Articles

NMR Solution Structure of ATT_p, an *Arabidopsis thaliana* Trypsin Inhibitor[†,‡]

Qin Zhao,[§] Young Kee Chae,[||] and John L. Markley*[,§]

Department of Biochemistry, University of Wisconsin−Madison, 433 Babcock Drive, Madison, Wisconsin 53706, and Department of Applied Chemistry and Molecular Biology, Sejong University, 98 Kunja-Dong, Kwangjin-Gu, Seoul, Korea 143−747

Received February 19, 2002; Revised Manuscript Received August 14, 2002

ABSTRACT: The three-dimensional structure of the precursor form of the *Arabidopsis thaliana* trypsin inhibitor (ATT_p, GenBank entry Z46816), a 68-residue (∼7.5 kDa) rapeseed class proteinase inhibitor, has been determined in solution at pH 5.0 and 25 °C by multinuclear magnetic resonance spectroscopy. The protein contains one α-helix and two strands of antiparallel β-sheet, with a type IV β-turn connecting the two strands. The α-helix and the inhibitory loop are connected to the β-sheet through three disulfide bridges; a fourth disulfide bridge connects the N- and C-termini. The overall structural topology of ATT_p is similar to those of the sweet tasting protein brazzein (rmsd of 3.0 Å) and the antifungal protein Rs-Afp1 [a knottin protein from radish (*Raphanus sativus*), rmsd of 2.7 Å]. The precursor segment in ATT_p is disordered, as visualized by the final 20-conformer ensemble and as confirmed by ^{15}N heteronuclear NOE analysis. The overall fold of ATT_p is distinct from those of other classes of serine proteinase inhibitors except in the inhibitor loop; therefore, it represents a new inhibitor fold.

Genomic DNA sequencing of *Arabidopsis thaliana*, a plant from the mustard family (*1*−*3*), revealed a group of isogenes (ATT isogenes) that belong (Figure 1) to the recently described rapeseed class of serine proteinase inhibitors, which includes MTI-2[1] purified from white mustard (*Sinapis alba* L.) seed (*4*) and RTI purified from oil rape (*Brassica napus* var. oieifera) seed (*5*). MTI-2 and RTI each inhibit trypsin and, at ∼1000-fold lower affinity, chymotrypsin. The physiological functions of this new class of serine proteinase inhibitors have not been fully explored. Because both RTI-III and MTI-2 and related proteins are found in plant seeds, Ryan speculated that they may regulate endogenous proteinases during seed dormancy and serve as a storage source of protein (*6*). The function of MTI-2 under certain circumstances has been attributed to plant defense (*7*, *8*). The ATT gene sequences encode N-terminal protein extensions that presumably are removed in the mature inhibitors as with MTI-2 (*9*). The precursor sequences of different ATT genes have slight differences; these may modulate protein maturation through the endoplasmic reticulum and lead to differential targeting within the organism (*10*).

The rapeseed class is one of a few serine proteinase inhibitor families for which a representative tertiary structure has not been reported (*11*). The reactive site loop in this class of inhibitors contains a unique P_3-P_3' sequence (-Cys-Ala-Pro-Arg-Ile-Phe/Tyr-Pro-) and cysteine as the P_6' residue (Figure 2). The primary sequence of this inhibitor contains eight conserved cysteine residues that form a disulfide arrangement similar to that seen in brazzein and a subgroup of knottin proteins (Q. Zhao, Y. K. Chae, M. M. Vestling, and J. L. Markley, manuscript in preparation). It has been difficult to obtain homogeneous samples of rapeseed inhibitors from natural sources because the plants produce multiple forms of the protein (*12*, *13*). Efforts have been made toward recombinant production of MTI-2, the rapeseed-type inhibitor from white mustard (*S. alba* L.), either by itself in *Pichia pastoris* (*10*) or as a fusion protein in *Escherichia coli* (*14*, *15*). However, no complete structural description of any rapeseed inhibitor has been reported. As described here, we have

[†] This work was supported by NIH Grants P50 GM 64598 and P41 RR 02301.

[‡] The coordinates and NMR constraints for the ATT_p structure have been deposited in the Protein Data Bank (entry 1JXC); the NMR chemical shifts have been deposited in the BioMagResBank (entry 5056).

* To whom correspondence should be addressed: Phone: +1 (608) 263−9349. Fax: +1 (608) 262−3759. E-mail: markley@nmrfam.wisc.edu.

[§] University of Wisconsin−Madison.

[||] Sejong University.

[1] Abbreviations: ATT, *A. thaliana* gene encoding a putative protein proteinase inhibitor (GenBank entry Z46816); ATT_p, recombinant protein produced from *E. coli* corresponding to the putative precursor gene product of ATT; ATT_m, mature form of the *Arabidopsis* trypsin inhibitor (ATT_m lacks the N-terminal hexapeptide CPEIEA of ATT_p); ATT_{m+4}, mature form of the *Arabidopsis* trypsin inhibitor containing the N-terminal tetrapeptide extension YVEF (the same gene was used in the production of ATT_m and ATT_{m+4}; under the conditions used to produce ATT_{m+4}, the tetrapeptide was not cleaved off); BLPNA, *N*-α-benzoyl-L-leucine *p*-nitroanilide; BNPNA, *N*-α-benzoyl-L-arginine *p*-nitroanilide; DTT, dithiothreitol; MTI-2, trypsin inhibitor from white mustard (*Sinapis alba* L.); MUGB, 4-methylumbelliferyl *p*-guanidinobenzoate; MWCO, molecular weight cutoff; RTI, family of trypsin inhibitors from oil-rape (*B. napus* var. oieifera); TAME, *N*-α-*p*-tosyl-L-arginine methyl ester hydrochloride; TFA, trifluoroacetic acid; GdmCl, guanidinium chloride; RP, reverse phase; ITI, trypsin/chymotrypsin isoinhibitors from *A. thaliana*; Rs-Afp1, *Raphanus sativus* antifungal protein 1; BBI, Bowman-Birk serine proteinase inhibitor; ATI, *Ascaris* trypsin inhibitor.

10.1021/bi025702a CCC: $22.00
Published on Web 09/19/2002

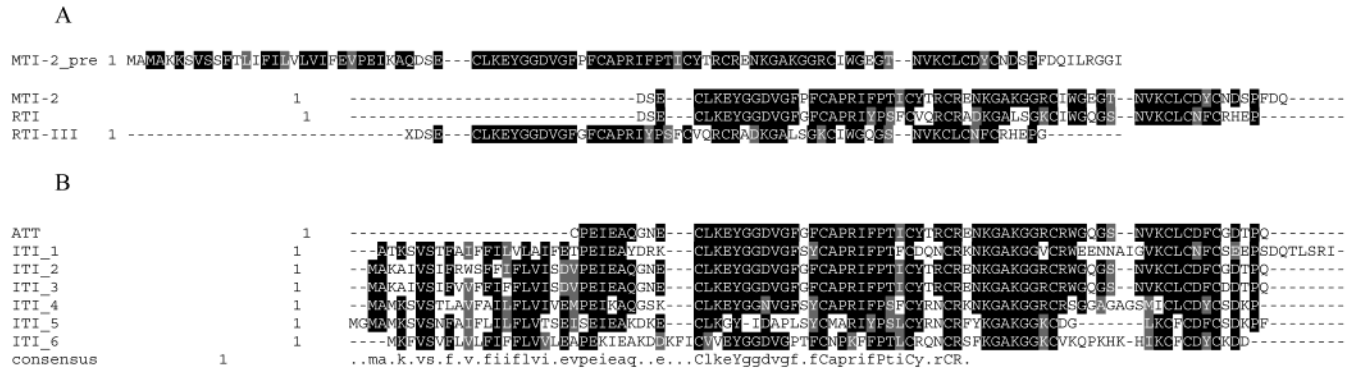

FIGURE 1: Sequence alignment of members of the rapeseed class of serine proteinase inhibitors: (A) MTI-2 and RTI isoinhibitors and (B) isogenes from *A. thaliana*. Dark shades indicate identical residues; light shades indicate similar residues. In the consensus sequence, uppercase letters indicate identical residues and lowercase letters indicate similar residues. MTI-2_pre, SWISS-PROT entry P26780; MTI-2 (*4*); RTI, SWISS-PROT entry P80301; RTI-III (*5*), X is pyroglutamate; ATT, GenBank entry Z46816; ITI_1, SWISS-PROT entry Q42330; ITI_2, SWISS-PROT entry Q42328; ITI_3, SWISS-PROT entry O22865; ITI_4, SWISS-PROT entry O22866; ITI_5, SWISS-PROT entry O22867; and ITI_6, SWISS-PROT entry O22869. This figure was generated with the programs CLUSTLAW (http://www.clustalw.genome.ad.jp/) and BOXSHADE (http://www.ch.embnet.org/software/BOX_form.html).

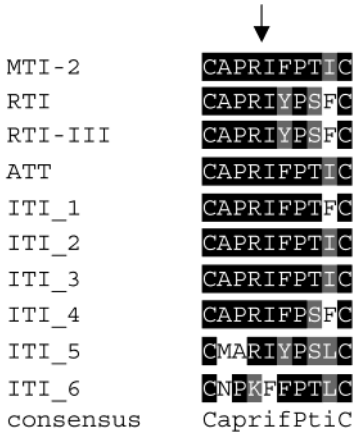

FIGURE 2: Reactive site loop residues (P_4-P_6') of members of the rapeseed family of serine protease inhibitors showing the level of conservation: black, fully conserved; and gray and white, functionally conserved residues. MTI-2_pre, SWISS-PROT entry P26780; MTI-2 (*4*); RTI, SWISS-PROT entry P80301; RTI-III (*5*); ATT, GenBank entry Z46816; ITI_1, SWISS-PROT entry Q42330; ITI_2, SWISS-PROT entry Q42328; ITI_3, SWISS-PROT entry O22865; ITI_4, SWISS-PROT entry O22866; ITI_5, SWISS-PROT entry O22867; and ITI_6, SWISS-PROT entry O22869. This figure was generated with the programs CLUSTALW (http://www.clustalw-.genome.ad.jp/) and BOXSHADE (http://www.ch.embnet.org/software/BOX_form.html). The arrow denotes the reactive site.

produced the putative precursor form of the inhibitor (ATT_p) heterologously in *E. coli*. Truncated forms of the inhibitor (ATT_m and ATT_{m+4}) lacking the precursor sequence have been produced in *P. pastoris*. We have carried out chemical shift assignments for ATT_p (^1H, ^{13}C, and ^{15}N), ATT_{m+4} (^1H and ^{15}N), and ATT_m (^1H, ^{15}N, and ^{13}C) and have determined that the three forms of the protein have very similar three-dimensional structures (Q. Zhao, Y. K. Chae, M. M. Vestling, and J. L. Markley, manuscript in preparation).

We report here the solution structure of ATT_p, the first three-dimensional structure of a member of the rapeseed family of serine proteinase inhibitors. The overall fold of ATT_p is similar to those of CS$\alpha\beta$ proteins and a subgroup of knottins. The conformation of the reactive site loop, although slightly noncanonical, is similar to the novel form observed for the high-pH conformer of the *Ascaris* class of serine proteinase inhibitors.

