THE JOURNAL OF BIOLOGICAL CHEMISTRY VOL. 284, NO. 39, pp. 26676–26684, September 25, 2009

The Ternary Structure of the Double-headed Arrowhead Protease Inhibitor API-A Complexed with Two Trypsins Reveals a Novel Reactive Site Conformation*

Received for publication, May 17, 2009, and in revised form, July 13, 2009 Published, JBC Papers in Press, July 28, 2009, DOI 10.1074/jbc.M109.022095

Rui Bao[‡], Cong-Zhao Zhou[§], Chunhui Jiang[‡], Sheng-Xiang Lin[¶1], Cheng-Wu Chi[‡¶2], and Yuxing Chen[§3]

From the [‡]Institute of Protein Research, Tongji University, Shanghai 200092, the [§]Hefei National Laboratory for Physical Sciences at Microscale and School of Life Sciences, University of Science and Technology of China, Hefei, Anhui 230026, and the [¶]Institute of Biochemistry and Cell Biology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Shanghai 200031, China

The double-headed arrowhead protease inhibitors API-A and -B from the tubers of *Sagittaria sagittifolia* (Linn) feature two distinct reactive sites, unlike other members of their family. Although the two inhibitors have been extensively characterized, the identities of the two P1 residues in both API-A and -B remain controversial. The crystal structure of a ternary complex at 2.48 Å resolution revealed that the two trypsins bind on opposite sides of API-A and are 34 Å apart. The overall fold of API-A belongs to the β-trefoil fold and resembles that of the soybean Kunitz-type trypsin inhibitors. The two P1 residues were unambiguously assigned as Leu^{87} and Lys^{145}, and their identities were further confirmed by site-directed mutagenesis. Reactive site 1, composed of residues P5 Met^{83} to P5′ Ala^{92}, adopts a novel conformation with the Leu^{87} completely embedded in the S1 pocket even though it is an unfavorable P1 residue for trypsin. Reactive site 2, consisting of residues P5 Cys^{141} to P5′ Glu^{150}, binds trypsin in the classic mode by employing a two-disulfide-bonded loop. Analysis of the two binding interfaces sheds light on atomic details of the inhibitor specificity and also promises potential improvements in enzyme activity by engineering of the reactive sites.

Protease inhibitors (PIs)[4] are ubiquitously distributed in all organisms, including plants, animals, and microorganisms (1). They play vital roles in regulating their corresponding proteases, which are involved in many biological processes such as protein digestion, cell signal transmission, inflammation, apoptosis, blood coagulation, and embryogenesis (2). The clinical applications of PIs are widespread, and there is great interest in developing more potent therapeutic PIs for treating human diseases related to cancer (3), pancreatitis (4), thrombosis (5), and AIDS (6). To this end, the soybean Kunitz-type serine proteases inhibitors have been extensively studied (1, 7–11). The inhibitors of this family generally contain 170–200 residues and have two disulfide bonds. Most members have only one reactive site located in the region of residues 60–70 (7, 10, 12–14). However, a few members possess two reactive sites that simultaneously bind two protease molecules and are thus termed double-headed inhibitors (15–18). All of these inhibitors are classified into family I3 of peptidase inhibitors (19). Most members are further grouped into subfamily I3A. However, the double-headed arrowhead PIs API-A and -B are grouped in subfamily I3B because of their very low sequence similarity to other members (19). In contrast to other double-headed PIs such as the Bowman-Birk and ovomucoid inhibitors, which have two identical reactive sites that have evolved by domain shuffling and gene duplication (1, 20–25), both API-A and -B have two distinct reactive sites.

API-A and -B were first purified from the tubers of *Sagittaria sagittifolia* (Linn) in 1979 (26). Both consist of 179 residues with three disulfide bonds and can inhibit a variety of serine proteases, including trypsin, chymotrypsin, and porcine tissue kallikrein (17, 26–28). Although the sequence identity of API-A and -B is as high as 91%, their inhibitory specificities differ. The former can bind one molecule of trypsin and one molecule of chymotrypsin, whereas the latter can simultaneously bind two molecules of trypsin (26). The two P1 residues of the reactive sites of API-A and -B were first predicted to be Lys^{44} and Arg^{76} based on their surrounding sequences, which are similar to those of the reactive sites of bovine pancreas trypsin inhibitor and soybean Kunitz trypsin inhibitor (29). However, their identities were later revised to Arg^{76} and Leu^{87} (for API-A) or Lys^{87} (for API-B) based on results from sited-directed mutagenesis studies (30).

To clarify these controversies, we solved the crystal structure of API-A in complex with two molecules of bovine trypsin. To the best of our knowledge, this is the first report on the three-dimensional structure of the double-headed Kunitz-type trypsin inhibitor in complex with two molecules of protease. On the basis of this structure, the two P1 residues have now been identified as Leu^{87} and Lys^{145} for reactive site 1 (RS1) and 2 (RS2), respectively. The results were further confirmed by site-directed mutagenesis. It was earlier shown that the first P1 residue Leu^{87} interacts preferentially with chymotrypsin (30). In our structure, Leu^{87} is snugly embedded in the S1 pocket of trypsin, as a consequence of the broad interface contributed by the surrounding residues. Comprehensive analyses of the two

* This work was supported by Ministry of Science and Technology of China Projects 2006CB910202 and 2006CB806501.

The atomic coordinates and structure factors (code 3E8L) have been deposited in the Protein Data Bank, Research Collaboratory for Structural Bioinformatics, Rutgers University, New Brunswick, NJ (http://www.rcsb.org/).

[1] On leave from the Laboratory of Molecular Endocrinology (CHUL, CHUQ) and Laval University.

[2] To whom correspondence may be addressed. Tel.: 86-21-54921165; E-mail: zwqi@sibs.ac.cn.

[3] To whom correspondence may be addressed. Tel.: 86-551-3602492; Fax: 86-551-3602491; E-mail: cyxing@ustc.edu.cn.

[4] The abbreviations used are: PI, protease inhibitor; RMSD, root mean square deviation; BBBI, Bowman-Birk protease inhibitor; STI, soybean trypsin inhibitor.

reactive site interfaces have provided functional insights into the novel inhibitory patterns of this unique double-headed protease inhibitor.

EXPERIMENTAL PROCEDURES

RNA Extraction, cDNA Cloning, and Site-directed Mutation—Total RNA was isolated from 100 mg of fresh arrowhead (*S. sagittifolia*, Linn) buds using TRIzol reagent (Invitrogen) according to the instruction protocol. The first chain of cDNA was reverse-transcribed from the total RNA using a specific primer 5′-CGCCGCGGCCGCTTAGAGTGCGTC-GRACTTMTG-3′ with a NotI site (where R stands for G/A, and M for A/C) and Superscript II reverse transcriptase (Invitrogen). The coding sequence of API-A was PCR-amplified from cDNA using the forward primer 5′-CTCGCATA-TGGATCCCGTCGTCGACAGC-3′ containing an NdeI site and specific primer as the reverse one. The PCR product was cloned into a T-easy vector (Promega) and transferred to DH5α. After the insert screening and DNA sequencing, the target gene was cloned into a pET28a-derived expression vector with a His_6 tag at the N terminus after the start codon. The site-direct mutagenesis (L87P or K145A) was carried out by PCR as described (31). Compared with the primary sequence of API-A reported previously (Swiss-Prot entry P31608), there are two mutations at the N and C termini (Arg^{39}/His^{39} and Gln^{172}/Arg^{172}) that might be due to polymorphism. Both mutations are far away from the reactive sites and thus have no influence on the inhibitory activity.

