doi:10.1016/j.jmb.2005.04.039 J. Mol. Biol. (2005) 350, 156–169

JMB

Available online at www.sciencedirect.com

SCIENCE DIRECT®

ELSEVIER

Extended Intermolecular Interactions in a Serine Protease–Canonical Inhibitor Complex Account for Strong and Highly Specific Inhibition

Krisztián Fodor[1], Veronika Harmat[2], Csaba Hetényi[3], József Kardos[3] József Antal[3], András Perczel[4], András Patthy[1], Gergely Katona[5] and László Gráf[1,3]*

[1]*Biotechnology Research Group of the Hungarian Academy of Sciences, Eötvös Loránd University, Budapest, H-1117 Hungary*

[2]*Protein Modeling Group of the Hungarian Academy of Sciences and Eötvös Loránd University Budapest, H-1117, Hungary*

[3]*Department of Biochemistry Eötvös Loránd University Budapest, H-1117, Hungary*

[4]*Department of Organic Chemistry, Eötvös Loránd University, Budapest, H-1117 Hungary*

[5]*Department of Biochemistry University of Leicester University Road, Leicester LE1 7RH, UK*

**Corresponding author*

We have previously shown that a trypsin inhibitor from desert locust *Schistocerca gregaria* (SGTI) is a taxon-specific inhibitor that inhibits arthropod trypsins, such as crayfish trypsin, five orders of magnitude more effectively than mammalian trypsins. Thermal denaturation experiments, presented here, confirm the inhibition kinetics studies; upon addition of SGTI the melting temperatures of crayfish and bovine trypsins increased 27 °C and 4.5 °C, respectively. To explore the structural features responsible for this taxon specificity we crystallized natural crayfish trypsin in complex with chemically synthesized SGTI. This is the first X-ray structure of an arthropod trypsin and also the highest resolution (1.2 Å) structure of a trypsin–protein inhibitor complex reported so far. Structural data show that in addition to the primary binding loop, residues P_3–P_3' of SGTI, the interactions between SGTI and the crayfish enzyme are also extended over the P_{12}–P_4 and P_4'–P_5' regions. This is partly due to a structural change of region P_{10}–P_4 in the SGTI structure induced by binding of the inhibitor to crayfish trypsin. The comparison of SGTI–crayfish trypsin and SGTI–bovine trypsin complexes by structure-based calculations revealed a significant interaction energy surplus for the SGTI–crayfish trypsin complex distributed over the entire binding region. The new regions that account for stronger and more specific binding of SGTI to crayfish than to bovine trypsin offer new inhibitor sites to engineer in order to develop efficient and specific protease inhibitors for practical use.

Keywords: serine protease; canonical inhibitor; X-ray crystallography; NMR; specificity

Introduction

Besides their involvement in extra- and intracellular breakdown of proteins to amino acids, serine proteases catalyze highly specific cleavages in a number of biological processes from blood clotting to the complement cascade of the immune system. They regulate the level of particular proteins in the organism or convert their inactive forms to active ones. While serine proteases perform a wide range of functions essential to life, they can also be harmful. This may be the reason why their activity is not only controlled by the proteolytic activation of their inactive forms, by their auto-inactivation (autolysis), and their transport, but also by their inhibition with specific protease inhibitors.

The control of trypsin activity in pancreas and intestine provides a good example of these mechanisms. As discovered more than 130 years ago[1] trypsin is produced by the pancreas in its inactive form. It was shown a few decades later that in addition to trypsinogen and other protease zymogens, pancreas also contains a protein inhibitor of pancreatic trypsin[2] that was shown to function as a protector of premature trypsinogen

Abbreviations used: DSC, differential scanning calorimetry; NOE, nuclear Overhauser enhancement; MD, molecular dynamics.

E-mail address of the corresponding author: graf@ludens.elte.hu

auto-activation. Since then several pairs of serine proteases and inhibitors have been discovered and become not only targets of physiological studies but also favorite structural models for protein–protein interactions. A further aspect that makes investigation of serine protease inhibitors extremely important is their potential use in therapy. Thrombin inhibitors in use (anticoagulants) are classic examples of the practical realization of this kind of research.[3–5]

Until recently, trypsin complexed with substrate-like canonical serine protease inhibitors provided the only models for protease–protease inhibitor interactions.[6,7] With the development of biotechnology and bioinformatics there is an increasing demand for higher resolution X-ray structures of serine protease–inhibitor complexes, which reveal the specific interactions responsible for the strong and selective inhibitory effect and provide scaffolds and reliable data sets for *in silico* and *in vitro* inhibitor design. Though there are at least 18 protein families in this class of inhibitors with different overall folds, all appear to share a distinct and similar conformation of the primary binding loop.[7,8] This loop has long been thought to exhibit the same main-chain conformation in both free form and in complex with the protease and to be the major determinant of inhibition.[6,7] Our recent NMR studies on the backbone dynamics of small inhibitors of the newly discovered pacifastin family,[9–12] however, have shown that the binding loops in these inhibitors are less defined and more flexible than all the remaining part of the molecule.[13,14] Another intriguing conclusion of these studies was that SGTI (trypsin inhibitor from *Schistocerca gregaria*) is taxon-specific, inhibiting arthropod trypsins, such as the crayfish one, orders of magnitudes more effectively than the mammalian ones.[15] Our interest in the structural basis of this strong interaction between SGTI and crayfish trypsin initiated the present study.

Here we report the results of an experimental approach (differential scanning calorimetry) to demonstrate the strength of interaction between SGTI and crayfish trypsin and the crystal structure of their complex. The three-dimensional structure reported here presents the first arthropod trypsin structure and one of the highest atomic resolution of a serine-protease–protein inhibitor complex determined so far. Results from our structure-based molecular dynamics calculations are in agreement with the experimental data showing that the intermolecular interactions in the crayfish trypsin–SGTI complex are much stronger than those in the bovine trypsin–SGTI one. Our data provide experimental support to the hypothesis[16,17] that taxon specificity of inhibitors of the pacifastin family like SGTI is at least partly due to their interaction with the protease outside the commonly used interaction site of a canonical protease inhibitor, the binding loop. Engineering these newly explored sites may allow the production of highly specific inhibitors of therapeutically relevant proteases.

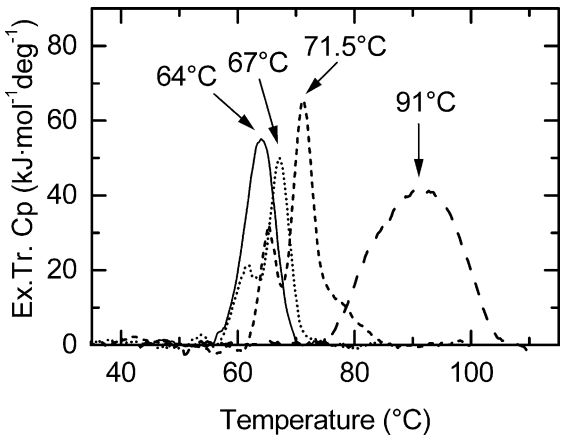

Figure 1. Thermal unfolding profiles of trypsins and their complexes with SGTI. Excess transition heat capacity of crayfish trypsin (–), bovine trypsin (· · · ·), bovine trypsin–SGTI complex (- - - -), and crayfish trypsin–SGTI complex (– – –) in 20 mM sodium phosphate, 100 mM NaCl (pH 7.0), using a heating rate of 1 deg.C/minute. The corresponding melting temperatures are indicated.

Results

Thermal stability of crayfish trypsin–SGTI and bovine trypsin–SGTI complexes

Bovine and crayfish trypsins without SGTI exhibited melting profiles with melting temperatures (T_m) around 67 °C and 64 °C, respectively, indicating somewhat lower structural stabilities of the crayfish enzyme (Figure 1). Addition of SGTI to bovine trypsin increased the T_m value to 71.5 °C, suggesting that stability of bovine trypsin is increased through interactions with SGTI. In the presence of SGTI, crayfish trypsin showed dramatically increased stability against thermal denaturation with a T_m value of 91 °C, indicating stronger enzyme–inhibitor interactions compared to the bovine trypsin–SGTI complex (Figure 1). Although the unfolding transitions were not reversible, differences in T_m values were independent of heating rate, suggesting that they reflect a real thermodynamic difference in stability.