EXPERIMENTAL PROCEDURES

Materials. Restriction enzymes were purchased from Promega (Madison, WI) and New England Biolabs (Beverly, MA). *Taq* DNA polymerase and T4 DNA ligase were from Promega. *Pfu* DNA polymerase was from Stratagene (La Jolla, CA). Porcine pancreatic trypsin type IX, proteinase substrates *N*-α-benzoyl-L-arginine *p*-nitroanilide (BNPNA) and 4-methylumbelliferyl *p*-guanidinobenzoate (MUGB), citric acid, K_2HPO_4, and NaH_2PO_4 were from Sigma Chemicals (St. Louis, MO). HPLC grade acetonitrile and methanol were from Fisher (Pittsburgh, PA), and HPLC-grade trifluoroacetic acid (TFA) was from Pierce (Rockford, IL). *E. coli* strain XL1Blue was used for DNA transformation and for routine plasmid storage. $^{15}NH_4Cl$, [^{13}C]glucose, and [^{13}C]methanol were from Isotech, Inc. (Miamisburg, OH).

DNA Synthesis and Sequencing. Oligonucleotides used as primers were synthesized at the University of Wisconsin Biotechnology Center. DNA sequencing was carried out at the Biotechnology Center with an ABI Prism automated DNA sequencer (Applied Biosystems, Foster City, CA).

Plasmid Constructions and Site-Directed Mutagenesis for ATT_p Production in E. coli. The original *A. thaliana* genomic DNA clone, BAC vector T1o024, containing ITI3 (GenBank entry AJ249958), an isogene of ATT, was used as the template in PCR amplification reactions. This gene was cloned into an expression system that led to the production of ATT_p as a fusion with staphylococcal nuclease with an engineered cyanogen bromide cleavage site.

Protein Production in E. coli. ATT_p was produced from *E. coli* cultures and labeled with ^{15}N and/or ^{13}C for NMR investigations. Full details will be published elsewhere.

MALDI MS Analysis of ATT. The molecular mass of purified ATT_p was determined by matrix-assisted laser desorption/ionization, time-of-flight (MALDI-TOF) mass spectrometry at the University of Wisconsin Biotechnology Center.

Inhibitory Activity of ATT_p. The inhibitory activity of ATT_p was determined against porcine pancreatic trypsin IX using

the chromogenic substrate *N*-α-benzoyl-L-arginine *p*-nitroanilide (BNPNA) (*16*). Trypsin activity was determined with the fluorogenic substrate 4-methylumbelliferyl *p*-guanidinobenzoate hydrochloride (MUGB) (*17*). The K_i of ATT_p toward porcine trypsin was calculated as described previously (*9*). The chymotrypsin activity test was performed on TLCK-treated α-chymotrypsin as previously described, with *N*-α-benzoyl-L-leucine *p*-nitroanilide (BLPNA) as the substrate (*16*). The inhibitory activities of ATT_p toward thrombin, plasmid, and papain were also tested. Full descriptions of these procedures are in the Supporting Information.

NMR Spectroscopy. NMR samples contained ~1 mM ATT_p, 50 mM perdeuterated sodium acetate at pH 5.0, and 0.1 mM DSS. NMR data were collected at 25 °C. The parameters shown in Supporting Information (Table S1) were used to record two- and three-dimensional (2D and 3D, respectively) NMR data sets for resonance assignments on Bruker DMX spectrometers equipped with triple-resonance ^1H, ^{13}C, ^{15}N probes and triple-axis pulse field gradients. Quadrature detection in the indirectly detected dimensions was achieved by either the States−TPPI (*19*) or echo/anti-echo (*20*) method. Gradient pulses, as a combination of *x*- and *z*-gradients at the "magic angle" (*21*), were used for coherence selection in experiments with sensitivity enhancement (SE) to maximize water suppression.

Two-dimensional NOESY spectra with a mixing time of 100 ms were acquired for unlabeled ATT_p samples, one dissolved in D_2O and one in H_2O. Two data sets were acquired for [U-^{15}N]ATT_p; a 3D $^1H-^{15}N$ NOESY-HSQC (*22*) data set obtained with a mixing time of 100 ms was used as a source of distance constraints, and an HNHA (*23*, *24*) data set was used in determining $^3J_{HNH\alpha}$ coupling constants, which were used as angular constraints. Triple-resonance ^{13}C-edited NOESY-HSQC (*25*) data sets with mixing times of 80, 100, and 150 ms were acquired for [U-^{13}C,^{15}N]ATT_p and were used to obtain additional stereospecific assignments and distance constraints. Pulsed field gradient pulses were applied along the magic angle to suppress the water resonance (*21*).

NMR data were processed with FELIX95 or FELIX98 software (Accelrys Inc., San Diego, CA). With three-dimensional NMR data sets, the indirect dimension with the least amount of digitization was extended with linear prediction equivalent to the original data size. Time domain convolution was used to remove the residual water signal during processing. A 90°-shifted square sine-bell window function was first applied in each dimension. Then the dimension was zero-filled to the final matrix size, Fourier transformed, and phase corrected. The initial values for incremented delays in the multidimensional experiments were set in a manner that allowed predictable phasing in each dimension and minimized rolling and offset of the baseline (*26*). All ^1H dimensions are referenced to internal DSS, and ^{13}C and ^{15}N dimensions are referenced indirectly to DSS (*27*). Processed NMR data were analyzed using the SPARKY [http://www.cgl.ucsf.edu/home/Sparky (*28*)] and XEASY (*29*) software packages.

Choice of a Temperature for Data Collection. Because fewer cross-peaks than predicted were resolved in the 2D homonuclear spectra of ATT_p, we suspected initially that resonances from aromatic side chains overlapped. However, because no additional ^1H signals from aromatic rings were obtained from analysis of the 3D ^{13}C TOCSY CT-HSQC and NOESY CT-HSQC data sets, we hypothesized that that the signal loss results from line broadening caused by ring flipping of aromatic rings at intermediate rates on the NMR chemical shift time scale. This assumption was confirmed through examination of one-dimensional ^1H NMR spectra of an ATT sample in 99% D_2O collected at different temperatures on a Bruker DMX 500 MHz spectrometer. The best resolution was obtained at 298 K, and signals were seen to broaden at higher or lower temperatures (spectra not shown).

Sequence Specific Backbone Assignments of ATT_p. HNCA and HN(CO)CA spectra of ATT_p were first analyzed to obtain sequential connectivities among ^1H, ^{15}N, and ^{13}C resonances (see Figure S1 of the Supporting Information). Data from the HNCO experiment were used in assigning backbone carbonyl carbon signals. The connectivities were verified through analysis of the HNCACB spectrum. The full BMRB statistics for chemical shifts of diamagnetic proteins (http://www.bmrb.wisc.edu/ref_info/statful.html) were used in distinguishing individual amino acid types, and the following rules were applied in the analysis. (i) No glycine spin system contains a $^{13}C^{\beta}$ peak. (ii) In serine and threonine spin systems, generally $\delta(^{13}C^{\beta}) > \delta(^{13}C^{\alpha})$. (iii) Proline spin systems are characterized by connectivity breaks because they lack an H^N atom.

Side Chain Assignments. Side chain ^{13}C and ^1H resonances of ATT_p were assigned by analyzing results from the following experiments: HNCACB, 3D C(CO)NH-SE, H(C-CO)NH-SE, HCCH-COSY, and HCCH-TOCSY. The combination of C(CO)NH-SE and HCCH-COSY proved to be very efficient for the unambiguous determination of assignments for side chain $^1H-^{13}C$ units. 3D ^{15}N-separated TOCSY HSQC and 3D HNHA data sets were collected from a sample of [U-^{15}N]ATT_p; these experiments provided complementary assignment information and precise proton chemical shift values.

Asn and Gln Side Chain Assignments. HNCO and HNCACB data sets collected for ATT_p with side chain-optimized delay values were used to confirm assignments to the side chains of the three asparagine and three glutamine residues.

Stereospecific Assignments. Stereospecific assignments of prochiral protons were determined in part by methods developed by Wüthrich and co-workers (*30*, *31*). The procedure is summarized as follows. (i) For proline, the amplitudes of the intraresidue NOE cross-peaks between the α-proton and the two β-protons are compared. Since H^{α} is always closer to $H^{\beta 3}$ than to $H^{\beta 2}$, the identification is straightforward. (ii) For Asn and Gln, one of the two amide N−H protons is always closer to the peripheral methylene group than the other; therefore, stereospecific assignments can be deduced from a comparison of NOE intensities between the H^{δ}s and $H^{\beta 2/\beta 3}$ signals (for Asn) or H^{ϵ}s and $H^{\gamma 2/\gamma 3}$ signals (for Gln). (iii) The prevalent χ^1 rotamer and the stereospecific assignments of the $H^{\beta 2/\beta 3}$ protons, when separately resolved, can be determined by comparison of ^3J coupling values and NOE intensities.

Relaxation Data. $^1H-^{15}N$ steady state NOE (htNOE) values were obtained by recording spectra with and without ^1H presaturation with a duration of 3 s (*32*). The SPARKY (*28*) software package was used in determining NOE cross-

peak intensities. The heteronuclear NOE was calculated from the ratio of resonance intensities in spectra recorded with and without saturation

$$\text{htNOE} = I_{sat}/I_{unsat} \quad (1)$$

where I_{sat} and I_{unsat} represent the measured intensities of a resonance in the presence and absence, respectively, of proton saturation.

Hydrogen Exchange Experiments. A solution of [U-^{15}N]-ATT_p was lyophilized to dryness and dissolved in 500 μL of H_2O. It was lyophilized again and dissolved in 500 μL of 100% D_2O. A series of $^1H-^{15}N$ HSQC spectra were collected at 25 °C at different time intervals following the addition of D_2O solvent. Residues whose H^N-N cross-peaks remained observable after >7 h were classified as having slowly exchanging backbone amide protons.

Experimental Restraints and Structure Calculations. The program SPARKY (*28*) was used in the manual assignment of the 3D $^1H-^{15}N$ NOESY-HSQC spectra and to measure peak intensities. These peak intensities were converted to distance constraints in XPLOR (*33*) format with a UNIX shell script created by Z. Zolnai in NMRFAM, and corrections for pseudoatoms were added. The above constraints were used subsequently as input to DYANA (*34*) and converted to DYANA upper limits for later rounds of calculations. $^3J_{HNH\alpha}$ coupling constants calculated from the cross-peak to diagonal-peak intensity ratio in the HNHA experiment were used as follows to constrain backbone dihedral angles: $^3J_{HNH\alpha} < 5.5$ Hz, $\phi = -57 \pm 20°$; $^3J_{HNH\alpha} > 7.7$ Hz, $\phi = -139 \pm 30°$. Hydrogen bond restraints were introduced for backbone amide protons only in cases where the amide proton exchanged slowly, the $^3J_{HNH\alpha}$ coupling constant was small, and medium-range NOEs were consistent with the presence of an α-helix. No disulfide bond restraints were imposed in the initial stages of the structure calculations.

The program DYANA1.5, which employs simulated annealing combined with molecular dynamics in torsion angle space (*33*), was used to generate structural models. An iterative process of calculation was employed in which the results of successive rounds of structure calculations were used to make further NOE assignments. During the initial rounds of structure calculations, 10 000 simulated annealing steps were used to generate each structure; ~100 such structures were determined, and among these, the 25 with the lowest target function values were retained for analysis of hydrogen bonds and long-range NOEs.

