Glotzbach, B., Schmelz, S., Reinwarth, M., Christmann, A., Heinz, D.W., Kolmar, H.

Structural characterization of Spinacia oleracea trypsin inhibitor III (SOTI-III)

(2013) Acta Crystallographica Section D: Biological Crystallography, 69 (1), pp. 114-120.

Person to contact for this submission:
Harald Kolmar (Kolmar@biochemie-tud.de)

IUCr

Acta Crystallographica Section D research papers

Structural characterization of *Spinacia oleracea* trypsin inhibitor III (SOTI-III)

Bernhard Glotzbach τ[a1], Stefan Schmelz τ[b], Michael Reinwarth[a], Andreas Christmann[a], Dirk W. Heinz[b] and Harald Kolmar[a]*

[a]Institute for Organic Chemistry and Biochemistry, Technische Universität Darmstadt, Petersenstraße 22, Darmstadt, 64287, Germany, and [b]Department of Molecular Structural Biology, Helmholtz Centre for Infection Research (HZI), Inhoffenstraße 7, Braunschweig, 38124, Germany

Correspondence email: Kolmar@biochemie-tud.de

[1] τ - These authors contributed equally.

Keywords: trypsin inhibitor, knottins, cysteine-knot miniproteins, SOTI-III, serine proteases

Synopsis

The crystal structure of SOTI-III in complex with bovine pancreatic trypsin has been solved to 1.7 Å resolution.

Abstract

In the last decades several canonical serine protease inhibitor families have been classified and characterized. Other than most trypsin inhibitors, those from garden four o´clock (*Mirabilis jalapa*) or spinach (*Spinacia oleracea*) do not share sequence similarities and were proposed to form the new *Mirabilis* serine proteinase inhibitor family. These 30-40 amino acid inhibitors comprise a defined disulfide bridge topology and belong to the cystine-knot miniproteins (knottins). So far no atomic structure of this inhibitor family has been solved. Herein, we report the first structure of *Spinacia oleracea* trypsin inhibitor III (SOTI-III) in complex with bovine pancreatic trypsin. The inhibitor was synthesized with solid phase peptide synthesis in a multi-milligram scale and assayed to test the inhibitory activity as well as the binding properties. The structure confirms the proposed cysteine bridge topology. The structural features of SOTI-III suggest that it belongs to a new canonical serine protease inhibitor family with promising properties for the use in protein engineering and medical applications.

1. Introduction

Proteases and their natural inhibitors play key roles in biochemical and physiological processes. They are involved in protein digestion, defense mechanisms of plants, degradation of misfolded proteins and turnover, signalling, enzyme activation, regulation, and many other processes (Antao & Malcata, 2005; Ehrmann & Clausen, 2004). Based on the reaction mechanism proteases are basically categorized into serine, cysteine, aspartic and metalloproteases. Protease activity has to be tightly controlled and two basic regulatory mechanisms have been described: production of inactive precursors (Khan & James, 1998) and active inhibition by proteins, peptides or non-proteinaceous compounds (Clardy *et al.*, 2006; Krowarsch *et al.*, 2003). For the latter mechanism the inhibitor blocks the active site of the proteolytic enzyme.

Serine protease inhibitors (SPIs) are classified based on their mode of inhibition either as canonical, non-canonical inhibitors or serpins (Krowarsch *et al.*, 2003; Otlewski *et al.*, 1999). Serpins form an irreversible covalent enzyme-inhibitor complex, causing large conformational changes of the enzyme and thus disrupting the active site (Stubbs *et al.*, 1990; Silverman *et al.*, 2001). In contrast the non-canonical inhibitors, such as hirudin, bind with their N-terminus *via* two interaction regions (Huntington & Carrell, 2001). This results in a two-step kinetic, mainly due to a substrate-mimic in the active site and an additional exosite binding on the surface of the protease (Farady *et al.*, 2008). Canonical inhibitors with a size of 3-21 kDa per domain obstruct the active site with a single convex inhibitor loop without causing conformational changes (Bode & Huber, 2000). The SPIs P_1 residue is essential for complex formation and has substantial influence on the association energy (Lu *et al.*, 1993).

From a structural point of view, there are various types of serine protease inhibitors such as α-helical proteins, β-sheet proteins, α/β-proteins and small disulfide rich proteins with different folds. While the serpin superfamily shares a highly ordered α_1-antitrypsin motif (Huber & Carrell, 1989; Loebermann *et al.*, 1984) the non-canonical SPIs have no well-defined structural motif. However, they share a disordered N-terminal region that forms a parallel β-sheet with the protease upon complex formation (Otlewski *et al.*, 1999). Some canonical serine protease inhibitors, such as the BPTI, Kazal, potato, cereal, SSI, STI or the ecotin family have typical secondary structure elements and a conserved hydrophobic core. Others such as squash, Bowman-Birk, grasshopper and Ascaris lack this hydrophobic core, but maintain a defined tertiary structure, which is mainly formed and stabilized by intramolecular disulfide bonds (Krowarsch *et al.*, 2003; Bateman & James, 2011).