Comparison of the amino acid sequence of crayfish and other trypsins

The nucleotide sequence was determined by DNA sequencing of recombinant crayfish trypsin. The amino acid sequences derived from this DNA sequence and from the X-ray structure of crayfish trypsin–SGTI complex using trypsin purified from *Astacus leptodactylus* were only different at a single position; in the X-ray structure a Val appears to replace an Ala residue at position 59 (Figure 2(a)). Comparing the amino acid sequences of crayfish trypsin to vertebrate trypsins, there is 41%–46%

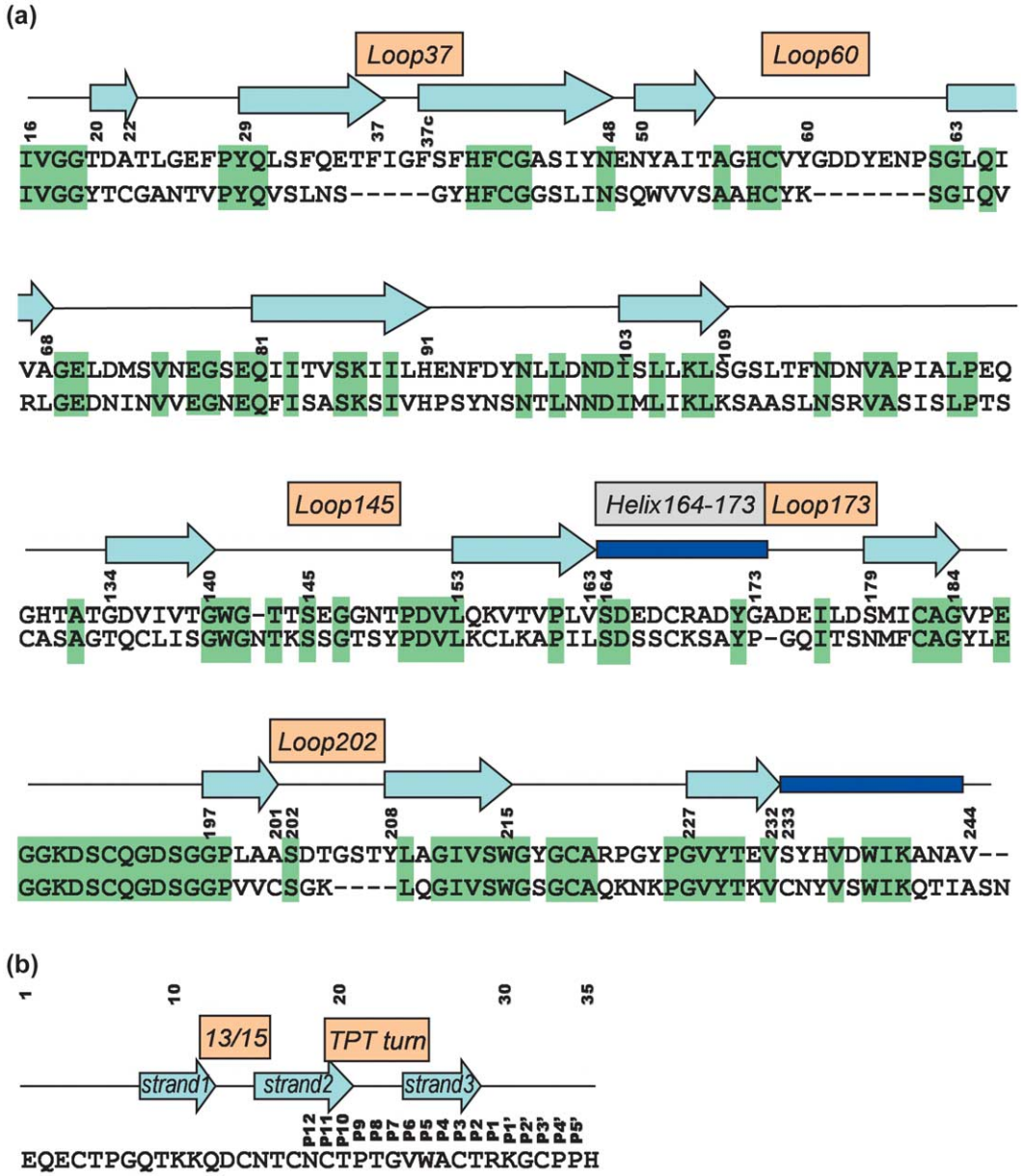

Figure 2. Amino acid sequences of crayfish trypsin and SGTI. (a) The sequence of crayfish trypsin (upper line) was aligned with the sequence of bovine trypsin (lower line, PDB id 3PTB) based on superposition of their 3D structures. Conserved residues are shown in green boxes. Secondary structure elements of crayfish trypsin are shown (α-helices as boxes, β-sheets as arrows) with their starting and end points marked (chymotrypsin numbering). Some regions showing structural differences between the crayfish and bovine enzymes are labeled in frames. (b) The sequence of SGTI with secondary structure elements shown (β-strands as arrows with labels). The two loops at the two ends of β-strand 2 are labeled. Residues forming contacts with crayfish trypsin are labeled as P_{12}–P_5'.

sequence homology between them (e.g. 41% for bovine anionic trypsin). In case of non-crustacean arthropod trypsins, amino acid sequence homology searches resulted in a 40%–53% homology while crustacean trypsins show a homology as high as 80%–98% with trypsin from *A. leptodactylus*.

General structural features of crayfish trypsin–SGTI complex

Crystals of natural crayfish trypsin in complex with SGTI synthesized by solid phase chemical synthesis were grown (see Materials and Methods) and diffracted to 1.2 Å resolution (PDB accession no. 1YR4). With the exception of the N and C termini of SGTI, the model is well defined in electron density (Figure 3(a)). Residues Ser79, Ser104 and Lys239 of crayfish trypsin as well as Cys27 of SGTI possess dual conformations. Crayfish trypsin exhibits the conserved core structure of the chymotrypsin fold consisting of two six-stranded β-barrel domains packed against each other, with the catalytic residues located at the junction of the two barrels (Figure 3(b) and (c)). The

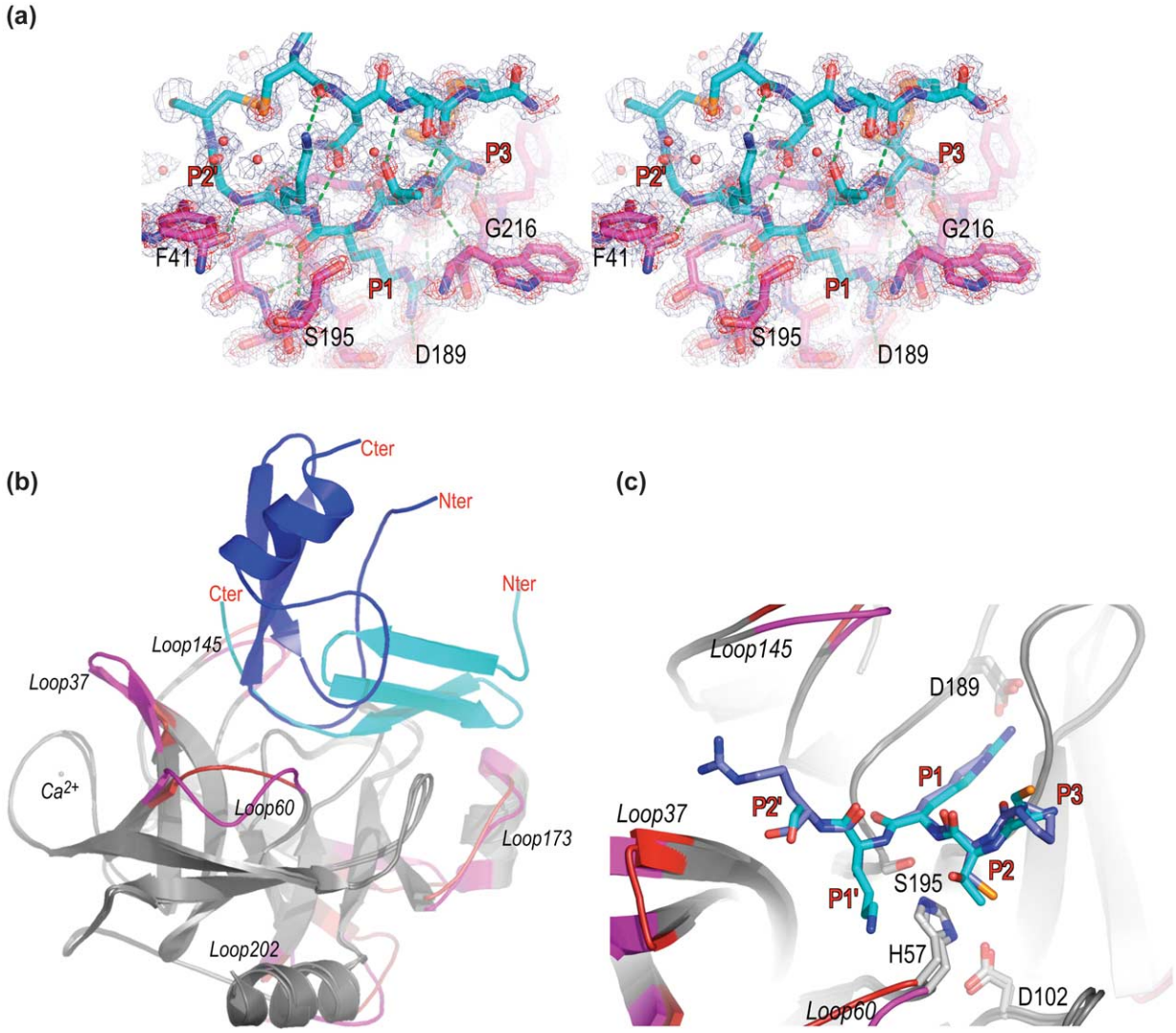

Figure 3. Crystal structure of the crayfish trypsin–SGTI complex. (a) Stereo view of the P_3–P_3' (carbon atoms in light blue), S_3–S_3' (carbon atoms in magenta) region with the $2F_o - F_c$ electron density map contoured at 1σ (blue) and 3σ (red). Protein–protein hydrogen bonds are shown as green shaded lines. (b) Overall conformation of the crayfish trypsin–SGTI complex compared with bovine trypsin–BPTI complex (PDB id 3BTK). Color codes are as described in (c). (c) Conformations of the P_2–P_2' regions of the inhibitors in the complexes, with the catalytic triad (carbon atoms in grey). In (b) and (c) the loop regions different for the two enzymes are colored magenta and red for the crayfish and bovine enzymes, respectively. Conserved structural elements are shown in grey. Carbon atoms of SGTI and BPTI are shown in light blue and dark blue, respectively. N, O and S atoms are shown in atomic colors. Black and red labels are used for the enzymes and inhibitors, respectively. For labeling of loop regions see Figure 2 and the text. Figures 3, 5(b) and 6(a) were generated by PyMOL.[51]

catalytic residues of trypsin are present in their active conformation. The overall structures of crayfish trypsin and bovine trypsin are similar. The Ca^{2+} binding loop characteristic to trypsin binds Cd^{2+} in the crystal structure as the crystallization medium contained Cd^{2+} in high concentration. (The ion in the binding loop had significantly stronger electron density than a Ca^{2+}, which was revealed by a positive $F_o - F_c$ difference Fourier peak. The nature of the bound ion was further evaluated by an anomalous difference Fourier map and *B*-factor analyses.) The conformation of SGTI is comparable to that determined by NMR spectroscopy (PDB accession no. 1KJ0).[13]

