Protein Science

Canonical or noncanonical? Structural plasticity of serine protease-binding loops in Kunitz-STI protease inhibitors.

Journal:	*Protein Science*
Manuscript ID	PRO-22-0382.R1
Wiley - Manuscript type:	Full-Length Papers
Date Submitted by the Author:	08-Dec-2022
Complete List of Authors:	Guerra, Yasel; Universidad de Las Américas, Ingeniería en Biotecnología, Facultad de Ingeniería y Ciencias Aplicadas.; Universidad de Las Americas, Grupo de Bio-Quimioinformática Armijos-Jaramillo, Vinicio; Universidad de Las Americas, Ingeniería en Biotecnología, Facultad de Ingeniería y Ciencias Aplicadas.; Universidad de Las Americas, Grupo de Bio-Quimioinformática Pons, Tirso; Centro Nacional de Biotecnologia, Departamento de Inmunología y Oncología Tejera, Eduardo; Universidad de Las Americas, Ingeniería en Biotecnología, Facultad de Ingeniería y Ciencias Aplicadas.; Universidad de Las Américas, Grupo de Bio-Quimioinformática Berry, Colin; Cardiff University, Cardiff School of Biosciences
Keywords:	Kunitz, protease inhibitors, canonical inhibitors, noncanonical inhibitors, serine proteases

Canonical or noncanonical? Structural plasticity of serine protease-binding loops in Kunitz-STI protease inhibitors.

Yasel Guerra[1,2*], Vinicio Armijos-Jaramillo[1,2], Tirso Pons[3], Eduardo Tejera[1,2], Colin Berry[4*]

[1]Ingeniería en Biotecnología, Facultad de Ingeniería y Ciencias Aplicadas. Universidad de Las Américas, Quito. Ecuador.

[2]Grupo de Bio-Quimioinformática. Universidad de Las Américas, Quito. Ecuador.

[3]Department of Immunology and Oncology, National Centre for Biotechnology (CNB-CSIC), Madrid, Spain

[4]Cardiff School of Biosciences, Cardiff University, Cardiff CF10 3AX, Wales, UK.

Correspondence:

Yasel Guerra,
Grupo de Bio-Quimioinformática.
Universidad de Las Américas,
Redondel del Ciclista, Antigua Vía a Nayón,
Quito EC 170124
Email: yasel.guerra.borrego@udla.edu.ec

Colin Berry
School of Biosciences
Cardiff University
Museum Avenue, Cardiff
CF10 3AX, UK
Email: Berry@cf.ac.uk

Total number of manuscript pages: 23 Supplementary material pages: 11

Figures: 6

Supplementary material:

Additional data of the structural analyses performed in this work including general information of the inhibitors analyzed, global and local root mean deviations of the structures superpositions, and values of peptide torsion angles.

Abstract

The Kunitz-Soybean Trypsin Inhibitor (Kunitz-STI) family is a large family of proteins with most of its members being protease inhibitors. The versatility of the inhibitory profile and the structural plasticity of these proteins, make this family a promising scaffold for designing new multifunctional proteins. Historically, Kunitz-STI inhibitors have been classified as canonical serine protease inhibitors, but new inhibitors with novel inhibition mechanisms have been described in recent years. Different inhibition mechanisms could be the result of different evolutionary pathways. In the present work, we performed a structural analysis of all the crystallographic structures available for Kunitz-STI inhibitors to characterize serine protease-binding loop structural features and locations. Our study suggests a relationship between the conformation of serine protease-binding loops and the inhibition mechanism, their location in the β-trefoil fold, and the plant source of the inhibitors. The classical canonical inhibitors of this family are restricted to plants from the Fabales order and bind their targets via the β4-β5 loop, while serine protease-binding loops in inhibitors from other plants lie mainly in the β5-β6 and β9-β10 loops. In addition, we found that the β5-β6 loop is used to inhibit two different families of serine proteases through a steric blockade inhibition mechanism. This work will help to change the general perception that all Kunitz-STI inhibitors are canonical inhibitors and proteins with protease-binding loops adopting noncanonical conformations are exceptions. Additionally, our results will help in the identification of protease-binding loops in uncharacterized or newly discovered inhibitors, and in the design of multifunctional proteins.

Author Summary

Proteinaceous protease inhibitors are promising molecules for the design of new products in biotechnology and biomedicine. Identification of the protease-binding region is essential in the design of new inhibitors. We have proved that binding regions in the Kunitz-STI family are more diverse than just a canonical conformation and that their conformation and location are related to the plant source of the inhibitors. Our findings will help researchers to identify the protease-binding regions in other inhibitors of this protein family and could help in the design of new inhibitors with the β-trefoil fold.

Keywords: Kunitz, protease inhibitors, canonical inhibitors, noncanonical inhibitors, serine protease.

1. Introduction

Protease inhibitors have proved to be efficient tools to regulate protease activity, and have been used to control proteases of importance in biomedicine and biotechnology.[1–3] Among families of proteinaceous protease inhibitors, the Kunitz-Soybean Trypsin Inhibitor (Kunitz-STI) family is one of the most versatile, with members able to inhibit proteases from different mechanistic classes like serine, aspartic, and cysteine proteases.[4–6] In addition, there are proteins within this family with other functions like lectin[7], water-soluble chlorophyll-binding[8,9], taste-modification[10], α-amylase inhibitors[11,12], and storage.[13] Proteins from the Kunitz-STI family have a β-trefoil fold consisting of 12 β-strands arranged in three structural units with three-fold pseudo-symmetry. Six β-strands form a β-barrel while the other six β strands are arranged like a cap for the β-barrel (Figure 1). Loops can be named based on the β strands to which they are attached (e.g., the β4-β5 loop connects strands β4 and β5). Versatility in the functions of this protein family has been linked to the diversity in sequence and length of the loops.[14] The structural plasticity of loops of the β-trefoil fold, together with multiple functionalities described for Kunitz-STI inhibitors make these proteins an interesting scaffold for inhibitor design and engineering. Identifying protease-binding loops is an essential step toward the design of more potent and selective inhibitors.

Historically, serine protease inhibitors from the Kunitz-STI family have been considered canonical inhibitors with the protease-binding region lying on loop β4-β5. Canonical serine protease inhibitors are proteins that interact with proteases through an exposed loop with a specific backbone conformation, named the “canonical” conformation, and their mechanism of action is called the Laskowski or standard mechanism.[15,16] In this inhibition mechanism, the inhibitor binds to the protease like a substrate and is slowly hydrolyzed by the protease. However, products are not released from the active site, and the peptide bond is resynthesized while hydrolyzed inhibitor is still bound to the active site of the protease.[4,17] This canonical backbone conformation of the protease-binding loop is adopted mainly by positions P3 to P3´ (Schechter & Berger notation)[18], and is present in 19 families of proteinaceous serine protease inhibitors with different structural folds described in the MEROPS database.[19] The canonical conformation of the protease-binding loop is stabilized by some scaffolding amino acids. In the Kunitz-STI inhibitors, there is a conserved asparagine in the N-terminus, which is essential to keep the ends of the hydrolyzed products in the right orientation and distance for the re-ligation reaction.[20]

The Laskowski mechanism is not the only way proteinaceous protease inhibitors can inhibit serine proteases. Some reversible inhibitors of serine proteases bind to the protease active site, blocking access to it but without any peptide bond being hydrolyzed. This type of inhibition mechanism is

known as the steric blockage of the active site.[21–23] The protease-binding region in these inhibitors adopts a conformation called “noncanonical” because is different to the canonical conformation described before, and the inhibitors are named “noncanonical inhibitors”.[16]

The Kunitz-STI inhibitors show a remarkable sequence variability and a large amount of gene duplications, however, these sequences maintain key structural conformations, like protease-binding loops.[24] This observation suggests a complex evolutionary scenario, which leads to the radiation of a protein family capable of inhibiting new targets. In that sense, a simple sequence analysis lacks the power to differentiate relationships between members of the family, and structural comparison could be a better starting point to classify Kunitz-STI.

In this work we analyzed the 3D-structures of all the serine protease inhibitors from the Kunitz-STI family through a combination of structural bioinformatics tools, to characterize the structural plasticity and constraints in the serine protease-binding loops. We have found that inhibitors with a canonical β4-β5 loop have been isolated only from plants from the order Fabales. On the other hand, inhibitors from monocotyledonous plants, seem to use exclusively β5-β6 and β9-β10 loops to interact with serine proteinases. Moreover, the subtilisin-binding conformation of the β5-β6 loop is widely extended among inhibitors from plants of different taxonomies, and it is very similar to the trypsin-binding conformation of the same loop present in some inhibitors, suggesting a potential evolutionary relationship in the inhibition mechanisms used by Kunitz-STI inhibitors against serine proteases of the families S1 and S8 (MEROPS database classification).

These findings are an important contribution to guide researchers in future identification of serine protease binding regions in new or uncharacterized inhibitors. Also, it can be used in the design of new multifunctional inhibitors with the β-trefoil fold targeting different serine proteases through different inhibition mechanisms.

2. Results

2.1 Overview of serine protease inhibitors from Kunitz-STI family

Seventy-one 3D-structures for 23 different serine protease inhibitors belonging to Kunitz-STI family have been deposited in the RCSB PDB database. Despite the fact that the Pfam database reports 170 3D-structures of members of the Kunitz legume family (PF00197), not all of these proteins are serine protease inhibitors or there is no functional data validating them as such. A cross-check with the RCSB PDB database revealed only one Kunitz-STI inhibitor 3D-structure (PDB ID: 1R8O) that is not listed in the Pfam database. Twenty-four (34 %) of the seventy-one PDB 3D-structures were discarded from further analysis because of validation problems such as an R/R_{free} gap higher than

5 %, sequence-engineered inhibitors, poor electron density in the loops of interest, low resolution (>2.85Å) (e.g., 5DSS, 3I2A, 2WBC). Forty-seven 3D-structures corresponding to 23 different inhibitors were selected for this study (Table S1).

Amino acid sequence identities among these 23 inhibitors are in general below 50 %, with a few exceptions (Table S2). The most conserved sequence regions are on β-sheets, with the loops being highly variable in sequence and length (Figure 1). Despite the low sequence identity all these 23 proteins share the same β-trefoil fold, with global RMSD-Cα values from 0.83 Å to 2.09 Å (Table S3). In general, all the inhibitors have twelve β-sheets typical of the fold, while in some cases there are extra short β-sheets, and a variable number of helices.

The taxonomic distribution of the source of the inhibitors reveals that most of them have been isolated from dicotyledonous plants (19), belonging to the orders Fabales (TKI, CTI, EcTI, WCI, ETI, STI, BbCI, BbKI, COTI, $CaTI_2$, MP-4, and DrTI), Vitales (vvMLP), Sapindales (mkMLP), Brassicales (CrataBL and PPI) and Solanales (API-11, API-3, and PSPI). Only four inhibitors are from monocotyledonous plants, two from the order Alismatales (Alocasin and API-A) and two from the order Poales (BASI and RASI).

2.2 Canonical inhibitors in the Kunitz-STI family

Canonical inhibitors are defined as inhibitors with a specific conformation (canonical conformation) in the protease-binding loop. In the Kunitz-STI family we find two types of canonical inhibitors based on the location of the protease-binding loop adopting the canonical conformation: β4-β5 or β9-β10.

2.2.1 Canonical serine protease binding site at β4-β5 loop

Usually, Kunitz-STI protease inhibitors have been described as canonical protease inhibitors with their protease-binding site located at the β4-β5 loop. There are 3D-structures for at least four different inhibitors with a canonical conformation in their protease-binding loop in complex with serine proteases of the S1 family, like trypsin (STI, TKI, EcTI, BbKI), chymotrypsin (BbKI), and kallikrein KLK4 (BbKI). Structural superposition of P4-P4´positions of β4-β5 loop for 9 inhibitors previously classified as canonicals showed RMSD values from 0.24 Å to 1.0 Å (mean RMSD: 0.57 Å, Table S4). However, when the conformations of these 9 β4-β5 loops are compared with 12 other serine protease inhibitors of the family, we obtained RMSD values higher than 1.0 Å in most of the cases (Table S4 and Figure 2A). Backbone torsion angles for inhibitors with canonical β4-β5 loops are very similar in the 3D-structures of protease-inhibitor complex and free inhibitors (Table S5). On the other hand, β4-β5 loops with noncanonical conformations have torsion angle differences of more of 90 degrees in comparison with canonical β4-β5 loops (Table S5). Sequence conservation in positions P4-P4´ of the β4-β5 loop is low among the 21 inhibitors analyzed, no matter whether this

loop adopts a canonical conformation or not (Figure 2B). In canonical β4-β5 loops, the most conserved position is P4´, with only two amino acids (Ile/Leu) present in the nine canonical inhibitors analyzed.

The canonical β4-β5 loops have a well-conserved atomic contact network with a conserved asparagine (Asn13 in STI, from now on Asn*), establishing hydrogen bonds (H-bonds) through its side chain with the backbone of positions P4, P2, P1' and P2' (Figure 2C), and only one main-chain H-bond between its NH and the carbonyl O of the residue at P2'. This H-bond network is conserved in all the canonical β4-β5 loops, despite the high variability of the amino acid sequence (Figure 2). In the inhibitor mkMLP, which has the most similar β4-β5 loop conformation to canonicals, the Asn* has been replaced by an alanine (Ala13), that makes only one H-bond through its NH with the carbonyl O of the Ile68 (P3'). The other amino acids interacting through H-bonds with P4-P4' residues are Leu3 and V11 with Ile69 (P4'), and Ala62 (P4) with Asn65 (P1) (Figure 2C). On the other hand, in inhibitor MP-4, the Asn* has been replaced by an arginine (Arg17) which establishes a H-bond network with several amino acids at P4-P4' positions in loop β4-β5 (Figure 2C). Residues interacting with Arg17 are Ile66 (P4), Arg67(P3), Gly68(P2), Leu70(P1'), and Arg72(P3'), with all these amino acids except Arg72 interacting through their carbonyl oxygen with the NH1 or NH2 of the Arg17. Despite this H-bond network, conformation of the MP-4 β4-β5 loop is different from canonical inhibitors (Figure 2A and Table S4). In inhibitors with noncanonical β4-β5 loops, the chemical groups of the amino acid replacing the Asn* makes the establishment of H-bonds with the β4-β5 loop impossible. For instance, inhibitor PPI has a tyrosine (Tyr16), inhibitors API-3, API-11, PSPI, and $CaTI_2$ have a proline (Pro16), and inhibitor DrTI has a leucine (Leu16) replacing Asn* (Figure 2A). There are other inhibitors where the amino acid replacing the Asn* has a side chain that could make H-bonds with P4-P4' positions, but no interactions were detected. This is the case for inhibitors CrataBL and vvMLP with a threonine (Thr4) and a serine (Ser17), respectively.