Among those, members of the cystine-knot miniprotein or knottin family show extraordinary chemical, thermal and proteolytic stability due to their unique disulfide linkage (Kolmar, 2009; Heitz *et al.*, 2008; Werle *et al.*, 2008). Knottins are broadly distributed in nature and can be isolated from various organisms like arthropoda, mollusca, porifera, vertebrata, fungi and plants (Kolmar, 2009).

They act as ion-channel blockers (Williams *et al.*, 2008), protease inhibitors (Avrutina *et al.*, 2005), have insecticidal (Jennings *et al.*, 2001), antimicrobial (Tam *et al.*, 1999), anti-HIV (Gustafson *et al.*, 2004) or hemolytic activity (Daly *et al.*, 1999). Several studies showed that the loop length as well as residue composition could be modified as long as the cysteine pattern is preserved (Avrutina *et al.*, 2005; Kimura *et al.*, 2011; Sommerhoff *et al.*, 2010). Due to their small size, their long-term stability and their likely oral availability, they are promising candidates for diagnostic and therapeutic applications (Kolmar, 2010; Werle *et al.*, 2008; Kimura *et al.*, 2011).

In 2007 a new family of serine protease inhibitors with a knottin-like disulfide bridge connectivity of CysI – CysIV, CysII – CysV and CysIII – CysVI has been described (Kowalska *et al.*, 2007). Based on their unique sequence and the C-terminal location of the inhibitor loop those SPIs were proposed to form the new *Mirabilis* SPI family (Figure 1; http://knottin.cbs.cnrs.fr/). Unfortunately, no subsequent new publication with additional information or more detailed description of this family has been reported so far.

Here we report the first multi-milligram solid phase synthesis of SOTI-III and a F14A mutant. The folded mini-proteins were utilized for structural elucidation in complex with pancreatic trypsin and kinetic analysis.

2. Materials and Methods

2.1. Synthesis of linear precursors and oxidative folding

Peptides were assembled on Fmoc-Gln preloaded TentaGel resin (*Rapp Polymere)* using standard Fmoc-SPPS chemistry on a fully automated microwave CEM *Liberty*® peptide synthesizer (see supporting information). Cleavage from the solid support was performed using 94 % (v/v) TFA, 2 % (v/v) H_2O, 2 % (w/v) DTT, 1 % (v/v) triethylsilane, and 1 % (v/v) anisole. After precipitation of the crude peptide in methyl-*tert*.-butyl ether (MTBE), the pellet was washed twice with MTBE, dried, dissolved in a mixture of acetonitrile and water (1:4, v/v), lyophilized and finally analyzed by RP-HPLC and LC-MS (see section below). Oxidative folding was performed as previously reported (Avrutina *et al.*, 2005) followed by RP-HPLC purification.

2.2. RP-HPLC and LC-MS analysis

Analytical RP-HPLC was performed with a Varian LC 920 system equipped with a *Phenomenex* Synergi 4µ Hydro-RP 80 Å (250 x 4.6 mm, 4 µm) column (see appendix) at a flow rate of 1 mL/min. Semi-preparative RP-HPLC was performed with a Varian LC 940 system equipped with an axia-packed *Phenomenex* Luna C18 (250 x 21.2 mm, 5 µm, 100 Å) column using linear acetonitrile gradients (Appendix A) with a flow rate of 18 mL/min.

LC-MS was performed with a *Shimadzu* LC-MS 2020 equipped with a *Phenomenex* Jupiter C4 (50 x 1 mm, 5 µm, 300 Å) column using linear acetonitrile gradients with a flow rate of 0.2 mL/min (Appendix B).