It is important to note that the major features of binding of crayfish trypsin to SGTI appear to be identical to the binding of bovine trypsin to bovine pancreatic trypsin inhibitor (BPTI) (PDB accession no. 3BTK),[18] despite the completely different fold of the inhibitors (Figure 3(c)); the antiparallel β-sheets formed between the proteases and the corresponding inhibitors are superimposable at sites S_3–S_2' (in the protease) and P_3–P_2' (in the inhibitor). A novel feature of protease–protease inhibitor interactions is that the interface region in the crayfish trypsin–SGTI complex is much more extended than in serine protease–inhibitor complexes of other inhibitor families (see Discussion).

Geometry of the scissile peptide bond in the inhibitor

The scissile peptide bond between Arg29 and Lys30 is present at full occupancy in the inhibitor complex. It is planar within the calculated coordinate error with the carbonyl carbon atom raising only 0.01 Å above the plane defined by the carbonyl oxygen and carbon alpha atom of Arg29 and the amid nitrogen of Lys30. The distance between the carbonyl carbon and the hydroxyl group of the catalytic serine is 2.69 Å, which is 0.09 Å shorter than the corresponding distance between the attacking water molecule and the carbonyl carbon of the ester bond in an atomic resolution elastase acyl-enzyme.[19] The shorter than van der Waals distance indicates significant orbital overlap between the two atoms although it is still too long for a true covalent bond. The active site is completely shielded from the solvent in the crayfish trypin–SGTI complex; the closest water molecule is 6.63 Å from the carbonyl carbon of the scissile peptide bond.

Loops of the enzyme

There are four loops of crayfish trypsin that are remarkably different in comparison with those of vertebrate trypsins (marked in Figures 2(a) and 3(b) as Loop37, Loop60, Loop145 and Loop202) from which two loops are important regarding inhibitor binding. (1) In contrast to bovine trypsin a more extended hydrophobic region is present in crayfish trypsin with a five residue insertion at position 37. The corresponding loop is referred to as Loop37. The insertion is manifested by an extension of two β-strands connected with a turn containing three phenylalanine residues and an isoleucine, which are oriented towards the inhibitor and interact with the C-terminal segment of SGTI. (2) Another insertion of seven residues occurs at position 60 (Loop60). Similar insertions could be found in some highly specified enzymes like those involved in the complement or blood clotting system (thrombin, mannose binding lectin associated serine protease, etc.).[20] However, while the so-called Loop60 of thrombin has direct influence on the S_2–P_2 interactions, this loop region of the crayfish enzyme turns away from the bound ligand and broadens the substrate binding groove, and thus plays a role in the formation of the S_1'–P_1' interaction.

Disulfide bridges of crayfish trypsin

Crayfish trypsin differs from vertebrate trypsins in its disulfide bond pattern. While bovine trypsin has six disulfide bonds, crayfish trypsin has only three. The conserved disulfide bridges are at positions 42–58, 168–182 and 191–220, respectively. All evolutionary conserved disulfide bonds are close to the active site of the enzyme, which was revealed by sequence comparison studies.[21] The 22–157 inter-domain disulfide bridge, which connects sequentially distant parts of the molecule and is suggested to be important in the structural stability of trypsins[22] is absent from crayfish trypsin. The two segments which are connected *via* this disulfide bond in bovine trypsin, are stabilized by a salt bridge between Lys157 and Glu26 in addition to main-chain hydrogen bonds in the crayfish enzyme. Absence of this disulfide bond may facilitate the relative motions of the two β-barrel domains. Another disulfide bridge is missing at positions 128–232. Ser232 connects *via* a hydrogen bond to the carbonyl oxygen of His128 while its carbonyl oxygen accepts another hydrogen bond from the Gly127 amide group. Consequently, the main-chain conformation of this part of the molecule is more similar to that of chymotrypsins that also lack this disulfide bridge. The disulfide bond at position 136–201 of bovine trypsin is also absent from crayfish trypsin. However, the main-chain conformation of this region is identical to that of the bovine enzyme. Despite the compensatory stabilizing interactions, the lack of the disulfide bridges, especially the one that connects the two domains, may cause the observed 3 °C drop in melting temperature of the crayfish trypsin compared to the bovine enzyme in the differential scanning calorimetry (DSC) study (see Figure 1).

Comparison of the interaction sites in crayfish trypsin–SGTI and bovine trypsin–SGTI complexes

The actual X-ray structure of the crayfish trypsin–SGTI complex was compared with the modeled bovine trypsin–SGTI complex, as well as with representative structures from the molecular dynamics run.

There is an important difference in the interaction pattern at the P_1' site. As seen in Figure 4(a) the P_1' lysine residue (light blue) of SGTI in the crayfish trypsin–SGTI complex is stabilized by a hydrogen bond with the Cys14 carbonyl oxygen atom of SGTI. Though its distance from Asp60b (located in Loop60) and Glu35 of crayfish trypsin (magenta) is about 8 Å in the crystal structure, the molecular dynamics simulations reveal that its position is stabilized closer to these negatively charged residues in solution establishing a weak electrostatic interaction (typical distance of charged groups of the P_1' lysine and Glu35 of crayfish trypsin is 5 Å; see Discussion). Our model of bovine trypsin–SGTI complex shows the P_1' lysine residue (green) located at a position similar to that in the crayfish trypsin–SGTI complex but the S_1' groove does not contain charged side-chains except for Lys60 (dark blue). The positively charged ε-amino group of this Lys60 forms a hydrogen bond to the Tyr39 side-chain that stabilizes its position at a distance of 6 Å from the P_1' Lys residue of SGTI.

Outside the primary binding region in both complexes there are further interactions. In crayfish trypsin–SGTI complex (Figure 4(b)) three phenylalanine residues of loop37 (magenta) bind Pro33 (P_4') of SGTI while Pro34 (P_5') turns outside. Phe39