After the initial structures were obtained, we used the ASSIGN function of DYANA to automate NOE assignments from three data sets: the 2D homonuclear NOESY of the sample in H_2O, the 2D homonuclear NOESY of the sample in D_2O, and the 3D ^{15}N-edited NOESY. Distances derived from the additional NOE peaks were calibrated separately for each data set. The newly generated distance constraints were added into further rounds of structure calculations. All violated constraints were investigated, and consistently violated constraints (never more than 5% of the constraints) were removed. The automated NOE assignment procedure was performed until no new constraints could be generated and no new NOE peaks could be assigned. These structures were used to identify additional hydrogen bonds. The arrangement of disulfide bonds consistent with these structures was in full agreement with the pattern determined biochemically (Q. Zhao, Y. K. Chae, M. M. Vestling, and J. L. Markley, manuscript in preparation). The input for subsequent structure determinations included additional constraints for disulfide bonds. The TALOS program (*35*) was used to generate additional backbone dihedral angle constraints in regions of secondary structure. The NOESY peaks were reinvestigated, and iterative rounds of structure calculations generated the final set of distance constraints.

In the following cycles of DYANA calculations, 50 conformers were obtained that satisfied the input restraints. Of these, the 10 conformers with the lowest overall energy were selected for analysis by the programs DYANA and PROCHECK (*36*), and the programs DYANA1.5 and MOLMOL (*37*) were used to determine their rmsds. Again, all violated constraints were investigated, and consistently violated constraints (less than 2% of total constraints used in the calculation) were removed. This process was repeated until all the distance and angle restraints produced a set of structures that had no NOE distance violations of >0.3 Å and no dihedral angle violations of >5°. Once the final set of restraints had been obtained, a new family of 50 structures was generated using DYANA1.5; the 20 conformers with the lowest penalty functions were retained for further structural analysis. The mean structure from the final ensemble of conformers was regularized by a 1000-step minimization in DYANA. MOLMOL and PROMOTIF (*38*) were used in structure visualization and analysis. The regularized mean structure and NMR restraints have been deposited with the Protein Data Bank (entry 1JXC).

RESULTS

ATT_p Produced from the E. coli Expression System. The yield of ATT_p after the final reverse phase (RP) HPLC purification step (at which the protein eluted as one sharp peak) was ~10 mg/L of cell culture. The inhibitory constant (K_i) of ATT against porcine trypsin was 1.15 ± 0.23 nM. ATT_p was found to inhibit chymotrypsin and thrombin but not plasmin or papain (details are provided in the Supporting Information).

The molecular mass of the ATT_p product from MALDI-TOF MS was 7452 ± 2 Da (within 0.01% of the calculated value of 7443 Da for the monomer); this clearly indicates that the protein is a monomer. The molecular mass determined for [^{13}C,^{15}N]ATTp was 7900 ± 10, which indicates a level of enrichment of more than 99% for each isotope.

Extent of the Assignments for ATT_p. It was possible to resolve signals and assign chemical shifts to ~80% of all the ^1H, ^{13}C, and ^{15}N nuclei of ATT_p. Stereospecific assignments of H^β protons for 31 of 47 residues and the side chain amides for two of three Asn residues and two of three Gln residues were obtained through analysis of ^{15}N-NOESY, ^{15}N-TOCSY, and ^{13}C-NOESY spectra. Partial assignments for the side chain resonances of seven of seven aromatic residues (two Tyr residues, four Phe residues, and one Trp residue) were achieved through analysis of 2D ^{1}H-NOESY and TOCSY, CT ^{13}C-HSQC, and 3D NOESY CT-HSQC and TOCSY CT-HSQC data sets. Regions of the molecule that remained unassigned are primarily those affected by ex-

change broadening. These include the N-terminal residue Cys^{1}, Pro^{26}, the side chains of several aromatic amino acids (Tyr^{15}, Phe^{21}, Phe^{23}, Tyr^{34}, Trp^{50}, and Phe^{62}), and the C^{ζ}, $N^{\eta 1}$, $N^{\eta 2}$, $H^{\eta 11}$, $H^{\eta 12}$, $H^{\eta 21}$, and $H^{\eta 22}$ atoms of the five arginine residues. NMR assignments have been deposited at BMRB (entry 5056).

Oxidation State of the Cysteine Residues. An excellent correlation was reported recently between cysteinyl $^{13}C^{\alpha}$ and $^{13}C^{\beta}$ chemical shifts and the oxidation state of the cysteine side chain (*39*). We have applied the two methods described by these authors to investigate the redox states of the eight cysteines of ATT_p that are conserved in all rapeseed inhibitors. Signals were not resolved for the nonconserved cysteine of ATT (Cys^{1}).

The first method is based on cysteine $^{13}C^{\beta}$ chemical shifts. Cysteine residues in which $\delta(^{13}C^{\beta}) > 35$ ppm can be classified as oxidized, while cysteine residues in which $\delta(^{13}C^{\beta}) < 32$ ppm can be classified as reduced. If only one $\delta(^{13}C^{\beta})$ lies in the gray area between 32 and 35 ppm, its oxidation state can be determined by considering the redox states of the other cysteine residues (*39*). All but one of the eight conserved cysteines in ATT_p has a $\delta(^{13}C^{\beta})$ of >35 ppm; the remaining conserved cysteine has a $\delta(^{13}C^{\beta})$ in the gray area [Cys^{60} $\delta(^{13}C^{\beta}) = 34.3$ ppm]. The second method for determining the oxidation state of cysteines utilizes the Cys $\delta(^{13}C^{\beta})$ and $\delta(^{13}C^{\alpha})$ chemical shifts as well as the local secondary structure. In a plot of $\delta(^{13}C^{\beta})$ versus $\delta(^{13}C^{\alpha})$ for each cysteine, oxidized and reduced cysteines fall into different regions, which are dependent on the secondary structure in which they are located. The analysis of ATT_p shows all conserved Cys residues in the oxidized region, with the exception of Cys^{60}, which is located in an indeterminate area but clearly not in the region of a reduced cysteine (see Figure S2 of the Supporting Information). Thus, we conclude that all eight conserved cysteines participate in disulfide bridges. The following disulfide bridges of ATT_{m+4}, a truncated form of ATT, were subsequently identified by chemical methods: C_1-C_8, C_2-C_5, C_3-C_6, and C_4-C_7 [where the numbers indicate the order of appearance of the conserved cysteine in the sequence (Q. Zhao, Y. K. Chae, M. M. Vestling, and J. L. Markley, manuscript in preparation)]. These disulfide bond constraints were applied in later rounds of structure calculation.

Secondary Structure. Secondary structural elements for ATT_p were deduced from the assigned chemical shifts and three-bond *J* couplings. The consensus chemical shift index (CSI) (*40*) was determined from the combined experimental $^{1}H^{\alpha}$, $^{13}C^{\alpha}$, $^{13}C^{\beta}$, and $^{13}C'$ chemical shifts (Figure 3). The following rules were applied in deriving secondary structure from *J* coupling constants. (i) Two or more consecutive $^{3}J_{HNH\alpha}$ values of >7.7 Hz imply β-strand, and (ii) more than three consecutive $^{3}J_{HNH\alpha}$ values of <5.0 Hz imply α-helix. The combined results (see Table S2 of the Supporting Information) indicate that ATT contains three β-strands, one short helical element, and one α-helix. Alignment of the secondary structural elements of ATT with those of other sequence-related proteins reveals an excellent match to brazzein and the group of scorpion toxins with known structure (Figure 4). A similar alignment of the secondary structural elements of ATT with those of snake venom toxins and epidermal growth factors domains with known structure (Figure 4) shows much less agreement. From this comparison, it appears to be clear that ATT is structurally distinct from snake venom toxins and epidermal growth factors.

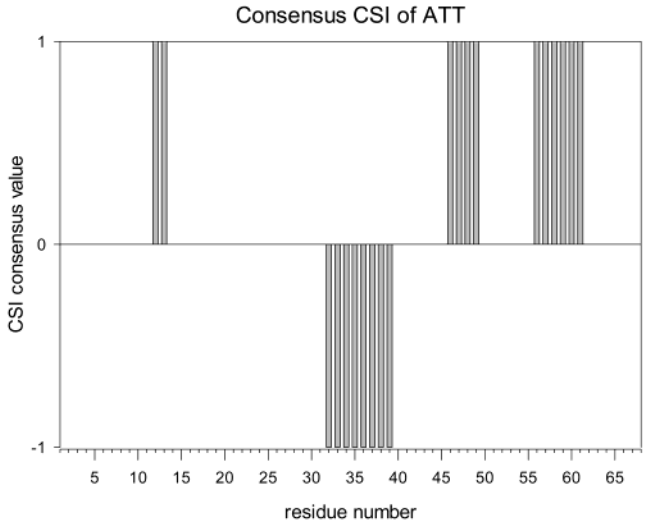

FIGURE 3: Consensus chemical shift index (CSI) (*40*) determined for ATT_p. Positive bars are indicative of stretches of β-sheet, and negative bars are indicative of α-helix.

The pattern of $d_{\alpha N}(i, i+3)$ and $d_{\alpha\beta}(i, i+3)$ connectivities (Figure 5) identified the presence of an α-helix. A series of strong $d_{\alpha N}$ sequential connectivities, together with the pattern of slowly exchanging amide hydrogens, led to elucidation of the β-strands. A short 3_{10}-helix (residues 21−24) was suggested by $d_{\alpha N}(i, i+2)$ connectivities, $^{3}J_{HN\alpha}$ values, and TALOS predictions; however, although observed in a few of the final structures, the 3_{10}-helix was not present in the final minimized mean structure. Residues 12−15, which correspond to a β-strand in brazzein and a subgroup of knottin proteins, were found not to adopt a β-strand conformation in ATT.

Comparisons of Chemical Shifts of ATT_p with Those of Other Proteins. CS$\alpha\beta$ proteins contain a $C_1\cdot\cdot C_2$-XXX-$C_3\cdot\cdot$G-X-$C_4\cdot\cdot C_5$-X-C_6 motif. The $^{1}H^{\alpha}$ resonances of the G, C_4, C_5, and C_6 residues within this motif are deshielded by the proximity of the disulfide bridges between C_2 and C_5 and between C_3 and C_6 (*26*). To compare the chemical shifts of ATT_p with those of other CS$\alpha\beta$ proteins for which NMR data are available, we calculated the deviations of the $^{1}H^{\alpha}$ chemical shifts from the standard random coil values used in CSI analysis (*40*), which are available from the BMRB website (http://www.bmrb.wisc.edu/ref_info/csishift.txt). The secondary $^{1}H^{\alpha}$ chemical shifts of the cysteines of ATT were compared with those of the sequence-aligned cysteines of snake venom toxins and EFG-like domains (see Figure S3 of the Supporting Information). Whereas seven of eight cysteines of the three snake venom toxins in the comparison exhibit very similar $^{1}H^{\alpha}$ chemical shifts, only two of the cysteines of ATT have comparable chemical shifts. Because it had been proposed that the first six cysteine residues of ATT have the same disulfide pairing topology as EGF domains (*12*), we compared the secondary chemical shifts of these six cysteines; however, the dissimilarity of the chemical shift patterns failed to support this hypothesis (see Figure S3 of the Supporting Information).