2.3. Crystallization and data collection

Trypsin from bovine pancreas (Sigma-Aldrich) was dissolved in 1 mM HCl (pH 2.0) and 10 mM $CaCl_2$. Protein was purified on a Superdex 75 16/60 column (GE Healthcare) in 25 mM MES pH 5.5, 50 mM NaCl and 10 mM $CaCl_2$ at 277 K and concentrated to 12 mg/mL. Prior to crystallization set up lyophilized SOTI-III was added to 12 mg/mL trypsin (in 25 mM MES pH 5.5, 50 mM NaCl and 10 mM $CaCl_2$) to a final concentration of 2.5 mM and kept on ice for 30 min. Initial crystallization screening was performed in Intelli-Plate™ 96-3 plates (Art Robins) using a HoneyBee 961 dispensing robot (Digilab Genomic Solutions). Initial crystals were obtained with the JCSG$^+$ screen (NeXtal; QIAGEN) in condition E12 (tube 60: 0.1 M imidazole pH 8.0 and 10% (w/v) PEG 8000). Diffraction-quality crystals were obtained after several days at 292 K by mixing equal volumes of protein complex solution with precipitant solution (0.1 M imidazole pH 7.5 and 12% (w/v) PEG 8000) by hanging-drop vapor-diffusion crystallization in 15 well plates (EasyXtal; QIAGEN) with a reservoir volume of 500 µL. Prior to flash-freezing in liquid nitrogen crystals (> 0.8 x 0.1 mm) were cryo-protected with 25 % (w/v) glycerol.

For SOTI-III F14A / trypsin co-complex crystals lyophilized miniprotein was added to 11.5 mg/mL trypsin to a final concentration of 1.5 mM and incubated for 30 min on ice. Best crystals were obtained directly in condition B3 of the JCSG$^+$ initial screen (tube 15: 0.1M BICINE pH 9, 20% (w/v) PEG 6000; NeXtal, QIAGEN) with equal volumes of complex protein and precipitant solution. Crystals grew at 292 K after several days and were cryo-protected with 33 % (w/v) glycerol prior flash freezing in liquid nitrogen. X-ray diffraction data was collected at the BESSY synchrotron beamline MX-14.1. Data was indexed, integrated and scaled with the XDS/XSCALE package (Kabsch, 2010). Phases were obtained with Phaser (McCoy *et al.*, 2007) using trypsin coordinates (PDB code: 2XTT) as a search model. The coordinates were refined with Phenix.refine (Adams *et al.*, 2004) and manually checked and corrected with COOT (Emsley & Cowtan, 2004). SOTI-III coordinates were manually built with COOT. The refined complex structure of SOTI-III/trypsin was used as molecular replacement model for SOTI-III F14A/trypsin complex. Figures and structure alignments were prepared using PyMOL (Schrodinger, 2010). Data and refinement statistics are summarized in Table 1.

2.4. Enzymatic methods

Bovine pancreatic trypsin (Sigma-Aldrich) was active-site titrated (Chase & Shaw, 1967; Jameson *et al.*, 1973) using 4-methylumbelliferyl guanidinobenzoate (MUGB; 1µM; Sigma-Aldrich) in reaction

buffer (50 mM Tris/HCl, 150 mM NaCl, 0.01% (v/v) Triton X-100, 0.01% (w/v) sodium azide, pH 7.6).

The concentration of inhibitory active SOTI was determined assuming a 1:1 interaction of enzyme and inhibitor (Avrutina *et al.*, 2005; Pohlig *et al.*, 1996). Several inhibitor concentrations (0.05 – 1.5 µM) were incubated with trypsin (0.2 µM) for 30 min at room temperature. Subsequently, the residual activity was quantified over 5 min by monitoring the relative fluorescence of the hydrolyzed carbobenzoxy-L-arginine-7-amino-4-methylcoumarin (75 µM; Sigma-Aldrich). The reactions were carried out in 96-well microtiter plates (Greiner Bion-One, flat bottom, black) and were monitored in a Tecan Genios reader (ex. 360 nm, em. 465 nm). The normalized reaction rates were fitted using equation (1), with unfixed E_0. The ratio of active sites (see above) and the fitted E_0 gives a correction factor for the active inhibitor concentration.

To determine the inhibition constants the reaction velocity was monitored after 30 min pre-incubation of enzyme and inhibitor. The substrate hydrolysis of Boc-QAR-pNA (250 µM, Bachem) was analyzed over a period of 30 min at 405 nm (Tecan Genios). The apparent inhibition constant (K_i^{app}) was calculated by fitting (Sigmaplot 11, Marquard-Levenberg algorithm) the relative reaction velocities using the Morrison equation (1) (Morrison, 1969).

$$(1) \quad \frac{v}{v_0} = 1 - \frac{\left(E_0 + I_0 + k_i^{app}\right) - \sqrt{\left(E_0 + I_0 + k_i^{app}\right)^2 - 4E_0I_0}}{2E_0} \qquad (2) \quad K_i = \frac{K_i^{app}}{\left(1 + \frac{[S]}{K_m}\right)}$$

All measurements were carried out in triplicate. The substrate independent inhibition constants (K_i) were calculated from K_i^{app} and K_m of the enzyme using equation (2). The Michaelis Menten constant K_m for the substrate and trypsin was determined previously (Tischler *et al.*, 2012).

