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

IUCr

Acta Crystallographica Section D research papers

Supporting Information:

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.

Appendix A. P-HPLC

For analytical HPLC *Varian* 920-LC analytical HPLC equipped with an UV-Vis detector for detection at 220 nm and 280 nm was used. For analytical RP-HPLC a *Phenomenex* Synergi 4u Hydro-RP 80 Å (250 x 4.6 mm, 4 µm, 8 nm) column was used at a flow rate of 1 mL/min. Preparative RP-HPLC was conducted on a *Varian* 940 LC equipped with a scale-up module and an axia-packed *Phenomenex* Luna C-18 (250 x 21.2 mm, 4 µm, 8 nm) column at a flow rate of 18 mL/min.

Eluent systems for both analytical and semi-preparative scales were performed using mixtures of Millipore grade water with 0.1% aq. TFA as eluent A and 90 % aq. acetonitrile containing 0.1% aq. TFA as eluent B.

Gradients started with isocratic 10 % eluent B for 2 (on analytical scale) or 5 (on semi-preparative scale) min followed by a linear gradient of 10 to 80 % B over 20 min at a flow rate of 1 mL/min or 18 ml/min, respectively.

Appendix B. LCMS

ESI mass spectra were recorded with a Shimadzu LCMS-2020 equipped with a Phenomenex Jupiter 5u C4 LC column (50 x 1 mm, 5 µm, 300 Å) and a Bruker-Franzen Esquire LC mass spectrometer. The eluent system consisted of eluent A (0.1% aq. formic acid, LC-MS grade) and eluent B (acetonitrile containing 0.1% formic acid, LC-MS grade). For clarity only the averaged mass spectra of the relevant segment is shown (section 3).

