Crystal Structure of Cleaved Bovine Antithrombin III at 3×2 Å Resolution

Lionel Mourey, Jean-Pierre Samama, Marc Delarue, Maurice Petitou, Jean Choay, Dino Moras

J. Mol. Biol. (1993) 232, 223–241

Crystal Structure of Cleaved Bovine Antithrombin III at 3·2 Å Resolution

Lionel Mourey[1]†, Jean-Pierre Samama[1]†, Marc Delarue[1], Maurice Petitou[2] Jean Choay[2] and Dino Moras[1]

[1]*UPR de Biologie Structurale*
Institut de Biologie Moléculaire et Cellulaire du CNRS
15 rue R. Descartes, 67084 Strasbourg Cedex, France

[2]*Sanofi Recherche-Centre Choay*
9, rue du Président Allende, 94256 Gentilly Cedex, France

(Received 1 October 1992; accepted 2 March 1993)

The crystal structure of cleaved antithrombin III (ATIII) has been determined to 3·2 Å resolution by single isomorphous replacement, real space density modification and phase extension protocols. The heavy-atom sites and the first molecular envelope were determined owing to the molecular replacement solution previously reported and partially refined. Refinement of the two molecules of the asymmetric unit led to a crystallographic R-factor of 0·212 for all reflections between 8·0 and 3·2 Å, without inclusion of water molecules. The root-mean-square deviation from ideal values is, respectively, 0·015 Å and 3·6° for bond lengths and bond angles. The topology of the molecule closely resembles that of cleaved serpins inhibitors with the two residues forming the reactive bond at opposite ends of the molecule. The most significant difference between ATIII and α_1-antitrypsin lies in the 45 residue N-terminal extension in ATIII which contribute to the definition of the heparin binding site. This loop region at the surface of the molecule is held by two disulphide bridges to the protein core and exhibits high temperature factor values. It forms a valley which restrains the possibilities for binding of heparin. Docking of the pentasaccharide unit which represents the minimum fragment of heparin able to bind to ATIII indicates a possible role for arginine 14 in the interaction of heparin and the protein.

Keywords: antithrombin III; crystal structure; serpins; heparin binding; SIR phasing

1. Introduction

Antithrombin III (ATIII‡) is an important constituent of blood plasma which has been studied for over one century (Waugh & Fitzgerald, 1956; for reviews, see Travis & Salvesen, 1983; Lane & Caso, 1989). It is considered critical for regulation of the hemostatic mechanism because its physiological function is clearly defined as a key inhibitor involved in the blood coagulation cascade. ATIII inhibits serine proteases of intrinsic coagulation pathway such as factor Xa and thrombin (Damus *et al.*, 1973) which are its principal physiological targets in consideration of the relative rates of inactivation (Travis & Salvesen, 1983).

ATIII is a member of the serpins superfamily (Carrell & Travis, 1985) which includes serine protease inhibitors as well as non-inhibitory proteins such as hormone binding globulins, angiotensinogen and egg-white ovalbumin (Hunt & Dayhoff, 1980; Carrell *et al.*, 1987; Huber & Carrell, 1989). Like other inhibitory serpins, ATIII forms with its target proteases stable stoichiometric complexes, which after dissociation release an inhibitor cleaved at the P1–P1′ reactive bond. In the ATIII and thrombin case, proteolysis of the ATIII molecule occurs between Arg394§ and Ser395 (R358-S359; Björk *et al.*, 1981). The cleaved inhibi-

† Present address: Laboratoire de Pharmacologie et de Toxicologie Fondamentales, CNRS, 205 route de Narbonne, 31077 Toulouse Cedex, France.

‡ Abbreviations used: ATIII, antithrombin III; A1AT, α_1-antitrypsin (α_1-proteinase inhibitor); A1ACT, α_1-antichymotrypsin; LEI, leucocyte elastase inhibitor; 1PAI-1, plasminogen activator inhibitor-1; OVA, ovalbumin; $F^{P}_{hkl}(F^{PH}_{hkl})$, observed native (derivative) structure factors; INDEL, insertion and deletion; SIR, single isomorphous replacement.

§ Bovine ATIII residue numbering followed by human A1AT numbering in parentheses (residues of sequence insertions are given consecutive letters with the number of the last aligned residue) based on sequence alignment from Huber & Carrell (1989).

0022-2836/93/130223-19 $08.00/0

tor has lost all or most of its ability to inhibit thrombin (Fish *et al.*, 1979) and undergoes a remarkable conformational change, since the P1 and P1′ residues that flank the scissile bond are found at opposite poles of the molecule, 70 Å apart (Mourey *et al.*, 1990). Such conformational rearrangement was observed in the three-dimensional structures of cleaved A1AT (Loebermann *et al.*, 1984) and modified A1ACT and LEI (Baumann *et al.*, 1991, 1992) but does not occur in the case of ovalbumin, which is not a protease inhibitor, as shown by the crystallographic structures of its cleaved (Wright *et al.*, 1990) and uncleaved (Stein *et al.*, 1990) form.

The peculiarity of the ATIII action is its relatively slow progressive inhibitory effect unless it binds a sulphated polysaccharide, heparin (Björk *et al.*, 1989). The formation of the thrombin–ATIII complex is accelerated by, at least, three orders of magnitude in the presence of catalytic amount of heparin (Hoylaerts *et al.*, 1984).

Bovine ATIII is a glycoprotein of molecular weight 56,600 Da, of which 12% represents the total carbohydrate (Kurachi *et al.*, 1976). Its primary structure was recently established by peptide sequencing from five different digests (Mejdoub *et al.*, 1991). It consists of 433 amino acid residues.

As previously described, we have grown crystals of bovine ATIII in its cleaved form (Samama *et al.*, 1989) and the solution from molecular replacement using the A1AT structure (Delarue *et al.*, 1990) could only be refined to an *R*-factor of 0·28 at 3·2 Å (Mourey *et al.*, 1990). The phases derived from this preliminary model were used to interprete a platinum derivative. We report here the structure determination of ATIII from SIR phasing followed by real space density modification and phase extension. This proceeding provided a high quality electron density map free of any bias which allowed interpretation and refinement of the protein structure to an *R*-factor of 0·212 at 3·2 Å resolution.

2. Experimental Methods

(a) *Crystals*

ATIII was purified from bovine plasma by affinity chromatography on heparin-agarose (Thaler & Schmer, 1975). Its purity was checked on gel electrophoresis and activity assays of the purified protein using chromogenic substrate S-2238 (Abildgaard *et al.*, 1977) were normal. Protein concentration was determined from measurements of absorbance at 280 nm, assuming an extinction coefficient of 6·0 for a 1% (w/v) solution of ATIII (Kurachi *et al.*, 1976). Pure protein was then separated in aliquots and stored at −20°C. Crystals were prepared as previously described by Samama *et al.* (1989) using the microdialysis method at 4°C from protein (13 mg/ml) in 70% saturated ammonium sulphate/30 mM phosphate buffer (pH 6·2) containing 1‰ sodium azide (NaN_3, w/v) and 1 mM divalent cations ($CdCl_2$, $HgCl_2$ or $ZnCl_2$). Crystals grew within 3 weeks as cubes of dimensions 0·3 mm × 0·3 mm × 0·3 mm. They belong to the tetragonal space group $P4_32_12$ with cell parameters $a = b = 91{\cdot}3$ Å, $c = 383{\cdot}1$ Å, and contain 2 molecules in the asymmetric unit. The fractional volume occupied by the solvent, as defined by Matthews (1968), is 65%.

Table 1
Statistics for ATIII film diffraction data

Compound	Native	PT1†
Resolution range (Å)	41·3·2	35·3·5
Number of measurements	61,000	29,930
Number of unique reflections	23,436	14,541
Completeness of data (%)		
Overall	84‡	68
Outer resolution shell§	42	46
Reflections with $F > 3{\cdot}0\sigma$ (%)		
Overall	90	91
Outer resolution shell	74	74
R_{merge}¶	0·091	0·076
R_{iso}‖	—	0·233

† Platinum derivative corresponding to native crystals soaked in a 2 mM solution of K_2PtCl_4 for 3 days.
‡ 94% at 3·5 Å resolution.
§ 3·42 to 3·20 Å in the case of the native data set, 3·74 to 3·50 Å for PT1.

¶ $R_{merge} = \dfrac{\Sigma\Sigma|\langle I_{hkl}\rangle - I^i_{hkl}|}{\Sigma\Sigma I^i_{hkl}}$ where I^i_{hkl} is the ith observation of the hkl reflection of average intensity $\langle I_{hkl}\rangle$.

‖ $R_{iso} = \dfrac{\Sigma|F_{PH} - F_P|}{\Sigma F_P}$.

Biochemical analyses of protein from washed, dissolved crystals revealed that the protein was cleaved at the peptide bond Ser395-Leu396 (S359-L360), one amino acid away from the known reactive site. However, the major species in the supernatant mother liquor corresponded to native ATIII (Samama *et al.*, 1989; Mejdoub *et al.*, 1991).

(b) *Diffraction data and molecular replacement*

Intensity data were collected on oscillation photographs using synchrotron radiation at LURE (Orsay, France) and CHESS (Cornell, Ithaca, NY, U.S.A.). The usable diffraction data extend to 3·19 Å resolution. Crystals soaked in different heavy-atom solutions were screened for intensity changes at our lab prior to data collection. Different derivatives were checked but only one derivative was useful. Diffraction data statistics are summarized in Table 1.

Rotation and translation functions using the three-dimensional structure of A1AT (Loebermann *et al.*, 1984; Protein Data Bank file 5API.BRK) as a model were solved and have been reported elsewhere (Delarue *et al.*, 1990). Refinement of this solution led to a crystallographic *R*-factor of 0·28 between 8·0 to 3·2 Å (Mourey *et al.*, 1990). This model will be referred as MOD1 in the following sections.

(c) *Single isomorphous replacement*

Derivative data were scaled to native data in spheres of reciprocal space using the local scaling method (Matthews & Czerwinski, 1975) as implemented in the program LOCAL of the CCP4 package†. Phases derived from MOD1 with program GENSFC were used to compute a difference Fourier map of the platinum derivative

† Unless otherwise stated with reference or author name, all programs used are from the CCP4 package (CCP4, 1979).

Table 2
Platinum heavy-atom sites for ATIII

Site no.†	x‡	y	z	Signal§	Amino acid environment‖
S1	0·3840	0·1069	0·2144	8·4σ	Cys129 molecule B
S2	0·3943	0·2683	0·2881	6·2σ	Met104, Ser117 molecule A
S3	0·3077	0·2913	0·2685	6·1σ	Cys22, Cys96 molecule A
S4	0·3786	0·3803	0·3409	5·2σ	Cys9, Cys129 molecule A
S5	0·3144	0·0792	0·4947	4·6σ	Met21 molecule B

† Difference Fourier map using the MOD1 calculated phases between 10·0 and 5·0 Å resolution. Initial interpretation with phases calculated from the crude molecular replacement solution (two A1AT molecules correctly oriented in the ATIII cell, R-factor of 0·46 between 10 and 7 Å resolution) gave 3 peaks (Delarue *et al.*, 1990). The missing sites, S2 and S5, were in the background, respectively, in the 17th and 27th positions.

‡ Fractional co-ordinates.

§ The height of the signal is given as the number of standard deviations (σ) above the mean value of the map. The first spurious peak is at 4·1σ, slightly less than for S5. Although its position was chemically in agreement with MOD1, it refined with a low occupancy and did not improve the phasing statistics. For these reasons, it was discarded.

‖ Localization on the refined structure. A and B designate the 2 molecules in the asymmetric unit which are related by the non-crystallographic symmetry (bovine ATIII sequence numbering).

(program FFT) and the largest peaks were identified using the program PEAK MAX. Five heavy-atom sites located on the surface of the protein were found (see Table 2). Two sites (S1 and S4) are related by the non-crystallographic symmetry. However, applying this operation to S3 and S5 resulted in positions which were 5 Å apart with respect to the equivalent initial positions. S2 is only occupied in one molecule of the asymmetric unit. These heavy-atom positions did not allow the interpretation of all peaks in the Harker sections of the difference Patterson map and, on the contrary, there was no Harker vector corresponding to some minor sites (data not shown).