3. Results and Discussion

3.1. SPPS & oxidative folding

SOTI-III and the F14A variant were synthesized by microwave-assisted solid phase peptide synthesis. The purity of the linear precursors was verified by HPLC (Figure 2) and LC-ESI-MS analysis (Figure S1). Oxidative folding of the crude peptide was conducted in an aqueous buffer system and monitored by analytical HPLC (Figure 2) and LC-ESI-MS (Figure S1). After the final semi-preparative HPLC purification a total yield of 23.8 % and 16.2 % for SOTI-III and the F14A variant respectively was obtained

3.2. Structural studies

Co-complex crystals of purified SOTI-III with bovine pancreatic trypsin had $P2_1$ symmetry and diffracted to 1.7 Å. The structure shows additional F_o-F_c electron density for the inhibitor covering the active site of trypsin (Figure S2). The asymmetric unit is packed with three trypsin molecules, each blocked by the inhibitor (Figure 3a). SOTI-III secondary structure shows the proposed C4 – C21, C11 – C25 and C20 – C36 disulfide linkage (Figure 3b and 4a (Kowalska *et al.*, 2007)). While the first cysteine bridge shapes a right handed hook, the second and last one forms a right and left handed spiral, respectively (Table S1). The inhibitor overall structure is mainly constructed by β-turns (Figure 4 and Table S2). The turns are interrupted by disulfide bonds, three short anti-parallel β-strands (residues 9-11, 26-28 and 33-36) and a short four amino acid long helix (residues 16-19). The inhibitor loop (residues 26-34), which contains the S_1 site filling Arg32 (P_1 residue), forms a β-hairpin with a γ-turn (Figure 3c and 4, Table S2b).

SOTI-III coordinates at the active site of each trypsin molecule without causing major dislocation of protease loops. Here the averaged RMS with 0.227±0.026 Å of all three trypsin/inhibitor complexes in the asymmetric unit (ASU) is in the same range as the RMS of each trypsin/inhibitor complex when compared with apo-trypsin (RMS = 0.259±0.008 Å) (Figure S3). SOTI-III is most flexible in region of loop 1, loop 3 and a three-residue helix as seen in a B-factor putty of all three monomers (Figure S4).

On average 15 residues of the inhibitor (~750 $Å^2$) interact with 24 residues of trypsin covering ~980 $Å^2$ of the protease surface (Table S3). In total 12 hydrogen bonds are involved in coordination of the inhibitor. Notably, the P_1 residue Arg32 forms eight of these interacting H-bonds (Figure S5).

Phe14 of SOTI-III binds to a cleft apart from the active site, which is formed by Ile78, Trp144 and Pro155 of trypsin (Figure 3d). Interestingly, a detailed search in the Protein Data bank showed that this cleft is not addressed by other trypsin inhibitors (Figure 5, appendix D).

To investigate the role of Phe14 upon trypsin binding and inhibition a F14A variant of SOTI-III was synthesized and co-crystallized. The co-complex with the SOTI-III F14A, which crystallized in the same $P2_1$ space group, shows that in monomer C of trypsin the residual space, which was formerly filled by the side chain of Phe14 (Figure 3d), is partly occupied by the protease residue Tyr154 (Figure S6). In one of the remaining two monomers the electron density is rather weak for the Tyr154 side chains, suggesting that the side chain might swap between those two rotamers.

The K_i of SOTI-III wild type and of the F14A variant were determined as described in materials and methods. They were found to inhibit trypsin in the nanomolar range. The substrate independent inhibition constants for SOTI-III and for SOTI-III F14A were calculated from respective inhibition curves (Figure 2c) with 60.9 ± 8.4 and 201.8 ± 27.7 nM, respectively. This indicates that Phe14

contributes to trypsin binding. However, the main inhibitory effect is mediated by the C-terminal inhibitor loop. This is supported by the fact that more than half of the inhibitor coordinating H-bonds are formed by Arg32 within this loop region. This is not surprising, since other P_1 residues were also reported to contribute strongly to inhibitor binding (Krowarsch *et al.*, 2005).

3.3. Comparison studies with other trypsin inhibitors

The structure of SOTI-III was previously modeled *in silico* (http://knottin.cbs.cnrs.fr/). Although the calculated structure is close to that of the co-crystallization with trypsin, significant conformational differences in the C-terminal inhibitory loop region (residues 25 – 35) are present. The average RMSD of the superposed structure is 2.33 Å with the largest RMS deviation of over 8 Å in the inhibitor loop region (Figure S7).