Structure Calculations. More than 500 $^{1}H-^{15}N$ NOESY cross-peaks, assigned manually with the program SPARKY (*28*), yielded 438 unique DYANA upper limit constraints.

plant proteinase inhibitors of the rapeseed family

1. ATT_p -CPEIEAQGNECLKEYGGDV-GFGFCAPRIFPT--*ICYTR*----------*CRENK*--GAKGGRCRWGQGSN---VKCLCDF----------CGDTPQ- 68
2. MTI-2 --------DSECLKEYGGDV-GFPFCAPRIFPT--ICYTR----------CRENK--GAKGGRCIWGEGTN---VKCLCDY----------CNDSPFD 63
3. RTI-III --------DSECLKEYGGDV-GFGFCAPRIYPS--FCVQR----------CRADK--GALSGKCIWGQGSN---VKCLCNF----------CRHEP-- 68

brazzein (a sweet protein)

4. brazzein --------XDKCKKVYENYP--VSKCQLA-----N*QCNYD*----------*CKLDK*--HARSGECFYDEKRN---LQCICDY----------CEY--- 54

scorpion toxin-like proteins

5. charybdoxin ----------FTNVSCTT-----*SKECWSV*----------*CQRLH*--NTSRGKCMN--------KKCRCYS---------- 36
6. pandinotoxin ----------TISCTNEKQ-----*CYPH*----------*CKKETG*-YPN-AKCMN--------RKCKCFGR---------- 35
7. toxin II KEGYLVDKNTGCKY----------ECLKLGDND--*YCLRE*----------*CKQQYG*-KGAGGYCYA--------FACWCTHLYEQAIVWPLPNKRCX---- 66
8. centruroides -KDGYLVEKTGCKK----------TCYKLGEN--*DFCNRE*----------*CKWKHI*-GGSYGYCYG--------FGCYCEGLPDSTQTWPLPNKTC----- 63

snake venom toxin-like proteins

9. cardiotoxin II ------LKCNKLVP---LFYKTCPAGKN----LCYKMFMVSNLTVPVKRGCID---------VCPKNSALVK--YVC-CNT----------DRCN---- 60
10. cardiotoxin III ------LKCNKLVP---LFYKTCPAGKN----LCYKMFMVATPKVPVKRGCID---------VCPKSSLLVKY--VC-CNT----------DRCN---- 60
11. toxin FS21 ------RICYSHKASLPRATKTCVEN------TCYKMFIRTHREYISERGCG----------CPTAMWPYQT--EC-CKG----------DRCNK--- 61
12. cardiotoxin A3 ------LKCNKLVP---LFYKTCPAGKN----LCYKMFMVATPKVPVKRGCID---------VCPKSSLLVKY--VC-CNT----------DRCN---- 60

epidermal growth factor domains

13. murine EGF ------NSYPGCPSSYD-----GYCLNG------GVCMHIESLD------SYTCN----------CVIGYSGD----RCQTRDLRWWELR---------- 53
14. fibrilin 1 ----SAVDMDECKEPD-------VCKH-------GQCINTDGSY--------RCE----------CPFGYILAGN--ECVD---------- 43
44 --------TDE*CS*VGN-------PCGN-------GTCKNVIGGF--------ECT----------CEEGFEPGPMM-TCE---------- 82
15. protein S ------KDVDECSLKPS------ICGT-------AVCKNIPGDF--------ECE----------CPEGYRYNLKSKSCEDI---------- 45

FIGURE 4: Alignment of the sequence of ATT_p with those of other proteinases of the rapeseed family and with members of other related families of proteins. The alignment takes into account known structural information for each protein. Italics denote helix, and underlines denote strand. The conserved cysteines and glycines for CS$\alpha\beta$ proteins are shaded. The numbers of amino acids in the sequences are shown at the right end of each row. The PDB codes (italic) for the tertiary structures and BMRB accession numbers (bold) for the chemical shift assignments are as follows: (1) ATT (**5056**), (4) brazzein (*1BRZ* and **4067**), (5) charybdoxin (*2CRD* and **114**), (6) pandinotoxin (*1C49* and **4760**), (7) toxin 2 from *Centruroides noxius* hoffmann (*1CN2* and **4218**), (8) *Centruroides sculpturatus* Ewing toxin I (*1B3C* and **4279**), (9) cardiotoxin II (*1CRE* and **1376**), (10) cardiotoxin III (*2CRS*), (11) toxin FS2 (*1TFS*), (12) cardiotoxin A3 (**4966**), (13) murine epidermal growth factor (*1EGF* and **2201**), (14) fibrilin (two domains) (*1EMO*), and (15) the EGF domain of protein S (**4728**) (chemical shift assignments are available for both major and minor conformational forms of the protein).

A total of 20 ϕ and ψ angle constraints were obtained from analyzing $^3J_{HNH\alpha}$ data; 14 of these were within the α-helix segment (Ile^{32}−Glu^{39}). The α-helical segment was further verified by the slow exchange of amide protons for residues Thr^{35}−Glu^{39} and by $d_{\alpha N}(i, i+3)$ connectivities deduced from NOE cross-peaks. Therefore, eight hydrogen bond constraints (four upper and four lower limits) were used in the initial calculations for the α-helical region (residues 32−39). The initial structures were investigated to identify hydrogen bonds in the β-sheet structure. These observations were compared with amino acid sequence alignments and were verified by hydrogen exchange measurements; hydrogen bond constraints for the β-sheet were added in later rounds of structure refinement.

After iteration of the structure calculations and NOE assignments, we obtained initial structures that suggested disulfide bridges between Cys^{11} and Cys^{63} and between Cys^{24} and Cys^{48}. In addition, we observed H^{β}−H^{β} NOE cross-peaks linking Cys^{11} to Cys^{63}, Cys^{24} to Cys^{48}, and Cys^{37} to Cys^{60}. These results confirmed those of an independent mass spectrometric study of proteolytic fragments ATT_{m+4}, a truncated version of ATT that lacks the precursor sequence (Q. Zhao, Y. K. Chae, M. V. Vestling, and J. L. Markley, manuscript in preparation). Constraints for the S_i−S_j and S_i−C^{β}_j distances of each of the four disulfide bonds (12 upper limit and 12 lower limit restraints) were added in the final

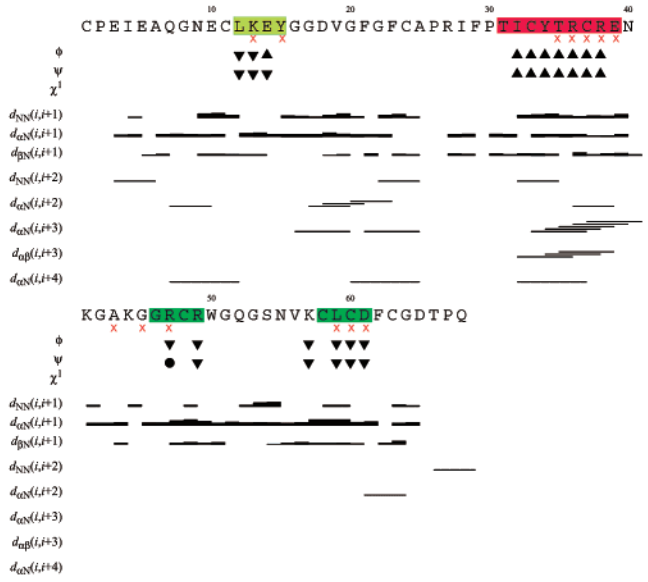

FIGURE 5: Amino acid sequence of ATT_p and NMR constraints used to elucidate its secondary structure. Residues with a slowly exchanging amide proton are indicated with an x. The color code for secondary structures is as follows: red for α-helix, dark green for β-strand, and light green for β-strand in structurally related proteins (brazzein and a subgroup of knottins). Other symbols: (●) dihedral angle for β-strands, (▲) dihedral angle for α-helix, and (▼) dihedral angle for coiled structure.

Table 1: Detailed Analysis of Residues in the Disallowed Regions of the Ramachandran Plot[a]

residue	angular order parameter S^{ϕ}	angular order parameter S^{ψ}	backbone dihedral angle ϕ	backbone dihedral angle ψ	no. of medium- and long-range constraints
Ala[25]	**0.210**	0.861	90.5 ± 97.1	126.5 ± 32.0	0
Arg[27]	**0.434**	0.873	−127.1 ± 78.2	−52.9 ± 35.7	1
Ile[28]	0.861	0.77	−128.8 ± 32.5	173.9 ± 45.2	0
Thr[31]	**0.442**	0.984	100.6 ± 66.3	42.6 ± 10.5	3
Lys[41]	0.996	**0.531**	62.2 ± 5.3	165.4 ± 65.4	1
Ala[43]	**0.491**	0.95	−165.7 ± 71.3	123.4 ± 18.8	3
Gln[52]	0.781	**0.574**	159.0 ± 42.6	73 ± 96.7	1
Ser[54]	0.826	0.796	−176.7 ± 39.0	10.5 ± 38.3	2

[a] Angular order parameters of <0.6 are in bold.

refinement stages. The TALOS program was used to generate additional backbone dihedral angle constraints from chemical shifts within individual secondary structural elements.

The final refinement made use of 953 upper limit constraints derived from NOE data (155 long-range, 123 medium-range, 286 short-range, and 389 intraresidue). This corresponds to an average of 14 restraints per residue. The NOE-derived restraints were supplemented by 26 upper limit constraints (14 for hydrogen bonds and 12 for disulfide bonds), 26 lower limit constraints (14 for hydrogen bonds and 12 for disulfide bonds), and 32 dihedral angle (ϕ and ψ) constraints. In the absence of lower bound disulfide constraints, several Cys S^{γ} atoms were too close to each other.

Structure Validation. Structural statistics are summarized in the Supporting Information (Table S3). The average value of the DYANA target function was 1.10 ± 0.17 Å. This indicates that the family of conformers agrees well with the experimental constraints. The minimized mean structure had a target function value of 1.15 Å. For residues 10−64, the average pairwise rmsd between the minimized structure and members of the 20-structure ensemble of conformers was 1.08 ± 0.31 Å for backbone atoms and 1.85 ± 0.42 Å for all heavy atoms.

Ramachandran plot analysis of residues 7−67 of the ensemble of 20 conformers showed that 54% of the nonglycyl residues fall into the most favored regions of torsion angle space, 28% of the residues are in other allowed regions, and 15% of the residues fall in the generously allowed region. All residues in the secondary structural elements of ATT_p lie in the allowed regions of torsion angle space. Eight residues (25, 27, 28, 31, 41, 43, 52, and 54) were found to be in disallowed regions in at least one of the 20 conformers (see Figure S4 of the Supporting Information). Few medium- or long-range constrains were found for residues 25, 27 (the P_1 residue of the reactive site), 28 (the P_1' residue of the reactive site), 41 (loop), and 52 (loop). Most of these residues have low angular order parameters ($S < 0.5$) for backbone dihedral angles (S^{ϕ} or S^{ψ}) and high standard deviations for angles ϕ and/or ψ (Table 1). Residues 27, 52, and 54 also have below average htNOE values (Figure 7A). Interestingly, almost all of these residues (except Ile[28]) are located adjacent to a Pro or Gly residue. For residues 10−64, the ensemble of conformers representing the ATT_p structure (Figure 6A) has an average rmsd with respect to mean structure values of 0.91 ± 0.25 Å for backbone atoms and 1.54 ± 0.28 Å for all heavy atoms. For the most structured regions (residues 12−15, 32−29, 45−48, and 58−61), average rmsd values with respect to the mean structure were 0.42 ± 0.14 Å for backbone and 1.06 ± 0.21 Å for heavy atoms. The regions of secondary structure are well-defined. The family of conformers has no distance constraint violations greater than 0.35 Å.