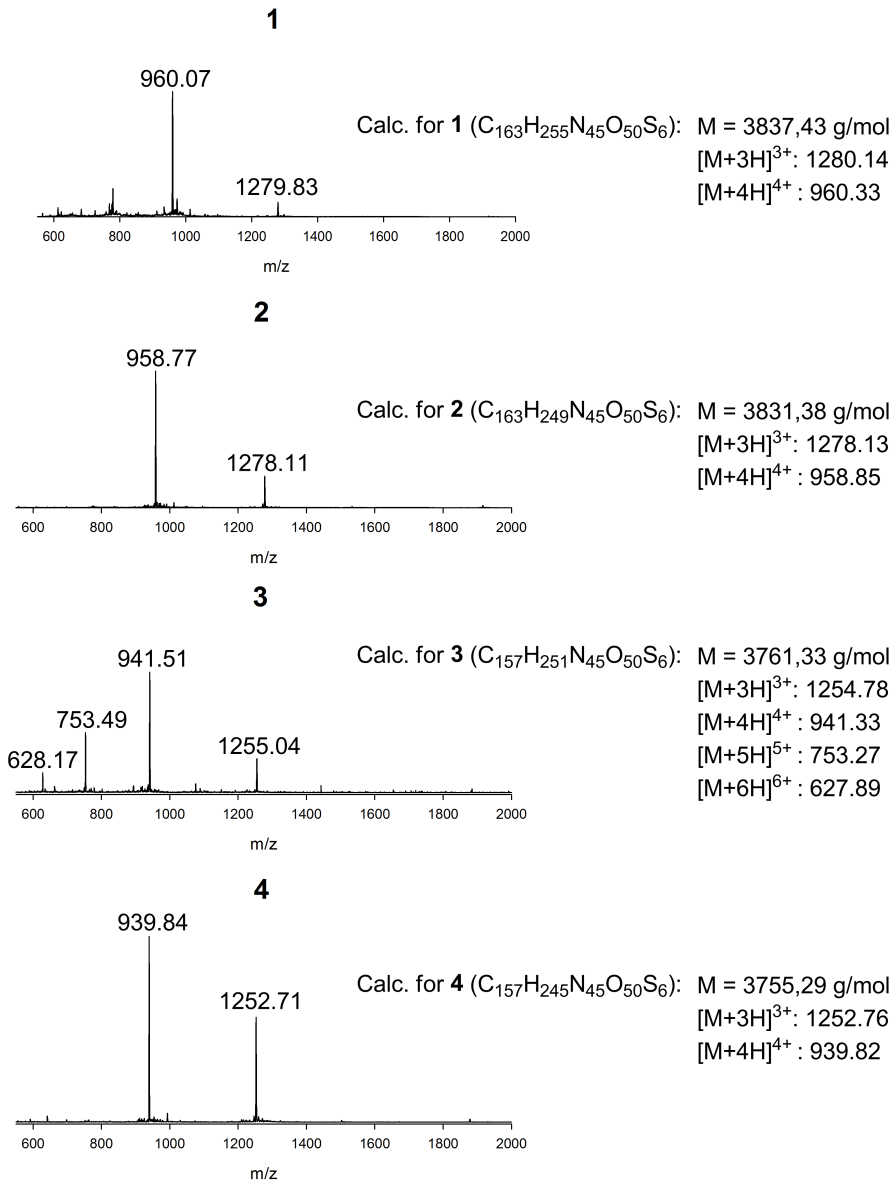


Figure S1. LC-ESI-MS analysis of SOT-III variants before and after oxidation and folding. (**1**) LC-ESI-MS of linear precursor of SOTI-III. (**2**) LC-ESI-MS of folded peptide. (**3**) LC-ESI-MS of linear precursor of SOTI-III F14A. (**4**) LC-ESI-MS of folded peptide.

Appendix C. Structural analysis

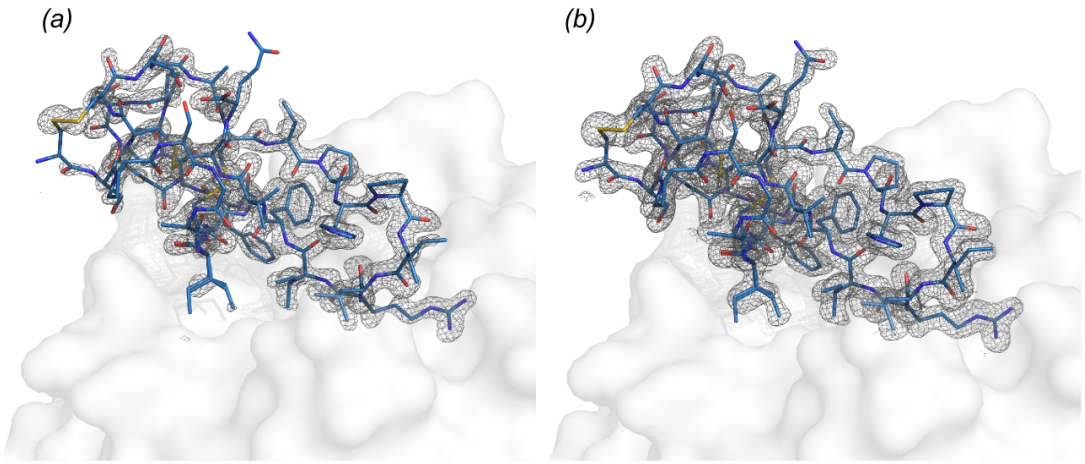
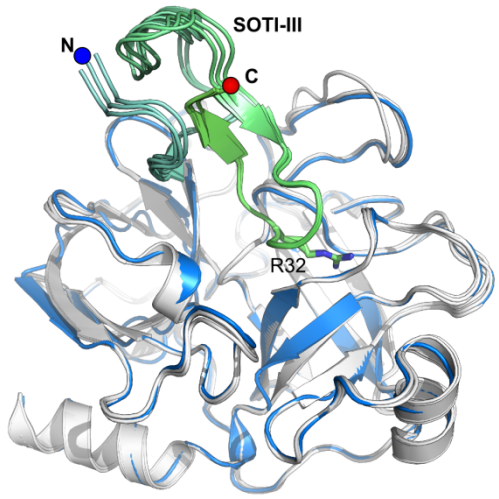


Figure S2. Electron density maps of SOTI-III. *(a)* Un-biased F_o-F_c electron density map of SOTI-III (chain E) contoured at 2.5 ð. SOTI-III residues are superposed from refined coordinates. *(b)* Refined $2F_o$-F_c electron density map of SOTI-III (chain B) contoured at 1.0 ð. Bovine pancreatic trypsin is shown in transparent surface representation.



trypsin/SOTI-III complex	RMS (Å^2)	trypsin apo vs. trypsin/SOTI-III	RMS (Å^2)
chain AD vs chain BE	0.200	apo - chain AD	0.267
chain AD vs chain CF	0.251	apo - chain BE	0.252
chain BE vs chain CF	0.229	apo - chain CF	0.259
average ± STD	0.227 ± 0.026	average ± STD	0.259 ± 0.008

Figure S3. Coordination of SOTI-III does not cause domain or loop shift of trypsin. Trypsin apo-form (blue) was superposed with all three monomers of trypsin (white, chain A, B and C) loaded with trypsin inhibitor each (green, chain D, E and F). Superposition is based on trypsin backbone residues. Root mean square (RMS) values are listed in the table on the right hand site. STD: standard deviation.