Heavy-atom parameters were refined with the program REFINE2 using least-squares procedure on centric reflections (Dodson, 1976) and are represented in Table 3.

Initial SIR phases were calculated from the refined heavy-atom parameters between 15 and 4·5 Å resolution using the program PHARE. This gave a figure of merit m of 0·42 and a phasing power of 1·34 for 7376 reflections. The corresponding electron density map was of poor quality and did not allow the definition of the molecular envelope nor secondary structure elements of the protein.

Table 3
Heavy-atom parameters

Site no.	Q†	x	y	z	B‡
S1	1·394	0·3899	0·0963	0·2133	40
S2	1·838	0·3949	0·2546	0·2894	60
S3	1·534	0·2973	0·2891	0·2681	60
S4	1·287	0·3848	0·3716	0·3412	50
S5	1·514	0·3126	0·0820	0·4967	50

† Relative occupancies obtained by least-squares refinement using 887 centric reflections between 13 and 4·5 Å resolution. During this refinement, positions and temperature factors were fixed to their previously refined values. Quality of the refinement is assessed by the following statistics (Dodson, 1976): an R-factor, equivalent to R_{Cullis} (Cullis *et al.*, 1961) of 0·626, a correlation coefficient of 0·378 and a gradient of the slope between the calculated and observed structure factors of 0·29.

‡ Temperature factors in Å^2 were obtained by refining occupancies against centric reflections in shells of data on $\sin^2\theta/\lambda^2$ *with fixed positions (Evans, 1983) and initial temperature factors* of 20 Å^2. They are given by the slope of $\ln(Q)$ against $\sin^2\theta/\lambda^2$.

(d) *Density modification between 15·0 and 4·5 Å*

For solvent flattening, the mask was computed using MOD1 *in the following way. Structure factors were calculated* between 200·0 and 4·5 Å resolution with the program GENSFC. Weighted structure factors were then calculated using the program HKLWEIGHT (Leslie, 1987, 1988) using an averaging sphere radius of 5 Å (this step is performed to obtain a slightly enlarged mask in order to prevent pernicious effects of an excessive flat*tening at the protein surface) and were used to compute* an averaged electron density map with the program FFT between 200·0 and 4·5 Å resolution. The final mask was obtained using the program MAKE_MASK (P. Dumas) by setting all points of density greater than a fixed threshold (0·01) to 1·0 and gives an effective solvent content of the crystals of 64%. This value is in good agreement with that obtained using the "Matthews' formula".

Starting from the mask and initial SIR phases, the solvent flattening procedure (Wang, 1985) was used to improve the quality of the phases and the interpretability of resultant electron density maps. All calculations were performed using programs of the CCP4 package as described by Leslie (1988); 6 cycles of solvent flattening were carried out between 15 and 4·5 Å. The final combined phases and the new figures of merit were used in a phase refinement (Blow & Matthews, 1973) of heavy-atom parameters (Rould *et al.*, 1989) with the PHASE option of the program REFINE2, using only reflections with m greater than 0·5. New SIR phases were calculated from those refined heavy-atom parameters and the whole procedure was iterated 5 times. The scale factor between native and derivative structure factors was obtained using the program RFACTOR and never refined.

All calculations for local symmetry averaging were performed in real space (Buehner *et al.*, 1974; Bricogne, 1974, 1976) using programs developed by J. E. Johnson (1978) and implemented in our laboratory by B. Rees in the RMOL package (Rees *et al.*, 1990). The procedure can be separated in 3 different steps.

(1) Calculation of the molecular envelope for both molecules (i.e. A and B) of the asymmetric unit. They were constructed using spheres centred around C^α and C^β atoms of MOD1 *following the protocol described by Rees et al.* (1990). After careful elimination of redundant points

Table 4
Non-crystallographic symmetry

Transformation†	κ (°)	Direction cosines			Translation vector (Å)		
A → B‡	−92·70	0·05621	−0·99837	0·00967	453.01541	51.14070	236.71820
B → A	92·70	0·05621	−0·99837	0·00967	−213·85081	−30·86634	466·03537

† With the cell origin taken as the rotation centre and the orientation matrix of the local molecular axis equal to the unit matrix, the non-crystallographic symmetry operations relating both molecules of the asymmetric unit are defined by the direction cosines of the non-crystallographic axis, the κ angle and the translation vector components in the crystallographic axes.

‡ Molecules A and B of the asymmetric unit which correspond to the most compact dimer.

(spheres intersection, overlapping points between crystallographic equivalent molecules), the largest envelopes without overlaps between non-crystallographic equivalents were obtained with a sphere radius of 4 Å. The fraction of crystal volume occupied by the solvent, estimated from the number of points falling outside the molecular envelopes, is 61%.

(2) Transformation of each envelope and determination of the non-crystallographic equivalent points into a single asymmetric unit using the non-crystallographic symmetry operator obtained by the best superposition of the C^{α} atoms of the 2 molecules of MOD1 (Table 4).

(3) Molecular averaging, carried out using the RMOL package and the programs FFT and SFC for the Fourier calculations. As the 2 molecules in the asymmetric unit are related by a screw axis (Delarue *et al.*, 1990), it was not possible to use a common envelope and the following steps had to be repeated: (1) density evaluation at each molecular envelope point of one monomer and at its non-crystallographic equivalent using linear interpolation; (2) averaging; (3) combination of the average electron density values with grid points co-ordinates within the asymmetric unit and attenuation of densities smaller than a lower threshold or larger than an upper threshold; (4) transformation into a standard density file (CCP4 type) with constant or initial density values affected at points not defined within the envelope. After each cycle, the phases calculated from the modified density were combined with the initial SIR phases (Bricogne, 1976). The combined phase probability density was obtained after multiplication of SIR and Sim (Sim, 1959) phase probability densities with the same weight for the 2 sources of phase information (i.e. the coefficients u and v were set to 1·0). During combination, unobserved reflections (i.e. with only calculated values) were rejected. For reflections with observed figures of merit greater than 0·94, the calculated phases were ignored. Molecular averaging was iterated 8 times after which convergence was reached.

The sampling used for the 3-dimensional grid was of about 1 Å and this value was kept during all the procedure.

The different steps of density modification and heavy-atom parameters refinement iterative procedure that were carried out at 4·5 Å resolution are presented in Appendix I. The results of density modification are given in Table 5.

Table 5
Results of density modification between 15 and 4·5 Å

Step	Cycle	C† mol. A	C mol. B	R-factor‡	Number of reflections	$\Delta\phi$§ (°)	m_{OPT}‖
SIR	—	—	—	—	7367	—	0·420
SF1	6	—	—	0·317	7411	34·1	0·769
SF2	6	—	—	0·320	7410	33·7	0·769
SF3	6	—	—	0·322	7411	33·5	0·768
SF4	6	—	—	0·322	7410	33·1	0·768
SF5	6	—	—	0·322	7411	33·5	0·768
	1	0·565	0·566	0·391		43·6	0·712
A1	2	0·779	0·780	0·378	7405	45·4	0·722
	8	0·855	0·857	0·338		50·1	0·747
	1	0·543	0·547	0·465		33·5	0·660
A2	2	0·797	0·799	0·401	7405	40·2	0·707
	8	0·863	0·866	0·338		48·9	0·747

† Weighted average correlation coefficient at non-crystallographic equivalent densities:

$$C = \frac{\sum_{i=1}^{n}\sum_{j=1}^{m}\rho_i\rho_j}{\sum_{i=1}^{n}\sum_{j=1}^{m}\rho_j^2} \times \frac{m}{n(m-1)}$$

for n grid points in the molecular envelope and m points related by the non-crystallographic symmetry operation.

‡ R-factor $= \dfrac{\sum_{hkl}|F_{hkl}^{obs} - F_{hkl}^{calc}|}{\sum_{hkl}F_{hkl}^{obs}}$.

§ Mean phase difference between initial SIR phases and combined phases after density modification.

‖ Combined figure of merit (except for step SIR).

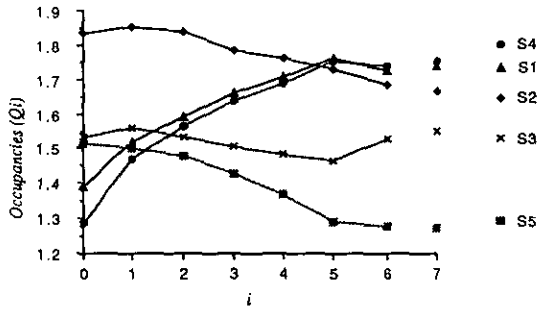


Figure 1. Evolution of the occupancies during the refinement of platinum heavy-atom parameters. $i = 0$ *corresponds to initial parameters,* $i = 1, 5$ *corresponds to* solvent flattening steps SF1 → SF5, $i = 6$ and 7 correspond to molecular averaging steps A1 and A2 (see *Appendix I*).

As it was first described for heavy-atom derivatives with closely related heavy-atom sites, MIR phasing can be improved by heavy-atom refinement using solvent flattened phases (Rould *et al.*, 1989, 1992). An identical procedure was applied in the SIR case of ATIII with modified phases coming both from solvent flattening and molecular averaging. After each density modification step (SF1 → SF5 and A1), combined phases were used to refine initial heavy-atom parameters (x_0, y_0, z_0, Z_0) (Table 3). The corresponding SIR phases were introduced in the procedure for the next step. As a test case, molecular averaging was also directly applied to initial SIR phases (step A2). Examination of the heavy-atom parameters evolution indicated small variations on co-ordinates and a good convergence at the end of the procedure. The minimum and maximum shifts on co-ordinates of the 5 heavy-atom sites were, respectively, 0·07 Å and 0·3 Å (data not shown). This proved the effectiveness of co-ordinates refinement using centric reflections. However, important changes were observed for the occupancies (Fig. 1). It is clear, from the comparison at points $i = 6$ and $i = 7$, that in our case, correct occupancies would be attained after a single step of molecular averaging and that iterative solvent flattening resulted in a slow convergence towards the correct values. Heavy-atom sites S1 and S4 almost had the same final occupancies corresponding to the best occupied sites ($Q \approx 1{\cdot}8$) though their initial values were the smallest. The similar behaviour for S1 and S4 consistently recalls that these 2 sites are the only ones related by the non-crystallographic symmetry. On the contrary, sites S3 and S5 whose positions did not obey the non-crystallographic symmetry had identical initial occupancies which diverged gradually. In that case, the maximum shift was obtained after non-crystallographic symmetry averaging.

The mean phase difference between the refined structure phases and the optimized phases obtained after solvent flattening (ϕ_i^{OPT}, $i = 1, 5$) was not significant. Starting from a mean phase error of 75·1°, the final value was 73·2° after 5 iterations. This should be compared to the 66° value, obtained after step *A*1 or *A*2 ($i = 6, 7$), which emphasizes the strong effect of local symmetry averaging (Cura *et al.*, 1991). On the contrary, the mean phase difference between the refined structure phases and SIR phases derived from heavy-atom parameters refinement (ϕ_i^{SIR}, $i = 1, 7$) was 81·2° at the start and showed no marked improvement with a stable value of 80·5°. It should be noted that this slight effect was completed after the first solvent flattening step (SF1), and that afterward there was no improvement even after molecular averaging.

Electron density maps were calculated after each density modification step using the combined phases and figures of merit. Maps evolution is in agreement with the above analysis, averaged maps displaying clearer features than solvent flattened maps which, indeed, did not vary from SF1 to SF5. Figures 2(a) and 3(a) show different regions of the averaged electron density map centred on portions of the final structure of ATIII.