Hydrogen Bonds. Slowly exchanging amide hydrogens were identified at positions 13, 15, 35−39, 43, 45, 47, and 59−61. With the exception of residue 13, all hydrogen bond acceptors were identified (Table 2). In seven of the 20 final structures, a hydrogen bond was indicated between residues 41 and 43 within the type IV β-turn (residues 40−43). However, this hydrogen bond was not indicated in the minimized mean structure. All hydrogen bonds within the α-helical region were identified. The hydrogen bonds forming the antiparallel β-sheet and the observed NOEs supporting this secondary structure are shown in Figure 7.

Order and Dynamics. Well-defined backbone torsion angles (angular order parameter $S > 0.8$) were observed within or adjacent to the regions of recognizable secondary structure (Ile[32]−Asn[40], Lys[44]−Arg[49], and Cys[56]−Cys[62]) and in other residues with low rmsd values and large numbers of distance constraints (Figure 7B−E). Regions with poorly defined backbone torsion angles included the precursor peptide (Cys[1]−Ala[6]), the three residues at the C-terminus (residues 65−67), and the connecting loop regions (residues 7, 8, 17, 41−43, and 50−52); these regions exhibited S angular order parameters of <0.6 (for ϕ and/or ψ) and larger rmsd values (2.25 ± 0.60 Å for the backbone and 2.88 ± 0.59 Å for heavy atoms).

The rmsd values for the inhibitor loop (residues 24−30) were 1.17 ± 0.44 Å for backbone and 2.34 ± 0.76 Å for heavy atoms. In the inhibitory loop, residues Cys[24] (P_4) to Arg[27] (P_1) yielded low angular order parameters, whereas residues Ile[28] (P_1') to Pro[30] (P_3') were better ordered (angular order parameter $S > 0.7$). The C-terminal residue (Gln[68]), which was constrained by few medium- and long-range NOEs, exhibited a moderate angular order parameter ($S^{\phi} = 0.63$); the rmsd values for backbone and heavy atoms at this position were >3 Å.

The internal mobility of ATT_p was investigated by measuring $^1H-^{15}N$ heteronuclear NOE (ht-NOE) values for most of the backbone hydrogen and nitrogen atoms. Negative ht-NOE values were obtained for residues 3−6; such values are indicative of large amplitude motions (on the picosecond to nanosecond time scale) and a lack of defined structure (*32*). Residues with below average htNOE values (indicative of higher mobility) included the region immediately adjacent to the precursor sequence (Gln[7]−Cys[11]), the reactive site P_1' residue (Ile[28]), the residues connecting the two β-strands (Gly[53]−Asn[55]), and residues near the C-terminus (Asp[65], Thr[66], and Gln[68]) (Figure 7A).

DISCUSSION

Characteristics of the ATT_p Structure. As determined by the program PROMOTIF (*38*) (Figure 6B), ATT_p contains one α-helix (residues 31−39) and an antiparallel β-sheet in a β-hairpin conformation, which consists of two β-strands (residues 45−48 and 58−61) connected by a type IV β-turn (residues 52−55). The putative reactive site loop is solvent-exposed, with the P_1 residue (Arg[27]) pointing outward to the solvent (Figure 6C). The side chains of the aromatic residues are mostly exposed to solvent (Figure 6D). The structure is

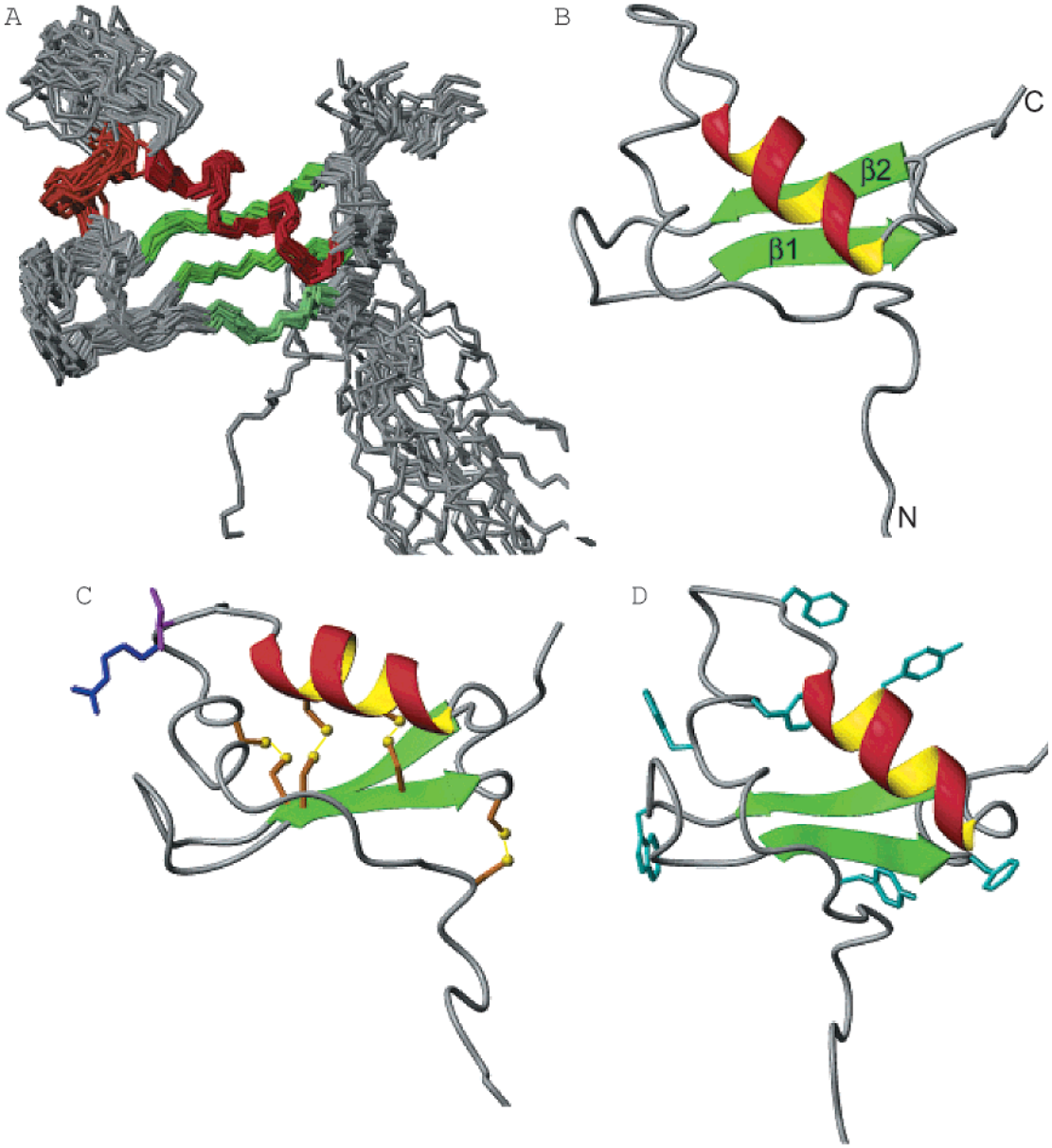

FIGURE 6: NMR structure of ATT_p. (A) Ensemble of 20 conformers with the lowest target function from the total of 50 final calculated conformers; each of these was superimposed on backbone atoms 12−15, 32−39, 45−49, and 57−61 of the mean structure. The β-strands are in green, and the α-helix is in red. Residues 12−14 are highlighted in light green. (B) Minimized mean structure of ATT_p shown as a ribbon diagram. The β-strands are in green, and the α-helix is in red and yellow. (C) Ribbon diagram of ATT_p showing the side chains of the inhibitory loop of residues P_1-P_1' (Arg^{27} in blue and Ile^{28} in coral) and disulfides (bonds in orange and sulfur atoms in yellow). (D) Side chains of the hydrophobic residues of ATT_p (cyan). Structures in panels A and B have the same orientation; those in panels C and D were rotated to better visualize the residues of interest.

stabilized by four disulfide bonds; the solvent accessibilities of the eight cysteines that form these linkages are lower than average.

The internal mobility of ATT_p was probed by analyzing the angular order parameters, the backbone rmsd values, and the htNOE data (Figure 7A−D). The regions of secondary structure are all well-defined, whereas the connecting loop regions and the precursor sequence are less ordered.

Comparison with Other Structures. When we determined the structure of brazzein, we predicted on the basis of sequence alignments that ATT might be structurally similar to the family of CSαβ proteins (*41*; Figure 8A). The structure presented here confirms this hypothesis. The ATT_p sequence is also similar to those of a subgroup of knottin proteins (represented by Rs-Afp1). The rmsd value calculated with the DALI software package (http://www.ebi.ac.uk/dali/) between ATT_p and brazzein (PDB entry 2BRZ, ∼28% identical and ∼38% homologous in primary sequence to ATT_p) was 3.0 Å; that between ATT_p and Rs-Afp1 (PDB entry 1AYJ, ∼28% identical and ∼36% homologous in primary sequence to ATT_p) was 2.7 Å. Within the regions of secondary structure, the similarity was still higher: 2.2 Å between ATT_p and brazzein and 1.9 Å between ATT_p and Rs-Afp.

