Structure-based insights into the repair of UV-damaged DNA

Meulenbroek, E.M.

7

Crystal structure of a heterodimeric, double-headed Kunitz-type serine protease inhibitor from potato.

The Potato Serine Protease Inhibitor (PSPI) constitutes about 22% of the total amount of proteins in potato tubers (cv. Elkana) making it the plant's most abundant protease inhibitor. PSPI is a heterodimeric, double-headed Kunitz-type serine protease inhibitor that can tightly and simultaneously bind two serine proteases by mimicking the enzyme's substrate with its reactive site loops. We report the crystal structure of PSPI, representing the first heterodimeric double-headed Kunitz-type serine protease inhibitor structure determined. PSPI has a β-trefoil fold and based on the structure, we have identified two reactive site loops bearing residues Phe75 and Lys95.

published in: E.M. Meulenbroek, E.A.J. Thomassen, L. Pouvreau, J.P. Abrahams, H. Gruppen, N.S. Pannu, *Structure of a post-translationally processed heterodimeric double-headed Kunitz-type serine protease inhibitor from potato.*, Acta Cryst. **D68**, 794-799 (2012)

7.1 Introduction

Protease inhibitors are abundantly present in plant seeds and tubers. In the tubers of the potato cultivar Elkana, for example, circa 50 % (w/w) of the total amount of soluble proteins are protease inhibitors (Pouvreau *et al*., 2001). Plant protease inhibitors prevent pathogens or predators feeding on the plant by inhibiting their proteases. Protease inhibitors are, therefore, also expressed in response to wounding of the plant (reviewed in Valueva & Mosolov, 2004). In addition to their importance in plant physiology, protease inhibitors are also reported to show anti-carcinogenic effects: they were shown to prevent carcinogenesis in many different model systems and are effective at very low concentrations (reviewed in Kennedy, 1998).

Kunitz-type protease inhibitors can bind and inhibit serine, cysteine and aspartic proteases (Oliva *et al*., 2010). Plant Kunitz-type protease inhibitors are small 20 kDa proteins that usually have two disulfide bonds and a single reactive site (single-headed), though two reactive site (double-headed) Kunitz-type inhibitors have also been described (e.g. Dattagupta *et al*., 1999 and Azarkan *et al*., 2011). The proteins inhibit proteases by binding tightly to the enzyme's active site via their reactive site loops in a substrate-like manner.

Potato serine protease inhibitor (PSPI) is the most abundant protease inhibitor group in potato tubers and constitutes 42 % of the protease inhibitors present in potato juice (Pouvreau *et al*., 2003). This group of Kunitz-type serine protease inhibitors is composed of seven different isoforms, differing slightly in their pI's. PSPI is reported to inhibit the serine proteases trypsin, chymotrypsin and human leukocyte elastase (Valueva *et al*., 2000). It is expressed as a single polypeptide, but six amino acids are deleted during post-translational processing to yield a protein consisting of a large (16.2 kDa) and small (4.2 kDa) subunit that are held together by a disulfide bridge and non-covalent interactions (Pouvreau *et al*., 2003); literature on Kunitz-type protease inhibitors refers to this multimeric complex as a "dimer", but for clarity, we refer to the complex as a heterodimer. PSPI is proposed to be a double-headed protease inhibitor, i.e. having two independent reactive site loops, since it can form a ternary complex with both trypsin and chymotrypsin (Valueva *et al*., 2000). These two loops, however, have never been identified. To be able to determine the location and residues involved in the reactive site loops and thus gain insight into the mechanism of PSPI and since no three-dimensional atomic structure of a heterodimeric double-headed Kunitz-type protease inhibitor is known, we set out to determine the crystal structure of this protein. Here we present the high resolution structure of PSPI. The structure shows a β-trefoil fold with large protruding loops and based on the structure, we identified the two reactive loops as being centered on Phe75 and on Lys95.