Protein Expression and Purification—All of the constructs were co-transformed with PKY206 (a plasmid containing the *GroESL* genes of *Escherichia coli*, for the synthesis of chaperones GroEL and GroES) (32) into *E. coli* BL21 (DE3). Transformant cells were grown in 2×YT medium (5 g of NaCl, 16 g of bactotryptone; 10 g of yeast extract in 1 liter of H_2O) at 37 °C up to an $A_{600\ nm}$ of 0.7. The protein expression was induced by adding isopropyl 1-thio-β-D-galactopyranoside to a final concentration of 0.2 mM, and the culture was incubated at 16 °C for further 20 h. The cells were collected by centrifugation and resuspended in the lysis buffer containing 100 mM NaCl, 20 mM Tris-HCl, pH 7.5, and lysed by sonication on ice. The soluble fraction was loaded to the nickel-nitrilotriacetic acid affinity resin (Qiagen) and washed with the lysis buffer plus 10–50 mM imidazole. The bound protein was then eluted with the lysis buffer plus 200 mM imidazole and further purified by gel filtration using a Hiload 16/60 Superdex 75 (Amersham Biosciences) column equilibrated with the lysis buffer. The purified protein was concentrated to 0.4 mg/ml in the same buffer plus 50% glycerol and stored at −80 °C for activity assay.

To prepare the protein complex, ~5 mg of purified API-A was mixed with 12 mg of bovine β-trypsin (USP grade; Ameresco) in the presence of 20 mM $CaCl_2$. After incubating for 2 h at 4 °C, the mixture was loaded to a Hiload 16/60 Superdex 200 (Amersham Biosciences) column equilibrated with the buffer of 100 mM NaCl, 20 mM Tris-HCl, pH 7.5, and 20 mM $CaCl_2$. The fractions containing the complex of API-A and trypsin were collected and concentrated to 15 mg/ml by ultrafiltration using an Amicon Ultra 10-kDa cut-off concentrator (Millipore).

Crystallization and X-ray Data Collection—Crystallization trials were carried out at 291 K by the hanging drop vapor diffusion method using crystal screen kits I and II (Hampton Research Inc.). Each drop containing 1 μl of reservoir solution and 1 μl of protein sample (13 mg/ml) was equilibrated against 0.5 ml of reservoir solution. Needle-like crystals appeared in several conditions. The crystals suitable for x-ray diffraction were optimized using a sitting drop method. In the optimized condition, a droplet was prepared by mixing 2 μl of protein sample with 2 μl of reservoir solution containing 20% polyethylene glycol 8000, 0.1 M sodium cacodylate, pH 6.5, 0.2 M $(NH_4)_2SO_4$. Plate-like crystals with dimensions of 0.5 × 0.2 × 0.03 mm^3 appeared in a week. A single crystal was transferred to the cryoprotectant containing the reservoir solution plus 20% glycerol and flash-frozen in the liquid nitrogen for data collection.

The diffraction data were collected at 100 K on an in-house Rigaku MM007 x-ray generator (λ = 1.54179 Å) with a Mar-Research 345 detector at School of Life Sciences, University of Science and Technology of China (Hefei, China). The data diffracted at 2.48 Å resolution were processed with program *AUTOMAR* 1.2 (33). The crystals belong to the space group $C222_1$ (see Table 1). There is a ternary complex of one API-A and two trypsin molecules in the asymmetric unit.

Structure Determination and Refinement—The initial phase was calculated by molecular replacement using the program PHASER, which integrated in the CCP4 suite (34). The starting search model for trypsin was derived from the bovine-trypsin complex with benzamidine (Protein Data Bank code 2J9N); two clear orientations of trypsin were obtained after rotation and translation search using the program MOLREP in CCP4 suite. Because of the low sequence identity (<25%) between API-A and homologs of known structure, we used a model consisting of the following segments: 30–37, 46–50, 60–64, 80–84, 125–130, 139–144, and 173–176 from barley α-amylase subtilisin inhibitor (endogenous protein inhibitor chain C, Protein Data Bank code 1AVA) as template for phased rotation and translation search using the phasing information obtained from these two molecules of trypsin. Further main chain tracing and manual rebuilding were performed with Coot 0.3.3 (35) based on the electron density from the initial solution, but residues 1, 29–32, 52–54, and 177–179 of API-A were not fitted because of the poor electron density. The refinement was carried out with Refmac5 (34). The stereochemistry of the model calculated with Molprobity (36) indicated that 97% of the residues in the model were in the most favored region of Ramachandran plot, and the rest were in the allowed regions. The data collection and refinement statistics are listed in Table 1. Final coordinates have been deposited in Protein Data Bank under the accession code of 3E8L.

Inhibitory Activity Assay—The protein concentration of API-A was determined by UV absorbance at 280 nm with an extinction coefficient of 22,920 $cm^{-1}\ M^{-1}$. The assay of inhibitory activity against trypsin was performed in 400 μl of 50 mM Tris-HCl, pH 7.8, 20 mM $CaCl_2$, 50 nM bovine trypsin (Ameresco), and various amounts of the wild-type API-A or variants using 500 nM *N*-α-benzoyl-DL-arginine-4-nitroanilide (Sigma) as the substrate. The absorbance at 410 nm was traced

TABLE 1
Data collection and refinement statistics

Data set	Parameters
Space group	$C222_1$
Unit cell parameters (Å)	
a	76.63
b	110.86
c	152.99
Resolution (Å)	30-2.48 (2.57-2.48)[f]
Unique reflections	23,471 (2,298)
Completeness (%)	99.7 (99.5)
$I/\sigma(I)$	9.9 (2.1)
R_{merge} (%)[a]	12.46 (47.80)
Refinement	
Resolution (Å)	30-2.48 (2.54-2.48)
R factor[b]	0.19 (0.25)
R_{free}[c]	0.24 (0.32)
Contents of asymmetric unit	
Protein molecules	3
Protein atoms	4,616
Water atoms	160
RMSD geometry[d]	
Bond lengths (Å)	0.011
Bond angles (°)	1.437
Average of B factors (Å^2)	37.08
Ramachandran plot[e]	
Most favored (%)	96.92
Additional allowed (%)	3.08
Outliers (%)	0
Protein Data Bank code	3E8L

[a] $R_{merge} = \Sigma_{hkl}\Sigma_i|I_i(hkl) - <I(hkl)>|/\Sigma_{hkl}\Sigma_i I_i(hkl)$, where $I_i(hkl)$ is the intensity of an observation, and $<I(hkl)>$ is the mean value for its unique reflection. Summations are over all reflections.
[b] *R* factor $= \Sigma_h|F_o(h) - F_c(h)|/\Sigma_h F_o(h)$, where F_o and F_c are the observed and calculated structure factor amplitudes, respectively.
[c] R_{free} was calculated with 6% of the data excluded from the refinement.
[d] RMSD, root mean square deviation from ideal values.
[e] Categories were defined by Molprobity.
[f] The values in parentheses refer to statistics in the highest bin.

for 5 min. The maximal slope of the plot of apparent absorbance *versus* time was used to determine the residual trypsin activity, which was plotted as a function of the molar ratio of inhibitor API-A to trypsin (29).