We also superimposed the structure of ATT_p (minimized mean structure) with those of five related plant proteins (PDB entries 1AYJ, 1MYN, 1GPT, 1BRZ, and 1BK8) using the best structure from each ensemble in the case of NMR structures. The superposition (Figure 8B) resulted in a good alignment of secondary structures. The major differences among the structures were in loop 1 connecting the first

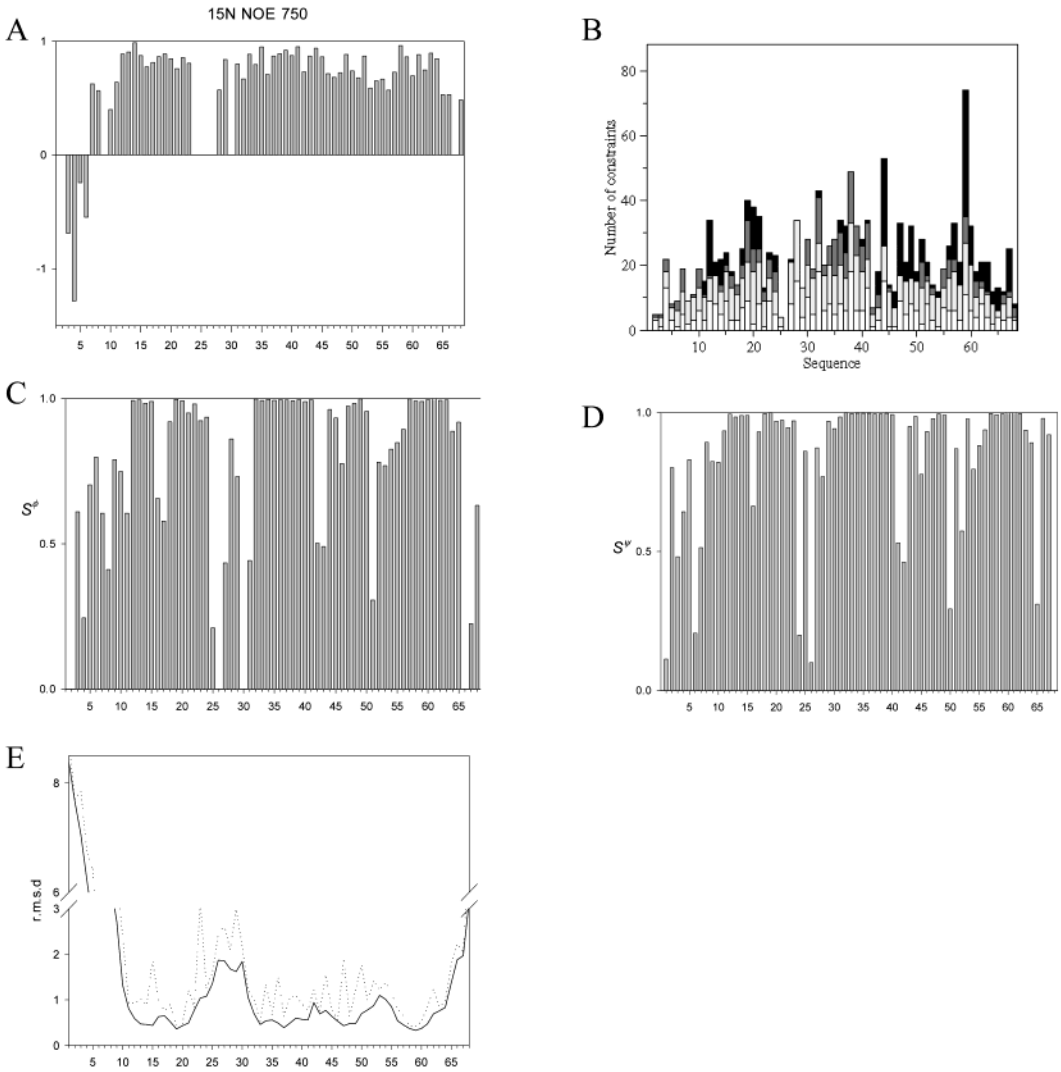

FIGURE 7: Internal dynamics and order in the ATT_p structure. (A) Heteronuclear $^{15}N-^{1}H$ NOE values vs sequence. Lower values indicate greater backbone mobility. The dashed line indicates the average value for positive NOEs. (B) Distribution of NOE constraints along the primary sequence: intraresidue (white), short-range (light gray), medium-range (dark gray), and long-range NOEs (black). (C) ϕ angular order parameters vs sequence [analyzed by MOLMOL (*37*)]. (D) ψ angular order parameters vs sequence (analyzed by MOLMOL). (E) Average backbone (—) and heavy atom (···) rmsd values for the final ensemble of 20 conformers relative to the mean structure.

Table 2: Hydrogen Bond Donors and Acceptors Identified in the Structure of ATT_p

		H^N-O distance (Å)	
residue number of the donor (H^N)	residue number of the acceptor (O)	range in the 20 conformers (minimum−maximum)	minimized structure
15[a]	58	1.77−2.45	1.92
35[b]	31	1.77−2.68	2.67
36[b]	32	1.72−2.72	1.79
37[b]	33	1.48−1.92	1.75
38[b]	34	1.80−2.29	2.28
39[b]	35	2.01−2.89	2.82
43[c]	41	1.78−2.44	—
45[a]	61	1.89−2.89	2.82
47[a]	59	1.96−2.82	2.14
59[a]	47	1.64−2.56	1.99
60[a]	13	1.67−2.29	2.02
61[a]	45	1.82−2.88	2.58

[a] Hydrogen bond in the β-strand. [b] Hydrogen bond in the α-helix. [c] Present in only seven of the 20 conformers.

conserved β-strand to the beginning of the α-helix (residues 16−30 in ATT_p) and loop 3 connecting the second and third conserved β-strands (a third strand was not observed in ATT_p). It has been suggested that these two regions are important for the antifungal and receptor binding activity of Rs-Afp (*42*). Furthermore, homologues of Rs-Afp that lack antifungal activity show the largest sequence divergence from Rs-Afp in these two regions (*43*). The largest structural differences between the two families of plant defensins exemplified by Ah-Amp1 (1BK8) and Rs-Afp are observed in these regions and have been implicated in their different functional activities (*44*). In the case of brazzein, single-amino acid substitutions of several residues in these two regions decreased its sweetness (*45*). In ATT_p, the reactive site (residues 25−30) is located in loop 1, which connects the first conserved β-strand of the CSαβ fold to the beginning of the α-helix. Therefore, the differential activities and specificities of homologous CSαβ proteins of plant origin may reside in these regions.

We also compared the topologies of the disulfide bonds (χ_{ss} angles) of ATT_p with those of brazzein and a subgroup of knottin proteins, a total of 14 structures (best structure of the ensemble in cases of NMR structures; see Table S4 of the Supporting Information). The disulfide topologies of

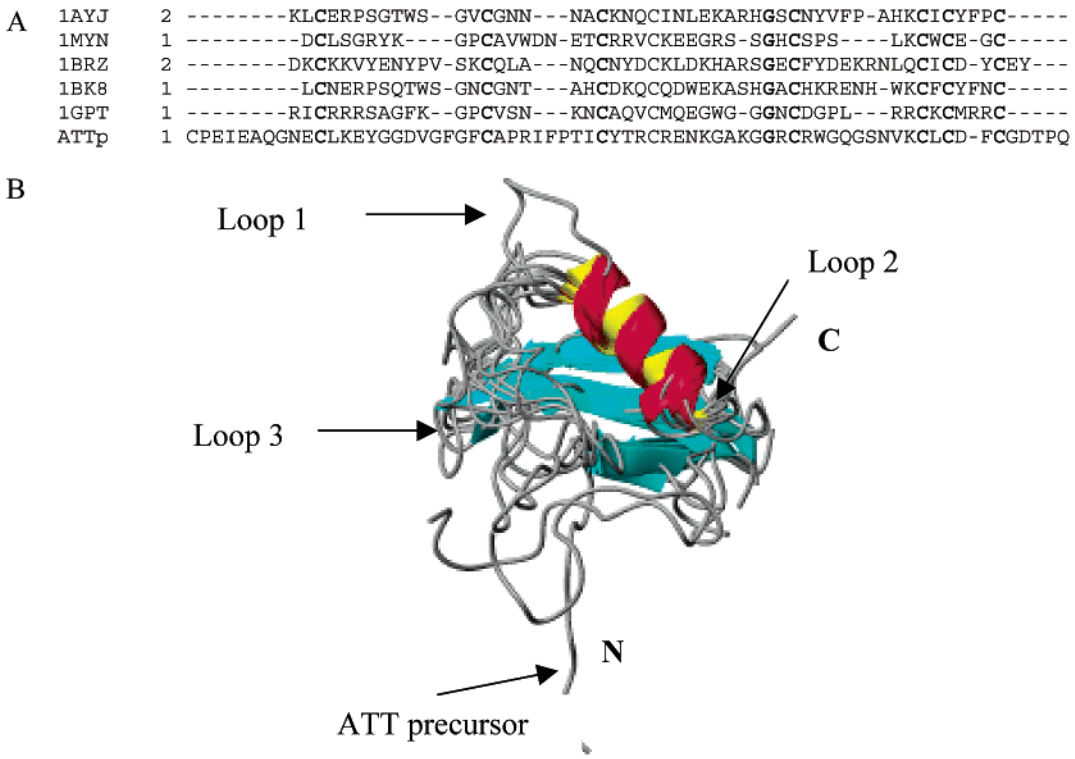

FIGURE 8: (A) Primary sequence alignment of ATT_p and five other plant CS$\alpha\beta$ proteins: Rs-Afp (1AYJ), drosomysin (1MYN), brazzein (1BRZ), Ah-Amp 1 (1BK8), and γ-1H thionin (1GPT). The conserved -C-$(X)_n$-C-$(X)_n$-CXXXC-$(X)_n$-GXC-$(X)_n$-CXC-$(X)_n$-C- motifs for these CS$\alpha\beta$ proteins are shown in bold letters. (B) Superposition of the minimized mean structure of ATT_p with those 1AYJ, 1myn, 1gpt, 1brz, and 1bk8 (NMR structures, best structure in the ensemble if applicable). The aligned residues are 33−37, 45−47, and 58−60 for ATT_p, 21−25, 34−36, and 45−47 for 1AYJ, 19−23, 31−33, and 39−41 for 1MYN, 20−24, 32−34, and 41−43 for 1GPT, 22−26, 35−37, and 47−49 for 1BRZ, and 20−24, 33−35, and 44−46 for 1BK8. The connecting loops are indicated by the arrows: loop 1, $\beta1-\alpha$; loop 2, $\alpha-\beta2$; and loop 3, $\beta2-\beta3$. In the case of ATT_p, the first β-strand is not well-defined.

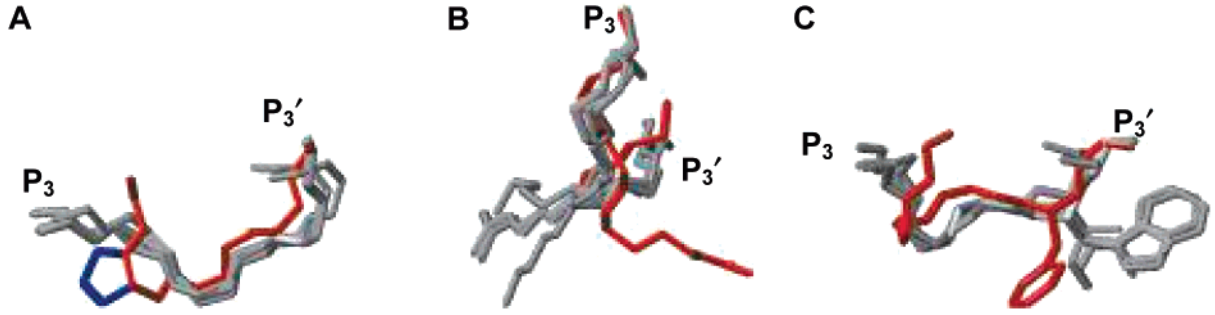

FIGURE 9: Comparison of the backbone conformation of the reactive site loop region of ATT_p (minimized mean structure) (residues 25−30) in red with those of members of the squash [1F2S (residues 303−308), 1MCT (residues 3−8), and 1PPE (residues 3−8)] and BBI [2BBI (residues 14−19) and 1DF9 (residues 518−523)] families. (A) Kink at the P_2 position of ATT_p (Pro^{26} ring shown in blue). (B) Side chain orientations of the P_1 residues. The side chain of Arg^{27} of ATT_p is in red; those of the other P_1 residues are in gray. (C) Side chain orientations of the P_2' residues: Phe^{29} of ATT_p in red and the other P_2' residues in gray.

