electronic reprint

STRUCTURAL BIOLOGY

ISSN: 2059-7983
journals.iucr.org/d

Crystal structures of the complex of a kallikrein inhibitor from *Bauhinia bauhinioides* with trypsin and modeling of kallikrein complexes

Mi Li, Jaroslav Srp, Alla Gustchina, Zbigniew Dauter, Michael Mares and Alexander Wlodawer

Acta Cryst. (2019). D**75**, 56–69

IUCr Journals
CRYSTALLOGRAPHY JOURNALS ONLINE

Copyright © International Union of Crystallography

Author(s) of this paper may load this reprint on their own web site or institutional repository provided that this cover page is retained. Republication of this article or its storage in electronic databases other than as specified above is not permitted without prior permission in writing from the IUCr.

For further information see http://journals.iucr.org/services/authorrights.html

STRUCTURAL BIOLOGY

ISSN 2059-7983

Crystal structures of the complex of a kallikrein inhibitor from *Bauhinia bauhinioides* with trypsin and modeling of kallikrein complexes

Mi Li,[a,b] Jaroslav Srp,[c,d] Alla Gustchina,[a] Zbigniew Dauter,[e] Michael Mares[c] and Alexander Wlodawer[a]*

[a]Macromolecular Crystallography Laboratory, Center for Cancer Research, National Cancer Institute, Frederick, MD 21702, USA, [b]Basic Science Program, Leidos Biomedical Research, Frederick National Laboratory for Cancer Research, Frederick, MD 21702, USA, [c]Institute of Organic Chemistry and Biochemistry, Czech Academy of Sciences, 166 10 Prague, Czech Republic, [d]Department of Biochemistry, Faculty of Science, Charles University, 128 40 Prague, Czech Republic, and [e]Synchrotron Radiation Research Section, Macromolecular Crystallography Laboratory, NCI, Argonne National Laboratory, Argonne, IL 60439, USA. *Correspondence e-mail: wlodawer@nih.gov

Received 24 August 2018
Accepted 19 November 2018

Edited by R. McKenna, University of Florida, USA

Keywords: proteases; inhibitors; kallikreins; protein complexes; *Bauhinia bauhinioides*; coagulation; fibrinolysis; inflammation; BbKI.

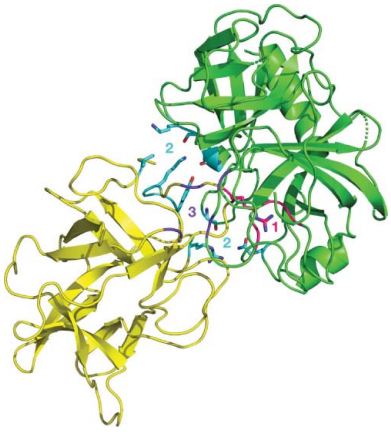
PDB references: BbKI, complex with bovine trypsin, space group $P6_4$, 6dwh; space group $P2_1$, 6dwu; L55R mutant of BbKI, complex with bovine trypsin, 6dwf

Supporting information: this article has supporting information at journals.iucr.org/d

Structures of a recombinant Kunitz-type serine protease inhibitor from *Bauhinia bauhinioides* (BbKI) complexed with bovine trypsin were determined in two crystal forms. The crystal structure with the L55R mutant of BbKI was determined in space group $P6_4$ at 1.94 Å resolution and that with native BbKI in the monoclinic space group $P2_1$ at 3.95 Å resolution. The asymmetric unit of the latter crystals contained 44 independent complexes, thus representing one of the largest numbers of independent objects deposited in the Protein Data Bank. Additionally, the structure of the complex with native BbKI was determined at 2.0 Å resolution from $P6_4$ crystals isomorphous to those of the mutant. Since BbKI has previously been found to be a potent inhibitor of the trypsin-like plasma kallikrein, it was also tested against several tissue kallikreins. It was found that BbKI is a potent inhibitor of human tissue kallikrein 4 (KLK4) and the chymotrypsin-like human tissue kallikrein 7 (KLK7). Structures of BbKI complexed with the catalytic domain of human plasma kallikrein were modeled, as well as those with KLK4 and KLK7, and the structures were analyzed in order to identify the interactions that are responsible for inhibitory potency.

1. Introduction

The Kunitz family of plant protease inhibitors shares its β-trefoil fold with a variety of proteins with different biological functions, including lectins, interleukins and fibroblast growth factors, as well as fragments of DNA-binding and actin cross-linking proteins (Renko *et al.*, 2012). The inhibitor family was named after the discoverer of the soybean trypsin inhibitor (STI), Moses Kunitz, who isolated and crystallized it (Kunitz, 1947). The inhibitors of the Kunitz type are usually (but are not always) single-chain proteins containing 160–180 amino-acid residues, although they may also form domains of larger proteins. These inhibitors usually contain one or two disulfides, but some have no cysteine residues (Oliva *et al.*, 2010; Oliva & Sampaio, 2008). The structures of a number of Kunitz-type inhibitors have been determined to date (Renko *et al.*, 2012). The family members contain a very similar core with pseudo-threefold symmetry, whereas they differ quite significantly in the structure of the loops connecting the central β-strands. The functional properties of these inhibitors are largely defined by the structures of their variable parts.



The inhibitory properties of the plant-derived Kunitz-type inhibitors have been studied very extensively. They have been

© 2019 International Union of Crystallography

electronic reprint

determined for enzymes involved in digestive processes, such as trypsin and chymotrypsin; for a number of blood-clotting enzymes, such as plasma kallikrein, factor XIIa, factor Xa and thrombin; and for enzymes involved in inflammatory processes, such as elastase (Batista *et al.*, 1996; Souza-Pinto *et al.*, 1996; Oliva & Sampaio, 2008; Vadivel *et al.*, 2014; Odei-Addo *et al.*, 2014).

BbKI is a Kunitz-type inhibitor that was originally isolated from the seeds of *Bauhinia bauhinioides*. This 18 kDa protein contains 165 amino acids with only one free cysteine and forms no disulfides. Originally designated as BbTI-II (Oliva *et al.*, 1999), it has been shown to be a potent inhibitor of trypsin, as well as of plasmin and human and porcine plasma kallikreins, with no detectable inhibition of thrombin or factor Xa (Oliva *et al.*, 1999, 2001).

BbKI, or its recombinant form rBbKI (in the following both are referred to as BbKI), has become an attractive molecule for studying pathological models of the circulatory system, since this protein acts on enzymes involved in coagulation, fibrinolysis and inflammation. One of the targets of BbKI is human plasma kallikrein, an enzyme that participates in the processes of blood coagulation, platelet aggregation and muscle contraction (Botos & Wlodawer, 2007; Pampalakis & Sotiropoulou, 2007; Turk, 2006). Indeed, BbKI was effective in inhibiting the viability of tumor cell lines (Nakahata *et al.*, 2013). Other important aspects of the inhibitor have been reported by Brito *et al.* (2014), who found that BbKI may prolong the formation of blood clots *in vitro* and that it exhibits antithrombotic activity in *in vivo* models of venous and arterial thrombosis. To the best of our knowledge, BbKI is the only plant-derived inhibitor of both plasma and tissue kallikreins (Oliva *et al.*, 2001, 2010). Different kallikreins have been implicated in a number of pathological processes, and interference with their activity by specific inhibitors is a promising new area of drug discovery (Sotiropoulou & Pampalakis, 2012; Prassas *et al.*, 2015). For example, kallikrein 7 (KLK7) is overexpressed in skin inflammation and atopic dermatitis, and its inhibition may prevent the penetration of skin by microorganisms and allergens (Murafuji *et al.*, 2017). KLK3 and KLK4 are upregulated in prostate cancer and both have been targeted for drug development (Mavridis *et al.*, 2014; Cereda *et al.*, 2015; Riley *et al.*, 2016). A detailed understanding of their mode of inhibition may help in such pursuits.

In this study, we evaluated the inhibitory properties of BbKI towards several kallikreins. We determined a high-resolution crystal structure of BbKI complexed with bovine trypsin, which was subsequently used to build models of the inhibitor complexes with selected kallikreins. These models were analyzed in order to reveal the important interactions that might be responsible for the differences in the inhibitory potency of BbKI towards various kallikreins. Since the initially grown crystal form of the complex of BbKI with trypsin contained a very large number of molecules in the asymmetric unit, we prepared an L55R mutant of BbKI, changing a hydrophobic surface residue that was not expected to be involved in enzyme–inhibitor interactions to a hydrophilic residue. This mutant protein, which was expected to be less prone to aggregation, was used in some of the experiments that are discussed here.

2. Materials and methods

2.1. Preparation and purification of BbKI and its complex with bovine trypsin

The ORF for BbKI was cloned from the previously used construct for BbKI (Araújo *et al.*, 2005) with the primers (Operon) p300-BbKIf (5′-GAGAACCTGTACTTCCAGTCGGTCGTTGTCGACACCAATGG-3′) and p301-BbKIr (5′-GGGGACCACTTTGTACAAGAAAGCTGGGTTATTACTCATCAGTTGCCTTCCTTATC-3′) using KOD Hot Start DNA polymerase (Novagen). The product was gel-purified and used as a template for a second PCR with the primers PE277 (5′-GGGGACAAGTTTGTACAAAAAAGCAGGCTCGGAGAACCTGTACTTCCAG-3′) and p301-BbKIr. This amplicon was inserted into pDONR201 using Gateway technology (Invitrogen) to generate the entry clone pZD226. The entry clone was then recombined with the destination vector pCB1426 to construct the His_6-MBP-TEV-BbKI fusion-protein expression vector pZD228. To facilitate protein purification, the pZD228 plasmid was modified using the QuikChange site-directed mutagenesis kit with the primers p303 (5′-GAGAACCTGTACTTCCAGGGTGGCGGTGGCGGTGGCTCGGTCGTTGTCGACACCAATGG-3′) and p305 (5′-CCATTGGTGTCGACAACGACCGAGCCACCGCCACCGCCACCCTGGAAGTACAGGTTCTC-3′). This generated the construct His_6-MBP-TEV-BbKI (pZD263) with six glycine residues inserted before the BbKI ORF. The plasmid was transformed into *Escherichia coli* Rosetta 2 (DE3) cells for expression. The cells were grown in Luria–Bertani medium supplemented with 25 µg ml^{-1} chloramphenicol (Invitrogen) at 37°C to an optical density $OD_{600\,nm}$ of 0.6, followed by induction of the expression of the fusion protein with isopropyl β-D-1-thiogalactopyranoside (IPTG; Invitrogen). IPTG was added to a final concentration of 1 m*M* and the culture was grown overnight at 16°C on an orbital shaker at 180 rev min^{-1}. Subsequently, the cells were harvested by centrifugation at 6000*g* for 20 min at 4°C (Sorvall Evolution RC, Thermo Scientific). The pellets were then resuspended in lysis buffer (50 m*M* Tris–HCl pH 8.0, 0.5 *M* NaCl, 0.5 m*M* $MgCl_2$, 0.5 m*M* $CaCl_2$) and lysed by adding Novagen BugBuster at a 1:50 volume ratio. After 40 min of lysis at 4°C, the lysis solution was centrifuged at 6000*g* for 30 min at 4°C. The supernatant was loaded onto an Ni–NTA affinity chromatography column, which was then eluted with an imidazole gradient (10–500 m*M*) in 20 m*M* Tris–HCl pH 7.5 buffer. Fractions containing the fusion protein were combined and dialyzed against 20 m*M* Tris–HCl pH 8.0, 200 m*M* NaCl buffer overnight to remove the imidazole. TEV protease (1 mg per 8 g of cells) was added to the dialyzed solution to cleave the MBP at 4°C overnight. As a consequence of the introduced TEV cleavage site in the expression vector, the BbKI contained six extra glycine residues at the N-terminus. The MBP (containing a His tag) was removed by a second run on an Ni–NTA column. The flow-

electronic reprint

through containing BbKI was concentrated and further purified by size-exclusion chromatography (Superdex 75 HR 10/30 column; GE Healthcare) in 20 m*M* Tris–HCl pH 7.5, 0.2 *M* NaCl.

