Supporting information

X-ray Structure of Native Scorpion Toxin BmBKTx1 by Racemic Protein Crystallography using Direct Methods

*Kalyaneswar Mandal, Brad L. Pentelute, Valentina Tereshko, Anthony A. Kossiakoff, and Stephen B. H. Kent**

Department of Chemistry, Department of Biochemistry and Molecular Biology, Institute for Biophysical Dynamics, The University of Chicago, Chicago, Illinois 60637

Experimental:

Materials

2-(1*H*-Benzotriazol-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate (HBTU) and protected amino acids (Peptide Institute, Osaka) were obtained from Peptides International. Side-chain protecting groups used were: Arg(Tos), Asn(Xan), Asp(OcHex), Cys(4-CH_3Bzl), Lys(2-ClZ), Ser(Bzl), Tyr(2-BrZ). Boc-L-Lys(2-ClZ)-OCH_2-phenylacetic acid was purchased from NeoMPS, Strasbourg, France. Boc-D-Lys(2-ClZ)-OCH_2-phenylacetic acid was prepared following the literature procedure.[1] Aminomethyl-(copolystyrene-divinylbenzene) resin was prepared from Biobeads S-X1 by published methods.[2] *N*,*N*-Diisopropylethylamine (DIEA) was obtained from Applied Biosystems. *N*,*N*-Dimethylformamide (DMF), dichloromethane, diethyl ether, HPLC-grade acetonitrile, and guanidine hydrochloride were purchased from Fisher. Trifluoroacetic acid (TFA) was obtained from Halocarbon Products, New Jersey. HF was purchased from Matheson. All other reagents were purchased from Sigma-Aldrich.

Chemical synthesis of BmBKTx1. The target amino acid sequence of BmBKTx1 is:

Ala[1]-Ala-Cys-Tyr-Ser-Ser-Asp-Cys-Arg-Val[10]-Lys-Cys-Val-Ala-Met-Gly-Phe-Ser-Ser-Gly[20]-Lys-Cys-Ile-Asn-Ser-Lys-Cys-Lys-Cys-Tyr[30]-Lys[31]

The D- and L- BmBKTx1 polypeptide chains were synthesized on Boc-Lys(2-ClZ)-OCH_2-Pam-resin[2] of the appropriate chirality using manual, in situ neutralization Boc chemistry protocols for stepwise SPPS,[3] at 0.1 and 0.2 mmol scale respectively. After removal of the N-terminal Boc group, the peptides were cleaved from the resin and simultaneously deprotected by treatment at 0 °C for 1 h with anhydrous HF containing {5% p-cresol + 5%thiocresol} as scavengers. After removal of HF by evaporation under reduced pressure, the crude peptide was precipitated and washed with diethyl ether, then dissolved in 50% aqueous acetonitrile containing 0.1% TFA and lyophilized. The crude lyophilized linear D- and L-BmBKTx1 peptides were obtained in 249 mg and 500 mg respectively. 120 mg of the crude linear D- and 150 mg of the crude linear L-BmBKTx1 were directly subjected to folding with simultaneous formation of disulfides, by dissolving the peptide at 3 mg/mL in 6 M guanidine hydrochloride followed by rapid 6-fold dilution with 100 mM tris-hydroxymethyl aminomethane aqueous buffer containing 9.2 mM L-cysteine and 1.2 mM L-cystine hydrochloride at pH 8.0, to give a final concentration of 0.5 mg/mL peptide and 1M guanidine hydrochloride in the folding buffer. The folding was essentially complete within 1 h at room temperature, as evidenced from the decrease in mass of ~6 Daltons for the product formed by LC-MS analysis. The final product was purified by reverse phase HPLC on a Silicycle spherical C-18, 10 × 250 mm column at 40 ºC using a gradient of 1%-41% acetonitrile over 80 minutes, at a flow rate of 10mL/min. Fractions containing the desired products were identified by LCMS, combined and lyophilized to furnish high purity D- BmBKTx1 and L-BmBKTx1 in 42.0 milligrams (26%, based on the amount (moles) of starting resin used) and 42.5 milligrams (21%, based on the amount (moles) of starting resin used) respectively. Analytical data are shown in Figure S1.