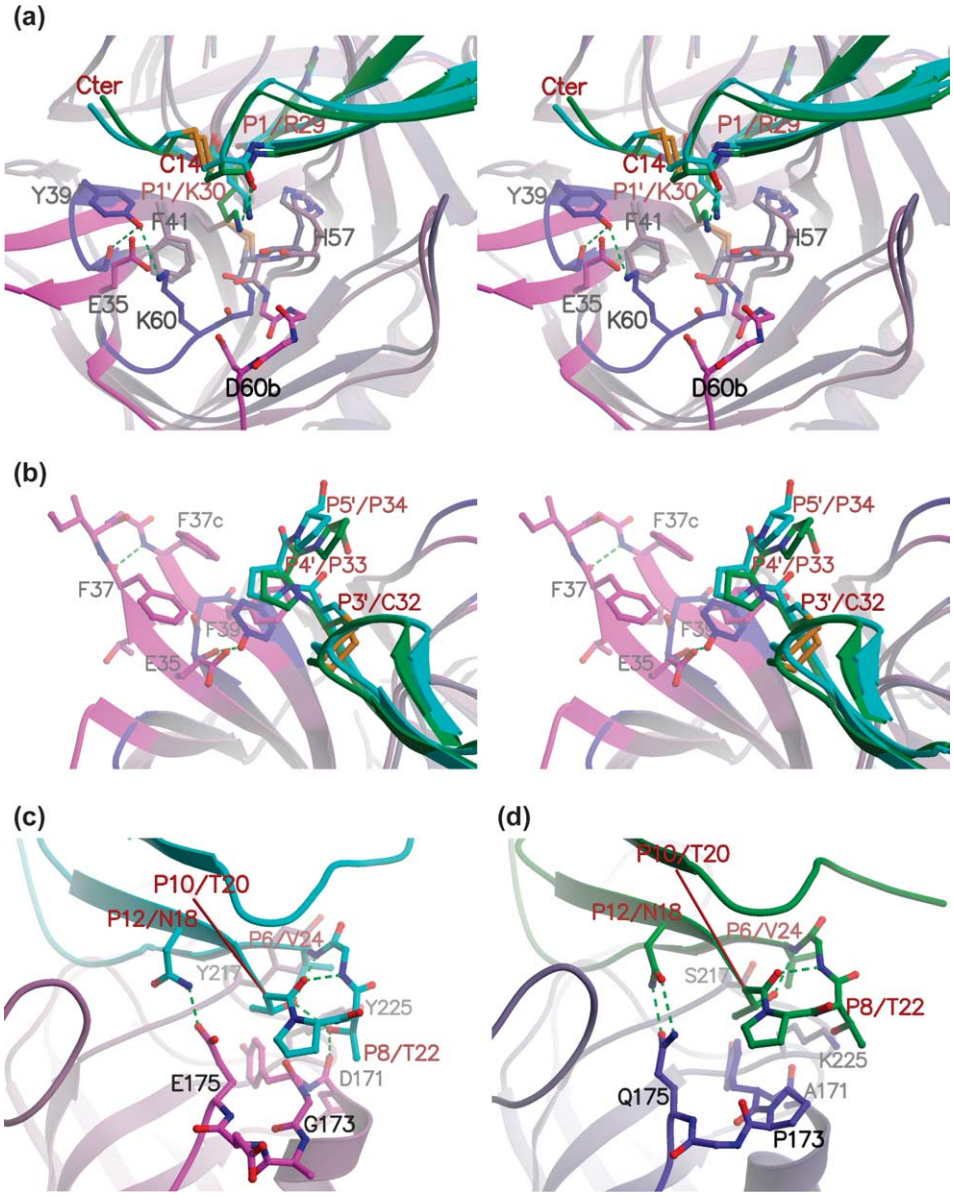

Figure 4. Extended binding region determining taxon specificity of SGTI. The crystal structure of the crayfish trypsin (magenta)–SGTI (light blue) complex superimposed over the representative model (MD, 180 ps step) of bovine trypsin (dark blue)–SGTI (green) complex. Conserved structural motives of trypsin are shown in grey. N, O and S atoms are shown in atomic colors. Hydrogen bonds are shown as green shaded lines. Black and red labels are used for the enzymes and inhibitors, respectively. (a) Stereo view of the $P1'$ residue accommodated in the $S1'$ site. Distances between the charged groups of the enzymes and the $P1'$ lysine amino group are 7.67 Å, 8.22 Å and 6.19 Å for E35 and D60b of crayfish trypsin and K60 of bovine trypsin, respectively. The conformation of the side-chain of the $P1'$ lysine is stabilized by an intramolecular hydrogen bond. (b) Binding of the P_4'–P_5' region (stereo view) is dominated by hydrophobic contacts. (c) Binding of the P_{12}–P_6 region by the crayfish enzyme is realized by several hydrogen bonds. (d) Binding of the P_{12}–P_6 region by the bovine enzyme. The hydrogen bonds of the P_8 threonine are lost, while stacking interaction is established between the P_9 proline and Pro173. The Figure was generated by MOLSCRIPT.[52]

of this cluster is in a key position, because it forms a stacking interaction with the Pro33 residue (P_4') (light blue). The shorter Loop37 (dark blue) of bovine trypsin is more rigid and contains only one aromatic residue, Tyr39, which forms a stacking interaction with the P_4' proline (light blue). Glu35 of crayfish trypsin, which was mentioned above as one of the electrostatic partners of the P_1' Lys residue of SGTI has another important role that it stabilizes Loop37.

Figure 4(c) illustrates the interaction between the P_{12}–P_6 region (light blue) of the inhibitor with crayfish trypsin (magenta). Val24 (P_6) forms a van der Waals interaction with Tyr217. Thr22 (P_8) forms hydrogen bonds with the 164–173 helix of crayfish trypsin. Thr22, important for the recognition of the enzyme, is stabilized by Thr20 (P_{10}) *via* a hydrogen bond. The hydroxyl group of Thr20 also stabilizes the backbone conformation of the P_{12}–P_6 loop.

In the bovine trypsin–SGTI complex the

interaction pattern is different (Figure 4(d)). At position 217 there is a serine residue instead of tyrosine that forms only a very weak hydrophobic interaction with Val24 of SGTI. In vertebrate trypsins, residue 173 is proline, which makes the 164–173 helix one residue shorter (dark blue). Differences in the backbone conformation of the 172–173 region cause loss of hydrogen bond interactions in the bovine trypsin–SGTI complex; the Thr22–enzyme hydrogen bond is missing, as well as the Thr20–Thr22 intra-molecular hydrogen bond.

Calculation of interaction energies of the proteins in the SGTI–crayfish trypsin and SGTI–bovine trypsin complexes

In order to find the structural basis of different inhibitory efficiencies of SGTI on different trypsins, structure-based calculations on the SGTI–trypsin complexes were performed.

Free energy of binding was calculated with the scoring function of AutoDock 3.0 (Materials and Methods). Scoring of the crystallographic and energy-minimized complexes of SGTI–crayfish trypsin and SGTI–bovine trypsin, respectively, resulted in a $\Delta\Delta G_b = -6.94$ kcal/mol lower binding free energy for the SGTI–crayfish trypsin complex. The difference between the complexes remained significant for conformations of the molecular dynamics trajectory (Supplementary Data; Figure 1). Contribution of each amino acid residue of SGTI to the interaction energy differences (free energy of binding) is depicted in Figure 5.

Extensive interactions facilitate conformational changes in the structure of SGTI

To search for any possible conformational changes in the inhibitor upon its binding to the enzyme we compared the X-ray structure of complexed SGTI with that of the solution structure of the free inhibitor.[13] Superpositions of atoms of SGTI in the complex and the average structure of the free forms yielded a backbone root-mean-square deviation of 1.80 Å in region 4–32 (Figure 6(a)).

Alignment of the NMR structure ensemble with the X-ray structure of complexed SGTI and a careful comparison of the backbone φ, ψ angles were carried out (Supplementary Data; Table 1). Additionally, NOE-derived restraints and corresponding distances in the complex are also compared (Table 2). The first (residues 8–12; see Figure 2(b) for secondary structure) and the second (residues 15–20) β-strands as well as the second loop (labeled as 13/15 in Figure 2(b)) interconnecting strands 1 and 2 have rather similar conformational properties both in the free and the complexed forms of the inhibitor.

The TPT turn (residues 20–22, P_{10}–P_8) shows conformational features resembling a somewhat distorted β-turn (a type II in solution and a type I in the complex) clearly stabilized both in solution and in the complex by the i-($i+2$) backbone hydrogen bond surrounding proline and by the Thr–Thr side-chain interactions. Nevertheless, an important structural change occurs in this part of the inhibitor. Both the above-mentioned 20–22 TPT segment with Gly23 and the Cys4–Thr5 region appear to be relatively stable in solution (with S^2 values among

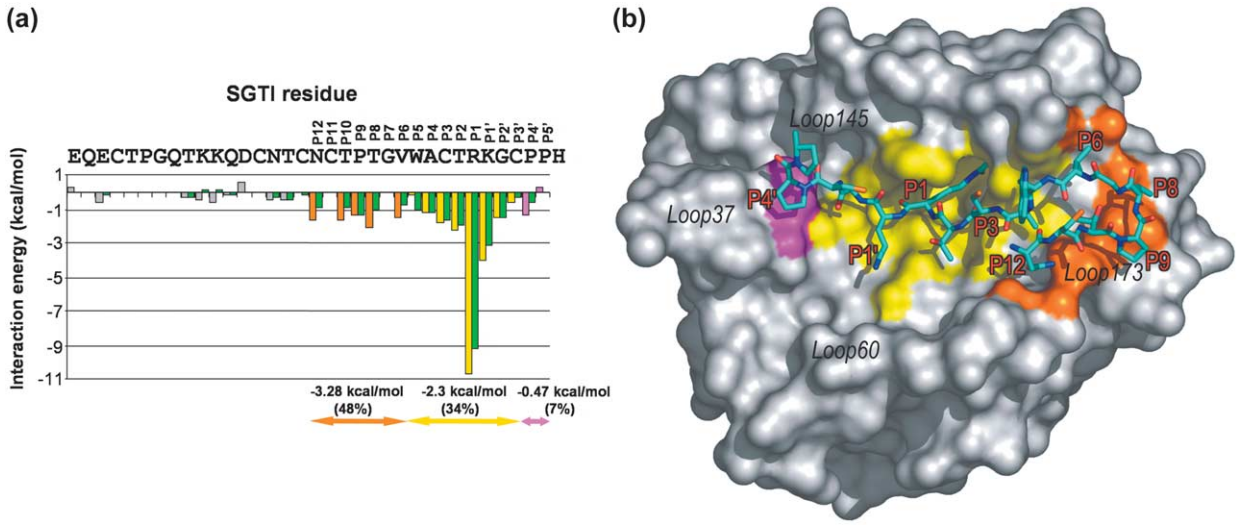

Figure 5. Energetic analysis of binding SGTI by crayfish and bovine trypsin. (a) Intermolecular interaction energy values corresponding to each residue of SGTI in the crystallographic structure of the SGTI–crayfish trypsin complex and energy-minimized structure ($t = 0$ ps) of the SGTI–bovine trypsin complex. The energy bars for the SGTI–bovine trypsin complex are colored green, while those of the SGTI–crayfish trypsin complex are colored magenta for P_4'–P_5', yellow for P_5–P_3', orange for P_{12}–P_6 and grey outside these regions. All the three binding regions contribute significantly to the taxon specificity of SGTI featured by the energy difference in interaction energy values shown for these regions. (b) The molecular surface of crayfish trypsin with the binding regions for P_4'–P_5', P_5–P_3' and P_{12}–P_6 of SGTI colored magenta, yellow and orange, respectively. The P_{12}–P_5' segment of SGTI is shown in atomic colors (C atoms in light blue).

