High-resolution structure of *Ascaris* trypsin inhibitor in solution: direct evidence for a pH-induced conformational transition in the reactive site

Bruce L Grasberger†, G Marius Clore* and Angela M Gronenborn*

Laboratory of Chemical Physics, Building 5, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, MD 20892, USA

Background: The *Ascaris* trypsin inhibitor (ATI) is a member of a new family of serine protease inhibitors isolated from the helminthic worm *Ascaris lumbricoides* var *suum*. This family comprises five chymotrypsin/elastase inhibitors and one trypsin inhibitor. Members are characterized by the presence of five disulfide bonds (two of which are located on either side of the reactive site) in a single small protein domain of 61–62 residues.

Results: The solution structure of ATI has been determined at pH 2.4 and pH 4.75 by NMR spectroscopy. Iterative refinement permitted the unambiguous assignment of the pairing of the five disulfide bridges (Cys5–Cys38, Cys15–Cys33, Cys18–Cys29, Cys22–Cys60, and Cys40–Cys54) which were previously unknown. The structure includes four short β-strands arranged in two approximately perpendicular β-sheets. The reactive site loop is bounded by two disulfide bridges (Cys15–Cys33 and Cys18–Cys29) and is part of the long loop (residues 15–25) connecting strands β1 and β2. Comparison of the nuclear Overhauser enhancement data at the two pH values revealed significant differences centered around the reactive site. While the reactive site at pH 2.4 closely resembles that of other protease inhibitors, at pH 4.75 the reactive site loop undergoes a major conformational rearrangement involving the ψ backbone torsion angles of the P_2, P_1 and P_1' residues (residues 30–32). This is associated with a change in the positions of the side chains of Arg31 and Glu32.

Conclusions: The overall three-dimensional structure of ATI posesses an unusual fold and, with the exception of the reactive site, shows no similarity to other serine protease inhibitors. The observation that the reactive site of the low pH form of ATI is similar to that of other serine proteases suggests that this is the active form of the protein.

Structure 15 July 1994, **2**:669–678

Key words: *Ascaris* trypsin inhibitor, NMR spectroscopy, pH-dependent conformational transition, solution structure

Introduction

Serine protease inhibitors function by binding to their cognate enzyme in a substrate-like manner, forming a stable complex. They are of broad interest because serine proteases play key roles in such functions as peptide hormone release, blood coagulation and complement fixation, and constitute pathogenic factors in numerous diseases, including some cancers, pulmonary emphysema and inflammatory processes such as glomerulonephritis and acute pancreatitis [1,2]. Serine protease inhibitors have been grouped into at least 10 families based primarily on sequence homology [3]. Variability at the reactive site within a family can lead to differences in specificity for proteases within a given mechanistic class. The three-dimensional structure of inhibitors from several of these families have been determined by X-ray crystallography [4–10] and NMR spectroscopy [11–16]. The observation that inhibitors from families having very different primary and tertiary structures can share a similar reactive site conformation and specificity for a given target protease indicates that the specificity is determined by a small region of the larger molecule [17]. Thus, these inhibitors serve as a good example of convergent evolution.

The primary sequence has been determined for six members of a new family of inhibitors isolated from the helminthic parasite *Ascaris lumbricoides var suum*, comprising five chymotrypsin/elastase inhibitors and one trypsin inhibitor [18–21]. This family of protease inhibitors, which shares no sequence homology with other protease inhibitor families, is characterized by the presence of five disulfide bonds in a single small protein domain of 61 to 62 residues. The disulfide bonds account for the unusual resistance to proteolysis and heat denaturation of these proteins. Although a preliminary report of the sequential NMR assignment and secondary structure determination of the *Ascaris* trypsin inhibitor (ATI) has appeared [22], no three-dimensional structure of any inhibitor from this class has yet been determined.

Ascaris lives in the hostile environment of the gut, and has developed specific mechanisms to protect itself from the host digestive enzymes. The protease inhibitors seem not to be secreted, but rather are bound on the surface of the worm gut and other tissues, as well as on the surface of eggs and developing larvae, where they form complexes with host proteases [21]. The presence of inactivated proteases on the surface

*Corresponding authors. †Present address: Sterling-Winthrop, Malvern, PA 19426, USA.

of the eggs and larvae ensure that the migrating larvae are not perceived as foreign, thereby permitting them to evade the host's immune system as they migrate from the intestines to the liver [21]. In addition, the *Ascaris* protease inhibitors inhibit both clot lysis by plasmin and streptokinase-activated fibrinolysis of human plasma clots, suggesting that the presence of inhibitors on the larval surface may also modify blood homeostatis during larval migration [21,23]. In this paper we report the first three-dimensional structure determination of a member of the *Ascaris* family of protease inhibitors, namely the *Ascaris* trypsin inhibitor, by means of NMR spectroscopy.

Results and discussion

The converged structures

Complete NMR data sets on ATI were collected at both pH 2.4 and pH 4.75, and examples of the quality of the data were provided in our previous paper dealing with the resonance assignments and secondary structure determination [22]. While the structure calculations were based on data collected at 27°C, data recorded at other temperatures (16°C, 40°C and 50°C at pH 2.4; 30°C, 34°C and 40°C at pH 4.75) were used to resolve ambiguities. We found that the nuclear Overhauser effect (NOE) intensities and cross peak patterns were different at the two pH values, particularly around the reactive site (Table 1). Thus, there were 12 cross peaks that appeared only at the lower pH, and 3 that were present only at pH 4.75. In addition, there was a very large chemical shift difference in the resonance of Gly17(NH): namely, 8.27 ppm at pH 2.4 and 11.09 ppm at pH 4.75. Subsequent examination of the structures revealed that the unusually large downfield shift of the resonance of Gly17(NH) at pH 4.75 was due to the presence of a hydrogen bond between the side chain carboxylate of Glu39 and the backbone amide of Gly17, an interaction which was absent at pH 2.4.

As in previous structure determinations from our laboratory [24–27], an iterative strategy was employed in which successive structure calculations included an increasing number of restraints. This method makes use of the information in the initial low-resolution structures to resolve ambiguities in the assignment of NOE cross peaks and to permit addditional stereospecific assignments to be made. In this particular case, the iterative approach was critical in identifying the correct pairing of the five disulfide bridges which was previously unknown. Thus, all initial calculations were carried out without disulfide bridges and these were only introduced once the structures were sufficiently well defined to permit the unambiguous assignment of the disulfide pairings. In addition, the NOE data provided an unambigous means of identifying the correct disulfide pairings as, with the exception of the Cys18–Cys29 pair (see below), at least two NOE cross peaks were observed between the partners in each pair. Confirmation of the validity of the pairings was then afforded by observing that the addition of the five covalent S–S bonds and associated bond angles to the covalent geometry restraints was achieved without any perturbation of the global fold and only minimal atomic root mean square (rms) shifts. The disulfide pairings found were as follows: Cys5–Cys38, Cys15–Cys33, Cys18–Cys29, Cys22–Cys60 and Cys40–Cys54.

Table 1. Differences in observed NOEs between the pH 4.74 and pH 2.4 structures within the reactive site.