(e) *Phase extension from 4·5 to 3·2 Å*

Phase extension from 4·5 to 3·2 Å was carried out very gradually using 2 c^* lattice units at a time (i.e. 18 resolu-

Table 6

Results of phase extension from 4·5 to 3·2 Å

Resolution (Å)	*C* mol. A	*C* mol. B	*R*-factor	*Number of* reflections	$\Delta\phi$ (°)	m_{OPT}
4·40	0·912	0·917	0·311	9902	52·7	0·735
4·30	0·918	0·923	0·295	10,607	54·3	0·739
4·21	0·922	0·928	0·284	11,327	55·3	0·741
4·12	0·925	0·932	0·275	12,095	56·0	0·743
4·03	0·927	0·935	0·268	12,914	56·7	0·743
3·95	0·928	0·938	0·261	13,685	57·5	0·744
3·87	0·930	0·939	0·257	14,549	58·2	0·744
3·79	0·931	0·941	0·253	15,445	58·9	0·742
3·72	0·932	0·942	0·246	16,330	59·4	0·745
3·65	0·932	0·943	0·243	17,220	59·7	0·745
3·58	0·932	0·944	0·240	18,191	60·0	0·745
3·51	0·933	0·946	0·237	19,152	60·2	0·746
3·45	0·932	0·945	0·237	19,928	60·4	0·744
3·39	0·932	0·945	0·237	20,565	60·4	0·742
3·33	0·931	0·945	0·238	21,120	60·5	0·740
3·27	0·931	0·945	0·238	21,593	60·5	0·737
3·22	0·931	0·945	0·240	21,956	60·5	0·735
3·19	0·930	0·945	0·239	22,056	60·5	0·734

Statistics are given for each resolution shell at the 8th cycle of density modification. Correlation coefficient on non-crystallographic equivalent densities *C*, *R*-factor, $\Delta\phi$ and m_{OPT} are defined in Table 5. The number of phased reflections is given. $\Delta\phi$ and m_{OPT} values are given for reflections with SIR phases (i.e. between 15·0 and 4·5 Å resolution).

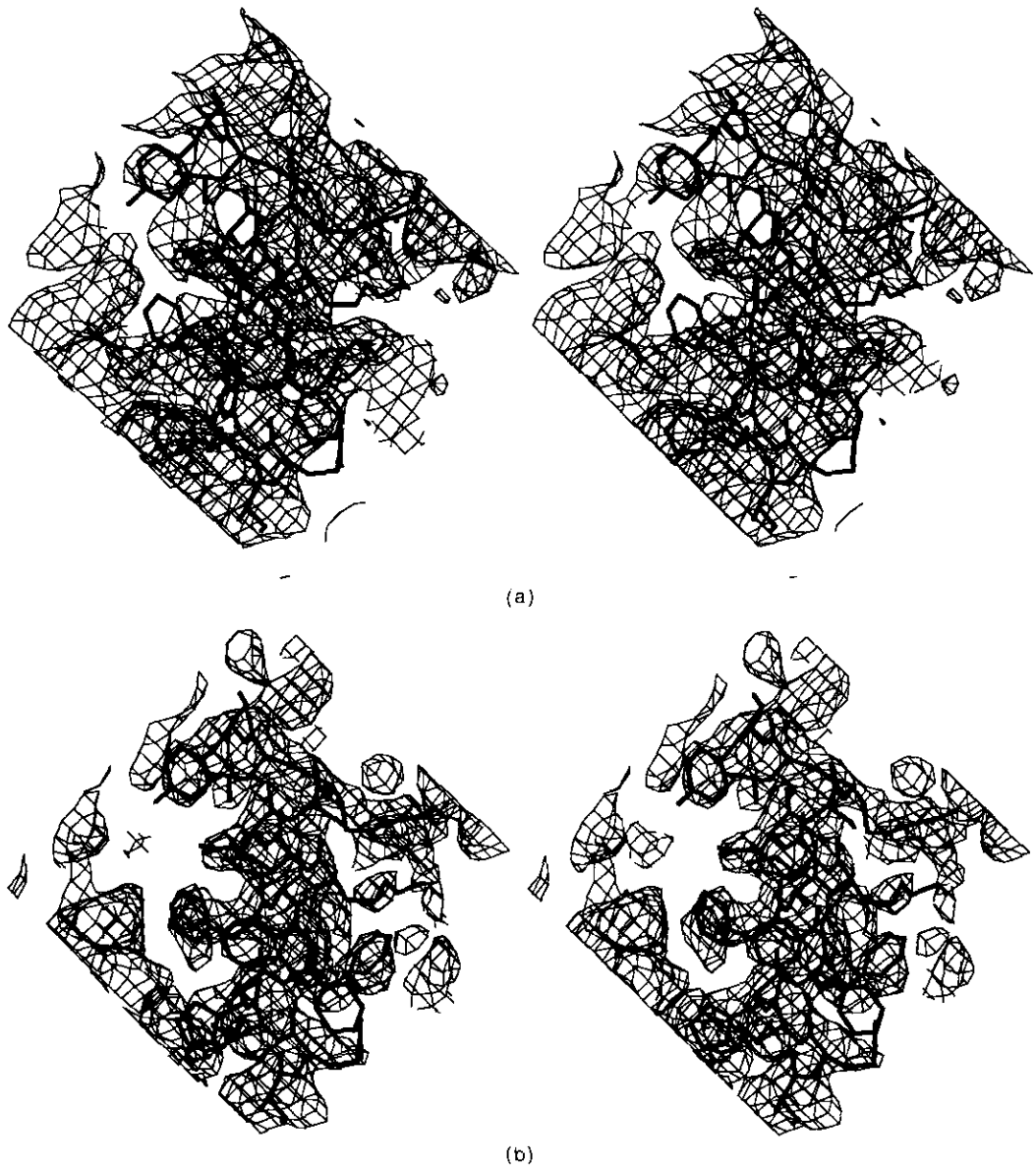


Figure 2. Residues 118 to 132 (bovine sequence numbering) of the final model of ATIII and averaged electron density maps obtained after: (a) step A1, (b) phase extension at 3·2 Å. The contour level is 1·0σ.

tion shells) interleaved with 8 cycles of molecular averaging at the current resolution with the RMOL package (Table 6). The same lower resolution limit of 15·0 Å was used. Using those reflections with modified phases obtained after *A1*, and after adding the first resolution shell, phase extension was conducted to 4·4 Å. 2497 new reflections, lacking SIR phase, were added. They corresponded to 1835 reflections between 15 and 4·5 Å, with measured F^{P}_{hkl} but without F^{PH}_{hkl}, and 662 reflections between 4·5 and 4·4 Å. It should be mentioned again that during all the density modification processes, unobserved reflections (i.e. F^{P}_{hkl} absent) were not replaced by their corresponding calculated values; 17 other steps were carried out until the final limit of 3·19 Å. At this stage, the mean difference between optimized phases and initial SIR phases was 60·5° for the 7405 reflections phased by the platinum heavy-atom derivative and a total number of 14,651 new phases were obtained. The optimized figure of merit and the *R*-factor were, respectively, 0·734 and 0·239. During phase extension, the 4·5 Å phases improved continuously to reach a mean phase error of 45·1° compared to the refined structure phases, while the mean phase error for reflections between 4·5 and 3·2 Å amounted to 50·0° (Table 7). The corresponding electron density map, computed between 8·0 and 3·2 Å, is of excellent quality (Figs 2(b) and 3(b)).

(f) *Model building and refinement*

Model building was performed on an Evans & Sutherland ESV3+ graphics system using the program FRODO (Jones, 1978). Crystallographic refinement was carried out using the X-PLOR package version 2.1 (Brünger, 1990) implemented on a VAX station 5000-200 and an HP station 720. Conventional energy minimization and simulated annealing with molecular dynamics (Brünger, 1988) were performed. In simulated annealing, the slow-cooling protocol was applied (Brünger *et al.*, 1990) starting from 3000 to 300 K for 2·7 ps. Refinement was carried out in 9 steps.

In steps 1 to 3, the 2 molecules in the asymmetric unit were assumed identical. Manual corrections were thus

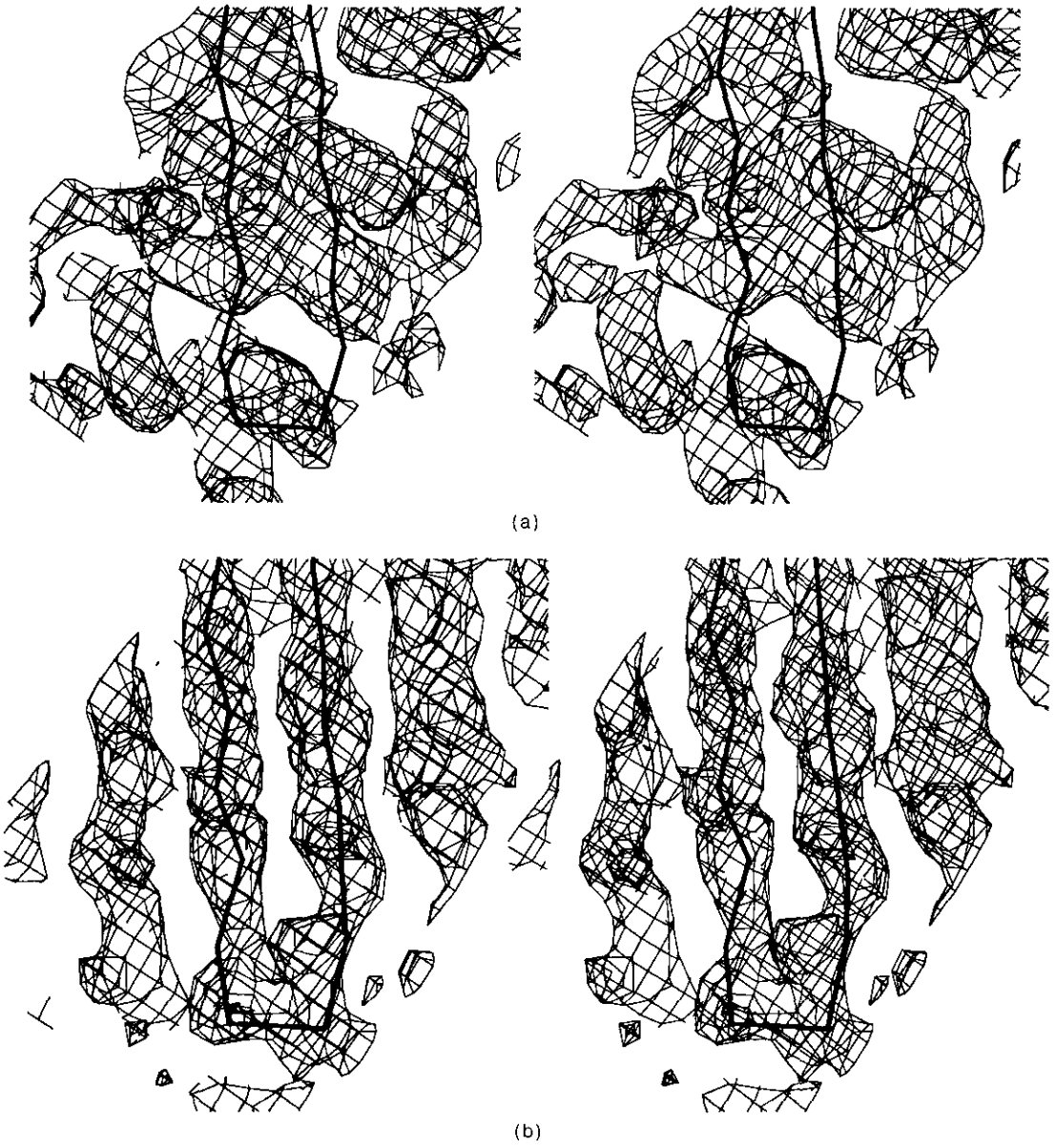


Figure 3. The C^{α} chain trace of residues 372 to 384 (bovine sequence numbering) of the final model of ATIII and averaged electron density maps obtained after: (a) step A1, (b) phase extension at 3·2 Å. The contour level is 1·0σ.

made on one monomer and the other molecule of the asymmetric unit was generated using the non-crystallographic symmetry operation described in Table 4. During refinement, non-crystallographic symmetry related atoms were maintained to their average positions by an effective force constant of 50·0 kcal/mol Å^2. In steps 4 to 9, the 2 ATIII copies were handled and treated independently.