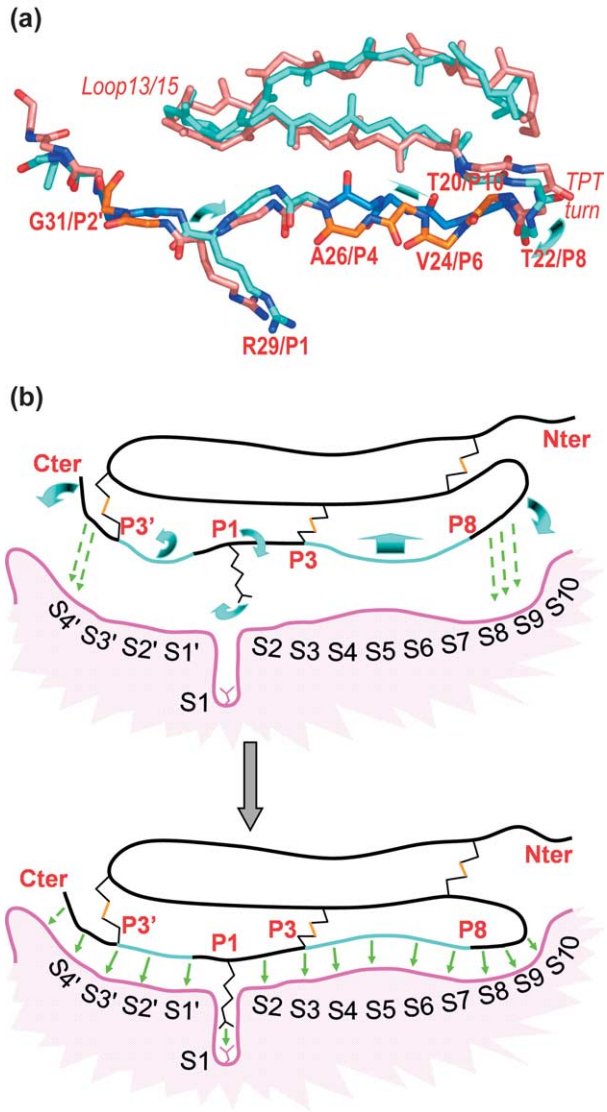

Figure 6. Shape adaptation upon binding of arthropod trypsin inhibitor SGTI to the surface of crayfish trypsin. (a) The structural alignment of the free (rose) and bound (light blue) forms of SGTI reveals three regions of major backbone conformation difference: the N-terminal segment (not shown), residues 20–26 (P_{10}–P_4) and residue 31 (P_2'). The latter two are parts of the binding region (O and N atoms shown in atomic colors). Carbon atoms of residues in the P_7–P_4 and P_2' regions are colored orange and dark blue for the free and bound form of SGTI, respectively. (b) Cartoon of SGTI binding to the enzyme. SGTI is shown in light blue (segments of the binding region with different backbone conformations in the free and bound form) and black (remaining parts). Cysteine and P_1 arginine side-chains of SGTI are shown, while some of its sub-sites are labeled in red. The enzyme surface is shown in magenta with black labels for the substrate binding sub-sites. Upper panel: conformation of the free form is preformed to recognize the S_{12}–S_8 and S_4'–S_5' sub-sites of the enzyme (shown as broken green arrows). In regions P_{10}–P_4 and P_2' conformation changes should occur, causing the rotation of the P_3–P_1' and P_4'–P_5' as well (light blue arrows). Lower panel: these conformational changes facilitate the build-up of an extended interaction network between SGTI and the enzyme (green arrows) in the complex.

the largest ones). However, considering the NOE-derived restraints between these two segments, a number of these are significantly (>0.5 Å) violated in the complex (Table 2). This indicates that these parts exhibit noticeable displacement with respect to each other upon protease binding.

Another important structural difference between the complexed and the free forms of SGTI is found in the 24–27 (P_6–P_4) region. This significant folding alteration is revealed by the values of φ, ψ angles of the complex structure. These angles are outside the entire folding range determined by the NMR ensemble.

Regarding the P_2–P_5' region, the folding similarity of Thr28–Arg29 dipeptide (P_2–P_1), especially the ψ angle of Thr28 and φ angle of Arg29 is very high. Nevertheless, Arg29 (P_1) is positioned farther from loop 13–15 in the complex than it is in the free form (Table 2). The C-terminal region ($31C^{\alpha}$–$34C^{\alpha}$) shows similar local conformations in the free and the bound form as well; however, the orientation of this unit is changed significantly upon complexation as it rotates around Gly31 (P_2'). Investigating the hydrogen bond system of the inhibitor we may conclude that H-bonds between strands 2 and 3 are well formed and shorter, therefore more stable in the complex than they are in the free form.

Discussion

The strength of interaction between crayfish trypsin and SGTI

SGTI, a protease inhibitor isolated from the haemolymph of desert locust, *S. gregaria*, is structurally homologous to the potent chymotrypsin inhibitor, SGCI, isolated from the same source. SGTI, however, contains an arginine residue instead of leucine at its P_1 site.[12] Despite this structural feature, which is favorable for trypsin inhibition, SGTI inhibits bovine trypsin relatively weakly, with an inhibitory constant (K_i) of 2.2×10^{-7} M. This value is five orders of magnitude larger than the equilibrium inhibitory constant of 6.2×10^{-12} M of SGCI determined on bovine chymotrypsin.[12,15] Unexpectedly, SGTI was found to be a potent inhibitor of crayfish trypsin with an eqilibrium K_i value of 0.7×10^{-12} M.[15]

Thermal denaturation experiments with crayfish trypsin–SGTI and bovine trypsin–SGTI complexes presented here have confirmed the results of the inhibition kinetics studies. As seen in Figure 1, the thermal stability of crayfish trypsin is increased dramatically upon binding of SGTI, resulting in a 27 °C higher melting temperature while the addition of SGTI to bovine trypsin led to only a 4.5 °C increase of T_m (Figure 1). The results are in line with the phylum specificity of SGTI and suggest that the observed difference in the K_i values may be realized in a thermodynamic stability difference in the enzyme–inhibitor interactions. We observed an increase in the peak areas, i.e. an

Table 1. Crystallographic data and refinement statistics

Resolution (Å)[a]		32.1–1.20 (1.26–1.20)	
Space group		$P2_12_12_1$	
Cell parameters (Å)		$\alpha=\beta=\gamma=90°$ $a=41.28$ $b=59.67$ $c=97.30$	
Number of observed reflections		878,624	
Number of unique reflections		93,027	
Completeness (%)[a]		91.2 (59.0)	
Mosaicity (°)		0.6	
$<I/\sigma>$[a]		13.1 (2.6)	
R_{merge} (%)[a,b]		5.9 (20.2)	
R_{work} (%)[c]		13.9	
R_{free} (%)[c]		18.2	
r.m.s. bond length (Å)		0.013	
r.m.s. bond angles (°)		2.242	
No. of non-hydrogen atoms			
Protein		1983	
Solvent		379	
Average *B*-factors (Å^2)	Protein	Water molecules	Overall
Main-chain	12.8±5.6		
Side-chain	15.7±7.7		
All	14.2±6.9	27.8±10.9	16.1±8.9
Anisotropy	0.43±0.14	0.54±0.17	0.44±0.15
Average estimated coordinate errors (esds) (pm)	5.3±3.9	9.7±6.1	5.9±4.6

[a] Values in parentheses indicate statistics for the highest resolution shell.
[b] $R_{merge}=\Sigma|I_o-\langle I\rangle|/\Sigma I_o\times100\%$, where I_o is the observed intensity of a reflection and $\langle I\rangle$ is the average intensity obtained from multiple observations of symmetry related reflections.
[c] R factor $=\Sigma||F_{obs}|-|F_{calc}||/\Sigma|F_{obs}|\times100\%$.

increased calorimetric enthalpy change of unfolding of the complexes compared to that of the single enzymes. This may be an outcome of intermolecular (enzyme–inhibitor) interactions rather than the simultaneous unfolding of SGTI, since the inhibitor alone shows no unfolding transitions up to 120 °C (not shown in Figure 1). In a previous study we pointed out the importance of the interdomain interactions in the function and structural stability of pancreatic serine proteases.[22] The X-ray structure of the complex revealed extensive interactions of SGTI with both β-barrel domains of crayfish trypsin, which may analogously explain the observed dramatic increase of thermal stability.

Based on the 3D structures of the crayfish trypsin–SGTI and the superimposed, energy minimized bovine trypsin–SGTI complexes, intermolecular interaction energies between the enzymes and SGTI were calculated (Figure 5(a)). Affinity of bovine trypsin to SGTI was found to be $\Delta\Delta G_b=-6.94$ kcal/mol smaller than affinity of crayfish trypsin. This structure-based free energy calculation is consistent with experimental stability difference between the investigated SGTI–trypsin complexes (6.6 kcal/mol) as converted from the inhibition constants.[12]

SGTI–crayfish trypsin interactions are extended over region P_{12}–P_5'

The crayfish trypsin–SGTI structure shows that the binding interface extends well beyond the primary binding region on both sides. Moreover, our energetic calculations reveal that all of these three regions are involved in stronger binding of SGTI to crayfish than to bovine trypsin (Figure 5(a) and (b)).

P_3–P_3': more favorable binding by the crayfish enzyme

Molecular dynamics (MD) simulation showed that the P_1–S_1 interaction is weaker in the bovine trypsin–SGTI complex than it is in the crayfish

Table 2. Some NMR restraints of SGTI and their violations in the SGTI–crayfish trypsin complex

Residue 1	Atom name	Residue 2	Atom name	NOE restraint (Å)	Violation (Å)
Cys4	H^{α}	Gly23	$H^{\alpha1}$	5	1.2
Cys4	H^{α}	Gly23	H_N	5	2.6
Thr5	H^{β}	Thr22	H_N	5	5.6
Thr5	$H^{\gamma2\#}$	Thr22	H^{α}	5	5.5
Thr5	$H^{\gamma2\#}$	Thr22	H_N	5	5.3
Thr5	H^{β}	Gly23	H_N	5	3.6
Thr5	H_N	Gly23	$H^{\alpha2}$	5	0.5
Asn15	$H^{\beta1}$	Arg29	H^{α}	5	0.5

$H^{\gamma2\#}$ is the pseudo-atom used for the γ-methyl group of Thr5.

trypsin–SGTI complex. In the former one, the distance between carboxylate 189 and the P_1 guanidino group is significantly longer during MD trajectory (Supplementary Data; Figure 2). A possible reason for the different behavior of the arginine side-chain is the looser fit of some neighboring sub-sites in bovine trypsin.