Residue	Atom	Residue	Atom	NOE intensity pH 4.75	NOE intensity pH 2.4
Pro28	$C^{\beta}H$	Cys29	NH	–	w
Cys29	NH	Thr30	NH	–	w
Cys29	$C^{\beta 2}H$	Thr30	NH	–	w
Cys29	$C^{\alpha}H$	Thr30	$C^{\gamma 2}H$	vw	w
Cys29	$C^{\alpha}H$	Arg31	NH	m	–
Cys29	$C^{\beta 1}H$	Arg31	NH	w	–
Cys29	$C^{\beta 2}H$	Arg31	NH	vw	–
Cys29	NH	Cys29	$C^{\alpha}H$	w	m
Cys29	NH	Cys29	$C^{\beta 1}H$	–	w
Thr30	$C^{\alpha}H$	Arg31	NH	w	s
Thr30	$C^{\beta}H$	Arg31	NH	–	w
Thr30	$C^{\gamma 2}H$	Glu32	NH	–	m
Thr30	$C^{\alpha}H$	Glu32	NH	–	w
Thr30	$C^{\gamma 2}H$	Glu32	$C^{\beta}H$	–	vw
Thr30	$C^{\gamma 2}H$	Glu32	$C^{\gamma}H$	–	vw
Arg31	$C^{\alpha}H$	Glu32	NH	m	s
Arg31	$C^{\beta}H$	Glu32	NH	m	w
Arg31	NH	Glu32	NH	–	m
Arg31	NH	Arg31	$C^{\alpha}H$	–	m
Glu32	$C^{\alpha}H$	Cys33	NH	m	s
Glu32	$C^{\gamma}H$	Cys33	NH	m	w
Glu32	NH	Cys33	NH	s	w
Glu32	NH	Glu32	$C^{\alpha}H$	–	m
Glu32	NH	Glu32	$C^{\gamma 2}H$	s	m

Abbreviations: s, strong; m, medium; w, weak; and vw, very weak.

The structures calculated from the pH 2.4 data were based on a total of 1083 experimental NMR restraints comprising 349 short range ($1 < |i-j| \leq 5$) and 216 long range ($|i-j| > 5$) inter-residue interproton distance restraints, 323 intra-residue interproton distance restraints, 46 distance restraints for 23 hydrogen bonds, and 59 ϕ, 49 ψ and 41 χ_1 torsion angle restraints. The structures calculated from the pH 4.75 data were based on a total of 1078 experimental NMR restraints comprising 343 short range ($1 < |i-j| \leq 5$) and 218 long range ($|i-j| > 5$) inter-residue interproton distance restraints, 320 intra-residue interproton distance restraints, 48 distance restraints for 24 hydrogen bonds, and 59 ϕ, 49 ψ and 41 χ_1 torsion angle restraints. At both pHs, stereospecific assignments were obtained for 25 of the 43 β-methylene protons and for the methyl groups of the two valine residues. The approximate interproton distance restraints were classified into four ranges, 1.8–2.7 Å, 1.8–3.3 Å, 1.8–5.0 Å and 1.8–6.0 Å, corresponding to strong, medium, weak and very weak

NOE intensities, respectively [11,28]. The minimum ranges employed for the ϕ, ψ and χ_1 torsion angle restraints, derived using a conformational grid search, were $\pm 30°$, $\pm 50°$ and $\pm 20°$, respectively [25].

A total of 32 structures were calculated using the hybrid distance geometry–dynamical simulated annealing method [29]. A summary of the structural statistics and atomic rms differences is provided in Tables 2 and 3, respectively, best fit superpositions of the backbone and selected side chains are shown in Figs 1 and 2, respectively, and plots of the atomic rms differences as a function of residue are given in Fig. 3. All the structures satisfy the experimental restraints within their errors

Table 2. Structural statistics.[a]

	pH 4.75		pH 2.4	
	$<SA_{high}>$	$(\overline{SA}_{high})_r$	$<SA_{low}>$	$(\overline{SA}_{low})_r$
Rms deviations from experimental distance restraints (Å)[b]				
All (929/934)	0.018 ± 0.001	0.019	0.019 ± 0.002	0.021
Short range ($1 < \|i - j\| \leq 5$) (343/349)	0.019 ± 0.001	0.022	0.022 ± 0.002	0.024
Long range ($\|i - j\| > 5$) (218/216)	0.022 ± 0.002	0.022	0.022 ± 0.003	0.023
Intra-residue (320/323)	0.010 ± 0.003	0.012	0.010 ± 0.002	0.010
Hydrogen bond (48/46)[c]	0.029 ± 0.003	0.027	0.032 ± 0.004	0.031
Rms deviations from experimental dihedral restraints (°) (149/149)[b]	0.108 ± 0.042	0.055	0.111 ± 0.048	0.106
F_{NOE} (kcal mol^{-1})[d]	9.0 ± 1.2	10.3	10.6 ± 1.7	11.3
F_{tor} (kcal mol^{-1})[d]	0.12 ± 0.10	0.03	0.13 ± 0.14	0.10
F_{repel} (kcal mol^{-1})[d]	12.9 ± 1.4	15.5	13.1 ± 1.8	20.5
E_{L-J} (kcal mol^{-1})[e]	−163 ± 8	−152	−158 ± 9	−146
Deviations from idealized covalent geometry[f]				
Bonds (Å) (925)	0.005 ± 0.000	0.005	0.005 ± 0.000	0.005
Angles (°) (1684)	1.759 ± 0.002	1.76	1.762 ± 0.003	2.851
Impropers (°) (343)	0.568 ± 0.010	0.580	0.571 ± 0.013	0.585

[a]The notation of the structures is as follows: $<SA_{high}>$ and $<SA_{low}>$ are the ensembles of 32 structures calculated at pH 4.75 and pH 2.4, respectively; $\overline{SA}_{high}$ and $\overline{SA}_{low}$ are the mean structures at pH 4.75 and pH 2.4, respectively, calculated by averaging the coordinates of the structures within the corresponding ensemble, best fitted to residues 5–60; $(\overline{SA}_{high})_r$ and $(\overline{SA}_{low})_r$ are the restrained regularized mean structures derived from $\overline{SA}_{high}$ and $\overline{SA}_{low}$, respectively [29]. The number of terms for the various restraints is given in parentheses with the values at pH 4.75 preceding those at pH 2.4. [b]None of the structures exhibited distance violations greater than 0.3 Å or dihedral angle violations greater than 3°. [c]For each hydrogen bond there are two restraints: r_{NH-O}, 1.7–2.3 Å; r_{N-O}, 2.4–3.3 Å. These restraints were only added in the final stages of the calculations. [d]The values of the square-well NOE and torsion angle potential energies (as per equations 2 and 3 in [28]) are calculated with force constants of 50 kcal mol^{-1} Å^{-2} and 200 kcal mol^{-1} rad^{-2}, respectively. The value of the quartic van der Waals repulsion energy (as per equation 5 in [29]) is calculated with a force constant of 4 kcal mol^{-1} Å^{-4} with the hard-sphere van der Waals radii set to 0.8 times the standard values used in the CHARMM empirical energy function [44]. [e]E_{L-J} is the Lennard–Jones van der Waals energy calculated with the CHARMM [44] empirical energy function. It is not included in the target function used during the simulated annealing. [f]The covalent geometry restraints include the bond and angle terms for the disulfide bonds which were only introduced in the final stages of the calculations. The improper torsion angles are used to maintain planarity and chirality. All peptide bonds are restrained to be planar and *trans*.

Table 3. Atomic root mean square differences (Å).[a]

	Residues 5–60			Residues 5–28 and 34–60			Residues 29–33	
	Backbone	All atoms	Ordered[b]	Backbone	All atoms	Ordered[b]	Backbone	All atoms
pH 4.75								
$<SA_{high}>$ *vs* $\overline{SA}_{high}$	0.52 ± 0.09	0.90 ± 0.09	0.60 ± 0.09	0.48 ± 0.09	0.83 ± 0.10	0.56 ± 0.09	0.25 ± 0.08	0.89 ± 0.20
$<SA_{high}>$ *vs* $(\overline{SA}_{high})_r$	0.54 ± 0.08	0.97 ± 0.10	0.63 ± 0.08	0.50 ± 0.08	0.91 ± 0.11	0.60 ± 0.08	0.26 ± 0.08	1.01 ± 0.28
$\overline{SA}_{high}$ *vs* $(\overline{SA}_{high})_r$	0.14	0.38	0.20	0.13	0.35	0.20	0.07	0.55
pH 2.4								
$<SA_{low}>$ *vs* $\overline{SA}_{low}$	0.50 ± 0.10	0.91 ± 0.09	0.58 ± 0.10	0.47 ± 0.10	0.84 ± 0.09	0.55 ± 0.10	0.38 ± 0.10	0.87 ± 0.16
$<SA_{low}>$ *vs* $(\overline{SA}_{low})_r$	0.54 ± 0.10	1.00 ± 0.09	0.63 ± 0.10	0.50 ± 0.10	0.93 ± 0.10	0.60 ± 0.09	0.42 ± 0.13	1.01 ± 0.18
$\overline{SA}_{low}$ *vs* $(\overline{SA}_{low})_r$	0.18	0.43	0.24	0.17	0.40	0.24	0.19	0.61
pH 4.75 *versus* pH 2.4								
$(\overline{SA}_{high})_r$ *vs* $(\overline{SA}_{low})_r$	0.96	2.08	1.05	0.55	0.74	0.60	1.84	4.71

[a]The notation is the same as in Table 2. [b]Atoms distal to C^β in residues Lys7, Lys14, Glu19, Arg31, Glu32, Glu39 and Glu58, and distal to C^γ in residues Glu10, Gln24, Lys25, Lys34, Arg37, Arg48, Gln51 and Lys56 are excluded.