7.2 Methods

Protein purification, crystallization and data collection have been reported previously (Thomassen *et al*., 2004). The crystal diffracted to 1.60 Å, but the data com-

pleteness at this resolution is very low, with only 18.4 % completeness in the highest resolution shell (1.69 - 1.60 Å). Although incomplete, we included this high-resolution data in refinement. The resolution of the data corresponding to a 100 % complete dataset, as calculated by *SFTOOLS* (Bart Hazes, unpublished) is 1.80 Å. The phase problem was solved by molecular replacement using *BALBES* (Long *et al.*, 2008) with the structure of soybean trypsin inhibitor (Song & Suh, 1998) as a search model. The molecular replacement solution was then rebuilt with *ARP/wARP* (Perrakis *et al.*, 1999) resulting in a model that was 87 % complete. Refinement was performed with *REFMAC* using TLS (Murshudov *et al.*, 2011) and manual fitting was performed using *COOT* (Emsley *et al.*, 2010). Data and refinement statistics are shown in Table 7.1. In the analysis of the structure, superpositions with other Kunitz-type serine proteases and root-mean-square deviation calculations were done by *THESEUS* (Theobald & Wuttke, 2006). Structure-based sequence alignment was done with *VAST* (Thompson *et al.*, 2009). Atomic coordinates and structure factors have been deposited in the RCSB Protein Data Bank (accession code 3tc2). Figures were made with *CCP4mg* (Potterton *et al.*, 2004).

7.3 Results and Discussion

7.3.1 Overall structure

The structure of PSPI was determined to high resolution and contains three molecules in the asymmetric unit. These three molecules are very similar with the root-mean-square deviations of the Cα-backbone being 0.23 Å between chains A and B, 0.33 Å between A and C and 0.28 Å between B and C.

The quality of the electron density map is excellent for nearly all of the structure (Figure 7.1(a)). The map is poor around the loop of amino acids 42-46 in chains B and C and around amino acids 72-77 in chains A and B. Nevertheless, for each of these loops, an NCS-related chain with moderate to good density (chain A for loop 42-46 and chain C for loop 72-77) was available to build a complete model. The electron density being well defined only in A for 42-46 and only in C for 72-77 could not be attributed to any obvious interface interactions in the crystal.

The overall structure of PSPI shows a β-trefoil fold (Figure 7.1(b)). It consists of twelve anti-parallel β-strands, which form six two-stranded β-hairpins. Three of these hairpins form a small β-barrel, while the other three hairpins close off the barrel in a triangular fashion. The β-strands are connected by long, protruding loops. Two disulfide bridges are present: between Cys48 and Cys97 and between Cys146 and Cys157. This fold is usually observed for Kunitz-type serine protease inhibitors (e.g. Soybean Trypsin Inhibitor, (Song & Suh, 1998)) and the reactive sites are commonly found in the protruding loops. The β-sheets provide a stable core that can hold together the structure even if incision by the protease takes place, both by the covalent interaction of the disulfide bridges and the numerous non-covalent interactions.

Several structures of Kunitz-type serine protease inhibitors have been deter-

Table 7.1: Refinement and validation statistics

Data collection	
Space group	$P12_11$
Cell dimensions	
a, *b*, *c* (Å)	54.82 x 93.92 x 55.44
α, β, γ (°)	90.00, 100.69, 90.00
Resolution (Å)	54.23–1.60 (1.69-1.60)[a]
Resolution of a 100 % complete data set (Å)	1.80[b]
Wilson plot B-factor	20.1
R_{merge}[c]	0.044 (0.370)
I/σI	14.0 (1.8)
Completeness (%)[b]	70.0 (18.4)
Redundancy	2.1 (1.8)
Total n°. observations	104501
N° unique reflections	50787
N° reflections in R_{free} set	2533
Refinement	
Resolution (Å)	54.23–1.60
N° reflections	48229
N° molecules in ASU	3
R_{work} / R_{free}	16.8/ 22.1
N° atoms	
Protein	4141
Ligand/ ion	0
Water	504
B-factors	
Protein	24.2
Ligand/ ion	N/A
Water	32.2
R.m.s. deviations	
Bond lengths (Å)	0.030
Bond angles (°)	2.25
No. TLS bodies	3
Ramachandran favored[d]	95.23 %
Ramachandran outliers[d]	0.38 %
Rotamer outliers[d]	0.88 %

[a] Values in parentheses are for highest-resolution shell.
[b] Data is complete to 1.80 Å.
[c] Multiplicity weighted R-merge, (Diederichs & Karplus, 1997; Weiss & Hilgenfeld, 1997)
[d] As determined by *Molprobity* (Davis *et al.*, 2007)

