

Structure of the Kunitz-Type Soybean Trypsin Inhibitor (STI): Implication for the Interactions Between Members of the STI Family and Tissue-Plasminogen Activator

PATRICE DE MEESTER,^a PETER BRICK,^a LESLEY F. LLOYD,^{a†} DAVID M. BLOW,^a SILVIA ONESTI^{a*}

^aBlackett Laboratory, Imperial College, London SW7 2BZ, England. E-mail: s.onesti@ic.ac.uk

(Received 21 July 1997; accepted 10 November 1997)

Abstract

The Kunitz-type soybean trypsin inhibitor (STI) has played a key role in the early study of proteinases, having been used as the main substrate in the biochemical and kinetic work that led to the definition of the standard mechanism of action of proteinase inhibitors. A partial structure of STI complexed with porcine trypsin has previously been reported, in which the first 93 residues of the inhibitor, including the region of contact with trypsin, were relatively well defined, whereas for the remaining part of the peptide chain only some C α atoms were located. The structure of the inhibitor in its free form has now been determined by molecular replacement to 2.5 Å, using the coordinates of the homologous *Erythrina* trypsin inhibitor as a search model. When the refined atomic coordinates of STI are compared with the partial model previously available, the conformation of the reactive-site loop and its position with respect to the main body of the molecule does not change when the inhibitor interacts with trypsin. There are instead, despite the high similarity in the overall tertiary structure, significant differences between STI and *Erythrina* trypsin inhibitor (ETI) in the region which is in contact with the enzyme in the STI:trypsin crystal structure. Some of these differences can explain the unique specificity of ETI and its ability to inhibit the fibrinolytic enzyme tissue-type plasminogen activator.

1. Introduction

Seeds from various Leguminosae are rich sources of inhibitors of trypsin and other proteinases. Their physiological role is not completely understood, but it has been proposed that they have a defensive function by inhibiting insect proteases and may be involved in the control of endogenous proteases during seed dormancy.

Proteinase inhibitors from soybean seeds can be classified into two different families, according to the molecular weight and the disulfide-bridge pattern: those of the Bowman–Birk type are small (with an M_r of about 8 kDa) with seven disulfide bridges, while those of the

Kunitz type have an M_r of about 20 kDa and two disulfide bridges. The Kunitz-type soybean trypsin inhibitor (STI) was first isolated and characterized by Kunitz (Kunitz, 1947*a,b*). It consists of 181 amino-acid residues and can be hydrolyzed by trypsin, the cleavage site being situated between Arg63 and Ile64 (Ozawa & Laskowski, 1966). This inhibitor was used as the main substrate for work that established the standard mechanism of action of serine proteinase inhibitors (Laskowski & Kato, 1980). The primary sequence was first determined by Koide & Ikenaka (1973); later Kim *et al.* (1985) identified and sequenced three genetically different variants of STI,‡ designated Ti^a, Ti^b, Ti^c. The sequence previously reported by Koide & Ikenaka corresponds to the Ti^a variant, with two amendments. The amino-acid sequences of a number of homologous seed proteins are known and have been classified as the ‘STI family’.

The structure of the complex between STI and porcine trypsin had previously been determined at 2.6 Å resolution (Blow *et al.*, 1974; Sweet *et al.*, 1974), but the quality of the diffraction data were poor and the molecular model for the inhibitor was not complete. Whereas atomic coordinates were determined and refined for residues 1–93, only a partial C α tracing could be defined for the remaining residues. Subsequently McLachlan (1979) showed that the structure displayed a pseudo threefold symmetry.

More recently the structure of one of the trypsin inhibitors from the seeds of the Leguminosae *Erythrina caffra* has been determined to 2.5 Å resolution by the isomorphous replacement method and refined to a crystallographic R factor of 20.8% (Onesti *et al.*, 1989, 1991). When this structure is compared with the incomplete STI model the two inhibitors, as could be expected from the high sequence homology [40% identity (Joubert & Dowdle, 1987) Fig. 1], have a similar tertiary structure but show significant differences in the conformation of the external loops. The crystal structure of the bifunctional α -amylase/subtilisin inhibitor from wheat (Zemke *et al.*, 1991) and of the chymotrypsin

‡ Abbreviations used: STI, soybean trypsin inhibitor; ETI, *Erythrina* trypsin inhibitor; r.m.s., root-mean-square; t-PA, tissue-type plasminogen activator; u-PA, urokinase-type plasminogen activator; PEG, polyethylene glycol.

† Present address: National Institute for Medical Research, The Ridgeway, Mill Hill, London NW7 1AA, England.

Table 1. *Data collection and processing for the triclinic P1 crystal from Na/K phosphate*

Resolution (Å)	Observed reflections	Completeness (%)	I/σ	R_{merge} (%)
6.32	690	95.0	11.0	7.3
5.00	949	96.7	11.1	6.0
4.26	1147	95.7	10.9	5.4
3.78	1341	97.2	10.7	5.2
3.43	1468	96.5	10.3	4.9
3.16	1593	96.5	9.8	6.1
2.95	1696	94.4	8.9	8.3
2.77	1717	92.5	7.9	10.9
2.63	1804	89.9	6.9	13.7
2.50	1914	88.6	6.1	18.1

$P2_12_12$. When sodium/potassium phosphate was used instead as a precipitating agent, irregularly shaped plates appeared. Some of these crystals have been characterized as belonging to space group $P1$ and others to space group $P2_12_12_1$ although it was not possible to distinguish the crystal system from the morphology. The two orthorhombic forms, $P2_12_12_1$ and $P2_12_12$, have very similar cell dimensions but the distinction from axial systematic absences between the two space groups is unambiguous.