We also found that all inhibitors with a canonical β4-β5 loop are from plants from the order Fabales. This is a very interesting finding considering the diversity of sources of Kunitz-STI inhibitors. We would like to note that there are also three inhibitors from plants from the order Fabales with their β4-β5 loops adopting a noncanonical conformation (DrTI, MP-4, and $CaTI_2$), but none of them have an Asn*.

2.2.2 Canonical serine protease-binding site at the β9-β10 loop

The 3D-structure of the double-headed inhibitor API-A in complex with two molecules of trypsin revealed that the inhibitor´s β9-β10 loop adopts a canonical conformation.[25] We compared this conformation with equivalent loops from other inhibitors to know if there are other inhibitors in the family with a canonical serine protease-binding site in their β9-β10 loop. Considering that the two

disulfide bonds in the β9-β10 loop of API-A are important for the stabilization of the canonical conformation, only inhibitors with both disulfide bonds in their β9-β10 loop were used for structural comparison. From the 23 inhibitors analyzed, six, including API-A, have two disulfide bonds at the β9-β10 loop (Alocasin, API-11, API-3, vvMLP, and mkMLP). The RMSD for the P4-P4´ positions shows values from 1.16 Å to 2.86 Å (Table S6). However, when comparison was performed using the P3-P3´ positions, RMSD values drop to 0.46 Å for Alocasin (Table S6). Figure 3A shows that the backbone of these six amino acids is very similar in both inhibitors (API-A and Alocasin), and have similar contact networks, including the two disulfide bonds involving cysteines in P2 and P3´positions (Figure 3). In addition, there is a basic amino acid at position P1 (Arg147) in Alocasin, which is a favorable amino acid for trypsin inhibition. It is worth mentioning that canonical conformation is conserved despite differences in loop length between both inhibitors. Structural superposition of P3-P3´positions of Alocasin on the API-A-trypsin complex structure showed that most of the interactions between API-A and trypsin, will be maintained in a potential complex between Alocasin and trypsin (Figure 3C).

2.3 Noncanonical inhibitors in the Kunitz-STI family

Besides the canonical loops described in previous sections, there is another loop involved in serine-protease binding in some Kunitz-STI inhibitors: the β5-β6 loop. This loop has been identified in at least two different inhibitors as the protease-binding loop for serine proteases from two different peptidase families of the MEROPS database: trypsin (Clan PA, family S1) and subtilisin (Clan SB, family S8). In the inhibitor API-A, its β5-β6 loop adopts a noncanonical conformation to bind the active site of bovine trypsin.[25] The inhibitor BASI inhibits Savinase subtilisin using a steric blockage mechanism using its β5-β6 loop with a noncanonical conformation.[21] We used the conformations of the loop β5-β6 in these two 3D-structures to identify other inhibitors that could interact with trypsin and subtilisin in the same way as API-A or BASI do, respectively.

2.3.1 Inhibitors with a noncanonical trypsin-binding loop

Some of the 23 inhibitors selected for analysis in this study were excluded from the conformational analysis of the β5-β6 loop as a trypsin-binding site, based on the functional data available. For example, the inhibitors BASI and RASI are not able to inhibit trypsin,[26] there is evidence that for the inhibitor CrataBL, its β4-β5 loop is involved in its interaction with trypsin[27], and inhibitors BbCI and BbKI have no cysteines, which are important for stabilization of the loop β5-β6 protease-binding conformation.

Structural comparison of the inhibitor API-A (P5-P2´ positions) trypsin-binding β5-β6 loop with other Kunitz-STI inhibitors, revealed four inhibitors with very similar loop conformations: three potato inhibitors (API-11, API-3 and PSPI) and the inhibitor STI (Table S7). In addition, the values of the

backbone torsion angles of these four inhibitors and API-A confirm the structural similarity between their β5-β6 loops (Table S8). The analysis of the contact network of this loop for these five inhibitors revealed a conserved H-bond between positions P2-P2´, and a disulfide bond between P2´and a cysteine at loop β2-β3 (Figures 1 and 4). In addition to this conserved H-bond, the three potato inhibitors and STI also share H-bonds between positions P7-P5, P1-P3´, and P2´and P5´, while API-A has a H-bond between I88 (P1´) and F115. We found that sequence conservation in those positions used for structural comparison (P5-P2´) is low (Figure 4C), except for positions P2´ (C) and P3 (T/V).

2.3.2 Inhibitors with a noncanonical subtilisin-binding loop

The inhibition mechanism of Savinase subtilisin by inhibitor BASI is very peculiar among subtilisin proteinaceous inhibitors. The inhibitor β5-β6 loop uses a steric blockade mechanism, with the amino acid in the P1 position out of reach from the protease catalytic residues.

Inhibitors RASI and BASI have very similar global 3D-structures, and a sequence identity of 60 % (Tables S2 and S3). Despite the lack of a 3D-structure for a complex between a subtilisin and the inhibitor RASI, the structural superimposition of its β5-β6 loop with the subtilisin-binding loop of BASI showed a very similar conformation (Table S9). Furthermore, most of the atomic contacts for the residues from P5 to P1, including the disulfide bond between P1 and a cysteine in the β2-β3 loop are conserved in both inhibitors. These structural similarities, together with conservation of those BASI amino acids identified as important for Savinase subtilisin inhibition (P5 (Ala), P3 (Thr) and P1 (Cys)),[21] suggest a similar interaction mode and subtilisin-binding loop in these two inhibitors. Structural comparison of conformation of the β5-β6 loop of inhibitor BASI with other Kunitz-STI inhibitors revealed several inhibitors with a similar conformation (Table S9). In the case of inhibitor PPI, the conformation of its β5-β6 loop is very similar (Figure 5A), even though only the cysteine residue at P1 is conserved in both inhibitors. There are other atomic contacts conserved in this loop in BASI and PPI, including a H-bond between amino acids at P7 and P5, in addition to the disulfide bond. Backbone torsion angles also reflects the structural similarity in the β5-β6 loop between BASI and PPI (Table S10). Additionally, we found another seven inhibitors (Alocasin, vvMLP, and mkMLP, API-A, API-3, API-11, and PSPI) with a similar conformation with the β5-β6 loop of inhibitor BASI (Table S9). For 4 of these inhibitors, the potential P1 residue is occupied by a cysteine forming a disulfide bond (PPI, Alocasin, vvMLP, and mkMLP) like in BASI, while in others this cysteine is in P2´ (API-A, API-3, API-11, and PSPI). Despite the two-residues shift of this cysteine, structural superposition of P5-P1 positions of API-A with equipositioned amino acids of inhibitor BASI in the 3D-structure of the complex subtilisin-BASI, showed that the amino acid at P1 (Leu87) in API-A is also pulled away from the active site of the protease but in a different direction than the P1 residue

of BASI (Figure 5A). There are several atomic contacts conserved in the subtilisin binding loop between BASI and these inhibitors with a similar loop β5-β6 conformation, including the H-bonds between residues, P7 and P5, P3 and P1´ (Figure 5B). Regarding the sequence similarity in P5-P1 positions, we have found that position P3 is conserved in seven of the nine inhibitors with a potential subtilisin binding loop, and one inhibitor (vvMLP) has four of the five amino acids conserved in comparison with BASI.

2.4 Phylogenetic patterns of Kunitz-STI sequences

Inhibitors with the canonical β4-β5 loop are clustered together in the amino acid sequence-based phylogenetic tree (Figure 6A), except for mkMLP. IAt the same time, the inhibitors from the order Fabales with a noncanonical β4-β5 loop (DrTI, $CaTI_2$ and MP-4) cluster in the same branch, showing that these proteins join in the tree following the structural similarity (presence/absence of a canonical β4-β5 loop) instead of a species tree pattern (Figure 6B). This observation indicates a certain level of evolutionary information associated with the inhibition mechanism of the proteins in this family. This idea is reinforced by the clustering observed in the rest of the proteins, following a species tree-like distribution (Figure 6B). In the other inhibitors from dicotyledonous plants, a less clear correlation between the topologies of the two trees is observed.

The amino acid sequence-based tree (Figure 6A) shows clearly that inhibitors from monocotyledonous plants (orders Alismatales and Poales) are clustered together. In these inhibitors, the protease-binding loops are β5-β6 and β9-β10, with the canonical β9-β10 loop only present in inhibitors from Alismatales. This pattern is highly congruent with the species tree.

3. Discussion

Protease inhibitors with the β-trefoil fold can inhibit proteases from different mechanistic classes and folds.[14] The versatility and plasticity of this fold make it a very interesting platform for the design of new protease inhibitors. There are at least four different proteinaceous protease inhibitors families with this fold in the MEROPS database: Kunitz-STI (I3), clitocypins (I48), cospins (I66), and macrocypins (I85). Structural and functional studies in clitocypins and macrocypins have shown that several loops, adopting noncanonical conformations, are used to inhibit at least three protease families.[28] In the inhibitors cospin and cnispin the trypsin-binding sites are located in the loops β2-β3 and β11-β12, which demonstrates the plasticity of the fold that can use several loops with different conformations to inhibit proteases.[29] This functional role and conformational plasticity of the loops seem to be patterns in all β-trefoil proteins.[30]

The inhibitors of the Kunitz-STI family are described as canonical inhibitors that inhibit serine peptidases by the Laskowski mechanism. However, we have found that 39 % (9) of the 23 inhibitors

analyzed in this study have a canonical β4-β5 loop, and all of them have been isolated from plants of the order Fabales. Only three inhibitors (DrTI, MP-4, $CaTI_2$) from Fabales do not have a canonical β4-β5 loop, and these three inhibitors inhibit trypsin with inhibition constants 10-1000-fold lower than the values reported for inhibitors with a canonical β4-β5 loop (Table S1).[31–34] Amino acid insertions in the β4-β5 loop of these three inhibitors, and the replacement of the stabilizing asparagine has been proposed as the reason behind the noncanonical conformation of the loop and weak inhibitory activity.[35] Previous studies with mutants of the inhibitor WCI have revealed that replacement of its Asn14 with other amino acids makes the resulting protein more susceptible to proteolytic hydrolysis and to a reduction in its inhibitory potency.[36] While in the nine inhibitors from Fabales with a canonical β4-β5 loop, the Asn* side chain makes H-bonds with the main chain of residues in P4, P2, and P1´, in inhibitors DrTI, $CaTI_2$, and MP-4 only the side chain of the Arg17 in MP-4 interacts through H-bonds with the main chain of residues P4, P3, P2, P1 and P1´. The absence of a stabilizing residue capable of making a specific H-bond network with the protease-binding loop could be argued as the primary reason for the lack of a canonical conformation in inhibitors DrTI and $CaTI_2$. In the case of MP-4 inhibitor, the larger side chain of the Arg17 in comparison with an asparagine seems to induce a noncanonical conformation in the β4-β5 loop of MP-4. This suggests that to adopt the canonical conformation, the size of the amino acid in the position of the Asn* is as important as its capacity to make H-bonds with the β4-β5 loop.

Based in our findings, we can say that in the Kunitz-STI family, inhibitors that use the β4-β5 loop to inhibit serine proteases through the standard mechanism seem to be restricted to those isolated from the order Fabales, although not all the inhibitors from plants of this order can be classified as canonicals. This means that the probability that new inhibitors isolated from plants not belonging to the Fabales order will have a canonical β4-β5 loop is low, at least based on the results of this study.

The presence of a canonical trypsin-binding β9-β10 loop is scarce among the 3D-structures analyzed in this study. Only inhibitors Alocasin and API-A have this conformation and, interestingly, both are from plants from the order Alismatales. Alocasin has been proposed to be classified as independent group inside the Kunitz-STI family together with inhibitors from other plants of the family Araceae.[37] This classification was suggested based on the length of the loop β9-β10 and the presence of the Arg147. The experimental stoichiometry calculated for Alocasin-trypsin interaction suggests the presence of only one trypsin-binding site, thus, considering our findings, we propose that Alocasin uses its canonical β9-β10 loop to inhibit trypsin through the standard mechanism. While canonical conformation in inhibitors from Eudicotyledons is restricted to the β4-β5 loop, in inhibitors from monocotyledons, canonical conformation is only present in the β9-β10 loop.

The conformational similarity between the β5-β6 loop of API and potato inhibitors API-11, API-3, and PSPI, was previously reported by us,[38] but, to the best of our knowledge, there are no previous

reports of the conformational similarity with inhibitor STI β5-β6 loop. This is interesting considering that STI can bind trypsin and chymotrypsin simultaneously, and amino acids M84 and L85 have been proposed as potential P1 and P1´positions in the second binding-site.[39] Our results give structural support to β5-β6 loop as a chymotrypsin-binding site in STI, which could use the same inhibitory mechanism used by API-A to inhibit trypsin. The mechanism used by API-A to inhibit trypsin using this noncanonical loop was not specified in the work publishing the 3D-structure of the trypsin-API-A complex.[25] From our structural analysis, it seems that the hydrolysis of the peptide bond between amino acids in P1 and P1´ seems not to be possible considering the distance (3.5 Å) from the catalytic serine (Ser192) to the carbonyl carbon of the P1 residue (Leu87). In several structures of canonical inhibitors in complex with serine proteases, this distance is from 2.5 to 2.8 Å.[17,40–42] This suggests the possibility that API-A uses a steric blockade mechanism.

