JMB

Crystal Structure of a 16 kDa Double-headed Bowman-Birk Trypsin Inhibitor from Barley Seeds at 1.9 Å Resolution

Hyun Kyu Song, Young Sil Kim, Jin Kuk Yang, Jinho Moon and Jae Young Lee and Se Won Suh*

Department of Chemistry College of Natural Sciences Seoul National University Seoul, 151-742, Korea

The Bowman-Birk trypsin inhibitor from barley seeds (BBBI) consists of 125 amino acid residues with two inhibitory loops. Its crystal structure in the free state has been determined by the multiwavelength anomalous diffraction (MAD) method and has been refined to a crystallographic R-value of 19.1 % for 8.0-1.9 Å data. This is the first report on the structure of a 16 kDa double-headed Bowman-Birk inhibitor (BBI) from monocotyledonous plants and provides the highest resolution picture of a BBI to date. The BBBI structure consists of 11 β-strands and the loops connecting these β-strands but it lacks α-helices. BBBI folds into two compact domains of similar tertiary structure. Each domain shares the same overall fold with 8 kDa dicotyledonous BBIs. The five disulfide bridges in each domain are a subset of the seven disulfide bridges in 8 kDa dicotyledonous BBIs. Two buried water molecules form hydrogen bonds to backbone atoms in the core of each domain. One interesting feature of this two-domain inhibitor structure is that the two P1 residues (Arg17 and Arg76) are approximately 40 Å apart, allowing the two reactive-site loops to bind to and to inhibit two trypsin molecules simultaneously and independently. The conformations of the reactive-site loops of BBBI are highly similar to those of other substrate-like inhibitors. This structure provides the framework for modeling of the 1:2 complex between BBBI and trypsin.

© 1999 Academic Press

Keywords: Bowman-Birk trypsin inhibitor; double-headed inhibitor; gene duplication; monocotyledonous plant; reactive-site loop

*Corresponding author

Introduction

Plant seeds contain a large number of different types of serine protease inhibitors, which block trypsin and chymotrypsin from animal, fungal, and bacterial origins. Among many types of protease inhibitors from plants, Kunitz-type and Bowman-Birk type inhibitors have been most extensively studied. Members of the Kunitz-type inhibitor family have a molecular mass of about 20 kDa and two disulfide bridges, whereas those of the Bowman-Birk family are smaller (typically ~8 kDa or ~16 kDa) and richer in disulfide bonds (Birk, 1987). A Bowman-Birk protease inhibitor (BBI) was first isolated from leguminous seeds (Bowman, 1946) and its molecular properties were characterized (Birk *et al.*, 1963). BBIs from dicotyledonous seeds such as soybean are small (~8 kDa), double-headed proteins. That is, the 8 kDa inhibitor contains two reactive sites, each of which

specifically inhibits trypsin or chymotrypsin. These BBIs are stable at cooking temperature and toward the acidic pH found in the digestive system of humans and animals (Birk, 1987). They have attracted much interest due to their anticarcinogenic activity (Kennedy, 1993) and immune-stimulating property (Harms-Ringdahl *et al.*, 1979). Human populations consuming a large amount of BBI in their diet have been shown to exhibit lower rates of colon, breast, prostate, and skin cancers (Birk, 1993). In contrast, monocotyledonous seeds such as barley contain 8 kDa single-headed inhibitors and 16 kDa inhibitors (Prakash *et al.*, 1996). In the evolutionary pathway, it is likely that one of the two reactive-sites of 8 kDa double-headed inhibitors from dicotyledonous seeds became non-functional, resulting in the 8 kDa single-headed inhibitors in monocotyledonous seeds (Tashiro *et al.*, 1990). Subsequently, 16 kDa inhibitors such as BBBI might have evolved in monocotyledons by gene duplication. There are no experimental data on the anticarcinogenic property of this inhibitor in the literature.

Several trypsin inhibitors are found in various tissues of barley (Mikola & Kirsi, 1972; Ogiso *et al.*, 1975; Boisen & Djurtoft, 1982; Odani *et al.*, 1983; Nagasue *et al.*, 1988). Among them, a 16 kDa trypsin inhibitor purified from the barley rootlet was characterized as a member of BBI family. It is double-headed, since it contains two reactive-sites for inhibiting trypsin in a molar ratio of 1:2. The amino acid sequence of this barley rootlet trypsin inhibitor (BRTI) was determined (Nagasue *et al.*, 1988) and it was found to consist of 124 amino acid residues. BRTI shows an intramolecular sequence identity of 55 % between the amino-terminal half (residues 1-62) and the carboxy-terminal half (residues 63-124). A comparison with soybean Bowman-Birk protease inhibitor (SBBI) indicated that BRTI has two reactive-sites for trypsin at Arg17 and Arg75.

Three-dimensional structures of several 8 kDa dicotyledonous BBIs have been reported. They include structures of tracy bean inhibitor PI-II (Chen *et al.*, 1992), peanut inhibitor A-II (Suzuki *et al.*, 1993), soybean inhibitor (Werner & Wemmer, 1992; Voss *et al.*, 1996), and pea seed inhibitor PsTI-IVb (de la Sierra *et al.*, 1999), which have been analyzed in the free form, and those of the inhibitors from adzuki bean (Tsunogae *et al.*, 1986) and mung bean (Lin *et al.*, 1993; Li *et al.*, 1994), which have been determined in complex with trypsin. However, no crystal structure of a 16 kDa BBI from monocots has been reported. It is necessary to compare the tertiary structure of BBIs in cereal grains, a monocot, with those in leguminous plants, a dicot, in order to understand the difference in structure and function, and the evolution of BBIs. Previously, we reported the purification, crystallization and preliminary X-ray crystallographic data of the 16 kDa BBI from barley seeds (Song & Suh, 1998b). Here, we have determined its crystal structure in the free state at 1.9 Å resolution

by the multiwavelength anomalous diffraction (MAD) method (Hendrickson & Ogata, 1997). This is the first report on the three-dimensional structure of a 16 kDa monocotyledonous BBI and, furthermore, it provides the highest resolution picture of a Bowman-Birk-type inhibitor to date. This structure of BBBI provides a basis for modeling studies to gain insight into the interactions between this 16 kDa inhibitor and the target protease.

Results and Discussion

Structure determination and model quality

The molecular replacement method was first tried to solve the structure of free BBBI but it failed to yield a solution. Next, the conventional multiple isomorphous replacement method was attempted. An extensive search of heavy-atom derivatives produced, however, only one useful derivative, that of K_2PtCl_4, possibly due to the scarcity of reactive side-chains in BBBI. The electron density map derived from this Pt-derivative alone was uninterpretable. Therefore, the phase problem was solved using MAD techniques (Hendrickson & Ogata, 1997). The resulting electron density map was readily interpretable in terms of the known chemical amino acid sequence, except that the MAD-phased electron density clearly indicated that a residue needs to be inserted between Pro41 and Met42 in the reported amino acid sequence (Nagasue *et al.*, 1988). The inserted residue could be identified as serine (Ser101 in Figure 1), which is consistent with the presence of the same residue at the corresponding position in the carboxy-terminal half (designated Ser100 by Nagasue *et al.*, 1988). The discrepancy may be due to the different barley tissues used (the rootlet for sequencing and the seeds for this study) or due to a possible error in the chemical sequencing. Here, Ser42 has been inserted and the residue numbers thereafter are increased by 1.