(no violations >0.3 Å for the distance restraints and 3° for the torsion angle restraints), display very small deviations from covalent geometry, and have good nonbonded contacts as evidenced by the large negative values of the Lennard–Jones van der Waals energies (Table 1). In addition, all the ϕ, ψ angles lie within the allowed regions of the Ramachandran plot (Fig. 4).

The amino terminus (residues 1–4) and carboxyl terminus (residues 61–62) are poorly defined by the data and appear to be disordered in solution. The remainder of the molecule (residues 5–60) is well defined for both the pH 2.7 and pH 4.75 structures with a precision of ~0.5 Å for the backbone atoms, ~0.9 Å for all atoms, and ~0.6 Å for all atoms which do not exhibit conformational disorder. (The precision is defined as the average atomic rms difference between the structures in a given ensemble and their mean coordinate positions.)

The global fold

Two orthogonal views of a schematic ribbon drawing of the structure of ATI are shown in Fig. 5. ATI has a wedge-like shape with the reactive site located at the apex of the wedge. The approximate dimensions are as follows: the molecule is 25 Å wide and 26 Å high in the view shown in Fig. 5a, and 17.5 Å wide at the bottom of the wedge and 5 Å wide at the top of the wedge in the view shown in Fig. 5b. The overall fold of ATI is unusual and bears no resemblance to that of any member of the other serine protease inhibitor families. There are four short β-strands arranged in two mini antiparallel β-sheets oriented approximately per-

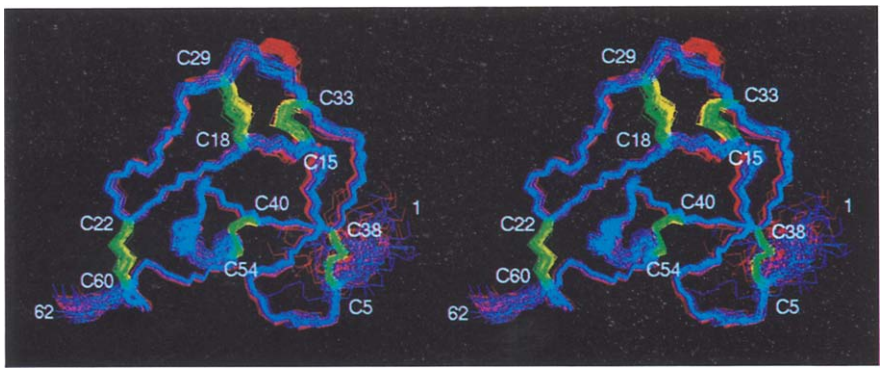

Fig. 1. Stereoview showing the best-fit (backbone residues 5–60) superposition of the backbone (N, C^{α}, C) atoms and cysteine side chains of the simulated annealing structures of ATI at pH 2.4 and pH 4.75. There are 32 structures for each pH value. The backbone and cysteine side chains are shown in red and yellow, respectively, for the pH 2.4 structures, and in blue and green, respectively, for the pH 4.75 structures.

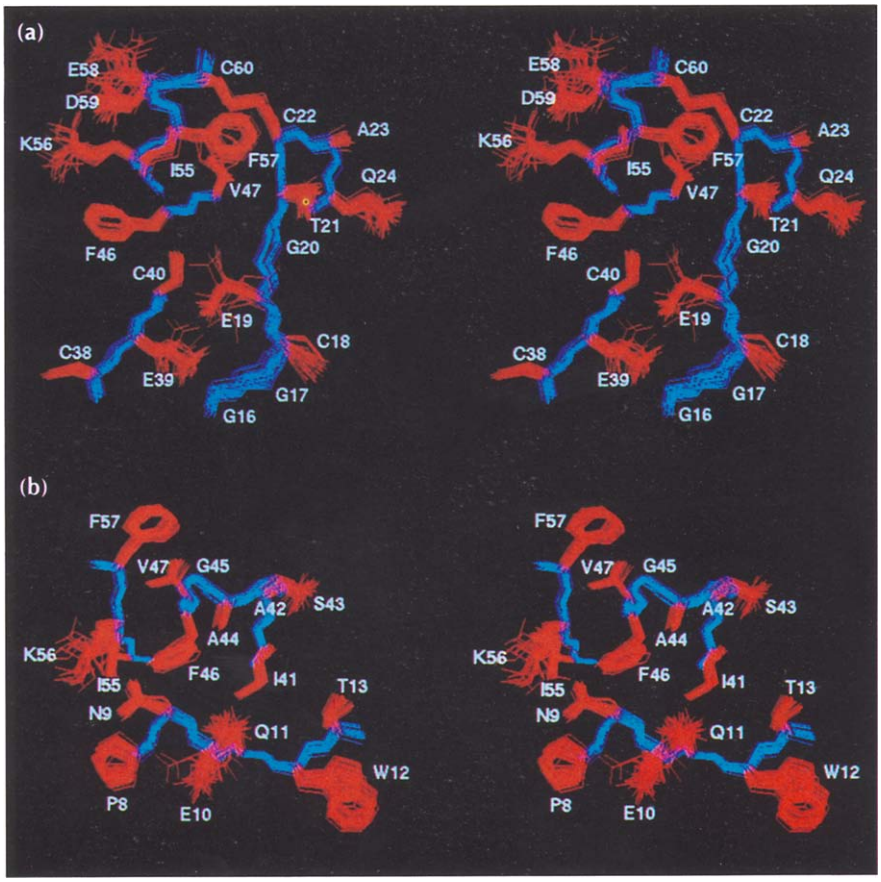

Fig. 2. Stereoviews showing a superposition of all atoms (except protons and backbone carbonyl oxygen atoms) of **(a)** residues 16–24, 38–40, 46–47 and 55–60 of the 32 simulated annealing structures of ATI at pH 2.4, and **(b)** residues 8–13, 41–47 and 55–57 of the 32 simulated annealing structures of ATI at pH 4.75. The backbone is shown in blue and the side chains in red.