SOTI-III exhibits the unique, but well-defined pseudo-knot motif, which is typical for the members of the "knottin" family (Figure 3b, Figure 4). This structural element determines the stability and rigidity, similar to the majority of known canonical protease inhibitors (squash, Bowman-Birk, grasshopper, hirustasin, chelonianin, and Ascaris) that lack a hydrophobic core and extensive secondary structure. It contains a canonical inhibitor loop located at the C-terminus thereby defining it as a member of the huge family of canonical serine protease inhibitors.

The C-terminal location of the inhibitor loop is a unique feature of SOTI-III. Due to the amino acid sequence, the additional interaction of Phe14 and the location of the inhibitor loop we support the notion of Kowalska (Kowalska *et al.*, 2007) that SOTI-III together with other *Mirabilis* and *Spinacea* serine proteinase inhibitors form a new family of inhibitors.

Acknowledgements

This work was supported by Deutsche Forschungsgemeinschaft (DFG) through SPP1623 and Bundesministerium für Bildung und Forschung (BMBF).

References

Adams, P. D., Gopal, K., Grosse-Kunstleve, R. W., Hung, L. W., Ioerger, T. R., McCoy, A. J., Moriarty, N. W., Pai, R. K., Read, R. J., Romo, T. D., Sacchettini, J. C., Sauter, N. K., Storoni, L. C. & Terwilliger, T. C. (2004). *J Synchrotron Radiat* 11, 53-55.

Antao, C. M. & Malcata, F. X. (2005). *Plant Physiol Biochem* 43, 637-650.

Avrutina, O., Schmoldt, H. U., Gabrijelcic-Geiger, D., Le Nguyen, D., Sommerhoff, C. P., Diederichsen, U. & Kolmar, H. (2005). *Biol Chem* 386, 1301-1306.

Bateman, K. S. & James, M. N. (2011). *Curr Protein Pept Sci* 12, 340-347.

Bode, W. & Huber, R. (2000). *Biochim Biophys Acta* 1477, 241-252.

Chase, T., Jr. & Shaw, E. (1967). *Biochem Biophys Res Commun* 29, 508-514.

Clardy, J., Fischbach, M. A. & Walsh, C. T. (2006). *Nat Biotechnol* 24, 1541-1550.

Daly, N. L., Love, S., Alewood, P. F. & Craik, D. J. (1999). *Biochemistry* 38, 10606-10614.

Ehrmann, M. & Clausen, T. (2004). *Annu Rev Genet* 38, 709-724.

Emsley, P. & Cowtan, K. (2004). *Acta Crystallogr D Biol Crystallogr* 60, 2126-2132.

Farady, C. J., Egea, P. F., Schneider, E. L., Darragh, M. R. & Craik, C. S. (2008). *J Mol Biol* 380, 351-360.

Gustafson, K. R., McKee, T. C. & Bokesch, H. R. (2004). *Curr Protein Pept Sci* 5, 331-340.

Heitz, A., Avrutina, O., Le-Nguyen, D., Diederichsen, U., Hernandez, J. F., Gracy, J., Kolmar, H. & Chiche, L. (2008). *BMC Struct Biol* 8, 54.

Huber, R. & Carrell, R. W. (1989). *Biochemistry* 28, 8951-8966.

Huntington, J. A. & Carrell, R. W. (2001). *Sci Prog* 84, 125-136.

Jameson, G. W., Roberts, D. V., Adams, R. W., Kyle, W. S. & Elmore, D. T. (1973). *Biochem J* 131, 107-117.

Jennings, C., West, J., Waine, C., Craik, D. & Anderson, M. (2001). *Proc Natl Acad Sci U S A* 98, 10614-10619.

Kabsch, W. (2010). *Acta Crystallogr D Biol Crystallogr* 66, 133-144.

Khan, A. R. & James, M. N. (1998). *Protein Sci* 7, 815-836.

Kimura, R. H., Jones, D. S., Jiang, L., Miao, Z., Cheng, Z. & Cochran, J. R. (2011). *PLoS One* 6, e16112.

Kolmar, H. (2009). *Curr Opin Pharmacol* 9, 608-614.

Kolmar, H. (2010). *Expert Rev Mol Diagn* 10, 361-368.

Kowalska, J., Pszczola, K., Wilimowska-Pelc, A., Lorenc-Kubis, I., Zuziak, E., Lugowski, M., Legowska, A., Kwiatkowska, A., Sleszynska, M., Lesner, A., Walewska, A., Zablotna, E., Rolka, K. & Wilusz, T. (2007). *Phytochemistry* 68, 1487-1496.