A data set to 2.5 Å resolution was collected at 291 K on an Enraf–Nonius FAST area detector for a triclinic crystal grown from sodium/potassium phosphate with cell dimensions $a = 39.2$, $b = 54.0$, $c = 58.3$ Å, $\alpha = 71.7$, $\beta = 90.0$, $\gamma = 83.8^\circ$. Another data set to 2.8 Å resolution was collected for an orthorhombic $P2_12_12$ crystal obtained from sulfate with cell dimensions $a = 58.3$, $b = 95.8$ and $c = 39.2$ Å. Data were evaluated with the program *MADNES* (Messerschmidt & Pflugrath, 1987), corrected for Lorentz and polarization effects and scaled according to the method of Fox & Holmes (1966) to produce a set of unique structure-factor amplitudes using the *CCP4* suite of programs (Collaborative Computational Project, Number 4, 1994). The merging R factor for the triclinic $P1$ crystal form was 6.3% for the 14 369 reflections (93.7% completeness) up to 2.5 Å resolution (18.1% for the last resolution shell, Table 1). The R_{merge} for the orthorhombic data set was 7% (17.8% in the outer shell) to 2.8 Å resolution, including 4 859 reflections which correspond to 87.3% completeness.

2.3. Solution and refinement of the structure

Attempts to solve first the $P2_12_12$ structure by molecular replacement, using a hybrid model (consisting of residues 1–93 from STI in the STI–trypsin complex and, beyond residue 93, all corresponding ETI residues with non-homologous side chains replaced by alanine residues) gave an unambiguous solution for the orientation of the molecule. Nevertheless, translation-function calculations did not give a clear maximum along the direction of the 2_1 axis parallel to the a cell dimension. Therefore, it appeared likely that only one parameter

Table 2. *Refinement statistics*

Resolution (Å)	R factor	R free
15.0–4.95	0.144	0.198
4.95–3.95	0.107	0.174
3.95–3.46	0.139	0.242
3.46–3.14	0.162	0.189
3.14–2.92	0.191	0.237
2.92–2.75	0.216	0.306
2.75–2.61	0.249	0.331
2.61–2.50	0.282	0.293
15.0–2.50	0.160	0.223

out of the six derived from the molecular replacement method, a translation along a 2_1 axis parallel to the a axis, was not determined. The orthorhombic structure was abandoned at that stage but that partial result was used to solve the triclinic structure. The conservation in the triclinic cell of the arrangement of the two molecules along the 2_1 axis was suggested by the relationship between the cell parameters of the two crystal forms and was consistent with the fact that a self-rotation function (Crowther, 1972) calculated using the triclinic data clearly indicated that the two independent molecules were related by a twofold rotation parallel to the triclinic 58.3 Å c axis. The coordinates of a second molecule were generated by applying a rotation of 180° and a translation of 29.1 Å. A cross-rotation function was then performed using these two molecules as a search object in the triclinic cell. Triclinic structure factors were calculated for the two molecules placed in a large orthogonal cell of dimensions $100 \times 70 \times 80$ Å. The cross-rotation function was calculated as a function of the Eulerian angles on a 5° grid using data between 11 and 3.5 Å and Patterson radial limits of 7 and 25 Å. The highest interpolated peak corresponding to the solution had a value 6.2 times the r.m.s. level and was 2.2 times higher than the next highest peak.

This model was then refined against the triclinic diffraction data in the resolution range 15–2.5 Å with the program *X-PLOR* (Brünger *et al.*, 1987). A random sample containing 5% of the data was excluded from the refinement and the agreement between calculated and observed structure factors for those reflections (R free) was used to monitor the course of the refinement (Brünger, 1992). The model was restrained with Engh and Huber stereochemical parameters (Engh & Huber, 1991). Low-resolution data to 15 Å were included and a bulk solvent correction applied throughout the refinement procedure.

