Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Structure of *Ecballium elaterium* trypsin inhibitor II (EETI-II): a rigid molecular scaffold

Ralph Krätzner,[a,b] Judit É. Debreczeni,[a]‡ Thomas Pape,[a] Thomas R. Schneider,[a]§ Alexander Wentzel,[b]¶ Harald Kolmar,[b]‡ George M. Sheldrick[a]* and Isabel Uson[a]‡‡

[a]Lehrstuhl für Strukturchemie, Georg-August-Universität Göttingen, Tammannstrasse 4, D37077 Göttingen, Germany, and [b]Institut für Mikrobiologie und Genetik, Georg-August-Universität Göttingen, Grisebachstrasse 8, D37077 Göttingen, Germany

‡ Present address: University of Oxford Structural Genomics Consortium, Botnar Research Centre, Oxford OX3 7LD, England.
§ Present address: FIRC Institute of Molecular Oncology Foundation, Via Adamello 16, 20139 Milan, Italy and European Institute of Oncology, Via Ripamonti 435, 20141 Milan, Italy.
¶ Present address: Selecore GmbH, Marie-Curie-Strasse 7, D37079 Göttingen, Germany.
‡‡ Present address: ICREA at Instituto de Biologia Molecular de Barcelona (IBMB-CSIC), Jordi Girona 18-26, 08034 Barcelona, Spain.

Correspondence e-mail: gsheldr@shelx.uni-ac.gwdg.de

Received 10 May 2005
Accepted 4 July 2005

PDB References: EETI-II, 1w7z, r1w7zsf; trypsin–EETI-II (wild type) complex, 1h9h, r1h9hsf; trypsin–EETI-II (mutant) complex, 1h9i, r1h9isf.

The *Ecballium elaterium* trypsin inhibitor II (EETI-II) belongs to the family of squash inhibitors and is one of the strongest inhibitors known for trypsin. The eight independent molecules of EETI-II in the crystal structure reported here provide a good opportunity to test the hypothesis that this small cystine-knot protein (knottin) is sufficiently rigid to be used as a molecular scaffold for protein-engineering purposes. To extend this test, the structures of two complexes of EETI-II with trypsin have also been determined, one carrying a four-amino-acid mutation of EETI-II. The remarkable similarity of these structures confirms the rigidity of the molecular framework and hence its suitability as a molecular scaffold.

1. Introduction

Rigid molecular 'scaffolds' are much sought after for protein engineering: they should enable the *de novo* design of proteins with specific properties without the complications introduced by conformational flexibility and the protein-folding problem (Skerra, 2000). Promising candidates for rigid molecular scaffolds include knottins (Le-Nguyen *et al.*, 1990), small proteins that share a common structural motif consisting of a cystine knot and a small triple-stranded antiparallel β-sheet. The lack of a hydrophobic core to provide structural stability is compensated by three disulfide bridges between the first and fourth, second and fifth, and third and sixth cysteines in the sequence. The first two disulfide bridges hold the polypeptide chain in a ring through which the third one passes (Fig. 1). This 'cystine knot' is found in a variety of small proteins from diverse organisms, and members of this family display diverse biological functionalities, mainly acting as inhibitors (Pallaghy *et al.*, 1994; Craik *et al.*, 2001). Examples of micro-proteins sharing this architecture are (i) ω-conotoxin MVIIa from the venom of the cone snail *Conus magus*, a 26-residue neurotoxin with a high affinity for voltage-gated Ca^{2+} channels (Kohno *et al.*, 1995), (ii) potato carboxypeptidase inhibitor (PCI), a 39-amino-acid peptide (Rees & Lipscomb, 1982), (iii) EETI-II from the jumping cucumber *Ecballium elaterium*, a 28-amino-acid member of the squash family of proteinase inhibitors (Favel *et al.*, 1989; Heitz *et al.*, 1989; Le-Nguyen *et al.*, 1990, 1993) and (iv) the cyclic 29-amino-acid inhibitor kalata B1 (Saether *et al.*, 1995).

Some members of the inhibitor cystine-knot family are extremely stable towards heat, digestion by proteases or denaturation by chaotropic agents. Recently, a systematic study on the thermal, chemical and enzymatic stability of cyclotide kalata B1 and variants was published (Colgrave &

Craik, 2004). It was shown that kalata B1 variants maintained their structural stability when subjected to heat, acids or chaotropic agents, even for acyclic mutants. In contrast, structural integrity was lost on guanidinium chloride treatment of a mutant with the first cystine changed to alanines, thereby undoing the cystine knot. The thermal stability of conotoxin PVIIA, also a cystine-knot protein, was found to be much lower than that of kalata B1. This result showed that also other structural features such as the strong hydrogen-bonding network of kalata B1 contribute to the overall stability of cystine-knot proteins. The thermal, chemical and enzymatic stability of kalata B1 might also be an indication of structural rigidity.

Biochemical assays by Christmann *et al.* (1999) have suggested that the cystine-knot protein EETI-II is indeed sufficiently rigid to be useful as a molecular scaffold for the presentation of combinatorial peptide libraries. EETI-II is one of the strongest inhibitors known for trypsin (binding constant $K_a = 8 \times 10^{11}\ M^{-1}$; Favel *et al.*, 1989) and folds spontaneously *in vitro* with proper formation of cystine bonds (Wentzel *et al.*, 1999). The fortuitous presence of eight independent molecules of recombinant EETI-II in the crystal structure reported here provides a good test of its conformational rigidity. We have also determined the structures of two of its complexes with trypsin for a further test of the conformational invariance. The sequences of the EETI-II variants used in this study are shown in Fig. 1. The variant EETI-II-β^{GPNG} is also considered as a wild type. A preliminary communication concerning the crystallization and diffraction properties of porcine trypsin in complex with EETI-II was published 16 years ago (Gaboriaud *et al.*, 1989), but no results of a structural analysis have been published or deposited to date.