The model accounts for 120 of 125 residues of BBBI and 41 water molecules. It has been refined to a crystallographic *R*-value of 19.1 % for 8.0-1.9 Å data with excellent stereochemistry. Missing from the inhibitor model are the three N-terminal residues (Ala1, Gly2, and Lys3) and two C-terminal residues (Pro124 and Arg125) due to missing electron density. The side-chains of Lys8, Lys37, Lys89, and Arg120, mostly in the loop regions, have poor electron density. Among 96 non-glycine and non-proline residues, the numbers of residues lying in the most favored, additionally allowed, generously allowed, and disallowed regions in the Ramachandran plot are 90 (93.8 %), 6 (6.3 %), 0, and 0, respectively. Table 1 summarizes the refinement statistics as well as model quality parameters. An average *B*-factor for the main-chain atoms of each residue in the BBBI model is plotted as a function of residue number in Figure 2. Besides N and C termini, several loops (34, 89-90, 103-107) show
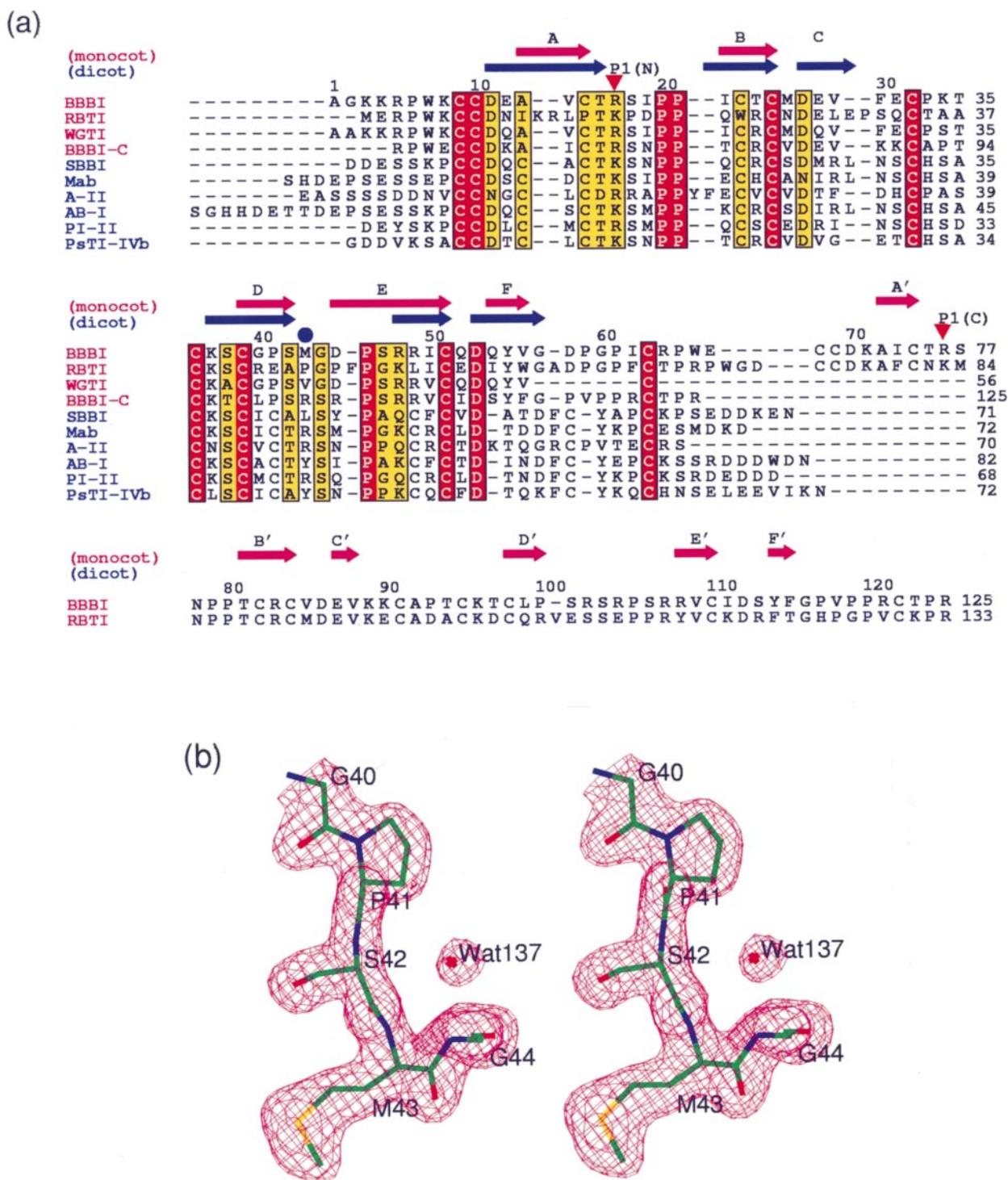
Figure 1. (a) Sequence alignment of monocotyledonous Bowman-Birk inhibitors (from barley seeds (BBBI), rice bran (RBTI), wheat germ (WGTI)) and dicotyledonous BBIs (from soybean (SBBI), mung bean (Mab), peanut (A-II), adzuki bean (AB-I), tracy bean (PI-II), and pea (PsTI-IVb)). The P1 residues in both the N and C domains of monocots and the N subdomain of dicots are marked by red triangles, while the P1 residues in the C subdomain of dicots are marked by a blue circle. Secondary structural elements of BBBI (monocot) and SBBI (dicot) are indicated above the sequence in magenta and blue, respectively. The secondary structure elements of BBBI were assigned by PROCHECK (Laskowski *et al.*, 1993), while those of SBBI follow the assignment made by Voss *et al.* (1996). For BBBI, β-strands are labeled sequentially A–F (C is missing) and A′–F′ in the N and C domains, respectively, to be consistent with the previous assignment for SBBI. This Figure was produced with ALSCRIPT (Barton, 1993). (b) Final $(2F_o − F_c)$ electron density map around Ser42, which is missing from the reported BRTI sequence (Nagasue *et al.*, 1988). The map was calculated using 15-1.9 Å data and contoured at 1.0 σ (0.33 e Å^{-3}).
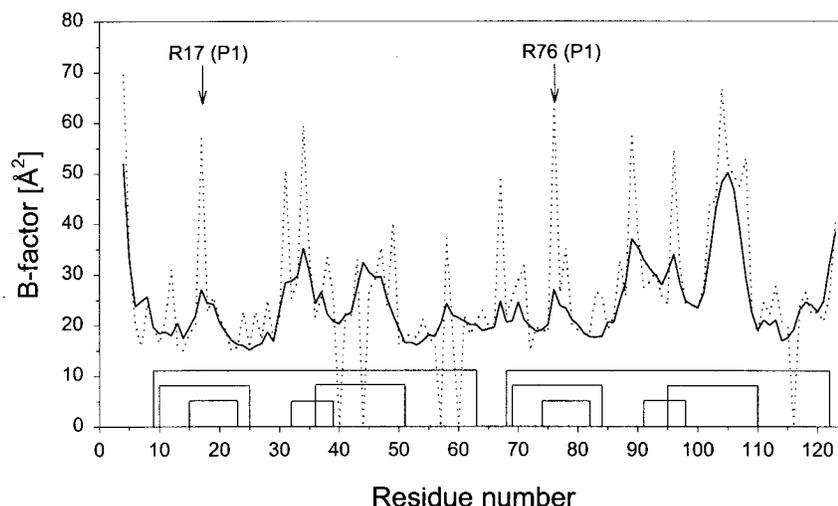
Figure 2. *B*-factor plot for the BBBI model. Average *B*-factors for the main-chain atoms (broken lines) and side-chain atoms (dotted lines) of each residue are plotted as a function of residue number. P1 residues (Arg17 and Arg76) are indicated as well as the ten disulfide bridges (9:63, 10:25, 15:23, 32:39, 36:51 in the N domain and 68:122, 69:84, 74:82, 91:98, 95:110 in the C domain).