Between each refinement cycle, model building was

Table 7

Evolution of the mean phase error during the molecular averaging and phase extension procedures

Resolution (Å)	Number of reflections	Mean phase error (°)		
		4·5 Å phases	Extended phases	All phases
15-4·5 (SIR)	7367	81·2		81·2
15-4·5 (A1)	7405	66·0	—	66·0
15-3·19	22,056	45·1†	50·0‡	47·9

† 9240 reflections. Including 1835 reflections with measured F^{P}_{hkl}, but without F^{PH}_{hkl}, which were not taken into account in step A1.

‡ 12,789 reflections between 4·5 and 3·19 Å resolution.

performed using different kind of maps.

(1) The averaged map was used for fitting in earlier stages and kept as a reference map at further stages.

(2) Electron density maps calculated with Fourier coefficients $F_{obs}+n(F_{obs}-F_{calc})$, ϕ^{calc} with $n = 1, 2$ were used all through the refinement.

(3) In steps 6 and 7, phase combination was performed with the program COMBINE. Maps were calculated with Fourier coefficients $m_{comb}[F^{obs}+Q_{comb}(F^{obs}-F^{calc})$, ϕ^{comb} (Stuart & Artymiuk, 1984) where m_{comb} and ϕ^{comb} correspond to the combination of the model calculated phases with phases obtained after density modification.

(4) In steps 8 and 9, Fourier coefficients $2mF_N-DF^{calc}$, ϕ^{calc} computed with the program SIGMAA (Read, 1986) were used for maps calculation.

MOD1 was used as initial chain tracing. It was globally in agreement with the electron density map although some protein regions appeared incorrectly positioned. The largest differences occurred between helices C and D: residues 109 to 116 (82 to 88) (the secondary structural elements classification is based on the molecular structure of A1AT), and for residues 286 to 314 (254 to 281) corresponding to helices G and H. Few errors occurred in areas where ATIII and A1AT display insertion or deletion, which were all located in loop regions. Finally, a large portion of the N-terminal domain and residues 134 to 140 (106 to 111), between helix D and strand 2A, were poorly defined in the electron density map.

337 residues were fitted in the averaged electron density map and regions: 1 to 19, 25 to 47 (−2 to 21), 134 to 140 (106 to 111), 264 to 267 (232 to 236), 286 to 314 (254 to 281), 358 to 360 (324 to 325a), 393 to 401 (357 to 363b) were omitted in refinement. The corresponding R-factor was 0·429 (step 0) and dropped to 0·320 after energy minimization (step 1). In step 2, the region 286 to 314 (254 to 281) and a few other residues were fitted in their electron density. The resulting model of 388 amino acids led to an R-factor of 0·283 after conventional refinement. The 10% missing residues belonged to the N-terminal part 1 to 14, 25 to 44, −2 to 18), loop 135 to 138 (107 to 109a) and residues 396 to 400 (360 to 363a), the second N terminus in the cleaved protein. At this stage, refinement and electron density maps did not show significant improvement and we suspected that information might have vanished at one stage of the density modification procedure. A new and enlarged molecular envelope was defined using the current model and extra dummy atoms, and the whole procedure of molecular averaging and phase extension was repeated. Those dummy atoms were placed in regions where it was clear, at that stage of refinement, that the initial envelope was too close to the protein atoms. The new averaged map indeed displayed improved electron density in the N-terminal region which allowed model building and refinement to proceed in step 3. The R-factor obtained after energy minimization and simulating annealing was 0·210 for 424 residues. From step 4 and on, several strategies were followed in order to define, at the best possible accuracy, the position of some residues involved in the N-terminal part and also in a few connecting loops of the protein. In this context, energy minimization was done at every step while simulating annealing was performed at steps 4, 7 and 8.

The final crystallographic structure consists of the 2 ATIII molecules of the asymmetric unit (denoted A and B). One molecule (A) contains 420 among 433 possible residues, lacking amino acids 1 to 4, 26 to 30 (−1 to 4), 396 to 399 (360 to 363). The other one has 410 well defined residues and lacks regions 1 to 11, 28 to 38 (2 to 12), 396 (360). Despite visible extra electron density that might account for additional chemical groups, neither sugars nor solvent molecules (water, sulphate and phosphate ions) and metal ions were included in refinement. Table 8 gives the final refinement statistics.

Table 8

Final refinement statistics

Resolution range (Å)	8·0-3·2
Number of protein atoms (excluding hydrogen)	6634
R-factor:	
19,543 reflections with $F > 3{\cdot}0\sigma$ (75% of total)	0·198
20,702 reflections (80% of total)	0·212
Average temperature factors B (Å^2)†:	
Molecule A	
Main-chain	18·4 (1·3)
Side-chain	20·2 (1·8)
Molecule B	
Main-chain	24·1 (1·3)
Side-chain	26·1 (1·7)
Root-mean-square deviation from ideal geometry	
Bonds (Å)	0·015
Angles (°)	3·63
Dihedrals (°)	26·09
Impropers (°)	1·51

† All values were calculated using the program BAVERAGE. The number in parentheses refers to the root-mean-square deviation in B values.

3. Antithrombin III Structure

(a) *Stereochemistry*

A Ramachandran plot for the two molecules is given in Figure 4. Three residues have an unfavourable conformation: Asp7, Asn97 (N70) and Thr402 (T363c). Thr402 is in the beginning of a β-strand and is preceded by Val401 (V363b) and Arg400 (R363a) of weak electron density in molecule A. Asp7, in the N-terminal domain, also has a poor electron density. On the contrary, Asn97 is well defined in both molecules but is in a high-energy conformation as in A1AT (Huber & Carrell, 1989), A1ACT (Baumann *et al.*, 1991) and ovalbumin (Stein *et al.*, 1991). It is located in an α-helix and it follows cysteine 96 (C69) which is disulphide linked. In addition it is involved in a turn and its side-chain is glycosylated (see below). Asn97 polar atoms are involved in a network of hydrogen bonds. Its main-chain carbonyl oxygen is close to Thr101 (T74) O^{γ}, the O^{δ} is in the vicinity of the side-chain nitrogen NH2 of Arg351 (R317a) and, in monomer B, the N^{δ} is in close contact to Thr154 (T125) from a crystallographic equivalent ATIII molecule. Cys96 (C69) main-chain oxygen is also hydrogen bonded to main-chain nitrogen of Thr99 (T72). About ten non-glycine residues have a left-handed α-helical conformation. Among those, Lys108 (K81) is at the beginning of a helix while residues: Asn179 (N149a), Glu196 (E166), Gln206 (Q176), Tyr241 (Y209), Glu266 (E234), Ser267 (S236), Asp278 (D246), Asn419 (N380) are involved in β-turns. Few other residues, found in this conformation, are located in

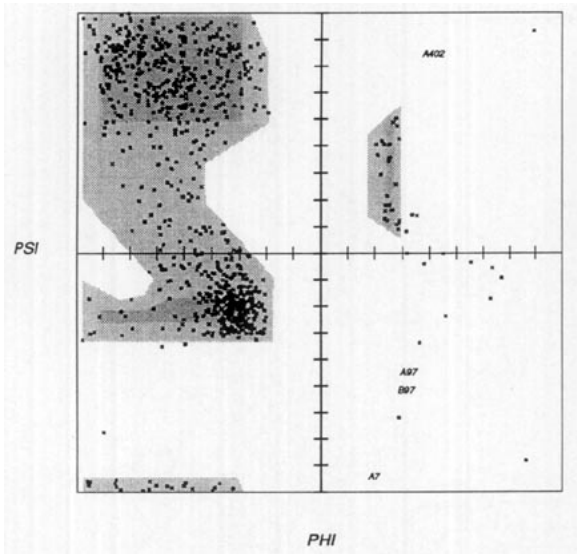


Figure 4. Ramachandran plot of main-chain torsion angles for the 2 molecules of ATIII in the asymmetric (A and B). All residues are denoted by crosses. Non-glycine residues falling outside "allowed regions" are labelled.

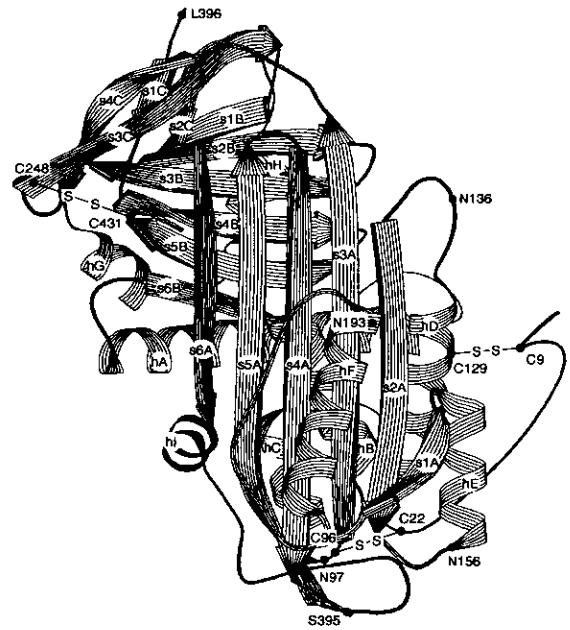


Figure 5. Schematic drawing of the ATIII structure produced using the program RIBBON (Priestle, 1988). Residues Ser395 (S359) and Leu396 (L360) form the cleavage site. The asparagine glycosylation sites and cysteine residues involved in disulphide linkage are shown (bovine ATIII sequence numbering and one letter code amino acid symbols).

loops. The expected co-ordinate error, obtained using a plot of the *R*-factor distribution (Luzatti, 1952), is between 0·30 and 0·35 Å.

(b) *Description of the structure*

ATIII is a globular protein of ellipsoidal shape with overall dimensions 55 Å × 55 Å × 65 Å (Fig. 5). Its secondary structural elements, given in Table 9, were assigned using the program DSSP (Kabsch & Sander, 1983) and labelled according to the A1AT nomenclature (Loebermann *et al.*, 1984; see also Table 1 of Huber & Carrell, 1989). The overall structure of ATIII resembles that of A1AT (see below). Most residues are involved in well-ordered secondary structures such as helices (31%), β-sheets (37%) and turns (15%). There are nine α-helices (denoted A to I), three short 3_{10}-helical segments

Table 9

Secondary structural elements in antithrombin III

Helices			Sheets			Turns		Bulges†	
α-Helices						Turns ($O_i \rightarrow N_{i+3}$)			
hA	47-71‡	(21-45)	s1A	169-175	(140-146)	72-75	(46-48)	366, 367, 390	(330, 331, 354)
hB	80-92	(53-65)	s2A	140-150	(111-121)	92-95	(65-68)	422, 423, 413	(383, 384, 374)
hC	97-107	(70-80)	s3A	213-226	(181-194)	96-99	(69-72)		
hD	116-130	(88-102)	s4A	379-393	(343-357)	150-153	(121-124)		
hE	156-167	(127-138)	s5A	362-378	(326-342)	195-198	(165-168)		
hF§	179-195	(149a-165)	s6A	323-332	(290-299)	204-207	(174-177)		
hG	292-299	(259-264)	s1B	260-265	(228-233)	209-212	(179-180)		
hH‖	304-311	(271-278)	s2B	267-275	(236-244)	242-245	(210-213)		
hI	332-340	(299-307)	s3B	279-288	(247-256a)	275-278	(244-246)		
3_{10}-Helices			s4B	408-416	(369-377)	288-291	(256-258)		
hC1	108-112	(81-84)	s5B	419-428	(380-389)	301-304	(268-271)		
hF1	231-234	(199-202)	s6B	76-80	(49-53)	346-349	(313-316)		
hI1	342-346	(309-313)	s1C	402-405	(363c-366)	377-380	(341-344)		
			s2C	315-322	(282-289)	α-Turns ($O_i \rightarrow N_{i+4}$)			
			s3C	245-258	(213-226)	175-179	(146-149a)		
			s4C	234-242	(202-210)	264-268	(232-237)		
						415-419	(376-380)		

† Classic β-bulges with residues 1, 2 and X, respectively, and structural properties as described by Richardson (1981).