The β5-β6 loop is used by subtilisin inhibitor BASI to interact with a serine protease of the subtilisin family. We have found that this subtilisin-binding conformation of the β5-β6 loop is present in several inhibitors of the family, including those where this loop is described as a trypsin-binding site (API-A and potato inhibitors). However, no inhibitors from the order Fabales have this subtilisin-binding conformation in their β5-β6 loop. The presence of a subtilisin-binding loop in Alocasin was suggested in a previous study, but it was considered as a unique structure because of the absence of an arginine in the potential subtilisin-binding site.[37] Nonetheless, inhibition of subtilisin by the Kunitz-STI inhibitors is based on a steric blockade that does not depend on any arginine. Therefore, discarding Alocasin as a subtilisin inhibitor due to the absence of an arginine seems not to be a valid argument for us. The conformational similarity with the subtilisin inhibitor BASI points to the β5-β6 loop in Alocasin being a subtilisin-binding site. In our opinion, identification of potential subtilisin inhibitors in the Kunitz-STI family should be based on the presence of a specific loop conformation and key residue contacts. Following this logic, we found eight inhibitors sharing a conformation in P5-P1 residues similar to that of the BASI subtilisin binding loop. At least one of these inhibitors (PPI) can inhibit Carlsberg subtilisin[43] while there are no reports for the other seven being tested against a subtilisin.

In some inhibitors, this putative subtilisin-binding loop is in the same region identified (API-A) or proposed (API-3, API-11, PSPI) as a trypsin-binding loop. Subtilisin inhibitors BASI and RASI are not able to inhibit trypsin but there is no evidence whether API-A can inhibit subtilisin.[44] The shifting of the cysteine residue from P1 to P2´in the API-A β4-β5 loop in comparison with BASI could have played a role in the transformation from a subtilisin-binding loop into a trypsin-binding loop. In any case, it appears that the β4-β5 loop is a functional noncanonical conserved region used to inhibit serine proteases from the S1 and S8 families, using a common inhibition mechanism.

Another interesting discovery is that the two inhibitors from Alismatales are double-headed inhibitors with the same loops (β5-β6 and β9-β10) as serine-protease binding sites, suggesting a conserved functionality in these regions among inhibitors from Alismatales.

The evolution of Kunitz-STI proteins sampled for this study reveals an interesting evolutionary pattern, highly congruent with the presence/absence and location of canonical and noncanonical protease-binding loops, but also with the species tree expectation (Figure 6B). In fact, the sequence-based tree pattern was unexpected, given the low level of similarity between amino acid sequences within the family members. When multiple sequence alignments were performed without structural guidance, phylograms less similar to the species tree were obtained, also with low bootstrap support values (data not shown). This observation reveals phylogenetic information contained in the proteins' 3D-structures.

More than 4000 sequences of Kunitz-STI have been identified and compiled in the Interpro database [45], many of them with several copies (highly duplicated) in the same species. This radiative pattern of evolution gives the opportunity to diversify the function of paralogues in order to gain the ability to inhibit new proteases. A similar pattern was described in the PIN-FORMED1 (PIN) auxin efflux carriers.[46] This behavior makes complex the phylogenetic reconstruction by loss and gain of copies throughout the species evolution. However, Kunitz-STI members still reconstruct a species tree-like pattern when structural information guides the alignments, despite the high number of copies carried by species genomes and the low sequence similarity.

We are aware that one of the main limitations of this study is the representativity of the structural data available, which do not necessarily describe the evolution of the entire Kunitz-STI family. It could be possible that inhibitors with different protease-binding loops in terms of location and conformation are present in one species. From more than 4000 protein sequences in the Kunitz-STI family, there are experimentally determined 3D-structures for only 23 serine protease inhibitors. The low sequence conservation in the protease-binding region makes it very difficult, if not impossible, to identify these regions using only the amino acid sequence of the loops. The availability of more 3D-structures of protease-inhibitor complexes or free inhibitors, especially for those inhibitors outside the order Fabales, will help to get a clearer picture of the evolution of the inhibitory functionality of this family. The recent release of the AlphaFold 3D-models for almost all Kunitz-STI inhibitors will be of great help in the identification of the protease-binding sites and the evolution this protein family.[47,48] Besides, the evaluation of known and new Kunitz-STI inhibitors against a wider array of serine proteases, specially subtilisin, will allow further refinement of the hypotheses proposed in this work.

4. Methods

4.1 Sequence and structural analysis

Coordinates of 3D-structures of inhibitors from the Kunitz-STI family were retrieved from the RCSB Protein Data Bank.[49] We also used the Pfam database[50] as proxy for all 3D structures available for inhibitors in this protein family.

Initial multiple sequence alignment was obtained with T-coffee[51] and manually edited in Jalview[52] guided by a structural superposition of inhibitor 3D-structures using PDBeFold[53], DALI[54], and PROMALS3D.[55] Percent sequence identity was calculated from pairwise alignments using Jalview. The root mean square deviation of Cα atoms (RMSD-Cα) for global structural alignments was calculated with PDBFold.[53]

To propose the inhibition mechanism for the inhibitors selected, the conformations of the protease-binding loop in 3D-structures of protease-inhibitor complexes were used as reference for each inhibition mechanism described in the Kunitz-STI family: Laskowski mechanism (PDB ID: 1AVW, 4AN6, 3E8L), steric blockade (PDB ID 3BX1), and uncharacterized mechanism (PDB ID 3E8L). The RMSD-Cα from structural comparison of loop conformations was calculated using the LSQ superpose tool from the program WinCoot v0.9.7.[56] Amino acids involved in the protease-inhibitor interaction were used for loop conformation superposition. Through the text, unless explicitly indicated, Schechter & Berger[18] nomenclature is used to indicate amino acid position relative to the scissile peptide bond at P1-P1', not necessarily implying that those residues are part of the protease-inhibitor interface.

Protein atomic contacts were identified using the Protein Contact Atlas [57], while identification of atomic contacts by type were calculated using Arpeggio.[58] In 3D-structures with more than one chain in the asymmetric unit (ASU), contacts were identified for each chain. Atomic contacts were considered conserved when present in all chains of each 3D-structure for a specific protein or a defined group of inhibitors (e.g., canonical inhibitors). Peptide torsion angles were calculated using VADAR.[59]

4.2 Tree reconstruction

The taxonomic tree (species tree) was constructed using the Common Tree Tool of the NCBI Taxonomy Database (https://www.ncbi.nlm.nih.gov/Taxonomy/CommonTree) and edited with iTOL v5.[60]

Modelgenerator was used to calculate a fitted model for the alignment.[61] The protein tree was reconstructed in PhyML 3.3[62] using the model predicted by Modelgenerator, allowing the proportion

of invariable sites and Gamma distribution parameter calculations by the same program. A hundred non-parametric bootstrap resampling supported the tree.

5. Supplementary material description.

Additional data of the structural analyses performed in this work including general information of the inhibitors analyzed, global and local root mean deviations of the structures superpositions, and values of peptide torsion angles.

Acknowledgments

This research was funded by Universidad de Las Américas, Quito, Ecuador, grant number BIO.YGB.22.01. All authors have read and agreed to the published version of the manuscript.

Author Contributions

Yasel Guerra: conceptualization (lead); investigation (lead); methodology (lead), writing original draft (lead); writing-review and editing. **Vinicio Armijos-Jaramillo:** investigation, methodology, writing original draft; writing-review and editing. **Tirso Pons:** Formal analysis, writing-review and editing. **Eduardo Tejera:** Formal analysis, writing-review and editing. **Colin Berry:** conceptualization (supporting), writing-review and editing.

References

1. Agbowuro AA, Huston WM, Gamble AB, Tyndall JDA (2018) Proteases and protease inhibitors in infectious diseases. Med. Res. Rev. 38:1295–1331. Available from: https://onlinelibrary.wiley.com/doi/10.1002/med.21475

2. Turk B (2006) Targeting proteases: successes, failures and future prospects. Nat. Rev. Drug Discov. [Internet] 5:785–799. Available from: https://www.nature.com/articles/nrd2092

3. Dunse KM, Stevens J a, Lay FT, Gaspar YM, Heath RL, Anderson M a (2010) Coexpression of potato type I and II proteinase inhibitors gives cotton plants protection against insect damage in the field. Proc. Natl. Acad. Sci. 107:15011–15015. Available from: https://pnas.org/doi/full/10.1073/pnas.1009241107

4. Laskowski M, Kato I (1980) Protein Inhibitors of Proteinases. Annu. Rev. Biochem. [Internet] 49:593–626. Available from: http://dx.doi.org/10.1146/annurev.bi.49.070180.003113

5. Araújo APU, Hansen D, Vieira DF, Oliveira C de, Santana LA, Beltramini LM, Sampaio CAM, Sampaio MU, Oliva ML V. (2005) Kunitz-type Bauhinia bauhinioides inhibitors devoid of disulfide

bridges: isolation of the cDNAs, heterologous expression and structural studies. Biol. Chem. [Internet] 386:561. Available from: file://www.degruyter.com/view/j/bchm.2005.386.issue-6/bc.2005.066/bc.2005.066.xml

6. Mareš M, Meloun B, Pavlík M, Kostka V, Baudyš M (1989) Primary structure of cathepsin D inhibitor from potatoes and its structure relationship to soybean trypsin inhibitor family. FEBS Lett. [Internet] 251:94–98. Available from: http://doi.wiley.com/10.1016/0014-5793%2889%2981435-8

7. Nascimento CO, Costa RMPB, Araújo RMS, Chaves MEC, Coelho LCBB, Paiva PMG, Teixeira JA, Correia MTS, Carneiro-da-Cunha MG (2008) Optimized extraction of a lectin from Crataeva tapia bark using AOT in isooctane reversed micelles. Process Biochem. [Internet] 43:779–782. Available from: https://linkinghub.elsevier.com/retrieve/pii/S1359511308000895

8. Satoh H, Nakayama K, Okada M (1998) Molecular cloning and functional expression of a water-soluble chlorophyll protein, a putative carrier of chlorophyll molecules in cauliflower. J. Biol. Chem. [Internet] 273:30568–30575. Available from: http://dx.doi.org/10.1074/jbc.273.46.30568

9. Prabahar V, Afriat-Jurnou L, Paluy I, Peleg Y, Noy D (2020) New homologues of Brassicaceae water-soluble chlorophyll proteins shed light on chlorophyll binding, spectral tuning, and molecular evolution. FEBS J. [Internet] 287:991–1004. Available from: https://onlinelibrary.wiley.com/doi/10.1111/febs.15068

10. Theerasilp S, Hitotsuya H, Nakajo S, Nakaya K, Nakamura Y, Kurihara Y (1989) Complete Amino Acid Sequence and Structure Characterization of the Taste-modifying Protein, Miraculin. J. Biol. Chem. [Internet] 264:6655–6659. Available from: https://linkinghub.elsevier.com/retrieve/pii/S0021925818834779

11. Mundy J, Hejgaard J, Svendsen I (1984) Characterization of a bifunctional wheat inhibitor of endogenous α-amylase and subtilisin. FEBS Lett. 167:210–214. Available from: http://doi.wiley.com/10.1016/0014-5793%2884%2980128-3

12. Nielsen PK, Bønsager BC, Fukuda K, Svensson B (2004) Barley α-amylase/subtilisin inhibitor: structure, biophysics and protein engineering. Biochim. Biophys. Acta - Proteins Proteomics 1696:157–164. Available from: http://www.sciencedirect.com/science/article/pii/S1570963903003261

13. Kortt AA, Strike PM, Jersey J (1989) Amino acid sequence of a crystalline seed albumin (winged bean albumin-1) from Psophocarpus tetragonolobus (L.) DC. Sequence similarity with Kunitz-type seed inhibitors and 7S storage globulins. Eur. J. Biochem. 181:403–408. Available from: https://onlinelibrary.wiley.com/doi/10.1111/j.1432-1033.1989.tb14739.x

14. Renko M, Sabotič J, Turk D (2012) β-Trefoil inhibitors – from the work of Kunitz onward. bchm 393:1043–1054. Available from: https://www.degruyter.com/document/doi/10.1515/hsz-2012-0159/html

15. Bode W, Huber R (1992) Natural protein proteinase inhibitors and their interaction with proteinases. Eur. J. Biochem. 204:433–451. Available from: https://onlinelibrary.wiley.com/doi/10.1111/j.1432-1033.1992.tb16654.x

16. Krowarsch D, Cierpicki T, Jelen F, Otlewski J (2003) Canonical protein inhibitors of serine proteases. Cell. Mol. Life Sci. C. 60:2427–2444. Available from: https://link.springer.com/10.1007/s00018-003-3120-x

17. Zakharova E, Horvath MP, Goldenberg DP (2009) Structure of a serine protease poised to resynthesize a peptide bond. Proc. Natl. Acad. Sci. 106:11034–11039. Available from: https://pnas.org/doi/full/10.1073/pnas.0902463106

18. Schechter I, Berger A (1967) On the size of the active site in proteases. I. Papain. Biochem. Biophys. Res. Commun. 27:157–162. Available from: http://www.sciencedirect.com/science/article/pii/S0006291X6780055X

19. Rawlings ND, Barrett AJ, Thomas PD, Huang X, Bateman A, Finn RD (2018) The MEROPS database of proteolytic enzymes, their substrates and inhibitors in 2017 and a comparison with peptidases in the PANTHER database. Nucleic Acids Res. 46:D624–D632. Available from: http://academic.oup.com/nar/article/46/D1/D624/4626772

20. Dasgupta J, Khamrui S, Dattagupta JK, Sen U (2006) Spacer Asn Determines the Fate of Kunitz (STI) Inhibitors, as Revealed by Structural and Biochemical Studies on WCI Mutants. Biochemistry 45:6783–6792. Available from: https://pubs.acs.org/doi/10.1021/bi060374q

21. Micheelsen PO, Vévodová J, De Maria L, Østergaard PR, Friis EP, Wilson K, Skjøt M (2008) Structural and Mutational Analyses of the Interaction between the Barley α-Amylase/Subtilisin Inhibitor and the Subtilisin Savinase Reveal a Novel Mode of Inhibition. J. Mol. Biol. 380:681–690. Available from: https://linkinghub.elsevier.com/retrieve/pii/S0022283608006001

22. Rydel TJ, Tulinsky A, Bode W, Huber R (1991) Refined structure of the Hirudin-thrombin complex. J. Mol. Biol. 221:583–601. Available from: https://linkinghub.elsevier.com/retrieve/pii/0022283691800745

23. Richardson JL, Fuentes-Prior P, Sadler JE, Huber R, Bode W (2002) Characterization of the Residues Involved in the Human α-Thrombin−Haemadin Complex: An Exosite II-Binding Inhibitor. Biochemistry 41:2535–2542. Available from: https://doi.org/10.1021/bi011605q