Krowarsch, D., Cierpicki, T., Jelen, F. & Otlewski, J. (2003). *Cell Mol Life Sci* 60, 2427-2444.

Krowarsch, D., Zakrzewska, M., Smalas, A. O. & Otlewski, J. (2005). *Protein Pept Lett* 12, 403-407.

Loebermann, H., Tokuoka, R., Deisenhofer, J. & Huber, R. (1984). *J Mol Biol* 177, 531-557.

Lu, W., Zhang, W., Molloy, S. S., Thomas, G., Ryan, K., Chiang, Y., Anderson, S. & Laskowski, M., Jr. (1993). *J Biol Chem* 268, 14583-14585.

McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C. & Read, R. J. (2007). ***J Appl Crystallogr*** **40, 658-674.**

Morrison, J. F. (1969). ***Biochim Biophys Acta*** **185, 269-286.**

Otlewski, J., Krowarsch, D. & Apostoluk, W. (1999). ***Acta Biochim Pol*** **46, 531-565.**

Pohlig, G., Fendrich, G., Knecht, R., Eder, B., Piechottka, G., Sommerhoff, C. P. & Heim, J. (1996). ***Eur J Biochem*** **241, 619-626.**

Schechter, I. & Berger, A. (1967). ***Biochem Biophys Res Commun*** **27, 157-162.**

Schrodinger, L. (2010).

Silverman, G. A., Bird, P. I., Carrell, R. W., Church, F. C., Coughlin, P. B., Gettins, P. G., Irving, J. A., Lomas, D. A., Luke, C. J., Moyer, R. W., Pemberton, P. A., Remold-O'Donnell, E., Salvesen, G. S., Travis, J. & Whisstock, J. C. (2001). ***J Biol Chem*** **276, 33293-33296.**

Sommerhoff, C. P., Avrutina, O., Schmoldt, H. U., Gabrijelcic-Geiger, D., Diederichsen, U. & Kolmar, H. (2010). ***J Mol Biol*** **395, 167-175.**

Stubbs, M. T., Laber, B., Bode, W., Huber, R., Jerala, R., Lenarcic, B. & Turk, V. (1990). ***EMBO J*** **9, 1939-1947.**

Tam, J. P., Lu, Y. A., Yang, J. L. & Chiu, K. W. (1999). ***Proc Natl Acad Sci U S A*** **96, 8913-8918.**

Tischler, M., Nasu, D., Empting, M., Schmelz, S., Heinz, D. W., Rottmann, P., Kolmar, H., Buntkowsky, G., Tietze, D. & Avrutina, O. (2012). ***Angew Chem Int Ed Engl*** **51, 3708-3712.**

Werle, M., Kolmar, H., Albrecht, R. & Bernkop-Schnurch, A. (2008). ***Amino Acids*** **35, 195-200.**

Williams, J. A., Day, M. & Heavner, J. E. (2008). ***Expert Opin Pharmacother*** **9, 1575-1583.**

Table 1 X-ray data collection and refinement statistics

*Values in parentheses are for the highest resolution shell.

Data collection	**Trypsin / SOTI-III *wt* co-complex**	**Trypsin / SOTI-III F15A co-complex**
Space group	$P2_1$	$P2_1$
Cell dimensions		
a, b, c (Å)	49.4, 66.8, 108.9	48.5, 68.4, 109.8
α, β, γ (°)	90.0, 90.02, 90.0	90.0, 93.3, 90.0
Resolution (Å)	32.0 - 1.7	19.9-2.0
(highest shell (Å))	(1.8 - 1.7)	(2.1 - 2.0)
R_{merge}	8.1 (49.1)*	14.5 (45.0)
I/σI	11.66 (2.85)	8.27 (2.90)
Completeness (%)	97.3 (93.7)	98.7 (98.5)
Redundancy	3.8 (3.8)	3.2 (3.2)
Matthews coefficient (Vm)	2.04	2.02
No. of molecules in unit cell (Z)	6	6
Refinement		
No. Reflections	75888	48006
R_{work}/R_{free}	17.9/22.1	16.7/22.0
B-factors		
- protein	16.7	11.0
- SOTI-III *wt*/F14A	21.0	16.2
R.m.s deviations		
Bond lengths (Å)	0.010	0.003
Bond angles (°)	1.30	0.83
Ramachandran		
Favoured /outliers	97.2/0	97.1/0
Molprobity score/percentile	$1.53/90^{th}$	$1.49/97^{th}$
PDB code	4AOR	4AOQ

*Values in parentheses refer to the highest resolution shell

Figure Legends

Figure 1

Sequence comparison of the *squash* (top) and *mirabilis* (bottom) serine protease inhibitor families. Cysteine connections are not shown for the sake of clarity. Light grey: inhibitory active loop of the compared inhibitors; dark grey: P_1 residue of inhibitors.