‡ Bovine ATIII sequence numbering followed by human A1PI numbering in parentheses. Residues at the termini of helices or β-strands are included if at least 1 of their main-chain conformation angles (ϕ, ψ) is canonical.

§ 180 to 195 (150 to 165) in molecule B.

‖ 305 to 310 (272 to 277) in molecule B.

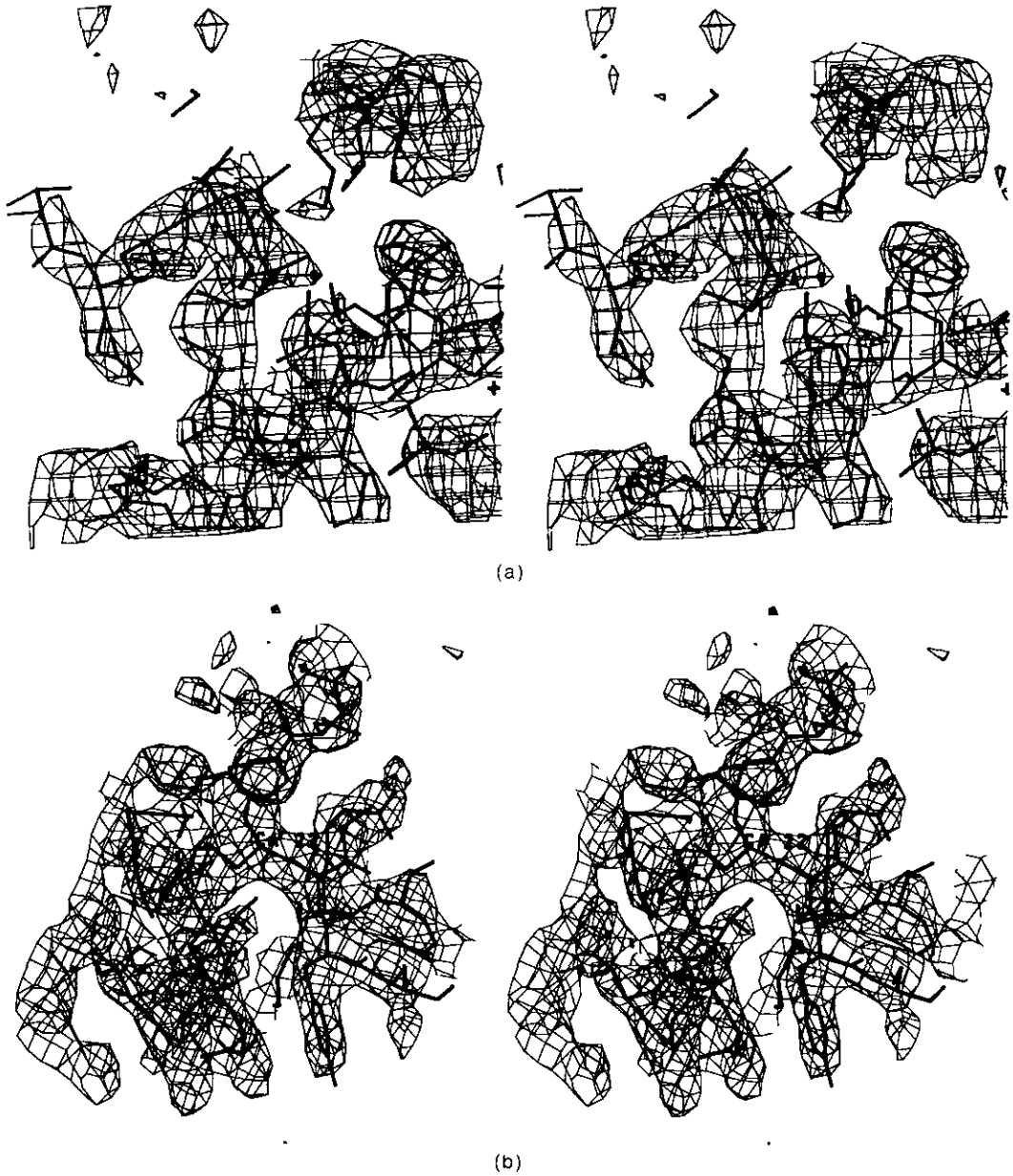


Figure 6. The final $2F_{obs}-F_{calc}$ electron density map, calculated from the refined model and contoured at 1·0σ, displayed around the cysteine residues of the N-terminal part: (a) Cys9, (b) Cys22.

(hC1, hF1 and hI1 which contain from 4 to 5 residues) and three β-sheets (A to C). There are six cysteine residues which form three disulphide linkages.

(c) *N-Terminal domain, cleavage site and disulphide bonds*

ATIII sequence has a specific N-terminal extension (26 to 48 residues when compared to A1AT and OVA) which is dedicated to heparin binding. All three-dimensional structures solved so far within the serpins family start at helix hA, the normal N-terminal end in the case of OVA (Stein *et al.*, 1991), PAI-1 (Mottonen *et al.*, 1992) and LEI (Baumann *et al.*, 1992) but 19 and 23 residues away from the respective N termini of A1AT (Loebermann *et al.*, 1984) and A1ACT (Baumann *et al.*, 1991). The N-terminal region of ATIII (molecule A) forms a long loop which is tied to the molecule by two disulphide bridges: Cys9 with Cys129 (C101), Cys22 with Cys96 (C69). The loop is at the protein surface, lies between helices D and E, passes nearly round hC1 and finally changes its direction before entering helix A. Figure 6 shows portions of the corresponding electron density map. Residues 1 to 4 and 26 to 30 (−1 to 4) have no interpretable electron density in molecule A. In molecule B, these disordered regions are more extended: residues 1 to 11 and 28 to 38 (2 to 12), and there is no electron density corresponding to the disulphide bridge Cys9–Cys129. Isomorphism defaults could arise from the chain flexibility in this area, as expressed

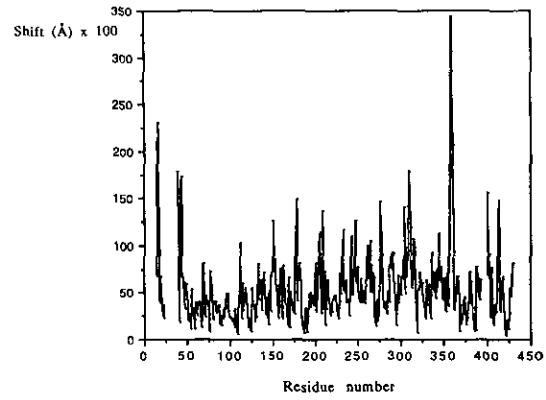


Figure 7. Distances between corresponding C^{α} atoms of each ATIII molecule in the asymmetric unit. Because both molecules are incomplete, 395 residues were taken into account omitting those at the edge.

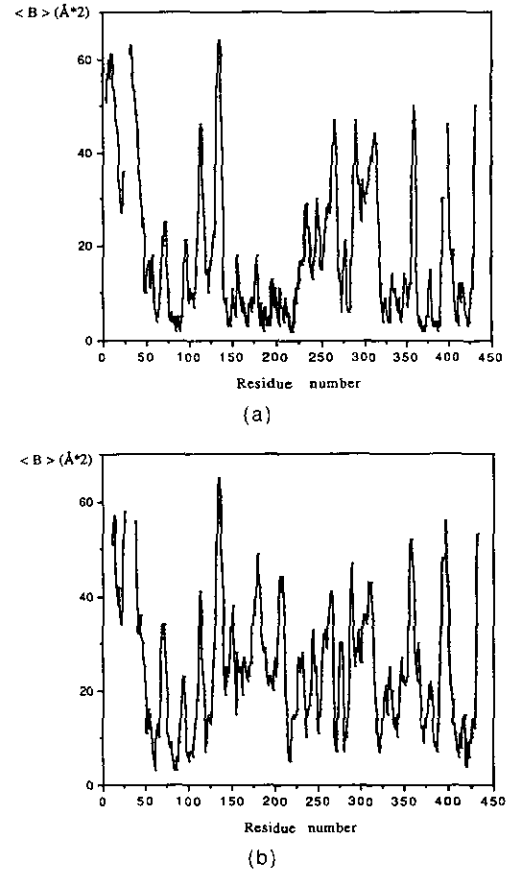


Figure 8. Average main-chain temperature factors ($Å^2$) calculated from the refined individual B values of N, C^{α}, C, O atoms for each ATIII molecule in the asymmetric unit: (a) molecule A, (b) molecule B. Residues 1 to 4, 26 to 30 (−1 to 4), 396 to 399 (360 to 363) of molecule A and residues 1 to 11, 28 to 38 (2 to 12), 396 (360) of molecule B are missing in the final model.

by the calculated high B values, and might explain the overall low diffraction power of the crystals. In addition, systematic errors in the initial phases could be expected from the fact that specific targets for heavy-atom binding are clustered in the N-terminal region of ATIII (Table 2).

The ATIII structure represents cleaved molecules. The cleavage site is located at the peptide bond Ser395–Leu396 (S359-L360), one residue after the known scissile bond, as shown by N-terminal sequencing of the molecular species present in the crystals (Samama *et al.*, 1989). There is no, even weak, density corresponding to residues 396 to 399 (360 to 363) in molecule A whereas the electron density is clearly defined from residue 397 in molecule B. After the cleavage site, the C-terminal domain (strands s1C, s4B, s5B) is part of the protein core and linked to the molecule by the disulphide bond between Cys248 (C216) and Cys431 (C392).

(d) *Crystal packing and comparison of the two ATIII molecules*

ATIII is a biologically active monomer in solution, but crystallizes as a dimer. The non-crystallographic dimer only results from packing constraints and cannot reflect any functional aspects since the molecules are cleaved. The distances between C^{α} atoms after least-squares superposition of both monomers are given as a function of residue number in Figure 7. The root-mean-square value is of 0·66 Å for 395 residues compared. The largest shifts occur at the end of the long loop by the N-terminal end of s5A. In this region, residues 358 to 361 (324 to 325b) are found in different conformations in both molecules as a clear result of different packing constraints. These residues are in the vicinity of Arg351 (R317a) side-chain and close to the putative glycosylation site Asn97 (N70). In the crystal, each molecule makes contacts with four other molecules. For instance, molecule A of the non-crystallographic dimer interacts with three molecules B and with a crystallographic equivalent of itself. The dominant packing interactions are by "head-to-head" and "tail-to-tail". The contact regions involve N terminus of hA, C terminus of hD, N terminus of s5A, s4C, and turns between s2A and hE, s1A and hF, s2B and s3B and hG, s4B and s5B, in both molecules and, in addition, s1A in molecule A and turn between hB and hC in molecule B. Packing constraints between homologous molecules involve topologically equivalent regions related by crystallographic 2-fold axes. The variation of the temperature factors averaged on main-chain atoms for each residue and for each monomer is shown in Figure 8. The higher values for residues 150 to 210 (121 to 179a) of molecule B compared to molecule A are due to different packing constraints and explain the higher global average B value for molecule B (Table 8). In both cases, the regions with the highest temperature factors correspond to the N-terminal domain of the protein and loop 133 to 138 (105 to 109b) which are close to each other.