24. Speranskaya AS, Krinitsina AA, Kudryavtseva A V, Poltronieri P, Santino A, Oparina NY, Dmitriev AA, Belenikin MS, Guseva MA, Shevelev AB (2012) Impact of recombination on polymorphism of genes encoding Kunitz-type protease inhibitors in the genus Solanum. Biochimie 94:1687–1696. Available from: http://www.sciencedirect.com/science/article/pii/S0300908412001034

25. 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. Available from: https://linkinghub.elsevier.com/retrieve/pii/S0021925820384982

26. Yoshikawa M, Iwasaki T, Fujii M, Oogaki M (1976) Isolation and Some Properties of a Subtilisin Inhibitor from Barley. J. Biochem. 79:765–773. Available from: https://academic.oup.com/jb/article/782578/Isolation

27. Ferreira R da S, Zhou D, Ferreira JG, Silva MCC, Silva-Lucca RA, Mentele R, Paredes-Gamero EJ, Bertolin TC, dos Santos Correia MT, Paiva PMG, et al. (2013) Crystal Structure of Crataeva tapia Bark Protein (CrataBL) and Its Effect in Human Prostate Cancer Cell Lines Pérez-Payá E, editor. PLoS One 8:e64426. Available from: https://dx.plos.org/10.1371/journal.pone.0064426

28. 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. Available from: https://linkinghub.elsevier.com/retrieve/pii/S0021925820661178

29. Avanzo Caglič P, Renko M, Turk D, Kos J, Sabotič J (2014) Fungal β-trefoil trypsin inhibitors cnispin and cospin demonstrate the plasticity of the β-trefoil fold. Biochim. Biophys. Acta - Proteins Proteomics 1844:1749–1756. Available from: https://linkinghub.elsevier.com/retrieve/pii/S1570963914001721

30. Blaber M (2022) Variable and Conserved Regions of Secondary Structure in the β-Trefoil Fold: Structure Versus Function. Front. Mol. Biosci. 9:1–11. Available from: https://www.frontiersin.org/articles/10.3389/fmolb.2022.889943/full

31. Pando SC, Oliva ML., Sampaio CA., Di Ciero L, Novello JC, Marangoni S (2001) Primary sequence determination of a Kunitz inhibitor isolated from Delonix regia seeds. Phytochemistry 57:625–631. Available from: https://linkinghub.elsevier.com/retrieve/pii/S0031942201000802

32. Kumar A, Kaur H, Jain A, Nair DT, Salunke DM (2018) Docking, thermodynamics and molecular dynamics (MD) studies of a non-canonical protease inhibitor, MP-4, from Mucuna pruriens. Sci. Rep. 8:689. Available from: http://www.nature.com/articles/s41598-017-18733-9

33. Jain A, Kumar A, Shikhi M, Kumar A, Nair DT, Salunke DM (2020) The structure of MP-4 from Mucuna pruriens at 2.22 Å resolution. Acta Crystallogr. Sect. F Struct. Biol. Commun. 76:47–57. Available from: http://scripts.iucr.org/cgi-bin/paper?S2053230X20000199

34. Bendre AD, Suresh CG, Shanmugam D, Ramasamy S (2019) Structural insights into the unique inhibitory mechanism of Kunitz type trypsin inhibitor from Cicer arietinum L. J. Biomol. Struct. Dyn. 37:2669–2677. Available from: https://www.tandfonline.com/doi/full/10.1080/07391102.2018.1494633

35. Krauchenco S, Pando SC, Marangoni S, Polikarpov I (2003) Crystal structure of the Kunitz (STI)-type inhibitor from Delonix regia seeds. Biochem. Biophys. Res. Commun. 312:1303–1308. Available from: https://linkinghub.elsevier.com/retrieve/pii/S0006291X03024173

36. Dasgupta J, Khamrui S, Dattagupta JK, Sen U (2006) Spacer Asn Determines the Fate of Kunitz (STI) Inhibitors, as Revealed by Structural and Biochemical Studies on WCI Mutants. Biochemistry 45:6783–6792. Available from: https://pubs.acs.org/doi/10.1021/bi060374q

37. Vajravijayan S, Pletnev S, Pletnev VZ, Nandhagopal N, Gunasekaran K (2018) Crystal structure of a novel Kunitz type inhibitor, alocasin with anti- Aedes aegypti activity targeting midgut proteases. Pest Manag. Sci. 74:2761–2772. Available from: https://onlinelibrary.wiley.com/doi/10.1002/ps.5063

38. Guerra Y, Valiente PA, Pons T, Berry C, Rudiño-Piñera E (2016) 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? J. Struct. Biol. 195:259–271. Available from: https://linkinghub.elsevier.com/retrieve/pii/S1047847716301204

39. Bösterling B, Quast U (1981) Soybean trypsin inhibitor (Kunitz) is doubleheaded. Biochim. Biophys. Acta - Enzymol. 657:58–72. Available from: https://linkinghub.elsevier.com/retrieve/pii/0005274481901303

40. Patil DN, Chaudhary A, Sharma AK, 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. Available from: https://onlinelibrary.wiley.com/doi/10.1111/febs.12042

41. Song HK, Suh SW (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. Available from: https://linkinghub.elsevier.com/retrieve/pii/S0022283697914698

42. Zhou D, Lobo YA, Batista IFC, Marques-Porto R, Gustchina A, Oliva ML V., Wlodawer A

(2013) Crystal Structures of a Plant Trypsin Inhibitor from Enterolobium contortisiliquum (EcTI) and of Its Complex with Bovine Trypsin Kobe B, editor. PLoS One 8:e62252. Available from: https://dx.plos.org/10.1371/journal.pone.0062252

43. 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. Available from: https://linkinghub.elsevier.com/retrieve/pii/S002192582068468X

44. Yamagata H, Kunimatsu K, Kamasaka H, Kuramoto T, Iwasaki T (1998) Rice Bifunctional α-Amylase/Subtilisin Inhibitor: Characterization, Localization, and Changes in Developing and Germinating Seeds. Biosci. Biotechnol. Biochem. 62:978–985. Available from: https://academic.oup.com/bbb/article/62/5/978-985/5947136

45. Blum M, Chang H-Y, Chuguransky S, Grego T, Kandasaamy S, Mitchell A, Nuka G, Paysan-Lafosse T, Qureshi M, Raj S, et al. (2021) The InterPro protein families and domains database: 20 years on. Nucleic Acids Res. [Internet] 49:D344–D354. Available from: https://doi.org/10.1093/nar/gkaa977

46. Bennett T, Brockington SF, Rothfels C, Graham SW, Stevenson D, Kutchan T, Rolf M, Thomas P, Wong GK-S, Leyser O, et al. (2014) Paralogous Radiations of PIN Proteins with Multiple Origins of Noncanonical PIN Structure. Mol. Biol. Evol. 31:2042–2060. Available from: https://academic.oup.com/mbe/article-lookup/doi/10.1093/molbev/msu147

47. Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, Tunyasuvunakool K, Bates R, Žídek A, Potapenko A, et al. (2021) Highly accurate protein structure prediction with AlphaFold. Nature 596:583–589. Available from: https://doi.org/10.1038/s41586-021-03819-2

48. Varadi M, Anyango S, Deshpande M, Nair S, Natassia C, Yordanova G, Yuan D, Stroe O, Wood G, Laydon A, et al. (2022) AlphaFold Protein Structure Database: massively expanding the structural coverage of protein-sequence space with high-accuracy models. Nucleic Acids Res. 50:D439–D444. Available from: https://doi.org/10.1093/nar/gkab1061

49. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE (2000) The Protein Data Bank. Nucleic Acids Res. 28:235–242. Available from: https://doi.org/10.1093/nar/28.1.235

50. Mistry J, Chuguransky S, Williams L, Qureshi M, Salazar GA, Sonnhammer ELL, Tosatto SCE, Paladin L, Raj S, Richardson LJ, et al. (2021) Pfam: The protein families database in 2021. Nucleic Acids Res. 49:D412–D419. Available from: https://academic.oup.com/nar/article/49/D1/D412/5943818

51. 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. Available from: http://nar.oxfordjournals.org/content/39/suppl_2/W13.abstract

52. Waterhouse AM, Procter JB, Martin DMA, Clamp M, Barton GJ (2009) Jalview Version 2--a multiple sequence alignment editor and analysis workbench. Bioinformatics 25:1189–1191. Available from: https://academic.oup.com/bioinformatics/article-lookup/doi/10.1093/bioinformatics/btp033

53. 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. Available from: http://dx.doi.org/10.1107/S0907444904026460

54. Holm L (2019) Benchmarking fold detection by DaliLite v.5 Elofsson A, editor. Bioinformatics 35:5326–5327. Available from: https://academic.oup.com/bioinformatics/article/35/24/5326/5526869

55. Pei J, Kim B-H, Grishin N V. (2008) PROMALS3D: a tool for multiple protein sequence and structure alignments. Nucleic Acids Res. 36:2295–2300. Available from: https://academic.oup.com/nar/article-lookup/doi/10.1093/nar/gkn072

56. Emsley P, Lohkamp B, Scott WG, Cowtan K (2010) Features and development of Coot. Acta Crystallogr. Sect. D Biol. Crystallogr. 6:486–501. Available from: https://scripts.iucr.org/cgi-bin/paper?S0907444910007493

57. Kayikci M, Venkatakrishnan AJ, Scott-Brown J, Ravarani CNJ, Flock T, Babu MM (2018) Visualization and analysis of non-covalent contacts using the Protein Contacts Atlas. Nat. Struct. Mol. Biol. 25:185–194. Available from: http://www.nature.com/articles/s41594-017-0019-z

58. Jubb HC, Higueruelo AP, Ochoa-Montaño B, Pitt WR, Ascher DB, Blundell TL (2017) Arpeggio: A Web Server for Calculating and Visualising Interatomic Interactions in Protein Structures. J. Mol. Biol. 429:365–371. Available from: http://www.sciencedirect.com/science/article/pii/S0022283616305332

59. Willard L, Ranjan A, Zhang H, Monzavi H, Boyko RF, Sykes BD, Wishart DS (2003) VADAR: a web server for quantitative evaluation of protein structure quality. Nucleic Acids Res. 31:3316–3319. Available from: https://doi.org/10.1093/nar/gkg565

60. Letunic I, Bork P (2021) Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. Nucleic Acids Res.] 49:W293–W296. Available from:

https://academic.oup.com/nar/article/49/W1/W293/6246398

61. Keane TM, Creevey CJ, Pentony MM, Naughton TJ, Mclnerney JO (2006) Assessment of methods for amino acid matrix selection and their use on empirical data shows that ad hoc assumptions for choice of matrix are not justified. BMC Evol. Biol. 6:29. Available from: https://doi.org/10.1186/1471-2148-6-29

62. Guindon S, Dufayard J-F, Lefort V, Anisimova M, Hordijk W, Gascuel O (2010) New Algorithms and Methods to Estimate Maximum-Likelihood Phylogenies: Assessing the Performance of PhyML 3.0. Syst. Biol. ~~[Internet]~~ 59:307–321. Available from: https://doi.org/10.1093/sysbio/syq010

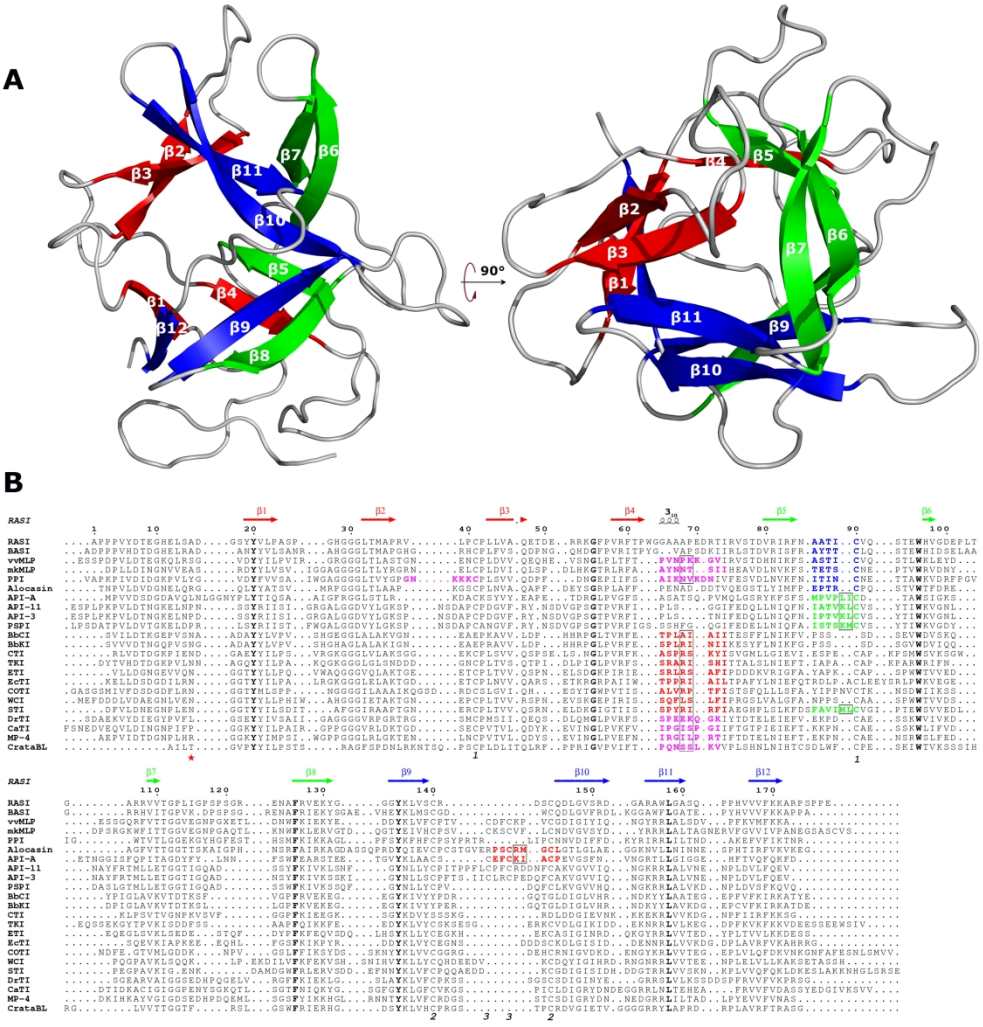

Figure 1. Features of Kunitz-STI inhibitors (A) General fold of the family, illustrated with inhibitor BbCI (PDB ID: 2GZB). β-sheets are colored according to the threefold pseudo-symmetry subdomains present in the 3D-structures. (B) Multiple sequence alignment of Kunitz-STI serine protease inhibitors with 3D-dimensional structures available. Red star: Position of the Asn* that stabilizes the canonical conformation of β4-β5 loop; Bold and blue: Positions P5-P1 in the noncanonical subtilisin-binding β5-β6 loop; Bold and green: Positions P5-P2´ in the noncanonical trypsin-binding β5-β6 loop; Bold and red: Positions P4-P4' in canonical trypsin-binding loops; Bold and purple: Proposed P4-P4´ positions in noncanonical β2-β3 and β4-β5 loops. Conserved amino acids are represented in bold and black. Black boxes highlight the P1-P1´residues in trypsin/chymotrypsin-binding loops. Black numbers at the bottom show the disulfide bond topology. The figure was made with ESPript 3.0 (https://espript.ibcp.fr/ESPript/ESPript/).