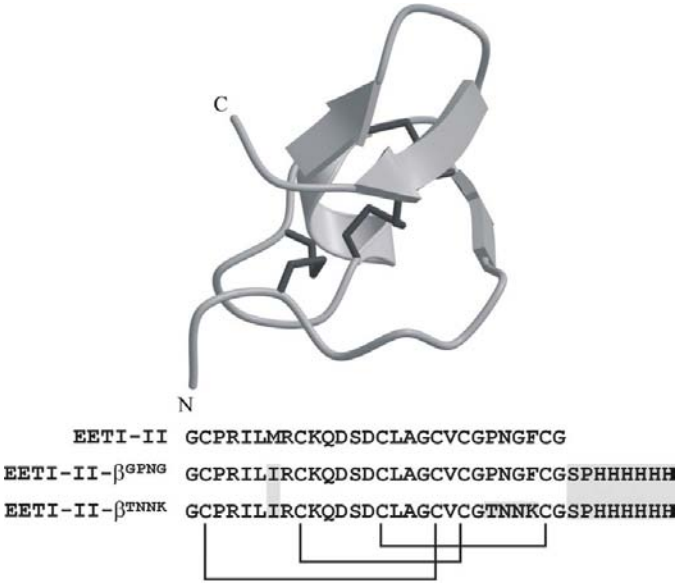

Figure 1
The cystine-knot motif in EETI-II and the sequences of the proteins investigated in this study. The sequence of EETI-II isolated from seeds of jumping cucumber *E. elaterium* was C-terminally extended by amino acids SPHHHHHH for affinity chromatography. The mutation M7I was introduced for purification purposes. Sequence deviations compared with the sequence of native EETI-II are highlighted in grey.

2. Materials and methods

2.1. Crystallization

Unless otherwise stated, chemicals were purchased from Fluka. Porcine trypsin (EC 3.4.21.4) was obtained from Sigma, Deisenhofen. Recombinant EETI-II wild type and variants were obtained by genetic engineering techniques as previously described (Wentzel *et al.*, 1999).

Uncomplexed EETI-II (wt) was crystallized at 277 K using the hanging-drop vapour-diffusion technique (McPherson, 1998). Extremely hygroscopic crystals (0.08 × 0.05 × 0.05 mm) could be obtained from droplets containing 1.5 µl protein solution (40 mg ml^{-1} EETI-II in double-distilled water) and 1.5 µl precipitant (4 *M* sodium formate) within several days.

Porcine trypsin was complexed with EETI-II in a molar ratio of 1:1.2. Complexation was performed in 100 m*M* bicine pH 8.0 for wild-type EETI-II and in 50 m*M* triethanolamine–HCl buffer pH 7.4 for the variants, using a total protein concentration of 10 mg ml^{-1}. Crystallization of the complexes was achieved by the hanging-drop vapour-diffusion method at 293 K. The drop was prepared by mixing equal volumes of the buffered complex solution and the reservoir solution, which contained 100 m*M* sodium citrate buffer pH 5.3, 20%(*v*/*v*) 2-propanol and 20%(*w*/*v*) polyethylene glycol 4000 for the wild-type complex, and 100 m*M* sodium citrate buffer pH 5.3, 100 m*M* triethanolamine–HCl buffer pH 7.4, 20%(*v*/*v*) 2-propanol and 20%(*w*/*v*) polyethylene glycol 4000 for the variants. Single crystals suitable for X-ray analysis grew within a few days. The crystals of the wild-type complex were octahedral with a maximum dimension of 0.25 mm; the trypsin–EETI-II-β^{TNNK} and trypsin–EETI-II-β^{NEDE} complexes gave rod-like crystals with a maximum size of 0.4 × 0.1 × 0.05 mm.

2.2. Data collection and processing

Native diffraction data were collected at beamline BL2 at BESSY in Berlin with a MAR345 image-plate detector using 2 *M* sodium malonate as cryoprotectant. Diffraction data for the trypsin–EETI-II-β^{GPNG} (wild type) complex were collected on beamline ID14-EH2 at ESRF, Grenoble using a cryoprotectant buffer composed of the reservoir solution plus 5% glycerol. Crystals of the remaining two complexes were mounted in sealed glass capillaries and data were collected at room temperature to 1.9 and 3.0 Å, respectively, with a MAR345 detector on a Siemens rotating-anode X-ray generator with Cu $K\alpha$ radiation and Osmic mirrors. Data were processed with *DENZO* and *SCALEPACK* (Otwinowski & Minor, 1997). The 3.0 Å data of the trypsin–EETI-II-β^{NEDE} complex were not analyzed further after we had established that the structure was isomorphous with that of the 1.9 Å trypsin–EETI-II-β^{TNNK} complex. Data quality and refinement statistics are summarized in Table 1.

2.3. Structure solution

All three structures were solved by molecular replacement. Porcine trypsin (Stubbs *et al.*, 1997; PDB code 1ldt) was used

Table 1
Data-collection and refinement statistics for EETI-II and trypsin–EETI-II complexes.

Values in parentheses are for the outer resolution shell.