RESULTS AND DISCUSSION

Previously, soluble forms of API-A and -B could only be obtained by heterogeneous expression in *Saccharomyces cerevisiae* (29). In this study, soluble expression was achieved by co-expressing API-A with the GroEL/ES chaperones in *E. coli*, which greatly simplified the procedure. The ternary complex was obtained in the presence of $CaCl_2$ and was readily crystallized. The crystals belong to the space group $C222_1$, and there is one ternary complex in the crystallographic asymmetric unit (Table 1). The structure was solved by molecular replacement searching for trypsin and API-A separately, and it was refined to 2.48 Å resolution (Table 1).

Overall Structure—The overall structure shows a ternary complex of the double-headed inhibitor API-A that is simultaneously bound to two trypsin molecules, *i.e.* trypsin-1 and -2 (Fig. 1*A*). The two reactive sites of API-A adopt obviously different conformations. This is in contrast to the ternary structure of Bowman-Birk protease inhibitors (BBBIs) from barley seeds in complex with porcine trypsin, in which two highly similar reactive site loops are located in the duplicated domains (Fig. 1*B*). The orientations of the trypsin molecules in the two complexes differ, but they share a typical area of the interface between the inhibitor and protease, which is 1295/1855 Å^2 in API-A and 1430/1824 Å^2 in BBBI (37). The distance between the two reactive sites, which is ~34 Å in BBBI and 39 Å in API-A, is also comparable with that in smaller inhibitors such as Bowman-Birk type protease inhibitors from snail medic seeds (7 kDa) or Bowman-Birk protease inhibitors from soybean (8 kDa). This has earlier been assumed to be the minimum distance required for the simultaneous binding of two protease molecules (21, 22, 24, 38, 39).

The overall fold of API-A resembles that of the Kunitz-type trypsin inhibitors. It belongs to the β-trefoil fold (40) with a core of six antiparallel β-sheets (β1 to β12) surrounded by 13 loops (Fig. 1*C*). The hydrophobic core consists of a shallow β-barrel formed by three β-sheets (β1-β12, β4-β5, and β8-β9) and a cap of three hairpins (β2-β3, β6-β7, and β10-β11) (Fig. 1*C*). Additionally, a 3_{10} helix is located between loops 8 and 9. Although they share a very similar hydrophobic core, the primary sequences of the Kunitz-type inhibitors vary dramatically because of the high degree of variation in the loop regions. In contrast to the root mean square deviation (RMSD) of 1.14 Å between the Cα atoms of the equivalent β-stranded cores and an overall RMSD of 2.17 Å, the sequence identity between API-A and soybean trypsin inhibitor (STI, Protein Data Bank code 1AVW) is only 23%. Structure-based multialignment of Kunitz-type trypsin inhibitors of known structure revealed that most conserved residues are located in the β strands (Fig. 1*D*), which are considered to be necessary for maintaining a universal rigid core (40).

In addition, API-A contains three disulfide bonds. The first bond (Cys^{43}–Cys^{89}) cross-links loops 3 and 6 of RS1, which is present in most Kunitz-type trypsin inhibitors (1). However, this disulfide bond may have little effect on the β-trefoil fold because the *Bauhinia bauhinioides* cruzipain inhibitor remains very stable even in the absence of any disulfide bond (41). The two other disulfide bonds are located in RS2, and one of these (Cys^{141}–Cys^{144}) forms an intraloop disulfide bond in loop 10, whereas the other (Cys^{139}–Cys^{148}) further stabilizes loop 10 with strand β9. This conformation of one reactive site stabilized by two disulfide bonds has not been found in any other Kunitz-type protease inhibitors of known structure (26). Taken together, these disulfide bonds might play a role in stabilizing the conformation of the reactive site, as observed in the structure of Bowman-Birk type protease inhibitors (22, 24, 38, 39).

Reactive Sites—As shown in Fig. 2 (*A* and *B*), the excellent continuous electron densities indicate well defined interfaces between the inhibitor and trypsins for RS1 and RS2 and intact scissile bonds between the P1 and P1′ residues in the two active sites (Leu^{87}–Ile^{88} for RS1 and Lys^{145}–Ile^{146} for RS2). Both reactive sites adopt a typical noncovalent "lock and key" inhibitory mechanism. The two reactive sites of API-A are located in loops 6 and 10, respectively, in contrast to the sole reactive site in loop 5 (residues 60–67), which is usually found in typical Kunitz-type inhibitors (8, 11, 13, 18, 42). The phenomenon of switching reactive site loops for double-headed inhibitors is not unique to API-A and has been proposed earlier for loop 3 of winged bean chymotrypsin inhibitor based on its crystal structure (18). Both reactive sites of API-A and winged bean chymotrypsin inhibitor

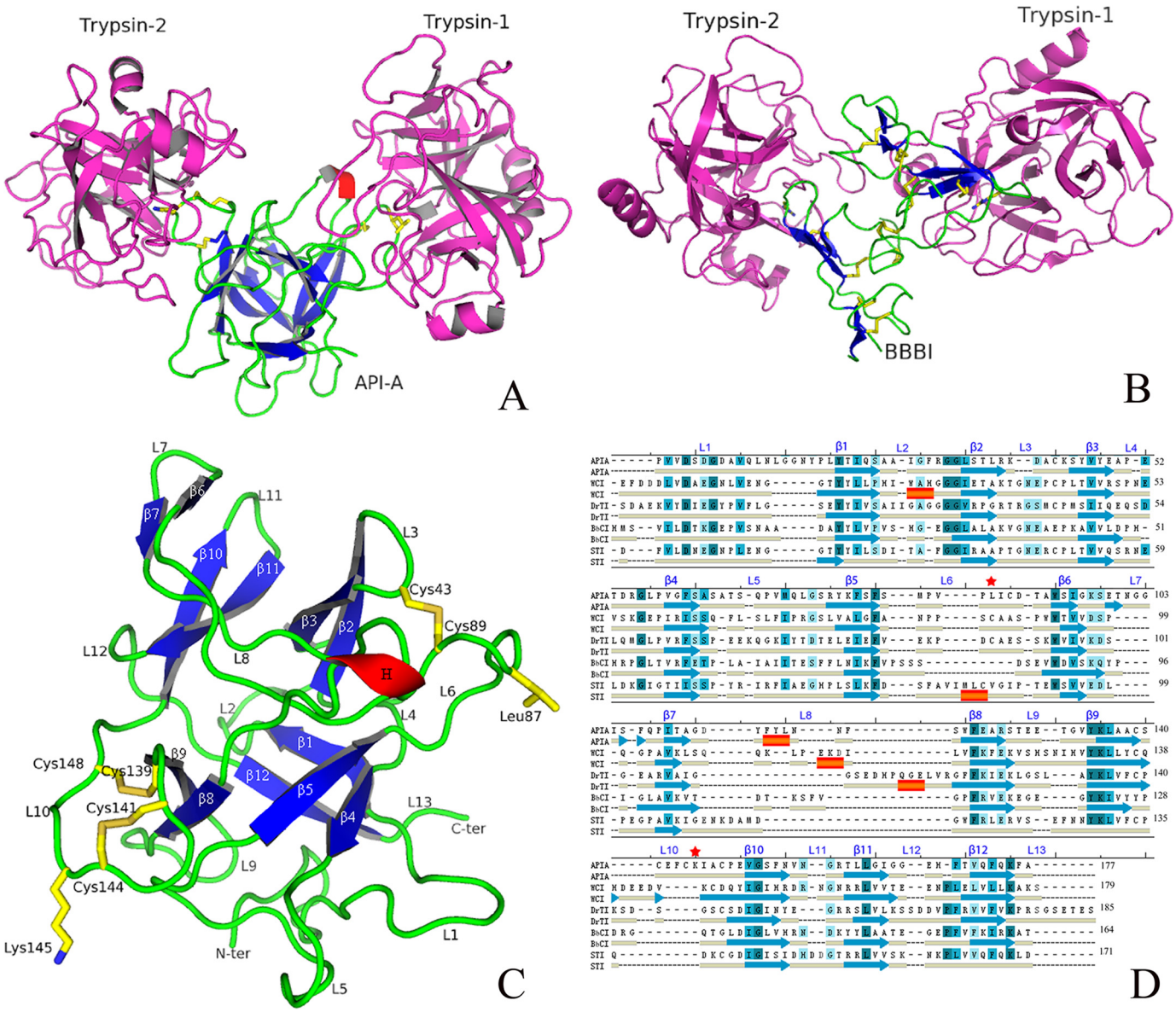
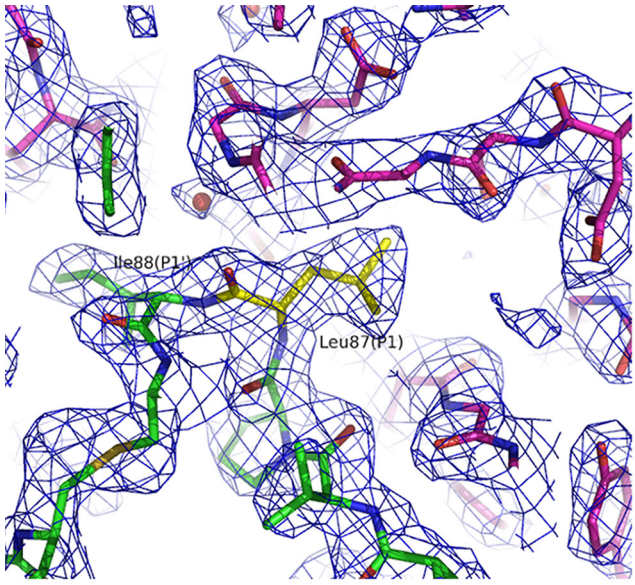