overall main-chain *B*-factors higher than 35 Å² (Figure 2). This BBBI exhibits a relatively low overall *B*-factor (26.6 Å²) compared with other protease inhibitors.

Overall structure

The BBBI structure consists of 11 β-strands and the loops connecting these β-strands without any α-helix (Figure 3). It folds into two compact domains (termed N and C domain, respectively) with approximate dimensions of 52 Å × 50 Å × 25 Å. Each domain shares the same overall fold with other 8 kDa dicotyledonous BBIs, which are composed of two subdomains (Voss *et al.*, 1996; Bode & Huber, 1992). And the five disulfide bridges in each domain (9:63, 10:25, 15:23, 32:39, 36:51 in the N domain and 68:122, 69:84, 74:82, 91:98, 95:110 in the C domain) are a subset of the seven in 8 kDa dicotyledonous BBIs. When the structure of BBBI was compared with those in the

Table 1. Crystallographic refinement statistics

Resolution range (Å)	8.0-1.9
Number of unique reflections (no cutoff)	13,294
Completeness (%)	87.9
R_{cryst}/R_{free} (%)[a]	19.1/21.9
Number of non-hydrogen atoms	
Protein (number of residues)	903 (120)
Water	41
Average *B*-factors (Å²)	
Main-chain (number of atoms)	24.2 (480)
Side-chain (number of atoms)	29.3 (423)
Water	33.3
r.m.s. deviation from ideal geometry	
Bond lengths (Å)	0.008
Bond angles (deg.)	1.09
Dihedral angles (deg.)	27.7
Improper angles (deg.)	1.71
Ramachandran outlier	0

[a] $R = \Sigma||F_o| - |F_c||/\Sigma|F_o|$, where R_{cryst} and free *R*-factors are calculated using the 95 % working set and 5 % test set of reflections, respectively.

database of DALI (Holm & Sander, 1993), a notable similarity (*Z* score > 2.0) was found with tracy bean PI-II inhibitor only (PDB code, 1PI2; Chen *et al.*, 1992), which is an 8 kDa dicotyledonous BBI. The C domain of BBBI is structurally more similar to PI-II than the N domain (C domain *versus* 1PI2, *Z* = 6.3; N domain *versus* 1PI2, *Z* = 5.5). In the N domain, each of the two subdomains is formed by an antiparallel β-sheet of three strands (A, B, and F) and another small antiparallel β-sheet of two strands (D and E), respectively. In the C domain, the second subdomain is formed by a small antiparallel β-sheet of three strands (C′, D′, and E′). The labels of the secondary structure elements in the C domain are primed here for the convenience of discussion. Each domain has approximate dimensions of 38 Å × 22 Å × 20 Å. There exist only a few interactions between the two domains; the side-chain oxygen atoms of Asp11 in the N domain make hydrogen bonds to the main-chain nitrogen atom of Arg64 and the side-chain hydroxyl group of Tyr114 in the C domain. The two reactive-sites (Arg17-Ser18 and Arg76-Ser77; P1-P1′) are located at opposite sides of the inhibitor structure on protruding loops between strands A and B in the N domain (or A′ and B′ in the C domain). Interestingly, the longest distance between the two P1 residues is approximately 40 Å, very similar to the case of the 8 kDa double-headed BBIs from dicotyledonous seeds (36-40 Å). The relative orientation of two domains of BBBI first observed in this structure is much different from the dimeric arrangement in the crystal structure of PsTI-IVb (de la Sierra *et al.*, 1999).

A superposition of the two domains of BBBI shows a high degree of similarity for the reactive-site loops and β-strands A and B, but not for the remaining parts (Figure 4(a)). The r.m.s. difference between 59 equivalent Cα atom pairs (for residues 5-63 in the N domain and 64-122 in the C domain) is as large as 2.2 Å, while that for each reactive-site loop (13 equivalent Cα atoms for residues 13-25 in

Figure 3. (a) A stereo ribbon diagram of BBBI. Arrows represent β-strands. P1 residues (Arg17 and Arg76) are indicated as well as the ten disulfide bridges (9:63, 10:25, 15:23, 32:39, 36:51 in the N domain and 68:122, 69:84, 74:82, 91:98, 95:110 in the C domain). The secondary structure elements are labeled and approximately every tenth residue is labeled and marked by a dot. (b) A stereo ribbon diagram of SBBI in the same orientation as the N domain of BBBI in (a). This Figure was drawn by MOLSCRIPT (Kraulis, 1991).

Figure 4. (a) Stereo diagram showing the superposition of C^α atoms between the N and C domains of the BBBI structure. Green and magenta lines represent the N and the C domains, respectively. Blue lines represent the P1 residue (Arg17) in the N domain, while red lines represent the P1 residue (Arg76) in the C domain. Disulfide bridges are in yellow. (b) A stereo diagram showing the superposition of C^α atoms in the structures of the N domain of BBBI and soybean BBI. Green and magenta lines represent BBBI and soybean BBI, respectively. Blue lines represent the P1 residue (Arg17) in N domain of BBBI, while red lines represent P1 residues (Lys16 and Leu43) in soybean BBI. Every tenth residue is labeled.

the N domain or residues 72-84 in the C domain) is only 0.50 Å. The largest difference occurs around the loops corresponding to the anti-chymotryptic reactive-site loop in 8 kDa (double-headed) BBIs (the regions around Met43 in the N domain and Arg102 in the C domain). These regions show high *B*-factors for the main-chain atoms (Figure 2), suggesting the inherent conformational flexibility of these regions. In the case of the 8 kDa BBIs from leguminous seeds, the flexibility of the second reactive-site loop is reduced by the extra disulfide bond.

Solvent structure

Our BBBI model has been refined at 1.9 Å, the highest resolution among the reported BBI structures to date, and a total of 41 ordered solvent molecules, all modeled as water, have been located. Most of them are bound to the protein surface. A water molecule (Wat161) forms a hydrogen bond with the side-chain oxygen atom (OG) of Ser77 in the reactive-site loop of the C domain (Figure 5(b)). However, it does not seem to affect the reactive-site loop conformation significantly. An equivalent water molecule is not present in either the N domain of BBBI or dicot BBI structures. Thus this

water molecule is not likely to play a significant role in complex formation with the target protease.

