Structures of a bi-functional Kunitz-type STI family inhibitor of serine and aspartic proteases: could the aspartic protease inhibition have evolved from a canonical serine protease-binding loop?

Yasel Guerra[1*], Pedro A. Valiente[2], Tirso Pons[3], Colin Berry[4] and Enrique Rudiño-Piñera[1*]

[1]Departamento de Medicina Molecular y Bioprocesos, Instituto de Biotecnología, Universidad Nacional Autónoma de México (UNAM), Avenida Universidad 2001, Colonia Chamilpa, Cuernavaca, Morelos, CP 62210, México. [2]Laboratorio de Biología Computacional y Diseño de Proteínas, Centro de Estudios de Proteínas (CEP), Facultad de Biología, Universidad de La Habana, Cuba. [3]Structural Biology and Biocomputing Programme, Spanish National Cancer Research Centre (CNIO), C/Melchor Fernández Almagro 3, Madrid 28029, Spain. [4]Cardiff School of Biosciences, Cardiff University, Cardiff CF10 3AT, Wales, UK. *Corresponding author: yaselg@ibt.unam.mx and rudino@ibt.unam.mx

Telephone:+52 7773291610 Fax: +52 777 317 2388

Abstract:

Bifunctional inhibitors from the Kunitz-type soybean trypsin inhibitor (STI) family are glycosylated proteins able to inhibit serine and aspartic proteases. Here we report six crystal structures of the wild-type and a non-glycosylated mutant of the bifunctional inhibitor E3Ad obtained at different pH values and space groups. The crystal structures show that E3Ad adopts the typical β-trefoil fold of the STI family exhibiting some conformational changes due to pH variations and crystal packing. Despite the high sequence identity with a recently reported potato cathepsin D inhibitor (PDI), three-dimensional structures obtained in this work show a significant conformational change in the protease-binding loop proposed for aspartic protease inhibition. The E3Ad binding loop for serine protease inhibition is also proposed, based on structural similarity with a novel non-canonical conformation described for the double-headed inhibitor API-A from the Kunitz-type STI family. In addition, structural and sequence analyses suggest that bifunctional inhibitors of serine and aspartic proteases from the Kunitz-type STI family are more similar to double-headed inhibitor API-A than other inhibitors with a canonical protease-binding loop.

Keywords: aspartic protease inhibitors, Kunitz-type STI family inhibitors, bi-functional inhibitors, plant protease inhibitors, β-trefoil fold

Abbreviations: BASI, barley alpha-amylase/subtilisin inhibitor; EcTI, *Enterolobium contortisiliquum* trypsin inhibitor; PSPI, potato serine protease inhibitor; API-A, Arrowhead proteinase inhibitor A; PDI, potato cathepsin D inhibitor; STI, soybean trypsin inhibitor; SLAPI, *Solanum lycopersicum* aspartic protease inhibitor; TKI, Tamarind Kunitz-type inhibitor.

Introduction

Protease inhibitors are an effective way to regulate the activity of target proteases, with great interest in those involved in biomedical and biotechnological processes. Among the different protease inhibitor families, proteinaceous serine protease inhibitors are probably the most studied and widespread in nature (Otlewski et al., 2001; Tyndall et al., 2005).

Inhibitors of the Kunitz-type soybean trypsin inhibitor (STI) family are among the most versatile protease inhibitors reported, being able to interact with proteases belonging to different mechanistic classes like serine, aspartic and cysteine proteases, as well as other enzymes: α-amylase and sucrose invertase (Azarkan et al., 2011; Cater et al., 2002; Fischer et al., 2015; Franco et al., 2002; Glaczinski et al., 2002; Keilová H. and Tomášek, 1976a). These inhibitors have a β-trefoil fold, composed of 12 β-strands arranged in three-fold pseudo-symmetry units or subdomains with the presence of helices in some cases (Broom et al., 2012; McLachlan, 1979; Murzin

et al., 1992; Renko et al., 2012). The pseudo-symmetry units are composed of four β-strands, two of them forming a six-stranded β-barrel with the other subdomains, and the other two of each unit forming a β-hairpin acting like the cap for the β–barrel. The structural regions reported to be involved in protease-inhibitor interactions are loops connecting the β-strand elements (i.e. loop L4 between β-strands 4 and 5 in STI), which can be highly variable in sequence and length (Renko et al., 2012). The most accepted model for the β-trefoil fold evolution proposes that the fold arose from a duplication and fusion of the monomeric unit from an ancestral trimeric protein (Broom et al., 2012; Ponting and Russell, 2000).

There are several structural and functional studies defining the inhibition mechanism for serine proteases (Laskowski and Kato, 1980; Radisky and Koshland, 2002; Song and Suh, 1998; Zakharova et al., 2009). Among the proteinaceous protease inhibitors, almost all Kunitz-type STI inhibitors have a “canonical” conformation in the protease-binding loop that interacts with the active site of the serine proteases forming a substrate-like “Michaelis” complex and thereby implicated directly in their biological activity (Song and Suh, 1998; Zhou et al., 2013). In the inhibition mechanism, the enzyme is able to hydrolyze the peptidic bond between P1-P1´ positions (Schechter & Berger Pn-Pn´nomenclature) (Schechter and Berger, 1967) but the product retains a similar conformation with respect to the intact inhibitor, trapping the enzyme in a hydrolysis/re-synthesis cycle (Zakharova et al., 2009). The amino acid sequence of these “canonical” protease-binding loops can be variable, although there are certain preferences for basic amino acids (Lys and Arg) in the P1 position for trypsin (E.C. 3.4.21.4) and the amino acids Tyr, Trp, Phe, Leu and Met in the case of chymotrypsin (E.C. 3.4.21.1) (Tyndall et al., 2005). Despite the variability among amino acid sequences, the conformation of the protease-binding loops are kept unchanged (Laskowski and Qasim, 2000). Generally, positions P3-P3´ are key determinants of the reactive loop conformation, with the P1 position comprising up to 50% of the enzyme-inhibitor interface contact area (Krowarsch et al., 2003). However, there is at least one case reported where a double-headed inhibitor has a binding loop with a non-canonical conformation, forming an enzyme-inhibitor complex with trypsin (Bao et al., 2009).

In the case of Kunitz-type STI inhibitors from legume seeds, a sub-classification has been proposed based on the numbers of cysteines in the sequence (Oliva et al., 2010). Specifically in potato, a different classification has been proposed to divide the proteinaceous inhibitors into five groups with the best characterized proteins belonging to groups A, B and C (Bauw et al., 2006; Heibges et al., 2003a). It has been shown that there is a great sequence variability in the Kunitz-type STI inhibitors in potato, which has been linked to an adaptive mechanism in response to pathogens or even the acquisition of new functions to enhance their adaptability (Heibges et al., 2003b). Recently, a cDNA library was obtained from several potato cultivars and some proteins were heterologously expressed and tested against different classes of proteases. This work showed that inhibition of aspartic proteases is also possible in potato protease inhibitors belonging to group B (Fischer et al., 2015).

Despite the wealth of data regarding the inhibition process for serine proteases, very little is known concerning the structural determinants for the inhibition of aspartic proteases by inhibitors from the Kunitz-type STI family. Although first studies on bi-functional inhibitors of the Kunitz-type STI family against aspartic proteases were performed using human cathepsin D (EC 3.4.23.5) (Keilová H. and Tomášek, 1976b), to date there is evidence that Kunitz-type STI family members are also able to inhibit other aspartic proteases of the clan AA, family A1, such as saccharopepsin (EC 3.4.21.41) (Cater et al., 2002), memapsin 1 (EC 3.4.23.45) (Fischer et al., 2015), and Plasmepsin II (EC 3.4.23.39) (this work).

Until recently, structural studies of proteinaceous inhibitors of aspartic proteases were limited with only four three-dimensional coordinates available in the PDB from four different families in the MEROPS database (Rawlings et al., 2014): inhibitor SQAPI, family I25 (Headey et al., 2010, PDB code: 2KXG); inhibitor PI3, family I33 (Ng et al., 2000, PDB codes: 1F32 and 1F34), inhibitor IA3, family I34 (Li et al., 2000, PDB code 1DPJ) and potato cathepsin D inhibitor (PDI), family I3 (Guo et al., 2015, PDB code: 5DZU). Only for the inhibitors PI3 and IA3 are

there three-dimensional coordinates of the inhibitors in complex with target aspartic proteases (Pepsin and saccharopepsin; PDB codes: 1F34 and 1DPJ, respectively). However, there are no similarities in the amino acid sequences, three-dimensional structures or inhibition binding modes between these distinct families, making extrapolations to other proteinaceous protease inhibitors impossible.

Previous computational analyses based on comparative protein modeling and docking studies, predicted loop L9 in the *Solanum lycopersicum* (tomato) aspartic protease inhibitor (SLAPI), as a key region involved in the inhibitor interaction with saccharopepsin´s active site (Guerra et al., 2012). A recent paper that describes the crystallographic structure of PDI also suggests the same structural region –the loop from Cys142 to Ala160 in PDI- as a binding loop for cathepsin D (Guo et al., 2015). Based on docking studies, Guo and colleagues, propose a second possible binding loop (loop L5) for cathepsin D running from Phe85 to Ile98.

In the present work we report six crystal structures, obtained at different pH values (3.0, 3.5, 7.0, 7.2, 7.4 and 9.0) and space groups ($C222_1$, $P22_12_1$ and $P4_322$), of a potato bi-functional inhibitor of the Kunitz-type STI family that is able to inhibit the serine protease bovine trypsin and the aspartic protease Plasmepsin II. Despite, the close sequence identity with the potato cathepsin D inhibitor recently reported by Guo and colleagues, our crystal structures show significant differences in loop conformations, especially in loop L9. The analysis of the structural similarities with other inhibitors of the same family, as well as docking-modeling studies suggests loop L5 and L9 as potential binding loops for serine and aspartic proteases, respectively.

Methods

DNA isolation and cloning

Genomic DNA from *Solanum tuberosum* (potato) Estima variety was isolated using the plant CTAB method (ABgene). Primers designed based on the API-8 inhibitor sequence (Strukelj et al., 1992) were used to amplify the inhibitor sequences and clone them in the pGEMT vector (Promega) and positive clones were sequenced. Among the sequences obtained, clone E3Ad, which does not contain introns and matched with GenBank entries XP_006362668.1 and XM_006362606.2, was selected for protein expression and re-amplified to introduce restriction sites for *XhoI* and *Sac II,* for subcloning in the vector pPICZαC, in frame with the α-factor secretion signal (Table S1). The mutant E3Ad_N19D was generated to delete the only glycosylation site on the E3Ad inhibitor using the construct pPICZαC-E3Ad as template (Table S1) with the QuickChange Multi-Site-Directed Mutagenesis Kit (Agilent). The clones were sequenced to verify the N19D mutation. The plasmid construct was linearized with the restriction enzyme *Sac I* and transformed into the yeast *Pichia pastoris* (GS115 strain) using the Pichia EasyComp Kit (Life Sciences). Positive clones were checked by colony PCR with 5´AOX1 and 3´AOX1 primers.

Expression and purification of inhibitors

A single colony expressing the inhibitor E3Ad or E3Ad-N19D was inoculated and pre-cultured in 120 mL of BMGY medium (2% tryptone, 1% yeast extract, 0.1 M potassium phosphate, pH 6.0, 1.34% yeast nitrogen base, 40 µM biotin, 1% glycerol) in baffled flasks at 30°C and 200 rpm for 24h. Cells were harvested by centrifugation at 3000 x g for 10 min at RT and washed with BMMY (2% tryptone, 1% yeast extract, 0.1 M potassium phosphate, pH 6.0, 1.34% yeast nitrogen base, 40 µM biotin, 1% methanol) medium and harvested as previously described. Expression of the inhibitors was achieved by re-suspension in BMMY medium to a final OD_{600nm} of 1.0. Recombinant *Pichia pastoris* cells were incubated with continuous shaking at 28 °C and 200 rpm for 72h. Methanol (100%) was added every 24 h to a final concentration of 1% to maintain induction. Cells were harvested by centrifugation at 3000 x g for 10 min at RT and discarded. The supernatant was used for protein purification.