FIGURE 1. **Overall structures and multialignment.** Overall structures of the ternary complex of API-A-trypsin (*A*) and BBBI-trypsin (*B*). Inhibitors are colored by secondary structure assignment: 3_{10} helix in *red*, strands in *blue*, and loops in *green*. The trypsin molecules are colored *magenta*. The cysteine residues are shown as *sticks* in *orange*, and P1 residues are in *yellow*. *C*, overall structure of API-A, with β-strands and loops numbered sequentially. Two P1 residues and six cysteine residues are highlighted as *yellow sticks*. *D*, structure-based multialignment of API-A and other Kunitz-type inhibitors, the Cα RMSD values to API-A were calculated: the winged bean chymotrypsin inhibitor (*WCI*, Protein Data Bank code 1EYL; RMSD, 2.19 Å), the *Delonix regia* seeds trypsin inhibitor (*DrTI*, Protein Data Bank code 1R8N; RMSD, 2.38 Å), the *B. bauhinioides Cruzipain* inhibitor (*BbCI*, Protein Data Bank code 2GZB; RMSD, 2.56 Å), and the STI (Protein Data Bank code 1AVW; RMSD, 2.17 Å). The secondary structure elements are indicated under the primary sequences, and the conserved residues are shaded in *blue*. The P1 residues in API-A are highlighted by *red stars*. All figures of protein structure were produced with PyMOL (46), and the multialignment was generated based on the structural superposition (47).

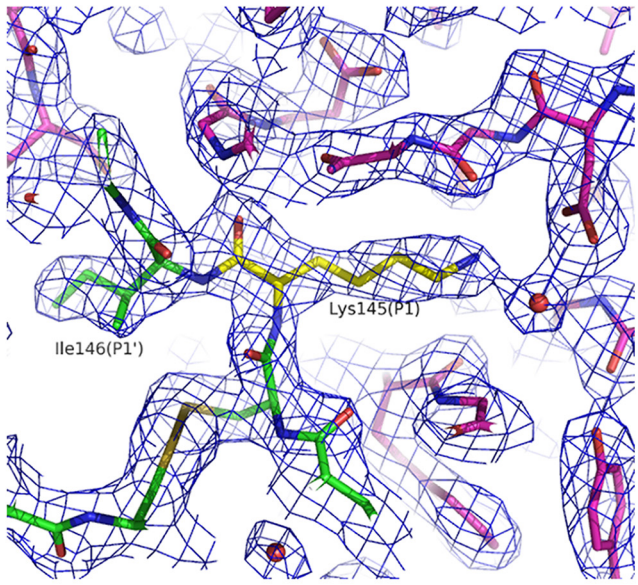
could have evolved from variable loops of ancestral Kunitz-type proteins (40).

Reactive Site 1—It has been predicted that Leu^{87} is a P1 residue that is more favorable for chymotrypsin than for trypsin (30), but this residue is completely embedded in the S1 pocket of trypsin in the present structure (Fig. 2*A*). To verify this crystallographic observation, Leu^{87} was mutated to proline, leading to a significant decrease in the inhibitory activity. The extrapolated molar ratio of inhibitor/protease increased from ~0.7 for the wild-type enzyme to ~1.0 for the L87P variant (Fig. 2*C*). This indicated that the inhibitor contains only the second reactive site (RS2) in which the P1 residue Lys^{145} is capable of binding one molecule of trypsin; this finding is in agreement with previously reported results (30).

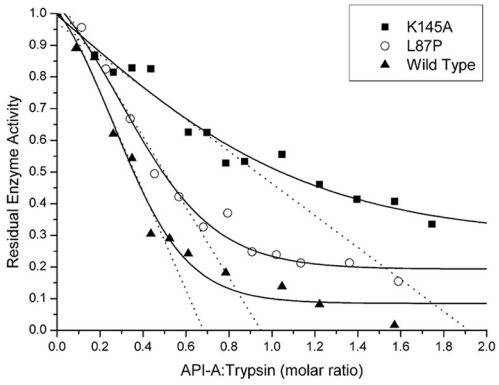
Superimposing of API-A with serine protease inhibitors of known structure revealed that the conformation of the RS1 loop was distinct from the typical canonical conformation of serine protease inhibitors (20). Among these structures, only the squash trypsin inhibitor MCTI-A (Protein Data Bank code 1F2S) could be partly superimposed on the residues from P4 to P1. To present more quantitative data, the dihedral angles (Φ/Ψ) of the residues in the reactive sites of both API-A and MCTI-A are listed in Table 2. The P1 Leu^{87} residue of API-A adopts the main chain conformation of the 3_{10} helix similar to