It was proposed in a previous study of ours that crayfish trypsin prefers positively charged residues at the P_1' position while bovine trypsin requires a neutral side-chain at the corresponding site.[13] Changing the P_1' residue lysine to methionine caused a one order of magnitude decrease in the inhibitory constant of SGTI on bovine trypsin. This result alludes to the importance of the P_1'–S_1' interaction for the determination of phylum specificity of SGTI (Figure 4(a)). Our MD study on the bovine trypsin–SGTI complex confirms this previous hypothesis regarding the conformation of Lys60 that was stabilized at the bottom of the S_1' cavity during the simulation. This residue is surrounded by hydrophobic residues and a positively charged one in the S_1' pocket in bovine trypsin, which is not favorable for adequate binding of SGTI. In contrast, the crayfish trypsin–SGTI structure shows that the broad S_1' cavity is more suitable for the positively charged P_1' lysine because it interacts with two negatively charged residues, Asp60b and Glu35 of the crayfish enzyme. MD shows high flexibility of Asp60b while Glu35 is stabilized at about 5 Å distance from the Lys30 N^{ζ} atom of SGTI. In the crayfish enzyme Glu35 seems to be especially important in respect of inhibitor binding, and it also stabilizes the Loop37 region that extends the primary binding region. The dual role of this residue ensures shaping of the binding surface in the primary binding region.

Hydrophobic binding patch at the S4′–S5′ region

A cluster of aromatic residues became inserted in Loop37 of arthropod trypsins. This cluster forms the binding surface for the Pro33–Pro34 (P_4'–P_5') region of SGTI (Figure 4(b)). MD shows that the main-chain conformation of this extended Loop37 and the three C-terminal residues of SGTI undergo only minor changes, and the movement of these two surface regions defined by Loop37 and the C-terminal residues of SGTI is restricted. The phenylalanine cluster interacts with the proline residues of SGTI; an alternative stacking interaction can be established by Phe37 and Pro34 (P_5') or Phe39 and Pro33 (P_4'). In the bovine trypsin–SGTI complex the weak interaction between Pro33 and Tyr39 is well maintained during simulation. This P_4'–P_5' region has the same conformation in all known complex structures of SGTI related peptides[10] while the mobility of these motifs in solution is relatively high. The phenylalanine cluster of crayfish trypsin might play a role in pre-orienting the inhibitor for the recognition of its C-terminal hydrophobic segment.

P_{12}–P_6 region stabilized by a network of hydrogen bonds

Both the interaction energy calculations and visual analysis of the contact interactions in trypsin–SGTI complexes have shown that the P_{12}–P_6 region forms significantly more favorable interactions with the crayfish enzyme (Figure 5(a) and (b)). The interactions in the crystal structure between P_{12}–P_6 and the 171–175 region of crayfish trypsin (Figure 4(c)) are practically unchanged in the MD trajectory, suggesting that these interactions may also be stable in solution.

Kellenberger and co-workers[16] proposed a hypothesis that the P_{10}–P_6 region of pacifastin-type trypsin inhibitors has an important role in its phylum selectivity. They suggested that the binding of P_{10}–P_6 to vertebrate trypsins is unfavorable because of the steric clash with Pro173 of these enzymes. Our present study confirms that Pro173 is indeed a key determinant for the binding difference, but rather than introducing an unfavorable steric effect it is disrupting the helical conformation at the 172–173 region of the enzyme. The key factor that determines the selective binding to trypsins of inferior or superior species is the molecular recognition of the C-terminal end of the 164–173 helix backbone. Trypsins with one residue insertion in the 164–175 region and glycine in position 173 (Figure 2), such as crayfish or *Fusarium oxysporum* trypsins, are likely to form a helix with a backbone conformation suitable for SGTI binding *via* a hydrogen bond network. Vertebrate trypsins have shorter loops in this region and Pro173 also breaks the helix. Our present molecular dynamics study also supports that Pro173 forms a stacking interaction with the Pro21 (P_9) residue of SGTI. These observations, when taken together, provide a circumstantial explanation of why vertebrate enzymes can form only less favorable interactions with SGTI.

SGTI binding to trypsin; anchor points and conformational adaptation

A comparison of the free and bound forms of SGTI reveals a conformational change upon binding to the enzyme (Figure 6(a)) that facilitates the emergence of an extended and strong interaction network. Local conformation of both P_{12}–P_4 and P_4'–P_5' regions of the inhibitor shows significant changes upon binding, suggesting that either or both of these regions may act as additional molecular recognition sites. The number of NMR distance (NOE) restraints (322 in total, ~10/residue; 123 long-range) provides adequate information to establish the overall conformation of the inhibitor in solution. In loop regions, however, short-range NOEs dominate and determine the local structure (e.g. 20–22 TPT). This is in agreement with the generalized order parameters (S^2), where those of T20 and T22 are somewhat higher than those belonging to their close vicinity, indicating

that the β-turn is likely to be involved in flip-flop type motions although not directly detected on the μs–ms time scale. This is consistent with the scarcity of long-range NOEs in this region despite of the short inter-atom distances in the solution structure. The suggested movement is strongly supported by the increased distance of the T20–G23 part and the N terminus clearly detectable as NOE restraint violations (Table 2). The P_6–P_4 region moves towards β-strand 2 of the inhibitor (the atom–atom distances become shorter and are consistent with the NOE experimental data; thus, this type of motion cannot be detected as restraint violation), influencing also local conformational preference, now forming a strong inter-strand H-bond network and adopting a conformation assuring a perfect match with the enzyme surface. As a consequence, the P_1 arginine residue is forced into the S_1 pocket of the enzyme and the P_1' lysine residue rotates into its binding groove. The C-terminal region preserves its conformation while it rotates around Gly31 (P_1') making a close fit with the enzyme surface (Figure 6(b)).

The extension and plasticity of crayfish trypsin–SGTI interaction offers new avenues for inhibitor specificity engineering

A great wealth of knowledge has been collected on the highly specific functions of serine proteases in the living organism. Development of computer-aided protein engineering opens a new inter-disciplinary route for the design of specific inhibitors for therapeutic use. Although there are a large number of efficient small molecule serine protease inhibitors, they are not sufficiently specific and often too toxic for medical use. The most selective and potent inhibitors provided by nature are either oligopeptides or proteins. Canonical or standard mechanism inhibitors represent an important subset of these protein protease inhibitors.[6,7] A common structural feature of this class of inhibitors is that they have a reactive peptide bond in a loop that binds to the protease in a standard manner, and that this loop of six to nine residue long (also called primary binding region) has a more or less similar conformation in inhibitors and also in the enzyme–inhibitor complexes.[7] The uniform structure and homologous binding mode of these loops, however, may not provide these inhibitors with an extreme selectivity of their action. The ovomucoid third domain is a good example of a typical serine protease inhibitor with a relatively broad specificity.[23] In the complex of human leucocyte elastase with ovomucoid third domain the interaction is extended to the P_5–P_3' region of the inhibitor. Regading molecular movements upon binding, only the N-terminal region shows minor movements, the binding loop preserves its conformation.[24] However, some protease inhibitors, in addition to their typical primary binding loops, possess secondary binding sites as well. Hirudin, the most active and specific natural thrombin inhibitor uses an even more sophisticated mode of binding; out of its 65 residues 27 directly interact with thrombin.[25] The examples of hirudin and some recently developed two-binding site protease inhibitors of Factor VII show that new binding sites of an inhibitor tremendously increase its specificity and strength of interaction with the target protease.[26,27] As our present study shows, SGTI uses an inhibition strategy somewhat different from the inhibitors described above. In complex with crayfish trypsin, SGTI exhibits, instead of more than one distinct binding site, more or less continuous contacts in an extended region (through sites P_{12}–P_5') of the molecule (Figure 5). Some of these contacts result from a conformational change of SGTI that was induced by its binding to the enzyme (Figure 6). This is strongly supported by the precise comparison of the atomic resolution crystal structure of the crayfish trypsin–SGTI complex with that of uncomplexed SGTI. In contrast to other serine protease–protein inhibitor complexes where secondary interactions are mostly van der Waals contacts and do not affect specificity,[28] our present study shows that the extension of the binding surface leads to an increased specificity and stabilization of the complex. The comparison of the complexes of bovine and crayfish trypsin with SGTI shows that more than half of the interaction energy difference originates from the differential binding of the extended regions in the two complexes (Figure 5). The high resolution structure presented here provides a good basis for further study of the structural aspects of protease inhibitor specificity and to introduce new interaction sites into the inhibitor to increase its specificity towards proteases of interest.