To prepare the complex of BbKI with trypsin, equimolar amounts of BbKI and bovine pancreatic trypsin (Sigma, catalog No. 8003) were mixed and incubated on ice for 1 h in 20 m*M* Tris pH 7.5, 0.2 *M* NaCl. The mixture was then applied onto a Superdex 75 column pre-equilibrated with the same buffer. Fractions of 1 ml volume were collected at a flow rate of 0.5 ml min^{-1}. The identity of the samples pooled from the different peaks was verified by SDS–PAGE. The fractions corresponding to the BbKI–trypsin complex were selected, concentrated to between 10 and 25 mg ml^{-1} and stored at −80°C for crystallization trials.

2.2. BbKI mutagenesis

A mutation of Leu55 to arginine was introduced into pZD263 using the QuikChange site-directed mutagenesis kit with the primers p359 (5′-CTCACCACCGTCCCGGTCGTC CGGTTAGATTTGAATCC-3′) and p260 (5′-GGATTCA AATCTAACCGGACGACCGGGACGGTGGTGAG-3′) to generate pZD264. The mutated gene for the L55R mutant of BbKI (BbKI L55R) was expressed and the resulting protein was purified using a procedure analogous to that used for native BbKI. The BbKI L55R–trypsin complex was prepared and purified using the same protocol as for native BbKI.

2.3. Enzyme-activity and inhibition assays

Proteolytic activities were measured using selective fluorogenic peptide substrates (Table 1) containing the fluorescent leaving groups 7-amino-4-methylcoumarin (Amc) or 7-amino-4-carbamoylmethylcoumarin (Acc). The reaction mixture included enzyme and substrate at the concentrations indicated in Table 1, and 0.1 *M* Tris–HCl pH 8.0, 0.1% PEG 1500 containing 0.15 *M* NaCl (for most KLKs), 1 *M* NaCl (for KLK3) or 10 m*M* $CaCl_2$ (for trypsin and chymotrypsin). The kinetics of product release were continuously monitored using an Infinite M1000 microplate reader (Tecan) at excitation wavelengths of 360 or 380 nm and emission wavelengths of 465 or 460 nm for substrates with Amc or Acc groups, respectively. For inhibition measurements, the enzyme mixture was pre-incubated with the BbKI inhibitor (at up to 10 μ*M* concentration) for 10 min followed by the addition of substrate. Reaction rates were obtained at various inhibitor concentrations, and IC_{50} values were determined by nonlinear regression using the *GraFit* software (Erithacus Software). The K_i values were calculated using the equation for competitive inhibition, $K_i = IC_{50}/\{1 + ([S]/K_m)\}$; substrate concentrations [S] and K_m values are presented in Table 1 (K_m values were determined by nonlinear regression of substrate-dependent velocity curves or were obtained from literature data). Each measurement was performed in triplicate. The concentrations of the BbKI inhibitors and the substrates were determined by amino-acid analysis. Kallikreins were purchased from R&D Systems and were activated according to the manufacturer's protocol; other enzymes were from Sigma. Substrates were purchased from Bachem; Ac-KHLY-Acc was synthesized as described in Horn *et al.* (2018).

Table 1
Enzyme-activity assay conditions used for inhibition kinetics.

Enzyme	Enzyme concentration (n*M*)	Substrate	Substrate concentration (μ*M*)	K_m (μ*M*)
Plasma kallikrein	10	PFR-Amc	25	28.7
hKLK2	40	PFR-Amc	50	3.1†
hKLK3	40	Suc-AAPF-Amc	50	>250
hKLK4	1	Boc-VPR-Amc	50	5.9
hKLK5	0.2	Boc-VPR-Amc	100	101.1‡
mKLK5	0.2	Boc-VPR-Amc	100	48.3‡
hKLK7	0.2	Ac-KHLY-Acc	25	99.1‡
mKLK7	2	Ac-KHLY-Acc	25	101.5‡
hKLK14	0.2	Boc-VPR-Amc	30	29.0‡
Trypsin	0.6	Cbz-FR-Amc	25	67.8
Chymotrypsin	0.024	Suc-AAPF-Amc	25	41.5

† Fogaça *et al.* (2004). ‡ Horn *et al.* (2018).

2.4. Protein crystallization and X-ray data collection and processing

Two crystal forms of the BbKI–trypsin complex were grown. Monoclinic crystals (space group $P2_1$) were grown by the hanging-drop technique, starting from a protein sample concentrated to 21 mg ml^{-1} in 20 m*M* Tris, 0.2 *M* NaCl pH 7.5. The well solution contained 17.5% PEG 3350 at pH 8.0. Each 4 μl hanging drop consisted of 2 μl sample and 2 μl well solution and was equilibrated against 500 μl well solution. Crystals of the complex of the L55R mutant of BbKI with trypsin grew under the same conditions as the BbKI–trypsin crystals, with the only difference being that the starting sample concentration was 13 mg ml^{-1}. Hexagonal crystals of the BbKI–trypsin complex (space group $P6_4$) were grown by an analogous procedure, starting with a protein sample at 16.5 mg ml^{-1} concentration in the same buffer. However, the well solution contained 1.6 *M* ammonium sulfate pH 4.2. Each hanging drop consisted of 4 μl sample and 2 μl well solution and was equilibrated against 500 μl well solution.

Diffraction data were collected on the Southeast Regional Collaborative Access Team (SER-CAT) beamline 22-ID at the Advanced Photon Source, Argonne National Laboratory, USA. Single crystals were transferred to a cryoprotectant solution (mother liquor supplemented with 25% glycerol) for approximately 2 min and were then flash-cooled at 100 K in a stream of cold nitrogen gas. All three data sets were collected at a wavelength of 1.000 Å using a Rayonix 300HS detector. Diffraction data were indexed, integrated and scaled with *HKL*-2000 (Otwinowski & Minor, 1997). Data-processing statistics are shown in Table 2.

2.5. Structure determination and refinement

The structure of the hexagonal crystals of the BbKI L55R–trypsin complex was solved first by molecular replacement using *Phaser* (McCoy *et al.*, 2007). The structures of BbKI (PDB entry 4zot; Zhou *et al.*, 2015) and of bovine trypsin (PDB entry 5eg4; Hinkes *et al.*, 2016) were used independently

electronic reprint

Table 2
Data collection and structure refinement.

Values in parentheses are for the highest resolution shell.

	BbKI–trypsin complex	BbKI–trypsin complex	BbKI L55R–trypsin complex
Data collection			
Space group	$P2_1$	$P6_4$	$P6_4$
Unit-cell parameters			
a (Å)	137.2	207.81	207.30
b (Å)	483.9	207.81	207.30
c (Å)	137.0	107.17	107.17
α (°)	90	90	90
β (°)	116.8	90	90
γ (°)	90	120	120
Resolution (Å)	50.0–3.96 (4.03–3.96)	50.0–2.00 (2.03–2.00)	50.0–1.94 (1.97–1.94)
$\langle I\rangle/\langle\sigma(I)\rangle$	8.16 (3.05)	14.4 (0.66)	26.5 (0.68)
Completeness (%)	88.7 (37.3)	96.5 (71.6)	98.2 (74.1)
Multiplicity	4.0 (2.1)	5.6 (1.8)	10.1 (2.3)
$R_{\rm merge}$† (%)	13.5 (27.9)	9.9 (110.5)	7.7 (115.1)
Refinement			
Resolution (Å)	50.0–3.95	50.0–2.00	50.0–1.94
No. of reflections	118575	140125	168232
$R_{\rm work}/R_{\rm free}$‡ (%)	22.60/28.15	18.99/22.3	19.6/22.3
R.m.s. deviations			
Bond lengths (Å)	0.004	0.008	0.009
Bond angles (°)	0.77	1.32	1.44
No. of atoms			
Protein	127116	17407	17407
Ligand/ion	—	8	0
Water	—	810	1642
PDB entry	6dwu	6dwh	6dwf

† $R_{\rm merge} = \sum_{hkl}\sum_i |I_i(hkl) - \langle I(hkl)\rangle| / \sum_{hkl}\sum_i I_i(hkl)$, where $I_i(hkl)$ is the observed intensity of the ith measurement of reflection hkl and $\langle I(hkl)\rangle$ is the average intensity of this reflection obtained from multiple observations. ‡ $R = \sum_{hkl} \big||F_{\rm obs}| - |F_{\rm calc}|\big| / \sum_{hkl} |F_{\rm obs}|$, where $F_{\rm obs}$ and $F_{\rm calc}$ are the observed and calculated structure factors, respectively. $R_{\rm free}$ is as defined in Brünger (1992).

as search models to locate the six complexes present in the asymmetric unit. The structure was refined with *REFMAC*5 (Murshudov *et al.*, 2011) in *CCP*4 (Winn *et al.*, 2011) and was rebuilt with *Coot* (Emsley *et al.*, 2010). The crystals of the BbKI L55R–trypsin complex were isomorphous to those of native BbKI; thus, the former structure was directly used as a starting point for refinement of the latter structure with *REFMAC*5. A single BbKI L55R–trypsin complex was used as a search model to solve the structure of the BbKI–trypsin complex in the monoclinic space group, with a total of 44 complexes located within the asymmetric unit. The structure was subsequently refined with *PHENIX* (Adams *et al.*, 2010).

2.6. Modeling of kallikrein complexes

A model of a complex of BbKI with human plasma kallikrein was built by superimposing the coordinates of the protease domain of the enzyme (PDB entry 5tjx; Li *et al.*, 2017) on the trypsin molecule (A) in the BbKI L55R–trypsin complex. Complexes with human tissue kallikrein 4 (KLK4; PDB entry 4k8y; Riley *et al.*, 2016) and kallikrein 7 (KLK7; PDB entry 2qxi; Debela *et al.*, 2007) were modeled in an analogous manner. The coordinates of all three complexes were subjected to energy minimization of the contact region. For this, an all-atom version (including all of the hydrogens) was created for each assembly using the *PSFGEN* plugin for *VMD* v.1.9.1 (Humphrey *et al.*, 1996). All of the charged residues were set to their default dissociation states at neutral pH. For residues with two alternate locations in the original assemblies, only the second alternate conformation (*i.e.* conformation B) was included in the final structures for minimization. The disulfides that were present in the original structures were secured with explicit covalent bonds (Cys419–Cys435, Cys517–Cys584, Cys548–Cys563 and Cys574–Cys602 cross-links for PDB entry 5tjx, and Cys22–Cys157, Cys42–Cys58, Cys128–Cys232, Cys136–Cys201, Cys168–Cys182 and Cys191–Cys220 cross-links for PDB entries 4k8y and 2qxi). Only a narrow interfacial region (a 10 Å layer centered at the interface) was allowed to move and adjust during the minimization, *i.e.* all of the atoms at the interface within a 5 Å cutoff from any atom of the other contact partner. The rest of the system was fixed.

The energy minimizations were performed with the *NAMD* package (Phillips *et al.*, 2005) using the CHARMM36 force field (MacKerell *et al.*, 1998; Klauda *et al.*, 2010) in a vacuum, using a conjugate energy-gradient algorithm with a 12 Å cutoff for noncovalent interactions. We performed 1000 steps of energy minimization, which was sufficient to resolve all of the steric conflicts.