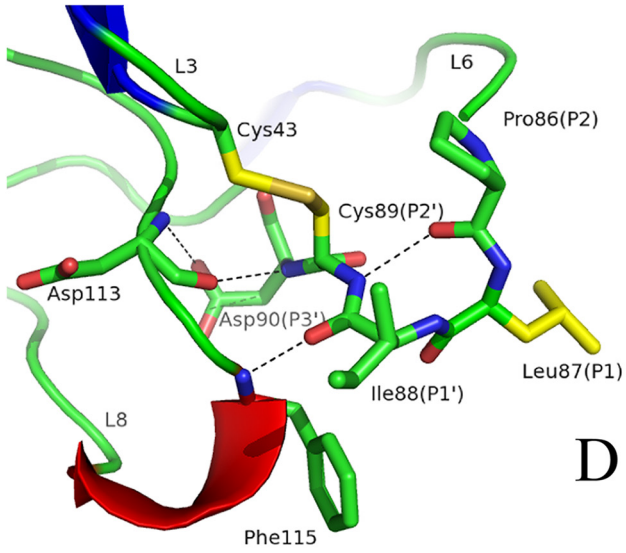
A



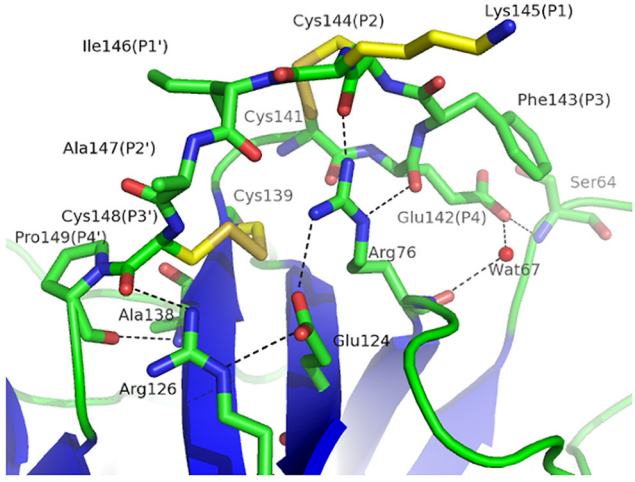
B



C



D



E

TABLE 2
Dihedral angles (Φ/Ψ) of residues of reactive sites
MCTI-A, squash trypsin inhibitor, Protein Data Bank code 1F2S (43); STI, Protein Data Bank code 1AVW (10).

	API-A (reactive site 1)	MCTI-A	API-A (reactive site 2)	STI
P4	(Pro^{84})-64/141	(Ile^{2})-92/109	(Glu^{142})-79/-52	(Ser^{60})-114/144
P3	(Val^{85})-155/146	(Cys^{3})-131/110	(Phe^{143})-114/-2	(Pro^{61})-58/-35
P2	(Pro^{86})-68/143	(Pro^{4})-57/142	(Cys^{144})-81/151	(Tyr^{62})-55/139
P1	(Leu^{87})-68/-14	(Arg^{5})-88/40	(Lys^{145})-94/27	(Arg^{63})-89/40
P1′	(Ile^{88})-69/-14	(Ile^{6})-92/141	(Ile^{146})-63/150	(Ile^{64})-85/148
P2′	(Cys^{89})-110/138	(Trp^{7})-80/123	(Ala^{147})-94/115	(Arg^{65})-68/-37
P3′	(Asp^{90})-96/-175	(Met^{8})-151/133	(Cys^{148})-79/151	(Phe^{66})-120/155

that of other inhibitors, whereas P1′ Ile^{88} continues in this conformation instead of the polyproline II conformation observed in the P1′ of typical inhibitors. Generally, the main chain atoms of P1′ interact directly with spacer residues as in the case of STI-Kunitz family inhibitors (10). Alternatively, they may interact indirectly with neighboring residues via a water molecule as in the case of the squash seed inhibitor family inhibitors (43). In contrast, residues P1′ to P3′ of API-A adopt a novel conformation, whereas a hydrogen bond network bridges loops 6 and 8 (Ile^{88}-O–Phe^{115}-N, Asp^{90}-N–Asp^{113}-O, and Asp^{90}-Oδ2–Asp^{113}-N) (Fig. 2*D*); such a conformation has not been reported for other Kunitz-type inhibitors. Loop 6 becomes more rigid because of the intraloop hydrogen bond between Pro^{86}-O and Cys^{89}-N. Moreover, the interloop disulfide bond Cys^{43}–Cys^{89} between loops 6 and 3 further stabilizes the first reactive site loop and thus also contributes to the novel conformation of RS1.

Reactive Site 2—Unexpectedly, the electron density for the P1 residue of RS2 matches perfectly with that of Lys^{145} (Fig. 2*B*), which has not been noted previously (29, 30). To confirm this assignment, we overexpressed and purified the K145A variant. In comparison with wild-type API-A, the inhibitory activity of the K145A variant dramatically decreased, resulting in a deduced inhibitor/protease ratio of ~1.9 (Fig. 2*C*). Because the presence of neither the P1 residue Leu^{87} nor Ala^{145} at the reactive site of the K145A variant is favorable for trypsin binding, the variant shows rather weak inhibitory activity.

RS2 adopts a canonical conformation even though the presence of two disulfide bonds (Cys^{139}–Cys^{148} and Cys^{141}–Cys^{144}) makes it different from other protease inhibitors that contain only one disulfide bond. The conformations of most residues in RS2 show no significant differences from those in STI, resulting in an RMSD value of 0.78 Å for the Cα atoms and similar dihedral angles (Table 2). It has been shown that the additional disulfide bond (Cys^{141}–Cys^{144}) is important for the thermostability but is not essential for the inhibitory activity (28). Notably, the three important spacer residues Arg^{76}, Glu^{124}, and Arg^{126} adopt a unique interaction pattern that is different from the patterns in other Kunitz-type inhibitors. Arg^{76} forms two direct hydrogen bonds with P2 (Arg^{76}-Nη2–Cys^{144}-O) and P4 (Arg^{76}-Nϵ–Glu^{142}-O) and one indirect hydrogen bond with P4 (Arg^{76}-O–Glu^{142}-Oϵ2) that is mediated by a water molecular, Wat^{67} (Fig. 2*E*). The hydrogen bonds contributed by Arg^{76} are important for stabilizing RS2, thus making it indispensable for maintaining the inhibitory activity. This is why Arg^{76} was earlier misassigned as the P1 residue on the basis of site-directed mutagenesis studies (29, 30). The two other spacer residues Glu^{124} and Arg^{126} also participate in two salt bridges and one hydrogen bond (Glu^{124}-Oϵ1–Arg^{76}-Nη1, Glu^{124}-Oϵ2–Arg^{126}-Nϵ, and Arg^{126}-Nη2–Cys^{148}-O). In addition, P4 Glu^{142} and P4′ Pro^{149} form hydrogen bonds with the neighboring residues Ser^{64} (Ser^{64}-N–Glu^{142}-Oϵ2) and Ala^{138} (Ala^{138}-N–Pro^{149}-O), respectively (Fig. 2*E*). This interaction network in addition to two disulfide bonds gives loop 10 a well defined conformation that aids in the inhibitory activity.

The Two Interfaces—The interfaces formed by the binding surfaces from both the protease and its inhibitor are important for the enhanced specificity and affinity of inhibitors. Therefore, detailed analyses of the residues at the interfaces could help in the development of more potent protease inhibitors.

Interface 1—Despite possessing a novel conformation, the interaction pattern at RS1 is conserved. The main contributions to inhibitor binding specificity and affinity are from the RS1 residues, and these residues are supplemented by the residues surrounding the reactive site. Apart from the P1 residue Leu^{87} that fits snugly into the S1 pocket of trypsin-1, the binding at interface 1 is also aided by the formation of a typical intermolecular antiparallel β-sheet stabilized by three main chain hydrogen bonds (Fig. 3*A* and Table 3).