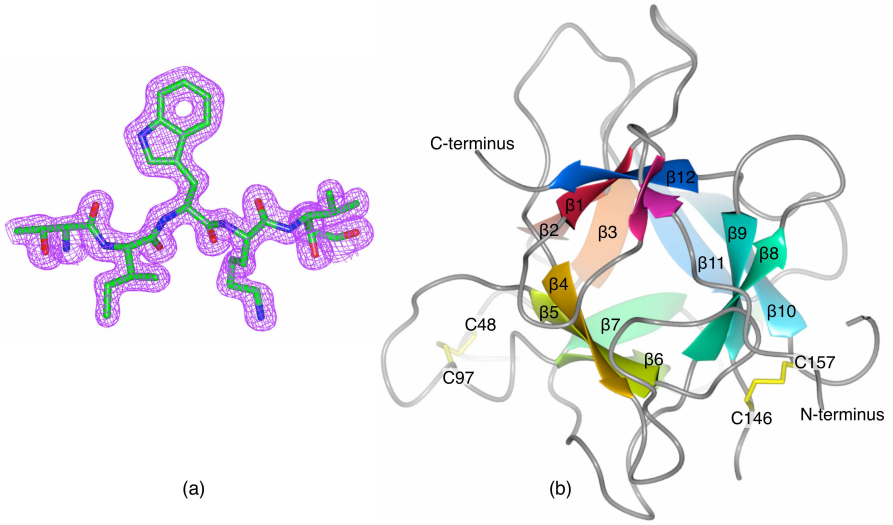

Figure 7.1: Overall structure of PSPI.
(a) Representative part of the electron density map of PSPI with a short stretch of the model (map clipped to show only this part of the model), demonstrating excellent electron density to a high resolution. Map is contoured at 1.5 σ.
(b) Ribbon diagram of one molecule of PSPI with β-sheets indicated, showing a β-trefoil fold. The N- and C-terminus are labelled as well as the two disulfide bridges (in yellow).

mined previously. The structures most similar to PSPI in terms of amino acid sequence are a miraculin-like protein with 32.7% identity (PDB entry 3iir, Gahloth *et al.*, 2010), a Kunitz-type inhibitor from *Delonix Regia* seeds with 31.7% identity (PDB entry 1r8n, Krauchenco *et al.*, 2003), a winged-bean chymotrypsin inhibitor with 27.9% identity (PDB entry 2wbc, Dattagupta *et al.*, 1999) and a soybean trypsin inhibitor with 27.4% identity (PDB entry 1avx, Song & Suh, 1998). A structure-based sequence alignment of PSPI with other similar proteins (Figure 7.2) shows that the sequences are most similar in the β-strand regions, which form the scaffold of the structure, compared with that of protruding loops, which determine the specificity of the inhibitors.

Structural superposition gives a similar pattern. The overall root-mean-square deviations of the Cα-backbone of PSPI with these four structures are (calculated with subunit A of PSPI): 1.47 Å for the miraculin-like protein, 1.22 Å for the Kunitz-type inhibitor from *Delonix regia* seeds, 1.21 Å for a winged-bean chymotrypsin inhibitor, and 1.35 Å for a soybean trypsin inhibitor. This indicates that the structures are indeed similar, but as expected, the similarity is mostly in the β-sheet core, whilst the protruding loops differ substantially.

An important difference between PSPI and the previously determined structures of Kunitz-type serine protease inhibitors is that, to the best of our knowledge, PSPI is the first structure of a post-translationally modified heterodimeric