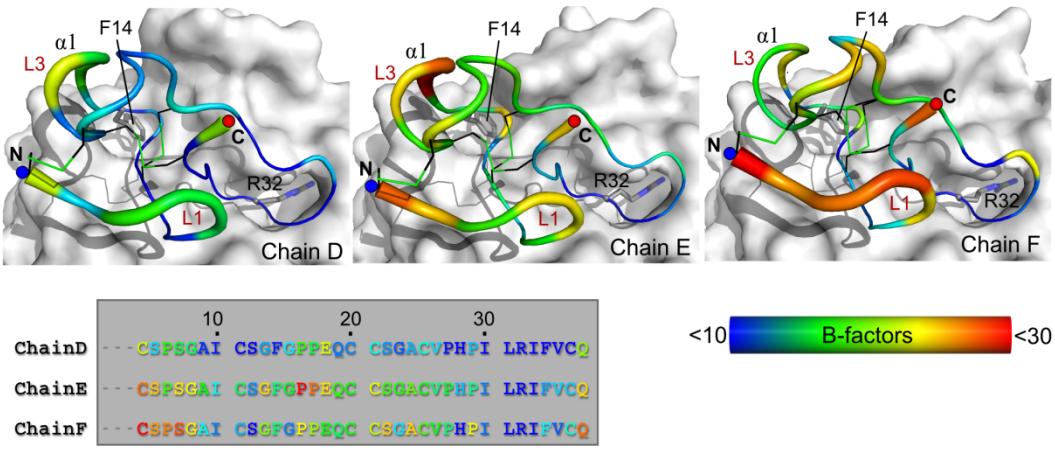


Figure S4. Comparison of the coordination of all three bound SOTI-III monomers. B-factors for the inhibitor residues are indicated by rainbow colors, whereas blue indicates a low and red a high B-factor (scale bar at bottom right corner). The sequence of each chain is colored according to the structural B-factors. Disulfide bridges are shown in black/green. Arg32 and Phe14 are shown in grey sticks. SOTI-III is most flexible in region of L1, α1 and L3. Note that the difference between the B-factors is in the range of 10 to 30 Å^2.

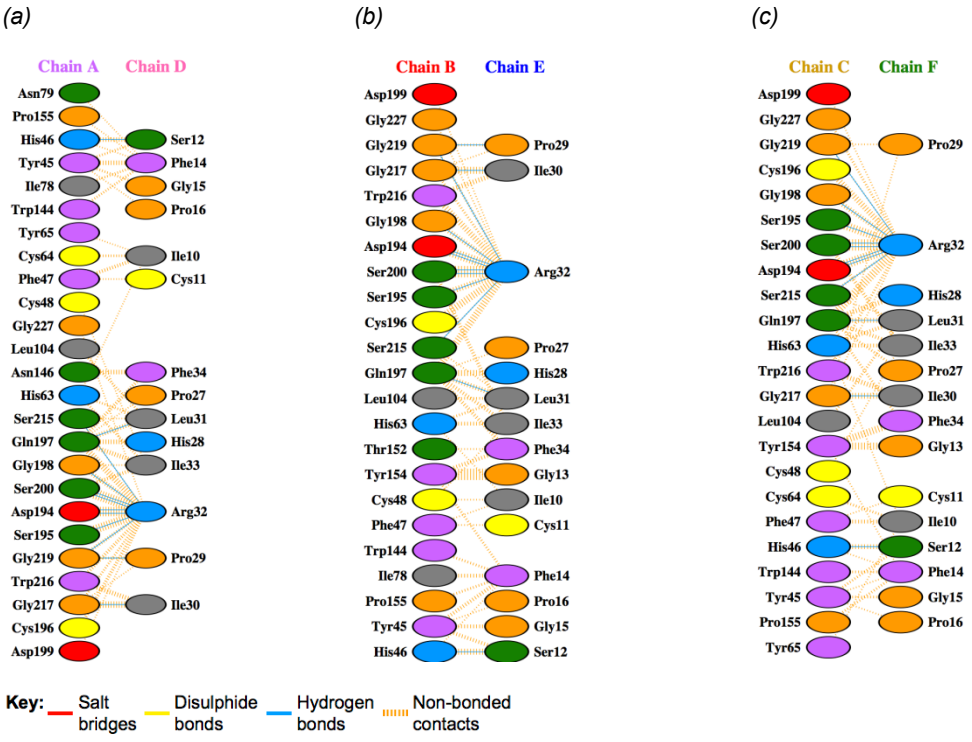


Figure S5. Schematic diagram of interactions between pancreatic trypsin and SOTI-III. Interactions for each of the three complexes are displayed (*(a)*, *(b)* and *(c)*). Diagrams were generated with PDBsum (Laskowski, 2001). Residue colors: Positive (H,K,R); negative (D,E); S,T,N,Q = neutral; A,V,L,I,M = aliphatic; F,Y,W = aromatic; P,G = Pro&Gly; C = cysteine.