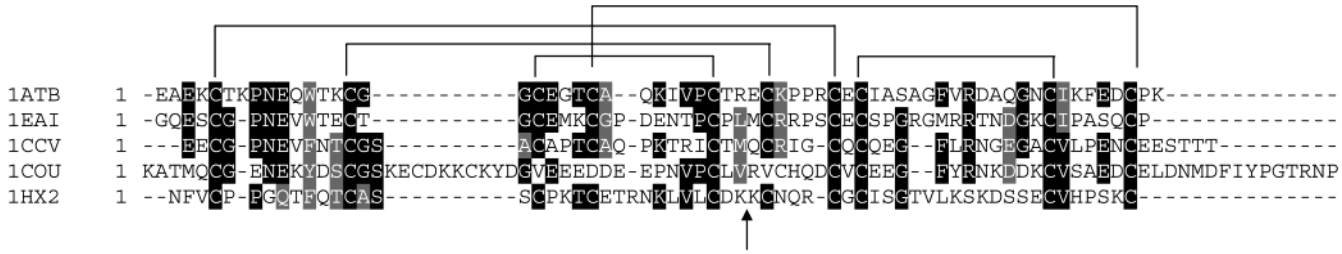

FIGURE 10: Alignment of the sequences of members of the ATI class of serine protease inhibitors (represented by their PDB codes), their disulfide bond arrangement, and their reactive sites (denoted with an arrow).

ATT_p can be characterized (*38*) as a right-handed hook (first disulfide bond, $Cys^{11}-Cys^{63}$), a left-handed spiral (second disulfide bond, $Cys^{24}-Cys^{48}$), and a right-handed hook (third disulfide bond, $Cys^{33}-Cys^{63}$). The conformation of the fourth disulfide bond ($Cys^{37}-Cys^{60}$) does not fit any of the more common topologies. The conformations observed in ATT_p, however, are observed in the homologous disulfide bridges of all 14 structures.

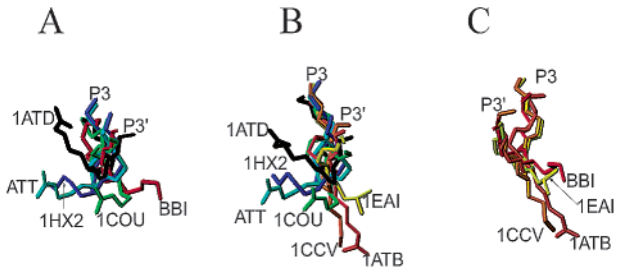

FIGURE 11: Comparison of the backbone conformations of the reactive site loop of ATT_p (minimized mean structure) (residues 25−30 aligned) with those of members of the ATI [1ATD (pH 4.5, residues 30−35), 1ATB (pH 2.5, residues 30−35), 1CCV (pH 2.5, residues 28−33), 1COU (pH 4.5, residues 42−47), 1HX2 (pH 5.0, residues 30−35), and 1EAI (residues 30−35)] and BBI [2BBI (residues 14−19)] classes of serine proteinase inhibitors. The backbone and P_1 side chains of different inhibitors are distinguished by colors: red for BBI, cyan for ATT, orange for 1CCV, green for 1COU, blue for 1HX2, black for 1ATD, coral for 1ATB, and yellow for 1EAI. All structures are of inhibitors in the free state, except for 1EAI, which is in complex with porcine elastase. (A) Reactive loop conformations of the high-pH forms (ATT, 1ATD, 1COU, 1HX2, and 2BBI) compared with that of BBI. (B) Comparison between ATT (at high pH) and the high- and low-pH forms of the ATI class. (C) Low-pH forms of the ATI class (1ATB, 1CCV, and 1EAI) compared to BBI (2BBI).

Conformation of the Reactive Site Loop. The reactive sites of canonical serine proteinase inhibitors share a conserved backbone conformation with common ϕ and ψ angles for residues P_3-P_3' (*11*, *46*, *47*). On comparing the P_3-P_3' loop conformation of ATT_p with those of other types of serine proteinase inhibitors Kazal, Kunitz, Bowman-Birk (BBI), and squash], we found that only the ψ angle of P_1 (Arg^{27}) in ATT_p lies outside the common region (see Table S5 of the Supporting Information). When all restraints (sequential and nonsequential) were removed from the loop region (residues 25−30) prior to structure determination, the ϕ and ψ angle values for Arg^{27} continued to lie outside the consensus range in all the calculated conformers, but with increased errors. The anomaly might be due to the paucity of constraints for this region resulting from the solvent exposure of the loop or of internal mobility. The htNOE value for the P_1 residue is below average.

Panels A−C of Figure 9 show superpositions of the P_3-P_3' loop region of ATT_p with those of five other serine proteinase inhibitors (three from the squash family and two from the BBI family). With the exception of ATT_p, all have nearly identical P_3-P_3' backbone conformations. The reactive site loop of ATT_p and the other five inhibitors exhibit two similarities: the orientation of the peptide bond connecting P_1 and P_1' and the conformation of the side chain of the P_1' residue (Ile^{28} in ATT_p). ATT_p differs from the others in three aspects. (i) Pro^{26}, a residue at the P_2 position only in ATT_p, causes a kink in the backbone of the reactive loop not observed in the other five inhibitors (Figure 9A). (ii) The orientation of the P_1 residue of ATT_p differs from those of the other inhibitors (Figure 9B). (iii) The side chain of the P_2' residue in ATT_p has an orientation different from those of the other inhibitors (Figure 9C).

Rapeseed and Bowman-Birk are the only two classes of serine proteinase inhibitors that have cysteines flanking the reactive site (P_4 and P_6' for ATT_p and P_3 and P_6' for BBI) as well as proline at the P_3' position. The P_3' Pro in BBI is always in the cis conformation, and the P_4' Pro (when present) is in the trans conformation. The cis conformation of the P_3' prolyl peptide bond probably orients the BBI inhibitory loop to the canonical form. The P_2 proline of ATT_p is not present in BBI, and both the P_2 and P_3' prolines of ATT_p occupy the trans conformation. The unique sequence of the ATT_p inhibitory loop (P_3-Cys-Ala-Arg-Ile-Phe-Pro-Thr-Ile-Cys-P_6') seems to be responsible for the uncommon conformation in this region. The inhibitory mechanism for this class of inhibitors needs to be investigated further.

Comparison of the Conformation of the Reactive Loop of ATT_p to Those of ATI-Type Inhibitors. The *Ascaris* trypsin inhibitor from the common round worm is the prototypical ATI inhibitor. Structures of ATI-type serine proteinase inhibitors have been determined by NMR (*48*) and X-ray (*49*) techniques. Sequence alignments and the common disulfide-paring pattern are shown in Figure 11. For ATI, the conformation of the reactive loop is dependent on pH; at pH $\leq$2.4, the reactive loop adopts the conventional conformation observed in BBI and Kazal inhibitors among others, and at pH >4.5, the conformation of the P_2, P_1, and P_1' residues is noncanonical. The P_3-P_3' backbone dihedral angles for this class are provided in the Supporting Information (Table S5).

In the structure of ATT_p, which was determined at pH 5.0, the backbone dihedral angles of the P_3-P_3' residues, particularly the ψ angle of P_1, are reminiscent of the ATI class of inhibitors at high pH. We have also found that the P_1 ψ angles in a couple of ATI class inhibitors, at pH $\geq$4.5, are also negative (*50*, *51*), similar to that in ATT_p. In contrast, at lower pH (2.4 and 2.5) (*48*, *52*) or in a complex (*49*), the P_1 ψ angles become more canonical. Upon superimposition of the reactive loops (P_3-P_3') of ATT_p, ATI inhibitors, and a Bowman-Birk serine proteinase inhibitor determined by NMR (*53*), it is apparent that the conformation of the P_1 side chain of ATT_p resembles those of ATI-type inhibitors at high pH (Figure 11A) rather than those of ATI-type inhibitors at low pH or when complexed with a proteinase (Figure 11B). In the ATI structures, the orientation of the P_1 side chain is nearly opposite in the low- and high-pH forms. In contrast, the ATI-type inhibitors at low pH (1ATB, 1CCV, and 1EAI) share the same reactive loop conformation with BBI (2BBI) (Figure 11C). This leads us to speculate that the reactive loop of ATT_p may undergo similar changes either at low pH or in complex with proteinase.

ACKNOWLEDGMENT

We thank Dr. Brian Volkman for advice on structure calculations by DYANA, Dr. Frits Abildgaard for discussions on data acquisition and processing, Dr. Zsolt Zolnai for software assistance, Dr. Ed Mooberry for hardware maintenance, Dr. Tom D. Goddard for advice on using SPARKY, and Dr. Craig Newman for assistance in preparing the manuscript.

SUPPORTING INFORMATION AVAILABLE

Full description of the inhibitory assays and the results, parameters used in recording the two- and three-dimensional NMR data sets used here (Table S1), comparison of secondary structure of ATT deduced from chemical shift index (CSI) analysis and ^3J coupling analysis (Table S2), structure statistics for the minimized mean structure and the

family of 20 conformers that represent the solution structure of ATT_p (Table S3), comparison of the topologies of the disulfide bonds (χ_{ss} angles) of ATT_p with those of brazzein and a subgroup of knottin proteins (Table S4), the P_3-P_3' backbone dihedral angles for protein proteinase inhibitors with noncanonical reactive loops (Table S5), representative sequential walk through HNCA and HN(CO)CA data (Figure S1), the relationship between C^α and C^β chemical shifts of structurally relevant cysteines in ATT_p (Figure S2), comparison of secondary $^1H^\alpha$ chemical shifts of cysteines and conserved glycines in ATT with those of related classes of proteins (Figure S3), and a Ramachandran plot for the final ensemble of 20 conformers representing the ATT_p structure as analyzed by PROCHECK (Figure S4). This material is available free of charge via the Internet at http://pubs.acs.org.