Two buried water molecules are present as an integral part of each domain (Figure 6). These water molecules are located symmetrically around the intramolecular pseudo 2-fold axis that relates the two subdomains of either the N or the C domain, and each water molecule donates two hydrogen bonds to two peptide $C = O$ groups and accepts one from the peptide N-H group (Figure 6). They show *B*-factors lower than the average (10.3 and 22.7 Å^2 for Wat128 and Wat147 in the N domain; 20.7 and 19.9 Å^2 for Wat131 and Wat141 in the C domain). This suggests that these two water molecules are tightly bound to specific pos-
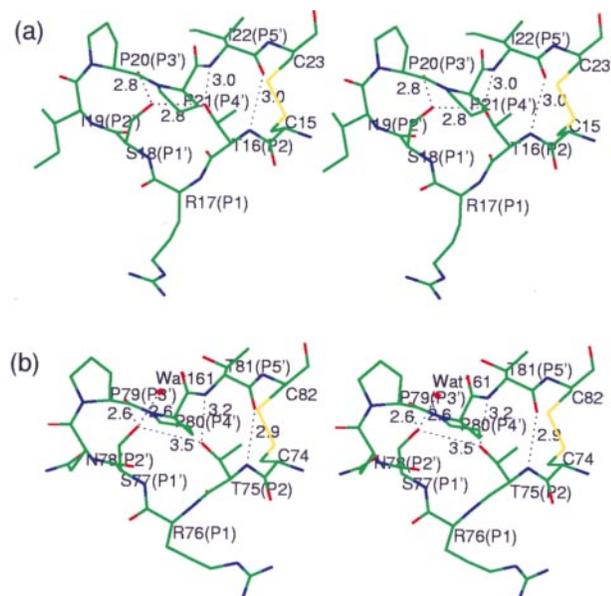
Figure 5. Stereo diagram showing the role of the P2 residue in stabilizing the reactive-site loop conformation of BBBI. (a) The N domain; (b) the C domain. The conserved threonine residue at P2 is involved in extensive hydrogen bondings with several residues in the reactive-site loop. The distances between the hydrogen bonding atoms are given.

Figure 6. Stereo diagram showing the interactions between two buried water molecules and surrounding backbone atoms. (a) The N domain; (b) the C domain. The distances between the water molecule and its hydrogen bonding atoms of the inhibitor are given.

itions within each domain. Similar water molecules have been observed in other 8 kDa dicotyledonous BBI structures (Voss *et al.*, 1996; Lin *et al.*, 1993; Suzuki *et al.*, 1993). They seem to be essential to maintain the tertiary structures of BBIs. Another example of a buried water molecule serving a similar structural role is found in the Kunitz-type soybean trypsin inhibitor (STI) family (Onesti *et al.*, 1991; Song & Suh, 1998a), in which a conserved water molecule is present in the center of a triangular structure made up of three β-strands, almost on the pseudo 3-fold axis.

Reactive-site loops

Each of the two reactive-site loops of BBBI is constrained by a disulfide bridge at their extremities (Cys15-Cys23 in the N domain and Cys74-Cys82 in the C domain). When viewed from the direction of Figure 5, the reactive-site loop displays a heart-like shape, as pointed out previously for peanut A-II inhibitor (Suzuki *et al.*, 1993). This characteristic structure is maintained by the polyproline-II conformation at the P3′ and P4′ positions, and the internal hydrogen bond network involving Thr16 or Thr75 at the P2 position and Ser18 or Ser77 at the P1′ position (Figure 5) (Bode & Huber, 1992; Flecker, 1995).

In general, the reactive-site loop of a standard mechanism protease inhibitor is relatively flexible in its free state and it becomes less mobile upon its binding to target proteases. This is reflected by a remarkable reduction of the atomic *B*-factors of the reactive-site loop upon complexation, which is not accompanied by any significant structural change in the reactive-site loop. Such behavior has been observed for the reactive-site loops of A-II (Suzuki *et al.*, 1993) and other inhibitors (Read & James, 1986; Song & Suh, 1998a). However, the overall *B*-factor of the reactive-site loops of BBBI is relatively low even in the free state (Figure 2). This may be partly contributed by the stabilization of the reactive-site loops through intermolecular contacts with the symmetry-related molecules in the crystal. These contacts involve direct hydrogen bonds (Cys15 (P3) O···N Arg102#, 3.0 Å; Arg17 (P1) N···O Pro100#, 2.6 Å; Cys74 (P3) O···OE1 Gln52#, 3.2 Å). However, the conformations of the two reactive-site loops of BBBI are not affected significantly by the crystal packing, as evidenced by the lack of any significant deviation when they are compared with those of other BBI structures (see below).

Comparison of inhibitor structures

The alignment of amino acid sequences of several BBIs is shown in Figure 1(a). The sequence identity among monocotyledonous BBIs is higher

than 50%, whereas that between monocotyledonous and dicotyledonous BBIs ranges between 30 and 33%. BBBI shows an intramolecular sequence identity of 55% between the N and C domains. As shown in Figure 1(a), the reactive-site loops (residues Cys15-Cys23 and Cys74-Cys82) are the most conserved regions of BBI sequences. Disulfide bridges are also well conserved; the five disulfide bridges in each domain of 16 kDa monocotyledonous BBIs are a subset of the seven in 8 kDa dicotyledonous BBIs. Among the two missing disulfide bridges (Cys12-Cys58 and Cys41-Cys49 in SBBI sequence), the latter constrains the anti-chymotryptic reactive-site loop in 8 kDa double-headed BBIs from dicotyledonous plants. The loss of this disulfide bond in BBBI may be partly responsible for the inactivity of its corresponding loop in each domain. The N domain of the BBBI model is superposed with its C domain and the whole molecule of SBBI in Figure 4(b). It can be seen that the structural variability is greater for the regions that are not involved in protease recognition. Large deviations in these regions (residues 40-49 and 99-108 in the N and C domains, respectively) are due to the loss of disulfide bridges and to high *B*-factors (Figure 2). Therefore, they are more divergent in tertiary structure as well as in primary structure among BBIs. The results of a disulfide-deletion analysis of the trypsin-reactive subdomain of SBBI pointed to the significance of the polar subdomain interface as a major refolding determinant (Philipp *et al.*, 1998). Arg28 and His33 of SBBI, which have been suggested to be responsible for inter-subdomain effects observed in a disulfide-deletion in SBBI, are replaced by non-polar residues in BBBI (Val29 and Pro33 in the N domain; Val88 and Ala92 in the C domain). And, consequently, the conformational distortions due to the missing disulfide bond (corresponding to Cys41-Cys49 in SBBI) are expected to be restricted to the subdomains around Met43 and Arg102 in BBBI (corresponding to Leu43 in SBBI).