3. Results and discussion

3.1. Crystallization of the complex of BbKI with bovine trypsin

The recombinant form of BbKI used to determine the structure of the free inhibitor (Zhou *et al.*, 2015) included two serine residues at the N-terminus (residues 19 and 20 of the precursor; GenBank AAR01967.1). Although Ser19 of the precursor is most likely to be part of the signal peptide and is not present in the native form of the inhibitor, it was retained in the recombinant form owing to the way that the protein was cloned. Since this residue was well ordered in the crystal structure, it was previously retained in the numbering scheme and thus the reported sequence of BbKI started with Ser1-Ser2-Val3 . . . This numbering scheme for the residues is followed in this work and the 'native BbKI' referred to below has an extra Ser1 at its N-terminus.

Our initial crystallization experiments utilized the complex of native BbKI with trypsin at pH 7.5 and resulted in the growth of monoclinic crystals with a unit cell of very large volume (see below). An initial estimate was that at least 30 molecules of the complex could be present in the asymmetric unit, and all efforts to determine the structure by molecular replacement using the coordinates of BbKI (PDB entry 4zot) and bovine trypsin (PDB entry 5eg4) separately were unsuccessful. In a search for a different crystal form, we prepared the L55R mutant of BbKI and continued crystallization efforts using this version of the inhibitor.

3.2. Hexagonal crystals of the BbKI–trypsin complex

Hexagonal crystals of the BbKI–trypsin complex were initially grown at low pH using the L55R mutant and were

electronic reprint

subsequently also grown, under the same conditions, utilizing the native inhibitor. The crystals of the former complex grew in the hexagonal space group $P6_4$, with unit-cell parameters $a = b = 207.30$, $c = 107.17$ Å, and diffracted to beyond 2 Å resolution. Analysis of the Matthews parameter suggested the presence of six complexes in the asymmetric unit. The structure was solved by molecular replacement (see §2.5) and was refined with *REFMAC*5 (Table 2). The electron density is clear and continuous for all molecules in the unit cell and the geometry of the model is acceptable. The isomorphous structure of the complex with native BbKI was subsequently solved by direct refinement starting with the coordinates of the complex of the L55R mutant. With the exception of the mutated residues, which were clearly visible in the electron density in both structures, no major differences are present between these two coordinate sets (r.m.s.d.s range between 0.14 and 0.17 Å when comparing C^{α} positions for each pair of the six molecules in the asymmetric unit). Since the structure of the L55R complex was determined at a slightly higher resolution and the differences between the six complexes in the asymmetric unit are minor, the coordinates of the complex consisting of molecule *A* of trypsin and molecule *G* of BbKI L55R (Fig. 1*a*) will be used in all comparisons reported below. The r.m.s.d.s for the other complexes compared with *AG* are 0.17 Å (*BH*), 0.10 Å (*CI*), 0.17 Å (*DJ*), 0.05 Å (*EK*) and 0.17 Å (*FL*), justifying this choice.

The structures of both components of the complex are very similar to their counterparts in the available structures of the individual proteins. Superposition of the bovine trypsin molecule in the complex on the highest resolution structure of this enzyme (PDB entry 4i8h) yields an r.m.s.d. of 0.55 Å for all 223 C^{α} atoms. Deviations exceeding 1 Å are only found in three loops: 76–79, 96–98 and 114–117 (chymotrypsin numbering is used throughout). In many published structures of bovine trypsin, including that used in this comparison, loop 76–79 is stabilized by a calcium ion, but calcium was not present in the crystallization buffers used to obtain the crystals of the complex and the ions are replaced by putative water molecules. Attempts to soak calcium into crystals of the complex were unsuccessful and adding calcium chloride to the crystallization buffer did not yield any crystals. The reasons for this behavior are not known. Loop 114–117 is located on the surface of the protein and its conformation may be influenced by crystal contacts, whereas loop 96–98 is found at the interface between the enzyme and the inhibitor (see below) and thus its conformation may be influenced by the presence of intermolecular interactions.

Superposition of native BbKI on molecule *G* of the complex yields an r.m.s.d. of 0.78 Å for all 163 C^{α} atoms of the inhibitor. Significant differences are only present in several loops, with the surface loops 26–28, 36–38, 49–50, 106–107 and 140–144 exhibiting maximum deviations between the positions of C^{α}

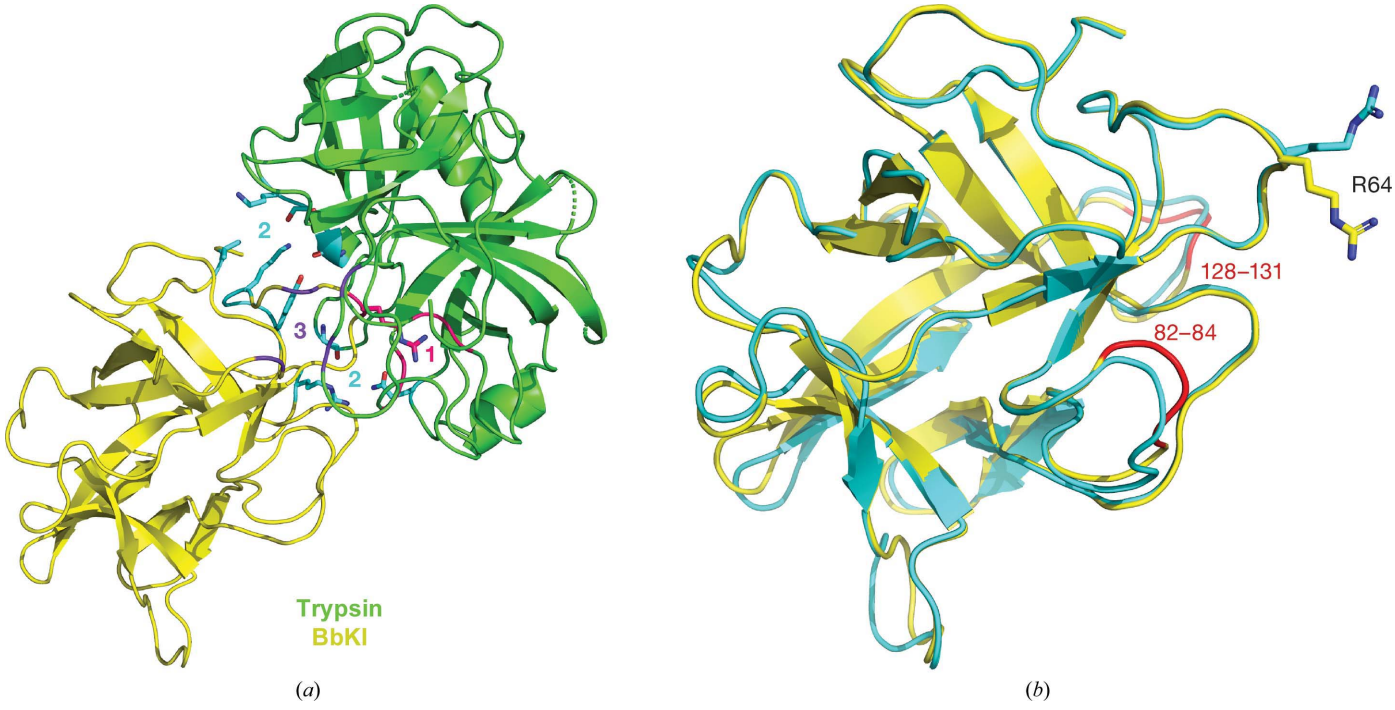


Figure 1
The crystal structure of BbkI in complex with trypsin and as a free inhibitor. (*a*) A cartoon representation of one of the six BbKI–trypsin complexes [consisting of molecules *A* (trypsin, green) and *G* (BbKI L55R, yellow)] present in the monoclinic crystals. The side chain of the specificity-determining Arg64 in BbKI is shown as magenta sticks. Three major contact areas of interaction between the inhibitor and the enzyme are marked in identical colors (1, magenta; 2, cyan; 3, purple) in both molecules. For clarity, selected side chains are shown as sticks to delineate the most extended epitope 2. (*b*) Superposition of BbKI in bound (yellow) and free (blue) states. Arg64 is shown as sticks in both structures. The two most divergent loops at the interface between the inhibitor and the enzyme (82–84 and 128–131) are colored red.

electronic reprint

atoms of up to 3 Å. Two loops at the interface between the inhibitor and the enzyme (82–84 and 128–131) were also shifted by up to 2.8 Å (Fig. 1*b*).

3.3. Monoclinic crystals of the BbKI–trypsin complex

The unit-cell parameters of this crystal form, a = 137.2, b = 483.9, c = 137.0 Å, β = 116.8°, and systematic absences with $k = 2n + 1$ for $0k0$ reflections point to the monoclinic space group $P2_1$, but the close similarity in the lengths of the a and c unit-cell axes could potentially lead to orthorhombic symmetry and the $C222_1$ space group, or indicate the existence of pseudo-merohedral twinning. However, all of the symmetry and twinning tests performed with *phenix.xtriage* excluded these possibilities and it was therefore assumed that this crystal is indeed monoclinic and not twinned.

The unit-cell volume of this crystal form suggested the presence of a large number of independent molecules in the unit cell, but initially it was not clear exactly how many. Molecular-replacement runs with *Phaser* (McCoy *et al.*, 2007), using the structure of the BbKI L55R–trypsin complex determined in this study (PDB entry 6dwf) first identified the positions of 22 complexes, but inspection of the molecular packing suggested that this was not a complete structure. Successive runs of *Phaser* eventually completed the structure, finding 44 complexes in the asymmetric unit of the cell. The structure was refined with the TLS option, with each of the 88 individual protein chains assigned as a separate TLS domain. The final refinement statistics are summarized in Table 2.

The 44 heterodimers (trypsin–inhibitor complexes) are arranged in the cell in four columns running along four independent 2_1 axes at $(0, y, 0)$, $(0, y, \frac{1}{2})$, $(\frac{1}{2}, y, 0)$ and $(\frac{1}{2}, y, \frac{1}{2})$. Each pair of successive dimers along each column are rotated with respect to each other by about 114° on average. This corresponds to a full-turn rotation between the first and 23rd dimer, or 180° rotation between the first and 12th dimer along the column axis. Each column, therefore, is built by dimers arranged according to an approximate 22_7 screw axis, since $114° \simeq 7 \times (360°/22)$. Two views of the content of the asymmetric unit are shown in Fig. 2.

Although some structures with as many as 60 molecules in an asymmetric unit are present in the PDB, they contain fewer biologically relevant units. For example, crystals of lumazine synthase (PDB entry 3mk3; Kumar *et al.*, 2011) contain 60 protein molecules in the asymmetric unit but only one biological unit. Similarly, several proteins with as many as 48 molecules in the asymmetric unit contain fewer biological units; for example, DUTPase (PDB entry 4apz; García-Nafría *et al.*, 2013) contains 16 trimers and ethanolamine-utilization microcompartment protein (PDB entry 3ia0; Tanaka *et al.*, 2010) contains eight hexamers. As this particular crystal form of the BbKI–trypsin complex appears to be merely a crystallographic curiosity, it was not analyzed further.

3.4. Interactions between the inhibitor and the enzyme

Three major epitopes on BbKI, each comprising several loops, are involved in contacts with bovine trypsin. An overall view of the interface is shown in Fig. 1(*a*). Contact areas within each epitope are marked in the same color in both the inhibitor and in trypsin. Details of the BbKI–trypsin contacts within each site can be seen in Fig. 3.