REFERENCES

1. Bevan, M., Mayer, K., White, O., Eisen, J. A., Preuss, D., Bureau, T., Salzberg, S. L., and Mewes, H. W. (2001) Sequence and analysis of the *Arabidopsis* genome, *Curr. Opin. Plant Biol. 4*, 105−110.
2. Pennisi, E. (2000) Sequence. Plants join the genome sequencing bandwagon, *Science 290*, 2054−2055.
3. Willmann, M. R. (2001) *Arabidopsis* enters the post-sequencing era, *Trends Plant Sci. 6*, 51.
4. Menegatti, E., Tedeschi, G., Ronchi, S., Bortolotti, F., Ascenzi, P., Thomas, R. M., Bolognesi, M., and Palmieri, S. (1992) Purification, Inhibitory Properties and Amino Acid Sequence of a New Serine Proteinase Inhibitor from White Mustard (*Sinapis alba* L.) Seed, *FEBS Lett. 301*, 10−14.
5. Ceciliani, F., Bortolotti, F., Menegatti, E., Ronchi, S., Ascenzi, P., and Palmieri, S. (1994) Purification, inhibitory properties, amino acid sequence and identification of the reactive site of a new serine proteinase inhibitor from oil-rape (*Brassica napus*) seed, *FEBS Lett. 342*, 221−224.
6. Ryan, C. (1990) Proteinase inhibitors in plants: genes for improving defenses against insects and pathogens, *Annu. Rev. Phytopathol. 28*, 425−449.
7. De Leo, F., Bonade-Bottino, M., Ceci, L. R., Gallerani, R., and Jouanin, L. (2001) Effects of a mustard trypsin inhibitor expressed in different plants on three lepidopteran pests, *Insect Biochem. Mol. Biol. 31*, 593−602.
8. De Leo, F., and Gallerani, R. (2002) The mustard trypsin inhibitor 2 affects the fertility of *Spodoptera littoralis* larvae fed on transgenic plants, *Insect Biochem. Mol. Biol. 32*, 489−496.
9. Ceci, L. R., Spoto, N., Devirgilio, M., and Gallerani, R. (1995) The Gene Coding For the Mustard Trypsin Inhibitor-2 Is Discontinuous and Wound-Inducible, *FEBS Lett. 364*, 179−181.
10. Volpicella, M., Schipper, A., Jongsma, M. A., Spoto, N., Gallerani, R., and Ceci, L. R. (2000) Characterization of recombinant mustard trypsin inhibitor 2 (MTI2) expressed in *Pichia pastoris*, *FEBS Lett. 468*, 137−141.
11. Laskowski, M., and Qasim, M. A. (2000) What can the structures of enzyme−inhibitor complexes tell us about the structures of enzyme substrate complexes? *Biochim. Biophys. Acta 1477*, 324−337.
12. Ascenzi, P., Ruoppolo, M., Amoresano, A., Pucci, P., Consonni, R., Zetta, L., Pascarella, S., Bortolotti, F., and Menegatti, E. (1999) Characterization of low-molecular-mass trypsin isoinhibitors from oil-rape (*Brassica napus* var. oleifera) seed, *Eur. J. Biochem. 261*, 275−284.
13. Ceci, L. R., De leo, F., Volpicella, M., Jounin, L., and Gallerani, R. (1999) A gene family coding for proteinase inhibitors in rapeseed (*Brassica napus*), in *10th International Rapeseed Congress*, Canberra, Australia.
14. Volpicella, M., Ceci, L. R., Gallerani, R., Jongsma, M. A., and Beekwilder, J. (2001) Functional expression on bacteriophage of the mustard trypsin inhibitor MTI-2, *Biochem. Biophys. Res. Commun. 280*, 813−817.
15. Trovato, M., Maras, B., Polticelli, F., Costantino, P., and Ascenzi, P. (2000) A chimeric mini-trypsin inhibitor derived from the oil rape proteinase inhibitor type III, *Biochem. Biophys. Res. Commun. 275*, 817−820.
16. Hinck, A. P., Walkenhorst, W. F., Westler, W. M., Choe, S., and Markley, J. L. (1993) Overexpression and purification of avian ovomucoid third domains in *Escherichia coli*, *Protein Eng. 6*, 221−227.
17. Jameson, G. W., Roberts, D. V., Adams, R. W., Kyle, W. S., and Elmore, D. T. (1973) Determination of the operational molarity of solutions of bovine α-chymotrypsin, trypsin, thrombin and factor Xa by spectrofluorimetric titration, *Biochem. J. 131*, 107−117.
18. Bieth, J. G. (1995) Theoretical and Practical Aspects of Proteinase Inhibition Kinetics, *Methods Enzymol. 248*, 59−84.
19. Marion, D., and Wüthrich, K. (1983) Application of phase sensitive two-dimensional correlated spectroscopy (COSY) for measurements of $^1H-^1H$ spin−spin coupling constants in proteins, *Biochem. Biophys. Res. Commun. 113*, 967−974.
20. Kay, L. E., Keifer, P., and Saarinen, T. (1992) Pure absorption gradient enhanced heteronuclear single quantum correlation spectroscopy with improved sensitivity, *J. Am. Chem. Soc. 114*, 10663−10665.
21. Warren, W. S., Richter, W., Andreotti, A. H., and Farmer, B. T., II (1993) Generation of impossible cross-peaks between bulk water and biomolecules in solution NMR, *Science 262*, 2005−2009.
22. Talluri, S., and Wagner, G. (1996) An Optimized 3D NOESY-HSQC, *J. Magn. Reson. 112*, 200−205.
23. Vuister, G. W., Yamazaki, T., Torchia, D. A., and Bax, A. (1993) Measurement of two- and three-bond $^{13}C-^1H$ *J* couplings to the C delta carbons of leucine residues in staphylococcal nuclease, *J. Biomol. NMR 3*, 297−306.
24. Kuboniwa, H., Grzesiek, S., Delaglio, F., and Bax, A. (1994) Measurement of HN−H alpha *J* couplings in calcium-free calmodulin using new 2D and 3D water-flip-back methods, *J. Biomol. NMR 4*, 871−878.
25. Muhandiram, D. R., Farrow, N. A., Xu, G. Y., Smallcombe, S. H., and Kay, L. E. (1993) A gradient C-13 NOESY−-HSQC experiment for recording NOESY spectra of C-13-labeled proteins dissolved in H_2O, *J. Magn. Reson. 102*, 317−321.
26. Bax, A., Ikura, M., Kay, L. E., and Zhu, G. (1991) Removal of F_1 Baseline Distortion and Optimization of Folding in Multidimensional NMR Spectra, *J. Magn. Reson. 91*, 174−178.
27. Markley, J. L., Bax, A., Arata, Y., Hilbers, C. W., Kaptein, R., Sykes, B. D., Wright, P. E., and Wüthrich, K. (1998) Recommendations for the presentation of NMR structures of proteins and nucleic acids, *Pure Appl. Chem. 70*, 117−142.
28. Goddard, T. D., and Kneller, D. G. (2001) *SPARKY*, University of California, San Francisco.
29. Bartels, C., Xia, T.-h., Billeter, M., Gunthert, P., and Wüthrich, K. (1995) The program XEASY for computer-supported NMR spectral-analysis of biological macromolecules, *J. Biomol. NMR 5*, 1−10.
30. Kline, A. D., Braun, W., and Wüthrich, K. (1988) Determination of the complete three-dimensional structure of the alpha-amylase inhibitor Tendamistat in aqueous solution by nuclear magnetic resonance and distance geometry, *J. Mol. Biol. 204*, 675−724.
31. Wagner, G., Braun, W., Havel, T. F., Schaumann, T., Go, N., and Wüthrich, K. (1987) Protein structures in solution by nuclear magnetic resonance and distance geometry. The polypeptide fold of the basic pancreatic trypsin inhibitor determined using two different algorithms, DISGEO and DISMAN, *J. Mol. Biol. 196*, 611−639.
32. Farrow, N. A., Zhang, O. W., Formankay, J. D., and Kay, L. E. (1994) A heteronuclear correlation experiment for simultaneous determination of N-15 longitudinal decay and chemical exchange rates of systems in slow equilibrium, *J. Biomol. NMR 4*, 727−734.
33. Brunger, A. T. (1992) *X-PLOR, version 3.1. A system for X-ray crystallography and NMR*, Yale University Press, New Haven, CT.
34. Guntert, P., Mumenthaler, C., and Wüthrich, K. (1997) Torsion angle dynamics for NMR structure calculation with the new program DYANA, *J. Mol. Biol. 273*, 283−298.
35. Cornilescu, G., Delaglio, F., and Bax, A. (1999) Protein backbone angle restraints from searching a database for chemical shift and sequence homology, *J. Biomol. NMR 13*, 289−302.
36. Laskowski, R. A., Rullmannn, J. A., MacArthur, M. W., Kaptein, R., and Thornton, J. M. (1996) AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR, *J. Biomol. NMR 8*, 477−486.

37. Koradi, R., Billeter, M., and Wüthrich, K. (1996) MOLMOL: a program for display and analysis of macromolecular structures, *J. Mol. Graphics 14*, 51−55, 29−32.
38. Hutchinson, E. G., and Thornton, J. M. (1996) PROMOTIF: a program to identify and analyze structural motifs in proteins, *Protein Sci. 5*, 212−220.
39. Sharma, D., and Rajarathnam, K. (2000) ^{13}C NMR chemical shifts can predict disulfide bond formation, *J. Biomol. NMR 18*, 165−171.
40. Wishart, D. S., Sykes, B. D., and Richards, F. M. (1992) The chemical shift index: a fast and simple method for the assignment of protein secondary structure through NMR spectroscopy, *Biochemistry 31*, 1647−1651.
41. Caldwell, J. E., Abildgaard, F., Džakula, Ž., Ming, D., Hellekant, G., and Markley, J. L. (1998) Solution structure of the thermostable sweet-tasting protein brazzein, *Nat. Struct. Biol. 5*, 427−431.
42. De Samblanx, G. W., Goderis, I. J., Thevissen, K., Raemaekers, R., Fant, F., Borremans, F., Acland, D. P., Osborn, R. W., Patel, S., and Broekaert, W. F. (1997) Mutational analysis of a plant defensin from radish (*Raphanus sativus* L.) reveals two adjacent sites important for antifungal activity, *J. Biol. Chem. 272*, 1171−1179.
43. Fant, F., Vranken, W., Broekaert, W., and Borremans, F. (1998) Determination of the three-dimensional solution structure of *Raphanus sativus* antifungal protein 1 by ^{1}H NMR, *J. Mol. Biol. 279*, 257−270.
44. Fant, F., Vranken, W. F., and Borremans, F. A. (1999) The three-dimensional solution structure of *Aesculus hippocastanum* antimicrobial protein 1 determined by ^{1}H nuclear magnetic resonance, *Proteins 37*, 388−403.
45. Assadi-Porter, F. M., Aceti, D. J., and Markley, J. L. (2000) Sweetness determinant sites of brazzein, a small, heat-stable, sweet-tasting protein, *Arch. Biochem. Biophys. 376*, 259−265.
46. Laskowski, M., Jr., and Kato, I. (1980) Protein inhibitors of proteinases, *Annu. Rev. Biochem. 49*, 593−626.
47. Bode, W., and Huber, R. (1992) Natural protein proteinase inhibitors and their interaction with proteinases, *Eur. J. Biochem. 204*, 433−451.
48. Grasberger, B. L., Clore, G. M., and Gronenborn, A. M. (1994) High-resolution structure of *Ascaris* trypsin inhibitor in solution: direct evidence for a pH-induced conformational transition in the reactive site, *Structure 2*, 669−678.
49. Huang, K., Strynadka, N. C., Bernard, V. D., Peanasky, R. J., and James, M. N. (1994) The molecular structure of the complex of *Ascaris* chymotrypsin/elastase inhibitor with porcine elastase, *Structure 2*, 679−689.
50. Duggan, B. M., Dyson, H. J., and Wright, P. E. (1999) Inherent flexibility in a potent inhibitor of blood coagulation, recombinant nematode anticoagulant protein c2, *Eur. J. Biochem. 265*, 539−548.
51. Rosengren, K. J., Daly, N. L., Scanlon, M. J., and Craik, D. J. (2001) Solution structure of BSTI: a new trypsin inhibitor from skin secretions of *Bombina bombina*, *Biochemistry 40*, 4601−4609.
52. Cierpicki, T., Bania, J., and Otlewski, J. (2000) NMR solution structure of *Apis mellifera* chymotrypsin/cathepsin G inhibitor-1 (AMCI-1): structural similarity with *Ascaris* protease inhibitors, *Protein Sci. 9*, 976−984.
53. Werner, M. H., and Wemmer, D. E. (1992) Three-dimensional structure of soybean trypsin/chymotrypsin Bowman-Birk inhibitor in solution, *Biochemistry 31*, 999−1010.

BI025702A