	EETI-II-β^{GPNG}	Trypsin–EETI-II-β^{GPNG}	Trypsin–EETI-II-β^{TNNK}
Space group	*C*2	$P4_132$	*I*4
Unit-cell parameters (Å, °)	$a = 71.16$, $b = 82.34$, $c = 62.93$, $\beta = 118.8$	$a = b = c = 121.05$	$a = b = 139.86$, $c = 33.67$
Temperature of data collection (K)	100	100	293
Resolution (Å)	1.68 (1.68–1.80)	1.5 (1.5–1.6)	1.9 (1.9–2.0)
Unique reflections	35516	49533	26024
Completeness (%)	96.8 (99.0)	99.7 (99.6)	99.1 (97.9)
Mean redundancy	4.42	8.05	6.60
R_{sym} (%)	6.9 (44.0)	5.3 (39.6)	6.8 (21.7)
Average $I/\sigma(I)$	13.7 (3.1)	19.2 (3.8)	17.8 (8. 9)
Protein atoms	1890	1861	1883
Solvent atoms	132	177	135
Mean e.s.d. in atom position† (Å)	0.16	0.12	0.15
R.m.s. bond-length deviation (Å)	0.020	0.009	0.007
R.m.s. bond-angle deviation (°)	1.8	1.2	1.1
Mean *B* value of main-chain atoms (Å^2)	31.1	31.1	21.5
Mean *B* value of side-chain atoms (Å^2)	32.8	34.4	24.8
Mean *B* value of solvent atoms (Å^2)	37.5	44.9	38.7
Ramachandran most favoured regions (%)	95.5	89.5	90.1
Ramachandran allowed regions (%)	100.0	100.0	100.0
R_{work}‡ (%)	19.5	21.1	13.6
R_{free}‡	23.5	25.4	17.0
PDB code	1w7z	1h9h	1h9i

† Calculated using the DPI formula from Cruickshank (1999). ‡ $R_{sym} = \sum|I_{hkl} - \langle I_{hkl}\rangle|/\sum I_{hkl}$. † $R_{free} = \sum||F_{obs}| - |F_{calc}||/\sum|F_{obs}|$ for the 5% of reflections in the test set; R_{work} is calculated in the same way for the remaining reflections.

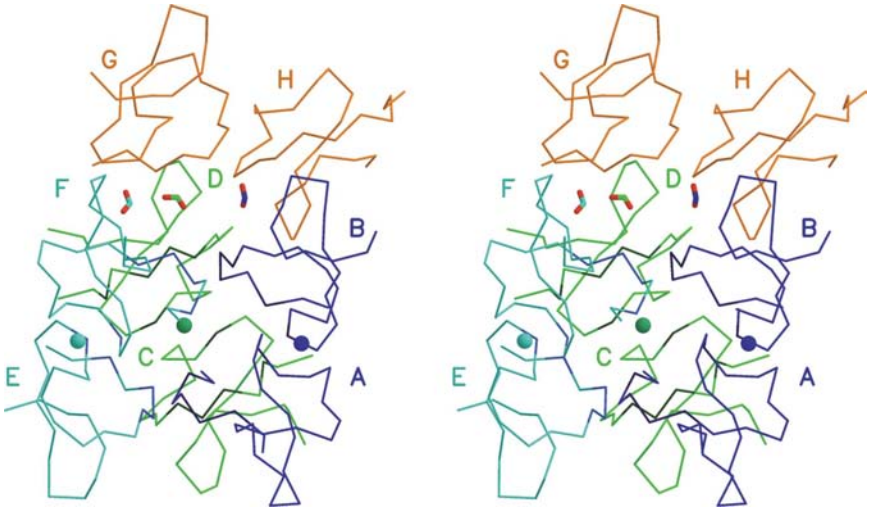

Figure 2
Stereoview of the overall structure of the uncomplexed EETI-II. Sodium ions are represented as solid balls and formate ions as sticks. Hydrophobic residues participating in the *A*–*F* hexamer formation are highlighted using darker colours.

as a search model for the complexes and the EETI-II model from the wild-type complex was then used to solve the native wild type. The structure of the trypsin–EETI-II-β^{GPNG} (wild type) complex was solved with *AMoRe* (Navaza, 1994) using data between 7 and 3.5 Å (other resolution ranges gave slightly poorer results). The best solution was identified by a correlation coefficient of 50.5% and a crystallographic *R* factor of 45.5%, which were well separated from the values of 23.4 and 54.7% for the next solution. The trypsin–EETI-II-β^{TNNK} complex was solved with *EPMR* (Kissinger *et al.*, 1999) using data between 15 and 4 Å. The correct solution was identified by a correlation coefficient of 64.8% and an *R* factor of 36.7%. After attempts with other programs had not been successful, six of the final eight molecules in the asymmetric unit of the native EETI-II could be located using *PHASER* (Read, 2001) with default settings. σ_A-weighted electron-density maps (Read, 1986) calculated from these six molecules indicated the presence of two additional molecules that could be fitted to the density by hand. The estimated Matthews coefficients (mc) and solvent contents (sc) for the reported structures were as follows: EETI-II wt, mc = 2.6 Å^3 Da^{-1}, sc = 52.6%; trypsin–EETI-II-β^{GPNG}, mc = 2.7 Å^3 Da^{-1}, sc = 54.5%; trypsin–EETI-II-β^{TNNK}, mc = 3.0 Å^3 Da^{-1}, sc = 58.9%.

2.4. Structure refinement

The uncomplexed structure consisting of eight EETI-II molecules was refined with *REFMAC*5 (Murshudov *et al.*, 1997) against *F*; NCS restraints were not used. The complex structures were refined with *SHELXL* (Sheldrick & Schneider, 1997) against F^2. In both cases, 5% of the data were flagged for R_{free} (Brünger, 1993) and $2mF_o - DF_c$ and $F_o - F_c$ maps were displayed with *XFIT* (McRee, 1999), which was also used for manual rebuilding of the structure and identifying disordered components. A bond-valence calculation for the three cations in the uncomplexed structure, using the methods of Müller *et al.* (2003), favoured Ca^{2+} and would also allow Na^+, but was not consistent with NH_4^+ (which would have accounted for the approximately tetrahedral coordination but had a bond-valence sum that was a factor of about four too high). Since the crystals were grown from 4 *M* sodium formate and neither Ca^{2+} nor NH_4^+ should have been present, these three ions were refined as sodium, which resulted in temperature factors similar to those of the surrounding atoms. The three V-shaped solvent particles were refined as formate ions, balancing the charge. In the complexes, omit maps were employed to establish the structure of the β-turn for residues 22–25 in EETI-II. As usual, the trypsin complexes included a Ca^{2+} ion. The quality of the final structures was assessed with *PROCHECK* (Laskowski *et al.*, 1993). Figures were drawn

with *MOLSCRIPT* (Kraulis, 1991), *BOBSCRIPT* (Esnouf, 1997) and *RASTER3D* (Merritt & Bacon, 1997).