Culture supernatant was extensively dialyzed against 10 mM sodium formate pH 4.0 buffer at 4 °C. The dialysate was adjusted to 150 mM NaCl and applied to a HiPrep™ SP-Sepharose HP column (GE Healthcare Life Sciences). Inhibitors were eluted using a linear gradient of 150-490 mM of NaCl in 10 mM sodium formate pH

4.0 buffer. Fractions were analyzed by SDS-PAGE, pooled and concentrated using 10 kDa cut-off centrifugal filter units (Millipore). Concentrated samples of the inhibitors were applied to a Superdex 75 10/30 HR column (GE Healthcare Life Sciences) equilibrated with 10 mM sodium formate, 150 mM NaCl pH 4.0, coupled to an Äktapure system (GE Healthcare Life Sciences). Fractions were analyzed by SDS-PAGE and those containing the inhibitor were pooled.

Crystallization

Purified inhibitors were concentrated up to 10 mg/mL using centrifugal filter units with a 10 kDa cut-off (Millipore).

Crystallization conditions were screened by the microbatch method using Wizard Classics (I-IV) (Emerald Biosciences) kits at 4 ºC and 18 ºC. The sitting-drop vapour-diffusion method was used for crystal optimization, with drops consisting of 1 µL of protein and 1 µL of reservoir solution against 0.5 mL of reservoir solution. Promising conditions were further optimized with the use of Hampton Research Additive Screen™. Final pH values were measured using a pH meter.

Diffraction quality crystals for mutant E3Ad_N19D were obtained from 15% (w/v) PEG 20000, 0.1 M HEPES pH 7.0, 30 mM glycyl-glycyl-glycine (Condition 1, final pH 3.0), 10% (v/v) MPD, 0.1 M Bicine pH 9.0 (Condition 2, final pH 9.0), 20% (w/v) PEG 8000, 0.1 M HEPES pH 7.5100 mM glycine (Condition 3, final pH 3.5) and 20% (w/v) PEG 8000, 0.1 M HEPES pH 7.5, 30 mM glycyl-glycyl-glycine (Condition 4, final pH 7.2). Crystals for E3Ad were grown in 10% (w/v) PEG 2000 MME, 0.1 M Tris-HCl pH 8.5, 0.2 M trimethylamine N-oxide30 mM glycyl-glycyl-glycine (Condition 5, final pH 7.4) and 20% (w/v) PEG 3350, 0.1 M Tris-HCl pH 7.0, 0.2 M $CaCl_2$ (Condition 6, final pH 7.0). Single crystals were soaked in reservoir solution containing a cryoprotectant agent (30% (v/v) glycerol or 35% (v/v) PEG 200) and mounted in loops, before flash-cooling in liquid nitrogen at 100K.

Data Collection and processing

Data were collected at beamline 19ID of the Advanced Photon Source (APS) on an ADSC Q315r CCD detector and at beamline BL14-1 of the Stanford Synchrotron Radiation Lightsource (SSRL) on a MARmosaic 325 CCD detector. Indexing, integration and scaling was performed with XDS and XSCALE (Kabsch, 2010). The reflections of the crystals obtained in conditions 5 and 6 were anisotropic scaled using the web server Diffraction Anisotropy Server (http://services.mbi.ucla.edu/anisoscale/) (Strong et al., 2006). The Matthews coefficient was calculated using MATTPROB program (Kantardjieff and Rupp, 2003; Weichenberger and Rupp, 2014). Molecular replacement was performed with PHASER (McCoy et al., 2007) from the CCP4 suite (Winn et al., 2011) using the crystallographic structure of the inhibitor PSPI (PDB code 3TC2, %ID 70.8) as the initial search model for data obtained in the APS. For diffraction data obtained in the SSRL, the structure obtained for inhibitor E3Ad_N19D at 2.45 Å (PDB code 5FNW) was used as template for molecular replacement. In all cases, a cycle of rigid-body refinement was performed with REFMAC5 (Murshudov et al., 2011). Model refinement was performed with alternate cycles of Phenix (Adams et al., 2010) and manual rebuilding with Coot (Emsley et al., 2010). Model validation was performed with MolProbity (Chen et al., 2010).

E**nzymatic assays**

The aspartic protease Plasmepsin II was expressed and purified as described previously (Ramírez et al., 2009). A continuous enzymatic assay was performed using the chromogenic peptide Lys-Pro-Phe-Glu-Phe-Nph-Arg-Leu (Nph: para-nitrophenylalanine) as substrate (84 µM final concentration, K_M: 50.8 µM) and 100 mM NaAc pH 4.4 as assay buffer. The reactions were followed at 310 nm in a Cary 60 spectrophotometer (Agilent, USA) at 37°C. The enzyme concentration in the assay was 180nM, which was determined by active-site titration with Pepstatin A.

Serine protease inhibition assay was performed with bovine trypsin (EC 3.4.21.4) from Sigma. A continuous enzymatic assay was performed using Nα-benzoyl-L-arginine p-nitroanilide (BAPNA) as substrate (1 mM final

concentration, K_M: 1 mM) and 20 mM Tris-HCl, 150 mM NaCl, 20 mM $CaCl_2$ pH 8.0 as assay buffer. The reactions were followed at 405 nm in a Cary 60 spectrophotometer (Agilent, USA) at 25 °C. The enzyme concentration in the assay was 66 nM.

For both assays, the enzyme and inhibitor were incubated for 5 min prior the addition of the substrate. The initial velocity was determined at different concentrations of inhibitor E3Ad_N19D using the Cary WinUV Kinetics v5.0 software. Fractional residual activity (v_i/v_0) was calculated from initial reaction velocities for different inhibitor concentrations with respect to an assay in the absence of inhibitor. All assays were performed in triplicate. K_i^{app} values were obtained by non-linear regression fitting of the data to the Morrison equation (Equation 1) (Bieth, 1995; Morrison, 1969) with the program Statistica v7.0 (Statsoft). K_i values were calculated using Equation 2, considering a competitive binding mode, as previously reported for other related bifunctional inhibitors (Keilová H. and Tomášek, 1976b).

$$\frac{v_i}{v_0} = 1 - \frac{([E]_0+[I]_0+K_i^{app})-\sqrt{([E]_0+[I]_0+K_i^{app})^2-4[E]_0[I]_0}}{2[E]_0} \qquad (1)$$

To calculate the stoichiometry of the Trypsin-inhibitor interaction, the enzyme concentration in the assay was increased to be under titration conditions ($[E]_0/K_i^{app}$=15). Under these conditions, the curve v_i/v_0 vs $[I]_0/[E]_0$ shows a biphasic behaviour with the x axis intercept being the point where all the enzyme is bound to the inhibitor present in the assay.

$$K_i^{app}=K_i(1+\frac{[S]_0}{K_M}) \qquad (2)$$

Sequence and structural comparison

Multiple sequence and structural alignments were performed using T-Coffee (http://tcoffee.crg.cat/apps/tcoffee/index.html) (Di Tommaso et al., 2011) and PDBeFold (http://www.ebi.ac.uk/msd-srv/ssm/) (Krissinel and Henrick, 2004) web servers, respectively. The final multiple sequence alignment (MSA) was manually parsed using the results obtained from both alignments. A phylogenetic consensus tree was calculated from the curated MSA obtained previously using the Maximum Likelihood method based on the JTT matrix-model and 1000 bootstrap replicates, using the program MEGA6 (Jones et al., 1992; Tamura et al., 2013). Pairwise whole protein and specific loop structural superpositions were performed with the Secondary Structure Matching (SSM) and Least-squares (LSQ) superposition modules respectively in the program Coot (Emsley et al., 2010),.

The root mean square fluctuation (RMSF) among the heavy atoms of all the X-ray structures determined for E3Ad and E3Ad_N19D inhibitors were calculated using the g_rmsf tool of the GROMACS software package (version 4.6.5) (Páll et al., 2015). For this analysis the structure of the inhibitor E3Ad_N19D at pH 3.0 (PDB code: 5FNW) was used as the reference.

The metaPPISP web server (http://pipe.scs.fsu.edu/meta-ppisp.html) (Qin and Zhou, 2007) was used to predict potential protein-protein interaction sites based on the three-dimensional structures obtained for the inhibitors E3Ad_N19D and PDI. All the residues with a final score above 0.34 were considered as a positive prediction, as suggested by the authors. Rigid body docking was performed with the ClusPro 2.0 web server (https://cluspro.bu.edu/login.php) (Comeau et al., 2004a) using a crystal structure of Plm II in its free form (PDB code: 1LF4) and the structure of the inhibitor E3Ad_N19D at pH 3.0 (PDB code: 5FNW).

Molecular Dynamics

Energy Minimizations (EM) and Molecular Dynamics (MD) simulations were carried out using the GROMACS software package (version 4.6.5) (Páll et al., 2015) with the AMBER99sb force field (Hornak et al., 2006), and the TIP3P water model (Price and Brooks, 2004). The simulation systems consist of each inhibitor (PDI or E3Ad_N19D) solvated in a dodecahedron box with 7414 water molecules. The protonation states of ionizable residues were assigned at pH 4.4 or pH 8.0 with the program PDB2PQR (http://nbcr-222.ucsd.edu/pdb2pqr_1.8/), which uses PROPKA for the prediction of pK_a values (Dolinsky et al., 2007). To achieve the electroneutrality in the system, Na^+ and Cl^- ions were added at an appropriate ratio to reach a NaCl concentration of 0.15 M. At each step, the electrostatic interactions were calculated with the particle-mesh Ewald method (Darden et al., 1993). Van der Waals interactions were described by a Lennard-Jones potential with a cutoff of 1.1 nm that was switched to zero between 1.0 and 1.1 nm. Dispersion corrections for energy and pressure were applied. The SETTLE algorithm (Miyamoto and Kollman, 1992) was used to constrain bonds and angles of water molecules and LINCS (Hess et al., 1997) was used for all other bonds, allowing a time step of 2 fs. The equilibrium simulations were sampled for 100 ns, using the leapfrog dynamic integrator (Hockney, 1970) at 310 K and a constant pressure of 1 atm, using the weak coupling to the velocity rescaling thermostat (Bussi et al., 2007) and Parrinello-Rahman barostat (Nosé and Klein, 1983; Parrinello and Rahman, 1981), respectively. Snapshots were saved at 10 ps intervals.

Results and Discussion

Overall structure of the bi-functional inhibitors E3Ad and mutant E3Ad_N19D

Inhibitors E3Ad and E3Ad_N19D crystallized in several conditions, with most of them belonging to tetragonal space group $P4_322$, with one molecule in the asymmetric unit, and two crystals belonging to orthorhombic space groups $C222_1$ and $P22_12_1$ with two and one molecules in the asymmetric unit, respectively. Even though space group $P22_12_1$ is non-standard, systematic absences and the final refinement of the model supported the space group assignation. Structures were obtained at different pH values and resolutions from 2.43 Å to 2.65 Å (Table 1). Coordinates for more than 98% of the residues were clearly defined by the $2F_o$-F_c electron density map, with the exception of residues Glu1 and Ser2 in all structures.