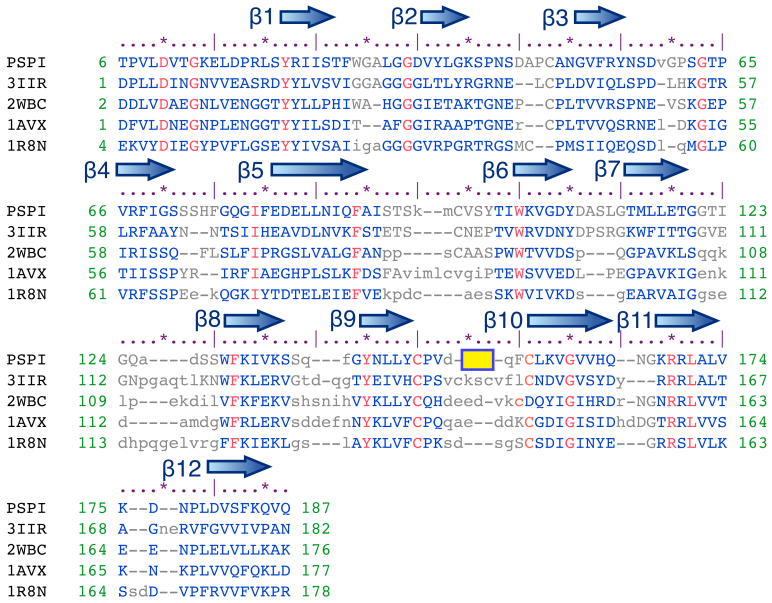

Figure 7.2: Structure-based sequence alignment.
Alignment of PSPI with other previously structurally characterized Kunitz-type serine protease inhibitors. β-strands are indicated with blue arrows and the yellow box shows where six residues are post-translationally deleted from PSPI.

double-headed protease inhibitor. Only crystal structures of *monomeric* Kunitz-type double-headed serine protease inhibitors are known, such as winged bean α-chymotrypsin inhibitor (Dattagupta *et al.*, 1999).

PSPI is expressed as a single polypeptide chain, but in a post-translational process around six residues after Thr150 are removed. Thus, there is no loop between Thr150 and Ser151 and despite the consecutive numbering of the amino acids, there is also no peptide bond between Thr150 and Ser151, because of the deletion. The residues just before the site of the deletion (Thr149 and Thr150) and just after (Ser151, Ser152 and Asp153) are not modelled in our structure due to poor density probably caused by structural flexibility. Despite being (hetero)dimeric, the conformation of the whole molecule including the region close to the deleted loop is very similar to other, monomeric, Kunitz-type serine protease inhibitors. The structure seems to be held together in this very similar conformation by the disulfide bridge Cys146-Cys157 close to the beginning and the end of the deleted loop, since only the residues between these two cysteines show a different backbone direction from monomeric Kunitz-type serine protease inhibitors. Apparently, the presence of the extra loop and the post-translational deletion of the loop has not affected the overall structure of the inhibitor. It is possible that the loop functions as a shield for an important interface of the protein leading to a (partially) inactive form, which gets

activated upon cleavage. The range in which the excised loop can be is indicated in Figure 7.3(a).

7.3.2 Identification of reactive site loops

The reactive site loops for PSPI have not yet been identified. Previous predictions based on sequence alignment with other Kunitz-type inhibitors located the reactive site loops around Arg67-Phe68 (with Arg67 being the P1 position) and around Met115-Leu116 (with Met115 occupying the P1 position; Valueva *et al*., 2000). However, both these proposed sites are located on the inside of the structure and not in protruding loops. Indeed, Arg67 is in the middle of a β-strand in the protein's core and hence cannot be the P1 residue of the reactive site loop. Similarly, there would also be a considerable amount of steric hindrance for a protease to bind if Met115 was a P1 amino acid.

To find more likely reactive site loops, we compared the PSPI structure to other Kunitz-type serine protease inhibitors with previously determined and verified reactive site loops. Five possibilities for reactive side loops arose. Based on comparison with double-headed arrowhead protease inhibitor, the reactive loops could be located around Asn50 and around Leu84 of PSPI (Xie *et al*., 1997). However, in the PSPI structure these residues are also not in protruding loops, but at sterically more inaccessible positions, and thus ruling out these sites.

Another possibility for a reactive site loop is near Asp45-Asn50 in PSPI (see Figure 7.3(a) and (b)). This loop corresponds to the Asn38-Leu43 loop that has been proposed to be one of the reactive site loops for winged-bean α-chymotrypsin inhibitor (Dattagupta *et al*., 1999) based on structural similarity to its other verified reactive site. The Asp45-Asn50 loop in PSPI looks similar in structure and partially similar in sequence to the winged-bean α-chymotrypsin inhibitor reactive loop. The P1 residue in this loop is proline (Pro47 in PSPI) for both proteins. In PSPI this loop is partially disordered in chains B and C and has higher than average B-factors for chain A. This is a typical property of substrate loops that probably aids in binding the protease. The reactive site loops adopt a more rigid conformation upon protease binding (Song & Suh, 1998).