Model building was carried out using the program *FRODO* (Jones, 1978). Residues 21–25 and 83–85 had to be rebuilt and the order Gly26–Phe27 and Pro60–Ser61 reversed with respect to the original STI structure, according to the revised sequences (Kim *et al.*, 1985). A few residues in external loops appeared to be completely disordered and the chain could not be traced in the electron-density maps. The missing residues are 124–128 (loop B5–C1), 139–144 (loop C1–C2) and the C-terminal

residues 177–181 in molecule I, and residues 125–127, 139–142 and 177–181 in molecule II. Some disordered side chains on the surface of the molecule could not be located and have, therefore, been omitted from the model. The final crystallographic R factor is 16.0% for 14 369 unique reflections between 15 and 2.5 Å (Table 2). The r.m.s. deviations from ideality for bond lengths and angles are 0.007 Å and 1.42°, respectively. A total of 157 ordered water molecules have been assigned to peaks larger than 3σ in the difference Fourier map when the putative molecule could make at least one good hydrogen bond to the protein. All residues have their main-chain torsion angles within or close to allowed regions of the Ramachandran plot (Laskowski *et al.*, 1993; Fig. 2). Fig. 3 shows a region of the electron-density map calculated with coefficients ($3F_{\text{obs}} - 2F_{\text{calc}}$; α_{calc}) and the corresponding refined model. This and subsequent figures were generated using the program *MOLSCRIPT* (Kraulis, 1991) as modified by Esnouf (1997). Structure factors and atomic coordinates have been deposited in the Protein Data Bank.†

3. Results and discussion

3.1. The structure

The STI molecule consists of 12 antiparallel β -strands separated by irregular loops. Six of the strands form an antiparallel β -barrel and the top of the barrel is capped by the other six strands, symmetrically arranged in three pairs around the barrel axis (Fig. 4). The bottom of the barrel is closed by the loop containing the reactive site and by the long N-terminal loop, wrapped on itself. The C-terminus also lies at the same end of the barrel.

Despite the deficiencies of the STI model previously available (Sweet *et al.*, 1974) the topology was sufficiently clear to allow McLachlan (1979) to identify the approximate threefold symmetry around the barrel axis. The repeating unit (or subdomain) is made up of four consecutive β -strands interconnected by loops. In each subdomain strands 1 and 4 belong to the β -barrel, while strands 2 and 3 are hydrogen bonded and form the cap. The nomenclature chosen for the β -strands is based on the topological equivalence between strands belonging to different subdomains. Thus, the strands are called A1, A2, A3, A4, B1 . . . C4, the letter indicating the subdomain to which the strands belong and the number indicating topologically equivalent strands. There is no sequence homology between the topologically equivalent residues of the three subdomains.

In the past few years the fold observed in the STI (Kunitz) family has been found in a number of unrelated

proteins such as interleukin-1 α (Graves *et al.*, 1990), interleukin-1 β (Finzel *et al.*, 1989; Priestle *et al.*, 1989), interleukin receptor antagonist (Vigers *et al.*, 1994), the acidic and basic fibroblast growth factors (Zhu *et al.*, 1991) and the B chain of the plant toxin ricin (Rutenber *et al.*, 1991). The only common feature between these unrelated proteins is that they are all involved in recognition. This topology has been described as β -trefoil and its structural determinants have been analysed (Murzin *et al.*, 1992; Swindells & Thornton, 1993). It has been suggested that this fold arose from gene triplication, but the position of the introns in interleukin-1 does not correspond to the subdomain structure.

3.2. Molecular packing

The two independent molecules in the triclinic asymmetric unit are related by a local pseudo 2_1 axis parallel to the crystallographic c direction. A total of 165 C α atoms can be superposed with a r.m.s. displacement of 0.35 Å, with the major differences located around His154; when the residues between 153 and 155 are not included in the superposition, the r.m.s. difference in the C α position is reduced to 0.22 Å. The two crystallographically independent copies of the reactive-site loop make similar interactions with adjacent molecules in the crystal lattice. A similar situation also occurs in crystals of the homologous ETI (Onesti *et al.*, 1991) where the reactive-site loop is involved in strong intermolecular contacts. This might be explained by the fact that this is an external loop exposed to the solvent and is, therefore, likely to interact with other molecules, but at the same time is clamped in a rigid conformation, in contrast to the other external loops which are much

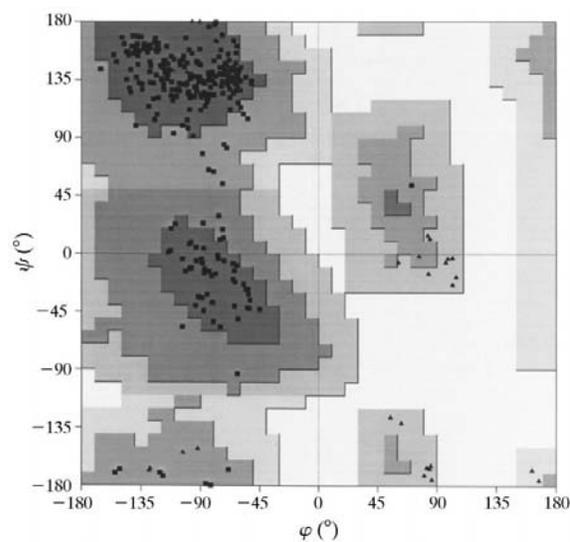


Fig. 2. Ramachandran plot of the main-chain torsion angles for the final model generated using the program *PROCHECK* (Laskowski *et al.*, 1993).

† Atomic coordinates and structure factors have been deposited with the Protein Data Bank, Brookhaven National Laboratory (Reference: 1BA7, R1BA7SF). Free copies may be obtained through The Managing Editor, International Union of Crystallography, 5 Abbey Square, Chester CH1 2HU, England (Reference: AD0031).

more flexible. The reactive-site loop offers, therefore, an ideal site for stable lattice contacts.