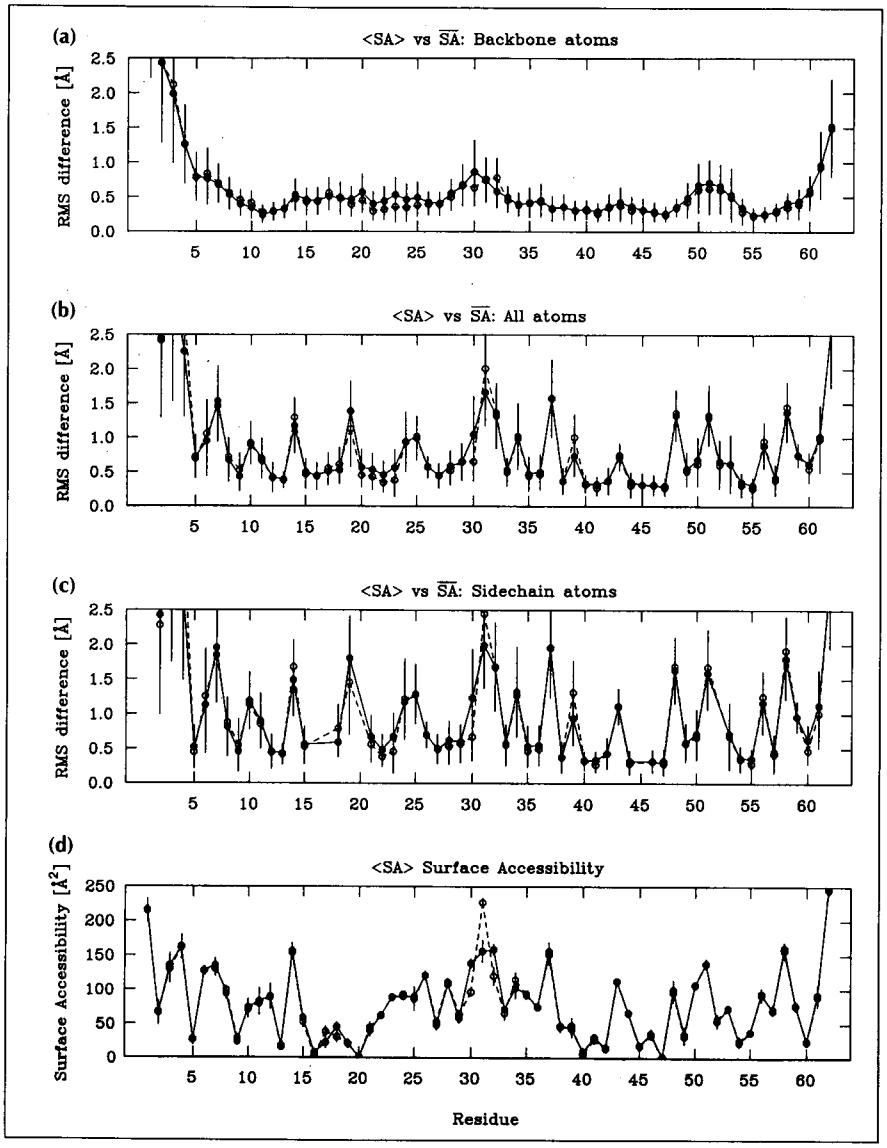

Fig. 3. Atomic rms distribution of the 32 individual structures at pH 2.4 (open circles) and pH 4.75 (closed circles) about their respective mean coordinate positions for **(a)** backbone atoms, **(b)** all atoms and **(c)** side chain atoms, best-fitted to the backbone atoms of residues 5–60. **(d)** Surface accessibility is shown as a function of residue number. The circles represent the mean values, and the bars the standard deviation.

pendicular to each other. Strand β1 (residues 11–14) is hydrogen bonded to strand β2 (residues 36–39), while strand β3 (residues 45–49) is hydrogen bonded to strand β4 (residues 52–57). Thus, strands 1 and 2 are connected by a long loop (residues 15–35) which includes the reactive site (residues 29–33). Strands β3 and β4 are flanked by two double turns. The turn preceding strand β3 includes two hydrogen bonds from the backbone amides of Ala45 and Gly45 to the backbone carbonyls of Ile41 and Ala42, respectively. The turn following strand β4 has hydrogen bonds from the backbone amides of Asp59 and Cys60 to the backbone carbonyls of Lys56 and Phe57, respectively. The turn connecting strands β3 and β4 includes both β-turn and β-bulge-type hydrogen bonds from the backbone carbonyl of Asp49 to the backbone amides of both Gly52 and Asn53. In addition, there is a hydrogen bond between the side chain amide of Asn53 and the backbone carbonyl of Gln51. A fourth turn, a type I β-turn, is found between residues 20–24 and it is stabilized by backbone hydrogen bonds from Gln24(NH) to Thr21(O) and from Thr21(NH) to Gln24(O).

The overall structure is dominated by the five disulfide bridges which clearly constitute the major stabilizing influence on the fold. The disulfide bridge between Cys5 and Cys39 orients the amino terminus with respect to strand β1. The three disulfide bridges, Cys15–Cys33, Cys18–Cys29 and Cys22–Cys60 stabilize the long loop connecting strands β1 and β2. The Cys15–Cys33 and Cys18–Cys29 disulfide bridges are located on each side of the reactive site, while the Cys22–Cys60 bridge orients the reactive-site loop with respect to the carboxyl terminus of the protein. Finally, the disulfide bridge between Cys40 and Cys54 serves to orient the two antiparallel β-sheets with respect to one other.

As stated above, the boundaries of the reactive site are formed by two disulfide bridges. Both are left-handed. The disulfide bridge between Cys15 and Cys33 is characterized by a short C^{α}–C^{α} distance (~4.7 Å) and χ_1,

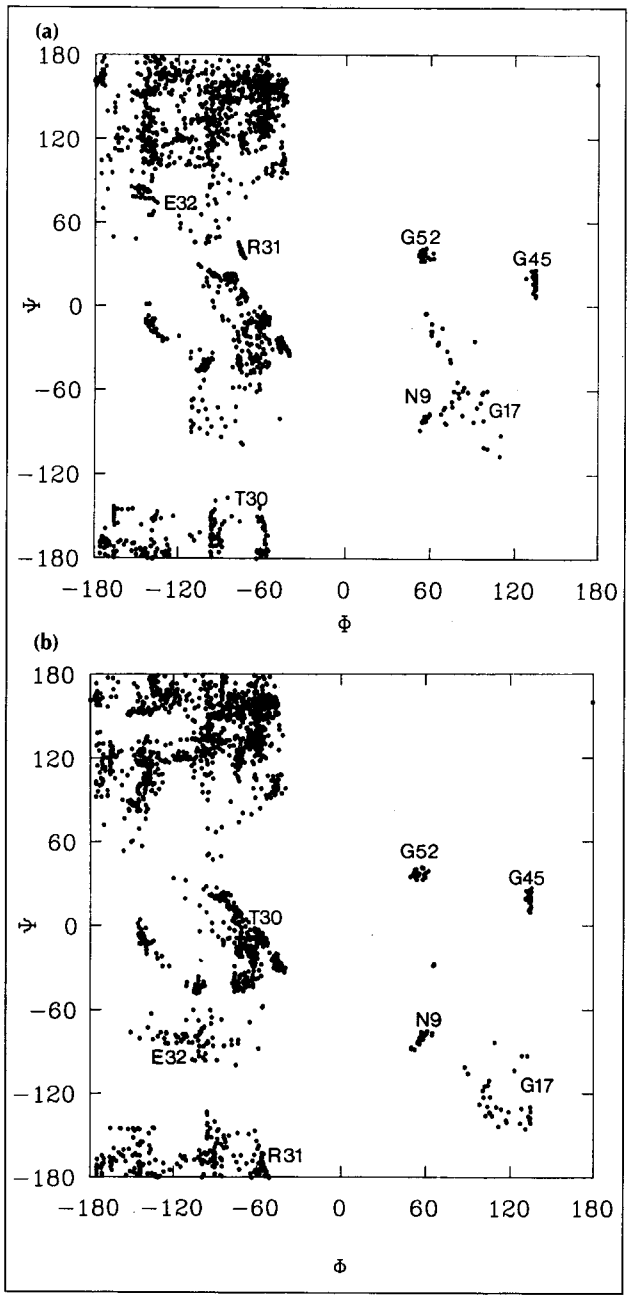

Fig. 4. Ramachandran ϕ, ψ plots of the 32 simulated annealing structures of ATI at **(a)** pH 2.4 and **(b)** pH 4.75. Note the large differences in the ψ angles of Thr30, Arg31 and Glu32 (which constitute the P_2, P_1 and P_1' residues of the reactive site) between the pH 2.4 and pH 4.75 structures.