The residues surrounding RS1 also contribute to the specific recognition toward trypsin-1 via hydrophobic and hydrophilic interactions. The residue Phe^{115} located in the vicinity of RS1 interacts with the hydrophobic patch of trypsin-1 composed of residues Tyr^{37}′, Tyr^{146}′, and Phe^{39}′, similar to the interaction observed when KD1 (Kunitz domain 1 of tissue factor pathway inhibitor-2) is complexed with trypsin (9). Moreover, an intermolecular hydrogen bond network between loop 1 and trypsin-1 (Asn^{14}-Nδ2–Ser^{142}′-Oγ, Gly^{17}-O–Gln^{214}′-Nϵ2, Asn^{18}-Oδ1–Lys^{215}′-N, and Tyr^{19}-N–Gln^{214}′-Oϵ1) further reinforces the enzyme-inhibitor interaction.

Interface 2—As shown in Fig. 3*B*, the interface between RS2 and trypsin-2 exhibits the classic protease-inhibitor interactions. In agreement with the canonical binding mode, the carbonyl group of the scissile peptide bond between P1 Lys^{145} and P1′ Ile^{146} is embedded in the oxyanion hole. However, one of

FIGURE 2. **The two P1 residues and reactive sites.** *A* and *B*, the $2F_o - F_c$ electron density maps around P1 residues of Leu^{87} (*A*) and Lys^{145} (*B*, 1 σ level). API-A was colored in *green* except for P1 (Leu^{87} and Lys^{145}) in *yellow* and trypsin molecules in *magenta*. *C*, inhibitory activity assays of the wild-type API-A and two variants (K87P and K145A). The assays were performed with 50 nM trypsin. *D* and *E*, RS1 and 2. The hydrogen bonds and ionic interaction are shown as *dashed lines*.

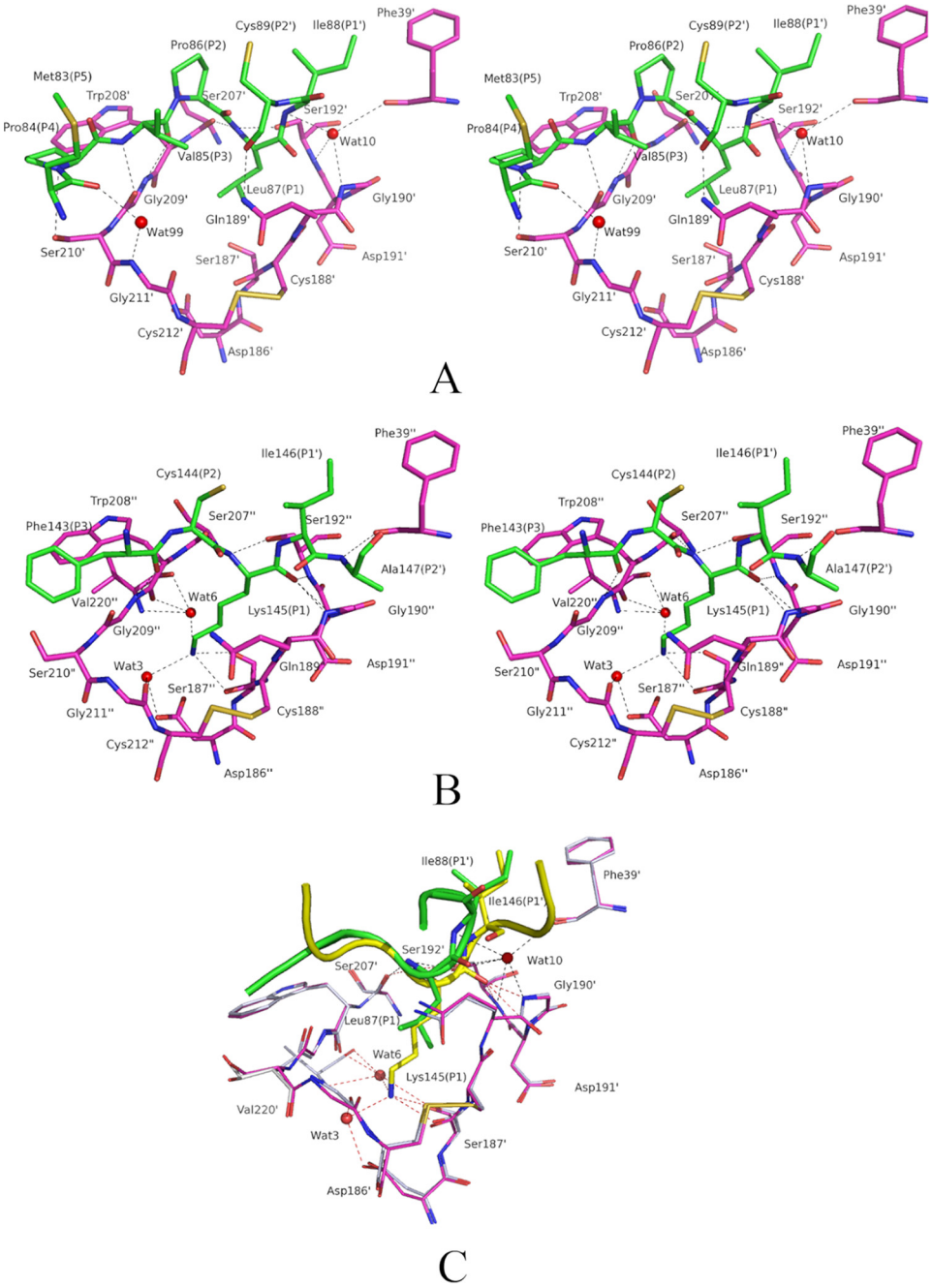


FIGURE 3. **The stereo views of the two interfaces.** *A* and *B*, interfaces 1 and 2 between reactive sites of API-A (in *green*) and trypsin (in *magenta*). The residues of trypsin-1 and -2 are labeled with a prime and a double-prime, respectively. All hydrogen bonds are shown as *dashed lines*. The water molecules (Wat^{3}, Wat^{6}, and Wat^{99}) are shown as *red spheres*. *C*, superposition of two interfaces. The residues of RS1 and 2 are colored in *green* and *yellow*, respectively. The P1 and P1′ residues are represented as *sticks*, and the reactive site loops are shown as a cartoon. All of the hydrogen bonds are shown as *dashed lines* (interface 1 in *black* and interface 2 in *red*).