High solvent content in most of the crystals may be responsible for the high B-values calculated for most of the structures (Table1). For the structures reported in this work, a sequential addition of the cryoprotectant to the crystallization drop was crucial in order to get well defined spots with a mosaicity below 1 degree. When the crystals were directly soaked in the cryoprotectant solution both resolution and mosaicity were always negatively affected, which seems to be a result of the high solvent content in most of our crystals.

Table 1. X-ray statistics for the E3Ad and E3Ad_N19D structures. Values in parenthesis are for the highest resolution shell.

Protein	E3Ad_N19D	E3Ad_N19D	E3Ad_N19D	E3Ad_N19D	E3Ad	E3Ad
Condition (pH)	1 (pH 3.0)	2 (pH 9.0)	3 (pH 3.5)	4 (pH 7.2)	5 (pH 7.4)	6 (pH 7.0)
Beamline	19ID (APS)	19ID (APS)	BL14-1 (SSRL)	BL14-1 (SSRL)	BL14-1 (SSRL)	BL14-1 (SSRL)
Wavelength (Å)	0.9792	0.9792	1.1807	1.1807	1.1807	1.1807
Space group	$P4_322$	$P4_322$	$C222_1$	$P4_322$	$P4_322$	$P22_12_1$
Unit-cell parameters						
a (Å)	77.22	77.94	102.95	77.14	77.18	47.58
b (Å)	77.22	77.94	114.24	77.14	77.18	56.45
c (Å)	94.24	95.01	98.77	94.14	95.90	72.55
α, β, γ (°)	90.0, 90.0, 90.0	90.0, 90.0, 90.0	90.0, 90.0, 90.0	90.0, 90.0, 90.0	90.0, 90.0, 90.0	90.0, 90.0, 90.0
Mosaic spread (°)[a]	0.14	0.25	0.21	0.13	0.38	0.55

Resolution (Å)	19.50-2.45 (2.50-2.45)	19.45-2.65 (2.75-2.65)	38.42-2.47 (2.57-2.47)	32.0-2.43 (2.53-2.43)	29.5-2.5 (2.57-2.5)	22.6-2.55 (2.62-2.55)
R_{merge} (%)[b]	5.3 (63.5)	4.9 (57.1)	4.7 (63.9)	4.6 (62.8)	5.5 (72.8)	6.5 (24.1)
R_{meas} (%)[c]	5.4 (65.2)	5.1 (58.6)	5.0 (67.9)	4.7 (64.8)	5.7 (75.1)	7.0 (25.8)
$CC_{1/2}$ (%)[d]	99.9 (96.2)	100 (95.4)	100 (95.1)	100 (94.3)	100 (94.2)	99.9 (99.2)
Completeness (%)	99.8 (100)	99.8 (100)	99.8 (100)	99.9 (100)	98.0 (74.1)	98.0 (75.8)
I/σ(I)	38.01 (5.58)	40.61 (5.98)	28.06 (3.24)	42.13 (5.21)	33.82 (4.89)	21.69 (8.08)
Multiplicity	18.5 (19.28)	18.3 (19.24)	8.59 (8.7)	16.4 (16.95)	16.2 (16.63)	7.73 (8.12)
No. of observed reflections	20 3134 (12 053)	163 456 (17 575)	182 351 (20 413)	184 496 (21 020)	167 335 (9378)	51 340 (2876)
No. of unique reflections	10 963 (625)	8 941 (913)	21 238 (2345)	11 223 (1240)	10 356 (564)	6 643 (354)
Wilson plot B factor ($Å^2$)	58.2	72.4	60.57	56.77	60.16	33.56
Solvent content (%)	64.1	65.0	65.24	63.99	59.62	40.67
R-factor (%)	20.9	20.8	22.53	19.85	22.80	18.1
Free R-factor (%)	26.2	26.0	27.42	23.75	24.37	24.6
RMSD bond lengths (Å)	0.010	0.012	0.010	0.009	0.010	0.009
RMSD bond angles (°)	1.21	1.50	1.39	1.22	1.35	1.16
No. of reflections in working set	10 458	8545	20 203	10 664	9857	6332
No. of reflections in test set	503	391	1035	559	499	311
Mean protein B-factor ($Å^2$)	58.0	70.0	65.10	58.10	67.40	34.10
Ramachandran plot[e]						
Most favored (%)	94.02	90.76	92.35	95.63	92.90	96.20
Additional allowed (%)	4.89	7.61	7.38	3.83	6.56	3.80
Outliers (%)	1.09	1.63	0.27	0.55	0.55	0.0
PDB code	5FNW	5FNX	5FZY	5FZU	5G00	5FZZ

[a]Calculated with XDS

[b]$R_{merge} = \Sigma_{hkl} \Sigma_i |I_i(hkl) - (I(hkl)| \Sigma_{hkl} \Sigma_i I_i (hkl)$, where $I_i(hkl)$and $(I(hkl))$ represent the diffraction-intensity values of the individual measurements and the corresponding mean values. The summation is over all unique measurements.

[c]Rmeas is a redundancy-independent version of Rmerge: Rmeas = Σ_{hkl} {N(hkl)/[N(hkl)-1]}$^{1/2}$ $\Sigma_l |I_i(hkl) - \{I(hkl)\}|/ \Sigma_{hkl}\Sigma_i I_i (hkl)$

[d]$CC_{1/2}$ values are the half-set correlation coefficients (Karplus and Diederichs, 2012).

[e]Categories defined by Molprobity

All structures determined show the β-trefoil fold typical of Kunitz-type STI family inhibitors, with 12 antiparallel β-strands arranged in three pseudo-symmetry elements composed of four β-strands each (Figure 1). The β-trefoil fold is composed of six β-strands forming a β-barrel (strands β1, β4, β5, β8, β9 and β12) and the other six β-strands arranged as a lid for the barrel. Two β-strands from the barrel and two from the lid form each pseudo-symmetry unit or subdomain. Three short 3_{10} helices have been assigned in the E3Ad_N19D structures at pH 3.0 (PDB code: 5FNW) and pH 3.5 (PDB code: 5FZY), as well as E3Ad structures at pH 7.0 (PDB code: 5FZZ) and pH 7.4 (PDB code: 5G00) comprising residues Arg29-Gly33 (3_{10}-A), Lys91-Tyr96 (3_{10}-B) and Cys152-Gln156 (3_{10}-C). However, in the E3Ad_N19D structures at pH 7.2 (PDB code: 5FZU) and pH 9.0 (PDB code: 5FNX) just two 3_{10} helices have been assigned, running from residues Arg29-Gly33 and Val90-Val94 (PDB code: 5FZU) or Val90-Val94 and Cys152-Gln156 (PDB code: 5FNX). The $2F_o$-F_c electron density maps of the E3Ad and E3Ad_N19D inhibitor structures clearly showed the presence of three disulfide bridges between positions Cys48-Cys93, Cys142-Cys158 and Cys148-Cys152. These disulfide bridges are present in all Kunitz-type STI family inhibitors able to inhibit aspartic proteases characterized to date, except those classified as part of the group B (Cater et al., 2002; Fischer et al., 2015; Keilová H. and Tomášek, 1976a; Mares et al., 1989).

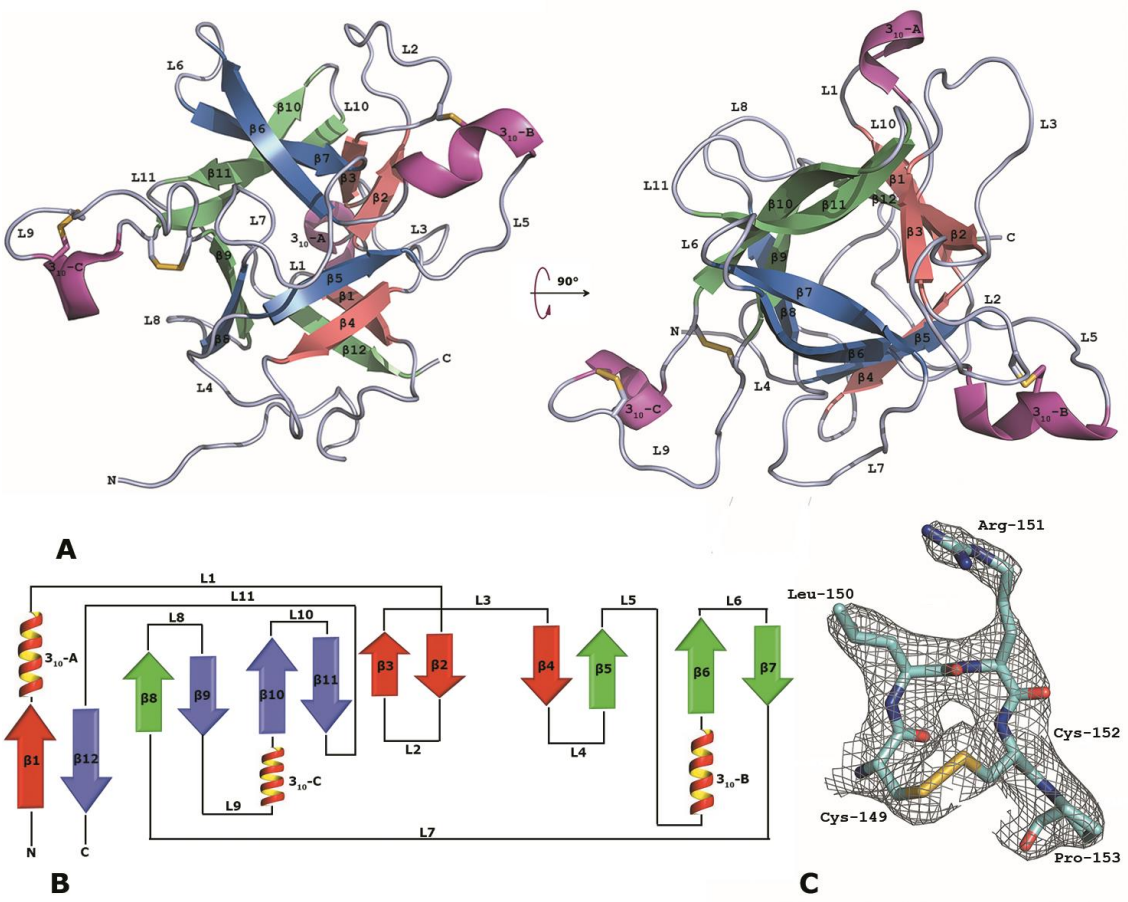

Figure 1. Overall three-dimensional structure of inhibitor E3Ad_N19D at pH 3.0. β-sheets are colored according the pseudo three-fold symmetry elements (A). Topology of the inhibitor E3Ad_N19D (B) Final ($2F_o$-F_c) electron density map around residues Cys149-Pro153 (Loop L9) contoured at 1.0σ (C). The figure was created with PyMOL Molecular Graphics System, Schrödinger, LLC (http://www.pymol.org).

In the coordinates of the inhibitor 3EAd (PDB code: 5G00), an N-Acetyl-glucosamine (NAG) molecule was modeled linked to the Asn19. Mass spectrometry analysis of the purified inhibitor E3Ad showed that several glycoforms are present with two molecules of NAG linked to the Asn19 and a variable number of mannose units (results not shown). However, electron density was just visible for one molecule of NAG.

The three-dimensional structures of the inhibitor E3Ad and the mutant E3Ad_N19D at different pH values and space groups are very similar, with RMSD values from 0.21 Å to 0.68 Å over 185 Cα atoms (Figure 2A). However, a more detailed analysis of RMSD differences among all the structures suggests that crystal packing is more important than pH for such differences (Supplementary Table S2). The main differences in the structures are located in loops L1 (28-32), L2 (42-46), and L5 (89-92) (Figure 2A and B). The presence of three 3_{10} helices present in four crystal structures reported in this work (PDB codes: 5FNW, 5FZY, 5FZZ, 5G00) are located in a pseudo-symmetry equivalent position connected to one β-strand from the β-barrel and one β-strand from the lid (Figures 1 and 3). To our knowledge, this is the first report of a protein of the Kunitz-type STI family with three helices, with most of the other family members having one α-helix, like STI (PDB code: 1AVW), API-A (PDB code: 3E8L), DrTI (PDB code: 1R8N) or one (PDI (PDB code: 5DZU)) or two (PSPI (PDB code: 3TC2)) 3_{10} helices. In this last structure, the two 3_{10} helices are in equivalent positions to our structures.