The primary interactions (epitope 1 in Fig. 1*a*) involve the so-called 'reactive loop' of the inhibitor (residues 61–66). The extensive interactions of Arg64, the key residue of the inhibitor that makes a strong ion pair with the carboxylate of Asp189 of trypsin, are of particular importance. The distance

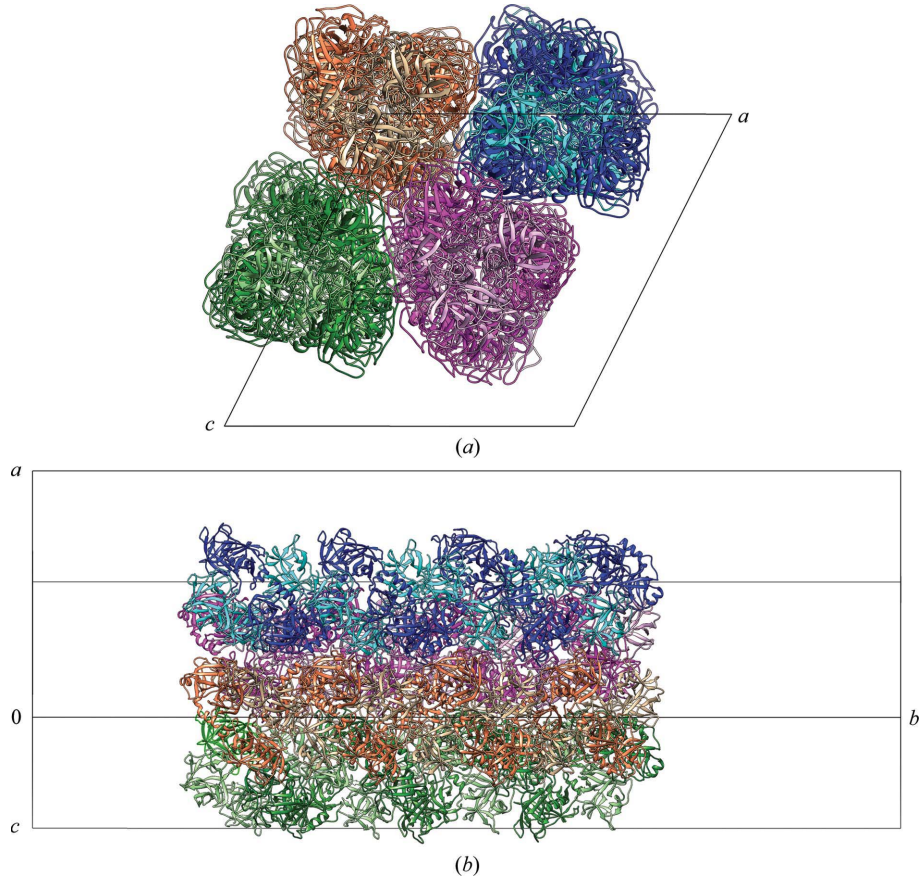


Figure 2
The packing of 44 independent BbKI–trypsin complexes in the monoclinic crystal. Each of the four helical columns of 11 complexes arranged along the four crystallographic 2_1 axes is presented in a different color, with the trypsin molecules in darker shades and the inhibitor molecules in paler shades. The unit cell is outlined in thin black lines. (*a*) The view along the length of the helical columns, *i.e.* along the monoclinic 2_1 axis. (*b*) The view perpendicular to the direction of the helices.

electronic reprint

between NH1 of Arg64 and OD1 of Asp189 is 2.9 Å, whereas the distance between Arg64 NH2 and Asp189 OD2 is 2.7 Å.

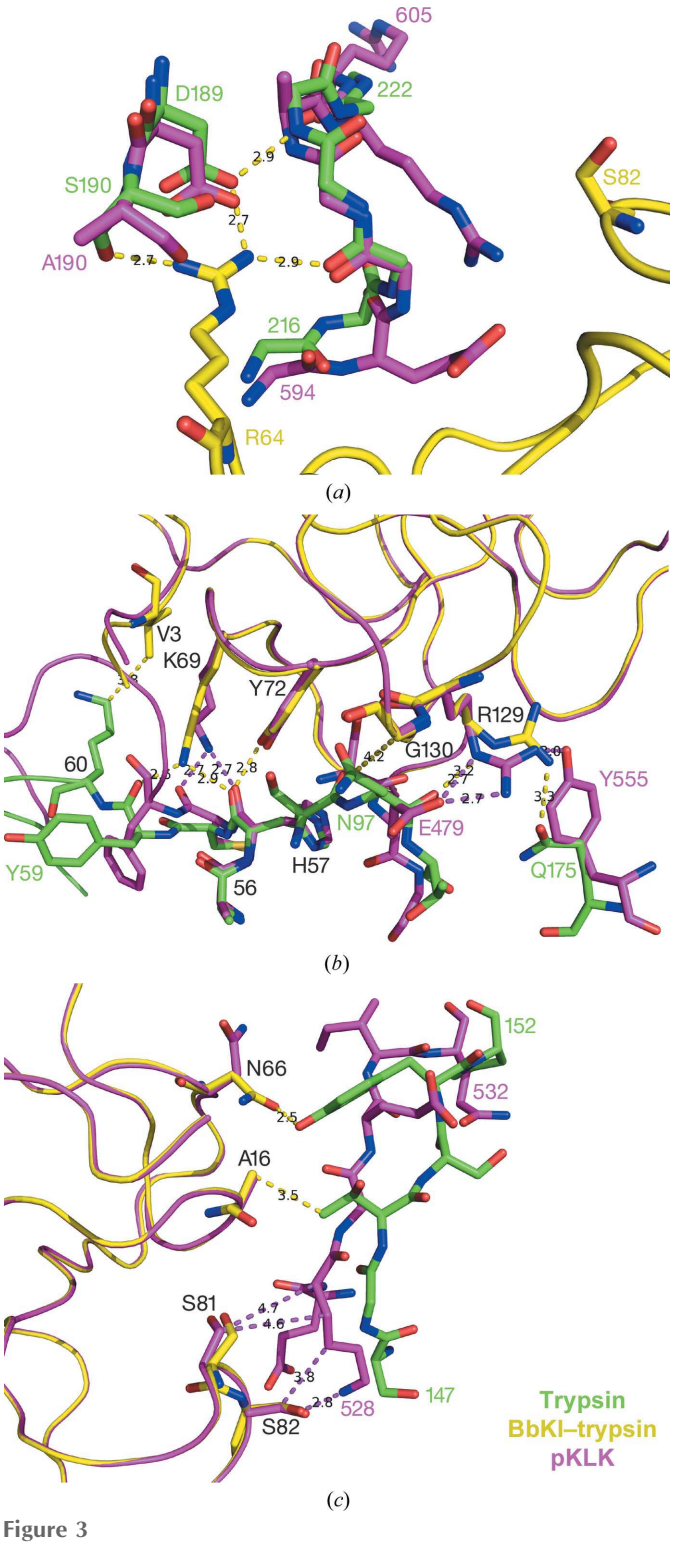


Figure 3
Modeling of the interactions between BbKI and pKLK. (*a*) Interactions in the vicinity of the specificity-providing P1 Arg64 in trypsin and plasma kallikrein (epitope 1). (*b*) Interactions of three loops in BbKI comprising Val3, Lys69 and Tyr72, and Arg129 and Gly130, respectively, with the corresponding fragments in trypsin and kallikreins (epitope 2). (*c*) Interactions of three loops in BbKI, which include Ser81–Ser82, Asn66 and Ala16, respectively, with trypsin and pKLK (epitope 3).

Other interactions of Arg64 include hydrogen bonds between Arg64 NH2 and Gly218 O (2.85 Å), and between Arg64 NH1 and Ser190 OG (2.7 Å). Hydrogen bonds between Arg64 O and Gly193 N (2.7 Å), between Arg64 N and Ser214 O (3.05 Å), and between Leu63 O and Gln192 NE2 (2.9 Å), as well as between Pro62 O and Gly216 N (3.1 Å), play crucial roles in positioning the main chain of the reactive loop in the active site of the enzyme and maximizing the interactions of the specificity-providing Arg64. Other interactions stabilizing the reactive loop involve putative hydrogen bonds between both OD1 and ND2 of Asn66 of BbKI and Tyr151 OH of trypsin.

A number of interactions outside the area of the reactive loop help to stabilize the complex and provide additional specificity. The guanidinium group of Arg129 of BbKI interacts with the side chain of Gln175 of trypsin, as well as with Asn97 O. Lys69 NZ and Tyr72 OH of BbKI are hydrogen-bonded to the carbonyl oxygens of Tyr59 and His57 of trypsin, respectively, with the former residue making hydrogen bonds to both carbonyls (epitope 2 in Fig. 1*a*).

The loops containing Ala16 and Asn66 in BbKI interact with trypsin as part of epitope 3 (Fig. 1*a*). Additional nonpolar interactions involve a number of residues from both proteins and presumably contribute to the stability of the complex.

3.5. A comparison with the structures of trypsin complexes of related inhibitors

The crystal structures of trypsin complexes of four inhibitors related to BbKI have previously been published. They are soybean trypsin inhibitor (STI; PDB entry 1avw; Song & Suh, 1998), tamarind Kunitz inhibitor (TKI; PDB entry 4an7; Patil *et al.*, 2012), *Enterolobium contortisiliquum* inhibitor (EcTI; PDB entry 4j2y; Zhou *et al.*, 2013) and an engineered version of winged-bean chymotrypsin inhibitor (WCI) in which the specificity-determining Leu65 was mutated to an arginine (PDB entry 3veq; Majumder *et al.*, 2012). STI and TKI were complexed with porcine trypsin, whereas EcTI and WCI were complexed with bovine trypsin, similarly to BbKI. (Parenthetically, the trypsin sequence in PDB entry 3veq contains an error in identifying Ile118 as a valine, unlike in other structures involving cationic bovine trypsin; the electron density clearly supports the presence of isoleucine.) However, the differences between the source of trypsin used in the structural studies seem to be immaterial, since the r.m.s.d.s after superposition of the trypsin molecules of the complexes were 0.52, 0.53, 0.55 and 0.48 Å for all C^{α} atoms in the complexes with STI, TKI, EcTI and WCI, respectively. The reactive loops of the inhibitors in all five complexes adopt very similar conformations and superimpose very well when trypsin is used as a reference, although such trypsin-based superposition leads to comparatively large shifts in most of the inhibitor molecules. This happens despite the fact that the inhibitors superimpose quite well by themselves. The r.m.s.d.s are 1.44 Å for the superposition of 143 C^{α} atoms of STI (sequence identity 32.2%), 1.55 Å for the superposition of 143 C^{α} atoms of TKI (sequence identity 36%), 1.35 Å for the superposition

electronic reprint

Table 3
Inhibitory potencies of wild-type BbKI and its mutant (BbKI L55R) against various kallikreins and archetypal proteases from the S1 family.

n.s.i. indicates that no significant inhibition was observed with 10 μ*M* inhibitor.

	K_i (n*M*)	
Enzyme	BbKI	BbKI L55R
Human kallikreins		
Plasma kallikrein	12.9 ± 3.2	14.4 ± 1.1
hKLK2	n.s.i.	n.s.i.
hKLK3	n.s.i.	n.s.i.
hKLK4	0.045 ± 0.002	0.036 ± 0.006
hKLK5	>500	>500
hKLK7	0.18 ± 0.03	2.2 ± 0.2
hKLK14	>5000	>5000
Mouse kallikreins		
mKLK5	>500	>500
mKLK7	9.8 ± 1.6	15.9 ± 3.4
Archetypal serine proteases		
Trypsin	0.20 ± 0.01	1.3 ± 0.1
Chymotrypsin	33.4 ± 2.3	50.6 ± 6.4

of 145 C^{α} atoms of EcTI (sequence identity 34.5%) and 1.65 Å for the superposition of 148 C^{α} atoms of WCI (sequence identity 32.5%). In all of these comparisons we find very accurate overall superposition of the β-sheets of the inhibitors, whereas many loops, including the reactive loops, are not well superimposed. The only way to keep the reactive loops superimposed is by rigid-body movement of the whole inhibitors in relation to the enzyme, thus affecting their interactions in areas distant from the active site. However, although rigid-body rotation has previously been invoked as a potential determinant of the secondary specificity of interaction (Zhou *et al.*, 2013), these adjustments of the quaternary structures may also be owing to crystal contacts, as suggested by the 7.25° rotation of STI in two high-resolution structures of the complex in different crystal forms (Song & Suh, 1998). A major difference between BbKI and the other four inhibitors is observed in the orientation of the loop 80–83, which faces the enzyme in the former inhibitor and faces away in the latter inhibitors.