299x312mm (300 x 300 DPI)

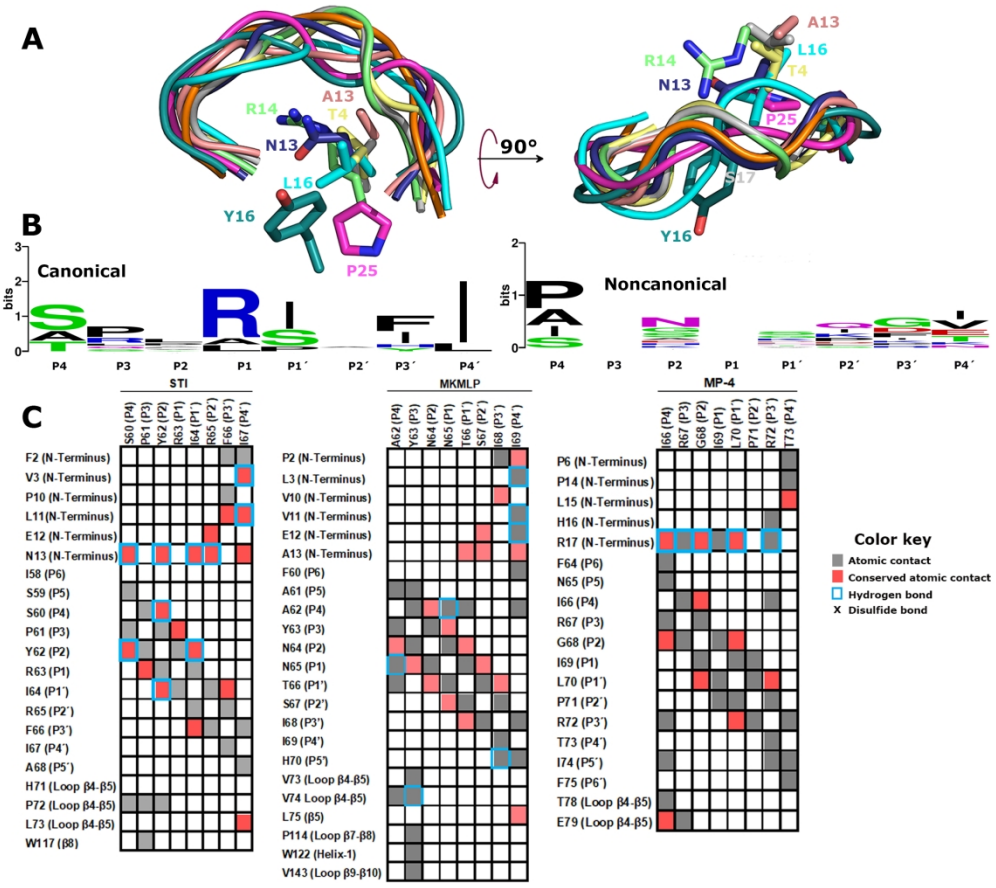

Figure 2. Structural features of loop β4-β5 in Kunitz-STI inhibitors. (A) Structural superposition of some noncanonical conformations with the canonical conformation of loop β4-β5 in inhibitor STI; STI (PDB ID: 1AVW, blue), DrTI (PDB ID: 1R8N, cyan), CaTI2 (PDB ID: 5XOZ, magenta), MP-4 (PDB ID: 6JBP, orange), mkMLP (PDB ID: 3ZC9, salmon), vvMLP (PDB ID: 5YH4, grey), CrataBL (PDB ID: 4IHZ, yellow), PPI (PDB ID: 3S8K, dark green), Alocasin (PDB ID: 5YCZ, light green). The side chain of Asn13 of STI and equivalent amino acids in the other inhibitors are shown in sticks. (B) Sequence conservation in canonical and noncanonical β4-β5 loops. (C) Contact matrices of the β4-β5 loop of inhibitor STI, mkMLP and MP-4.

215x189mm (300 x 300 DPI)

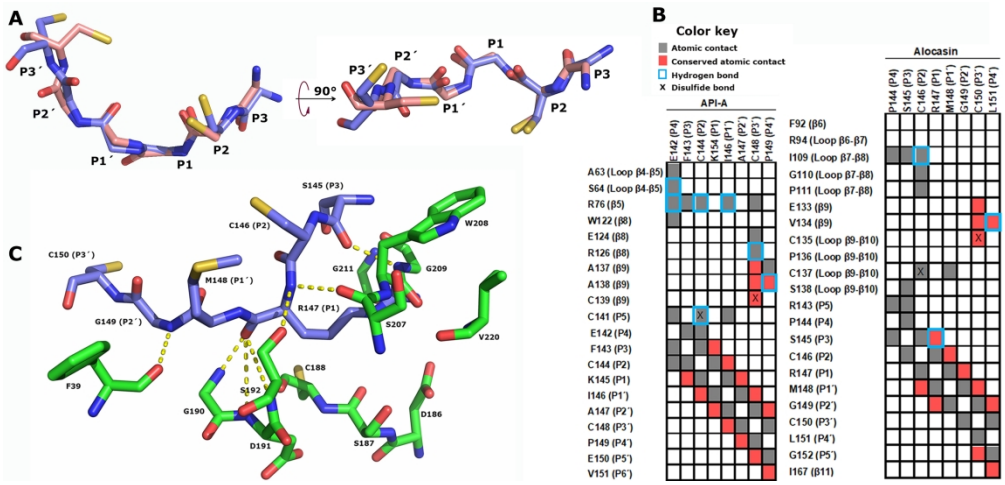

Figure 3. Structural features of canonical loop β9-β10 in Kunitz-STI inhibitors. (A) Structural superposition of canonical β9-β10 loops of inhibitors API-A (3E8L, salmon) and Alocasin (5YCZ, blue). (B) Contact matrix of the canonical β9-β10 loops of inhibitors API-A and Alocasin. (C) Model of a potential trypsin-Alocasin complex with the canonical loop β9-β10 as the trypsin-binding region. Alocasin is shown in blue and interacting residues from trypsin shown in green.

279x133mm (300 x 300 DPI)

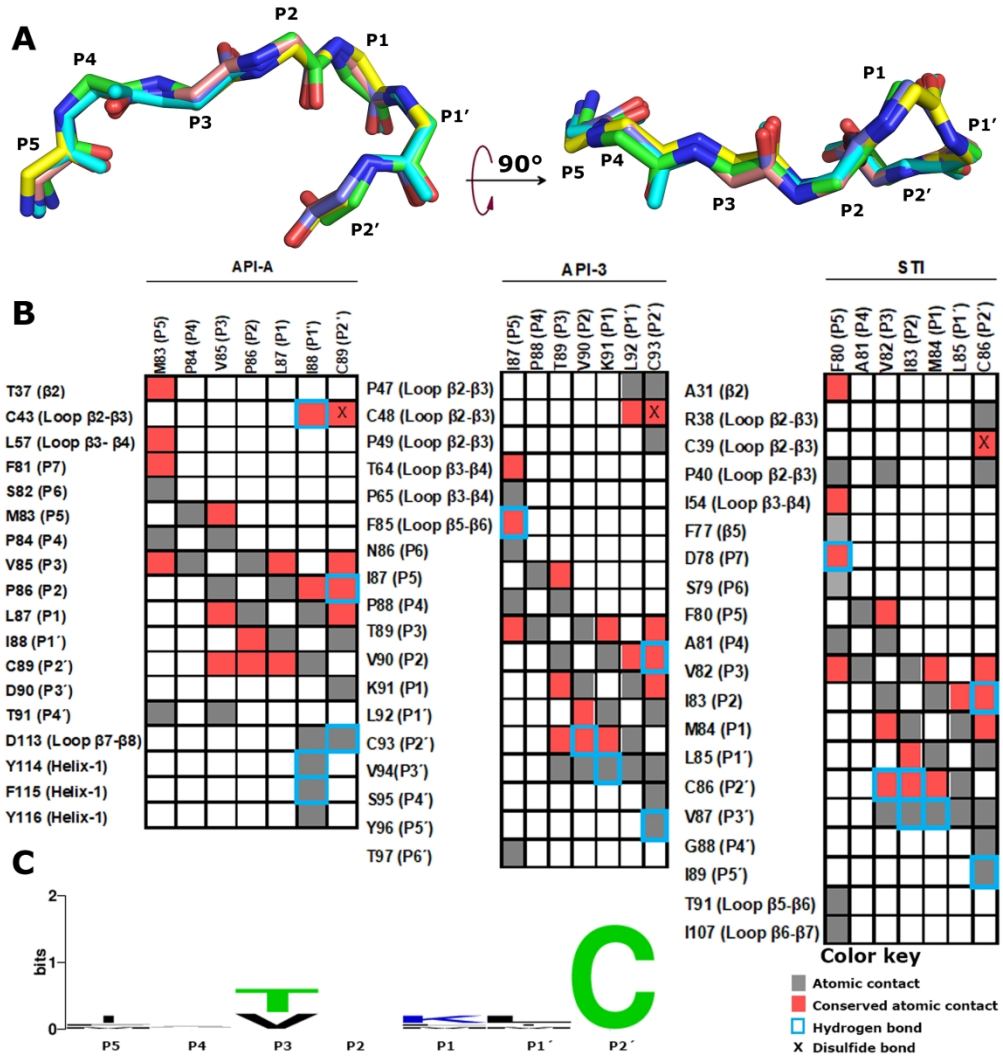

Figure 4. Structural features of noncanonical loop β5-β6 as a trypsin-binding region. (A) Structural superposition of β5-β6 loops of inhibitors API-A (3E8L, green), API-3 (5FZU, yellow), API-11 (5DZU, salmon), PSPI (3TC2, blue), and STI (1AVW, cyan). (B) Contact matrix of the trypsin-binding β4-β5 loops of inhibitors API, API-3, and STI. (C) Sequence conservation among inhibitors with trypsin-binding β4-β5 loops.

215x226mm (300 x 300 DPI)

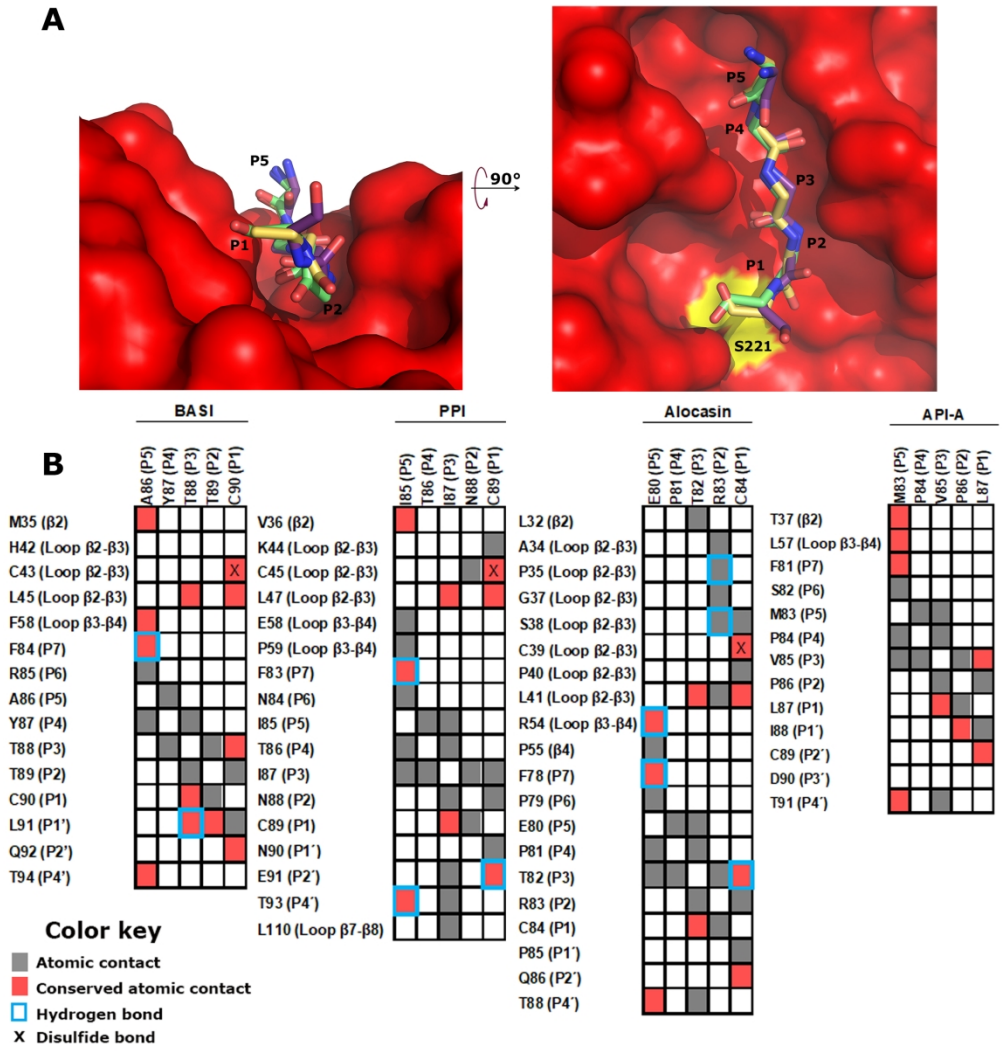

Figure 5. Structural features of noncanonical loop β5-β6 as a subtilisin-binding region. (A) Structural superposition of noncanonical β5-β6 loops of inhibitors PPI (3S8K, yellow), and API-A (3E8L, purple) with inhibitor BASI (3BX1, green) in complex with Savinase subtilisin (red). (B) Contact matrix of the subtilisin-binding β4-β5 loops of inhibitors BASI, PPI, Alocasin, and API-A.