Figure 2

Oxidative folding of SOTI-III variants and inhibition curves of bovine pancreatic trypsin. *(a)* HPLC traces of *wild type* linear precursor (**1**) and folded knottin (**2**) at 220 nm. *(b)* HPLC traces of F14A linear precursor (**3**) and folded knottin (**4**) at 220 nm. *(c)* The inhibition curves of SOTI-III (triangles) and SOTI-III F14A (squares) were obtained by fitting data to equation 1, as described in materials and methods. Error bars were calculated from three independent experiments.

Figure 3

Structure of SOTI-III in complex with pancreatic trypsin. *(a)* Trimeric complex of trypsin (surface representation) with SOTI-III (cartoon representation) in the asymmetric unit. *(b)* SOTI-III binding pocket. Cysteine residues are colored green. Disulfide bonds are labeled with residue numbers. *(c)* Close-up on SOTI-III inhibitor loop (red) blocking the catalytic triad (blue). Arg32 is coordinated in the S_1 pocked in a substrate like fashion (Schechter & Berger, 1967). Dashed lines indicate H-bonds. Closest H-bonds of catalytic Ser200 to inhibitor backbone are labeled red. *(d)* Close-up on sub-site where Phe14 of SOTI-III is coordinated. Tyr154 of trypsin (purple) is pointing away from residue 14.

Figure 4

Secondary structure and topology of SOTI-III. *(a)* Secondary structure elements of SOTI-III. The four loops (mainly β-turns) are labeled with L1-L4. Green arrows indicate the three β-strands. The short helix α1 is colored purpled. Disulfide bonds are drawn in black lines. Involved cysteine residues are numbered I-VI. The red U-shaped line symbolizes the β-hairpin with a γ-turn of the inhibitor loop L4. *(b)* Topology of SOTI-III. Residue numbers for boundaries of secondary structure elements are shown. Coloring of secondary structure elements is the same as in *(a)*.

Figure 5

Comparison of SOTI-III coordination with other peptide trypsin inhibitors. *(a)* Superposition of SOTI-III with other canonical inhibitors. Trypsin is shown in surface representation. SOTI-III (chain A) is colored purple, STI in lime (PDB code: 1AVW), SFTI-I in black (PDB code: 1SFI), CPTI-II in blue (PDB code: 2BTC) and BPTI in orange (PDB code: 1F7Z). Inhibitors were superposed based on trypsin main chain atoms. *(b)* Close up on active site residues and the inhibitor loop with P1 residues arginine or lysine. Coloring of inhibitors is as in *(a)*. Active site residues are shown as sticks with carbons, nitrogen and oxygen in white, blue and red, respectively. *(c)* Superposition of 99 deposited trypsin inhibitor structures with the co-complex that are close to residues 12-15 of SOTI-III in chain D (< 3 Å) (see appendix D). SOTI-III is shown in purple and trypsin in white. Superposed trypsin molecules (blue lines) align well, while superposed inhibitors (grey lines) do not align or address the F14 binding site of SOTI-III. Tyr45 of trypsin in the SOTI-III co-complex has a different rotamer conformation. *(d)* Close up on the inhibitor loop region with P1 residue Arg32 of panel *(c)* (red box). In this region the inhibitors superpose well with the SOTI-III inhibitor.