3.6. Inhibitory specificity of BbKI

We investigated the inhibitory specificity of BbKI using a panel of nine kallikreins of human and mouse origin, in addition to trypsin and chymotrypsin; the latter enzymes represent archetypal serine proteases from the S1 family. BbKI was screened using a kinetic activity assay and the inhibition constants K_i were determined (Table 3). The results show that BbKI is a potent inhibitor of human kallikreins KLK4 (K_i of ~0.045 n*M*), KLK7 (K_i of ~0.18 n*M*) and plasma kallikrein (K_i of ~12.9 n*M*). In contrast, it displayed only weak inhibition of KLK5 and KLK14 (K_i values of >0.5 and 5 μ*M*, respectively) and no detectable inhibition of KLK2 and KLK3. The relative sensitivities of mouse KLK7 and KLK5 to BbKI inhibition were rather similar to those of their human orthologs. Furthermore, we compared the inhibitory profiles of both native BbKI and its L55R mutant, indicating that this substitution did not influence the binding properties to serine proteases in any significant way (Table 3).

These results demonstrated that BbKI is able to effectively inhibit kallikreins with trypsin-like cleavage specificities (human plasma kallikrein and KLK4), as well as chymotrypsin-like enzymes (KLK7). This is in line with the inhibition of the archetypal proteases trypsin and chymotrypsin by BbKI (K_i values of ~0.2 and 33 n*M*, respectively). A versatile reactive-site loop with a basic P1 residue that is responsible for the inhibition of trypsin and also chymotrypsin has been proposed for the plant Kunitz inhibitor STI (De Vonis Bidlingmeyer *et al.*, 1972) and was subsequently demonstrated crystallographically for the animal Kunitz inhibitor BPTI (Helland *et al.*, 1999; Scheidig *et al.*, 1997). We provide evidence that BbKI is a highly potent inhibitor of human KLK4 that is effective in the low-picomolar range. To our knowledge, BbKI is one of the best natural proteinaceous inhibitors of KLK4 (and kallikreins in general) with the noncovalent, Laskowski mechanism of action. A similar inhibitory potency was reported for a rationally designed derivative of the sunflower trypsin inhibitor (SFTI) interacting with KLK4 (Swedberg *et al.*, 2011).

3.7. Modeling of the interactions between BbKI and selected kallikreins

A model of the complex between human plasma kallikrein (pKLK) and BbKI was built by superimposing the coordinates of pKLK refined at 1.4 Å resolution (PDB entry 5tjx; Li *et al.*, 2017) on those of trypsin in the BbKI L55R–trypsin complex (r.m.s.d. of 1.11 Å for 205 C^{α} pairs). Unlike in the previous model of such a complex, which was based on the structure of free BbKI (Zhou *et al.*, 2015), in the new energy-minimized model the specificity-determining residue on the reactive loop of BbKI, Arg64, makes a strong ion pair with Asp572 of kallikrein (Fig. 3*a*).

Secondary contributions to the specific binding of BbKI to pKLK are provided by the interactions between Lys69 NZ and Tyr72 OH of BbKI with the carbonyl oxygens of His434 and Cys435 of pKLK, respectively. The interaction pattern for the residues that follow is unique to pKLK, owing to an insertion in the main chain of this enzyme. This is an area in which the structures of trypsin and pKLK diverge significantly (Fig. 3*b*). Several other specific interactions between BbKI and pKLK are seen in the modeled complex. Strong charge–charge interactions are formed between Arg129 of BbKI and Glu479 of pKLK. Also, Arg129 NH1 is hydrogen-bonded to Tyr555 OH of pKLK (Fig. 3*b*).

In the pKLK–BbKI complex the loop containing residues 81–82 becomes part of epitope 3 of the interface. Ser81 and Ser82 in BbKI interact extensively with the side chain of Lys528 of pKLK (Fig. 3*c*). The hydroxyl group of Ser82 is hydrogen-bonded to Lys528, whereas the C^{β} atoms of both Ser81 and Ser82 make hydrophobic contacts with this residue in pKLK. This loop does not make similar contacts with trypsin. Some other interactions include an ion pair between Lys106 of BbKI and Glu600 of pKLK in another contact area. Close to the specificity-defining Arg64, Asn66 of BbKI is hydrogen-bonded to the guanidinium group of Arg416 in

electronic reprint

pKLK, as well as to a carbonyl oxygen of Leu418. The latter interaction is unique to this complex, since the counterpart of Arg416 in pKLK is Tyr39 in trypsin.

Since the inhibitory properties of BbKI evaluated with a panel of kallikreins were shown to be best for human KLK4, a model of a complex between KLK4 and BbKI was prepared in a manner analogous to the model of a complex of pKLK by superimposing the published atomic resolution model of KLK4 (PDB entry 4k8y; Riley *et al.*, 2016) on the coordinates of trypsin in the BbKI–trypsin complex (r.m.s.d. of 1.09 Å for 215 C^{α} pairs). Comparative analysis of the interactions of BbKI in the experimental structure of the complex with trypsin and in the model of the complex with KLK4 does not provide clear justification for the much higher inhibitory potency of BbKI towards KLK4. The interactions of the P1 Arg64 of the inhibitor with part of the primary binding pocket comprising residues 189–190, as well as residues 216–218, seem to be very similar for trypsin and KLK4. However, loop 216–222 in KLK4 and several other kallikreins, such as KLK2, KLK3, KLK5, KLK7 and KLK14, adopts a different conformation owing to the presence of the Pro218A insertion in this protein region (Figs. 4 and 5*a*). The changes in the conformation do not affect the formation of the specificity-determining ion pair between Arg64 of BbKI and Asp189 of the enzyme. The insertion of Pro218A changes the conformation of this loop in a similar way, creating an identical pattern of BbKI–kallikrein interactions in all of these enzymes (Figs. 5*a* and 6*a*). Kallikreins that contain a Pro218A insertion gain additional hydrophobic contacts with the loop in BbKI that carries Ser82, placing this loop within epitope 1 of the interface (Fig. 5*a*).

Three other segments of BbKI, one containing Val3 and another containing Lys69 and Tyr72, as well as the loop that includes Arg129 and Gly130, consistently maintain contacts with the corresponding fragments comprising residues 56–60, 95–98 and 175 in trypsin, KLK4 and all other kallikreins. Only two hydrogen bonds, which include that between Lys69 NZ in BbKI and Tyr59 O in trypsin, as well as that between the hydroxyl of Tyr72 OH in BbKI and the carbonyl oxygen of His57, are conserved in all modeled kallikrein complexes (Fig. 5*b*). The other interactions vary in the different complexes. The hydrogen bond between Arg129 in BbKI and Gln175 in trypsin is not present in the complexes with KLK4 owing to the substitution Q175L.

In the absence of an experimental structure, very little can be said about inhibitor interactions with residues 96–98 in KLK4. The structure of KLK4, which was used for modeling (PDB entry 4k8y), includes the bound inhibitor SFTI-1. Loop 96–98 in this complex adopts a different conformation to that in trypsin, KLK5, KLK7 and pKLK.

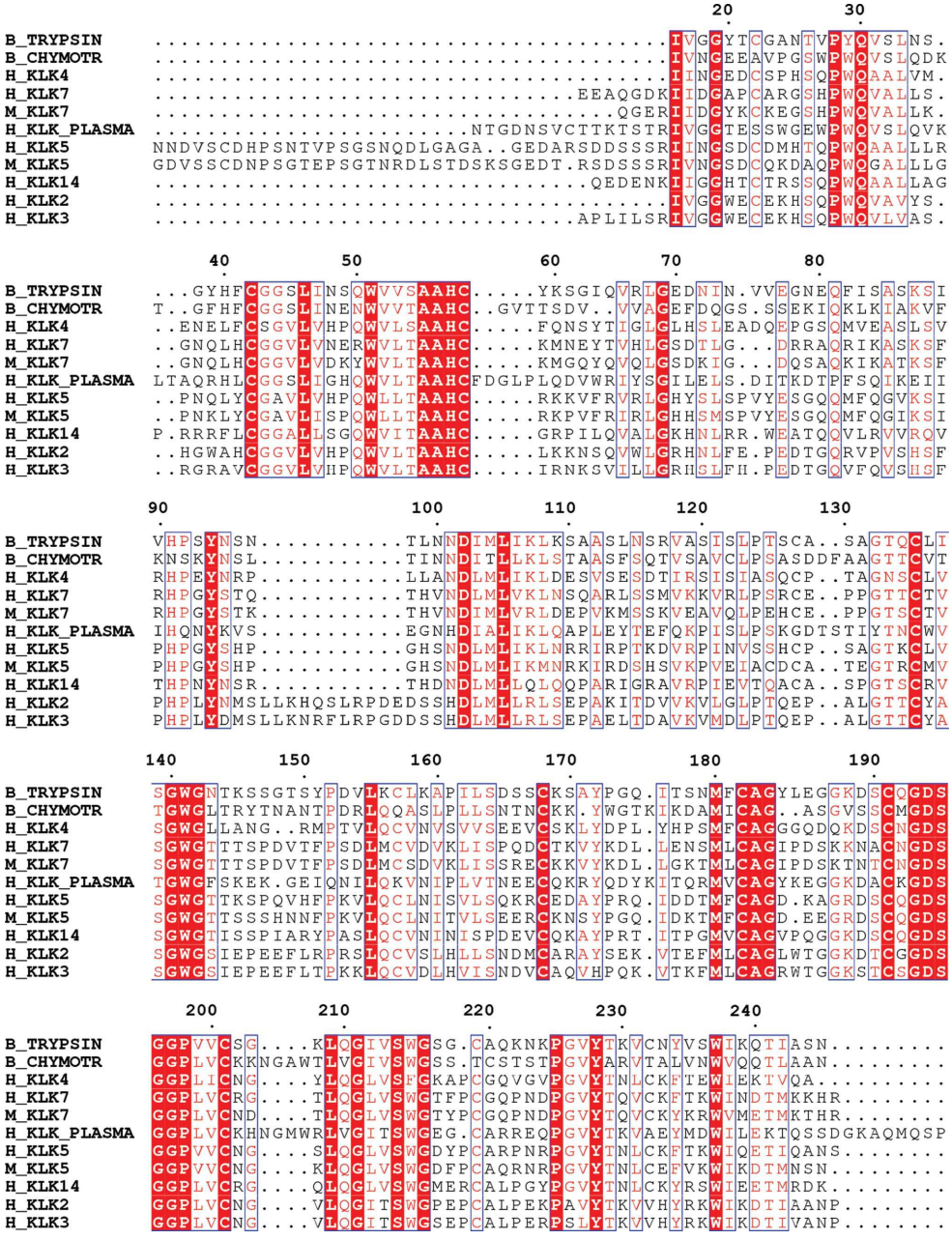


Figure 4
Structure-based alignment of the amino-acid sequences of trypsin, chymotrypsin and selected kallikreins.

electronic reprint

It interacts with the bulky side chain of Phe12 of SFTI-1 and is also involved in crystal contacts with its counterpart from a symmetry-related molecule. As a result, the tip of the loop is shifted to a different position (Fig. 5*b*). Therefore, the contact area for this loop in the complex with BbKI could not be properly modeled.