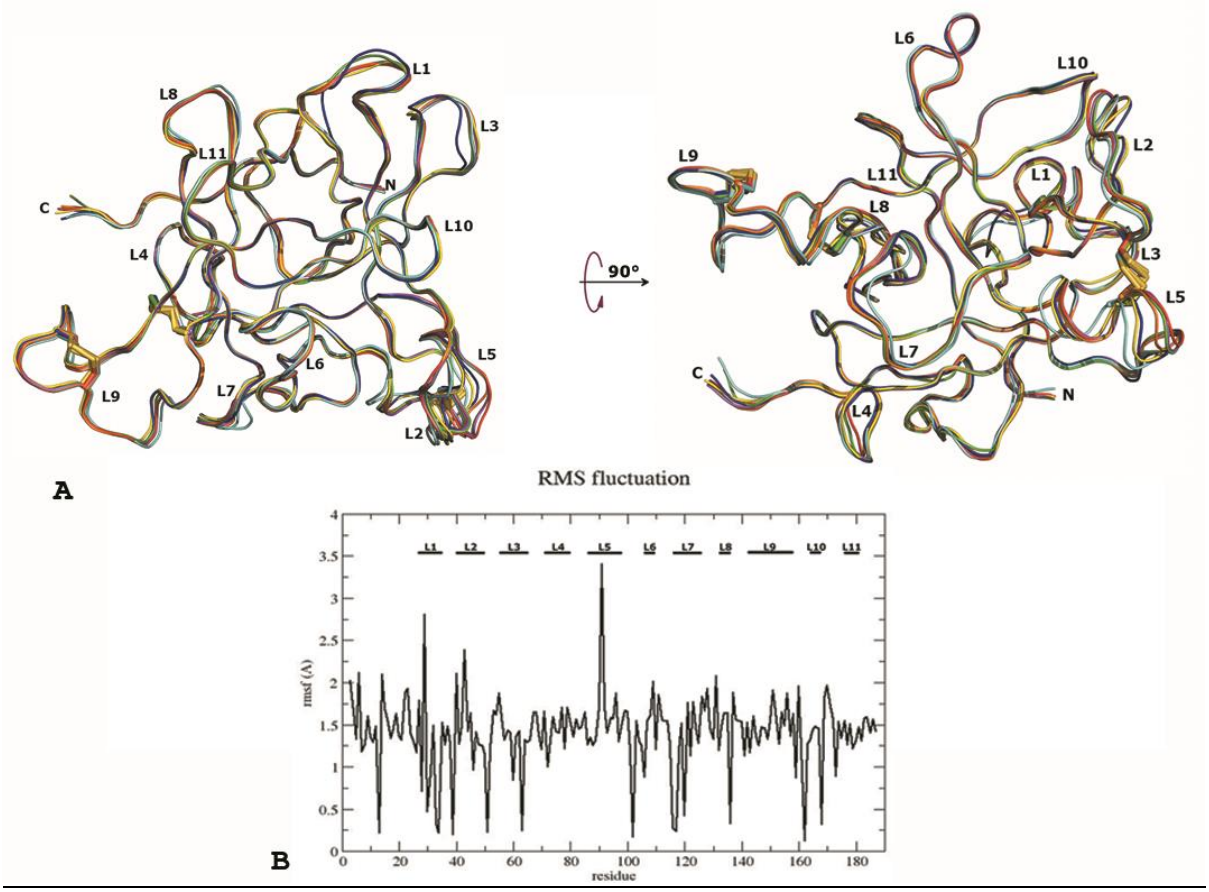

Figure 2. Structural superposition of all structures of inhibitor E3Ad and mutant E3A_N19D. 5FNW (red), 5FNX (green), 5FZY chain A (blue), 5FZY chain B (yellow), 5FZU (magenta), 5FZZ (cyan) and 5G00 (orange) (A). The root mean square fluctuation (RMSF) among the heavy atoms of all the X-ray structures determined using E3Ad_N19D structure with PDB code 5FNW as the reference (B).

Sequence and structural analysis

The amino acid sequence of inhibitor E3Ad matches with a computationally predicted protein from *Solanum tuberosum* (potato) (UniProt identifier: M1AKE5_SOLTU). The closest functionally characterized Kunitz-type STI inhibitor deposited in the UniProt database is API9_SOLTU (95.7% sequence identity) (Ritonja et al., 1990), which inhibit porcine pancreatic trypsin (E.C. 3.4.21.4), bovine chymotrypsin (E.C. 3.4.21.1), human leukocyte elastase (EC 3.4.21.37), human and bovine cathepsin D (EC 3.4.23.5) and Saccharopepsin (EC 3.4.21.41) (Cater et al., 2002; Pouvreau et al., 2001). The sequence identity of E3Ad with other inhibitors of the Kunitz-type STI family characterized so far are: PDI (90.4%) (Keilová H. and Tomášek, 1976a), SLAPI (76.1%) (Cater et al., 2002), PSPI (71.2%) (Meulenbroek et al., 2012), *Murraya koenigii* miraculin-like protein (27.7%) (Gahloth et al., 2010), TKI (26.1%) (Patil et al., 2012) EcTI (25.5%) (Zhou et al., 2013), API-A (24.5%) (Bao et al., 2009), BASI (23.6%) (Vallée et al., 1998) and STI (22.6%) (Song and Suh, 1998). Details of amino acid substitutions in key regions for Kunitz-type STI inhibitors are shown in figure 3.

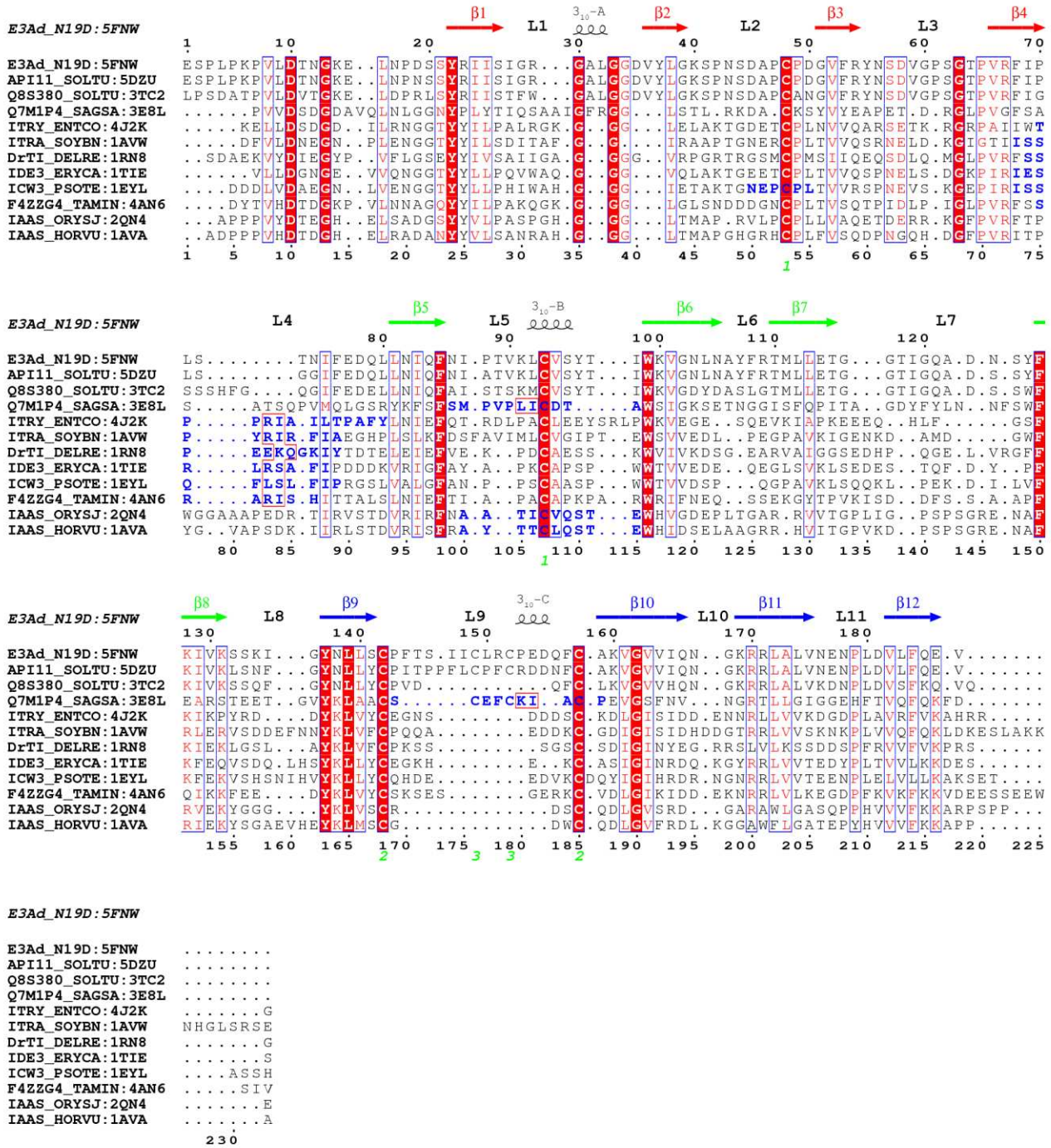

Figure 3. Multiple sequence alignment of inhibitor E3Ad_N19D with other members of the Kunitz-type STI family for which three-dimensional structures are available, with the protease-binding loops identified. β-sheets are coloured according to the three-fold pseudo-symmetry sub-domain to which they belong. Blue and bold: protease-binding loops for serine proteases, red box: P1-P1´ residues, green numbers at the bottom show the disulfide bridge topology. The following UniProt identifiers were used for inhibitors: PDI (API11_SOLTU), PSPI (Q8S380_SOLTU), API-A (Q7M1P4_SAGSA), EcTI (ITRY_ENTCO), STI (ITRA_SOYBN), DrTI (DRTI_DELRE), α-amylase /subtilisin inhibitor from rice (IAAS_ORYSJ), ETI (IDE3_ERYCA), BASI (IAAS_HORVU), double-headed winged bean alpha-chymotrypsin inhibitor (ICW3_PSOTE) and Tamarind Kunitz inhibitor (F4ZZG4_TAMIN). The figure was made with the ESPript 3.0 webserver (http://espript.ibcp.fr/ESPript/cgi-bin/ESPript.cgi).

The three-dimensional structure of, a potato cathepsin D inhibitor PDI that is closely-related to E3Ad (90.4% sequence identity) was reported very recently (Guo et al., 2015). The structural superposition of E3Ad_N19D structure (PDB code: 5FZU) with PDI, showed that they are very similar in the protein core but differ in the conformation of some loops, like L5 (Ans86-Ile98) L7 (Gly116-Tyr126) L8 (Ser132-Gly136) and L9 (Cys142-Cys158). The E3Ad_N19D structure used for the superposition was selected considering similarity with the pH of the crystallization condition for PDI. One of most interesting features of this structural comparison is the significant conformational difference of the loop L9 (S141-A159) in the two structures (Figure 4). Most of the sequence differences between the inhibitors are located in this loop (10 out 17 residues) and also crystallographic packing is different in both cases, which could explain this difference in conformations. Interestingly, this loop was predicted in a previous study, using a combination of computational tools, as the region involved in aspartic protease inhibition (Guerra et al., 2012). In addition, in that work it was also predicted that residues I144, V148, L49, P151, F152 and R154 (tomato inhibitor numbering) could be involved in the difference in potency against the proteases cathepsin D and saccharopepsin reported previously for tomato and potato inhibitors (Cater et al., 2002), and all these residues are localized in loop L9.