228x240mm (300 x 300 DPI)

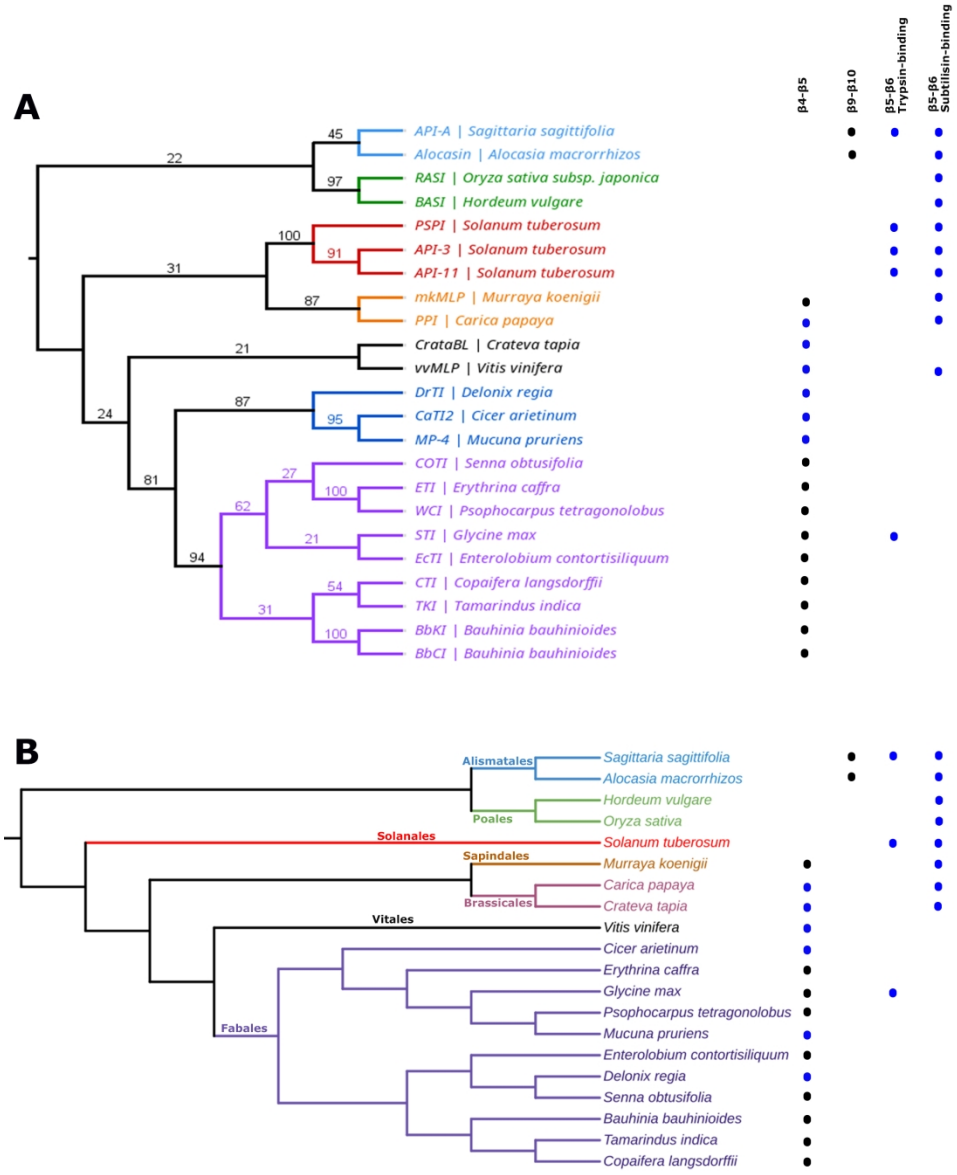

Figure 6. Relationship among Kunitz-STI inhibitors and plant sources. (A) The sequence-based tree of the 23 inhibitors analyzed in this work. (B) Species tree of the inhibitors analyzed. Black dots indicate the canonical conformation. Blue dots indicate the presence of a noncanonical conformation. Branches are colored according the taxonomic order they belong to (as indicated in part B). The NCBI common taxonomy tree does not provide bootstrap values for the tree branches.

215x272mm (300 x 300 DPI)

Table S1. Structural and functional data for the serine protease inhibitors analyzed in this study.

Common name[a]	UniProt accession	Organism	PDB ID[b]	Resolution (Å)	Complex Structure	K_i (M)	Method (K_i)	Serine proteases inhibited	References
RASI	P29421	*Oryza sativa subsp. japonica* (Rice)	2QN4	1.8	No	-	-	S	1
BASI	P07596	*Hordeum vulgare* (Barley)	3BX1/1AVA	1.85/1.9	Yes	4.5×10^{-9}	Enzymatic	S	2
API-A	Q7M1P4	*Sagittaria sagittifolia* (Arrowhead)	3E8L	2.48	Yes	-	-	T	3
Alocasin	P35812	*Alocasia macrorrhizos* (Giant taro)	5YCZ	2.5	No	1.41×10^{-9}	Enzymatic	T	4
API-11	P16348	*Solanum tuberosum* (Potato)	5DZU	2.12	No	8.9×10^{-9}	Enzymatic	T	5,6
API-3*	M1AKE5		5FZU/5FNW/5FNX/5G00/5FZY/5FZZ	2.43/2.45/2.65/2.5/2.47/2.55	No	4.3×10^{-8}	Enzymatic	T	7
PSPI	Q8S380		3TC2	1.6	No	1.62×10^{-9}	Enzymatic	T	8
vvMLP	A5BCQ0	*Vitis vinifera* (Grape)	5YH4	1.3	No	1.37×10^{-5}	Enzymatic	T	9
mkMLP	D2YW43	*Murraya koenigii* (Curry leaf tree)	3ZC8/3ZC9/3IIR	2.24/2.24	No	7×10^{-9}	Enzymatic	T	10
CrataBL	U3KRG0	*Crateva tapia* (Garlic-pear tree)	4IHZ/4II0	1.5/1.75	No	4.3×10^{-5}	Enzymatic	T	11
PPI	P80691	*Carica papaya* (Papaya)	3S8K/3S8J	1.7/2.6	No	3.4×10^{-13}/ 8.22×10^{-8}	SPR	T/S	12
BbCI	P83051	*Bauhinia bauhinioides* (Perlebia bauhinoides)	2GZB	1.70	No	5.3×10^{-9}	Enzymatic	NE	13
BbKI	P83052		4ZOT	1.40/1.87/2.0	No	2.8×10^{-8} / 2×10^{-9}	Enzymatic	T/PK	13,14
	Q6VEQ7		2GO2/6DWH/7JOD/7JOS/7JOE	1.87/2.0/1.33/2.1/2.6	Yes	2×10^{-10}/3.33×10^{-8}/4.5×10^{-11}	Enzymatic	T/C/ hKLK4	15,16
CTI[c]	-	*Copaifera langsdorffii* (Diesel tree)	1R8O	1.83	No	1.2×10^{-9}	Enzymatic	T	17
TKI	F4ZZG4	*Tamarindus indica* (Tamarind)	4AN6/4AN7	1.94/2.23	Yes	3.2×10^{-9}	Enzymatic	T	18
DrTI	P83667	*Delonix regia* (Royal poinciana)	1R8N	1.75	No	2.19×10^{-8} / 5.25×10^{-9}	Enzymatic	T/PK	19,20
EcTI	P86451	*Enterolobium contortisiliquum* (Pacara earpod tree)	4J2K/4J2Y	1.75/2.0	Yes	8.8×10^{-10}	Enzymatic	T	21
CaTI2	Q9M3Z7	*Cicer arietinum* (Chickpea)	5XOZ	2.8	No	6.31×10^{-7}	Thermophoresis	T	22
WCI	P10822	*Psophocarpus tetragonolobus* (Winged bean)	1EYL/1XG6/2QYI/3VEQ	1.9/2.15/2.6/2.2.25	No	1.14×10^{-9}	Enzymatic	C	23
STI	P01070	*Glycine max* (Soybean)	1AVW/1AVU/1AVX/1BA7/6O1F	1.75/2.3/1.9/2.5/2.15	Yes	1.29×10^{-9}	Enzymatic	T	21
ETI	P09943	*Erythrina caffra* (Kaffir tree)	1TIE	2.5	No	4.3×10^{-10}	Enzymatic	T	24,25
COTI	A0A097P6E1	*Senna obtusifolia* (sicklepod)	6KV2	2.0	No	-	-	T	26
MP-4	A0A371E4L6	*Mucuna pruriens* (Velvet bean)	6JBP	2.22	No	2.11×10^{-6} / 2.59×10^{-6}	SPR	T/C	27

[a]Common name used in publications or UniProt.

[b]PDB entries used in most of the analyses and protease-inhibitor complex when available.

[c]No entry was found in UniProt. The name used in primary citation is indicated

*Previously published as E3Ad

T: Trypsin
NE: Neutrophil Elastase
PK: Plasma Kallikrein
C: Chymotrypsin
hKLK4: Human kallikrein related peptidase 4
S: Subtilisins
SPR: Surface Plasmon Resonance

References

1. Yamagata H, Kunimatsu K, Kamasaka H, Kuramoto T, Iwasaki T (1998) Rice bifunctional alpha-amylase/subtilisin inhibitor: characterization, localization, and changes in developing and germinating seeds. Biosci. Biotechnol. Biochem. 62:978–985.
2. Micheelsen PO, Vévodová J, De Maria L, Østergaard PR, Friis EP, Wilson K, Skjøt M (2008) Structural and Mutational Analyses of the Interaction between the Barley α-Amylase/Subtilisin Inhibitor and the Subtilisin Savinase Reveal a Novel Mode of Inhibition. J. Mol. Biol. 380:681–690.
3. Bao R, Zhou CZ, Jiang C, Lin SX, Chi CW, 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.
4. Vajravijayan S, Pletnev S, Pletnev VZ, Nandhagopal N, Gunasekaran K (2018) Crystal structure of a novel Kunitz type inhibitor, alocasin with anti-Aedes aegypti activity targeting midgut proteases. Pest Manag. Sci. 74:2761–2772.
5. Guo J, Erskine PT, Coker AR, Wood SP, Cooper JB (2015) Structure of a Kunitz-type potato cathepsin D inhibitor. J. Struct. Biol. [Internet] 192:554–560. Available from: http://www.sciencedirect.com/science/article/pii/S1047847715300976
6. Keilová H. and Tomášek V (1976) Isolation and properties of Cathepsin D inhibitor from potatoes. Collect Czech Chem Commun 41:489–497.
7. Guerra Y, Valiente PA, Pons T, Berry C, Rudiño-Piñera E (2016) 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? J. Struct. Biol. [Internet] 195:259–271. Available from: http://dx.doi.org/10.1016/j.jsb.2016.06.014
8. Valueva T a., Revina T a., Mosolov V V., Mentele R (2000) Primary structure of potato Kunitz-type serine proteinase inhibitor. Biol. Chem. 381:1215–1221.
9. Ohkura S, Hori M, Saitoh K, Okuzawa T, Okamoto I (2018) BBA - Proteins and Proteomics Structural and functional analysis of miraculin-like protein from Vitis vinifera. 1866:1125–1130.
10. Shee C, Sharma A (2007) Purification and characterization of a trypsin inhibitor from seeds of Murraya koenigii. J. Enzyme Inhib. Med. Chem. 22:115–120.
11. Dias GB, Gomes VM, Pereira UZ, Ribeiro SFF, Carvalho AO, Rodrigues R, MacHado OLT, Fernandes KVS, Ferreira ATS, Perales J, et al. (2013) Isolation, characterization and antifungal activity of proteinase inhibitors from Capsicum chinense Jacq. seeds. Protein J. 32:15–26.
12. 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.
13. Zhou D, Hansen D, Shabalin IG, Gustchina A, Vieira DF, de Brito M V, Araújo APU, Oliva ML V, Wlodawer A (2015) Structure of BbKI, a disulfide-free plasma kallikrein inhibitor. Acta Crystallogr. Sect. F [Internet] 71:1055–1062. Available from: http://dx.doi.org/10.1107/S2053230X15011127
14. Araujo APU, Hansen D, F VD, De OC, A SL, M BL, M SCA, U SM, V OML (2005) Kunitz-type Bauhinia bauhinioides inhibitors devoid of disulfide bridges: isolation of the cDNAs, heterologous expression and structural studies. Biol. Chem. [Internet] 386:561. Available from: file://www.degruyter.com/view/j/bchm.2005.386.issue-6/bc.2005.066/bc.2005.066.xml
15. Li M, Srp J, Mares M, Wlodawer A, Gustchina A (2021) Structural studies of complexes of kallikrein 4 with wild-type and mutated forms of the Kunitz-type inhibitor BbKI. Acta Crystallogr. Sect. D Struct. Biol. 77:1084–1098.
16. Li M, Srp J, Gustchina A, Dauter Z, Mares M, Wlodawer A (2019) Crystal structures of the complex of a kallikrein inhibitor from Bauhinia bauhinioides with trypsin and modeling of kallikrein complexes. Acta Crystallogr. Sect. D [Internet] 75:56–69. Available from: https://doi.org/10.1107/S2059798318016492
17. Silva JA, Macedo MLR, Novello JC, Marangoni S (2001) Biochemical characterization and N-terminal sequences of two new trypsin inhibitors from Copaifera langsdorffii seeds. J. Protein Chem. 20:1–7.
18. Patil DN, Chaudhary A, Sharma AK, 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.
19. Pando SC, Oliva MLV, Sampaio CAM, Di Ciero L, Novello JC, Marangoni S (2001) Primary sequence determination of a Kunitz inhibitor isolated from Delonix regia seeds. Phytochemistry 57:625–631.
20. Krauchenco S, Pando SC, Marangoni S, Polikarpov I (2003) Crystal structure of the Kunitz (STI)-type inhibitor from Delonix regia seeds. Biochem. Biophys.