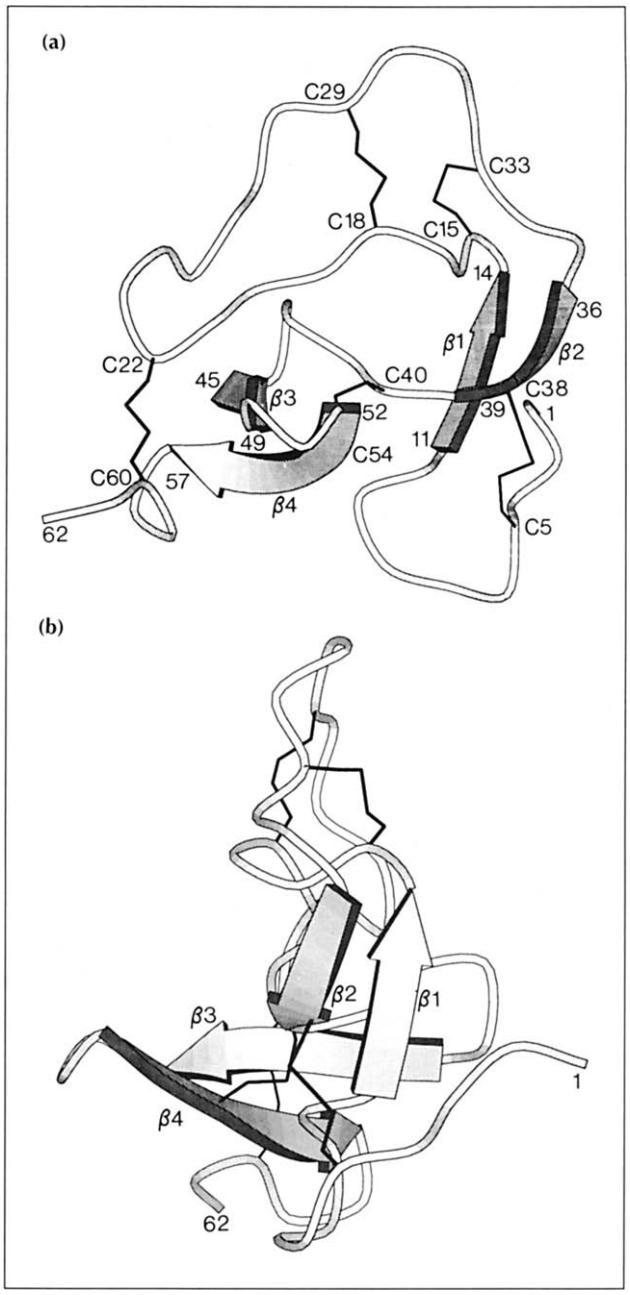

Fig. 5. Ribbon diagrams of the restrained minimized mean structure of ATI at pH 2.4 shown in two approximately orthogonal views in **(a)** and **(b)**. Arrows indicate the region of β-sheet (residues 11–13, 37–39, 45–49, and 53–57). The disulfide bridges are indicated by the solid lines connecting the labeled cysteine C^α atoms. The reactive site in (a) is located in the upper right hand corner between Cys29 and Cys33, and the orientation in (a) is approximately the same as that shown in Fig. 1. (Figures generated with the program MOLSCRIPT [46].)

χ_2, χ_3, χ_2' and χ_1' torsion angles of 150°, 110°, −60°, −100° and −80°, respectively. The disulfide bridge between Cys18 and Cys29, on the other hand, is characterized by a long C^α–C^α distance of ~6.8 Å and a conformation reminiscent of the disulfide bridges that span β-barrels in immunoglobulins [30]. (The χ_1, χ_2, χ_3, χ_2' and χ_1' torsion angles for the Cys18–Cys29 disulfide have values of −60°, −130°, −90°, −150° and −80°, respectively, for the pH 2.4 structure, and values of −40°, −180°, −110°, −160° and −80°, respectively, for the pH 4.75 structure.) It is worth noting that the alternative disulfide pairing of Cys15 with Cys29 and Cys18 with Cys33 is not feasible as the C^α–C^α separation between Cys15 and Cys29 is ~11 Å in the pH 2.4 structure and ~13 Å in the pH 4.75 structure.

Other significant interactions include the sequestration of hydrophobic side chains on either side of the β-sheet formed by strands β3 and β4. On the one side, Val47 is flanked by Phe57, Ile55 and Thr21 (Fig. 2a).

Similar hydrophobic interactions exist between Phe46, Ala44, and Ile41 on the opposite side of the sheet (Fig. 2b). In addition, there are a number of noteworthy side chain–backbone hydrogen bonds: the hydrogen bond between Asn9(O^{δ}) and Lys56(NH) orients strand β1 relative to strand β4 and accounts for the positive φ angle of Asn9 (Fig. 4); the hydrogen bond between Thr13(O^{γ}) and Cys15(NH) stabilizes the turn following strand β1; and the two hydrogen bonds between Glu19(O^{ε}) and Ala42(NH) and between Glu39(O^{ε}) and Gly17(NH) stabilize the interaction between the loop following strand β1 and the loop connecting strands β2 and β3. Of the latter two, only the former is present in the pH 2.4 structure.

Comparison of the pH 2.4 and pH 4.75 structures

A superposition of the restrained minimized mean structures at pH 2.4 and pH 4.75 is shown in Fig. 6. The fold is identical with the exception of the conformation at the reactive site. The overall backbone atomic rms difference for residues 5–60 is 0.96 Å which is significantly higher than the average backbone atomic rms difference between the structures in each ensemble and their respective means which is only ~0.5 Å (Table 3). However, when the reactive site (residues 29–33) is excluded from this comparison, there is no difference between the structures at the two pH values, and the backbone atomic rms difference between the two restrained minimized mean structures (0.55 Å) is comparable with the precision of the two sets of structures (Table 3). Thus, the two ensembles of structures are superimposable for residues 5–28 and 34–60 and there is no difference between them within the error of the coordinates. On the other hand, the backbone rms difference for the reactive site between the two structures, is large (1.84 Å).

The difference in the reactive site conformation between the pH 2.4 and pH 4.75 structures is characterized by large changes (140–160°) in the backbone ψ angles of the P_2, P_1 and P_1' residues (residues 30–32) (Table 4). The resulting difference in the disposition of the reactive site side chains is shown in Fig. 6b. The most critical difference is the orientation of the P_1 and P_1' side chains flanking the scissile peptide bond, as these determine the specificity of the inhibitor for its cognate enzyme. It can be seen that the Arg31 and Glu32 side chains 'swap' positions at the different pHs. At pH 4.75, Glu32 points away from the body of the molecule, while Arg31 is closer to the center. At pH 2.4, the carboxylate side chain of Glu32 comes close to packing against the hydroxyl group of Thr30, and the Arg31 side chain points away from the core of the molecule. The orientation of the Arg31 side chain in the low pH conformation should permit the interaction of the positively charged guanidinium group with Asp189 in the 'specificity pocket' of the protease, trypsin. It is possible that the carboxylate group of Glu32 is protonated at pH 2.4, and the absence of charge may decrease its solvation.