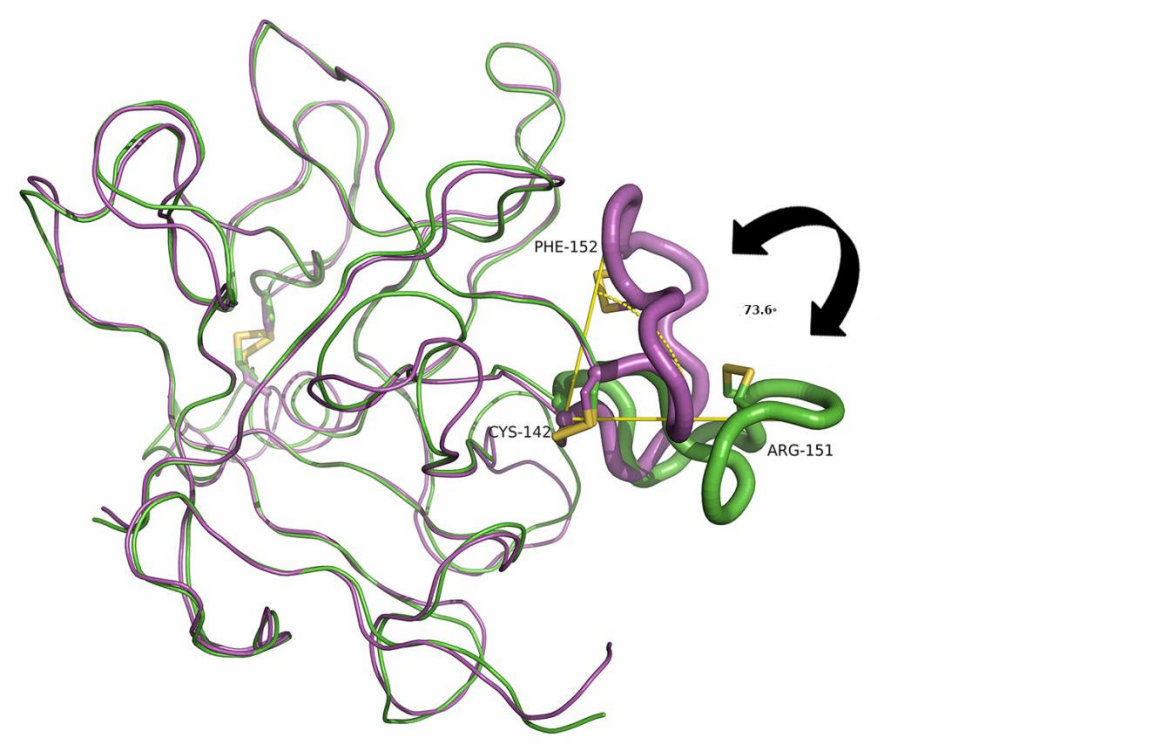

Figure 4. Structural superposition of E3Ad_N19D (5FZU, green) and PDI (5DZU chain A, magenta). Loop L9 is represented with a higher ribbon radius and the angle between equivalent residues Arg151 (5FZU) and Phe152 (5DZU) is shown.

A molecular dynamics simulation study was performed to explore whether these conformations could be a result of the crystal packing or could be possible in solution. This analysis did not reveal large conformational changes at loop L9 of the E3Ad_N19D inhibitor but the results showed that loop L9 in PDI structure (5DZU) changes its conformation after 20 ns at pH 4.4 and 310 K (Figure S3). This conformational rearrangement was not observed in simulations of PDI at pH 8. This is quite interesting because it suggests that loop L9 in PDI could adopt a conformation more similar to the equivalent loop in the E3Ad inhibitor reported in this work.

The analysis performed with the PISA server based on the crystal arrangement, suggested that the E3Ad_N19D inhibitor could exist as a dimer in solution, however, size exclusion chromatography profiles show that, at least at pH 4.0, E3Ad_N19D behaves as a monomer (Figure S2). A similar prediction was made for inhibitor PSPI (PDB code: 3TC2), which has been proved to be a monomer in solution (Meulenbroek et al., 2012).

Enzyme Inhibition

Inhibitor E3Ad_N19D showed a typical tight-binding behavior against the aspartic protease Plm II as well as the serine protease bovine trypsin with K_i values of 10.5 and 43 nM respectively (Figure 5). These results prove that glycosylation is not essential for aspartic protease activity. Previous work has demonstrated that glycosylation was not essential for the inhibition of the serine protease trypsin by Kunitz-type STI family inhibitors (Bao et al., 2009; Speranskaya et al., 2006). The structural superposition of coordinates of the inhibitor E3Ad and E3Ad_N19D showed that the presence of the sugar moiety does not make any significant change to the overall three-dimensional structure of the protein, which correlates well with the inhibitory activity assay results.

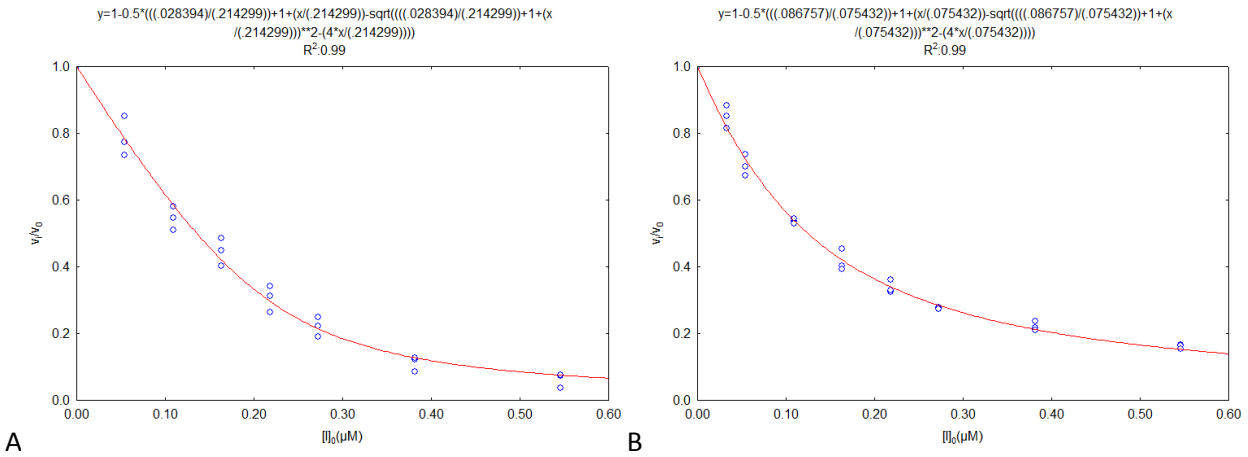

Figure 5. Inhibition curves fitted to Morrison´s equation for inhibitor E3Ad_N19D against the proteases Plm II (A) and bovine trypsin (B).

Based on structural analysis and the possibility that E3Ad_N19D could be able to interact with two molecules of trypsin in a similar way to that described for the inhibitor PSPI (PDB code: 3TC2), the stoichiometry of the trypsin-E3Ad_N19D binding was calculated using an enzyme titration assay, and an $[I]_0/[E]_0$ value of 1.18 was obtained (Figure S1). This result proved that the inhibitor E3Ad_N19D interacts with just one molecule of trypsin.

Protease-binding loop for serine proteases

The Kunitz-type STI inhibitor protease-binding loops involved in serine protease inhibition are highly variable in sequence and length (Figure 3) with almost all cases adopting a canonical conformation at positions P3-P3´. For most of the Kunitz-type STI family inhibitors structurally characterized so far, the protease-binding loop for serine proteases is located between strands β4 and β5 (Loop L4), however there are some inhibitors where it lies between strands β2- β3 (Loop L2), β5-β6 (Loop L5) or β9-β10 (Loop L9) (Figure 3) (Azarkan et al., 2011; Bao et al., 2009). In addition, there is just one double-headed inhibitor (API-A) where one of its protease-binding loops (loop L5) adopts a novel, non-canonical conformation, and even has an unusual Leu P1 site, that was confirmed by site-directed mutagenesis experiments (Bao et al., 2009).

The loop L4 in the E3Ad_N19D inhibitor is shorter and does not contain any equivalent amino acids for positions P1-P1´ in comparison with other members with a canonical conformation in this region (Figure 3). Besides, in all three-dimensional structures of the E3Ad/E3Ad_N19D inhibitors determined in this work, the conformation of the loop L4 is such that the end of the loop is oriented toward the N-terminus (or N-terminal part) of the protein and not protruding to the solvent, an orientation that might make it difficult to allow any interaction with the active site of a protease. Structural comparison of the E3Ad_N19D loop L5 with identified protease-binding loops for serine proteases present in other Kunitz-type STI family inhibitors, showed that the closest structural similarity is with the non-canonical loop L5 of API-A. The RMSD values ranged from 0.71 Å to 0.94 Å for residues Val85-Asp90 (P3-P3') for the API-A inhibitor and residues Thr89-Val94 of the E3Ad and E3Ad_N19D inhibitors (Figure 6A, C).

These RMSD values become even lower if we only consider positions P3-P2´ (0.13Å to 0.42 Å). In contrast, when the structural superposition was made with the canonical binding loops (mainly L4), the RMSD values were much higher: 2.75 Å (PDB code: 4J2K), 2.79 Å (PDB code: 4AN6) and 2.72 Å (PDB code: 1AVW). Interestingly, the loop L5 of the inhibitors PDI and PSPI (both also isolated from potato), not only have a similar conformation (Figure 6A), but also have a 3_{10} helix in equivalent residues to API-A and E3Ad_N19D. Phylogenetic analysis showed that inhibitors E3Ad, PDI and PSPI are closer to API-A than to the canonical Kunitz-type STI family inhibitors such as, for example STI (Figure 6B). These results are similar to those obtained by Azarkan et al., 2011, where they showed that API-A is evolutionarily closer to API8_SOLTU (an aspartic protease inhibitor from the Kunitz-type STI family) than to the canonical inhibitors of the Kunitz-type STI family. It is worth noting that proposed P1 residues for inhibitors PSPI and PDI are also located in loop L5, and, in both cases, they have a Lys in this position as in the E3Ad inhibitor (Figure 6C) (Guo et al., 2015; Meulenbroek et al., 2012). Lys is a preferred amino acid for the P1 position in protease-binding loop for serine proteases (Krowarsch et al., 2003; Tyndall et al., 2005). On the other hand, despite the fact that amino acid Leu92 present in E3Ad and PDI is not a common P1´ residue, the similar amino acid Ile is present in that position in inhibitors with canonical conformations like STI (PDB code: 1AVW), EcTI (PDB code: 4J2K), TKI (PDB code: 4AN4) (Patil et al., 2012) and in both protease-binding loops of the double-headed inhibitor API-A (PDB code: 3TC2) (Figure 4).