Comparison of the Two Interfaces—By superimposing trypsin-1 and -2, the differences in the two interfaces become apparent (Fig. 3*C*). As expected, the two trypsin molecules are well aligned with an RMSD of 0.31 Å for the overall structure and 0.26 Å for the active site. The conformations of the two reactive site loops of API-A show significant differences. The side chain of P1 Lys^{145} at interface 2 forms a salt bridge with Asp^{186} and is also in contact with the main chain atoms of four residues including Gly$^{190''}$, Asp$^{191''}$, Ser$^{192''}$, and Ser$^{207''}$ (Table 3 and Fig. 3*B*). Taken together, P1 Lys^{145} in API-A contributes to the majority of the highly conserved primary contacts, which are known to be pivotal for the inhibitory specificity (20). In contrast, the carboxyl oxygen atom of P1 Leu^{87} shifts by ~2.43 Å in comparison with that of P1 Lys^{145}, resulting in breakage of the major main chain contacts with trypsin-1. This leaves only two main chain hydrogen bonds (Leu^{87}-N–Ser$^{207'}$-O and Leu^{87}-O–Gly$^{190'}$-N) (Table 3 and Fig. 3*A*). Instead, the main chain nitrogen atom of P1′ Ile^{88} forms three indirect hydrogen bonds with Phe$^{39'}$, Gly$^{190'}$, and Ser$^{192'}$; these are mediated by the well ordered water molecule Wat^{10}. The nitrogen atom and carbonyl oxygen of P3 Val^{85} form two main chain hydrogen bonds with Gly$^{209'}$, whereas P5 Met^{83} forms an indirect hydrogen bond with Gly$^{211'}$ (Table 3 and Fig. 3*A*). These hydrogen bonds at RS1 complement the nonfavored P1 residue Leu^{87} of API-A for binding to trypsin. In contrast, API-B has an Arg as the corresponding P1 residue that favors its simultaneous binding to two molecules of trypsin (26, 30). Moreover, the hydrophobic interactions and hydrogen bond network beyond RS1 are also involved in the formation of interface 1, resulting in a much larger area than that of interface 2 (1855 *versus* 1295 Å^2). Thus, the residues surrounding P1 Leu^{87} rather than P1 itself mainly contribute to the inhibitory activity, leading to the broad specificity of RS1.

the classic hydrogen bonds (44) (P3 Phe^{143}-N–Gly$^{209''}$-O) is absent because of the distorted dihedral angle of P3 Phe^{143}. As shown in Table 3, the main chain hydrogen bonds are mainly contributed by P1 Lys^{145} (the residues of trypsin-2 are labeled with a double-prime). In addition to direct hydrogen bonds, the Nζ atom of P1 Lys^{145} forms five indirect hydrogen bonds mediated by two water molecules Wat^{3} and Wat^{6}, which are highly conserved in most protease inhibitors that have lysine as the P1 residue (10, 45).

Kunitz-type inhibitors are characterized by loops that vary in length, sequence, and conformation. These loops form the

TABLE 3

Interactions between API-A and trypsins

The residues of trypsin-1 are labeled with a prime, and those of trypsin-2 are labeled with a double-prime.

API	Trypsin	Water	Distance
			Å
Interface 1: Reactive site 1 against tryspin-1			
Asn^{14}-Nδ2	$Ser^{141\prime}$-O		3.52
	$Ser^{142\prime}$-Oγ		3.30
Gly^{17}-O	$Gln^{214\prime}$-Nϵ2		3.27
Asn^{18}-Oδ1	$Lys^{215\prime}$-N		2.29
Tyr^{19}-N	$Gln^{214\prime}$-Oϵ1		3.13
Met^{83}-O		Wat^{99}	2.68
	$Gly^{211\prime}$-N[a]	Wat^{99}	2.66
Val^{85}-N	$Gly^{209\prime}$-O		2.89
Val^{85}-O	$Gly^{209\prime}$-N		3.15
Leu^{87}-N	$Ser^{207\prime}$-O		3.10
	$Ser^{192\prime}$-Oγ		3.14
Leu^{87}-O	$Gly^{190\prime}$-N		3.31
Ile^{88}-N	$Ser^{192\prime}$-Oγ		3.40
Ile^{88}-N		Wat^{10}	3.28
	$Phe^{39\prime}$-O[a]	Wat^{10}	3.19
	$Gly^{190\prime}$-N[a]	Wat^{10}	3.05
	$Ser^{192\prime}$-N[a]	Wat^{10}	3.20
Cys^{89}-O	$Gln^{189\prime}$-Nϵ2		3.03
Tyr^{116}-Oη<ρυλ;1>	$Lys^{58\prime}$-Nζ[b]		2.61
Interface 2: Reactive site 2 against trypsin-2			
Phe^{143}-O	$Gly^{209\prime\prime}$-N		3.08
Lys^{145}-N	$Ser^{207\prime\prime}$-O		2.98
	$Ser^{192\prime\prime}$-Oγ		2.82
Lys^{145}-O	$Gly^{190\prime\prime\prime}$-N		2.48
	$Asp^{191\prime\prime}$-N		3.27
	$Ser^{192\prime\prime}$-N		3.12
Lys^{145}-Nζ	$Ser^{187\prime\prime}$-O		2.83
	$Asp^{186\prime\prime}$-Oδ1[b]		3.20
		Wat3	2.85
	$Asp^{186\prime\prime}$-Oδ2[a]		2.55
	$Gly^{211\prime\prime}$-O[a]		2.98
		Wat6	2.85
	Ser^{187}-Oγ[a]		3.22
	$Val^{220\prime\prime}$-O[a]		2.66
	$Trp^{208\prime\prime}$-O[a]		2.85
Ala^{147}-N	$Phe^{39\prime\prime}$-O		3.11

[a] Interaction mediated by water molecules.
[b] Ionic interaction.

major part of the protein surface and are responsible for the diversity in inhibitory activities that has arisen with evolution (40). Because of an unknown evolutionary mechanism, API-A has two reactive site loops that are distinct from each other. More notably, RS1 is quite different from the classic Kunitz-type inhibitors because of its unique conformation and extraordinary recognition pattern for trypsin. Although they share a classic conformation and binding mode, RS2 and its P1 residue were once misassigned in previous studies; however, both have been clearly identified in the present report.

The abundance of API-A in the tubers of arrowhead, its strong inhibitory effect reflected in its nanomolar scale inhibition constant (K_i), and the distinct inhibitory activities of the two reactive sites present in this enzyme make API-A a promising therapeutic PI with potential clinical applications (30). Moreover, the present crystal structure would enable us to engineer the two reactive site loops for further improving the affinity and specificity against various proteases.

Acknowledgments—We are grateful to Prof. Joel Janin (Université Paris-Sud, Paris, France) and Dr. Di Xia (NCI, National Institutes of Health) for critical reading of the manuscript.