The BbKI loop containing residues Ser81-Ser82, as well as two loops that include Asn66 and Ala16, respectively, make almost no contacts with the fragment of KLK4 comprising residues 147–152, whereas a corresponding fragment in KLK5 makes several interactions with the same loops (Fig. 5*c*), although BbKI is a very weak inhibitor of KLK5 (Table 3).

Whereas pKLK and KLK4 are trypsin-like enzymes, KLK7 is chymotrypsin-like; thus, the reasons for the potent inhibitory activity of BbKI against this enzyme were not obvious. A model of a complex between KLK7 and BbKI was prepared in a manner analogous to the model of the complex with pKLK by superimposing the published atomic resolution model of KLK7 (PDB entry 2qxi; Debela *et al.*, 2007) on the coordinates of trypsin in the BbKI–trypsin complex (r.m.s.d. of 1.14 Å for 218 C^{α} pairs). A crucial difference between the KLK7 on one hand and trypsin and some other kallikreins on the other is the presence of Asn in place of Asp189 (or Ser189 in chymotrypsin and KLK3). Both carboxyl oxygens of Asp189 in trypsin and other kallikreins are hydrogen-bonded to NH1 and NH2 of Arg64 of BbKI. In KLK7 only Asn189 OD1 is the acceptor of a hydrogen bond from Arg64 NH1, in addition to an intramolecular hydrogen bond provided by Gly220 N. Asn189 ND2 is stabilized by an intramolecular hydrogen bond to Lys188 O. Other interactions of Arg64 observed in the structure of the BbKI–trypsin complex include hydrogen bonds between Arg64 NH1 and Ser190 OG (2.73 Å) and between Arg64 NH2 and Gly218 O (2.85 Å). Both interactions are not preserved in the KLK7 complex owing to the presence of Ala190 in the place of serine, as well as the insertion of a proline residue at position 218A that changes the conformation of the loop around it (Fig. 6*a*). The same insertion is also found in KLK4, KLK5, KLK2 and KLK3, with Arg218A present in KLK14. The conformational changes induced by this insertion are very similar in all compared kallikreins, leading to the loss of a hydrogen bond between Arg64 NH2 and Gly218 O, while gaining hydrophobic interactions between the side chains of the inserted proline and Ser82 (Fig. 6*a*).

Substitution of Ser190 by alanine combined with a change of Asp189 to Asn reorients the side chain of Arg64, leading to dramatic changes in the network of interactions of the guanidinium group of Arg64 with KLK7 (Fig. 6*a*). An extensive network of interactions involving the NH1, NH2 and NE atoms of Arg64 with different counterparts in the enzyme might contribute to the high inhibitory potency of BbKI against KLK7.

The fact that the changes of the orientation of Arg64 in the KLK7 complex are directly connected to the substitutions D189N and S190A, and not to the Pro218A insertion, is supported by a comparison with KLK4. While the same insertion of a proline residue is present in KLK4 as in KLK7, Asp189 and Ser190 are not replaced. Fig. 5(*a*) very clearly shows the same orientation and a very similar interaction pattern of Arg64 of BbKI in the complexes with KLK4 and trypsin, whereas the orientation of Arg64 is different in the complex with KLK7 (Fig. 6*a*). The inhibition constants for these two enzymes are also very similar.

The interactions between the loop in BbKI containing Lys69 and Tyr72 do not change much in the complex with KLK7, whereas the longer side chain of Gln97 enables a much better interaction pattern with Arg129 in KLK7 than in trypsin. Numerous contacts that include hydrophobic interactions, as well as hydrogen bonds, may compensate for the loss of contacts with Val3 and Leu175 in KLK7 (Fig. 6*b*).

Three loops in BbKI, containing Ser81-Ser82, Asn66 and Ala16, respectively, make more extensive interactions with KLK7 than with trypsin (Fig. 6*c*). Although a strong hydrogen bond between Asn66 and Tyr151 in the trypsin complex is not present in the KLK7 complex owing to substitution by Phe151, the hydrophobic contacts between these residues still remain. An extra hydrogen bond is formed between the carbonyl oxygen of Pro147 and the hydroxyl of Ser81, and several hydrophobic contacts are made between Pro147 and Ser81-Ser82 (Fig. 6*c*). These interactions appear to be present in all kallikreins that simultaneously contain a proline insertion (Pro218A) and Pro147 in the corresponding loop. In these complexes the loop containing Ser82 will simultaneously contribute to epitopes 1 and 3 of the interface.

The loops that contain Pro147 and Pro218A are located just above the entrance to the specific P1 pocket of the enzyme. The interactions of the Ser81-Ser82 loop of the inhibitor with these two loops in KLK7 may reduce their mobility and thus facilitate docking of the side chain of Arg64 in the primary recognition pocket (Fig. 7).

3.8. Structural basis for the lack of inhibition of several other kallikreins

As seen in Table 3, BbKI shows only very low (if any) inhibitory potency against KLK5 and KLK14, whereas it does not inhibit KLK2 and KLK3 at all. To justify the lack of inhibition in structural terms, the available structure of human KLK5 (PDB entry 2psx) was superimposed on trypsin in complex with BbKI in order to examine the differences in the enzyme–inhibitor interactions compared with the trypsin complex.

In the complex with trypsin, Arg64, the specificity-determining P1 basic residue on the reactive loop of BbKI, contacts Asp189 and Ser190, and is hydrogen-bonded to the carbonyl oxygen of Gly218. The latter interaction is most likely not present in KLK5 and KLK14 owing to the insertion of Pro218A in the former and Arg218A in the latter that changes the conformation of the main chain of the enzyme around this residue (Fig. 8*a*). However, an equivalent insertion is also present in KLK4 and KLK7. While the loss of the hydrogen bond between P1 Arg64 and O of Gly218 might explain the lower inhibitory activity of BbKI against KLK5 and KLK14, this effect is compensated by other interactions in

electronic reprint

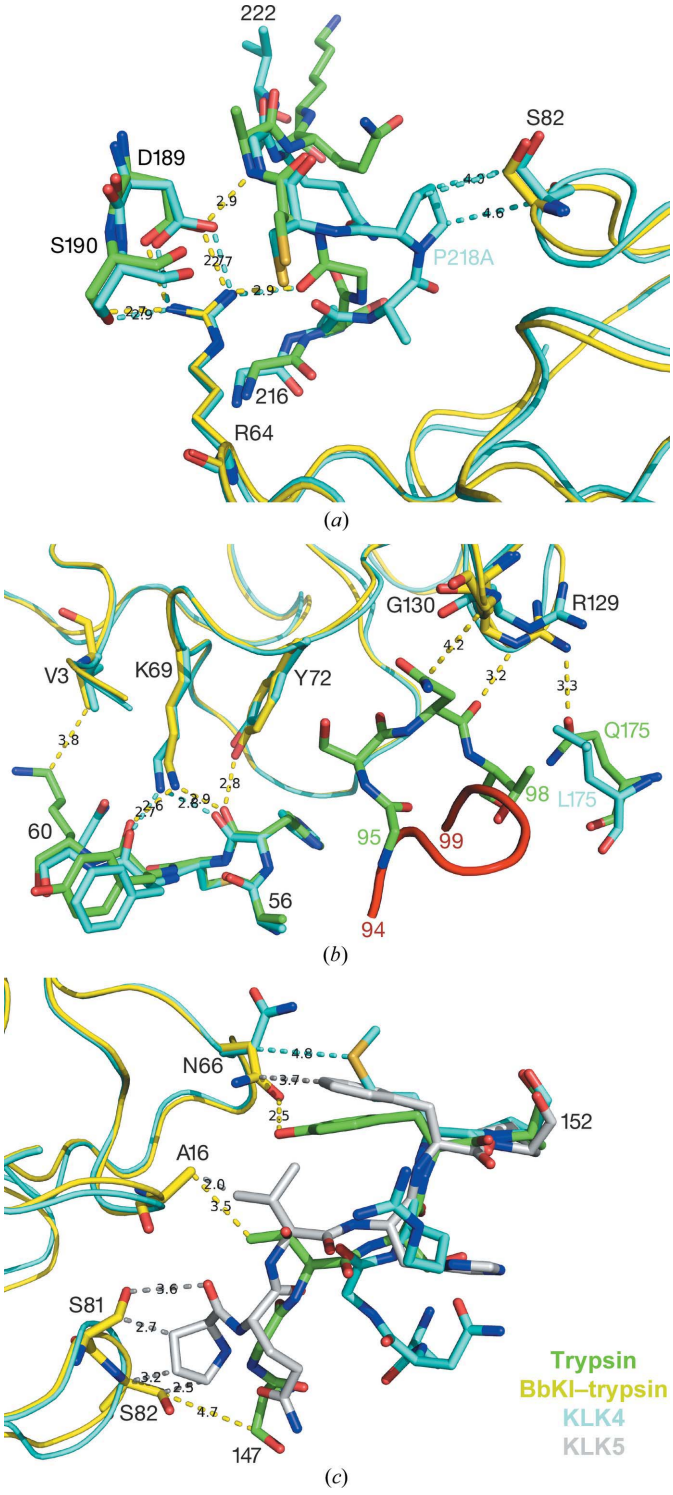


Figure 5
Modeling of the interactions between BbKI and KLK4. (*a*) Interactions in the vicinity of Arg64. The inserted Pro218A in KLK4 maintains additional hydrophobic contacts with the loop in BbKI carrying Ser82. (*b*) Interactions of three loops in BbKI comprising Val3, Lys69 and Tyr72, and Arg129 and Gly130, respectively, with trypsin and KLK4. Interactions of BbKI with residues 96–98 in KLK4 (shown as a red ribbon) cannot be modeled. (*c*) The open conformation of the loop 96–98 in KLK4 complexed with the sunflower trypsin inhibitor (PDB entry 4k8y). (*d*) Interactions of the BbKI loop containing Ser81-Ser82, as well as the two loops that include Asn66 and Ala16, respectively, with KLK4 and trypsin.