Comparison with the reactive sites of other protease inhibitors

The differences in the reactive site conformation of the structures of ATI at low and high pH are especially interesting when compared with other protease inhibitors. It has previously been noted that the structures of serine protease inhibitors from different fami-

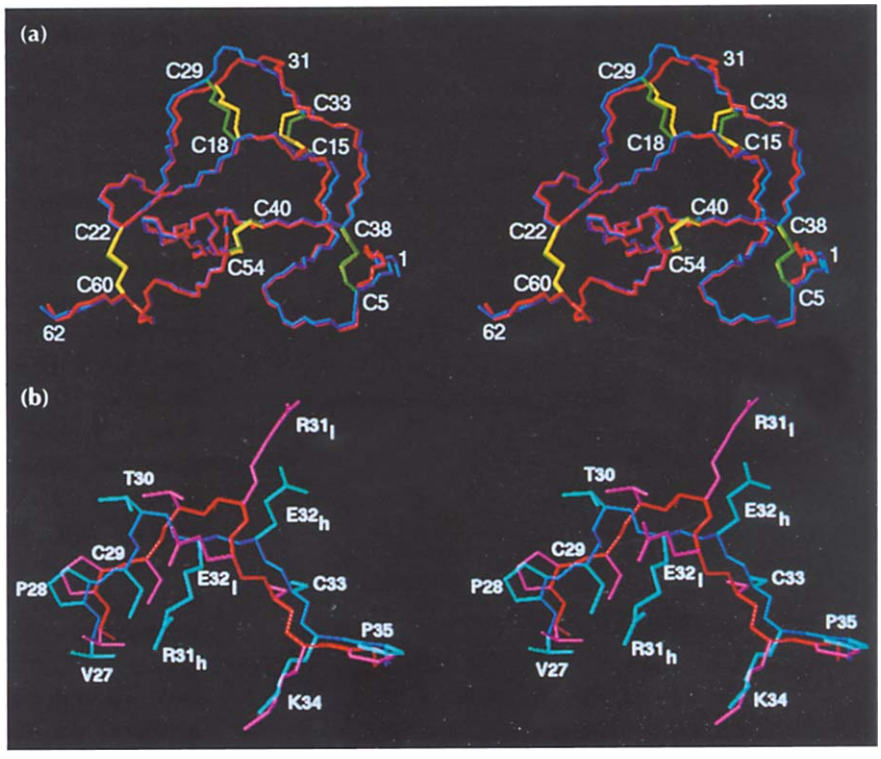

Fig. 6. Best-fit superposition of the restrained minimized mean structures of ATI at pH 2.4 and pH 4.75. **(a)** Backbone (N, C^{α}, C) atoms and cysteine side chains. (The backbone and side chains are shown in red and yellow, respectively, for the pH 2.4 structure, and in blue and green, respectively, for the pH 4.75 structure.) **(b)** Reactive site (residues 27–35) showing all atoms except protons. (The backbone and side chains are shown in red and pink, respectively, for the pH 2.4 structure, and in blue and light blue, respectively, for the pH 4.75 structure.) The subscripts h and l indicate the side chains of Arg31 and Glu32 at high pH (pH 4.75) and low pH (pH 2.4), respectively. Note that the positions of these two side chains are dramatically different at low and high pH.

Table 4. Backbone ϕ and ψ angles (°) in protease inhibitor active sites.[a]

Inhibitor[b]	P_3 (29)		P_2 (30)		P_1 (31)		P_1' (32)		P_2' (33)	
	ϕ	ψ	ϕ	ψ	ϕ	ψ	ϕ	ψ	ϕ	ψ
$<SA_{high}>$	−160 ± 9	119 ± 9	−68 ± 13	−10 ± 9	−63 ± 9	−174 ± 17	−117 ± 15	−77 ± 5	−170 ± 4	114 ± 15
$<SA_{low}>$	−156 ± 10	143 ± 28	−85 ± 33	−154 ± 5	−74 ± 1	40 ± 3	−108 ± 10	87 ± 32	−122 ± 29	121 ± 16
BPTI	−84 ± 2	−7 ± 3	−86 ± 4	159 ± 4	−110 ± 6	21 ± 9	−77 ± 1	171 ± 2	−129 ± 1	81 ± 4
2OVO	−131	155	−87	173	−96	9	−58	139	−99	93
2SSI	−130	147	−72	145	−92	89	−118	167	−116	90
1CSE	−139	168	−62	143	−115	45	−97	168	−117	110

[a]P_x is the nomenclature of Schechter and Berger [45] for the residues surrounding the P_1–P_1' scissile bond. The numbers in parentheses correpond to the residues in ATI. [b]$<SA_{high}>$ and $<SA_{low}>$ are the averages for the ensemble of structures calculated for ATI at pH 4.75 and pH 2.4, respectively. The values listed under bovine pancreatic trypsin inhibitor (BPTI) are the averages of crystal forms I (4PTI; [5]), II (5PTI; [9]) and III (6PTI; [10]). 2OVO is ovomucoid third domain of silver pheasant [7]; 2SSI is *Streptomyces* subtilisin inhibitor [6]; and 1CSE is eglin-c complexed with subtilisin Carlsberg [8].

lies are quite different, with the notable exception of the conformation of the reactive site, which is preserved throughout [8]. As shown in Table 4, the ϕ and ψ angles of various serine protease inhibitors are very similar when aligned about the P_1–P_1' scissile bond. While the reactive site of the pH 2.4 ATI structure shares the same pattern of ϕ, ψ angles as the other inhibitors, the ψ angles of the reactive site of the pH 4.75 structure differ significantly at the P_2, P_1 and P_1' positions. The similarity of the ϕ and ψ angles is also reflected in the better fit of the low pH backbone C^α carbons to protease inhibitor reactive sites, as shown in Table 5. Thus, the C^α backbone of the reactive site of the pH 2.4 ATI structure lies within the ensemble of reactive sites found for other serine protease inhibitors, whereas that of the pH 4.75 structure lies well outside this ensemble.

Table 5. Backbone C^α atomic rms differences (Å) among the reactive sites (residues P_2–P_2') of different serine protease inhibitors.[a]

	1CSE	2SSI	2OVO	4PTI	$(\overline{SA_{low}})_r$
$(\overline{SA_{high}})_r$	1.61	1.15	1.27	1.45	1.74
$(\overline{SA_{low}})_r$	0.41	0.88	0.59	0.52	
4PTI	0.21	0.47	0.22		
2OVO	0.40	0.45			
2SSI	0.55				

[a]The correspondence of PDB accession codes to serine protease inhibitors is given in Table 4.

Biological implications

It is estimated that 25% of the world's population, including 4 million Americans, are infected with *Ascaris lumbricoides*, with prevalence rates of 80–100% in tropical and less developed countries [31]. While disease caused by *Ascaris* is infrequent, it generally correlates with the intensity of infection. Consequently, clinical manifestations are generally found in the developing world. Symptomatic cases fall into two categories: pulmonary disease associated with the migration of larvae in the small vessels of the lung, and intestinal disease arising from obstruction due either to the presence of a large number of parasites in the small intestine or to the migration of adult worms to the biliary tree or pancreatic ducts.

The trypsin and chymotrypsin/elastase inhibitors of *Ascaris* play an important role in the life cycle of the parasite and belong to a new family of serine protease inhibitors which is characterized by the presence of five disulfide bridges in a single small protein domain of 61–62 residues. Two half cysteines are located at the P_2 and P_3' positions in the reactive site. In this paper we present the first three-dimensional structure determination of a member of this family of protease inhibitors, namely the *Ascaris* trypsin inhibitor (ATI) using NMR spectroscopy. The three-dimensional structure bears no resemblance to that of other serine protease inhibitors with the exception of the reactive site (residues 29–33). At pH 2.4, the reactive site conformation is very similar to that of other protease inhibitors. At pH 4.75, however, the reactive site undergoes a conformational transition involving changes of 140–160° in the ψ backbone torsion angles of the P_2, P_1 and P_1' residues (residues 30–32). This results in a change in the positioning of the side chains of Arg31 and Glu32. At low pH, the side chain carboxylate of Glu32 packs against the hydroxyl group of Thr30, and the Arg31 side chain points into the solvent, thereby making a positive charge available for interaction with Asp189 in the specifity pocket of trypsin. At high pH, on the other

hand, Glu32 points away from the molecule, while Arg31 points towards the main body of the protein.