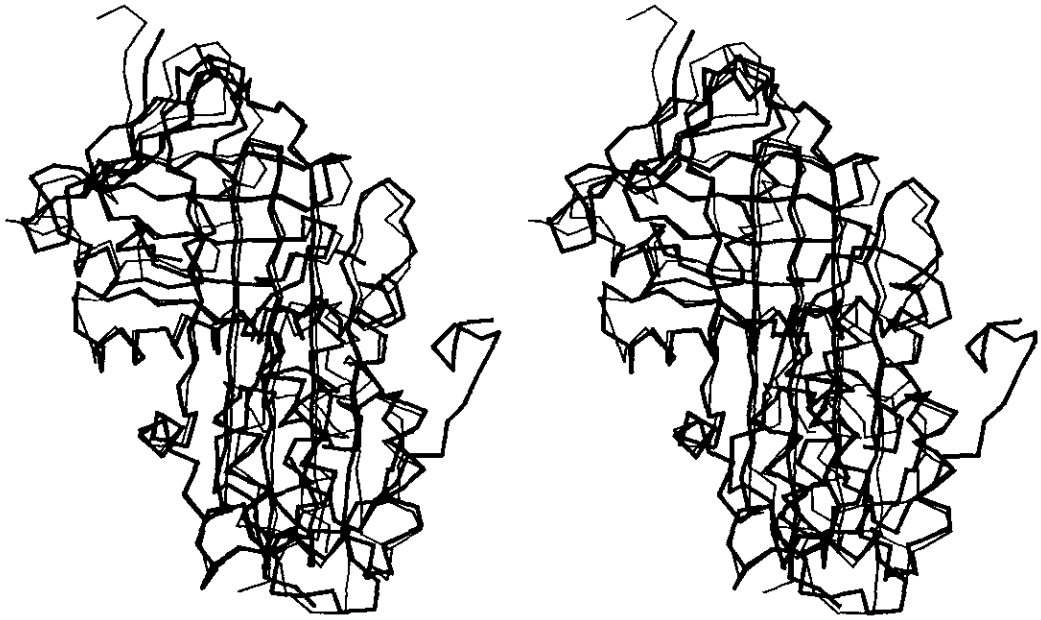


Figure 9. Stereoscopic view of the superposition of the C^{α} backbones of ATIII (thick line) and A1PI (thin line).

(e) *Post-translational modifications*

There are four N-X-T/S tripeptides, corresponding to the consensus sequence for N-glycosylation, in bovine ATIII: N-N-T 97 to 99 (70 to 72), N-K-S 136 to 138 (108 to 109a), N-E-T 156 to 158 (127 to 129) and N-K-T 193 to 195 (163 to 165). For all of them, it has been shown from amino acid sequencing that the corresponding asparagine residues are glycosylated (Mejdoub *et al.*, 1991) as is the case in human ATIII (Franzén *et al.*, 1980). The attachment sites are located at the surface of the protein: Asn97 lies in the N-terminal part of helix C, Asn136 is located in a loop between helix D and strand 2A, Asn156 is the first residue of helix E and Asn193 is at the C-terminal end of helix F (see Fig. 5). As earlier reported (Mourey *et al.*, 1990), both ATIII molecules of the asymmetric unit show very clear electron density at the expected locations of the carbohydrates at positions 156 and 193. Those densities, however, only allow the assignment of two to three sugar units of the biantennary carbohydrate substituent. There is weak density for the sugar group above Asn136 in molecule A and no density in molecule B. It should, however, again be pointed out that loop 133 to 138 (105 to 109a) displays very high *B* factors which require cautious interpretation of the electron density in this area. Weak and discontinuous electron density is found in the continuity of Asn97 side-chain in molecule A which might correspond to sugar groups. On the contrary, the absence of extra density and the tight packing contact found for this residue in the second molecule of the asymmetric unit strongly suggest that Asn97 is not glycosylated in molecule B. This microheterogeneity explains that crystallization of ATIII is dependent upon protein batches, as reported earlier (Samama *et al.*, 1989).

(f) *Comparison of ATIII and A1AT*

Antithrombin III was superimposed to A1AT using, as a reference, highly conserved secondary structure elements: hB, s4B, s5B and s6B. The root-mean-square deviation for the 35 corresponding C^{α} atoms is 0·61 Å (1·50 Å when strands s1B, s2B and s3B are included, 80 C^{α} atoms) and the resulting transformation matrix (Table 10) was applied to all atoms (Fig. 9). The structures of the two molecules are strikingly similar despite the small sequence identity of 28%. The sequence-conserved residues in the serpins as described by Huber & Carrell (1989) occupy equivalent positions in ATIII with the same spatial environment except for Leu93 (L66), close to conserved Tyr159 (Y130) and Tyr167 (Y138), and located in a surface niche in both molecules but in the vicinity of the N-terminal Pro20 in ATIII. There are 14 insertions and one deletion when ATIII is compared to A1AT. After superposition of both X-ray structures, the INDEL regions were located and analysed and the sequence alignment between A1AT and ATIII (Huber & Carrell, 1989) was revised (Appendix II). Insertion of Asn74 (N47a) occurs in a peculiar sequence stretch where three asparagine and one aspartic acid residues are found. Several INDEL locations are not

Table 10

Transformation from A1AT to ATIII (molecule A)

Rotation matrix†			Translation *T*
0·68998	0·71097	−0·13579	68·064
−0·56130	0·64401	0·51980	39·412
0·45701	−0·28244	0·84343	180·458

† *x*ATIII (Å) = Matrix × *x*A1AT (Å) + *T* (Å).

Figure 10. C^{α}-overlay of the refined structure of ATIII (thick line) and MOD1 (thin line).

in agreement with the sequence alignment. The most striking differences are about the proposed insertions at positions 210-211 (179a-179b), 351 (317a) and 360-361 (325a-325b) which actually occur at positions 206-207 (175a-175b), 348 (314a) and 357-358 (322a-322b). In addition, segment 204 to 207 (174-175b) is shifted by about 3·0 Å and loop 356 to 361 (322 to 325) is in a different conformation. All those INDEL appear in a connecting loops or tight turns and preserve integrity of secondary structures. Insertion of Thr111 (T83a) occurs at the C-terminal end of hC1. This helix is formed by residues 108 to 112 (81 to 84) in ATIII and 110 to 115 (83 to 87) in A1AT. This motif together with residues 113-114 (85-86) are shifted towards helix A. This movement helps packing of the N-terminal segment to the protein core. As a consequence, Tyr24 phenol group and residues 31-32 (5-6) in ATIII, respectively, occupy the positions of Leu112 (L84) side-chain and of residues 113-114 (85-86) of A1AT. Those differences are highly relevant for the functionality of the ATIII molecule because they affect Lys115 (K87), which has been shown to be involved in the heparin binding site (Liu & Chang, 1987). Helix E is flanked in ATIII by the N-terminal segment 16 to 21. The side-chains of the lysine residues 158 (K129) and 165 (K136) in A1AT would interfere with Ile16 and Pro20. They are mutated to threonine and valine in ATIII. Helix H is shorter and shifted by about 6 Å as it is the case in A1ACT (Baumann *et al.*, 1991). It is well defined in both molecules of the asymmetric unit. Crystal packing constraints do not explain the shift of helix H but explain the different location of the following region: residues 312-313 (279-280). Strand 1C is shorter in ATIII and residues Asn397, Ser398 and Asp399 (360a to c) are inserted just before it. They are only visible in molecule B.

(g) *Comparison of the refined structure of ATIII and MOD1*

MOD1, the molecular replacement solution, whose refinement could not be pursued below an *R*-factor of 0·28 has been compared to the final structure reported here. Figure 10 shows an overlay of their C^{α} backbones. As previously mentioned, a large number of differences between both structures occurs and only 75% of the residues in MOD1 were in agreement with the 3·2 Å SIR averaged map. The errors were about: (1) INDEL locations between A1AT and ATIII, (2) loop regions, (3) shift of secondary structure elements, (4) the N-terminal part. The root-mean-square deviation for all C^{α} atoms between both structures is 3·7 Å while a value of 1·2 Å is found when regions corresponding to items (3) (residues 109 to 116, 265 to 314) and (4) are omitted. In our case, it is clear that refinement of the molecular replacement solution by molecular dynamics did not converge to the correct structure which may be due to the absence of high resolution diffraction data. A critical analysis of the reduction of model bias in macromolecular structure determination appeared recently (Hodel *et al.*, 1992).

4. Discussion

The 3·2 Å resolution refined structure of cleaved bovine ATIII displays a topology similar to that of

post-complexed A1AT (Loebermann *et al.*, 1984), A1ACT and LEI (Baumann *et al.*, 1991, 1992) and confirms the unusual and drastic structural transition that results upon cleavage. Cleavage of molecular species came as a surprise during the crystallization trials of ATIII and sequence analysis of dissolved crystals showed that it had occurred at the P1′–P2′ peptide bond (Samama *et al.*, 1989). Residue Leu396 (L360), the P2′ site, is not visible in the electron density maps but there is no doubt about the positions of the residues delimiting the cleavage region which are found at opposite ends of the molecule. This rearrangement after cleavage, the so-called S → R transition which causes the complete insertion of strand 4A as an additional strand in β-sheet A, is a characteristic feature of inhibitory serpins (for a review, see Huber & Carrell, 1989) and has not been found in the case of the non-inhibitory serpin ovalbumin as demonstrated by the crystal structures of plakalbumin (Wright *et al.*, 1990) and intact ovalbumin (Stein *et al.*, 1990, 1991). In the latter, the residues corresponding to strand 4A fold into a helix over β-sheet C at the surface of the protein.

As no atomic structure of an uncleaved inhibitory serpin was available, a current working hypothesis was that ovalbumin provides a model for intact serpins in general (Stein & Chothia, 1991). Nevertheless, peptide annealing experiments (Schulze *et al.*, 1990) and/or investigations on mutants of A1AT and C1̄ inhibitor (Schulze *et al.*, 1991; Skriver *et al.*, 1991) provide evidence that the inhibitory conformation includes the partial insertion of the reactive loop into sheet A. Furthermore, it has been recently shown that insertion of a single residue, P14 (T345 with A1AT sequence and numbering), is sufficient for the inhibitory action of A1AT (Schulze *et al.*, 1992).

The case of ATIII is a more complex one: this serpin exhibits different conformational states, in its uncleaved form, which are under the control of glycosaminoglycans like heparin or heparan sulphate. The situation can be summarized as follows: in plasma, under normal conditions, ATIII is in a quiescent form (Q) if unbound to any glycosaminoglycan, or in a stressed form (S) if bound to heparan sulphate (at the surface of endothelial cells) or heparin. The S form, represented by the partially inserted canonical conformation, corresponds to the serpins active inhibitory form. The structural difference between the Q and S forms most probably lies in the degree of insertion of the reactive site loop (Carrell & Evans, 1992). When coagulation factors (factor Xa, thrombin) are generated, S-ATIII is recognized and it complexes to these proteinases. After inhibition, ATIII is finally released in the cleaved, relaxed, R form in which the P1 to P14 residues are fully inserted in the A sheet. Another conformational state, L (for locked), apparently not relevant for the biological function of ATIII, has been revealed by Carrell *et al.* (1991) after exposure of the protein to dilute guanidinium chloride. Its properties resemble cleaved R-ATIII and its conformation results from the insertion of the loop, beyond that of the inhibitory conformation, which can be represented by the extreme case of latent (and inactive) PAI-1 (Mottonen *et al.*, 1992).

Heparin binds to ATIII through a well defined pentasaccharide sequence that we have synthesized (Petitou *et al.*, 1987). The interaction of this compound and ATIII has been studied by stop-flow kinetics (Olson *et al.*, 1992). In a first step the pentasaccharide and Q-ATIII recognize each other and assemble with low affinity (20 to 30 μM) but a rapid conformational change (520 to 700 s^{-1}) then leads to S-ATIII which has a much higher affinity for the pentasaccharide (K_D = 30 to 50 μM). Thus the present data, relative to R-ATIII, are not *a priori* pertinent for the study of the recognition of heparin which would require the structure of uncleaved Q-ATIII. However, they allow us to draw some conclusions concerning the interaction of heparin and ATIII.

(1) The residues known, from biochemical experiments, to be involved in the interaction (Koide *et al.*, 1984; Owen *et al.*, 1987; Liu & Chang, 1987; Peterson *et al.*, 1987; Borg *et al.*, 1988; Chang, 1989; Smith & Knauer, 1987) were mapped on the X-ray structure of A1AT onto which the sequence of ATIII had been introduced (Huber & Carrell, 1989; Carrell & Evans, 1992; Grootenhuis & van Boeckel, 1991). The structure of R-ATIII presented here confirms the main features of the heparin binding site of ATIII deduced from the molecular modelling approach.