The most common reactive site loop for monomeric Kunitz-type serine protease inhibitors is the loop starting at Ser71 to Phe80 with Phe75 being the P1 residue in PSPI. This sequence was proposed to be the reactive site loop of other serine protease inhibitors including miraculin-like protein (Gahloth *et al*., 2010), for a mutant trypsin inhibitor based on a co-crystal with trypsin (PDB entry 3i29), for double-headed winged bean chymotrypsin inhibitor based on biochemical evidence (Shibata *et al*., 1988) and, finally, for soybean trypsin inhibitor based on a co-crystal with trypsin (Song & Suh, 1998). In PSPI this loop has poor density in chains A and B and has high B-factors in chain C, indicating flexibility. As can be seen in Figure 7.3(a), Phe75 sticks out into the solvent in chain C (the only NCS-related molecule for which there is moderately good electron density; see Figure 7.3(c) for an OMIT map), which would facilitate its recognition by a target protease.

The last possibility for the reactive site loop is around Lys95 in PSPI. The

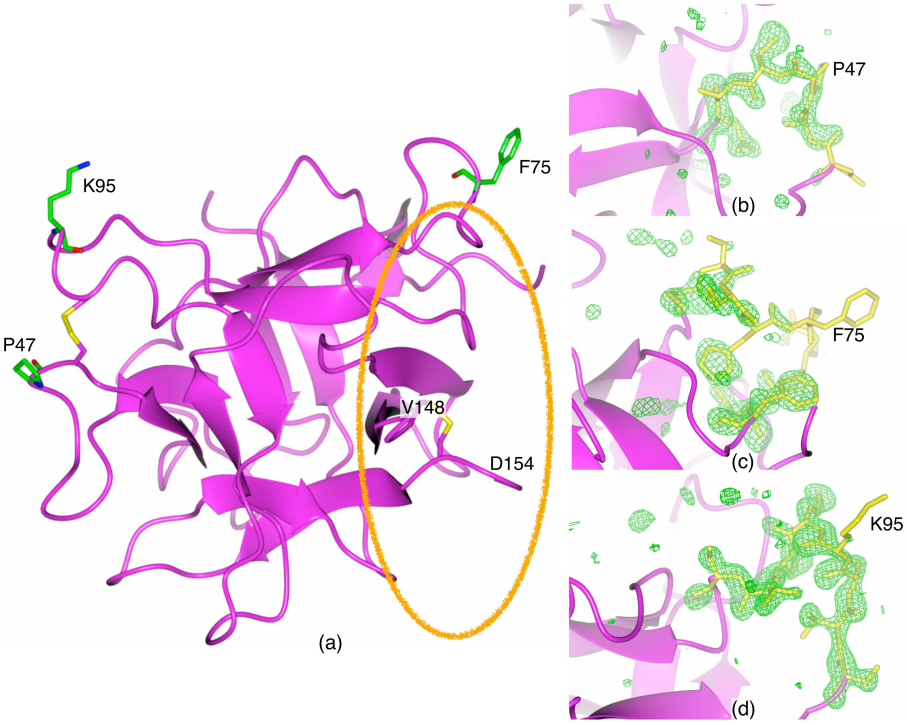

Figure 7.3: Reactive site loops of PSPI.
(a) P1 residues of the three most probable candidates for reactive site loops of PSPI highlighted on the structure of PSPI, showing their position to be in loops and showing how they stick out into the solvent. Disulfide bridges are indicated in yellow. The orange circle shows the range in which the excised loop can be (on the right side of the protein in this view) and the last and first ordered residue in the structure at the beginning and end of the excised loop are indicated (V148 and D154).
(b) Omit map of putative reactive site loop of Asp45-Asn50 (yellow).
(c) Omit map of putative reactive site loop of Ser71-Phe80 (yellow).
(d) Omit map of putative reactive site loop of Ser92-Ser99 (yellow).
The electron difference density maps are calculated based on phases from the model without the residues shown here in yellow and are contoured at 3σ.

corresponding loop was identified to be a reactive site loop in both double-headed arrowhead protease inhibitor (PDB entry 3e8l) based on mutagenesis studies and a co-crystal with two trypsins (Bao *et al*., 2009) and also in barley α-amylase/subtilisin inhibitor based on a co-crystal with subtilisin savinase (Micheelsen *et al*., 2008). The latter inhibitor is identified as fifth structurally most related protein to PSPI by the *VAST* program. The loop around Lys95 has a similar conformation to the loop from the double-headed arrowhead protease inhibitor and protrudes from the structure with Lys95 sticking out to the solvent as can be expected for a P1 reactive site residue (Figure 7.3(a) and (d)).