3. Results and discussion

The EETI-II molecules adopt the canonical conformation of the cystine-knot inhibitor family. The core of the molecule is made up by a three-stranded antiparallel β-sheet consisting of β-strands spanning over the peptide stretches 6–8, 20–22 and 26–28 (Fig. 1). These are connected by short coils, of which the second is interrupted by a four-residue 3_{10}-helix. The EETI-II variants used in this study differ from the wild-type protein in as much as they all contain the mutation Met7Ile and a C-terminal extension with the amino-acid sequence SPHHHHHH, both of which were introduced to simplify the preparation and purification procedure (Wentzel *et al.*, 1999). The six-histidine tag was not visible in the electron density in any of these structures. Although this is the normal state of affairs in the many crystal structures that have been reported of proteins containing such tags, the implicit assumption that the tags behave as disordered solvent and have little structural influence is a matter of some concern. As we show later, the CMTI-I structure that did not contain this tag is very similar to that of EETI-II.

3.1. The wild-type structure

As shown in Fig. 2, the eight molecules in the asymmetric unit of the uncomplexed EETI-II (referred to as molecules *A*–*H*) form a 6+2-type octamer in such a way that six molecules (*A*–*F*) associate into a hexamer, whereas molecules *G* and *H* are attached to this hexamer in a rather irregular fashion. The structure of the hexamer can be visualized as follows. Molecule pairs *AB*, *CD* and *EF* form dimers with twofold non-crystallographic symmetry between the molecules. This molecular association is mediated by three metal ions refined as sodium ions (Fig. 3) that are coordinated by the main-chain O atoms of residues Ser13 and Cys15 of both molecules with an average sodium–oxygen distance of 2.22 Å. The dimerization is also promoted by hydrophobic contacts between Leu16 side chains from both monomers that are at an average minimum distance of 3.74 Å apart. These six molecules further associate into a hexamer so that a threefold non-crystallographic axis relates molecules *A*, *C* and *E*. Molecules *B*, *D* and *F* are also arranged as a quasi-trimer but are translated by about 12 Å along the same axis. The hexamer is held together by hydrophobic forces. Water-repulsive side chains of residues Pro3, Ile5, Leu6, Ile7 and Leu16 of molecules *A*–*F* form the hydrophobic core of the hexamer, where a water-excluded volume of about 4 Å diameter can be located. The three formate ions attached to molecules *B*, *D* and *F* neutralize the charge of the three Arg8 residues and can be located at the interface between the hexamer and the molecules *G* and *H*.

In contrast to molecules *A*–*F*, *G* and *H* are bound to the hexamer predominantly *via* hydrophilic interactions. Residues Arg8(*B*) and Arg8(*F*) are within hydrogen-bonding distance of the main-chain O atom of Gly1(*G*), Arg4(*G*) and Ala17(*G*), and hydrogen bonds were found between the side-chain N atom of Lys10(*H*) and the main-chain O atom of Cys21(*D*). The interactions between atom pairs Cys9(*H*)—O···H_2N—Arg4(*B*) and Cys9(*H*)—O···O—Cys9(*D*) are mediated by a water molecule. Molecules *G* and *H* occupy a relative position to each other that was observed in the case of molecules of the hexamer but are further apart and therefore lack any *G*–*H* intermolecular interactions; a sodium ion typical of molecules *A*–*F* could not be found between these molecules either.

Comparison of the average temperature factors of all molecules in the asymmetric unit revealed a correlation between the thermal motion and the position of the molecule

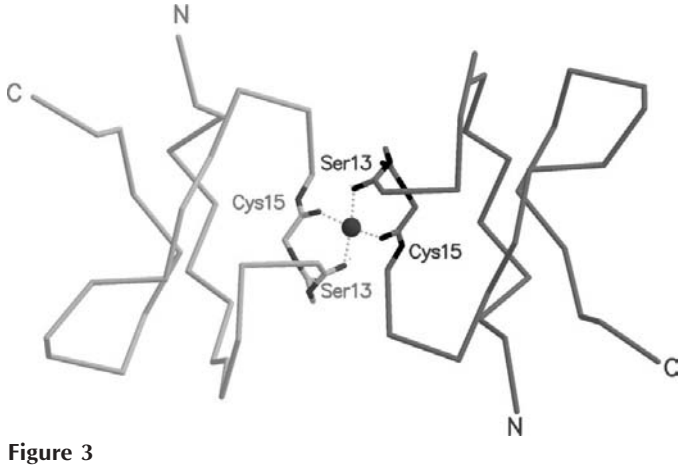

Figure 3
Dimer formed by the molecules *A* and *B*. Molecule pairs *CD* and *EF* display a similar arrangement. The sodium ion is shown as a solid ball.