Despite large overall structural variations among BBIs, the inhibitory-loop regions show a highly similar conformation. And the magnitude of the deviations is roughly what is expected from the estimated coordinate errors. For a detailed comparison of the inhibitory loop conformations, six BBI models are superposed in Figure 7(a) and the results of the superpositions are summarized in Table 2, which lists the r.m.s. differences for nine C^α atom pairs (P3-P6'). Table 2 gives the results for comparing the inhibitor structures in the free and complexed states. When the reactive-site loops of the N and C domains are compared, the r.m.s. deviation is 0.46 Å. Superpositions of the BBBI reactive-site loops in the N and C domains with the corresponding loop in SBBI, a prototypic dicotyledonous BBI, give r.m.s. differences of 0.24 and 0.40 Å (Table 2). When the free BBBI is compared with BBIs in complex with trypsin (AB-I and Mab in Table 2), the r.m.s. deviations are between 0.22 Å and 0.40 Å. This indicates that the conformations of the reactive-site loops of different BBIs are highly similar and are little affected by complex formation.

When C^α atoms of P2-P3' residues of BBBI are superposed with the corresponding atoms of other inhibitors, the r.m.s. deviations are 0.80 Å for BPTI (PDB code 4PTI), 0.66 Å for STI (PDB code 1AVW), and 0.96 Å for OMTKY3 (PDB code 1CHO), respectively (Figure 7(b)). Therefore, the conformations of the central portions of the inhibitory loops are very similar. In contrast, the conformations outside the P2-P3' residues can be much different among the inhibitors. The extension of the reactive-site loop of BBI family members (as measured by the distance between C^α atoms of residues P3 and P4') is among the smallest (5.8 Å for BBBI and 5.9 Å for SBBI), compared with 15.1 Å for BPTI (Kunitz family), 10.3 Å for STI (STI-Kunitz family), 13.8 Å for MCTI (squash family), 13.5 Å for OMTKY3 (Kazal family), and 15.2 Å for eglin c (PI-1 family). This narrow reactive-site loop, instead of an extended conformation, is a consequence of the *cis*-proline residue at the P3' position in BBI family members (Figure 1) (Tsunogae *et al.*, 1986). The reactive-site loops of different classes of inhibitors are stabilized in different ways. The reactive-site loops of BPTI (Kunitz), Kazal, squash, PI-1, and BBI families are constrained at their extremities by disulfide bridges that could limit their conformational freedom. No disulfide bridge adjacent to the reactive-site is present in STI or eglin c (PI-1 family). In the case of STI, the reactive-site loop is held in an inhibitory conformation by an extensive hydrogen bond network involving the conserved residue Asn13. The conformation of the reactive-site loops of BBBI is also stabilized by the hydrogen bonding network

Table 2. r.m.s. differences (in Å) of the nine matching C^α atoms of P3-P6' residues in the inhibitory loop of Bowman-Birk inhibitors

	BBBI-C	SBBI-N	SBBI-C	AB-I[a]	Mab[a]
BBBI-N	0.46	0.24	0.39	0.31	0.22
BBBI-C		0.40	0.49	0.35	0.40
SBBI-N			0.29	0.30	0.32
SBBI-C				0.49	0.46
AB-I					0.27

[a] Only the first of the two inhibitory loops is contained in these models. These two are in the complexed state.
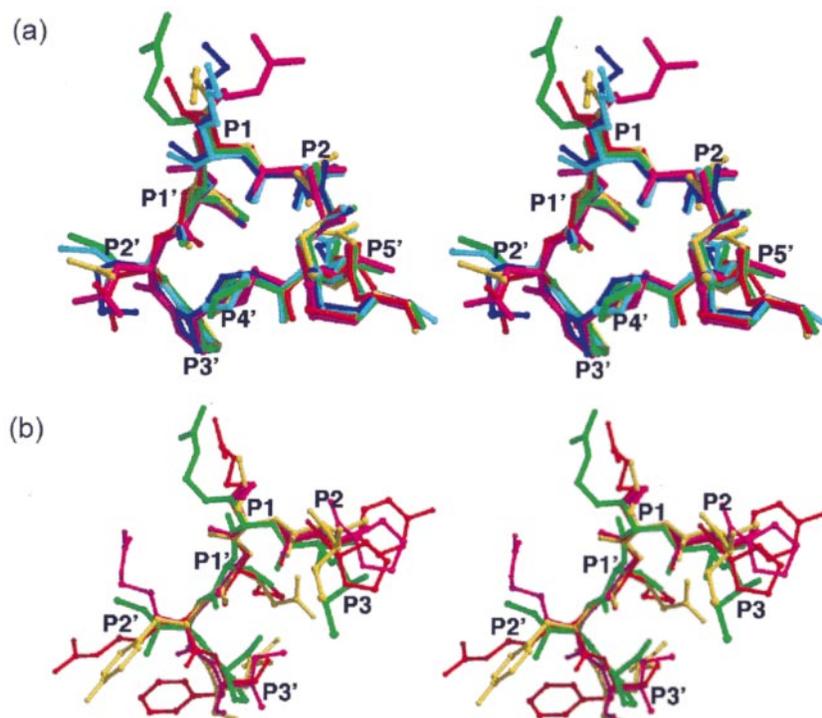
Figure 7. (a) Stereo diagram showing the superposition of the reactive-site loops of known BBI structures. Green, violet, yellow, red, blue, and cyan lines represent the N domain of BBBI, the C domain of BBBI, the N subdomain of SBBI, the C subdomain of SBBI, AB-I and Mab, respectively. (b) Stereo diagram showing the superposition of the reactive-site loops of other inhibitor family structures. Green, violet, yellow, and red lines represent the N domain of BBBI, BPTI, OMTKY3, and STI, respectively.

formed by the P2 threonine residue in each domain (Figure 5). In short, the central portions (P2-P3′) of the reactive loops in the above inhibitors have very similar main-chain conformations despite large structural differences elsewhere.

Interactions with trypsin molecules

A single polypeptide chain of this double-headed inhibitor can bind up to two molecules of trypsin. In order to understand the interactions between the inhibitor and trypsin, modeling studies were undertaken. The inhibitory loop of the N domain is most similar to Mab, while that of the C domain is most similar to AB-I (Table 2). Therefore, the trypsin molecules bound to BBBI have been modeled assuming that they bind to the inhibitor in a manner similar to that observed in the crystal structures of the Mab:bovine pancreatic trypsin (BPT) complex (Li *et al.*, 1994) or the AB-I:BPT complex (Tsunogae *et al.*, 1986). The two trypsins will be referred to as E (interacting with the N domain) and F (interacting with the C domain), respectively. The hypothetical model of the BBBI:BPT complex was further optimized by the program MULTIDOCK (Jackson *et al.*, 1998) and manual adjustments of a few side-chains. The results of modeling studies indicate that each of the two exposed reactive-site loops of BBBI readily slot into the active-site cleft of BPT (Figure 8). In our model of the 1:2 complex, bound trypsin molecules come very close, making slight contacts with each other. One side of a trypsin molecule (Ser170E and Pro173E) makes van der Waals con-

tacts with the other side of another trypsin molecule (Ser147F, Gly148F, and Thr149F) (Figure 8). There is no hydrogen bond interaction between the two trypsin molecules. It is interesting to note that the two P1 residues (Arg17 and Arg76) are approximately 40 Å apart in this two-domain inhibitor structure, allowing the two reactive-site loops to bind to and to inhibit two trypsin molecules simultaneously and independently. It has been suggested that a minimum distance requirement between the two binding sites of 8 kDa double-headed BBIs for simultaneously binding two protease molecules is 39 Å (de la Sierra *et al.*, 1999). However, the relative spatial orientations of two protease molecules bound to double-headed BBIs from dicotyledonous and monocotyledonous seeds are markedly different (Figure 3). This is because the two reactive-site loops in 8 kDa double-headed dicotyledonous BBIs are related by a pseudo 2-fold axis relating the two subdomains, whereas those in 16 kDa double-headed BBBI are not.