Materials and Methods

DNA preparation, amplification and sequencing

Crayfish trypsin mRNA was obtained from *A. leptodactylus* hepatopancreas. Tissue (100 mg) was homogenized in 1 ml TRI-REAGENT (Sigma Chemical Co., Hungary) and RNA isolated according to the protocol of the manufacturer. Due to the known complete sequences of species *A. fluviatilis* and *Pacifastacus leniusculus* that are relatives of *A. leptodactylus* we could design oligonucleotide primers for amplifying the coding region of crayfish trypsin. RT-PCR was performed using the following primers: CFT3′: 5′-GGAGCTCAGACTGC ATTTGCTTTGAT-3′, CFT5′: CCGAAGCTTTTCCCGTG GATGATGATGACAAGATCGTTGGTG. These primers include a HindIII site at the 5′ end and a SacI site at the 3′ end. Additionally, since the propeptide sequence of crayfish trypsin is unknown we attached a rat trypsin propeptide sequence at the 5′ region. The amplified DNA was cloned into a pET17b vector. The sequence of the chimeric trypsin was determined by automated dideoxy sequencing (ABI Prism) using the Big Dye Terminator Kit (GenBank accession no. AY906961). Since our experiments that were aimed at producing recombinant crayfish trypsin yielded only a low amount of enzyme, for the

crystallization procedure natural crayfish trypsin was used.

Isolation and characterization of crayfish trypsin

Narrow-clawed crayfish (*A. leptodactylus*) trypsin was purified by the procedure as described by Zwilling *et al.*,[29] with some modifications. A stock of about 500 crayfish cardia fluid was collected by introducing a teflon capillary tube attached to a syringe into the cardia of the animal. The collected cardia fluid was ultrafiltrated on AMICON (AMICON Corp., Beverly, MA, USA) membranes with 50 kDa and 10 kDa cut-off. The fraction between 10 kDa and 50 kDa was loaded onto a CNBr-Sepharose-4B soybean-trypsin inhibitor column. The column was washed with three volumes of distilled water and then eluted with dilute NH_3 solution (pH 11.0). Fractions containing crayfish trypsin were pooled and loaded on a MONO Q (Pharmacia, Sweden) ion exchange column equilibrated with 10 mM Mes (pH 6.0) and eluted with a linear gradient of 0 M to 1 M NaCl. Fractions containing different forms of crayfish trypsin were collected, and checked by SDS-PAGE, 2D-SDS-gel electrophoresis and activity measurements. All izoenzymes of crayfish trypsin were found to be identical regarding their enzymatic activities and sensitivities to inhibition. For further study the most abundant form was chosen, and concentrated by ultrafiltration using Centricon-10 concentrators (AMICON Corp., Beverly, MA, USA).

Differential scanning calorimetry (DSC)

Calorimetric measurements were performed on a VP-DSC (MicroCal) differential scanning calorimeter. Equimolar mixtures of trypsin and SGTI were used for studying the effect of the inhibitor on bovine and crayfish trypsins. The protein concentration was set to 0.1 mg/ml. Samples were dialyzed against 20 mM sodium phosphate (pH 7.0), 100 mM NaCl, and the dialysis buffer was used as a reference. Denaturation curves were recorded between 10 °C and 120 °C at a pressure of 2.5 atm, using a scanning rate of 1 deg.C/minute. The thermal unfolding curves were analyzed using MicroCal Origin 7.0 software. We note that bovine trypsin and its complex exhibited additional minor components at lower temperatures, which is a consequence of heterogeinity, probably due to an autolysis product of the commercial enzyme sample.

Chemical synthesis of SGTI

The inhibitor was synthesized, oxidized and purified as described.[12]

Preparation of the crayfish trypsin–SGTI complex and its crystallization

A fourfold molar excess of SGTI was added to crayfish trypsin and incubated for 15 minutes at room temperature. The complex was loaded to a HiPrep S-100 gel filtration column (Amersham Biosciences, UK), and eluted with 10 mM Mes (pH 6.0). The pure crayfish trypsin–SGTI complex was collected and concentrated to 11 mg/ml. Crystals of the crayfish trypsin–SGTI complex were grown by the hanging drop method at 20 °C. Equal amount of protein solution (11 mg/ml protein in 10 mM Mes (pH 6.0)) and precipitant solution (30% (w/v) polyethylene glycol (PEG) 400, 0.1 M cadmium chloride, 0.1 M sodium acetate (pH 4.6)) were mixed and equilibrated against 0.5 ml of precipitant solution. Crystals were grown in two days.

X-ray diffraction studies

Two datasets were collected from a single crystal at ESRF on beamline ID 14 EH2 at cryogenic temperature (100 K). Crystallograpic intensities were integrated and scaled to a resolution of 1.2 Å using Mosflm[30] and Scala[31] of the CCP4 package V5.0.[32] Completeness of the data was 91.2% at 1.2 Å resolution. The structure was solved by molecular replacement using the program Molrep[33] from the CCP4 package. A polyalanine search model was used which was derived from the X-ray structure of human trypsin IV (PDB entry 1H4W).[34] The asymmetric unit contains one trypsin–inhibitor 1:1 complex. Automated model building was carried out with Arp/wArp.[35] The model was systematically improved using iterative cycles of manual rebuilding with the program O[36] and restrained least-squares refinement with SHELX.[37] Atomic *B*-factors were refined anisotropically and this step reduced the *R*-factor and R_{free}[38] values by 4.3% and 3.3%, respectively. Finally, all except for hydroxyl and His $N^{\varepsilon 2}$ and $N^{\delta 1}$ riding hydrogen atoms were added to the structure. The geometry of the P_2–P_2' residues of the inhibitor were not restrained in the final rounds of the refinement. The final model contains residues 16–244 (chymotrypsin numbering system[39]) of crayfish trypsin, and residues 2–34 of SGTI. The stereochemistry of the structure was assessed with Whatcheck[40] and PROCHECK.[41] The distribution of anisotropic *B*-factors was monitored with the program Parvati.[42] Data collection and refinement statistics are shown in Table 1.

Calculations

SGTI–bovine trypsin complex was derived from superposition of the structure of bovine trypsin (PDB entry 3PTB)[43] on the crayfish trypsin complex of the present study using the LSQMAN program[44] from Uppsala Software Factory (USF). The GROMACS[45] program package was applied for generation of a simulation box, addition of explicit water molecules and counter ions. Both complexes (SGTI–crayfish trypsin and SGTI–bovine trypsin) and SGTI alone were energy minimized using the GROMOS[46] force field implemented in the program package. The 100 ps long position restrained and 500 ps long unrestrained molecular dynamics simulations (MD) were performed to equilibrate the surrounding molecules and to generate conformations for calculations of intermolecular interaction energy (E_i).

Intermolecular energy terms of scoring function of AutoDock 3.0 program[47] were applied to 50 conformations of the complexes (sampled at every 10 ps of the 500 ps trajectories). Extra penalty constants of H-bonds were not used. Thus, scaled Coulombic, Lennard-Jones terms and the desolvation free energy term[48] were involved in calculation of E_i. Difference of the E_i-s is considered as $\Delta\Delta G_b$ (difference in free energy of binding) of the interactions of SGTI with the two trypsins. Preparation of trypsin and SGTI molecules and grid calculations were done as described in our previous studies[49,50] for each conformation using shell scripts. AMBER charges were applied for all molecules.

Acknowledgements

This work was supported by the Hungarian National Science Foundation (OTKA), grant TS044730, the ICGEB-HUN04-03 fund and an EU Improving Human Potential program, BBSRC grant. We thank Zoltán Gáspári for helpful discussions. We also thank Hanna-Kirsti Schrøder Leiros for the expert assistance at beamline ID14-EH2, ESRF. G.K. thanks Richard Neutze, Nigel S. Scrutton and David Leys for their support. C.H. is a Békésy Fellow of the Hungarian Ministry of Education. This paper is dedicated to Professor András Lipták and his outstanding scientific and professional activities on his 70th birthday.

Supplementary Data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.jmb.2005.04.039