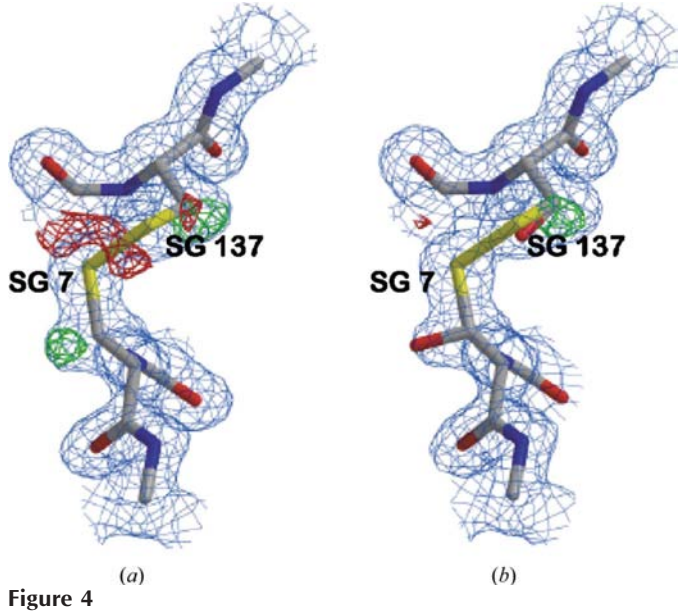

Figure 4
The consequences of radiation damage (wild-type complex). (*a*) A σ_A-weighted difference map shows negative density (red) at the disulfide bridge and positive density (green) consistent with an extra atom attached to C^{β} of Cys7 in trypsin; (*b*) the interpretation as a partially occupied disulfide bridge and a partially occupied serine instead of Cys7. The contouring levels for the density maps were 1σ (σ_A map) and 3σ for the difference density map.

within the octamer, as molecules *A*–*F* possess significantly lower *B* values than *G* and *H* (29.0 and 37.2 Å^2 for molecules *A*–*F* and *G*–*H*, respectively). This finding can be explained by the higher number of intermolecular interactions observed in the hexamer and by the lack of the sodium ion that would promote the dimerization of molecules *G* and *H*. This also explains the difficulties in locating molecules *G* and *H* by molecular replacement.

3.2. Trypsin complexes of EETI-II

Porcine trypsin was co-crystallized with three different sequences for the β-turn of EETI-II involving residues 22–25: GPNG (the wild-type sequence), TNNK and NEDE, yielding cubic crystals in space group $P4_132$ for the former sequence and tetragonal crystals in space group $I4$ for the latter two. These two mutants were selected for crystallization and structural investigation because they show different behaviour in folding-kinetics studies to that of wild-type EETI-II (Wentzel *et al.*, 1999). Since the NEDE complex was isostructural with that of the TNNK mutant and diffracted to lower resolution (3.0 Å in-house), it was not investigated further.

The main chain of the inhibitor is well defined in electron-density maps from Gly1 to Pro30 (the start of the His tag). Owing to the high intensity of the synchrotron beam used for the cubic crystals (ID14-EH2 at ESRF, Grenoble) some radiation damage is apparent in this structure (Fig. 4). Some of the disulfide cysteines in the trypsin molecule were modelled as partially oxidized to serine, in accordance with high-resolution studies of radiation damage (Burmeister, 2000; Weik *et al.*, 2000). For Cys128 no density was present for S$^{\gamma}$. The main chain in this area (residues 127–132) has an average *B* value of 36.4 Å^2 and is poorly defined. This indicates the cleavage of the disulfide bridge Cys128–Cys232 which otherwise would constrain this loop. Also, the C-terminal α-helix (Val235–Asn245) shows somewhat high main-chain *B* values (mean 36.4 Å^2) and residual difference electron-density peaks, indicating some disorder.

Table 2
R.m.s.d.s (Å) between the eight independent EETI-II molecules *A*–*H* in the wild-type structure and between *A*–*H* and the two complexes, the crystal structure of the related sequence 1luo and the NMR structure of EETI-II 2eti.

	A	*B*	*C*	*D*	*E*	*F*	*G*	*H*
B	0.39							
C	0.33	0.25						
D	0.33	0.37	0.32					
E	0.32	0.33	0.25	0.42				
F	0.35	0.32	0.28	0.31	0.35			
G	0.58	0.60	0.54	0.52	0.54	0.51		
H	0.51	0.61	0.62	0.44	0.70	0.57	0.54	
Trypsin–EETI-II-β^{GPNG}	0.49	0.53	0.45	0.40	0.57	0.49	0.53	0.43
Trypsin–EETI-II-β^{TNNK}	0.55	0.59	0.51	0.51	0.63	0.59	0.59	0.48
1lu0	0.81	0.87	0.86	0.73	0.96	0.86	0.74	0.66
2eti	2.12	2.07	2.10	2.18	1.88	2.07	2.02	2.20

Fig. 5 shows the different packing in the two structures. In the cubic form the GPNG β-turn is involved in packing contacts with its symmetry equivalents, but in the tetragonal form the EETI-II mutants are located in the interior of wide solvent-filled channels, with the β-turn exposed to the solvent. There are 18 polar and 14 non-polar short interatomic interactions between the EETI-II inhibitor loop, Cys2–Leu6, and the trypsin molecule, consistent with the very high binding constant ($K_a = 8 \times 10^{11}\ M^{-1}$; Favel *et al.*, 1989).

3.3. Comparison of the various EETI-II structures

Least-squares superposition of all main-chain atoms in the eight crystallographically independent molecules in the uncomplexed EETI-II structure gives r.m.s.d.s (root-mean-