3.3. The reactive-site loop and comparison with the STI–trypsin complex

The reactive-site loop containing the scissile bond Arg63–Ile64 is located at the bottom of the molecule between strands A4 and B1, which belong to the β -barrel (Fig. 4). Although the overall quality of the Fourier map for this loop is good, there is only partial electron density for the side chain of Arg63 in both copies of the molecule. Despite the lack of disulfide bridges or strong electrostatic interactions, the reactive-site loop is maintained in a well ordered conformation by a network of hydrogen bonds involving the N-terminal loop (residues 1–14). A key role is played by the side chain of Asn13, which makes hydrogen bonds with both the main-chain carbonyl O atoms of Ile64 and Tyr62 and the amide N atom of Ser60, on either side of the scissile bond (Fig. 5). Asn13 is a highly conserved residue in the sequences of the inhibitors which belong to the STI family. One side of the reactive-site loop packs against a number of bulky side chains, including Tyr62, Ile64, Ile67 and His71.

In the electron-density map for the STI–trypsin complex (Sweet *et al.*, 1974) the trypsin molecule was clearly defined, but some parts of the inhibitor molecule were not clear, due to the poor quality of the diffraction data. Whereas atomic coordinates could be determined and partially refined for residues 1–93, only the $C\alpha$ -atom positions could be defined for residues 94–106 and 130–176, and a tentative assignment was made for residues 116–122. In addition the quality of the model was not very good and some residues showed energetically unfavourable torsion angles. The conformation of the chain in the refined uncomplexed STI model is very similar to the atomic model previously reported for the region 1–93. In particular, the conformation of the reactive-site loop and its position with respect to the main body of the inhibitor molecule is identical. All the

residues which lie at the trypsin interface in the crystal structure of the complexed STI (Asp1, Phe2, Asn13, Ser60, Pro61, Tyr62, Arg63, Ile64, Arg65, Phe66, His71) have very similar side-chain torsion angles in the two structures. A slightly different conformation of Asn13^{STI} in the complexed structure could be due to the interaction with Gln192^{Try}. In the crystal structure of the complex, the carbonyl group of Asp1^{STI} forms a salt bridge with Lys60^{Try}. The position and conformation of Asp1^{STI} is identical in the uncomplexed inhibitor, while the conformation of Lys60^{Try} is altered from that observed in other trypsin structures. The erroneous STI sequence Pro60–Ser61 reported by Koide & Ikenaka (1973) that was used by Sweet *et al.* (1974) to fit the electron density has been amended subsequently to Ser60–Pro61 (Kim *et al.*, 1985). The conformation of the side chain of Ser60 is such that at this resolution the two sequences are not easy to distinguish.

When the refined structure of STI is overlapped onto the $C\alpha$ tracing available beyond residue 93 of STI in the STI–trypsin complex, although the position of the main chain in the β -strands is correct, there are a number of places where the sequence was previously misaligned.

3.4. Comparison with ETI

Despite the high sequence and structural homology to STI, ETI has some unique properties. In addition to trypsin it has the ability to bind and inhibit tissue-type plasminogen activator (t-PA) (Onesti *et al.*, 1992), but not the homologous urokinase-type plasminogen activator (u-PA). Neither STI nor other homologous inhibitors are known to bind to t-PA or u-PA.

When the crystal structure of STI is overlapped onto the ETI structure, 129 $C\alpha$ positions can be superposed with an r.m.s. deviation of 0.86 Å (Fig. 5). Not only are all the β -strands conserved, but also several β -turns and β -bulges are present in topologically equivalent positions in the two structures. Most of the water molecules acting as bridges between β -strands are also located in similar positions.

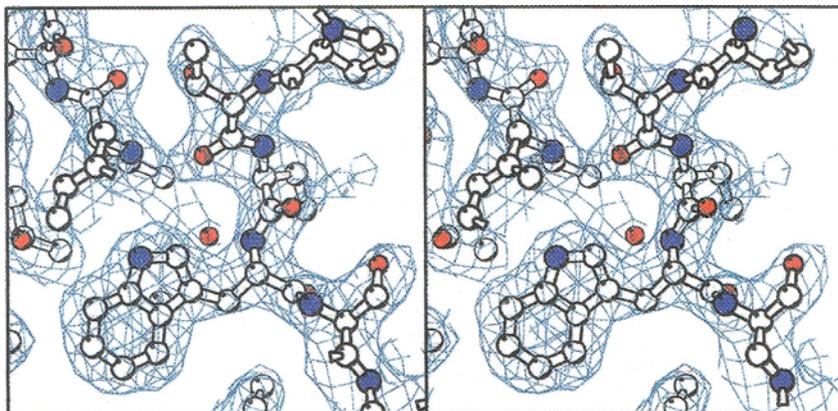


Fig. 3. A region of the final electron-density map calculated with coefficients ($3F_{\text{obs}} - 2F_{\text{calc}}$; α_{calc}), and the corresponding refined model.