The low pH conformation is important for the developing *Ascaris*. The initial infection occurs by ingestion of the embryo. The larvae pass through the intestine to the liver, and then to the lungs where they are swept up the trachea to the mouth and ingested a second time. Thus, the developing worm is exposed twice to the low pH environment of the stomach and its attendant proteases. Physiological studies have shown that ATI and the other *Ascaris* protease inhibitors are bound to the surface of the worm gut and other tissues and to the surface of the developing eggs and larvae where they form complexes with the host proteases [21]. This not only inactivates the host proteases (thereby protecting the worm, eggs and larvae from digestion) but also ensures that the migrating larvae are not perceived as foreign. This permits them to evade the host immune system as they migrate from the intestines to the liver. The finding that the conformation of ATI at pH 2.4 is very similar to that of other serine protease inhibitors supports the physiological studies and strongly suggests that the active from of ATI is the low pH form. The origin for the conformational transition in the reactive site at higher pHs is unknown. One possible explanation is that at low pH (i.e. pH 2.4) the carboxylate side chain of Glu32 is protonated. The resulting absence of a negative charge which would be present at higher pH values (pH 4.75), may decrease its solvation.

Materials and methods

Sample preparation

ATI was prepared from *Ascaris lumbricoides* var *suum* collected from the intestinal contents of hogs [32] and prepared as described previously [22]. Samples for NMR spectroscopy contained approximately 3.3 mM ATI in either 90 % H_2O/10 % D_2O or 99.9 % D_2O. Spectra were recorded using samples at pH 2.4 (16°C, 27°C, 40°C and 50°C) and pH 4.75 (27°C, 30°C, 34°C and 40°C).

NMR spectroscopy

All spectra were recorded on a Bruker AM-600 spectrometer. Two-dimensional NMR experiments were recorded in the pure-phase absorption mode using the time-proportional incrementation method [33]. Complete assignment of the ^{1}H-NMR spectrum of ATI at pH 2.4 has been reported [22]. Sequential assignment of the ^{1}H-NMR spectrum at pH 4.5 was carried out as described in [22] using standard procedures [34,35]. Examples of the quality of the NMR spectra are provided in [22]. Intra-residue and sequential interproton distance restraints were derived from 50 ms nuclear Overhauser effect (NOESY) [36] spectra, while all other interproton distance restraints were derived from 150 ms mixing time NOESY spectra. $^3J_{HN\alpha}$ and $^3J_{\alpha\beta}$ coupling constants were measured from H_2O primitive correlated (P.COSY) [37] and D_2O primitive exclusive correlated (PE.COSY) [38] spectra, respectively.

Structure calculations

All interproton distances were classified into one of four ranges: 1.8–2.7 Å, 1.8–3.3 Å, 1.8–5.0 Å, or 1.8–6.0 Å, corresponding to strong, medium, weak, and very weak NOE intensities, respectively [11,28]. Upper limits for the distance constraints involving methyl protons and non-stereospecifically assigned methylene protons were corrected appropriately for center averaging [39]. In addition, the apparent higher intensity of methyl protons was corrected by the addition of 0.5 Å to the upper distance bound [12,13].

Stereospecific assignment of β-methylene protons and ϕ, ψ and χ torsional angle restraints were obtained on the basis of $^3J_{HN\alpha}$ and $^3J_{\alpha\beta}$ coupling constants and the intra-residue and sequential NOEs involving the NH, $C^\alpha H$, and $C^\beta H$ protons using the conformational grid search program STEREOSEARCH [40]. Stereospecific assignments of the methyl groups of valine were obtained as described in [41].

Three-dimensional structures were calculated on the basis of the experimental NMR restraints using the hybrid distance–geometry simulated annealing method of Nilges *et al.* [29] with minor modifications [26] using the program X-PLOR [42,43]. The target function minimized by the simulated annealing protocol comprises square-well quadratic potentials for the experimental distance and torsion angle restraints, quadratic harmonic potentials for the covalent geometry, and a quartic van der Waals repulsion term to prevent atoms from coming too close together. No Lennard–Jones van der Waals, electrostatic or hydrogen bonding potentials were included in the target function.

The coordinates for the 32 simulated annealing structures of the low and high pH forms, as well as the corresponding restrained minimized mean structures, and the complete set of experimental NMR restraints have been deposited in the Brookhaven Protein Data Bank.

Note added in proof

The structure of the *Ascaris* chymotrypsin/elastase inhibitor complexed with porcine elastase has recently been solved independently by X-ray crystallography (K Huang, NCJ Strynadka, VD Bernard, RJ Peanasky, MNG James, *Structure* 1994, **2**:679–689). The fold and disulfide pairings are identical to those of the *Ascaris* trypsin inhibitor reported in this paper.

Acknowledgements: This work was supported by the AIDS Targeted Anti-Viral Program of the Office of the Director of the National Institutes of Health (GMC and AMG). We thank Professor Robert Peanasky for providing us with a sample of ATI.

References

1. Hörl, H. & Heidland, A. (1982). Proteases: potential role in health and disease. In *Advances in Experimental Medicine and Biology*. Vol. 167, Plenum Press, New York.
2. Schnebli, H.P. & Braun, N.J. (1986). Proteinase inhibitors as Drugs. In *Proteinase Inhibitors*. (Barrett, A.J. & Salvesen, A., eds), pp. 613–627, Elsevier, The Netherlands.
3. Laskowski, M. & Kato, I. (1980). Protein inhibitors of proteinases. *Annu. Rev. Biochem.* 49, 593–626.
4. Read, R.J. & James, M.N.G. (1986). Introduction of the protein inhibitors: X-ray crytallography. In *Proteinase Inhibitors*. (Barrett, A.J. & Salvesen, A., eds), pp. 301–336, Elsevier, The Netherlands.
5. Deisenhofer, J. & Steigemann, W. (1975). Crystallographic refinement of the structure of bovine pancreatic trypsin inhibitor at 1.5 Å resolution. *Acta Crystallogr. B* **31**, 238–250.