Most of the contacts between BBBI and BPT involve six residues (P4-P2′) in the reactive-site loops. And the mode of interaction is very similar to other BBI:protease complexes (Tsunogae *et al.*, 1986; Lin *et al.*, 1993; Li *et al.*, 1994; de la Sierra *et al.*, 1999; Luckett *et al.*, 1999). In our hypothetical model, each of the P1 residues (Arg17 and Arg76) makes the most extensive hydrogen bonds with BPT and its side-chain atoms occupy their expected positions in the primary binding pocket of BPT (S1 pocket). The guanidium group of P1 residues makes an ionic interaction with the carboxylate
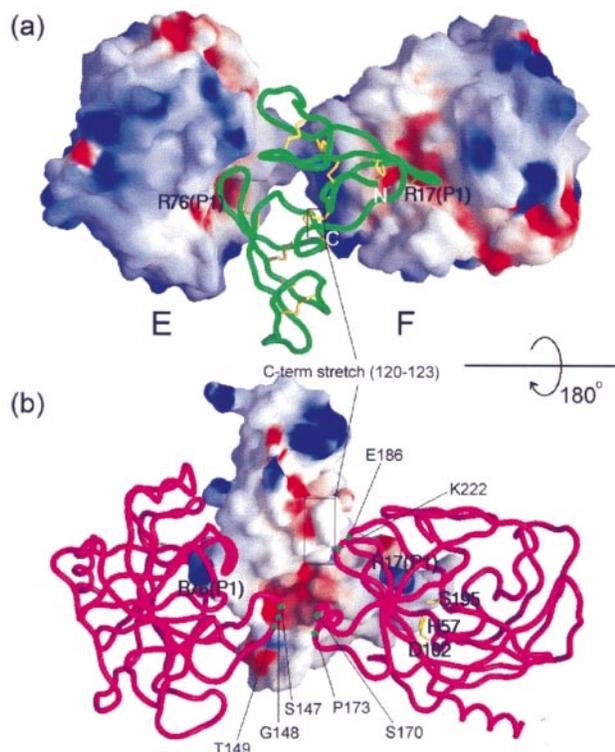
Figure 8. A hypothetical model for the 1:2 complex between BBBI and two trypsin molecules. (a) The backbone model of BBBI (green tubes) and the electrostatic potential surface of trypsin molecules. (b) The backbone model of trypsins (magenta tubes) and the electrostatic potential surface of BBBI. The view in (b) is obtained by a 180° rotation of (a) around a horizontal axis. Positively charged regions are blue and negatively charged regions are red. The N and C-terminal and P1 residues in BBBI and the catalytic triad (Asp102-His57-Ser195) in trypsin are labeled. This Figure was generated using GRASP (Nicholls, 1992).

group of Asp189 in each BPT. P1 carbonyl carbon atoms are in the ''sub-van der Waals'' distances from trypsin Ser195 OG atom (2.90 and 2.60 Å, respectively). Extensive mutagenesis studies on the P2 and P2′ positions of BBI showed that a threonine residue at the P2 position provides an optimal inhibition and the best residue at the P2′ position is isoleucine (McBride et al., 1998; Gariani et al., 1999). All available sequences of BBIs showed that this threonine residue at the P2 position is highly conserved (Figure 1(a)). In BBBI, the P2 residues are Thr16 and Thr75, respectively. They form the hydrogen bond network that maintains the characteristic conformation of the inhibitory loops. Our modeling study shows that the P2′ residues (Ile19 and Asn78) make the second most significant interactions with trypsin next to the P1 residues. This is consistent with the suggestion that the P2′ residue is important in determining the specificity of the inhibitor towards its cognate enzyme (Gariani et al., 1999). In addition to the reactive-site, a C-terminal

stretch of BBBI (residues Pro120, Arg121, Cys122, and Thr123; Pro124 and Arg125 are missing from our model) makes contact with trypsin molecule E (Figure 8(a)). The major interacting residues in trypsin E are Glu186 and Lys222, respectively. The C-terminal tail of the pea seed inhibitor, PsTI-IVb, also makes contact with the trypsin molecule in the model of the PsTI-IVb:trypsin:chymotrypsin ternary complex (de la Sierra et al., 1999). However, the isoform PsTI-II displays the same K_M value for trypsin as PsTI-IVb, despite the missing C-terminal tail. In the crystal structure of the AB-I:trypsin complex, only the peptide 55-63 of the inhibitor was visible in the electron density map and thus it is not clear whether the C-terminal tail interacts with trypsin (Tsunogae et al., 1986). Therefore, a structural analysis of the complex is necessary to delineate more precisely the role of the C-terminal tail of BBBI in trypsin binding.

Materials and Methods

Preparation of Pt-derivative crystals

Purification, crystallization, and X-ray data collection of native BBBI have been reported (Song & Suh, 1998b). The Pt-derivative for MAD phasing (Hendrickson & Ogata, 1997) was prepared as follows. A crystal was soaked at a final concentration of 2.5 mM K_2PtCl_4 for six days in the hanging drop where it had grown. The crystal turned brown when derivatized. For the MAD data collection, the derivative crystal was equilibrated in the cryoprotectant solution (reservoir solution (23 % PEG 3 K, 100 mM sodium citrate, pH 6.23) plus 20 % glycerol for a few minutes). The crystal was then flash-frozen in the nitrogen gas stream at 100 K (Oxford Cryosystems, Oxford, UK) for data collection. Upon flash-freezing, the lengths of a and c axes were decreased by about 0.5 and 1.5 Å, respectively (Table 3).

MAD data collection and structure determination

MAD data were collected to 2.2 Å from a Pt-derivative crystal at 100 K on an ADSC Quantum 4R detector at the beamline X8-C of the National Synchrotron Light Source, Brookhaven National Laboratory. The data were collected at three wavelengths: 1.07127 Å (''peak'', maximum of f''), 1.07159 Å (''edge'', maximum of $|f'|$), and 1.04192 Å (''remote''). The inverse beam geometry was used for measuring Bijvoet pairs simultaneously. All data were reduced and scaled using DENZO and SCALEPACK (Otwinowski & Minor, 1997). This set of three-wavelength MAD data were collected within six hours. A single pt site was located by the program SOLVE (Terwilliger & Berendzen, 1999; http://www.solve.lanl.gov). For scaling and subsequent steps, the remote wavelength data were chosen as the ''native'' data set (Ramakrishnan & Biou, 1997). Heavy-atom parameters were further refined with SHARP (de La Fortelle & Bricogne, 1997) and initial MAD phases were improved by solvent-flattening and histogram matching with DM (CCP4, 1994). A summary of X-ray data collection and phasing statistics is given in Table 3.