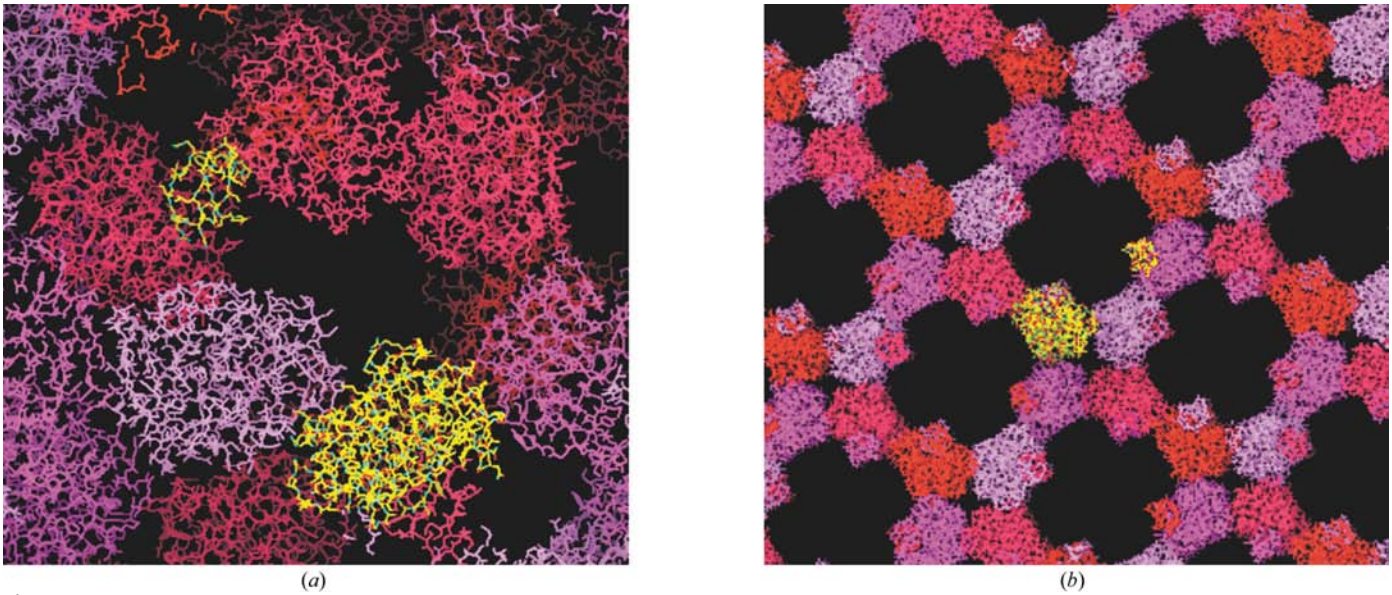

Figure 5
Packing of the trypsin–EETI-II complexes. (*a*) The cubic crystal form of the wild-type complex; (*b*) the tetragonal crystal form of the EETI-II–β^{TNNK} complex.

square deviations) in the range 0.25 Å (molecules *C* and *E*) to 0.70 Å (*E* and *H*). Molecules *A*–*F* differ from each other to a lesser extent than from molecules *G* and *H* (Table 2). These deviations are significantly larger than the mean coordinate errors (Table 1), which are in the range 0.12–0.16 Å. The program *ESCET* (Schneider, 2002) was used to establish which regions of the protein molecule are flexible and which parts can be regarded as a rigid framework. As illustrated in Figs. 6(*a*) and 7(*a*), short regions around positions 4 and 24 were found to be somewhat flexible, but the rest of the molecule seems to be rather rigid. The structural flexibilty of EETI-II at residue 4 between the complexed and uncomplexed inhibitor molecules can be explained by strong binding of Arg4 to Asp189 in the trypsin catalytic site.