References

1. Kühne, W. (1867). Über die Verdaung der Eiweistoffe durch den Pankreassaft. *Virchows Arch.* **39**, 130–174.
2. Kunitz, M. & Northrop, J. H. (1936). Isolation from beef pancreas of crystalline trypsinogen, trypsin, trypsin inhibitor and an inhibitor trypsin compound. *J. Gen. Physiol.* **19**, 991–1007.
3. Bajusz, S., Széll, E., Bagdy, D., Barabás, E., Horváth, G., Diószegi, M. *et al*. (1990). Highly active and selective anticoagulants: D-Phe-Pro-Arg-H, a free tripeptide aldehyde prone to spontaneous inactivation, and its stable N-methyl derivative, D-MePhe-Pro-Arg-H. *J. Med. Chem.* **33**, 1729–1735.
4. Hirsh, J., O'Donnell, M. & Weitz, J. I. (2005). New anticoagulants. *Blood*, **105**, 453–463.
5. Kaplan, K. L. & Francis, C. W. (2002). Direct thrombin inhibitors. *Semin. Hematol.* **39**, 187–196.
6. Bode, W. & Huber, R. (1991). Proteinase–protein inhibitor interaction. *Biomed. Biochim. Acta*, **50**, 437–446.
7. Laskowski, M. & Qasim, M. A. (2000). What can the structures of enzyme–inhibitor complexes tell us about the structures of enzyme substrate complexes? *Biochim. Biophys. Acta*, **1477**, 324–337.
8. Schechter, I. & Berger, A. (1968). On the site of the active site in proteases. *Biochem. Biophys. Res. Commun.* **27**, 157–162.
9. Boigegrain, R. A., Mattras, H., Brehelin, M., Paroutaud, P. & Coletti-Preverio, M. A. (1992). Insect immunity: two proteinase inhibitors from hemolymph of *Locusta migratoria*. *Biochem. Biophys. Res. Commun.* **189**, 790–793.
10. Kellenberger, C., Boudier, C., Bermudez, I., Bieth, J. G., Luu, B. & Hietter, H. (1995). Serine protease inhibition by insect peptides containing a cysteine knot and a triple-stranded beta-sheet. *J. Biol. Chem.* **270**, 25514–25519.
11. Hamdaoui, A., Schoofs, L., Wateleb, S., Bosch, L. V., Verhaert, P., Waelkens, E. & De Loof, A. (1997). Purification of a novel, heat-stable serine protease inhibitor protein from ovaries of the desert locust, *Schistocerca gregaria*. *Biochem. Biophys. Res. Commun.* **238**, 357–360.
12. Malik, Z., Amir, S., Pál, G., Buzás, Z., Várallyay, E., Antal, J. *et al*. (1999). Proteinase inhibitors from desert locust, *Schistocerca gregaria*: engineering of both P(1) and P(1)′ residues converts a potent chymotrypsin inhibitor to a potent trypsin inhibitor. *Biochim. Biophys. Acta*, **1434**, 143–150.
13. Gáspári, Z., Patthy, A., Gráf, L. & Perczel, A. (2002). Comparative structure analysis of proteinase inhibitors from the desert locust, *Schistocerca gregaria*. *Eur. J. Biochem.* **269**, 527–537.
14. Szenthe, B., Gáspári, Z., Nagy, A., Perczel, A. & Gráf, L. (2004). Same fold with different mobility: backbone dynamics of small protease inhibitors from the desert locust, *Schistocerca gregaria*. *Biochemistry*, **43**, 3376–3384.
15. Patthy, A., Amir, S., Malik, Z., Bódi, A., Kardos, J., Asbóth, B. & Gráf, L. (2002). Remarkable phylum selectivity of a *Schistocerca gregaria* trypsin inhibitor: the possible role of enzyme-inhibitor flexibility. *Arch. Biochem. Biophys.* **398**, 179–187.
16. Kellenberger, C., Ferrat, G., Leone, P., Darbon, H. & Roussel, A. (2003). Selective inhibition of trypsins by insect peptides: role of P6-P10 loop. *Biochemistry*, **42**, 13605–13612.
17. Roussel, A., Mathieu, M., Dobbs, A., Luu, B., Cambillau, C. & Kellenberger, C. (2001). Complexation of two proteic insect inhibitors to the active site of chymotrypsin suggests decoupled roles for binding and selectivity. *J. Biol. Chem.* **276**, 38893–38898.
18. Helland, R., Otlewski, J., Sundheim, O., Dadlez, M. & Smalas, A. O. (1999). The crystal structures of the complexes between bovine beta-trypsin and ten P1 variants of BPTI. *J. Mol. Biol.* **287**, 923–942.
19. Katona, G., Wilmouth, R. C., Wright, P. A., Berglund, G. I., Hajdu, J., Neutze, R. & Schofield, C. J. (2002). X-ray structure of a serine protease acyl-enzyme complex at 0.95-Å resolution. *J. Biol. Chem.* **277**, 21962–21970.
20. Bode, W., Turk, D. & Karshikov, A. (1992). The refined 1.9-A X-ray crystal structure of D-Phe-Pro-Arg chloromethylketone-inhibited human alpha-thrombin: structure analysis, overall structure, electrostatic properties, detailed active-site geometry, and structure-function relationships. *Protein Sci.* **1**, 426–471.
21. Roach, J. C., Wang, K., Gan, L. & Hood, L. (1997). The molecular evolution of the vertebrate trypsinogens. *J. Mol. Evol.* **45**, 640–652.
22. Kardos, J., Bódi, A., Závodszky, P., Venekei, I. & Gráf, L. (1999). Disulfide-linked propeptides stabilize the structure of zymogen and mature pancreatic serine proteases. *Biochemistry*, **38**, 12248–12257.
23. Lu, W., Apostol, I., Quasim, M. A., Warne, N., Wynn, R., Zhang, W. L. *et al*. (1997). Binding of amino acid side-chains to S1 cavities of serine proteinases. *J. Mol. Biol.* **266**, 441–461.
24. Bode, W., Wei, A. Z., Huber, R., Meyer, E., Travis, J. & Neumann, S. (1986). X-ray crystal structure of the complex of human leukocyte elastase (PMN elastase) and the third domain of the turkey ovomucoid inhibitor. *EMBO J.* **5**, 2453–2458.
25. Grutter, M. G., Priestle, J. P., Rahuel, J., Grossenbacher, H., Bode, W., Hofsteenge, J. & Stone, S. R. (1990). Crystal structure of the thrombin-hirudin complex: a novel mode of serine protease inhibition. *EMBO J.* **9**, 2361–2365.

26. Lee, G. F., Lazarus, R. A. & Kelley, R. F. (1997). Potent bifunctional anticoagulants: Kunitz domain-tissue factor fusion proteins. *Biochemistry*, **36**, 5607–5611.
27. Roberge, M., Peek, M., Kirchhofer, D., Dennis, M. S. & Lazarus, R. A. (2002). Fusion of two distinct peptide exosite inhibitors of Factor VIIa. *Biochem. J.* **363**, 387–393.
28. Bode, W. & Huber, R. (1992). Natural protein proteinase inhibitors and their interaction with proteinases. *Eur. J. Biochem.* **204**, 433–451.
29. Zwilling, R., Pfleiderer, G., Sonneborn, H.-H., Kratz, V. & Stucky, I. (1969). The evolution of endopeptidases-V. Common and different traits of bovine and crayfish trypsin. *Comp. Biochem. Physiol.* **28**, 1275–1287.
30. Leslie, A. G. W. (1992). Recent changes to the MOSFLM package for processing film and image plate data. *Joint CCP4 and ESF-EAMCB Newsletter on Protein Crystallography*, **26**, 1–9.
31. Evans, P. R. (1997). SCALA. *Joint CCP4 and ESF-EAMBC Newsletter on Protein Crystallography*, **33**, 22–24.
32. CCP4. (1994). The CCP4 suite: programs for protein crystallography. *Acta Crystallog. sect. D*, **50**, 760–763.
33. Vagin, A. & Teplyakov, A. (1997). MOLREP: an automated program for molecular replacement. *J. Appl. Crystallog.* **30**, 1022–1025.
34. Katona, G., Berglund, G. I., Hajdu, J., Gráf, L. & Szilágyi, L. (2002). Crystal structure reveals basis for the inhibitor resistance of human brain trypsin. *J. Mol. Biol.* **315**, 1209–1218.
35. Lamzin, V. S. & Wilson, K. S. (1993). Automated refinement of protein models. *Acta Crystallog. sect. D*, **49**, 129–147.
36. Jones, T. A., Zou, J. Y., Cowan, S. W. & Kjeldgaard, M. (1991). Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallog. sect. A*, **47**, 110–119.
37. Scheldrick, G. & Schneider, T. (1997). SHELXL: High-resolution refinement. *Methods Enzymol.* **277**, 319–343.
38. Brunger, A. T. (1992). Free R value: a novel statistical quantity for assessing the accuracy of crystal structures. *Nature*, **355**, 472–475.
39. Hartley, B. S. & Kauffman, D. L. (1966). Corrections to the amino acid sequence of bovine chymotrypsinogen A. *Biochem. J.* **101**, 229–231.
40. Hooft, R. W., Vriend, G., Sander, C. & Abola, E. E. (1996). Errors in protein structures. *Nature*, **381**, 272.
41. Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallog.* **26**, 283–291.
42. Merritt, E. A. (1999). Expanding the model: anisotropic displacement parameters in protein structure refinement. *Acta Crystallog. sect. D*, **55**, 1109–1117.
43. Marquart, M., Walter, J., Deisenhofer, J., Bode, W. & Huber, R. (1983). The geometry of the reactive site and of the peptide groups in trypsin, trypsinogen and its complexes with inhibitors. *Acta Crystallog. sect. B*, **39**, 480.
44. Madsen, D. & Kleywegt, G. J. (2002). Interactive motif and fold recognition in protein structures. *J. Appl. Crystallog.* **35**, 137–139.
45. Lindahl, E., Hess, B. & van der Spoel, D. (2001). GROMACS 3.0: A package for molecular simulation and trajectory analysis. *J. Mol. Model.* **7**, 306–317.
46. Stocker, U. & van Gunsteren, W. F. (2000). Molecular dynamics simulation of hen egg white lysozyme: a test of the GROMOS96 force field against nuclear magnetic resonance data. *Proteins: Struct. Funct. Genet.* **40**, 145–153.
47. Morris, G. M., Goodsell, D. S., Halliday, R. S., Huey, R., Hart, W. E., Belew, R. K. & Olson, A. J. (1998). Automated docking using a Lamarckian genetic algorithm and and empirical binding free energy function. *J. Comp. Chem.* **19**, 1639–1662.
48. Stouten, P. F. W., Frömmel, C., Nakamura, H. & Sander, C. (1993). An effective solvation term based on atomic occupancies for use in protein simulations. *Mol. Simul.* **10**, 97–120.
49. Hetényi, C. & van der Spoel, D. (2002). Efficient docking of peptides to proteins without prior knowledge of the binding site. *Protein Sci.* **11**, 1729–1737.
50. Hetényi, C., Maran, U. & Karelson, M. (2003). A comprehensive docking study on the selectivity of binding of aromatic compounds to proteins. *J. Chem. Inf. Comput. Sci.* **43**, 1576–1583.
51. DeLano, W. L. (2002). *PyMOL Molecular Graphics System*, DeLano Scientific, San Carlos CA, USA.
52. Kraulis, Per J. (1991). MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallog.* **24**, 946–950.

Edited by R. Huber

(Received 10 February 2005; received in revised form 8 April 2005; accepted 12 April 2005)