Table 3. Data collection and phasing statistics

			K$_2$PtCl$_4$ derivative[a]		
	Native1	Native2	Remote	Edge	Peak
X-ray source	SNU[b]	PF[c]	NSLS[d]	NSLS	NSLS
Resolution (Å)	2.5	1.9	2.2	2.2	2.2
Wavelength (Å)	1.54178	1.000	1.04192	1.07159	1.07127
Temperature (K)	293	283	100	100	100
Unit cell (*a,c*; Å)	62.58, 94.85	62.48, 94.63		61.96, 93.03	
Total reflections	28,815	133,929	151,296	145,107	147,156
Unique reflections	6,464	13,611	9,585	9,500	9,517
Completeness (%)[e]	98.1 (97.4)	88.0 (74.1)	98.9 (93.4)	97.9 (86.1)	98.0 (87.2)
R_{merge} (%)[f]	5.9	5.8	4.9	5.6	5.5
Cullis-R_{iso} (acen/cen)				0.49/0.48	0.57/0.56
Cullis-R_{ano}			0.66	0.62	0.65
Phasing power$_{iso}$ (acen/cen)				3.10/1.84	2.31/1.62
Phasing power$_{ano}$			2.79	3.01	2.93
FOM (acen/cen)	0.61/0.46 for 28-2.2 Å data				
FOM (after DM)	0.80 for 30-2.2 Å data				

[a] The number of Pt sites is 1 (Met26).
[b] SNU, Seoul National University, Seoul, Korea.
[c] PF, Photon Factory, Tsukuba, Japan.
[d] NSLS, National Synchrotron Light Source, Brookhaven National Laboratory, U.S.A.
[e] Values in parentheses are for reflections in the highest resolution bin of 0.1 Å thickness.
[f] $R_{merge} = \Sigma_h \Sigma_i |I(h,i) - \langle I(h) \rangle| / \Sigma_h \Sigma_i I(h,i)$, where $I(h,i)$ is the intensity of the *i*th measurement of *h* and $\langle I(h) \rangle$ is the corresponding average value for all *i* measurements.

Model building and refinement

The MAD-phased electron density map at 2.5 Å resolution was of sufficient quality to allow building of a nearly complete model using O (Jones *et al.*, 1991). The identification of ten disulfide bridges was helpful in connecting secondary structural elements. The Pt atom bound to the sulfur atom of Met26 was also clearly visible. Phases computed from the partially refined models were combined with the experimental MAD phases in each cycle of refinement using SIGMAA (CCP4, 1994). The model was initially refined against the remote wavelength data with X-PLOR (Brünger, 1992). Atomic positions were refined by the conventional conjugate gradient minimization, with higher resolution data to 2.5 Å being added in steps. After a round of simulated annealing refinement employing the standard slow-cooling protocol, the model was subject to refinement against native2 data (Table 3), extending the high-resolution limit to 1.9 Å in steps. Individual isotropic *B*-factors, initially set to 20 Å^2, were refined in the last stages of the refinement with restraint. Solvent molecules were placed by searching the model-phased ($F_o - F_c$) maps and a bulk solvent correction was applied. A summary of the refinement statistic is given in Table 1.

Structure analysis

The stereochemistry of the model was assessed with PROCHECK (Laskowski *et al.*, 1993). Models were displayed with O (Jones *et al.*, 1991) and CHAIN (Sack, 1988). Model superpositions were done by LSQKAB in CCP4 (CCP4, 1994). The putative complex model was built by the superposition of reactive-site loops based on the trypsin-inhibitor complex structures (PDB codes 1TAB and 1SMF) and was further optimized by the program MULTIDOCK (Jackson *et al.*, 1998). To analyze the protein-protein interface of the BPT-BBBI complex model we used the protein-protein interaction server (Jones & Thornton, 1996; http://www.biochem.ucl.ac.uk/bsm/PP/server). The DALI server was used for searching structural homology (Holm & Sander, 1993; http://www2.ebi.ac.uk/dali).

Protein Data Bank accession number

The atomic coordinates have been deposited with the Protein Data Bank for immediate release, accession code 1C2A.

References

Barton, G. J. (1993). ALSCRIPT: a tool to format multiple sequence alignments. *Protein Eng.* **6**, 37-40.

Birk, Y. (1987). Proteinase inhibitors. In *Hydrolytic Enzymes* (Neuroberger, A. & Brocklehurst, K., eds), pp. 257-300, Elsevier Science Publishers B.V., Amsterdam.

Birk, Y. (1993). Protease inhibitors of plant origin and role of protease inhibitors in human nutrition. In *Protease Inhibitors as Cancer Chemopreventive Agents* (Troll, W. & Kennedy, A. R., eds), pp. 97-106, Plenum Press, New York.

Birk, Y., Gertler, A. & Khalef, S. (1963). A pure trypsin inhibitor from soya beans. *Biochem. J.* **87**, 281-284.

Bode, W. & Huber, R. (1992). Natural protein proteinase inhibitors and their interactions with proteinases. *Eur. J. Biochem.* **204**, 433-451.

Boisen, S. & Djurtoft, R. (1982). Protease inhibitor from barley embryo inhibiting trypsin and trypsin-like microbial proteases. Purification and characterization of two isoforms. *J. Sci. Food Agric.* **33**, 431-440.

Bowman, D. E. (1946). Differentiation of soy bean anti-tryptic factors. *Proc. Soc. Expt. Biol. Med.* **63**, 547-550.

Brünger, A. T. (1992). *X-PLOR Manual, Version 3.1: A System for Crystallography and NMR*, Yale University Press, New Haven.

CCP4 (1994). Collaborative Computational Project Number 4. The CCP4 suite: programs for protein crystallography. *Acta Crystallog. sect. D*, **50**, 760-763.

Chen, P., Rose, J., Love, R., Wei, C. H. & Wang, B.-C. (1992). Reactive sites of an anticarcinogenic Bowman-Birk proteinase inhibitor are similar to other trypsin inhibitors. *J. Biol. Chem.* **267**, 1990-1994.

de La Fortelle, E. & Bricogne, G. (1997). Maxium-likelihood heavy-atom parameter refinement for multiple isomorphous replacement and multiwavelength anomalous diffraction methods. *Methods Enzymol.* **276**, 472-494.

de la Sierra, I. L., Quillien, L., Flecker, P., Gueguen, J. & Brunie, S. (1999). Dimeric crystal structure of a Bowman-Birk protease inhibitor from pea seeds. *J. Mol. Biol.* **285**, 1195-1207.

Flecker, P. (1995). Template-directed protein folding into a metastable state of increased activity. *Eur. J. Biochem.* **232**, 528-535.

Gariani, T., McBride, J. D. & Leatherbarrow, R. J. (1999). The role of the P_2' position of Bowman-Birk proteinase inhibitor in the inhibition of trypsin: studies on P_2' variation in cyclic peptides encompassing the reactive-site loop. *Biochim. Biophys. Acta*, **1431**, 232-237.

Harms-Ringdahl, M., Forsberg, J., Fedorcsák, I. & Ehrenberg, L. (1979). Trypsin inhibitory activity of a polypeptide isolated from red kidney beans that also enhances lymphocyte stimulation. *Biochem. Biophys. Res. Commun.* **86**, 492-499.