6. Mitsui, Y., Satow, Y., Wanatabe, Y. & Iitaka, Y. (1979). Crystal structure of a bacterial protein proteinase inhibitor (*Streptomyces* subtilisin inhibitor) at 2.6 Å resolution. *J. Mol. Biol.* **131**, 697–724.
7. Bode, W., Epp, O., Huber, R., Laskowski, M. & Ardelt, W. (1985). The crystal and molecular structure of the third domain of silver pheasant ovomucoid (OMSVP3). *Eur. J. Biochem.* **147**, 387–395.
8. Bode, W., Papamokos, E. & Musil, D. (1987). The high resolution X-ray crystal structure of the complex formed between subtilisin Carlsberg and eglin c, an elastase inhibitor from the leech *Hirudo medicinalis*. *Eur. J. Biochem.* **166**, 673–692.
9. Wlodawer, A., Walter, J., Huber, R. & Sjolin, L. (1984). Structure of bovine pancreatic trypsin inhibitor: results of joint neutron and X-ray refinement of crystal form II. *J. Mol. Biol.* **180**, 301–329.
10. Wlodawer, A., Nachman, J., Gilliland, G.L, Gallagher, W. & Woodward, C. (1987). Structure of form III crystals of bovine pancreatic trypsin inhibitor. *J. Mol. Biol.* **198**, 469–480.
11. Williamson, M.P., Havel, T.F. & Wüthrich, K. (1985). Solution conformation of proteinase inhibitor IIA from bull seminal plasma by ^{1}H nuclear magnetic resonance and distance geometry. *J. Mol. Biol.* **182**, 295–315.
12. Wagner, G., Braun W., Havel, T.F., Schaumann, T., Go, N. & Wüthrich, K. (1987). Protein structures in solution by nuclear magnetic resonance and distance geometry: the polypeptide fold of basic pancreatic trypsin inhibitor determined using two different algorithms, DISGEO and DISMAN. *J. Mol. Biol.* **196**, 611–639.
13. Clore, G.M., Gronenborn, A.M., Nilges, M. & Ryan, C.A. (1987). Three-dimensional structure of potato carboxypeptidase inhibitor in solution: a study using nuclear magnetic resonance, distance geometry and restrained molecular dynamics. *Biochemistry* **26**, 8012–8023.
14. Clore, G.M., Gronenborn, A.M., Kjaer, M. & Poulsen, J.M. (1987). The determination of the three-dimensional structure of barley serine proteinase inhibitor 2 by nuclear magnetic resonance, distance geometry and restrained molecular dynamics. *Protein Eng.* **1**, 313–318.
15. Folkers, P.J.M., Clore, G.M., Driscoll, P.C., Dodt, J., Kohler, S. & Gronenborn, A.M. (1989). The solution structure of recombinant hirudin and the Lys47→Glu mutant: a nuclear magnetic resonance and hybrid distance geometry–dynamical simulated annealing study. *Biochemistry* **28**, 2601–2617.
16. Holak, T.A., Gondol, D., Otlewski, J. & Wilusz, T. (1989). Determination of the complete three-dimensional structure of the trypsin inhibitor from squash seeds in aqueous solution by nuclear magnetic resonance and a combination of distance geometry and dynamical simulated annealing. *J. Mol. Biol.* **210**, 635–648.
17. Mitsui, Y, Satow, Y., Wanatabe, Y. Hirono, S. & Iitaka, Y. (1979). Crystal structures of *Streptomyces* subtilisin inhibitor and its complex with subtilisn BPN'. *Nature* **277**, 447–452.
18. Peanasky, R.J., Bentz, Y., Paulson, B., Graham, D.L. & Babin, D.R. (1984). The isoinhibitors of chymotrypsin/elastase from *Ascaris lumbricoides*: isolation by affinity chromatogrpahy and association with enzymes. *Arch. Biochem. Biophys.* **232**, 127–134.
19. Peanasky, R.J., Bentz, Y., Homandberg, G.A., Minor, S.T. & Babin, D.R. (1984). The isoinhibitors of chymotrypsin/elastase from *Ascaris lumbricoides*: the reactive site. *Arch. Biochem. Biophys.* **232**, 135–142.
20. Babin, D.R., Peanasky R.J. & Goos, S.M. (1984). The isoinhibitors of chymotrypsin/elasatase from *Ascaris lumbricoides*: the primary structure. *Arch. Biochem. Biophys.* **232**, 143–161.
21. Peanasky, R.J., Martzen, M.R., Homandberg, G.A., Cash, J.M., Babin, D.R. & Litweiler, B. (1987). Proteinase inhibitors from intestinal parasitic helminths: structure and indications of some possible functions. In *Paradigms for Eradicating Helminthic Parasites*. (Macinnis, A.J., ed), pp. 349–366, Alan R Liss, New York.
22. Gronenborn, A.M., Nilges, M., Peanasky, R.J. & Clore, G.M. (1990). Sequential resonance assignment and secondary structure determination of the *Ascaris* trypsin inhibitor, a member of a novel class of proteinase inhibitors. *Biochemistry* **29**, 183–189.
23. Crawford, G.P.M., Howse, D.J. & Grove, D.I. (1982). Inhibition of human blood clotting by extracts of *Ascaris suum*. *J. Parasitol.* **68**, 1044–1047.
24. Clore, G.M. & Gronenborn, A.M. (1991). Structures of larger proteins in solution: three- and four-dimensional heteronuclear NMR spectroscopy. *Science* **252**, 1390–1399.
25. Kraulis, P.J., *et al.*, & Gronenborn, A.M. (1989). Determination of the three-dimensional solution structure of the C-terminal domain of cellobiohydrolase I from *Thrichoderma reesei*: a study using nuclear magnetic resonance and hybrid distance geometry–dynamical simulated annealing. *Biochemistry* **28**, 7241–7257.
26. Forman-Kay, J.D., Clore, G.M., Wingfield, P.T. & Gronenborn, A.M. (1991). High-resolution three-dimensional structure of reduced recombinant human thioredoxin in solution. *Biochemistry* **30**, 2685–2698.
27. Clore, G.M., Wingfield, P.T. & Gronenborn, A.M. (1991). High-resolution three-dimensional structure of interleukin-1 in solution by three- and four-dimensional nuclear magnetic resonance spectroscopy. *Biochemistry* **30**, 2315–2323.
28. Clore, G.M., Nilges, M., Sukumaran, D.K., Brünger, A.T., Karplus, M. & Gronenborn, A.M. (1986). The three-dimensional structure of α1-purothionin in solution: combined use of nuclear magnetic resonance, distance geometry and restrained molecular dynamics. *EMBO J.* **5**, 2729–2735.
29. Nilges, M., Gronenborn, A.M., & Clore G.M. (1988). Determination of three-dimensional structures of proteins from interproton distance data by hybrid distance geometry–dynamical simulated annealing calculations. *FEBS Lett.* **229**, 129–136.
30. Richardson, J.S. (1981). The anatomy and taxonomy of protein structure. *Adv. Protein Chem.* **34**, 167–339.
31. Harrison's Principles of Internal Medicine (1991) (Wilson, J.D., *et al.*, & Root, RK., eds), Vol. 1, pp. 818–819, McGraw Hill, London.
32. Goodman, R.B. & Peanasky, R.J. (1982). Isolation of the trypsin inhibitors in *Ascaris lumbricoides* var *suum* using affinity chromatography. *Anal. Biochem.* **120**, 387–393.
33. Marion, D. & Wüthrich, K. (1983). Application of phase sensitive two-dimensional correlated spectroscopy (COSY) for measurements of ^{1}H–^{1}H spin–spin coupling constants in proteins. *Biochem. Biophys. Res. Commun.* **113**, 967–974.
34. Wüthrich, K. (1986). *NMR of Proteins and Nucleic Acids*. Wiley, New York.
35. Clore, G.M. & Gronenborn, A.M. (1989). Determination of three-dimensional structures of proteins in solution by nuclear magnetic resonance spectroscopy. *Protein Eng.* **1**, 275–288.
36. Ernst, R.R., Bodenhausen, G. & Wokaun, A. (1987). *Principles of Nuclear Magnetic Resonance in One and Two Dimensions*. Clarendon Press, Oxford.
37. Marion, D. & Bax, A. (1988). P.COSY: a sensitive alternative for double quantum filtered COSY. *J. Magn. Reson.* **80**, 528–533.
38. Mueller, L. (1987). PE.COSY: a simple alternative to E.COSY. *J. Magn. Reson.* **72**, 191–196.
39. Wüthrich, K. Billeter, M. & Braun, W. (1983). Pseudo-structures for the 20 common amino acids for use in studies of protein conformations by measurements of intramolecular proton–proton distance constraints with nuclear magnetic resonance. *J. Mol. Biol.* **169**, 949–961.
40. Nilges, M., Clore, G.M. & Gronenborn, A.M. (1990). ^{1}H-NMR stereospecific assignments by conformational data-base searches. *Biopolymers* **29**, 813–822.
41. Zuiderweg, E.R.P., Boelens, R. & Kaptein, R. (1985). Stereospecific assignments of ^{1}H-NMR methyl lines and conformation of valyl residues in the lac repressor headpiece. *Biopolymers* **24**, 601–611.
42. Brünger, A.T., Clore, G.M., Gronenborn, A.M. & Karplus, M. (1986). Three-dimensional structure of proteins determined by molecular dynamics with interproton distance restraints: application to crambin. *Proc. Natl. Acad. Sci. USA* **83**, 3801–3805.
43. Brünger, A.T. (1992). *X-PLOR Version 3.1 Manual*. Yale University, New Haven, CT.
44. Brooks, B.R., Bruccoleri, T.E., Olafson, B.D., States, D.J., Swaminathan, S. & Karplus, M. (1983). CHARMM: a program for macromolecular energy minimization and dynamics calculations. *J. Comput. Chem.* **4**, 187–217.
45. Schechter, I. & Berger, A. (1967). On the size of the active site in proteases. I. Papain. *Biochem. Biophys. Res. Commun.* **27**, 157–162.
46. Kraulis, P.J. (1991). MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallogr.* **24**, 946–950.

Received: **25 Apr 1994**; revisions requested: **12 May 1994**; revisions received: **16 May 1994**. Accepted: **17 May 1994**.