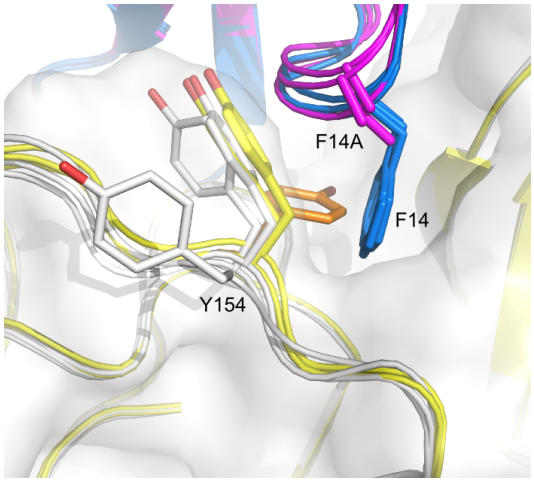


Figure S6. Superposition of *wt* and F14A SOTI-III coordination with pancreatic trypsin at the mutated site. In monomer C of the SOTI-III F14A/trypsin co-complex Tyr154 partly blocks the cavity that was formerly occupied by Phe14 of SOTI-III. SOTI-III *wt* and the F14A variant are colored blue and purple, respectively. Trypsin of the *wt* co-complex is shown in white, with a white transparent surface of chain C. Trypsin of the F14A variant is shown in yellow, except Y154 of chain C, which is highlighted in orange.

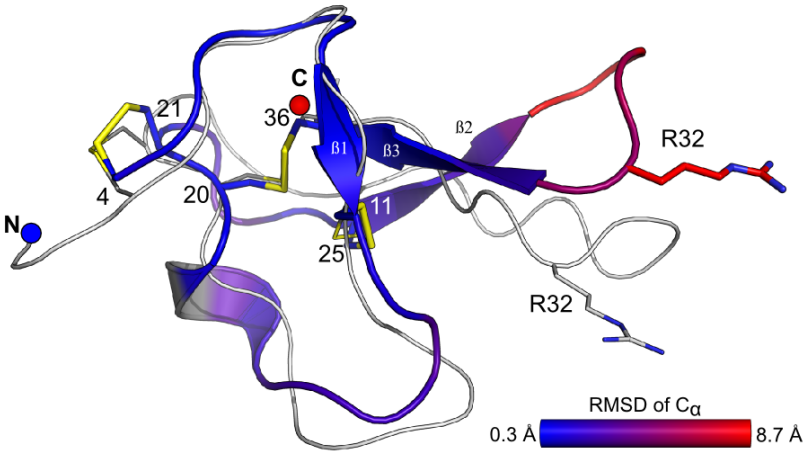


Figure S7. Superposition of SOTI-III and the *in-silico* model of SOTI-III. SOTI-III structure from the co-complex is colored in a blue-red gradient, which correlates to RMSD of C_α atoms with those of the *in-silico* model colored in grey. The active site loop with Arg32 shows the largest discrepancies between the structure and the *in-silico* model.

Table S1. Characterization of SOTI-III disulfide bonds. The type of bridge is abbreviated (RHH: right hand hook; LHS: left handed spiral; RHS: right handed spiral). Chi1, chi2, chi3, chi2' and chi1' values are also recorded. Statistics were acquired with PDBsum (Laskowski, 2001).