Res. Commun. 312:1303–1308.
21. Zhou D, Lobo YA, Batista IFC, Marques-Porto R, Gustchina A, Oliva ML V., Wlodawer A (2013) Crystal Structures of a Plant Trypsin Inhibitor from Enterolobium contortisiliquum (EcTI) and of Its Complex with Bovine Trypsin Kobe B, editor. PLoS One [Internet] 8:e62252. Available from: https://dx.plos.org/10.1371/journal.pone.0062252
22. Bendre AD, Suresh CG, Shanmugam D, Ramasamy S (2019) Structural insights into the unique inhibitory mechanism of Kunitz type trypsin inhibitor from Cicer arietinum L. J. Biomol. Struct. Dyn. 37:2669–2677.
23. Kortt AA (1981) Specificity and stability of the chymotrypsin inhibitor from winged bean seed (Psophocarpus tetragonolobus (L) Dc.). BBA - Enzymol. 657:212–221.
24. Lehle K, Wrba A, Jaenicke R (1994) Erythrina caffra trypsin inhibitor retains its native structure and function after reducing its disulfide bonds. J. Mol. Biol. 239:276–284.
25. Onesti S, Brick P, Blow DM (1991) Crystal structure of a Kunitz-type trypsin inhibitor from Erythrina caffra seeds. J. Mol. Biol. 217:153–176.
26. Zhou J, Li C, Chen A, Zhu J, Zou M, Liao H, Yu Y (2020) Structural and functional relationship of Cassia obtusifolia trypsin inhibitor to understand its digestive resistance against Pieris rapae. Int. J. Biol. Macromol. [Internet] 148:908–920. Available from: https://doi.org/10.1016/j.ijbiomac.2020.01.193
27. Jain A, Kumar A, Shikhi M, Kumar A, Nair DT, Salunke DM (2020) The structure of MP-4 from Mucuna pruriens at 2.22 Å resolution. Acta Crystallogr. Sect. F Struct. Biol. Commun. 76:47–57.

Table S2. Sequence identity matrix of Kunitz-STI protease inhibitors with 3D-structures analyzed in this study.

		BASI	vvMLP	MKMLP	PPI	Alocasin	API-A	API-11	E3Ad	PSPI	BbCI	BbKI	CTI	TKI	ETI	EcTI	COTI	WCI	STI	DrTI	$CaTI_2$	MP-4	CrataBL
RASI	2QN4:A	60.4	42.5	29.2	25.5	29.6	25.0	29.2	29.5	28.8	28.1	27.0	27.3	29.8	28.2	28.8	21.1	26.5	27.9	30.7	28.7	32.2	28.7
BASI	3BX1:C		38.7	28.9	28.4	32.2	24.7	22.5	22.6	23.5	26.9	24.1	27.7	31.4	27.8	28.3	28.6	26.6	30.7	26.3	28.0	29.1	26.3
vvMLP	5YH4:A			43.7	34.2	33.0	25.7	33.6	32.3	36.0	32.0	32.0	36.0	35.0	33.0	35.2	27.3	31.0	34.6	37.2	37.5	38.5	32.9
mkKMLP	3ZC8:A				37.4	29.2	23.5	27.2	27.9	31.9	28.4	28.4	26.0	29.4	25.4	24.3	21.1	26.3	28.7	33.0	30.1	28.1	25.8
PPI	3S8K:A					29.0	25.8	27.9	26.8	27.4	25.8	27.5	26.9	29.8	23.6	27.0	26.6	25.5	26.7	30.0	28.3	31.4	27.7
Alocasin	5YCZ:A						24.1	28.7	30.8	28.1	21.5	22.0	25.0	28.7	23.5	26.3	27.0	20.3	23.2	27.2	25.1	26.3	28.6
API-A	3E8L:C							23.4	24.3	23.2	25.4	20.0	24.6	28.6	22.0	26.0	24.9	23.5	24.2	23.8	27.3	18.7	22.9
API-11	5DZU:A								89.9	71.4	24.2	24.6	25.4	27.3	22.4	27.1	23.0	24.0	26.1	26.6	29.2	28.0	24.7
API-3	5FZU:A									70.2	23.8	25.8	24.9	28.7	21.4	26.7	24.5	21.3	25.1	26.7	27.8	27.3	25.6
PSPI	3TC2:A										26.7	28.3	24.2	29.0	29.4	26.6	26.3	27.4	27.8	29.0	28.6	29.0	25.9
BbCI	2GZB:A											82.3	31.6	31.8	30.8	28.8	25.1	28.7	26.7	25.4	27.6	30.0	25.8
BbKI	4ZOT:A												33.9	34.7	32.0	29.4	25.7	30.3	28.3	24.4	27.9	28.5	25.6
CTI	1R8O:A&B													39.7	36.8	37.9	33.0	33.0	30.7	34.9	31.1	32.6	27.9
TKI	4AN6:A														38.1	43.1	29.8	27.2	33.0	35.9	34.7	37.5	24.3
ETI	1TIE:A															38.0	31.6	56.5	42.0	31.5	32.6	29.9	25.0
EcTI	4J2K:A																33.5	33.3	38.4	29.8	33.7	36.2	30.4
COTI	6KV2:A																	32.5	33.3	25.0	26.1	29.0	23.1
WCI	1EYL:A																		43.6	25.8	28.8	29.1	24.1
STI	1AVW:B																			30.3	30.9	32.3	25.9
DrTI	1R8N:A																				47.0	45.6	28.1
$CaTI_2$	5XOZ:A																					59.8	26.2
MP-4	6JBP:B																						27.9
CrataBL	4IHZ:A																						

Table S3. Global root mean square deviation of Cα atoms for the Kunitz-STI inhibitors analyzed in this study.

		3BX1:C	5YH4:A	3ZC8:A	3S8K:A	5YCZ:A	3E8L:C	5DZU:A	5FZU:A	3TC2:A	2GZB:A	4ZOT:A	1R8O:A&B	4AN6:A	1TIE:A	4J2K:A	6KV2:A	1EYL:A	1AVW:B	1R8N:A	5XOZ:A	6JBP:B	4IHZ:A
RASI	2QN4:A	1.02	1.44	1.82	1.59	1.54	1.76	1.52	1.40	1.59	2.11	2.07	1.33	1.66	1.36	1.49	1.64	1.60	1.41	1.52	1.24	1.50	1.73
BASI	3BX1:C		1.42	1.62	1.55	1.53	1.66	1.62	1.59	1.79	2.14	2.16	1.48	1.64	1.39	1.61	1.62	1.60	1.45	1.50	1.36	1.64	1.91
vvMLP	5YH4:A			1.37	1.19	1.38	1.94	1.50	1.48	1.80	2.18	2.04	1.53	1.81	1.52	1.74	1.62	1.80	1.66	1.84	1.56	1.62	1.74
mkMLP	3ZC8:A				1.49	1.68	2.09	1.76	1.75	1.94	2.19	2.24	1.91	1.79	1.73	2.09	1.92	1.96	1.95	1.78	1.72	1.78	1.97
PPI	3S8K:A					1.59	2.04	1.60	1.57	1.76	2.21	2.15	1.53	1.83	1.59	1.89	1.67	1.84	1.68	1.88	1.53	1.79	1.77
Alocasin	5YCZ:A						1.88	1.56	1.55	1.70	2.18	2.05	1.60	1.75	1.56	1.79	1.78	1.82	1.64	1.83	1.52	1.64	1.84
API-A	3E8L:C							1.85	2.03	2.06	2.09	2.16	1.76	1.99	1.73	1.79	1.85	1.85	1.82	1.88	1.80	1.91	1.90
API-11	5DZU:A								1.18	1.33	2.12	2.07	1.60	1.83	1.57	1.80	1.59	1.74	1.49	1.87	1.45	1.73	1.74
API-3	5FZU:A									1.03	2.24	2.10	1.540	1.83	1.66	1.82	1.77	1.90	1.59	1.91	1.44	1.74	1.88
PSPI	3TC2:A										2.31	2.10	1.70	1.96	1.75	1.96	1.89	1.97	1.65	1.94	1.49	1.80	1.90
BbCI	2GZB:A											1.33	1.70	1.68	1.78	1.74	1.88	1.88	1.96	2.03	1.92	1.89	2.31
BbKI	4ZOT:A												1.73	1.90	1.75	1.71	1.95	1.86	1.99	2.08	1.94	1.98	2.28
CTI	1R8O:A &B													1.23	1.24	1.16	1.41	1.42	1.23	1.57	1.28	1.52	1.73
TKI	4AN6:A														1.48	1.41	1.67	1.59	1.45	1.65	1.40	1.67	2.01
ETI	1TIE:A															1.27	1.28	0.84	1.19	1.37	1.19	1.31	1.66
EcTI	4J2K:A																1.38	1.36	1.26	1.64	1.48	1.65	1.92
COTI	6KV2:A																	1.44	1.25	1.70	1.47	1.60	1.74
WCI	1EYL:A																		1.22	1.54	1.42	1.59	1.93
STI	1AVW:																			1.50	1.13	1.47	1.83
DrTI	1R8N:A																				1.32	1.15	1.89
$CaTI_2$	5XOZ:A																					1.21	1.80
MP-4	6JBP:B																						1.81
CrataBL	4IHZ:A																						

Table S4. Root mean square deviation of Cα atoms in loop β4-β5 for Kunitz-STI inhibitors.

										CaTI2	DrTI	MP-4	vvMLP	mkMLP	CrataBL	PPI	Alocasin	API-11	API-3	PSPI	API-A
										I74-I81	S65-K72	I66-T73	P65-V72	A62-I69	P55-V62	A64-N71	P60-T67	P70-E77	P70-E77	S72-I79	A63-V70
Inhibitor name	PDB	P4-P4´	4ZOT:A	1R8O:A	4AN6:A	6KV2:A	4J2K:A	1EYL:A	1AVW:B	1TIE:A	5XOZ:A	1R8N:A	6JBP:A	5YH4:A	3ZC9:A	4IHZ:A	3S8K:A	5YCZ:A	5DZU:A	5FZU:A	3TC2:A	3E8L:C
BbCI	2GZB:A	T60-I67	0.27	0.53	0.26	0.90	0.49	0.88	0.61	1.00	1.43	1.68	1.47	1.66	1.37	1.67	1.82	1.59	1.65	1.62	1.89	2.00
BbKI	4ZOT:A	S61-I68		0.36	0.37	0.69	0.35	0.84	0.37	0.92	1.49	1.73	1.41	1.54	1.28	1.63	1.8	1.44	1.61	1.51	1.97	1.97
CTI	1R8O:A	A61-I68			0.47	0.56	0.36	0.54	0.32	0.60	1.41	1.78	1.35	1.55	0.95	1.60	1.54	1.48	1.76	1.65	1.78	2.00
TKI	4AN6:A	S63-I70				0.84	0.55	0.74	0.61	0.95	1.52	1.74	1.45	1.65	1.32	1.67	1.83	1.57	1.67	1.63	1.86	2.06
COTI	6KV2:A	A83-I90					0.56	0.50	0.45	0.39	1.47	1.82	1.10	1.50	0.71	1.47	1.25	1.47	1.74	1.73	1.71	2.11
EcTI	4J2K:A	T61-L68						0.67	0.27	0.67	1.32	1.7	1.14	1.40	0.98	1.42	1.46	1.37	1.74	1.58	1.76	1.90
WCI	1EYL:A	S65-I72							0.55	0.48	1.63	2.06	1.23	1.58	0.97	1.68	1.59	1.55	1.84	1.67	1.9	2.11
STI	1AVW:B	S60-I67								0.58	1.36	1.82	1.25	1.43	0.89	1.49	1.41	1.33	1.82	1.68	1.8	1.95
ETI	1TIE:A	S60-I67									1.42	2.02	1.21	1.65	0.76	1.66	1.25	1.62	1.83	1.8	1.92	2.08

Table S5. Peptide torsion angles phi and psi for positions P4 to P4' in loop β4-β5 adopting canonical and noncanonical conformations.