As stated above, the loops around Pro47, Phe75 and Lys95 are the three most likely candidates for reactive site loops. Since in the trypsin active site a large positively charged amino acid fits best (to make contacts with Asp189), we propose that the Lys95 is the trypsin interacting loop. Chymotrypsin is specific for sequences with bulky hydrophobic amino acids; hence it is plausible that chymotrypsin binds to the Phe75 loop. Moreover, these two reactive site loops have been confirmed by co-crystallization in other Kunitz-type serine protease inhibitors providing stronger evidence for its potential biological function than structural similarity of a loop in an uncomplexed structure, as is the case with Pro47. The Phe75 reactive site loop might be hidden before the post-translational processing of the loop between Thr150 and Ser151 (Figure 7.3(a)).

It has been reported that one molecule of PSPI can bind a trypsin and a chymotrypsin molecule at the same time (Valueva *et al*., 2000). To verify whether this is indeed possible with the suggested reactive site loops, we modelled two proteases on the reactive site loops based on co-crystal structures in which a trypsin is bound to one of these loops (Song & Suh, 1998 and Bao *et al*., 2009) using the secondary-structure matching (SSM) function in *COOT*. As can be seen in Figure 7.4, it is possible for the two proteases to simultaneously bind PSPI.

It has also been reported that PSPI inhibits human leukocyte elastase (Valueva *et al*., 2000). Since elastase is specific for a small neutral amino acid, we are unsure whether it can act on either of the loops proposed for trypsin or chymotrypsin. Thus, it might need another binding site such as the Pro47 loop. Unfortunately, no elastase structure with bound Kunitz-type serine protease inhibitor is available for comparison.

7.3.3 Interfaces for aggregation

The *PISA* (Protein Interfaces, Surfaces and Assemblies) server (Krissinel & Henrick, 2007) was used to investigate potential interfaces for interaction in the PSPI crystal structure. The surface including residues 107, 127 and 148 with the plane formed by residues 30, 58 and 169 was predicted to be a strong interface and have a high probability to occur in solution. Since PSPI was reported to exist as a single molecule in solution (Pouvreau *et al*., 2005), this interface is apparently not seen under ambient conditions. However, it has been reported that PSPI aggregates upon heating into a product with four molecules of PSPI (Pouvreau *et al*., 2005). The interface that is detected by *PISA* could be the starting interface of such aggrega-

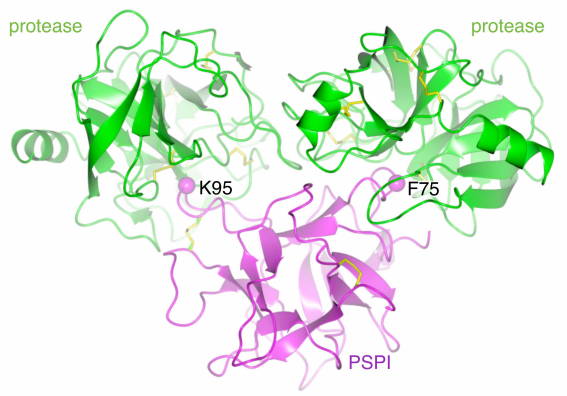

Figure 7.4: Model of PSPI with two proteases.
Model of PSPI (magenta) with two proteases (green) on it, showing that it is possible for two proteases to bind simultaneously at the reactive sites F75 and K95 to one molecule of PSPI. In this picture two trypsins were modelled, but the picture will be equivalent for one chymotrypsin and one trypsin, due to the structural similarity of trypsin and chymotrypsin.

tion when the conditions for solubility are changed, as happens in a crystallization experiment with increasingly high concentrations, after which two seeding two-molecule-PSPI-complexes join to form an aggregate containing four molecules.