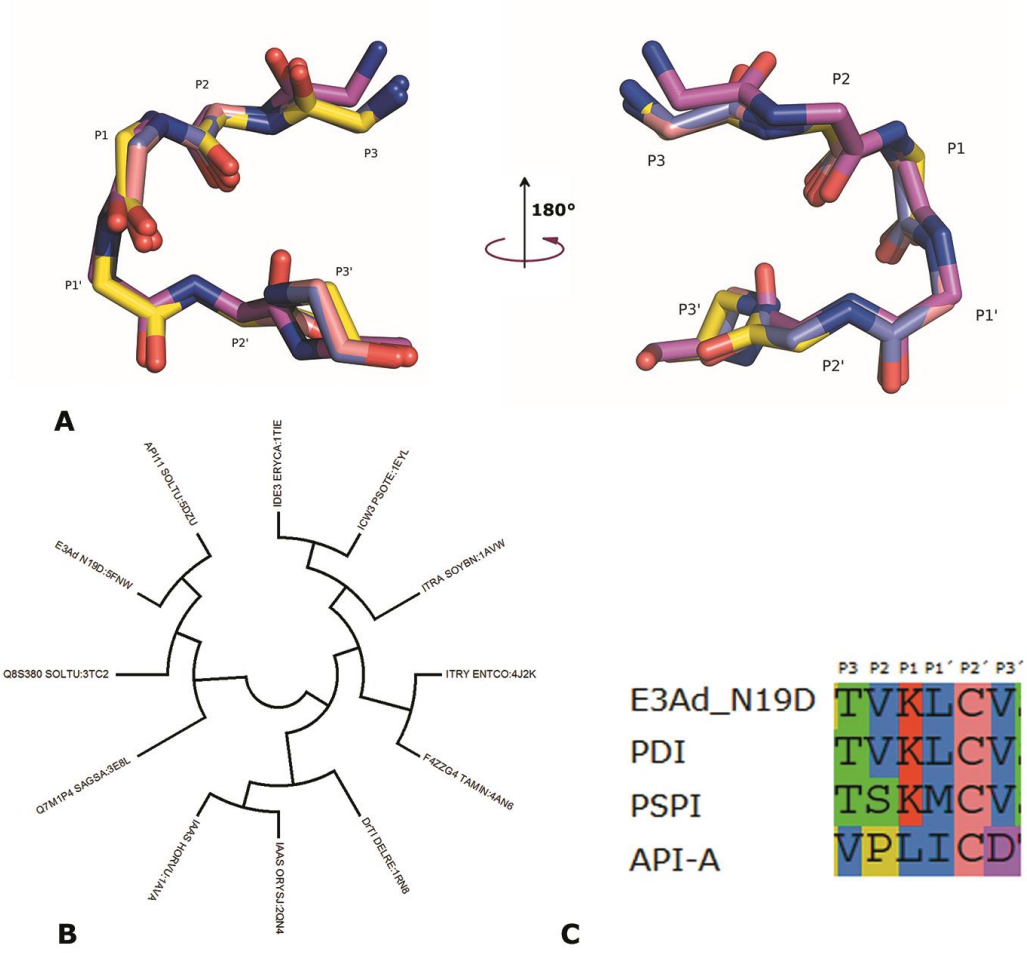

Figure 6. Structural superposition of loop L5 of inhibitor E3Ad_N19D (5FZU, yellow) with non-canonical protease-binding loop of API-A (3E8L, magenta), loop L5 from PDI (5DZU, salmon) and PSPI (3TC2, blue) (A). Consensus tree (1000 replicates) using the maximum likelihood method calculated for the multiple sequence alignment of Figure 3. (B). Sequences of the proposed P3-P3´ positions in loop L5 for inhibitors E3Ad_N19D, PDI, PSPI and API-A (C).

Considering the phylogenetic relationship between E3Ad and API-A, a structural superposition of loop L9 of both inhibitors was performed taking into account that API-A´s loop L9 is a canonical protease-binding loop for serine proteases that has been validated by site-directed mutagenesis (Bao et al., 2009). However, there are several insertions in the loop L9 of E3Ad_N19D with respect to API-A and the RMSD values between both loops, considering Phe143-Cys148 (P3-P3´) of API-A and Ile148-Pro153 of E3Ad, is higher than 2.0 Å. This difference indicates that loop L9 in E3Ad inhibitor does not adopt a canonical conformation.

In the case of inhibitor PSPI, Meulenbroek et al., 2012 proposed Phe75 as a P1 residue in a secondary protease-binding loop, however, there is no equivalent amino acid in inhibitor E3Ad_N19D (Figure 3), which could explain why PSPI is able to interact with two serine proteases simultaneously while E3Ad_N19D interacts with just one (Figure S1). Considering all this evidence, we propose that inhibitor E3Ad (and E3Ad_N19D) could interact with trypsin through a non-canonical protease-binding loop (loop L5), with Lys91 and Leu92 as the P1-P1´residues.

Protease-binding loop for aspartic proteases

The case of proteinaceous inhibitors of aspartic proteases is quite different from serine proteases. To date, there are only two three-dimensional structures of protease-inhibitor complexes, with no amino acid sequence or structural similarity between the two inhibitors, or with other inhibitor families (Li et al., 2000; Ng et al., 2000). Previously, combining different computational tools, loop L9 of a tomato aspartic protease inhibitor of the Kunitz-type STI family was predicted as the region responsible for the inhibition of aspartic proteases (Guerra et al., 2012). More recently, Guo et al., 2015 also suggested that equivalent loop L9 in PDI inhibitor could be implicated in the inhibition of cathepsin D, although their favoured docking models place residue Lys91 (loop L5) interacting with catalytic aspartic acids of cathepsin D.

The analysis of three-dimensional structures of inhibitor E3Ad_N19D and PDI with the metaPPISP web server predicted that in both cases loop L9 could be part of a Protein-Protein interface (Figure S4). In addition, inhibitor E3Ad_N19D (PDB code: 5FNW) was docked to a three-dimensional apo structure of Plm II (PDB code: 1LF4) using the ClusPro 2.0 web server (Comeau et al., 2004b). The solutions of the balanced and electrostatic coefficients were analyzed, and the top solutions for both coefficients placed the residue Arg151 (loop L9) at interaction distance from both catalytic aspartic acids (Asp34 and Asp214) (Figure 7A). The roles of basic residues in the inhibition of aspartic proteases have been described for self-inhibition of the aspartic proteases progastricin (Moore et al., 1995), pepsinogen (James and Sielecki, 1986), cathepsin D (Lee et al., 1998) and also for the complex saccharopepsin-inhibitor IA3 (Li et al., 2000). Loop L9 has been previously proposed as the region involved in aspartic protease inhibition (Guerra et al., 2012). Also, it is important to note that Cater et al., 2002 reported that bond hydrolysis between residues Arg154-Gly155 (loop L9) of the tomato aspartic protease inhibitor SLAPI affects aspartic protease inhibition, doubling the K_i value against saccharopepsin. It is also interesting that loop L9 is in an equivalent sequence position to a canonical protease-binding loop used by the double-headed inhibitor API-A to interact with trypsin (Figure 3). Considering the present structural and computational analyses, results from Guo et al., 2015 may be explained by a conformational change of loop L9.

Despite the fact that the docking results obtained by Guo et al., 2015 with the inhibitor PDI suggested a different protease-binding loop (loop L5) as their favoured interaction site for aspartic proteases, the results discussed in the present work reinforce our previous hypothesis about the protease-binding loop for aspartic protease inhibition (Guerra et al., 2012). A plausible explanation is that the specific conformation observed in the PDI structure for the loop L9 led to docking solutions where Lys91 is interacting with the active site of cathepsin D. To explore this option we performed a docking study with the coordinates of a molecular dynamics simulation (93.77 ns) with the structure 5DZU at pH 4.4 that showed a change in the conformation of the loop L9 and the same Plm II apo structure used in docking with E3Ad_N19D. The results showed that in top solutions PDI´s loop L9 is placed in the active site of Plm II, although there is no direct interaction with catalytic aspartic acids as shown for E3AD_N19D (Figure 7B). Even though this approach is theoretical, it is remarkable that the simulation

suggested that if PDI´s loop L9 could adopt an extended conformation it could be possible for it to interact with the active site of an aspartic protease, just as has been modeled for the E3Ad_N19D inhibitor.

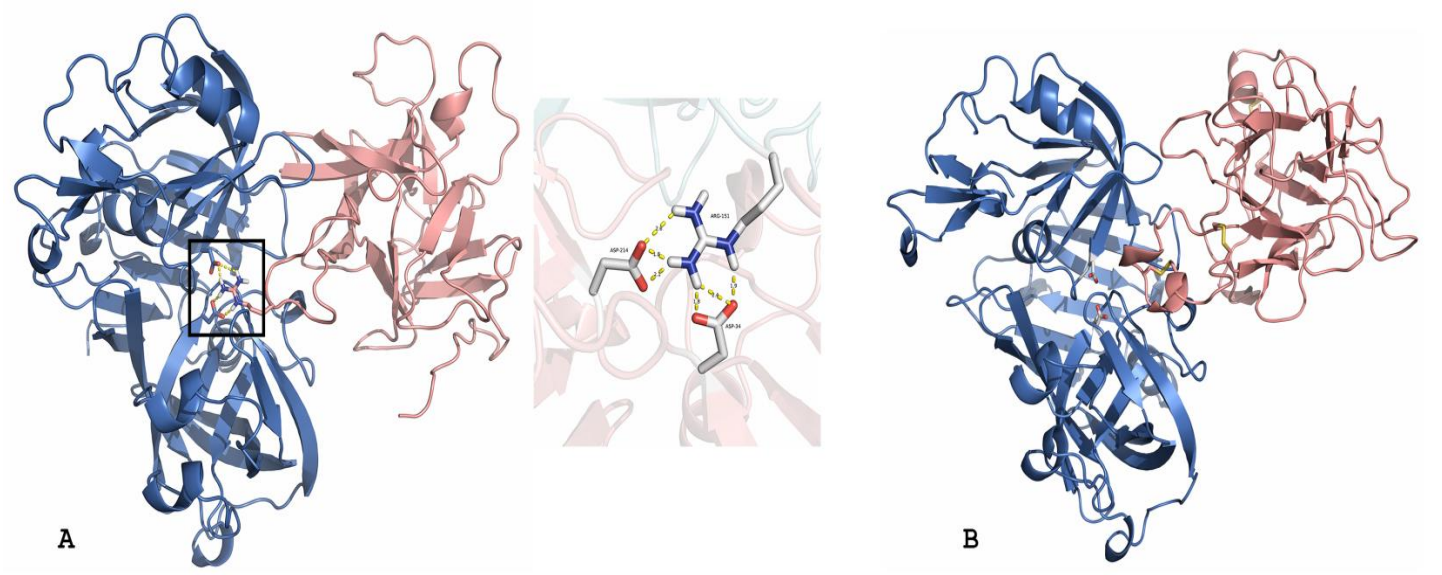

Figure 7. Docking models of the complex between inhibitors E3Ad_N19D and PDI with Plm II. Interactions of the Arg151 of the inhibitor E3Ad_N19D (PDB code: 5FNW) with Plm II´s catalytic aspartic acids (A). Modeled interaction of coordinates of PDI (PDB code: 5DZU), after 93.77 ns of molecular dynamic simulation, with Plm II (B). Plm II and inhibitors (E3Ad_N19D and PDI) are shown in blue and salmon, respectively.

The presence of a pseudo three-fold symmetry, as a result of double gene duplication and fusion events, has been suggested as a selective advantage to develop new functions (Eisenbeis and Höcker, 2010). Bi-functional inhibitors of the I85 family which shares the β-trefoil fold, although having low sequence similarity with Kunitz-type STI family inhibitors, use loops located in different pseudo-symmetry elements to interact with serine and cysteine proteases (Renko et al., 2010). Maybe this fact could be related to a hypothesis proposing that ancestral reactive sites could evolve independently allowing the acquisition of a new functionality or sub-specialization (Christeller, 2005). Considering the present structural and computational analysis of this work, this hypothesis could explain what happened with loop L9 in bifunctional inhibitors of aspartic and serine proteases of the Kunitz-type STI family. Of course this will need experimental studies to confirm or refute the relationship between aspartic and canonical serine protease-binding loops.