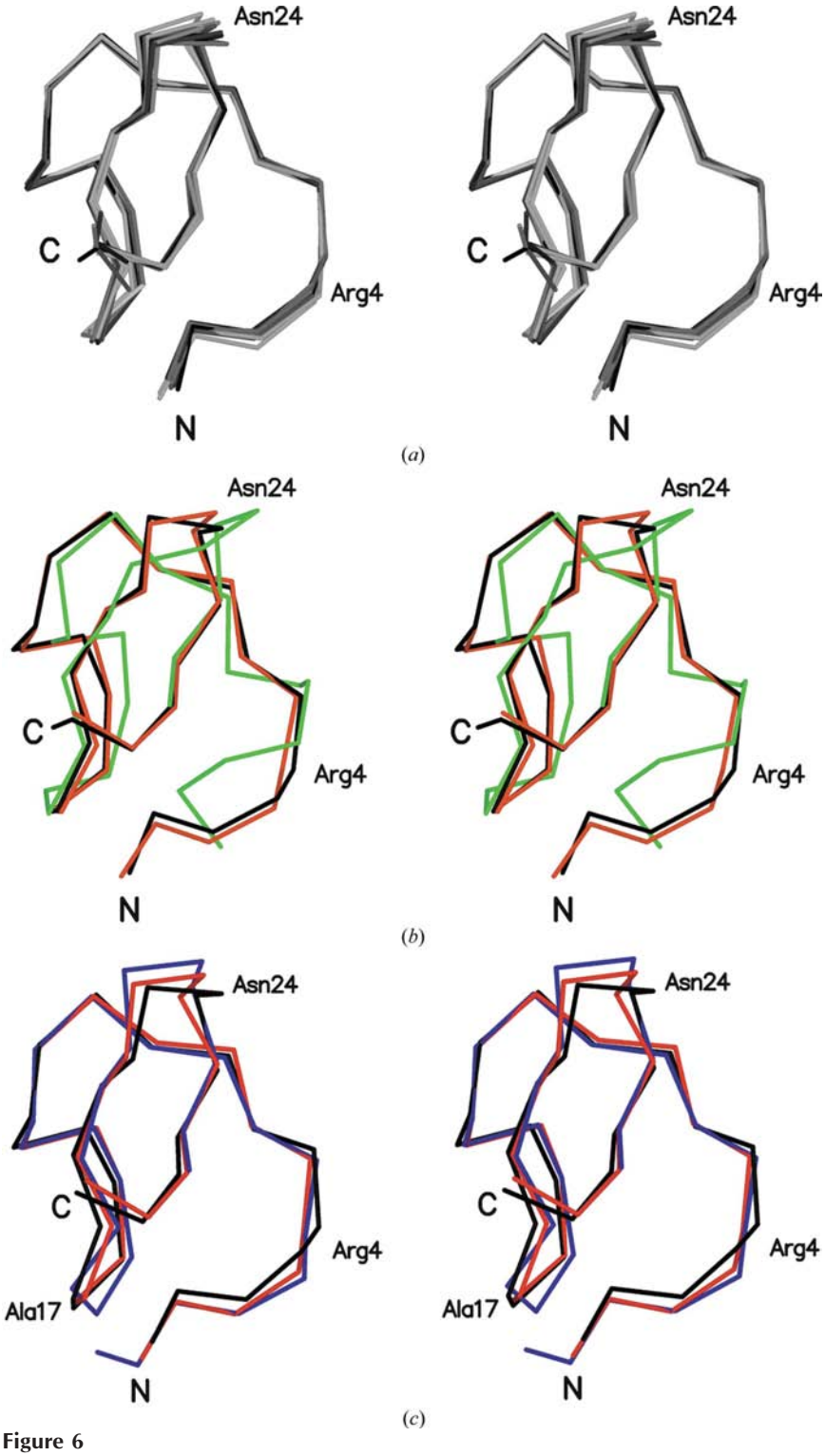

Figure 6
Least-squares superpositions of all main-chain atoms of EETI-II. (*a*) All eight independent molecules in the uncomplexed structure; (*b*) uncomplexed (black) and complexed (red) EETI-II and the NMR model (green); (*c*) uncomplexed EETI-II (black), the trypsin–EETI-II-β^{TNNK} complex (red) and CMTI-I (blue).

A similar result was obtained on superimposing the uncomplexed EETI-II structures and the wild-type EETI-II–trypsin complex (Fig. 6*b*). In this case, the mean r.m.s.d. varies between 0.40 Å (molecule *D*) and 0.57 Å (molecule *E*). Significant side-chain reorientations were observed, mostly in the Arg4–Arg8 and the Asn24–Phe26 regions, owing to different intermolecular interactions. For instance, in the complex the EETI-II residue Arg4 is involved in a salt bridge to the trypsin Asp189, whereas Arg8 forms hydrogen bonds to His40 and Ser39 through a water molecule. Since the uncomplexed EETI-II lacks those interactions, these side chains are free to adopt different conformations. Similarly, Ile5 and Leu6 take part in hydrophobic intermolecular interactions in the uncomplexed structure that are not possible in the complex, leading to different side-chain conformations.

Comparison with the NMR model of EETI-II (Chiche *et al.*, 1989; PDB code 2eti) revealed striking differences even in the main-chain conformation of the molecules (mean r.m.s.d. 2.08 Å). Large variations in the torsion angles could be observed in regions 1–7, 11–15 and 21–26 as illustrated in Fig. 6(*b*). In particular, the large difference between the conformation of the inhibitor-binding loop of the NMR and all the X-ray structures, despite the fact that according to the X-ray structures it scarcely changes on complex formation, is difficult to explain. In contrast to all the X-ray EETI-II structures, which have about 90% of residues in the most favoured regions and 100% in the allowed regions of the Ramachandran plot (calculated using *PROCHECK*; Laskowski *et al.*, 1993), the deposited NMR structure (2eti) has only 62% in the most favoured regions and 90% in the allowed regions.

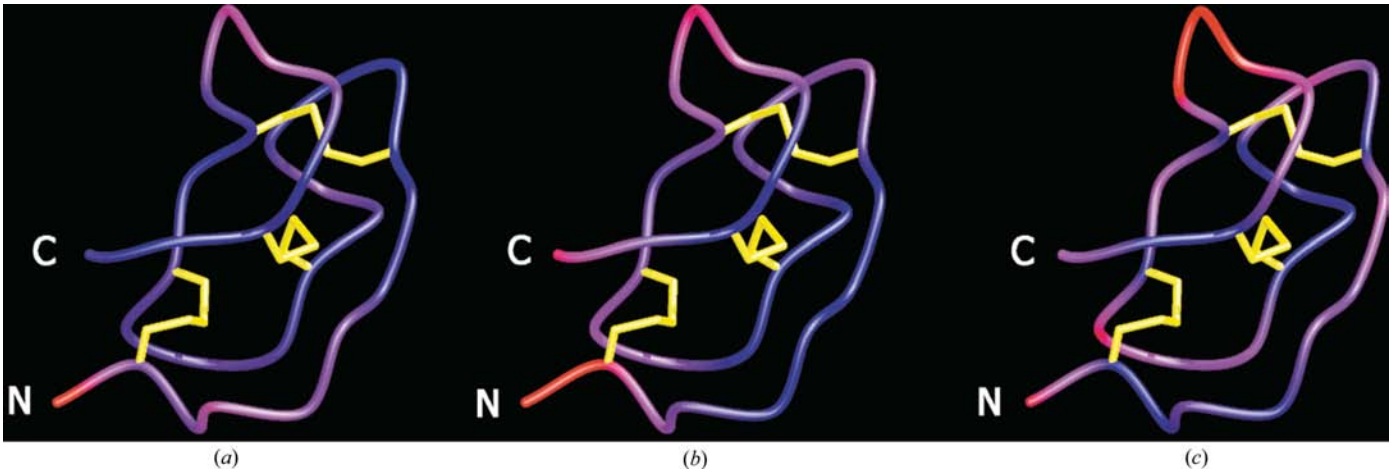

Figure 7
Colour-coded variations in the EETI-II structure (largest, red; smallest, blue). (*a*) Mean r.m.s.d.s in main-chain atomic positions for the nine structures with the wild-type sequence; (*b*) mean *B* values of the eight independent molecules in the uncomplexed structure; (*c*) the sequence variation (relative to EETI-II) of 30 squash inhibitors based on the SWISS-PROT database.

The superposition of the averaged structure of the eight independent uncomplexed EETI-II molecules with the EETI-II-β^{TNNK} mutant in complex with trypsin and with the related inhibitor CMTI-I (Thaimattam *et al.*, 2002; PDB code 1lu0) shown in Fig. 6(*c*) highlights similar small variations as in the comparison of the complexed and uncomplexed wild-type molecules, with some additional variation around position 17. The mean r.m.s.d.s are 0.48 and 0.81 Å, respectively. It is worthy of note that these regions also show the largest thermal motion (Fig. 7*b*) and are the least conserved parts of the sequence (Fig. 7*c* based on 30 EETI-II-related sequences; see also Leluk, 2000), *i.e.* the higher conformational flexibility is coupled with larger variability of the sequence.