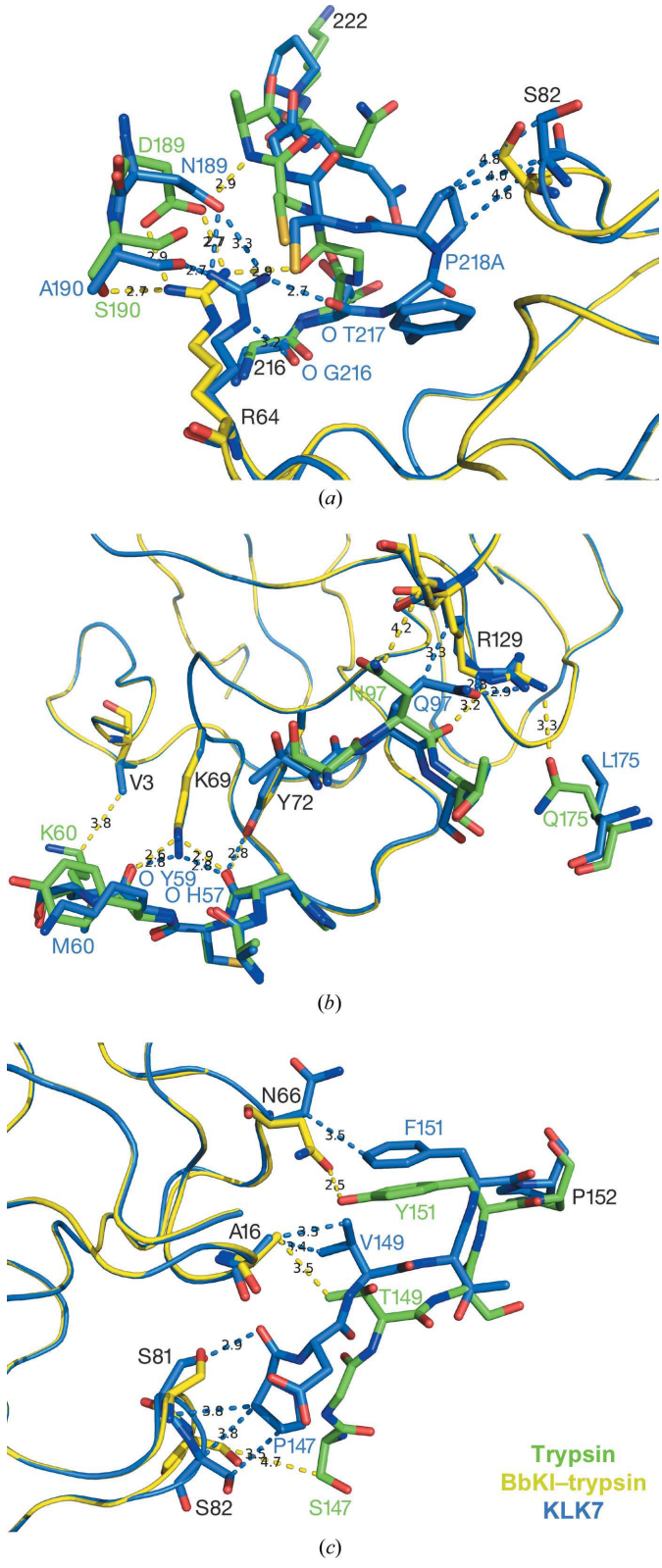


Figure 6
Modeling of the interactions between BbKI and KLK7. (*a*) Differences in the interactions of Arg64 of BbKI in the primary pocket of KLK7 and its counterparts in trypsin and KLK4. (*b*) Interactions of three loops in BbKI comprising Val3, Lys69 and Tyr72, and Arg129 and Gly130, respectively, with trypsin and KLK7. (*c*) The extensive interactions between three loops in BbKI which include Ser81-Ser82, Asn66 and Ala16, respectively, with KLK7 compared with those with trypsin.

electronic reprint

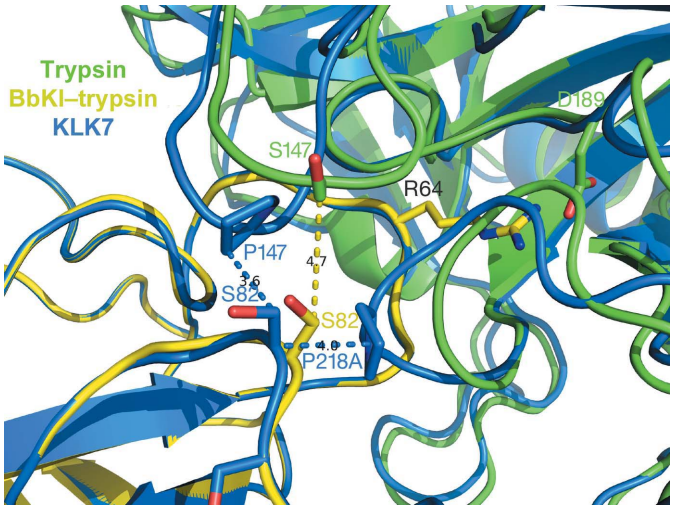


Figure 7
Hydrophobic interactions of Ser81-Ser82 of the inhibitor with two loops in kallikreins simultaneously containing Pro147 and Pro218A.

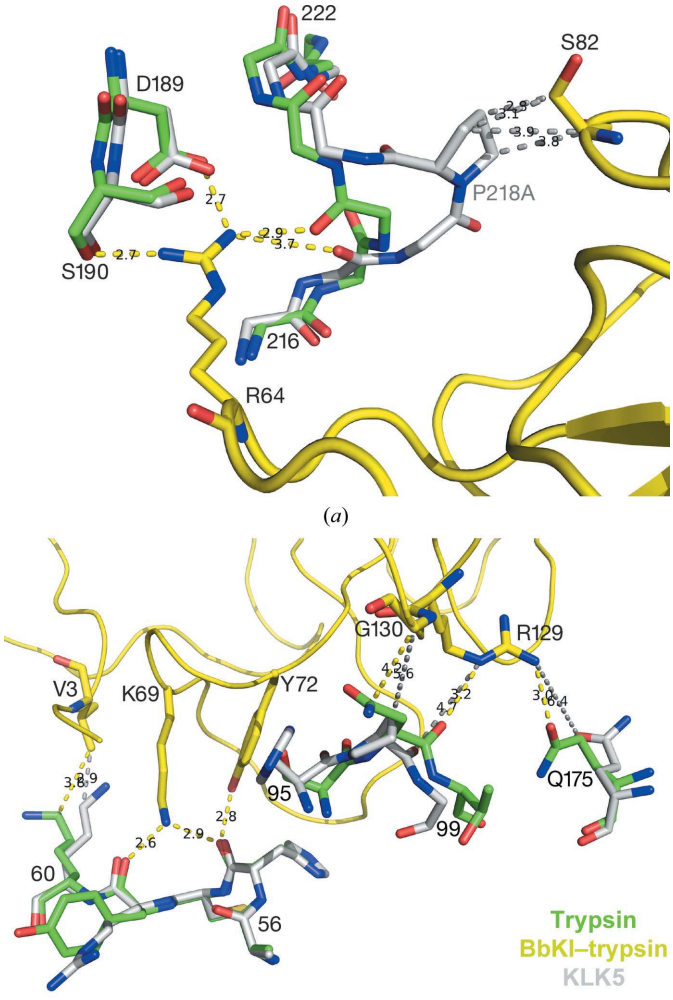


Figure 8
Modeling of the interactions between BbKI and KLK5. (*a*) Insertion of Pro218A in KLK5 leads to the loss of a hydrogen bond between Arg64 in BbKI and Gly218. (*b*) Modified interactions between BbKI and KLK5 owing to the amino-acid substitution in the loop containing residue 97.

KLK4 and KLK7. Two other interactions might be lost in KLK5 owing to the substitution of Asn97 (trypsin) by Pro97 (a hydrogen bond between the carbonyl oxygen of Asn97 and NE of Arg129 in BbKI, as well as a hydrophobic contact between the CG atom of Asn97 and the C^{α} atom of Gly130; Fig. 8*b*). In addition, the bridging interactions between the side chain of Arg129 in BbKI and two loops, 174–175 and 96–97 in trypsin, will most likely be lost in the complex with KLK5 (Fig. 8*b*).

Close contacts occurred between the loop containing Ser80-Ser81 in BbKI and the two loops in KLK5 containing Pro147 (<2 Å) and Tyr218 (~1.5 Å) (Fig. 9*a*) that may contribute to the lack of inhibition. In addition, a strong hydrogen bond between Asn66 in BbKI and Tyr151 in trypsin (2.5 Å) cannot be present in the complex with KLK5 owing to the Y151F substitution.

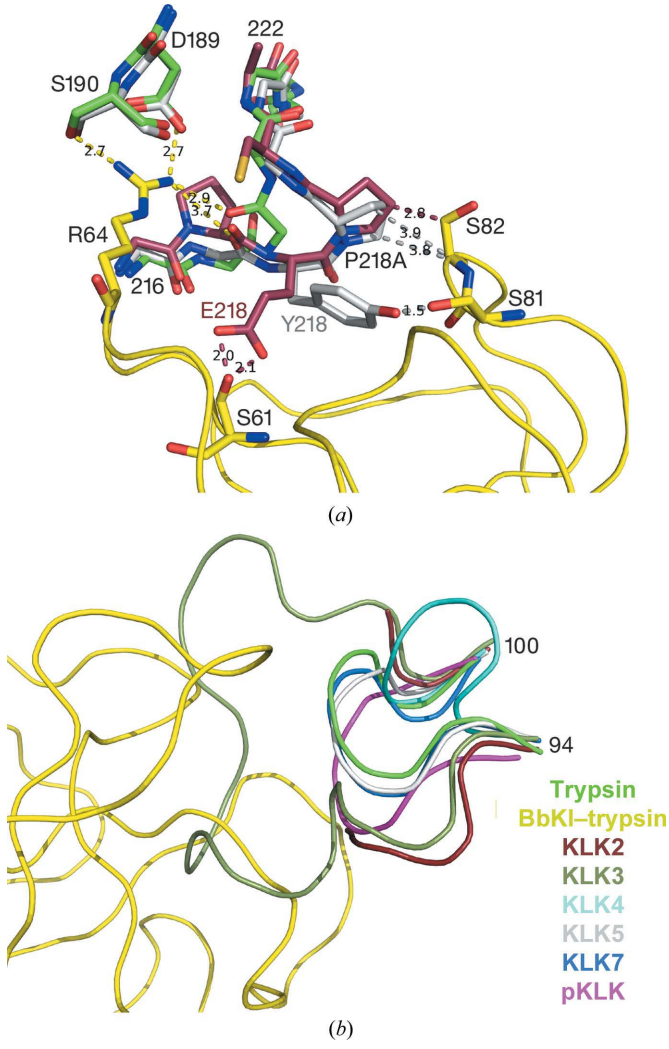


Figure 9
Structural basis for the lack of inhibition of KLK2 and KLK3. (*a*) Extensive clashes of Glu218 in KLK2 and KLK3 with the loop containing Ser61 in BbKI may contribute to the lack of inhibition, in a manner similar to the conflict between Tyr218 in KLK5 and the Ser81-Ser82 loop in BbKI. (*b*) The large insertion in the loop 94–100 in both enzymes leads to collision with several loops in BbKI, leading to the inability of BbKI to inhibit KLK2 and KLK3.

electronic reprint

We attribute the complete inability of BbKI to inhibit KLK2 and KLK3 predominantly to the large insertion in the loop 94–100 in both enzymes (Fig. 9*b*), which leads to severe collisions and overlaps with several loops in BbKI (Fig. 9*b*). In particular, two points of collision are found in the vicinity of residues 62 and 71 in BbKI, which are part of the loop that includes the P1 Arg64.

The side chain of Glu218 in the loop that includes the Pro218A insertion in KLK2 and KLK3 collides with the loop containing Ser61 in BbKI in a manner similar to Tyr218 in KLK5 described above (Fig. 9*a*). There appears to be fairly good correlation between the size of the side chain of residue 218 and the potency of the inhibitor (Table 3).

4. Conclusions

BbKI is a plant-derived inhibitor of various proteolytic enzymes. In particular, it is a potent inhibitor of plasma kallikrein and several tissue kallikreins, towards which it exhibits various levels of inhibitory activity. Whereas BbKI is indeed the most potent inhibitor of some enzymes from this family among all proteinaceous inhibitors studied to date, it is not particularly specific and its activity towards some tissue kallikreins is fairly low or totally absent. Although BbKI is called a 'kallikrein inhibitor', it is also active against both trypsin and chymotrypsin.

In this study, we determined a high-resolution structure of the complex of BbKI and trypsin. The structure allowed the mapping of three major interfaces between the enzyme and the inhibitor. Based on this structure, we constructed models of the complexes of several kallikreins with BbKI, evaluating the structural basis of the inhibition of these enzymes, as well as the lack of inhibition of some other members of the family. The models revealed some areas of the enzymes that are important determinants of the potency of inhibition. Substitutions in the enzymes in the specificity-determining pocket that accommodates the P1 Arg64 modify the network of interactions in this area. Another important interface maintains the interactions with the loop in BbKI carrying Ser81-Ser82. As an example, interactions in these two areas act in a concerted way in KLK7, enhancing the inhibition potency of BbKI for this chymotrypsin-like enzyme.