Cysteine linkage	1st cysteine	2nd cysteine	Type	Chi1	Chi2	Chi3	Chi2'	Chi1'
1-4	D4	D21	RHH	69.2	79.5	89.0	-42.5	-63.7
2-5	D11	D25	RHS	57.6	101.3	104.5	68.1	177.8
3-6	D20	D36	LHS	-64.0	-73.5	-77.4	-60.3	-66.9

Table S2. Table of β- and γ-turns of SOTI-III inhibitor. *(a)* The following data about each β-turn are shown: The residue numbers of residues i and i+3 in the turn, the one-letter amino acid code of residues i, i+1, i+2 and i+3 in the turn, and the turn type. For each of the central two residues (i+1 and i+2) phi, psi and chi1 are recorded. The final columns show the distance between the C_α atoms of residues i and i+3 and whether or not a hydrogen bond exists between these two residues. *(b)* The following data about the γ-turn are shown: start and end residues of the gamma turn (residues i and i+2), the amino acid sequence of the residues in the turn and the turn type. Phi, psi and chi1 dihedral angles are given for the central residue (i+1). The final column gives the distance between the C_α atoms of residues i and i+2. Statistics were acquired with PDBsum (Laskowski, 2001).

(a)

Loop	Turn	Sequence	Turn type	Residue i+1			Residue i+2			i to i+3 C_α-dist	H-bond
				Phi	Psi	Chi	Phi	Psi	Chi		
L1	Pro6-Ala9	PSGA	II	-56.8	132.8	-60.6	94.1	-11.3	-	5.7	yes
L2	Ser12-Gly15	SGFG	I	-62.6	-14.3	-	-92.0	-12.9	-66.0	5.3	yes
L3	Cys20-Gly23	CCSG	I	-53.4	-37.5	-63.7	-77.2	-14.1	70.9	5.1	yes
L4	His28-Leu31	HPIL	IV	-72.1	-18.7	27.0	-125.7	-56.9	-63.0	5.6	

(b)

No.	Start	End	Sequence	Turn type	Residue i+1			i to i+2 Cα-dist
					Phi	Psi	Chi1	
L4	Leu31	Ile33	L R I	INVERSE	-91.8	48.4	-64.0	5.6

Table S3. Interface statistics of trypsin and SOTI-III co-complex structure. Chain A, B and C correspond to trypsin monomers, respectively. Chain D, E and F are SOTI-III monomers, respectively. Statistics were acquired with PDBsum (Laskowski, 2001).

Chains	No. of interface residues	Interface area (Å^2)	No of H-bonds	No of non-bonded contacts
A : D	25 : 14	696 : 935	12	124
B : E	23 : 15	777 : 994	12	135
C : F	23 : 15	764 : 1012	11	125
average	**24 : 15**	**746 : 980**	**12**	**128**

Appendix D. Supporting Methods

Comparison of SOTI-III/trypsin co-complex coordination with all known trypsin inhibitors structures

Alignment of the SOTI-III/trypsin co-complex with other trypsin inhibitors deposited at the Protein Data was performed to select those close to Phe14 of SOTI-III. An initial list with 494 structures was generated using the advanced search of the database by searching for 'trypsin' with two or more different protein/peptide chains. All structures in this list were superposed with PyMol to trypsin in chain A of the SOTI-III/trypsin co-complex. Using PyMol's selection algebra only those structures were kept, which are blocking the active site of trypsin and which are simultaneously in close proximity (<3Å) to residues 12-15 of SOTI-III in chain D. This resulted in a list of 110 trypsin-inhibitor structures. After manual inspection only the following 99 structures remained: 1AN1, 1AVW, 1AVX, 1BRC, 1BZX, 1C9P, 1C9T, 1CO7, 1D6R, 1EJA, 1EJM, 1EJM, 1EZS, 1EZU, 1EZU, 1F2S, 1F7Z, 1G9I, 1H9H, 1H9I, 1K9O, 1LDT, 1MCT, 1OPH, 1OX1, 1P2I, 1P2J, 1P2K, 1PPE, 1SBW, 1SFI, 1SLV, 1SLW, 1SLX, 1SMF, 1TAB, 1TAW, 1TFX, 1TFX, 1TPA, 1YKT, 1YLC, 1YLD, 1Z7K, 1ZR0, 1ZR0, 2BTC, 2F3C, 2F91, 2FI3, 2FI4, 2FI5, 2FTL, 2FTM, 2G81, 2ILN, 2ILN, 2O9Q, 2PLX, 2PTC, 2QN5, 2QYI, 2RA3, 2RA3, 2STA, 2STB, 2TGP, 2TPI, 2UUY, 2XTT, 3BTD, 3BTE, 3BTF, 3BTG, 3BTH, 3BTK, 3BTM, 3BTQ, 3BTT, 3BTW, 3D65, 3FP6, 3FP7, 3FP8, 3I29, 3M7Q, 3MYW, 3MYW, 3OTJ, 3P92, 3RDZ, 3RDZ, 3TGI, 3TGJ, 3TGK, 3TPI, 4ABI, 4ABJ, 4TPI.