Surprisingly, one of the regions where the structures of the two inhibitors show the largest differences is at the bottom of the molecule and comprises both the N-terminal loop (residues 1–14) and the loop between strands *A4* and *B1* (residues 59–72) containing the scissile bond Arg62–Ile63 (Fig. 6). At the N-terminus not only is STI one residue longer than ETI but also the main-chain conformation of the second residue is different. In the reactive-site loop, although the torsion angles for the residues close to the scissile bond are similar, residues 61 and 66 appear to be at the hinge of a movement that shifts the $C\alpha$ atom of the reactive-site

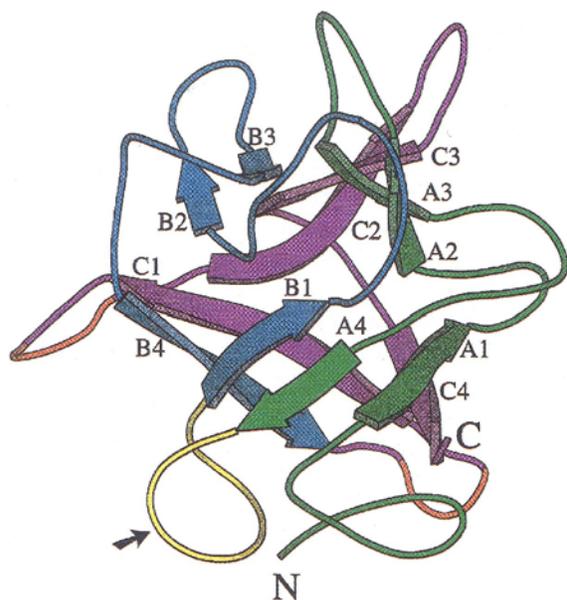


Fig. 4. A ribbon representation of the overall fold of the molecule. The pseudo threefold axis is vertical and corresponds to the axis of the barrel. The three subdomains *A*, *B*, *C* are shown in green, blue and magenta, respectively. In order to give a clear overview of the topology, the main chain has been modelled for the omitted residues (loops *B4*–*C1* and *C1*–*C2*, shown in orange). The reactive-site loop, between strands *A4* and *B1*, is shown in yellow with the scissile bond indicated by an arrow.

arginine 3.5 Å away from the corresponding ETI position (Fig. 7). The presence of Pro61 in STI affects the conformation of residues 62–65. The conformation of the side chain of Phe66 is also very different. The presence of a proline at position 68 in ETI might account for the observed 2 Å shift in the position of the β -bend between residues 68 and 71. When the crystal structure of ETI was determined and compared with the partial model available for the trypsin-bound STI, the same displacement of the reactive-site loop was observed, but it was not clear whether this difference resulted from the fact that the structure of STI had been determined in its complexed form or it was an intrinsic feature of ETI or was simply an artifact of the crystal contacts involving the reactive-site loop. The similarity between the conformation of the STI reactive-site loop in the complex and free form rules out the former hypothesis.

The conserved Asn13 (Asn12 in ETI) is in a very similar position, but makes different hydrogen bonds with the reactive-site loop because of the slightly different positions of these residues. In ETI the asparagine side-chain hydrogen bonds to the hydroxyl of Ser60 and to the main-chain carboxyl group of residue 62. Neither hydrogen bond is conserved in the STI structure and Asn13 instead makes interactions with other reactive-site loop residues (Fig. 5). The conformation of the scissile bond carbonyl group differs from that observed for ETI but is in agreement with that of other proteinase inhibitors. Due to the absence of bulky residues such as Tyr62, Ile64 and His71, one side of the reactive-site loop is more accessible in ETI than in STI.

The region between residues 108 and 117 (loop *B3*–*B4*) also adopts a rather different conformation in the two inhibitors.

3.5. Implications for binding to *t*-PA

Both the main-chain shifts and the differences in sequence and conformation of the residues at the interface with trypsin in the structure of the complexed

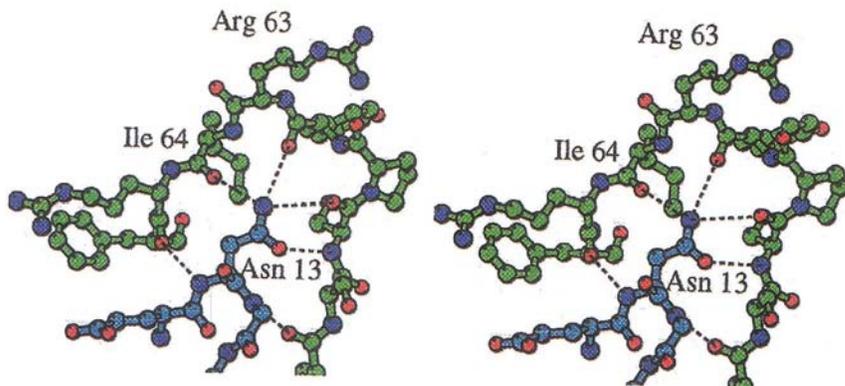


Fig. 5. Stereoview of the reactive-site loop of STI, with the region between residues 60 and 66 shown in green. The scissile bond is between Arg63 and Ile64. The hydrogen-bonding pattern involving Asn13 plays a key role in maintaining the reactive-site loop in the correct conformation.