Conclusions

The crystallographic structures reported here show that inhibitor E3Ad maintains basically the same conformation at different pH values and crystal packings, with only minor changes in some of its loops L1, L2 and L5. Moreover, the structural comparison with the closely-related inhibitor PDI suggests that loop L9 can adopt different conformations depending on pH, loop sequence and also crystal packing. Based on sequence and structural analysis we propose that protease-binding loops for serine and aspartic protease inhibition in E3Ad could be the loops L5 and L9, respectively. The proposed binding loops for serine and aspartic proteases for the bifunctional inhibitor E3Ad are in equivalent loops with two binding loops for serine proteases in the double-headed inhibitor API-A. In addition, phylogenetic analysis suggests that inhibitor E3Ad (and also PDI and PSPI) could have a closer evolutionary relationship with double-headed inhibitor API-A than with members of the Kunitz-type STI-family with canonical protease-binding loops. Whether the inhibition of aspartic proteases in the particular case of the Kunitz-type STI inhibitor family could have evolved from a loop used for serine protease inhibition or not, is an intriguing question that a three-dimensional structure of a complex between an aspartic protease and a bifunctional inhibitor like E3Ad could help to answer.

Acknowledgements

ER-P acknowledges financial support from CONACYT project No. 204639 and PAPIIT IN209114. YG is supported by a PhD fellowship from CONACYT and by Grant F/4927-1(YG) from International Foundation for Science (Sweden). X-ray diffraction data were collected at beamline 19-ID at the Argonne National Laboratory, Structural Biology Center at the Advanced Photon Source (APS). Argonne is operated by UChicago Argonne, LLC, for the U.S. Department of Energy, Office of Biological and Environmental Research under contract DE-AC02-06CH11357. The authors thank the Argonne staff for their support during data collection, particularly Dr. Norma Duke. This research was also performed at BL14-1 at the SSRL under the agreement between SSRL-SMB and NSLSII-LSBR. SMB operates under NIH-NIGMS grant P41GM103393 and a DOE-BER; and SSRL operates under DOE-BES contract No. DE-AC02-76SF00515. LSBR program is supported by DOE BER contract DE-SC0012704 and by NIH-NIGMS grant P41GM111244. NSLSII is operated under DOE BES contract # DE-AC02-98CH10886. The authors thank the SSRL and NSLS staff for their support during data collection, particularly Dr. Vivian Stojanoff.

References

Azarkan, M., Martinez-Rodriguez, S., Buts, L., Baeyens-Volant, D., Garcia-Pino, A., 2011. The plasticity of the β-trefoil fold constitutes an evolutionary platform for protease inhibition. J. Biol. Chem. 286, 43726–43734. doi:10.1074/jbc.M111.291310

Bao, R., Zhou, C.Z., Jiang, C., Lin, S.X., Chi, C.W., Chen, Y., 2009. The ternary structure of the double-headed arrowhead protease inhibitor API-A complexed with two trypsins reveals a novel reactive site conformation. J. Biol. Chem. 284, 26676–26684. doi:10.1074/jbc.M109.022095

Bauw, G., Nielsen, H. V., Emmersen, J., Nielsen, K.L., Jørgensen, M., Welinder, K.G., 2006. Patatins, Kunitz protease inhibitors and other major proteins in tuber of potato cv. Kuras. FEBS J. 273, 3569–3584. doi:10.1111/j.1742-4658.2006.05364.x

Bieth, J.G., 1995. Theoretical and practical aspects of proteinase inhibition kinetics. Methods Enzymol. 248, 59–84. doi:10.1016/0076-6879(95)48007-2

Broom, A., Doxey, A.C., Lobsanov, Y.D., Berthin, L.G., Rose, D.R., Howell, P.L., McConkey, B.J., Meiering, E.M., 2012. Modular evolution and the origins of symmetry: reconstruction of a three-fold symmetric globular protein. Structure 20, 161–71. doi:10.1016/j.str.2011.10.021

Bussi, G., Donadio, D., Parrinello, M., 2007. Canonical sampling through velocity rescaling. J. Chem. Phys. 126. doi:http://dx.doi.org/10.1063/1.2408420

Cater, S. a., Lees, W.E., Hill, J., Brzin, J., Kay, J., Phylip, L.H., 2002. Aspartic proteinase inhibitors from tomato and potato are more potent against yeast proteinase A than cathepsin D. Biochim. Biophys. Acta - Protein Struct. Mol. Enzymol. 1596, 76–82. doi:10.1016/S0167-4838(02)00206-6

Chen, V.B., Arendall, W.B., Headd, J.J., Keedy, D.A., Immormino, R.M., Kapral, G.J., Murray, L.W., Richardson, J.S., Richardson, D.C., 2010. MolProbity: all-atom structure validation for macromolecular crystallography. Acta Crystallogr. Sect. D 66, 12–21. doi:10.1107/S0907444909042073

Christeller, J.T., 2005. Evolutionary mechanisms acting on proteinase inhibitor variability. FEBS J. 272, 5710–5722. doi:10.1111/j.1742-4658.2005.04975.x

Comeau, S.R., Gatchell, D.W., Vajda, S., Camacho, C.J., 2004a. ClusPro: A fully automated algorithm for protein-protein docking. Nucleic Acids Res. 32, 96–99. doi:10.1093/nar/gkh354

Comeau, S.R., Gatchell, D.W., Vajda, S., Camacho, C.J., 2004b. ClusPro: An automated docking and discrimination method for the prediction of protein complexes. Bioinformatics 20, 45–50.

doi:10.1093/bioinformatics/btg371

Darden, T., York, D., Pedersen, L., 1993. Particle mesh Ewald: An N·log(N) method for Ewald sums in large systems. J. Chem. Phys. 98.

Di Tommaso, P., Moretti, S., Xenarios, I., Orobitg, M., Montanyola, A., Chang, J.-M., Taly, J.-F., Notredame, C., 2011. T-Coffee: a web server for the multiple sequence alignment of protein and RNA sequences using structural information and homology extension. Nucleic Acids Res. 39, W13–W17. doi:10.1093/nar/gkr245

Dolinsky, T.J., Czodrowski, P., Li, H., Nielsen, J.E., Jensen, J.H., Klebe, G., Baker, N.A., 2007. PDB2PQR: expanding and upgrading automated preparation of biomolecular structures for molecular simulations. Nucleic Acids Res. 35 , W522–W525. doi:10.1093/nar/gkm276

Eisenbeis, S., Höcker, B., 2010. Evolutionary mechanism as a template for protein engineering. J. Pept. Sci. 16, 538–544. doi:10.1002/psc.1233

Emsley, P., Lohkamp, B., Scott, W.G., Cowtan, K., 2010. Features and development of Coot. Acta Crystallogr. Sect. D Biol. Crystallogr. 66, 486–501. doi:10.1107/S0907444910007493

Fischer, M., Kuckenberg, M., Kastilan, R., Muth, J., Gebhardt, C., 2015. Novel in vitro inhibitory functions of potato tuber proteinaceous inhibitors. Mol. Genet. Genomics 290, 387–398. doi:10.1007/s00438-014-0906-5

Franco, O.L., De Grossi Sá, M.F., Sales, M.P., Mello, L. V., Oliveira, A.S., Rigden, D.J., 2002. Overlapping binding sites for trypsin and papain on a Kunitz-type proteinase inhibitor from Prosopis juliflora. Proteins Struct. Funct. Genet. 49, 335–341. doi:10.1002/prot.10228

Gahloth, D., Selvakumar, P., Shee, C., Kumar, P., Sharma, A.K., 2010. Cloning, sequence analysis and crystal structure determination of a miraculin-like protein from Murraya koenigii. Arch. Biochem. Biophys. 494, 15–22. doi:10.1016/j.abb.2009.11.008

Glaczinski, H., Heibges, A., Salamini, F., Gebhardt, C., 2002. Members of the Kunitz-type protease inhibitor gene family of potato inhibit soluble tuber invertase in vitro. Potato Res. 45, 163–176. doi:10.1007/BF02736112

Guerra, Y., Valiente, P. a., Berry, C., Pons, T., 2012. Predicting functional residues of the Solanum lycopersicum aspartic protease inhibitor (SLAPI) by combining sequence and structural analysis with molecular docking. J. Mol. Model. 18, 2673–2687. doi:10.1007/s00894-011-1290-2

Guo, J., Erskine, P.T., Coker, A.R., Wood, S.P., Cooper, J.B., 2015. Structure of a Kunitz-type potato cathepsin D inhibitor. J. Struct. Biol. 192, 554–560. doi:http://dx.doi.org/10.1016/j.jsb.2015.10.020

Headey, S.J., MacAskill, U.K., Wright, M. a., Claridge, J.K., Edwards, P.J.B., Farley, P.C., Christeller, J.T., Laing, W. a., Pascal, S.M., 2010. Solution structure of the squash aspartic acid proteinase inhibitor (SQAPI) and mutational analysis of pepsin inhibition. J. Biol. Chem. 285, 27019–27025. doi:10.1074/jbc.M110.137018

Heibges, a, Glaczinski, H., Ballvora, a, Salamini, F., Gebhardt, C., 2003a. Structural diversity and organization of three gene families for Kunitz-type enzyme inhibitors from potato tubers (Solanum tuberosum L.). Mol. Genet. Genomics 269, 526–34. doi:10.1007/s00438-003-0860-0

Heibges, a, Salamini, F., Gebhardt, C., 2003b. Functional comparison of homologous members of three groups of Kunitz-type enzyme inhibitors from potato tubers (Solanum tuberosum L.). Mol. Genet. Genomics 269, 535–41. doi:10.1007/s00438-003-0861-z

Hess, B., Bekker, H., Berendsen, H.J.C., Fraaije, J.G.E.M., 1997. LINCS: A linear constraint solver for molecular simulations. J. Comput. Chem. 18, 1463–1472. doi:10.1002/(SICI)1096-987X(199709)18:12<1463::AID-JCC4>3.0.CO;2-H

Hockney, R.W., 1970. The potential calculation and some applications, in: Methods in Computational Physics. pp. 136–211.

Hornak, V., Abel, R., Okur, A., Strockbine, B., Roitberg, A., Simmerling, C., 2006. Comparison of multiple Amber force fields and development of improved protein backbone parameters. Proteins Struct. Funct. Bioinforma. 65, 712–725. doi:10.1002/prot.21123

James, M.N.G., Sielecki, A.R., 1986. Molecular structure of an aspartic proteinase zymogen, porcine pepsinogen, at 1.8 Å resolution. Nature 319, 33–38.

Jones, D.T., Taylor, W.R., Thornton, J.M., 1992. The rapid generation of mutation data matrices from protein sequences. Comput. Appl. Biosci. CABIOS 8 , 275–282. doi:10.1093/bioinformatics/8.3.275

Kabsch, W., 2010. Xds. Acta Crystallogr. Sect. D Biol. Crystallogr. 66, 125–132. doi:10.1107/S0907444909047337

Kantardjieff, K.A., Rupp, B., 2003. Matthews coefficient probabilities: Improved estimates for unit cell contents of proteins, DNA, and protein–nucleic acid complex crystals. Protein Sci. 12, 1865–1871. doi:10.1110/ps.0350503

Karplus, P. a., Diederichs, K., 2012. Linking Crystallographic Model and Data Quality. Science (80-.). 336, 1030–1033. doi:10.1126/science.1218231

Keilová H. and Tomášek, V., 1976a. Isolation and properties of Cathepsin D inhibitor from potatoes. Collect Czech Chem Commun 41, 489–497.

Keilová H. and Tomášek, V., 1976b. Further characteristics of cathepsin D inhibitor from potatoes. Collect. Czech. Chem. Commun. 41, 2440–2447.