Figure 1

Figure 2

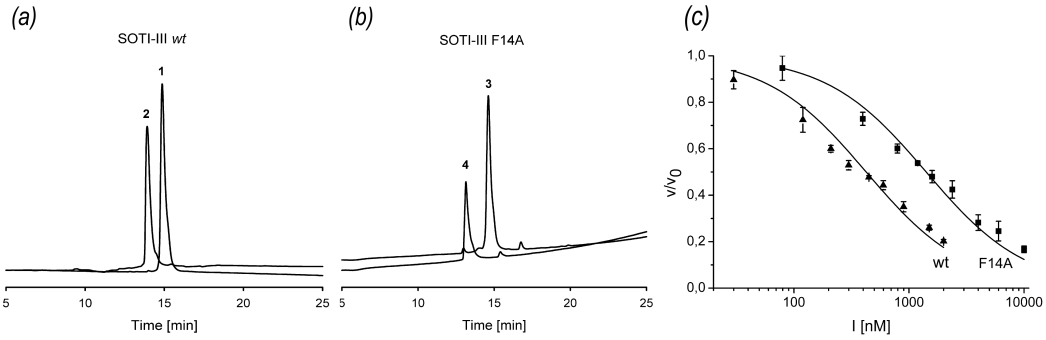

Figure 3

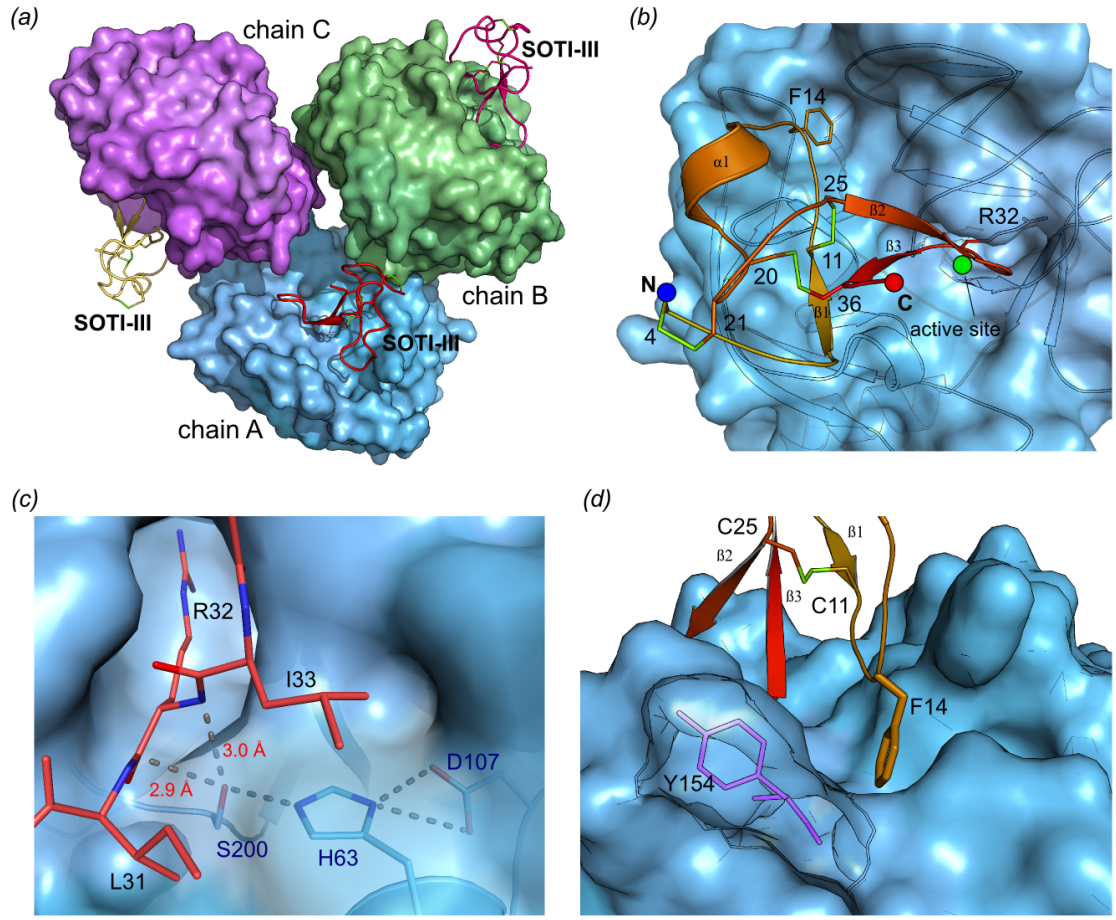

Figure 4

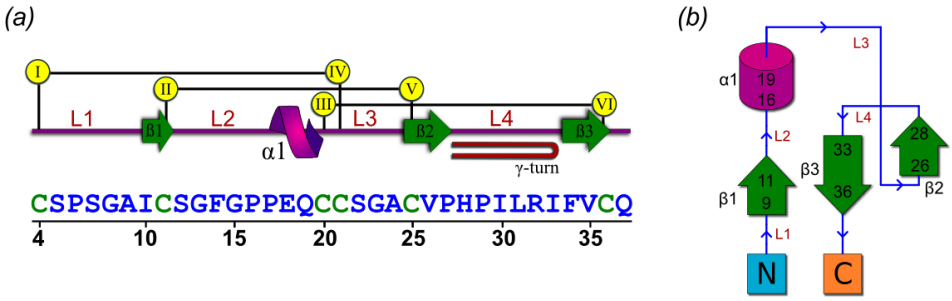

Figure 5