STI are probably important in explaining the different specificities of the two inhibitors.

The crystal structure of the complex between porcine trypsin and STI was used as a template to study the interactions between both inhibitors and t-PA, whose crystal structure has been determined recently (Lamba *et al.*, 1996). A model has been built by docking the reactive-site loop of ETI into the active-site cleft of t-PA. For the *Erythrina* inhibitor two different loop conformations were compared: the one observed in the crystal structure and a hypothetical 'modified ETI' where the relative orientation of the reactive-site loop was similar to the corresponding loop in STI, to account for the possibility that the conformation of the loop in ETI could be an artifact resulting from the strong crystal lattice contacts. The model presented in Fig. 8 corresponds to the latter 'modified' inhibitor, but most of the interactions on the *P'* side are also possible when docking the original ETI crystal structure. The side-chain torsion angles have been modified for a few amino-acid residues, in order to optimize the enzyme-inhibitor interactions, and the side chain of Lys135,

which was disordered in the ETI crystal structure, has been built. No attempt has been made to energy minimize the model of the complex.

Although the absence of experimental structural data makes any attempt to understand the interactions between ETI and t-PA speculative, a number of observations can be made which could explain the specificity of the inhibitory activity of ETI (Fig. 8). In particular the different conformation of the loop *B3-B4* in ETI places it in the ideal position to interact with the loop 170-174 of t-PA. This loop assumes a different conformation because of a two-residue insertion in t-PA (residues 169A and 169B) with respect to trypsin. A salt bridge between Asp112^{ETI} and Arg174^{tPA} can be easily modelled. At the interface of the complex another potential bridge between Arg61^{ETI} and either Asp96^{tPA} or Asp97^{tPA} can easily be accommodated: the corresponding residues in STI and trypsin are Ser61^{STI}, Ser96^{Try}, Asn97^{Try}. The interaction between Tyr62^{STI} and Leu99^{Try} observed in the crystal structure of the STI-trypsin complex is replaced by the complementary interaction Tyr99^{tPA} and Leu62^{ETI}, while when STI and

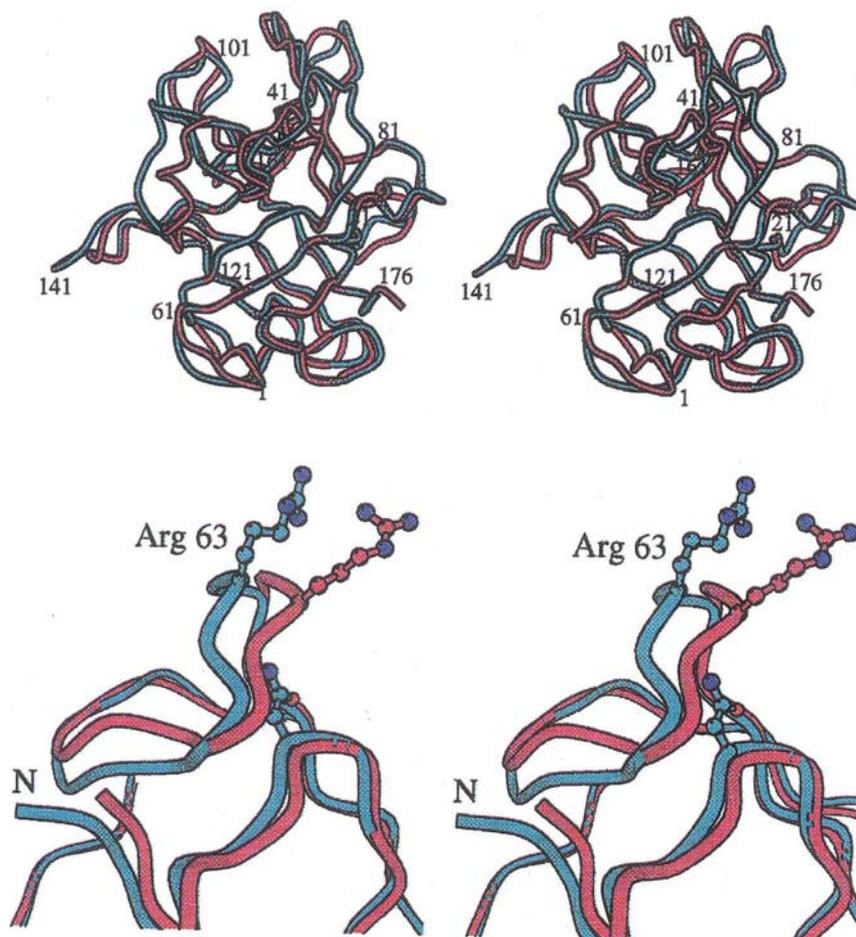


Fig. 6. Stereoscopic superposition of the α -atom backbones of STI (blue) and ETI (magenta). The molecules are shown with the threefold axis vertical and the numbering corresponding to the STI sequence.

Fig. 7. Stereoview of the α -atom backbone of STI (in blue) and ETI (in magenta) showing the different orientations of the reactive-site and the N-terminal loops. The positions of Arg63 and Asn13 are shown.