4. Conclusions

The remarkable similarity of the structures of all nine crystallographically characterized molecules with the EETI-II-β^{GPNG} sequence and the close similarity with one mutant and with the previously reported CMTI-I (Thaimattam *et al.*, 2002), despite different molecular environments in the native and complexed forms, provide convincing evidence that EETI-II is a rather rigid molecule, making it a suitable choice for a molecular scaffold.

We are grateful for data-collection time at the BESSY synchrotron, Berlin (beamline BL2) and ESRF, Grenoble (beamline ID14-EH2). This work was supported by the Deutsche Forschungsgemeinschaft (SFB416) and by the Fonds der Chemischen Industrie. IU thanks the Spanish MEC for grant BIO2003-06653.

References

Brünger, A. T. (1993). *Acta Cryst.* D**49**, 24–36.
Burmeister, W. P. (2000). *Acta Cryst.* D**56**, 328–341.
Chiche, L., Gaboriaud, C., Heitz, C., Mornon, J. P., Castro, B. & Kollman, P. A. (1989). *Proteins*, **6**, 405–417.
Christmann, A., Walter, K., Wentzel, A., Krätzner, R. & Kolmar, H. (1999). *Protein Eng.* **12**, 797–806.
Colgrave, M. L. & Craik, D. J. (2004). *Biochemistry*, **43**, 5965–5975.
Craik, D. J., Daly, N. L. & Waine, C. (2001). *Toxicon*, **39**, 43–60.
Cruickshank, D. W. J. (1999). *Acta Cryst.* D**55**, 583–601.
Esnouf, R. M. (1997). *J. Mol. Graph. Model.* **15**, 132–134.
Favel, A., Mattras, H., Coletti-Preiero, M. A., Zwilling, R., Robinson, E. A. & Castro, B. (1989). *Int. J. Pept. Protein Res.* **33**, 202–208.
Gaboriaud, C., Vaney, M. C., Bachet, B., Le-Nguyen, D., Castro, B. & Mornon, J. P. (1989). *J. Mol. Biol.* **210**, 883–884.
Heitz, A., Chiche, L., Le-Nguyen, D. & Castro, B. (1989). *Biochemistry*, **28**, 2392–2398.
Kissinger, C. R., Gehlhaar, D. K. & Fogel, D. B. (1999). *Acta Cryst.* D**55**, 484–491.
Kohno, T., Kim, J. I., Kobayashi, K., Kodera, Y., Maeda, T. & Sato, K. (1995). *Biochemistry*, **34**, 10256–10265.
Kraulis, J. P. (1991). *J. Appl. Cryst.* **24**, 946–950.
Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). *J. Appl. Cryst.* **26**, 283–291.
Le-Nguyen, D., Heitz, A., Chiche, L., Castro, B., Boigegrain, R. A., Favel, A. & Coletti-Previero, M. A. (1990). *Biochimie*, **72**, 431–435.
Le-Nguyen, D., Heitz, A., Chiche, L., el Hajji, M. & Castro, B. (1993). *Protein Sci.* **2**, 165–174.
Leluk, J. (2000). *Cell. Mol. Biol. Lett.* **5**, 91–106.
McPherson, A. (1998). *J. Crystal. Growth*, **122**, 161–167.
McRee, D. E. (1999). *J. Struct. Biol.* **125**, 156–165.
Merritt, E. A. & Bacon, D. J. (1997). *Methods Enzymol.* **277**, 505–524.
Müller, P., Köpke, S. & Sheldrick, G. M. (2003). *Acta Cryst.* D**59**, 32–37
Murshudov, G. N., Vagin, A. A. & Dodson, E. J. (1997). *Acta Cryst.* D**53**, 240–255.
Navaza, J. (1994). *Acta Cryst.* A**50**, 157–163.
Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
Pallaghy, P. K., Nielsen, K. J., Craik, D. J. & Norton, R. S. (1994). *Protein Sci.* **3**, 1833–1839.
Read, R. J. (1986). *Acta Cryst.* A**42**, 140–149.
Read, R. J. (2001). *Acta Cryst.* D**57**, 1373–1382.
Rees, D. C. & Lipscomb, W. N. (1982). *J. Mol. Biol.* **160**, 475–498.
Saether, O., Craik, D. J., Campbell, I. D., Sletten, K., Juul, J. & Norman, D. G. (1995). *Biochemistry*, **34**, 4147–4158.
Schneider, T. R. (2002). *Acta Cryst.* D**58**, 195–208.

Sheldrick, G. M. & Schneider, T. R. (1997). *Methods Enzymol.* **277**, 319–343.
Skerra, A. (2000). *J. Mol. Recognit.* **13**, 167–187.
Stubbs, M. T., Morenweiser, R., Sturzebecher, J., Bauer, M., Bode, W., Huber, R., Piechottka, G. P., Matschiner, G., Sommerhoff, C. P., Fritz, H. & Auerswald, E. A. (1997). *J. Biol. Chem.* **272**, 19931–19937.
Thaimattam, R., Tykarska, E., Bierzynski, A., Sheldrick, G. M. & Jaskolski, M. (2002). *Acta Cryst.* D**58**, 1448–1461.
Weik, M., Ravelli, R. B., Kryger, G., McSweeney, S., Raves, M. L., Harel, M., Gros, P., Silman, I., Kroon, J. & Sussman, J. L. (2000). *Proc. Natl Acad. Sci. USA*, **97**, 623–628.
Wentzel, A., Christmann, A., Krätzner, R. & Kolmar, H. (1999). *J. Biol. Chem.* **274**, 21037–21043.