(2) The A1AT molecule, and consequently the derived models of ATIII, lack the 45 residue N-terminal region. This region, although displaying high isotropic temperature factors, has a clear chain tracing, except for nine disordered residues. It defines a valley which restrains the possibilities for binding of the pentasaccharide. We thus attempted to position the pentasaccharide unit, in its conformations deduced from ^{1}H-NMR spectroscopy (Ragazzi *et al.*, 1990), on the protein. Side-chains of

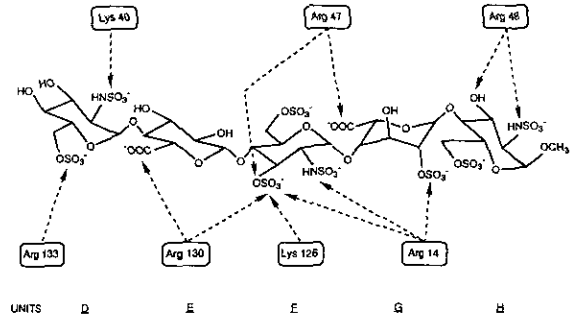
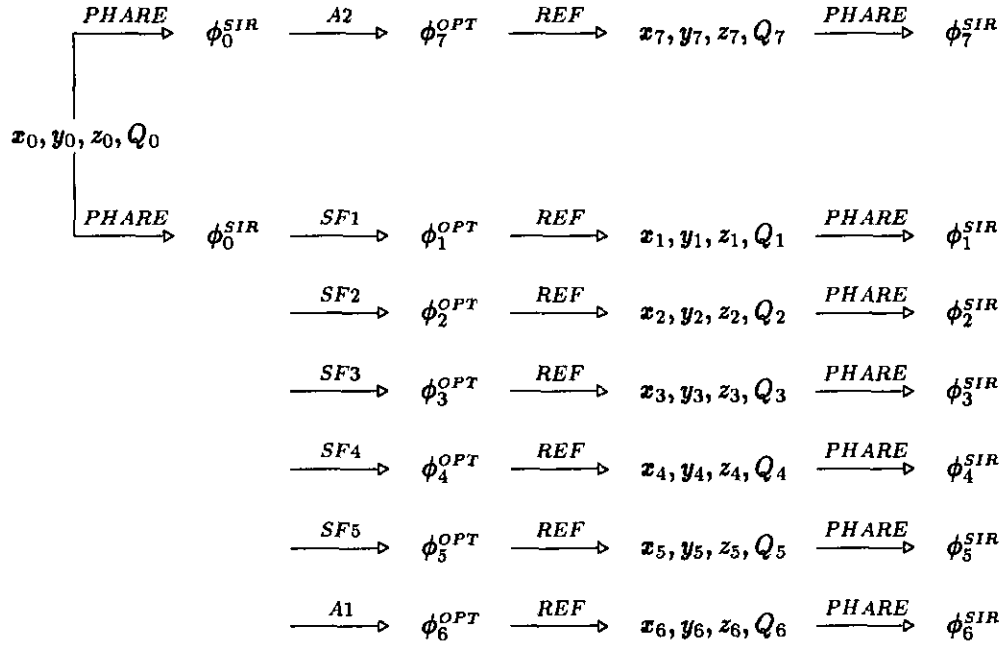


Figure 11. Schematic diagram of the interactions network in the ATIII-pentasaccharide complex. Interactions between basic residues of the protein (bovine ATIII sequence numbering) with sulphate and carboxylate groups of the oligosaccharide are denoted by arrows. Monosaccharidic subunits are labelled D to H according to Petitou *et al.* (1988).

essential heparin-binding amino acids as well as other basic residues found in their vicinity were used as guides in positioning the oligosaccharide. No attempts were made to minimize the energy of the docked structure. Figure 11 shows the result of our investigation. Thus in our view this latter first interacts through the sulphates and the carboxylate of subunits D, E and F, with the positively charged residues Lys126 (K98), Arg130 (R102) in helix D, Arg133 (R105) in the loop between hD and s2A and Arg14 (R-13) in the N-terminal part of the protein. Further binding involves units G and H, at the reducing end of the oligosaccharide, and the loop preceding the N-terminal part of helix A. Longer oligosaccharides would have additional interactions with other basic amino acids found in the area already stressed by Carrell & Evans (1992). It is worth mentioning that the interactions depicted in Figure 11 can only occur altogether if one slightly distorts the pentasaccharide molecule at the level of the iduronic acid unit which constitutes a hinge between the two parts (DEF on one side and GH on the other side) of the pentasaccharide (Petitou *et al.*, 1988). This is in agreement with recent findings, based on enzyme kinetics studies, revealing a two-step binding process for heparin and analogues (Olson *et al.*, 1992). Compared to the model of interaction proposed by Grootenhuis & van Boeckel (1991) we found that, taking into account interaction points within the N-terminal domain of the protein, requires different contact points at both the protein and the pentasaccharide sides.

The conformational change triggered by the binding of the pentasaccharide to ATIII (Q → S) leads to a considerable reinforcement of the interaction which, one can imagine, reflects an increase in the number of contact points between the two molecules, as indicated above. Thus, on the one hand, the first event (recognition) could well be described using the structure of a molecule that has only low affinity for the pentasaccharide. However, on the other hand, one must keep in mind that the form "seen" by the pentasaccharide, and bonded with low affinity, is the Q form. We do not know to which extent its structure is related to the structure of the present R-form. As to the structure of the S-form of ATIII, it will only be definitively settled after one has obtained the three-dimensional structure of the pentasaccharide–ATIII complex.

Appendix I: Density Modification and Heavy-atom Parameters Refinement Scheme Between 15·0 and 4·5 Å Resolution



ϕ_i^{SIR}, SIR phases (ϕ^{best}) calculated from heavy-atom parameters with program PHARE at step i.

ϕ_i^{OPT}, optimized phases at step i.

SF_i, solvent flattening.

Ai, non-crystallographic symmetry averaging.

REF, heavy-atom parameters (x_0, y_0, z_0, Q_0) refinement.

In protocol 1, non-crystallographic symmetry averaging (A1) was applied as the final step of density modification after iterative solvent flattening and heavy-atom parameters refinement (SF1 → SF5). In protocol 2, it was performed as the only step of density modification (A2).

Appendix II: Sequence Alignment from X-ray Structures Superposition

Human A1AT: line 1, A1AT numbering; line 2, secondary structure assignment; line 3, amino acid sequence. Bovine ATIII: line 4, amino acid sequence; line 5, secondary structure assignment; line 6, ATIII numbering. The symbol ~~~~~~ indicates loop regions where INDEL occurs. The residues at each end are nearly superimposed.

We thank A. Mitschler for help in data collection and J. Cavarelli for the film data processing programs. We are grateful to B. Rees for assistance with the use of the RMOL package and for helpful discussions. R. Ripp is gratefully acknowledged for his expertise in computing. We also thank the scientific staffs of LURE (Orsay, France) and CHESS (Ithaca, NY, U.S.A.) for excellent data collection facilities. L.M. thanks the Société de Secours des Amis des Sciences for a post-doctoral grant. Co-ordinates have been deposited at the Brookhaven Data Bank and are available directly from the authors on request until they have been processed and released.

References

Abildgaard, U., Lie, M. & Odegard, O. R. (1977). Antithrombin (heparin cofactor) assay with "new" chromogenic substrates (S2238 and Chromozym TH). *Thromb. Res.* **11**, 549–556.

Baumann, U., Huber, R., Bode, W., Grosse, D., Lesjak, M. & Laurell, C. B. (1991). Crystal structure of cleaved human α_1-antichymotrypsin at 2·7 Å resolution and its comparison with other serpins. *J. Mol. Biol.* **218**, 595–606.

Baumann, U., Bode, W., Huber, R., Travis, J. & Potempa, J. (1992). Crystal structure of cleaved equine leucocyte elastase inhibitor determined at 1·95 Å resolution. *J. Mol. Biol.* **226**, 1207–1218.

Björk, I., Danielsson, Å., Fenton II, J. W. & Jörnvall, H. (1981). The site in human antithrombin for functional proteolytic cleavage by human thrombin. *FEBS Letters*, **126**, 257–260.

Björk, I., Olson, S. T. & Shore, J. D. (1989). Molecular mechanisms of the accelerating effect of heparin on the reactions between antithrombin and clotting proteinases. In *Heparin: Chemical and Biological Properties, Clinical Applications* (Lane, D. A. & Lindahl, U., eds), pp. 229–255, Edward Arnold, London.

Blow, D. M. & Matthews, B. W. (1973). Parameter refinement in the multiple isomorphous replacement method. *Acta Crystallogr. sect A*, **29**, 56–62.

Borg, J. Y., Owen, M. C., Soria, C., Caen, J. & Carrell, R. W. (1988). Proposed heparin binding site in antithrombin based on arginine 47. *J. Clin. Invest.* **81**, 1292–1296.

Bricogne, G. (1974). Geometric sources of redundancy in intensity data and their use for phase determination. *Acta Crystallogr. sect. A*, **30**, 395–405.

Bricogne, G. (1976). Methods and programs for direct-space exploitation of geometric redundancies. *Acta Crystallogr. sect. A*, **32**, 832–847.

Brünger, A. T. (1988). Crystallographic refinement by simulated annealing. Application to a 2·8 Å resolution structure of aspartate aminotransferase. *J. Mol. Biol.* **203**, 803–816.

Brünger, A. T. (1990). *X-PLOR Manual*, version 2.1. The Howard Hughes Medical Institute and Department of Molecular Biophysics and Biochemistry, Yale University, New Haven, U.S.A.

Brünger, A. T., Krukowski, A. & Erickson, J. W. (1990). Slow-cooling protocols for crystallographic refinement by simulated annealing. *Acta Crystallogr. sect. A*, **46**, 585–593.

Buehner, M., Ford, G. C., Moras, D., Olsen, K. W. & Rossmann, M. G. (1974). Structure determination of crystalline lobster D-glyceraldehyde-3-phosphate dehydrogenase. *J. Mol. Biol.* **82**, 563–585.

Carrell, R. & Travis, J. (1985). α_1-Antitrypsin and the serpins: variation and countervariation. *Trends Biochem. Sci.* **10**, 20–24.

Carrell, R. W. & Evans, D. L. (1992). Serpins: mobile conformations in a family of proteinase inhibitors. *Curr. Opin. Struct. Biol.* **2**, 438–446.

Carrell, R. W., Pemberton, P. A. & Boswell, D. R. (1987). The serpins: evolution and adaptation in a family of protease inhibitors. *Cold Spring Harbor Symp. Quant. Biol.* **52**, 527–535.

Carrell, R. W., Evans, D. L. & Stein, P. E. (1991). Mobile reactive centre of serpins and the control of thrombosis. *Nature (London)*, **353**, 576–578.

CCP4 (1979). The SERC (UK) Collaborative Computing Project No. 4, a suite of programs for protein crystallography, distributed from Daresbury Laboratory, Warrington WA4 4AD, U.K.

Chang, J. Y. (1989). Binding of heparin to human antithrombin III activates selective chemical modification at lysine 236. Lys107, Lys125 and Lys136 are situated within the heparin-binding site of antithrombin III. *J. Biol. Chem.* **264**, 3111–3115.

Cullis, A. F., Muirhead, H., Perutz, M. F., Rossmann, M. G. & North, A. C. T. (1961). Structure of haemoglobin. VIII. A three-dimensional Fourier synthesis at 5·5 Å resolution; determination of the phase angles. *Proc. Roy. Soc. ser. A*, **265**, 15–38.

Cura, V., Podjarny, A. D., Khrishnaswamy, S., Rees, B., Rondeau, J. M., Tete, F., Mourey, L., Samama, J. P. & Moras, D. (1991). Heavy atom refinement against solvent-flattened and local-symmetry averaged phases. In: *Isomorphous Replacement and Anamolous Scattering. Proceedings of the Daresbury study weekend, 25–26 January 1991* (Science and Engineering Research Council of the Daresbury Laboratory, ed.), pp. 107–115.