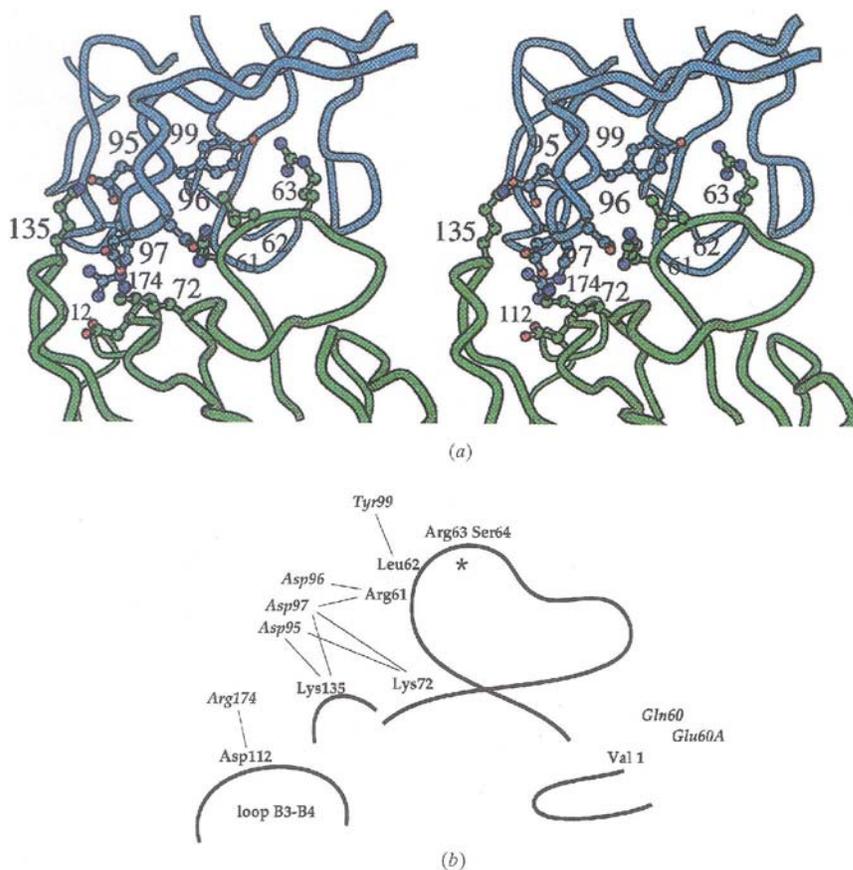


Fig. 8. (a) Stereoview of the modelled interface between ETI (in green) and t-PA (in blue). The position and orientation of the residues which may be responsible for the specificity of the interactions are shown. (b) A schematic diagram of the modelled interface between ETI and t-PA. The thick line represents the backbone of the inhibitor with an asterisk (*) indicating the position of the scissile bond. Residues belonging to t-PA are shown in italic; thin lines indicate the unique interactions between t-PA and ETI.

t-PA are docked there is a steric clash between Try99^{tPA} and Tyr62^{STI}. Site-directed mutagenesis studies carried out on recombinant ETI from the closely related *Erythrina variegata* (Kouzuma *et al.*, 1997) show that the mutants R61P and L62F both have a significant reduction in t-PA inhibitory activity while the double mutant R61P and L62F lacks any activity towards t-PA, despite retaining the ability to inhibit trypsin.

A systematic biochemical study of inhibitors from the seeds of different *Erythrina* species is reported by Joubert *et al.* (1987). The inhibitors can be classified in three groups (*a*, *b*, *c*) according to their different specificity. Only group *c* inhibitors bind and inhibit t-PA, although the scissile bond is Arg-Ser in all groups, and the *P'* sequences show high similarity. All the group *a* and *b* inhibitors have blocked amino-termini, while group *c* inhibitors show the same N-terminal amino acid (Val). Both STI and winged bean trypsin inhibitor have one extra residue (Glu or Asp) at the amino terminus with respect to *Erythrina* inhibitors. As mentioned previously, the N-terminal residue of STI (Asp1^{STI}) makes a salt bridge with Lys60^{Trv}. When the t-PA crystal structure is overlapped with that of trypsin, the insertion occurring after residue 60 in t-PA causes both Glu60^{tPA}

and Glu60A^{tPA} to sterically clash with Asp1^{STI} and Phe2^{STI}. ETI is one residue shorter than STI at the N-terminus and the conformation of Val1^{ETI} is such as to easily accommodate the insertion at residue 60 in t-PA. A modified form of ETI in which the N-terminal residue is mutated to an aspartic acid shows no inhibitory activity towards t-PA (Teixeira *et al.*, 1994).

References

- Blow, D. M., Janin, J. & Sweet, R. M. (1974). *Nature (London)*, **249**, 54–57.
- Brünger, A. T. (1992). *Nature (London)*, **355**, 472–474.
- Brünger, A. T., Kuriyan, J. & Karplus, M. (1987). *Science*, **235**, 458–460.
- Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* **D50**, 760–763.
- Crowther, R. A. (1972). *The Molecular Replacement Method*, edited by M. G. Rossmann, pp. 173–178. New York: Gordon & Breach.
- Dattagupta, J. K., Podder, A., Chakrabarti, C., Sen, U., Dutta, S. K. & Singh, M. (1996). *Acta Cryst.* **D52**, 521–528.
- Engh, R. A. & Huber, R. (1991). *Acta Cryst.* **A47**, 392–400.
- Esnouf, R. M. (1997). *J. Mol. Graphics*, **15**, 133–138.