Hendrickson, W. A. & Ogata, C. M. (1997). Phase determination from multiwavelength anomalous diffraction measurements. *Methods Enzymol.* **276**, 494-523.

Holm, L. & Sander, C. (1993). Protein structure comparison by alignment of distance matrices. *J. Mol. Biol.* **233**, 123-138.

Jackson, R. M., Gabb, H. A. & Sternberg, M. J. E. (1998). Rapid refinement of protein interfaces incorporating solvation: application to the docking problem. *J. Mol. Biol.* **276**, 265-285.

Jones, S. & Thornton, J. M. (1996). Principles of protein-protein interactions derived from structural studies. *Proc. Natl Acad. Sci. USA*, **93**, 13-20.

Jones, T. A., Zou, J. Y., Cowan, S. W. & Kjeldgaard, M. (1991). Improved methods for the building of protein models in electron-density maps and the location of error in these maps. *Acta Crystallog. sect. A*, **47**, 110-119.

Kennedy, A. R. (1993). Anticarcinogenic activity of protease inhibitors. In *Protease Inhibitors as Cancer Chemopreventive Agents* (Troll, W. & Kennedy, A. R., eds), pp. 9-64, Plenum Press, New York.

Kraulis, P. J. (1991). MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallog.* **24**, 946-950.

Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallog.* **26**, 283-291.

Li, Y., Huang, Q., Zhang, S., Liu, S., Chi, C. & Tang, Y. (1994). Studies on an artificial trypsin inhibitor peptide derived from the mung bean trypsin inhibitor: chemical synthesis, refolding, and crystallographic analysis of its complex with trypsin. *J. Biochem. (Tokyo)*, **116**, 18-25.

Lin, G., Bode, W., Huber, R., Chi, C. & Engh, R. A. (1993). The 0.25-nm X-ray structure of the Bowman-Birk-type inhibitor from mung bean in ternary complex with porcine trypsin. *Eur. J. Biochem.* **212**, 549-555.

Luckett, S., Santiago, Garcia R., Barker, J. J., Konarev, A. I. V., Shewry, P. R., Clarke, A. R. & Brady, R. L. (1999). High-resolution structure of a potent, cyclic proteinase inhibitor from sunflower seeds. *J. Mol. Biol.* **290**, 525-533.

McBride, J. D., Brauer, A. B. E., Nievo, M. & Leatherbarrow, R. J. (1998). The role of threonine in the P_2 position of Bowman-Birk proteinase inhibitors: studies on P_2 variation in cyclic peptides encompassing the reactive-site loop. *J. Mol. Biol.* **282**, 447-457.

Mikola, J. & Kirsi, M. (1972). Differences between endo-spermal and embryonal trypsin inhibitors in barley, wheat, and rye. *Acta Chem. Scand.* **26**, 787-795.

Nagasue, A., Fukamachi, H., Ikenaga, H. & Funatsu, G. (1988). The amino acid sequence of barley rootlet trypsin inhibitor. *Agric. Biol. Chem.* **52**, 1505-1514.

Nicholls, A. (1992). *GRASP: Graphical Representation and Analysis of Surface Properties*, Columbia University, New York.

Odani, S., Koide, T. & Ono, T. (1983). The complete amino acid sequences of barley trypsin inhibitor. *J. Biol. Chem.* **258**, 7998-8003.

Ogiso, T., Noda, T., Sako, Y., Kato, Y. & Aoyama, M. (1975). Studies on trypsin inhibitor in barley. *J. Biochem. (Tokyo)*, **78**, 9-17.

Onesti, S., Brick, P. & Blow, D. M. (1991). Crystal structure of a Kunitz-type trypsin inhibitor from *Erythrina caffra* seeds. *J. Mol. Biol.* **217**, 153-176.

Otwinowski, Z. & Minor, W. (1997). Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307-326.

Philipp, S., Kim, Y.-M., Dürr, I., Wenzl, G., Vogt, M. & Flecker, P. (1998). Mutational analysis of disulfide bonds in the trypsin-reactive subdomain of a Bowman-Birk-type inhibitor of trypsin and chymo-trypsin. Cooperative versus autonomous refolding of subdomains. *Eur. J. Biochem.* **251**, 854-862.

Prakash, B., Selvaraj, S., Murthy, M. R. N., Sreerama, Y. N., Rao, D. R. & Gowda, L. R. (1996). Analysis of the amino acid sequences of plant Bowman-Birk inhibitors. *J. Mol. Evol.* **42**, 560-569.

Ramakrishnan, V. & Biou, V. (1997). Treatment of multi-wavelength anomalous diffraction data as a special

case of multiple isomorphous replacement. *Methods Enzymol.* **276**, 538-557.

Read, R. J. & James, M. N. G. (1986). Introduction to protein inhibitors: X-ray crystallography. In *Proteinase Inhibitors* (Barrett, A. J. & Salvesen, G., eds), pp. 301-335, Elsevier, Amsterdam.

Sack, J. S. (1988). CHAIN - a crystallographic modeling program. *J. Mol. Graph.* **6**, 244-245.

Song, H. K. & Suh, S. W. (1998a). Kunitz-type soybean trypsin inhibitor revisited: Refined structure of its complex with porcine trypsin reveals an insight into the interaction between a homologous inhibitor from *Erythrina caffra* and tissue-type plasminogen activator. *J. Mol. Biol.* **275**, 347-363.

Song, H. K. & Suh, S. W. (1998b). Preliminary X-ray crystallographic analysis of Bowman-Birk trypsin inhibitor from barley seeds. *Acta Crystallog. sect. D,* **54**, 441-443.

Suzuki, A., Yamane, T., Ashida, T., Norioka, S., Haram, S. & Ikenazka, T. (1993). Crystallographic refinement of Bowman-Birk type protease inhibitor A-II from peanut (*Arachis hypogaea*) at 2.3 Å resolution. *J. Mol. Biol.* **234**, 722-734.

Tashiro, M., Asao, T., Hirata, C., Takahashi, K. & Kanamoru, M. (1990). The complete amino acid sequence of a major trypsin inhibitor from seeds of foxtail millet (*Setaria italica*). *J. Biochem. (Tokyo),* **108**, 669-672.

Terwilliger, T. C. & Berendzen, J. (1999). Automated structure solution for MIR and MAD. *Acta Crystallog. sect. D,* **55**, 849-861.

Tsunogae, Y., Tanaka, I., Yamane, T., Kikkawa, J., Ashida, T., Ishikawa, C., Watanabe, K., Nakamura, S. & Takahashi, K. (1986). Structure of the trypsin-binding domain of Bowman-Birk type protease inhibitor and its interaction with trypsin. *J. Biochem. (Tokyo),* **100**, 1637-1646.

Voss, R.-H., Ermler, U., Essen, L.-O., Wenzl, G., Kim, Y.-M. & Flecker, P. (1996). Crystal structure of the bifunctional soybean Bowman-Birk inhibitor at 0.28-nm resolution. *Eur. J. Biochem.* **242**, 122-131.

Werner, M. H. & Wemmer, D. E. (1992). Three-dimensional structure of soybean trypsin/chymotrypsin Bowman-Birk inhibitor in solution. *Biochemistry,* **31**, 999-1010.