REFERENCES

1. Laskowski, M., Jr., and Kato, I. (1980) *Annu. Rev. Biochem.* **49,** 593–626
2. Leung, D., Abbenante, G., and Fairlie, D. P. (2000) *J. Med. Chem.* **43,** 305–341
3. Kobayashi, H., Yagyu, T., Inagaki, K., Kondo, T., Suzuki, M., Kanayama, N., and Terao, T. (2004) *Cancer* **100,** 869–877
4. Ishikura, H., Nishimura, S., Matsunami, M., Tsujiuchi, T., Ishiki, T., Sekiguchi, F., Naruse, M., Nakatani, T., Kamanaka, Y., and Kawabata, A. (2007) *Life Sci.* **80,** 1999–2004
5. Bunnage, M. E., Blagg, J., Steele, J., Owen, D. R., Allerton, C., McElroy, A. B., Miller, D., Ringer, T., Butcher, K., Beaumont, K., Evans, K., Gray, A. J., Holland, S. J., Feeder, N., Moore, R. S., and Brown, D. G. (2007) *J. Med. Chem.* **50,** 6095–6103
6. Ng, T. B., Lam, S. K., and Fong, W. P. (2003) *Biol. Chem.* **384,** 289–293
7. Nielsen, P. K., Bønsager, B. C., Fukuda, K., and Svensson, B. (2004) *Biochim. Biophys. Acta* **1696,** 157–164
8. Onesti, S., Brick, P., and Blow, D. M. (1991) *J. Mol. Biol.* **217,** 153–176
9. Schmidt, A. E., Chand, H. S., Cascio, D., Kisiel, W., and Bajaj, S. P. (2005) *J. Biol. Chem.* **280,** 27832–27838
10. Song, H. K., and Suh, S. W. (1998) *J. Mol. Biol.* **275,** 347–363
11. Sweet, R. M., Wright, H. T., Janin, J., Chothia, C. H., and Blow, D. M. (1974) *Biochemistry* **13,** 4212–4228
12. Vallée, F., Kadziola, A., Bourne, Y., Juy, M., Rodenburg, K. W., Svensson, B., and Haser, R. (1998) *Structure* **6,** 649–659
13. Krauchenco, S., Pando, S. C., Marangoni, S., and Polikarpov, I. (2003) *Biochem. Biophys. Res. Commun.* **312,** 1303–1308
14. Ravichandran, S., Dasgupta, J., Chakrabarti, C., Ghosh, S., Singh, M., and Dattagupta, J. K. (2001) *Protein Eng.* **14,** 349–357
15. Strukelj, B., Pungercar, J., Mesko, P., Barlic-Maganja, D., Gubensek, F., Kregar, I., and Turk, V. (1992) *Biol. Chem. Hoppe-Seyler* **373,** 477–482
16. Ishikawa, A., Ohta, S., Matsuoka, K., Hattori, T., and Nakamura, K. (1994) *Plant Cell Physiol.* **35,** 303–312
17. Yang, H. L., Luo, R. S., Wang, L. X., Zhu, D. X., and Chi, C. W. (1992) *J. Biochem.* **111,** 537–545
18. Dattagupta, J. K., Podder, A., Chakrabarti, C., Sen, U., Mukhopadhyay, D., Dutta, S. K., and Singh, M. (1999) *Proteins* **35,** 321–331
19. Rawlings, N. D., Tolle, D. P., and Barrett, A. J. (2004) *Biochem. J.* **378,** 705–716
20. Bode, W., and Huber, R. (1992) *Eur. J. Biochem.* **204,** 433–451
21. Capaldi, S., Perduca, M., Faggion, B., Carrizo, M. E., Tava, A., Ragona, L., and Monaco, H. L. (2007) *J. Struct. Biol.* **158,** 71–79
22. Koepke, J., Ermler, U., Warkentin, E., Wenzl, G., and Flecker, P. (2000) *J. Mol. Biol.* **298,** 477–491
23. Lin, G., Bode, W., Huber, R., Chi, C., and Engh, R. A. (1993) *Eur. J. Biochem.* **212,** 549–555
24. Park, E. Y., Kim, J. A., Kim, H. W., Kim, Y. S., and Song, H. K. (2004) *J. Mol. Biol.* **343,** 173–186
25. Raj, S. S., Kibushi, E., Kurasawa, T., Suzuki, A., Yamane, T., Odani, S., Iwasaki, Y., Yamane, T., and Ashida, T. (2002) *J. Biochem.* **132,** 927–933
26. Chang, H. Y., Chi, C. W., and Lo, C. Q. (1979) *Sci. Sin.* **22,** 1443–1454
27. Luo, M. J., Lu, W. Y., and Chi, C. W. (1997) *J. Biochem.* **121,** 991–995
28. Xie, Z. W., Luo, M. J., and Chi, C. W. (1996) *Sci. Sin.* **28,** 700–702
29. Xie, Z. W., Luo, M. J., Xu, W. F., and Chi, C. W. (1997) *Biochemistry* **36,** 5846–5852
30. Li, J., Ruan, K. C., and Chi, C. W. (2002) *Sci. Sin.* **34,** 662–666
31. Kunkel, T. A. (1985) *Proc. Natl. Acad. Sci. U.S.A.* **82,** 488–492
32. Mizobata, T., Akiyama, Y., Ito, K., Yumoto, N., and Kawata, Y. (1992) *J. Biol. Chem.* **267,** 17773–17779
33. Bartels, K., and Klein, C. (2003) *The AUTOMAR User's Guide,* Version 1.04. MAR Research; GmbH; http://www.marresearch.com/automar/
34. Collaborative Computational Project, Number 4 (1994) *Acta Crystallogr. D Biol. Crystallogr.* **50,** 760–763
35. Emsley, P., and Cowtan, K. (2004) *Acta Crystallogr. D Biol. Crystallogr.* **60,** 2126–2132
36. Davis, I. W., Leaver-Fay, A., Chen, V. B., Block, J. N., Kapral, G. J., Wang, X., Murray, L. W., Arendall, W. B., 3rd, Snoeyink, J., Richardson, J. S., and

Richardson, D. C. (2007) *Nucleic Acids Res.* **35,** 375–383
37. Lo Conte, L., Chothia, C., and Janin, J. (1999) *J. Mol. Biol.* **285,** 2177–2198
38. Song, H. K., Kim, Y. S., Yang, J. K., Moon, J., Lee, J. Y., and Suh, S. W. (1999) *J. Mol. Biol.* **293,** 1133–1144
39. Li de la Sierra, I., Quillien, L., Flecker, P., Gueguen, J., and Brunie, S. (1999) *J. Mol. Biol.* **285,** 1195–1207
40. Murzin, A. G., Lesk, A. M., and Chothia, C. (1992) *J. Mol. Biol.* **223,** 531–543
41. Araújo, A. P., Hansen, D., Vieira, D. F., Oliveira, C., Santana, L. A., Beltramini, L. M., Sampaio, C. A., Sampaio, M. U., and Oliva, M. L. (2005) *Biol. Chem.* **386,** 561–568
42. De Meester, P., Brick, P., Lloyd, L. F., Blow, D. M., and Onesti, S. (1998) *Acta Crystallogr. D Biol. Crystallogr.* **54,** 589–597
43. Zhu, Y., Huang, Q., Qian, M., Jia, Y., and Tang, Y. (1999) *J. Protein Chem.* **18,** 505–509
44. Fodor, K., Harmat, V., Neutze, R., Szilágyi, L., Gráf, L., and Katona, G. (2006) *Biochemistry* **45,** 2114–2121
45. Helland, R., Berglund, G. I., Otlewski, J., Apostoluk, W., Andersen, O. A., Willassen, N. P., and Smalås, A. O. (1999) *Acta Crystallogr. D Biol. Crystallogr.* **55,** 139–148
46. DeLano, W. L. (2002) *The PyMOL User's Manual*, Palo Alto, CA
47. Bauer, R. A., Bourne, P. E., Formella, A., Frommel, C., Gille, C., Goede, A., Guerler, A., Hoppe, A., Knapp, E. W., Poschel, T., Wittig, B., Ziegler, V., and Preissner, R. (2008) *Nucleic Acids Res.* 36 **(Web Server issue)**, W47-W54

Protein Structure and Folding:

The Ternary Structure of the Double-headed Arrowhead Protease Inhibitor API-A Complexed with Two Trypsins Reveals a Novel Reactive Site Conformation

Rui Bao, Cong-Zhao Zhou, Chunhui Jiang, Sheng-Xiang Lin, Cheng-Wu Chi and Yuxing Chen

J. Biol. Chem. 2009, 284:26676-26684.
doi: 10.1074/jbc.M109.022095 originally published online July 28, 2009