- Finzel, B. C., Clany, L. L., Holland, D. R., Muchmore, S. W., Watenpaugh, K. D. & Einspahr, H. M. (1989). *J. Mol. Biol.* **209**, 779–791.
- Fox, G. C. & Holmes, K. C. (1966). *Acta Cryst.* **A20**, 886–891.
- Graves, B. J., Hatada, M. H., Hendrickson, W. A., Miller, J. K., Madison, V. S. & Satow, Y. (1990). *Biochemistry*, **29**, 2679–2684.
- Jones, T. A. (1978). *J. Appl. Cryst.* **11**, 268–272.
- Joubert, F. G., Merrifield, E. H. & Dowdle, E. B. (1987). *Int. J. Biochem.* **19**, 601–606.
- Joubert, F. J. & Dowdle, E. B. (1987). *Thromb. Haem.* **57**, 356–360.
- Kim, S. H., Hara, S., Hase, S., Ikenada, T., Toda, H., Kitamura, K. & Kaizuma, N. (1985). *J. Biochem.* **98**, 435–448.
- Koide, T. & Ikenaka, T. (1973). *Eur. J. Biochem.* **32**, 417–431.
- Kouzuma, Y., Yamasaki, N. & Kimura, M. (1997). *J. Biochem.* **121**, 456–463.
- Kraulis, P. J. (1991). *J. Appl. Cryst.* **24**, 946–950.
- Kunitz, M. (1947a). *J. Gen. Physiol.* **30**, 291–310.
- Kunitz, M. (1947b). *J. Gen. Physiol.* **30**, 311–320.
- Lamba, D., Bauer, M., Huber, R., Fischer, S., Rudolph, R., Kohnert, U. & Bode, W. (1996). *J. Mol. Biol.* **258**, 117–135.
- Laskowski, M. J. & Kato, I. (1980). *Ann. Rev. Biochem.* **49**, 593–626.
- Laskowski, R. A., MacArthur, M. V., Moss, D. S. & Thornton, J. M. (1993). *J. Appl. Cryst.* **26**, 283–291.
- Lee, J., Song, H. K., Hwang, K. Y., Kim, K. K. & Suh, S. W. (1993). *Mol. Cells*, **3**, 335–337.
- McLachlan, A. D. (1979). *J. Mol. Biol.* **133**, 557–563.
- Messerschmidt, A. & Pflugrath, J. W. (1987). *J. Appl. Cryst.* **20**, 306–315.
- Murzin, A. G., Lesk, A. M. & Chothia, C. (1992). *J. Mol. Biol.* **223**, 531–543.
- Onesti, S., Brick, P. & Blow, D. M. (1991). *J. Mol. Biol.* **217**, 153–176.
- Onesti, S., Lloyd, L. F., Brick, P. & Blow, D. M. (1989). *J. Mol. Biol.* **210**, 241–242.
- Onesti, S., Matthews, D. J., Aducci, P., Amiconi, G., Bolognesi, M., Menegatti, E. & Ascenzi, P. (1992). *J. Mol. Recogn.* **5**, 105–114.
- Ozawa, K. & Laskowski, M. Jr (1966). *J. Biol. Chem.* **241**, 3955–3961.
- Priestle, J. P., Schar, H. P. & Grutter, M. G. (1989). *Proc. Natl Acad. Sci. USA*, **86**, 9667–9671.
- Rutenber, E., Katzin, B. J., Ernst, S., Collins, E. J., Mlsna, D., Ready, M. P. & Robertus, J. D. (1991). *Proteins*, **10**, 240–250.
- Schechter, I. & Berger, A. (1967). *Biochem. Biophys. Res. Commun.* **27**, 157–162.
- Sweet, R. M., Wright, H. T., Chothia, C. H. & Blow, D. M. (1974). *Biochemistry*, **13**, 4212–4228.
- Swindells, M. B. & Thornton, J. M. (1993). *Protein Eng.* **6**, 711–715.
- Teixeira, A. V., Dowdle, E. B. D. & Botes, D. P. (1994). *Biochim. Biophys. Acta*, **1217**, 23–28.
- Vigers, G. P. K., Caffes, P., Evans, R. J., Thompson, R. C., Eisenberg, S. P. & Brandhuber, B. J. (1994). *J. Biol. Chem.* **269**, 12874–12879.
- Zemke, K. J., Muller-Fahrnow, A., Jany, K. D., Pal, G. P. & Saenger, W. (1991). *FEBS Lett.* **279**, 240–242.
- Zhu, X., Komiya, H., Chirino, A., Faham, S., Fox, G. M., Arakawa, T., Hsu, B. T. & Rees, D. C. (1991). *Science*, **251**, 90–93.