	Inhibitor	PDB ID	P4	P3	P2	P1	P1´	P2´	P3´	P4´
Canonical Inhibitor in complex with Serine Proteases	STI	1AVW	-114.5/144.4	-58.5/-35.2	-55.1/138.9	-89.3/39.6	-84.5/148.1	-68.0/-37.5	-119.4/155.3	-75.0/123.4
	TKI	4AN7	-109.4/148.1	-69.6/-19.4	-66.6/141.4	-94.1/39.4	-82.3/151.5	-72.0/-28.5	-138.1/172.7	-82.3/129.7
	EcTI	4J2Y	-141.55/148.5	-60.2/-39.5	-65.7/143.5	-95.3/42.9	-79.9/156.4	-79.5/-23.4	-123.9/175.7	-122.4/143.3
	BbKI	7JOS	-125.7/144.6	-71.3/-14.0	-58.9/142.4	-86.7/43.1	-111.9/156.5	-101.8/5.5	-130.7/166.1	-117.0/129.5
Canonical Inhibitors	TKI	4AN6	-130.3/154.4 -129.2/151.2	-66.9/-17.3 -63.3/-36.7	-61.8/134.4 -73.0/116.9	-81.8/68.5 -100.3/48.2	-151.4/155.0 -78.2/167.9	-86.8/-11.7 -73.8/-31.5	-137.4/160.2 -122.4/153.9	-97.8/141.9 -83.4/131.5
	EcTI	4J2K	-128.3/150.7 -142.3/143.8	-62.0/-37.5 -60.2/-39.8	-49.9/133.1 -59.2/156.6	-86.1/63.5 -82.1/76.3	-113.8/162.8 -133.8/139.9	-83.0/-34.5 -77.5/-26.5	-123.5/171.3 -120.8/176.8	-112.7/150.9 -118.6/144.9
	BbKI	4ZOT	-120.8/142.8	-80.8/2.1	-69.9/153.4	-73.7/-13.6 -91.7/74.6	-65.1/153.4 -155.1/147.0	-99.1/5.0	-129.8/167.3	-113.4/139.7
	BbCI	2GZB	-142.9/144.4 -142.5/148.2	-66.6/-4.3 -74.0/0.1	-60.3/136.9 -60.7/135.8	-77.3/77.6 -76.2/78.4	-157.8/147.3 -159.7/145.7	-87.2/-6.8 -84.2/-8.9	-133.4/167.4 -135.1/164.1	-112.0/157.5 -110.7/153.5
	CTI	1R8O	-114.5/138.8	-69.9/-42.5	-46.8/143.3	-82.3/83.7	-121.5/147.7	-76.7/-27.5	-130.7/142.5	-82.6/162.9
	ETI	1TIE	-88.3/171.7	-66.3-30.9	-112.7/150.9	-66.6/135	-169.1/151.6	-74.7/-46.1	-106.3/163.5	-102.7/134.1
	WCI	1EYL	-89.9/172.9	-71.3/-17.5	-97.8/168.1	-82.4/-9.1	-55.4/158.4	-73.7/-29.2	-119.6/-174.2	-128.8/107.3
	COTI	6KV2	-87.4/168.5 -84.9/171.1 -82.7/158.5 -85.7/157.7	-84.9/-12.8 -90.6/-11.8 -68.7/-42.0 -69.2/-41.3	-107.9/173.6 -102.9/169.4 -75.3/154.0 -73.1/148.8	-117.1/37.8 -116.7/36.0 -119.9/40.4 -114.7/35.8	-74.0/166.5 -73.9/170.6 -77.8/177.3 -79.7/176.3	-71.6/-29.4 -71.1/-33.5 -68.0/-39.6 -65.3/-40.3	-124.5/149.6 -120.3/152.7 -101.8/178.8 -99.8/-178.7	-90.6/133.1 -91.2/125.8 -127.9/130.2 -129.3/131.7
Noncanonical Inhibitors	mkMLP	3ZC9	-64.4/154.8	-59.6/-38.1	-97.9/16.3	56.8/42.5	-75.5/149.1	-64.0/-38.3	-97.3/137.6	-94.8/126.4
	$CaTI_2$	5XOZ	-109.2/125.6	-77.8/-31.6	-71.6/-155.5	-80.0/134.0	-102.1/134.0	-86.2/-6.9	98.9/1.9	-142.3/147.2
	DrTI	1R8N	-136.2/148.9	-81.0/1.1	-111.8/-24.3	-151.8/148.1	-86.6/-23.4	-142.6/162.4	132.9/175.4	-80.3/150.2
	MP-4	6JBP	-130.4/150.9	-48.8/142.6	97.7/-171.6	-133.5/-13.7	-66.7/142.6	-54/142.6	-161.3/157.9	-79.5/147.8
	vvMLP	5YH4	-60.8/150.3	-58.0/-34.0	-88.7/93.0	-35.3/-49.5	-56.8/-39.2	-60.9/135.6	-106.9/-23.2	-90.0/135.9
	CrataBL	4IHZ	-63.6/154.3 -61.2/148.9	-57.6/-34.9 -57.4/-37.9	-88.2/90.8 -79.4/110.8	-70.7/-6.6 -71.4/-18.9	-83.5/-13.0 -81.2/-17.0	-67.1/140.7 -71.9/121.2	-91.6/-25.7 -83.7/-30.3	-106.5/132.9 -106.2/133.9
	PPI	3S8K	-77.4/148.7 -78.3/148.1	-90.0/-46.4 -79.9/-54.0	-94.4/169.1 -90.5/170.5	-70.9/126.7 -68.2/122.1	-119.9/103.3 -118.8/97.7	-63.8/-40.0 -66.7/-33.3	-87.5/1.3 -81.3/-1.8	62.4/18.8 60.5/20.5
	Alocasin	5YCZ	-59.9/137.8 -58.7/135.9	-63.5/-23.0 -56.5/-38.0	-86.3/121.6 -71.5/106.0	-73.6/-24.8 -50.9/-26.1	-87.1/-17.7 -93.8/-13.9	-65.1/123.0 -66.7/114.5	-117.6/5.8 -87.7/-31.3	-119.2/161.5 -102.4/134.1
	API-A	3E8L	-151.2/170.1	-65.9/154.7	-44.1/157.5	81.2/-11.2	-65.8/144.2	-71.2/156.2	-74.4/7.7	-125.5/143.5
	API-11	5DZU	-65.6/147.9 -72.4/148.1	-54.1/-46.0 -65.7/-45.4	-142.5/132.6 -135.6/98.2	81.3/-142.3 98.1/-111.3	-89.4/179.8 -106.7/174.4	-91.5/110.1 -91.9/107.7	-129.3/156.3 -129.0/157.5	-61.2/150.1 -61.9/144.2
	API-3	5FZU	-75.1/162.0	-60.2/-46.3	-64.2/-149.9	-119.6/-3.0	-126.6/140.9	-89.5/103.6	-111.1/157.8	-58.6/156.9
	PSPI	3TC2	-56.3/-6.8	-102.5/129.6	58.8/-69.5	61.4/-65.3	86.5/19.2	85.9/-60.1	-100.2/162.4	-114.9/125.6

* The angles phi and psi for all inhibitor chains are shown for those 3D-structures with more than one chain for the inhibitor

Table S6. Structural comparison of the conformation of loops β9-β10 with the canonical trypsin binding loop of inhibitor API-A

		RMSD-Cα (Å)				
Inhibitor		Alocasin	API-11	API-3*	vvMLP	mkMLP*
API-A	P4-P4´	P144-L151	P151-F158	L150-F157	D145-G152	K145-N152
	E142-P149	1.75	2.86	3.14	1.58	1.2
	P3-P3´	S145-C150	F152-N157	R151-Q156	F146-C151	S146-C151
	F143-C148	0.46	1.68	2.67	1.37	1.05

*3D-structures used in the analysis: API-3 (PDB ID: 5FZU) and mkMLP (PDB ID: 3ZC9)

Table S7. Structural comparison of the conformation of β5-β6 loops with noncanonical trypsin binding loop of inhibitor API-A (M83-C89).

Inhibitors	PDB ID	Amino acid residues	RMSD-Cα
API-3	5FZU:A	I87-C93	**0.46**
API-11	5DZU:A	I87-C93	**0.31**
PSPI	3TC2:A	I91-C97	**0.32**
PPI	3S8K:A	F83-C89	2.00
Alocasin	5YCZ:A	F78-C84	2.13
vvMLP	5YH4:A	F83-C89	2.2
mkMLP	3ZC9:A	F79-C85	2.1
CTI	1R8O:A	V78-C84	2.76
TKI	4AN6:A	F80-C86	2.58
ETI	1TIE:A	F77-C83	2.71
EcTI	4J2K:A	T80-C86	2.29
COTI	6KV2:A	Y102-C108	1.97
STI	1AVW:B	F80-C86	**0.31**
WCI	1EYL:A	F82-C88	2.78
DrTI	1R8N:A	F83-C89	2.57
$CaTI_2$	5XOZ:A	F92-C98	2.53
MP-4	6JBP:A	F84-C90	2.52

Table S8. Dihedral angles for P5-P2´ positions of noncanonical trypsin-binding β5-β6 loops.

Inhibitor	Amino Acid	Position	Phi (ϕ)	Psi (ψ)	Phi (ϕ)	Psi (ψ)	Phi (ϕ)	Psi (ψ)
			First Chain*		Second Chain*		Third Chain*	
API-A (3E8L:C)	M83	P5	-151.2	148.7	-	-	-	-
	P84	P4	-64.8	140.8	-	-	-	-
	V85	P3	-155.1	145.6	-	-	-	-
	P86	P2	-67.6	143.1	-	-	-	-
	L87	P1	-68.3	-14	-	-	-	-
	I88	P1´	-68.7	-13.8	-	-	-	-
	C89	P2´	-109.7	138.4	-	-	-	-
API-3 (5FZU:A)	I87	P5	-143.8	150.1	-	-	-	-
	P88	P4	-62.4	154.7	-	-	-	-
	T89	P3	-147.8	151	-	-	-	-
	V90	P2	-62.9	153.2	-	-	-	-
	K91	P1	-46.0	-46.7	-	-	-	-
	L92	P1´	-56.6	-41.3	-	-	-	-
	C93	P2´	-97.0	8.6	-	-	-	-
PSPI (3TC:A/B/C)	I91	P5	-137.1	156.9	-139.4	158.5	-132.3	153.1
	S92	P4	-72.1	136.4	-75.2	137.0	-76.7	145.9
	T93	P3	-140.1	177.5	-139.1	170.5	-131.3	171.0
	S94	P2	-72.8	160.9	-62.2	161.8	-57.9	160.6
	K95	P1	-49.9	-47.6	-55.4	-52.6	-58.3	-43.2
	M96	P1´	-64.3	-25.1	-52.1	-22.1	-54.8	-38.6
	C97	P2´	-92.7	3.2	-99.6	0.5	-99.4	7.2
API-11 (5DZU:A/B)	I87	P5	-136.1	150.3	-140.3	164.9	-	-
	A88	P4	-64.1	137.9	-91.1	130.3	-	-
	T89	P3	-144.4	169.6	-122.8	169.6	-	-
	V90	P2	-61.1	156.7	-64.2	151.1	-	-
	K91	P1	-53	-42.4	-38.9	-46.9	-	-
	L92	P1´	-65.1	-27.4	-51.0	-45.5	-	-
	C93	P2´	-92.8	-0.3	-90.6	11.4	-	-
STI (1AVW:B)	F80	P5	-136.3	163.1	-	-	-	-
	A81	P4	-76.5	125.9	-	-	-	-
	V82	P3	-126.1	163.7	-	-	-	-
	I83	P2	-65	157.9	-	-	-	-
	M84	P1	-58.8	-30.4	-	-	-	-
	L85	P1´	-67.7	-12.8	-	-	-	-
	C86	P2´	-115.1	-0.5	-	-	-	-

*According to the order of the inhibitor chains in the PDB file.

Differences of more than 90 degrees are highlighted in red

Table S9. Root mean square deviations of the conformation of loops β5-β6 with noncanonical subtilisin binding loop of inhibitor BASI.

			BASI	
			3BX1:C	3BX1:D
Inhibitor	PDB	P5-P1	A86-C90	
RASI	2QN4:A	A86-C90	0.44	0.37
	2QN4:B		0.30	0.27
CrataBL	4IHZ:A	D76-C80	0.92	0.84
	4IHZ:B		0.89	0.81
PPI	3S8K:A	I85-C89	0.42	0.32
	3S8K:B		0.42	0.33
	3S8J:A	I85-C89	0.50	0.42
	3S8J:B		0.48	0.41
Alocasin	5YCZ:A	E80-C84	0.27	0.25
	5YCZ:B		0.17	0.18
vvMLP	5YH4	A85-C89	0.43	0.47
mkMLP	3ZC9	T81-C85	0.39	0.4
API-A	3E8L:C	M83-L87	0.33	0.26
API-3	5FZU:A	I87-K91	0.26	0.19
API-11	5DZU:A	I87-K91	0.44	0.39
	5DZU:B		0.48	0.45
PSPI	3TC2:A	I91-C95	0.46	0.4
	3TC2:B		0.41	0.35
	3TC2:C		0.72	0.68
CTI	1R8O	E80-C84	1.47	1.46
TKI	4AN6:A	I82-C86	1.44	1.43
	4AN6:B		1.47	1.46
ETI	1TIE:A	Y79-C83	1.59	1.58
EcTI	4J2K:A	R81-A85	0.65	0.58
	4J2K:B		0.61	0.52
COTI	6KV2:A	Y102-N106	1.32	1.31
	6KV2:B		1.32	1.3
	6KV2:C		1.34	1.32
	6KV2:D		1.38	1.36
WCI	1EYL	N84-C88	1.65	1.64
STI	1AVW:B	F80-M84	0.93	0.87
	1R8N	E85-C89	1.25	1.24
$CaTI_2$	5XOZ:A	K94-C98	1.3	1.28
	5XOZ:B		1.3	1.28
MP-4	6JBP:B	E86-C90	1.1	1.08

Table S10. Dihedral angles for P5-P1 positions of noncanonical subtilisin-binding β5-β6 loops.

Inhibitor	Amino acid	Position	Phi (ϕ)	Psi (ψ)	Phi (ϕ)	Psi (ψ)
			First Chain*		Second Chain	
BASI (3BX1)	A86	P5	-167.1	161.5	-167.0	155.6
	Y87	P4	-132.1	141.1	-119.0	139.2
	T88	P3	-133.9	156.7	-131.4	154.0
	T89	P2	-68.1	-22.8	-68.9	-18.8
	C90	P1	-68.2	-26.4	-70.3	-30.0
RASI (2QN4)	A86	P5	-164.4	146.1	-173.5	152.7
	A87	P4	-67.8	143.2	-99.3	147.8
	T88	P3	-119.1	140.8	-128.4	134.7
	I89	P2	-71.4	-8.9	-51.7	-23.1
	C90	P1	-100.4	-8.1	-73.7	-15.9
PPI (3S8K)	I85	P5	-139.3	170.3	-143.3	171.8
	T86	P4	-101.9	142.5	-102.9	143.6
	I87	P3	-126.1	161.4	-128.2	165.6
	N88	P2	-107.5	36.4	-113.3	36.4
	C89	P1	-119.1	150.8	-118.1	152.2
Alocasin (5YCZ)	E80	P5	-141.3	160.6	-139.8	156.9
	P81	P4	-73.8	169.0	-67.3	172.0
	T82	P3	-152.6	156.8	-158.0	150.5
	R83	P2	-74.5	-6.0	-70.2	-0.2
	C84	P1	-114.7	151.8	-115.1	143.5
vvMLP (5YH4)	A85	P5	-85.2	157.7	-	-
	S86	P4	-68.4	163.0	-	-
	T87	P3	-137.8	-177.8	-	-
	I88	P2	-69.6	-10.3	-	-
	C89	P1	-106.0	143.8	-	-
mkMLP (3ZC9)	T81	P5	-147.7	153.4	-	-
	E82	P4	-88.1	156.9	-	-
	T83	P3	-130.4	163.2	-	-
	S84	P2	-61.4	-7.4	-	-
	C85	P1	-86.0	-7.8	-	-
API-A (3E8L)	M83	P5	-151.2	148.7	-	-
	P84	P4	-64.8	140.8	-	-
	V85	P3	-155.1	145.6	-	-
	P86	P2	-67.6	143.1	-	-
	L87	P1	-68.3	-14.0	-	-
API-3 (5FZU)	I87	P5	-143.8	150.1	-	-
	P88	P4	-62.4	154.7	-	-
	T89	P3	-147.8	151.0	-	-
	V90	P2	-62.9	153.2	-	-
	K91	P1	-46.0	-46.7	-	-

*According to the order of the inhibitor chains in the PDB file.