. (1982a) *J. Mol. Biol. 155*, 347–366.

Wagner, G., & Wuthrich, K. (1982b) *J. Mol. Biol. 160*, 343–361.

Wagner, G., Kumar, A., & Wuthrich, K. (1981) *Eur. J. Biochem. 114*, 375–384.

Wlodawer, A., Deisenhofer, J., & Huber, R. (1987) *J. Mol. Biol. 193*, 145–156.

Wun, T.-C., Kretzmer, K. K., Girard, T. J., Miletich, J. P., & Broze, G. J., Jr. (1988) *J. Biol. Chem. 263*, 6001–6004.

Wuthrich, K. (1986) *NMR of Proteins and Nucleic Acids*, John Wiley & Sons, New York.

Zarbock, J., Oschkinat, H., Hannappel, E., Kalbacher, H., Voelter, W., & Holak, T. A. (1990) *Biochemistry 29*, 7814–7821.