Krissinel, E., Henrick, K., 2004. Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. Acta Crystallogr. Sect. D 60, 2256–2268.

Krowarsch, D., Cierpicki, T., Jelen, F., Otlewski, J., 2003. Canonical protein inhibitors of serine proteases. Cell. Mol. Life Sci. 60, 2427–2444. doi:10.1007/s00018-003-3120-x

Laskowski, M., Kato, I., 1980. Protein Inhibitors of Proteinases. Annu. Rev. Biochem. 49, 593–626. doi:10.1146/annurev.bi.49.070180.003113

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 - Protein Struct. Mol. Enzymol. 1477, 324–337. doi:10.1016/S0167-4838(99)00284-8

Lee, a Y., Gulnik, S. V, Erickson, J.W., 1998. Conformational switching in an aspartic proteinase. Nat. Struct. Biol. 5, 866–871. doi:10.1038/2306

Li, M., Phylip, L.H., Lees, W.E., Winther, J.R., Dunn, B.M., Wlodawer, a, Kay, J., Gustchina, a, 2000. The aspartic proteinase from Saccharomyces cerevisiae folds its own inhibitor into a helix. Nat. Struct. Biol. 7, 113–117. doi:10.1038/72378

Mares, M., Meloun, B., Pavlik, M., Kostka, V., Baudys, M., 1989. Primary structure of cathepsin D inhibitor from potatoes and its structure relationship to soybean trypsin inhibitor family. FEBS Lett. 251, 94–98. doi:10.1016/0014-5793(89)81435-8

McLachlan, a D., 1979. Three-fold structural pattern in the soybean trypsin inhibitor (Kunitz). J. Mol. Biol. 133, 557–563. doi:10.1016/0022-2836(79)90408-X

Meulenbroek, E.M., Thomassen, E. a J., Pouvreau, L., Abrahams, J.P., Gruppen, H., Pannu, N.S., 2012. Structure of a post-translationally processed heterodimeric double-headed Kunitz-type serine protease inhibitor from potato. Acta Crystallogr. Sect. D Biol. Crystallogr. 68, 794–799. doi:10.1107/S090744491201222X

Miyamoto, S., Kollman, P.A., 1992. Settle: An analytical version of the SHAKE and RATTLE algorithm for rigid water models. J. Comput. Chem. 13, 952–962. doi:10.1002/jcc.540130805

Moore, S.A., Sielecki, A.R., Chernaia, M.M., Tarasova, N.I., James, M.N.G., 1995. Crystal and Molecular Structures of Human Progastricsin at 1.62 Å Resolution. J. Mol. Biol. 247, 466–485. doi:http://dx.doi.org/10.1006/jmbi.1994.0154

Morrison, J.F., 1969. Kinetics of the reversible inhibition of enzyme-catalysed reactions by tight-binding inhibitors. Biochim. Biophys. Acta - Enzymol. 185, 269–286. doi:http://dx.doi.org/10.1016/0005-2744(69)90420-3

Murshudov, G.N., Skubák, P., Lebedev, A. a., Pannu, N.S., Steiner, R. a., Nicholls, R. a., Winn, M.D., Long, F., Vagin, A. a., 2011. REFMAC5 for the refinement of macromolecular crystal structures. Acta Crystallogr. Sect. D Biol. Crystallogr. 67, 355–367. doi:10.1107/S0907444911001314

Murzin, A.G., Lesk, A.M., Chothia, C., 1992. β-Trefoil fold. J. Mol. Biol. 223, 531–543. doi:http://dx.doi.org/10.1016/0022-2836(92)90668-A

Ng, K.K., Petersen, J.F., Cherney, M.M., Garen, C., Zalatoris, J.J., Rao-Naik, C., Dunn, B.M., Martzen, M.R., Peanasky, R.J., James, M.N., 2000. Structural basis for the inhibition of porcine pepsin by Ascaris pepsin inhibitor-3. Nat. Struct. Biol. 7, 653–657. doi:10.1038/77950

Nosé, S., Klein, M.L., 1983. Constant pressure molecular dynamics for molecular systems. Mol. Phys. 50, 1055–1076. doi:10.1080/00268978300102851

Oliva, M.L. V, Silva, M.C.C., Sallai, R.C., Brito, M. V., Sampaio, M.U., 2010. A novel subclassification for Kunitz proteinase inhibitors from leguminous seeds. Biochimie 92, 1667–1673. doi:10.1016/j.biochi.2010.03.021

Otlewski, J., Jaskólski, M., Buczek, O., Cierpicki, T., 2001. Struc ture-function re la tion ship of serine pro te ase – pro tein in -. Acta Biochim. Pol. 48, 419–428.

Páll, S., Abraham, M.J., Carsten, K., Hess, B., Lindahl, E., 2015. Tackling Exascale Software Challenges In Molecular Simulations with GROMACS 8759, 3–27. doi:10.1007/978-3-319-15976-8

Parrinello, M., Rahman, A., 1981. Polymorphic transitions in single crystals: A new molecular dynamics method. J. Appl. Phys. 52.

Patil, D.N., Chaudhary, A., Sharma, A.K., Tomar, S., Kumar, P., 2012. Structural basis for dual inhibitory role of tamarind Kunitz inhibitor (TKI) against factor Xa and trypsin. FEBS J. 279, 4547–4564. doi:10.1111/febs.12042

Ponting, C.P., Russell, R.B., 2000. Identification of distant homologues of fibroblast growth factors suggests a common ancestor for all β-trefoil proteins1. J. Mol. Biol. 302, 1041–1047. doi:http://dx.doi.org/10.1006/jmbi.2000.4087

Pouvreau, L., Gruppen, H., Piersma, S.R., van den Broek, L. a, van Koningsveld, G. a, Voragen, a G., 2001. Relative abundance and inhibitory distribution of protease inhibitors in potato juice from cv. Elkana. J. Agric. Food Chem. 49, 2864–2874. doi:10.1021/jf010126v

Price, D.J., Brooks, C.L., 2004. A modified TIP3P water potential for simulation with Ewald summation. J. Chem. Phys. 121.

Qin, S., Zhou, H.-X., 2007. meta-PPISP: a meta web server for protein-protein interaction site prediction. Bioinformatics 23, 3386–7. doi:10.1093/bioinformatics/btm434

Radisky, E.S., Koshland, D.E., 2002. A clogged gutter mechanism for protease inhibitors. Proc. Natl. Acad. Sci. U. S. A. 99, 10316–10321. doi:10.1073/pnas.112332899

Ramírez, A.R., Guerra, Y., Otero, A., García, B., Berry, C., Mendiola, J., Hernández-Zanui, A., Chávez, M. de los A., 2009. Generation of an affinity matrix useful in the purification of natural inhibitors of plasmepsin II,

an antimalarial-drug target. Biotechnol. Appl. Biochem. 52, 149–157. doi:10.1042/BA20080015

Rawlings, N.D., Waller, M., Barrett, A.J., Bateman, A., 2014. MEROPS: The database of proteolytic enzymes, their substrates and inhibitors. Nucleic Acids Res. 42, 503–509. doi:10.1093/nar/gkt953

Renko, M., Sabotič, J., Mihelič, M., Brzin, J., Kos, J., Turk, D., 2010. Versatile loops in mycocypins inhibit three protease families. J. Biol. Chem. 285, 308–316. doi:10.1074/jbc.M109.043331

Renko, M., Sabotič, J., Turk, D., 2012. β-Trefoil inhibitors - From the work of Kunitz onward. Biol. Chem. 393, 1043–1054. doi:10.1515/hsz-2012-0159

Ritonja, a, Krizaj, I., Mesko, P., Kopitar, M., Lucovnik, P., Strukelj, B., Pungercar, J., Buttle, D.J., Barrett, a J., Turk, V., 1990. The amino acid sequence of a novel inhibitor of cathepsin D from potato. FEBS Lett. 267, 13–15.

Schechter, I., Berger, A., 1967. On the size of the active site in proteases. I. Papain. Biochem. Biophys. Res. Commun. 27, 157–162. doi:http://dx.doi.org/10.1016/S0006-291X(67)80055-X

Song, H.K., Suh, S.W., 1998. Kunitz-type soybean trypsin inhibitor revisited: refined structure of its complex with porcine trypsin reveals an insight into the interaction between a homologous inhibitor from Erythrina caffra and tissue-type plasminogen activator. J. Mol. Biol. 275, 347–363. doi:10.1006/jmbi.1997.1469

Speranskaya, a S., Krinitsina, a a, Revina, T. a, Gerasimova, N.G., Keruchen'ko, Y.S., Shevelev, a B., Valueva, T. a, 2006. Heterologous expression, purification, and properties of a potato protein inhibitor of serine proteinases. Biochemistry. (Mosc). 71, 1176–1182. doi:10.1134/S0006297906110022

Strong, M., Sawaya, M.R., Wang, S., Phillips, M., Cascio, D., Eisenberg, D., 2006. Toward the structural genomics of complexes: Crystal structure of a PE/PPE protein complex from Mycobacterium tuberculosis. Proc. Natl. Acad. Sci. 103 , 8060–8065. doi:10.1073/pnas.0602606103

Strukelj, B., Pungercar, J., Mesko, P., Barlic-Maganja, D., Gubensek, F., Kregar, I., Turk, V., 1992. Characterization of aspartic proteinase inhibitors from potato at the gene, cDNA and protein levels. Biol. Chem. Hoppe Seyler 373, 477–482.

Tamura, K., Stecher, G., Peterson, D., Filipski, A., Kumar, S., 2013. MEGA6: Molecular Evolutionary Genetics Analysis Version 6.0. Mol. Biol. Evol. 30 , 2725–2729. doi:10.1093/molbev/mst197

Tyndall, J.D.A., Nall, T., Fairlie, D.P., 2005. Proteases universally recognize beta strands in their active sites. Chem. Rev. 105, 973–999. doi:10.1021/cr040669e

Vallée, F., Kadziola, A., Bourne, Y., Juy, M., Rodenburg, K.W., Svensson, B., Haser, R., 1998. Barley α-amylase bound to its endogenous protein inhibitor BASI: crystal structure of the complex at 1.9 å resolution. Structure 6, 649–659. doi:http://dx.doi.org/10.1016/S0969-2126(98)00066-5

Weichenberger, C.X., Rupp, B., 2014. Ten years of probabilistic estimates of biocrystal solvent content: new insights via nonparametric kernel density estimate. Acta Crystallogr. Sect. D 70, 1579–1588.

Winn, M.D., Ballard, C.C., Cowtan, K.D., Dodson, E.J., Emsley, P., Evans, P.R., Keegan, R.M., Krissinel, E.B., Leslie, A.G.W., McCoy, A., McNicholas, S.J., Murshudov, G.N., Pannu, N.S., Potterton, E. a., Powell, H.R., Read, R.J., Vagin, A., Wilson, K.S., 2011. Overview of the CCP4 suite and current developments. Acta Crystallogr. Sect. D Biol. Crystallogr. 67, 235–242. doi:10.1107/S0907444910045749

Zakharova, E., Horvath, M.P., Goldenberg, D.P., 2009. Structure of a serine protease poised to resynthesize a peptide bond. Proc. Natl. Acad. Sci. U. S. A. 106, 11034–11039. doi:10.1073/pnas.0902463106

Zhou, D., Lobo, Y. a., Batista, I.F.C., Marques-Porto, R., Gustchina, A., Oliva, M.L. V, Wlodawer, A., 2013. Crystal Structures of a Plant Trypsin Inhibitor from Enterolobium contortisiliquum (EcTI) and of Its Complex with Bovine Trypsin. PLoS One 8. doi:10.1371/journal.pone.0062252