Damus, P. S., Hicks, M. & Rosenberg, R. D. (1973). Anticoagulant action of heparin. *Nature (London)*, **246**, 355–357.

Delarue, M., Samama, J. P., Mourey, L. & Moras, D. (1990). Crystal structure of bovine antithrombin III. *Acta Crystallogr. sect. B*, **46**, 550–556.

Dodson, E. J. (1976). A comparison of different heavy atom refinement procedures. In *Crystallographic Computing Techniques* (Ahmed, F. R., ed.), pp. 259–268, Munksgaard, Copenhagen.

Evans, P. (1983). The refinement of heavy atom sites. *Informat. Quart. Protein Crystallogr.* **11**, 15–27.

Fish, W. W., Orre, K. & Björk, I. (1979). The production of an inactive form of antithrombin through limited proteolysis by thrombin. *FEBS Letters*, **98**, 103–106.

Franzén, L. E., Svensson, S. & Larm, O. (1980). Structural studies on the carbohydrate portion of human antithrombin III. *J. Biol. Chem.* **255**, 5090–5093.

Grootenhuis, P. D. J. & van Boeckel, C. A. A. (1991). Constructing a molecular model of the interaction between antithrombin III and a potent heparin analogue. *J. Amer. Chem. Soc.* **113**, 2743–2747.

Hodel, A., Kim, S.-H. & Brünger, A. T. (1992). Model bias in macromolecular structures. *Acta Crystallogr. sect. A*, **48**, 851–858.

Hoylaerts, M., Owen, W. G. & Collen, D. (1984). Involvement of heparin chain length in the heparin-catalyzed inhibition of thrombin by antithrombin III. *J. Biol. Chem.* **259**, 5670–5677.

Huber, R. & Carrell, R. W. (1989). Implications of the three-dimensional structure of α_1-antitrypsin for structure and function of serpins. *Biochemistry*, **28**, 8951–8966.

Hunt, L. T. & Dayhoff, M. O. (1980). A surprising new protein superfamily containing ovalbumin, antithrombin III, and α_1-proteinase inhibitor. *Biochem. Biophys. Res. Commun.* **95**, 864–871.

Johnson, J. E. (1978). Averaging of electron density maps. *Acta Crystallogr. sect. B*, **34**, 576–578.

Jones, T. A. (1978). A graphics model building and refinement system for macromolecules. *J. Appl. Crystallogr.* **11**, 268–272.

Kabsch, W. & Sander, C. (1983). Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical structures. *Biopolymers*, **22**, 2577–2637.

Koide, T., Odani, S., Takahashi, K., Ono, T. & Sakuragawa, N. (1984). Antithrombin III Toyama: replacement of arginine-47 by cysteine in hereditary abnormal antithrombin III that lacks heparin-binding ability. *Proc. Nat. Acad. Sci., U.S.A.* **81**, 289–293.

Kurachi, K., Schmer, G., Hermodson, M. A., Teller, D. C. & Davie, E. W. (1976). Characterization of human, bovine and horse antithrombin III. *Biochemistry,* **15**, 368–373.

Lane, D. A. & Caso, R. (1989). Antithrombin: structure, genomic organization, function and inherited deficiency. *Bailière's Clin. Haematol.* **2**, 961–998.

Leslie, A. G. W. (1987). A reciprocal-space method for calculating a molecular envelope using the algorithm of B.C. Wang. *Acta Crystallogr. sect. A,* **43**, 134–136.

Leslie, A. G. W. (1988). A reciprocal-space algorithm for calculating molecular envelope using the algorithm of B. C. Wang. In *Improving protein phases, Proceedings of the Daresbury study weekend, 5–6 February 1988* (Bailey, S., Dodson, E. & Phillips, S., eds), pp. 25–31. SERC Daresbury Laboratory.

Liu, C. S. & Chang, J. T. (1987). The heparin binding site of human antithrombin III. Selective chemical modification at Lys^{114}, Lys^{125}, and Lys^{287} impairs its heparin cofactor activity. *J. Biol. Chem.* **262**, 17356–17361.

Loebermann, H., Tokuoka, R., Deisenhofer, J. & Huber, R. (1984). Human α_1-proteinase inhibitor. Crystal structure analysis of two crystal modifications, molecular model and preliminary analysis of the implications for function. *J. Mol. Biol.* **177**, 531–556.

Luzatti, P. V. (1952). Traitement statistique des erreurs dans la détermination des structures cristallines. *Acta Crystallogr.* **5**, 803–810.

Matthews, B. W. (1968). Solvent content of protein crystals. *J. Mol. Biol.* **33**, 491–497.

Matthews, B. W. & Czerwinski, E. W. (1975). Local scaling: a method to reduce systematic errors in isomorphous replacement and anomalous scattering measurements. *Acta Crystallogr. sect. A,* **31**, 480–487.

Mejdoub, H., Le Ret, M., Boulanger, Y., Maman, M., Choay, J. & Reinbolt, J. (1991). The complete amino acid sequence of bovine antithrombin (ATIII). *J. Protein Chem.* **10**, 205–212.

Mottonen, J., Strand, A., Symersky, J., Sweet, R. M., Danley, D. E., Geoghegan, K. F., Gerard, R. D. & Goldsmith, E. J. (1992). Structural basis of latency in plasminogen activator inhibitor-1. *Nature (London),* **355**, 270–273.

Mourey, L., Samama, J. P., Delarue, M., Choay, J., Lormeau, J. C., Petitou, M. & Moras, D. (1990). Antithrombin III: structural and functional aspects. *Biochimie,* **72**, 599–608.

Olson, S. T., Björk, I., Sheffer, R., Craig, P. A., Shore, J. D. & Choay, J. (1992). Role of the antithrombin-binding pentasaccharide in heparin acceleration of antithrombin-proteinase reactions. Resolution of the antithrombin conformational change contribution to heparin rate enhancement. *J. Biol. Chem.* **267**, 12528–12538.

Owen, M. C., Borg, J. Y., Soria, C., Soria, J., Caen, J. & Carrell, R. W. (1987). Heparin binding defect in a new antithrombin III variant: Rouen, 49 Arg to His. *Blood,* **69**, 1275–1279.

Peterson, C. B., Noyes, C. M., Pecon, J. M., Church, F. C. & Blackburn, M. N. (1987). Identification of a lysyl residue in antithrombin which is essential for heparin binding. *J. Biol. Chem.* **262**, 8061–8065.

Petitou, M., Duchaussoy, P., Lederman, I., Choay, J., Jacquinet, J. C., Sinaÿ, P. & Torri, G. (1987). Synthesis of heparin fragments: a methyl α-pentaoside with high affinity for antithrombin III. *Carbohydrate Res.* **167**, 67–75.

Petitou, M., Lormeau, J. C. & Choay, J. (1988). Interaction of heparin and antithrombin III. The role of *O*-sulfate groups. *Eur. J. Biochem.* **176**, 637–640.

Priestle, J. P. (1988). RIBBON: a stereo cartoon drawing program for proteins. *J. Appl. Crystallogr.* **21**, 572–576.

Ragazzi, M., Ferro, D. R., Perly, B., Sinaÿ, P., Petitou, M. & Choay, J. (1990). Conformation of the pentasaccharide corresponding to the binding site of heparin for antithrombin III. *Carbohydrate Res.* **195**, 169–185.

Read, R. J. (1986). Improved Fourier coefficients for maps using phases from partial structures with errors. *Acta Crystallogr. sect. A,* **42**, 140–149.

Rees, B., Bilwes, A., Samama, J. P. & Moras, D. (1990). Cardiotoxin V_4^{II} from *Naja mossambrica mossambica*. The refined crystal structure. *J. Mol. Biol.* **214**, 281–297.

Richardson, J. S. (1981). The anatomy and taxonomy of protein structure. *Advan. Protein Chem.* **34**, 167–339.

Rould, M. A., Perona, J. J., Söll, D. & Steitz, T. A. (1989). Structure of *E. coli* glutaminyl-tRNA synthetase complexed with $tRNA^{Gln}$ and ATP at 2·8 Å resolution. *Science,* **246**, 1135–1142.

Rould, M. A., Perona, J. J. & Steitz, T. A. (1992). Improving multiple isomorphous replacement phasing by heavy atom refinement using solvent-flattened phases. *Acta Crystallogr. sect. A,* **48**, 751–756.

Samama, J. P., Delarue, M., Mourey, L., Choay, J. & Moras, D. (1989). Crystallization and preliminary crystallographic data for bovine antithrombin III. *J. Mol. Biol.* **210**, 877–879.

Schulze, A. J., Baumann, U., Knof, S., Jaeger, E., Huber, R. & Laurell, C. B. (1990). Structural transition of α_1-antitrypsin by a peptide sequentially similar to β-strand s4A. *Eur. J. Biochem.* **194**, 51–56.

Schulze, A. J., Huber, R., Degryse, E., Speck, D. & Bischoff, R. (1991). Inhibitory activity and conformational transition of α_1-proteinase inhibitor variants. *Eur. J. Biochem.* **202**, 1147–1155.

Schulze, A. J., Frohnert, P. W., Engh, R. A. & Huber, R. (1992). Evidence for the extent of insertion of the active site loop of intact α_1 proteinase inhibitor in β-sheet A. *Biochemistry,* **31**, 7560–7565.

Sim, G. A. (1959). The distribution of phase angles for structures containing heavy atoms. II. A modification of the normal heavy-atom method for non-centrosymmetrical structures. *Acta Crystallogr.* **12**, 813–815.

Skriver, K., Wikoff, W., Patston, P. A., Tausk, F., Schapira, M., Kaplan, A. P. & Bock, S. C. (1991). Substrate properties of $C\bar{1}$ inhibitor MA (Alanine 434 → Glutamic acid). Genetic and structural evidence suggesting that the P12-region contains critical determinants of serine protease inhibitor/substrate status. *J. Biol. Chem.* **266**, 9216–9221.

Smith, J. W. & Knauer, D. J. (1987). A heparin binding site in antithrombin III. Identification, purification, and amino acid sequence. *J. Biol. Chem.* **262**, 11964–11972.

Stein, P. & Chothia, C. (1991). Serpin tertiary structure transformation. *J. Mol. Biol.* **221**, 615–621.

Stein, P. E., Leslie, A. G. W., Finch, T., Turnell, W. G.,

McLaughlin, P. J. & Carrell, R. W. (1990). Crystal structure of ovalbumin as a model for the reactive centre of serpins. *Nature (London)*, **347**, 99–102.

Stein, P. E., Leslie, A. G. W., Finch, J. T. & Carrell, R. W. (1991). Crystal structure of uncleaved ovalbumin at 1·95 Å resolution. *J. Mol. Biol.* **221**, 941–959.

Stuart, D. & Artymiuk, P. (1984). The use of phase combination in crystallographic refinement: the choice of amplitude coefficients in combined syntheses. *Acta Crystallogr. sect. A*, **40**, 713–716.

Thaler, E. & Schmer, G. (1975). A simple two-step isolation procedure for human and bovine antithrombin II/III (heparin cofactor): a comparison of two methods. *Brit. J. Haematol.* **31**, 233–243.

Travis, J. & Salvesen, G. S. (1983). Human plasma proteinase inhibitors. *Annu. Rev. Biochem.* **52**, 655–709.

Wang, B. C. (1985). Resolution of phase ambiguity in macromolecular crystallography. *Methods Enzymol.* **115**, 90–112.

Waugh, D. F. & Fitzgerald, M. A. (1956). Quantitative aspects of antithrombin and heparin in plasma. *Amer. J. Physiol.* **184**, 627–639.

Wright, H. T., Qian, H. X. & Huber, R. (1990). Crystal structure of plakalbumin, a proteolytically nicked form of ovalbumin. Its relationship to the structure of cleaved α-1-proteinase inhibitor. *J. Mol. Biol.* **213**, 513–528.

Edited by R. Huber