Kallikreins have been found to be important in the progression and metastasis of various types of cancer, and their activity is also associated with other diseases, making them important targets for investigation. The crystal structure of the complex of BbKI with trypsin and the modeling of its interactions with selected kallikreins for which experimental structures are available allowed us to analyze the structural basis of its inhibition of this important family of serine proteases. Differences in the inhibitor–enzyme interactions in the corresponding sites can provide insight into the specificity of individual kallikreins. Comparative analysis of such structural data combined with inhibitory measurements might guide us in the design of a more potent and specific inhibitor targeting a particular kallikrein.

Acknowledgements

We would like to thank Dr Andriy Anishkin (University of Maryland) for the calculations of the energy-minimized structures of selected kallikrein complexes. These models are available from the corresponding author of this manuscript. We acknowledge the use of beamline 22-ID of the Southeast Regional Collaborative Access Team (SER-CAT) located at the Advanced Photon Source, Argonne National Laboratory. Use of the Advanced Photon Source was supported by the US Department of Energy, Office of Science, Office of Basic Energy Sciences under Contract No. W-31-109-Eng-38. The content of this publication is solely the responsibility of the authors and does not necessarily represent the official views or policies of the Department of Health and Human Services, nor does the mention of trade names, commercial products or organizations imply endorsement by the US Government.

Funding information

This work was supported in part by the Intramural Research Program of the NIH, National Cancer Institute, Center for Cancer Research and with Federal funds from the National Cancer Institute, NIH under Contract No. HHSN261200800001E (to ML). JS and MM were supported by project ChemBioDrug CZ.02.1.01/0.0/0.0/16_019/0000729 from the European Regional Development Fund (OP RDE) and the institutional project RVO 61388963.

References

Adams, P. D., Afonine, P. V., Bunkóczi, G., Chen, V. B., Davis, I. W., Echols, N., Headd, J. J., Hung, L.-W., Kapral, G. J., Grosse-Kunstleve, R. W., McCoy, A. J., Moriarty, N. W., Oeffner, R., Read, R. J., Richardson, D. C., Richardson, J. S., Terwilliger, T. C. & Zwart, P. H. (2010). *Acta Cryst.* D**66**, 213–221.

Araújo, A. P. U., Hansen, D., Vieira, D. F., Oliveira, C., Santana, L. A., Beltramini, L. M., Sampaio, C. A. M., Sampaio, M. U. & Oliva, M. L. V. (2005). *Biol. Chem.* **386**, 561–568.

Batista, I. F. C., Oliva, M. L. V., Araujo, M. S., Sampaio, M. U., Richardson, M., Fritz, H. & Sampaio, C. A. M. (1996). *Phytochemistry*, **41**, 1017–1022.

Botos, I. & Wlodawer, A. (2007). *Curr. Opin. Struct. Biol.* **17**, 683–690.

Brito, M. V., de Oliveira, C., Salu, B. R., Andrade, S. A., Malloy, P. M. D., Sato, A. C., Vicente, C. P., Sampaio, M. U., Maffei, F. H. A. & Oliva, M. L. V. (2014). *Thromb. Res.* **133**, 945–951.

Brünger, A. T. (1992). *Nature (London)*, **355**, 472–475.

Cereda, V., Formica, V., Menghi, A., Pellicori, S. & Roselli, M. (2015). *Expert Opin. Investig. Drugs*, **24**, 929–947.

Debela, M., Hess, P., Magdolen, V., Schechter, N. M., Steiner, T., Huber, R., Bode, W. & Goettig, P. (2007). *Proc. Natl Acad. Sci. USA*, **104**, 16086–16091.

De Vonis Bidlingmeyer, U., Leary, T. R. & Laskowski, M. Jr (1972). *Biochemistry*, **11**, 3303–3310.

Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. (2010). *Acta Cryst.* D**66**, 486–501.

Fogaça, S. E., Melo, R. L., Pimenta, D. C., Hosoi, K., Juliano, L. & Juliano, M. A. (2004). *Biochem. J.* **380**, 775–781.

García-Nafría, J., Timm, J., Harrison, C., Turkenburg, J. P. & Wilson, K. S. (2013). *Acta Cryst.* D**69**, 1367–1380.

Helland, R., Otlewski, J., Sundheim, O., Dadlez, M. & Smalås, A. O. (1999). *J. Mol. Biol.* **287**, 923–942.

electronic reprint

Hinkes, S., Wuttke, A., Saupe, S. M., Ivanova, T., Wagner, S., Knörlein, A., Heine, A., Klebe, G. & Steinmetzer, T. (2016). *J. Med. Chem.* **59**, 6370–6386.

Horn, M., Zbodakova, O., Kasparek, P., Srp, J., Haneckova, R., Hradilek, M., Mares, M. & Sedlacek, R. (2018). *Biol. Chem.* **399**, 1085–1089.

Humphrey, W., Dalke, A. & Schulten, K. (1996). *J. Mol. Graph.* **14**, 33–38.

Klauda, J. B., Venable, R. M., Freites, J. A., O'Connor, J. W., Tobias, D. J., Mondragon-Ramirez, C., Vorobyov, I., MacKerell, A. D. Jr & Pastor, R. W. (2010). *J. Phys. Chem. B*, **114**, 7830–7843.

Kumar, P., Singh, M. & Karthikeyan, S. (2011). *Acta Cryst.* D**67**, 131–139.

Kunitz, M. (1947). *J. Gen. Physiol.* **30**, 311–320.

Li, Z., Partridge, J., Silva-Garcia, A., Rademacher, P., Betz, A., Xu, Q., Sham, H., Hu, Y., Shan, Y., Liu, B., Zhang, Y., Shi, H., Xu, Q., Ma, X. & Zhang, L. (2017). *ACS Med. Chem. Lett.* **8**, 185–190.

MacKerell, A. D. Jr, Bashford, D., Bellott, M., Dunbrack, R. L., Evanseck, J. D., Field, M. J., Fischer, S., Gao, J., Guo, H., Ha, S., Joseph-McCarthy, D., Kuchnir, L., Kuczera, K., Lau, F. T., Mattos, C., Michnick, S., Ngo, T., Nguyen, D. T., Prodhom, B., Reiher, W. E., Roux, B., Schlenkrich, M., Smith, J. C., Stote, R., Straub, J., Watanabe, M., Wiórkiewicz-Kuczera, J., Yin, D. & Karplus, M. (1998). *J. Phys. Chem. B*, **102**, 3586–3616.

Majumder, S., Khamrui, S., Dasgupta, J., Dattagupta, J. K. & Sen, U. (2012). *Biochim. Biophys. Acta*, **1824**, 882–890.

Mavridis, K., Avgeris, M. & Scorilas, A. (2014). *Expert Opin. Ther. Targets*, **18**, 365–383.

McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C. & Read, R. J. (2007). *J. Appl. Cryst.* **40**, 658–674.

Murafuji, H., Sakai, H., Goto, M., Imajo, S., Sugawara, H. & Muto, T. (2017). *Bioorg. Med. Chem. Lett.* **27**, 5272–5276.

Murshudov, G. N., Skubák, P., Lebedev, A. A., Pannu, N. S., Steiner, R. A., Nicholls, R. A., Winn, M. D., Long, F. & Vagin, A. A. (2011). *Acta Cryst.* D**67**, 355–367.

Nakahata, A. M., Mayer, B., Neth, P., Hansen, D., Sampaio, M. U. & Oliva, M. L. V. (2013). *Planta Med.* **79**, 227–235.

Odei-Addo, F., Frost, C., Smith, N., Ogawa, T., Muramoto, K., Oliva, M. L. V., Gráf, L. & Naude, R. (2014). *J. Enzyme Inhib. Med. Chem.* **29**, 633–638.

Oliva, M. L. V., Mendes, C. R., Juliano, M. A., Chagas, J. R., Rosa, J. C., Greene, L. J., Sampaio, M. U. & Sampaio, C. A. M. (1999). *Immunopharmacology*, **45**, 163–169.

Oliva, M. L. V., Mendes, C. R., Santomauro-Vaz, E. M., Juliano, M. A., Mentele, R., Auerswald, E. A., Sampaio, M. U. & Sampaio, C. A. M. (2001). *Curr. Med. Chem.* **8**, 977–984.

Oliva, M. L. V. & Sampaio, U. M. (2008). *Biol. Chem.* **389**, 1007–1013.

Oliva, M. L. V., Silva, M. C. C., Sallai, R. C., Brito, M. V. & Sampaio, M. U. (2010). *Biochimie*, **92**, 1667–1673.

Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.

Pampalakis, G. & Sotiropoulou, G. (2007). *Biochim. Biophys. Acta*, **1776**, 22–31.

Patil, D. N., Chaudhary, A., Sharma, A. K., Tomar, S. & Kumar, P. (2012). *FEBS J.* **279**, 4547–4564.

Phillips, J. C., Braun, R., Wang, W., Gumbart, J., Tajkhorshid, E., Villa, E., Chipot, C., Skeel, R. D., Kalé, L. & Schulten, K. (2005). *J. Comput. Chem.* **26**, 1781–1802.

Prassas, I., Eissa, A., Poda, G. & Diamandis, E. P. (2015). *Nature Rev. Drug Discov.* **14**, 183–202.

Renko, M., Sabotič, J. & Turk, D. (2012). *Biol. Chem.* **393**, 1043–1054.

Riley, B. T., Ilyichova, O., Costa, M. G. S., Porebski, B. T., de Veer, S. J., Swedberg, J. E., Kass, I., Harris, J. M., Hoke, D. E. & Buckle, A. M. (2016). *Sci. Rep.* **6**, 35385.

Scheidig, A. J., Hynes, T. R., Pelletier, L. A., Wells, J. A. & Kossiakoff, A. A. (1997). *Protein Sci.* **6**, 1806–1824.

Song, H. K. & Suh, S. W. (1998). *J. Mol. Biol.* **275**, 347–363.

Sotiropoulou, G. & Pampalakis, G. (2012). *Trends Pharmacol. Sci.* **33**, 623–634.

Souza-Pinto, J. C., Oliva, M. L. V., Sampaio, C. A. M., Damas, J., Auerswald, E. A., Limãos, E., Fritz, H. & Sampaio, M. U. (1996). *Immunopharmacology*, **33**, 330–332.

Swedberg, J. E., de Veer, S. J., Sit, K. C., Reboul, C. F., Buckle, A. M. & Harris, J. M. (2011). *PLoS One*, **6**, e19302.

Tanaka, S., Sawaya, M. R. & Yeates, T. O. (2010). *Science*, **327**, 81–84.

Turk, B. (2006). *Nature Rev. Drug Discov.* **5**, 785–799.

Vadivel, K., Ponnuraj, S. M., Kumar, Y., Zaiss, A. K., Bunce, M. W., Camire, R. M., Wu, L., Evseenko, D., Herschman, H. R., Bajaj, M. S. & Bajaj, S. P. (2014). *J. Biol. Chem.* **289**, 31647–31661.

Winn, M. D., Ballard, C. C., Cowtan, K. D., Dodson, E. J., Emsley, P., Evans, P. R., Keegan, R. M., Krissinel, E. B., Leslie, A. G. W., McCoy, A., McNicholas, S. J., Murshudov, G. N., Pannu, N. S., Potterton, E. A., Powell, H. R., Read, R. J., Vagin, A. & Wilson, K. S. (2011). *Acta Cryst.* D**67**, 235–242.

Zhou, D., Hansen, D., Shabalin, I. G., Gustchina, A., Vieira, D. F., de Brito, M. V., Araújo, A. P. U., Oliva, M. L. V. & Wlodawer, A. (2015). *Acta Cryst.* F**71**, 1055–1062.

Zhou, D., Lobo, Y. A., Batista, I. F. C., Marques-Porto, R., Gustchina, A., Oliva, M. L. V. & Wlodawer, A. (2013). *PLoS One*, **8**, e62252.

electronic reprint