Crystallization.

Crystallization screenings were conducted at 19.5±0.5 °C using the commercially available Hampton index. Crystallization screens were done by the hanging drop vapor diffusion method. Aqueous protein

solution was centrifuged to remove minor amounts of particulate matter and then used directly for crystallizations. The drops were generated by mixing 1 μL of protein solution with 1 μL of reservoir solution and placed over 1mL of reservoir solution. Attempt to crystallize L-BmBKTx1 alone was performed in 100 mg/mL protein concentration in water using 96 Hampton index conditions and resulted no crystal formation even after several months. Crystallization of racemic BmBKTx1 was performed by mixing equal amounts (by weight) of lyophilized D- BmBKTx1 and L-BmBKTx1 in water at the following concentrations: 25 mg/mL (12.5 mg of D- and 12.5 mg L-), 50 mg/mL (25 mg of D- and 25 mg L-), 100 mg/mL (50 mg of D- and 50 mg L-), and 150 mg/mL (75 mg of D- and 75 mg L-). Crystals appeared after one day; within a week, ~10-20% of the conditions examined had produced microcrystals. One set of conditions (0.1 M citric acid pH = 3.5, 2 M ammonium sulfate) was further optimized by varying the concentration of ammonium sulfate in the precipitant solution to get crystals suitable for X-ray diffraction. X-ray diffraction data was collected from a crystal grown from 0.1 M citric acid pH = 3.5, 0.9 M ammonium sulfate.

Data collection. For low temperature data collection, selected crystals were briefly transferred to the cryoprotectant (reservoir solution plus 20% (v/v) glycerol) and flash-frozen in liquid nitrogen. The atomic resolution X-ray diffraction data (Supporting information Figure S2) were collected at 100K at the Argonne National Laboratory (Advanced Photon Source, beamline 23ID equipped with a MARCCD 300 detector, using 0.97911 Å synchrotron radiation. Crystal diffractions images were integrated, scaled, and merged with HKL2000.[4]

X-ray structure determination. The structure of racemic BmBKTx1 was solved by direct methods using SHELXS; after an 8-hour run, the positions for most of the atoms were revealed in the best solution with figure-of-merit of 0.23. The high-symmetry of the BmBKTx1 crystals that belong to tetragonal space group $I41/a$ rather then $P\bar{1}$ (found for other racemic proteins) significantly increased the success rate of the direct methods approach (Figure S3). Preliminary refinement was done in SHELXL.[5] Electron density map examination and model building were done using TURBO-FRODO.[6] The restrained positional and anisotropic B-factor refinement was performed in REFMAC5.[7] Hydrogen

atoms were included in riding positions. Molecular graphics were generated using Pymol[8] and Ribbons[9]. The main chain torsion angles for all residues are in the allowed regions and additional allowed regions of the Ramachandran plot. The data collection and refinement statistics are summarized in Table S1. The atomic coordinates have been deposited in the protein data bank. The PDB accession number for the crystal structure is 3E8Y.

The first residue (Ala1) was removed from the model due to disorder. Two additional atoms (chloride and sulfate) had been assigned and placed in the electron density map. It is interesting to note that a second centrosymmetric interface between L- and D-enantiomers present in the crystals may involve a chloride ion that presumably originated from the folding buffer (Figure S4 and Table S3). The observed crystal contacts between symmetry related L and L or D and D molecules in the crystals are shown in Table S4.

In order to explore the sensitivity of the use of direct methods to the resolution of the data used, we truncated the data to 1.2, 1.3, 1.4 and 1.5 Å. We then ran a series of SHELXS computational runs and examined the solutions from these calculations to determine the minimum resolution that was required to obtain a reasonable structural model. We found that for this molecule and data set, we could build a reasonable model up to 1.2 Å; above that, useful solutions were not obtained.

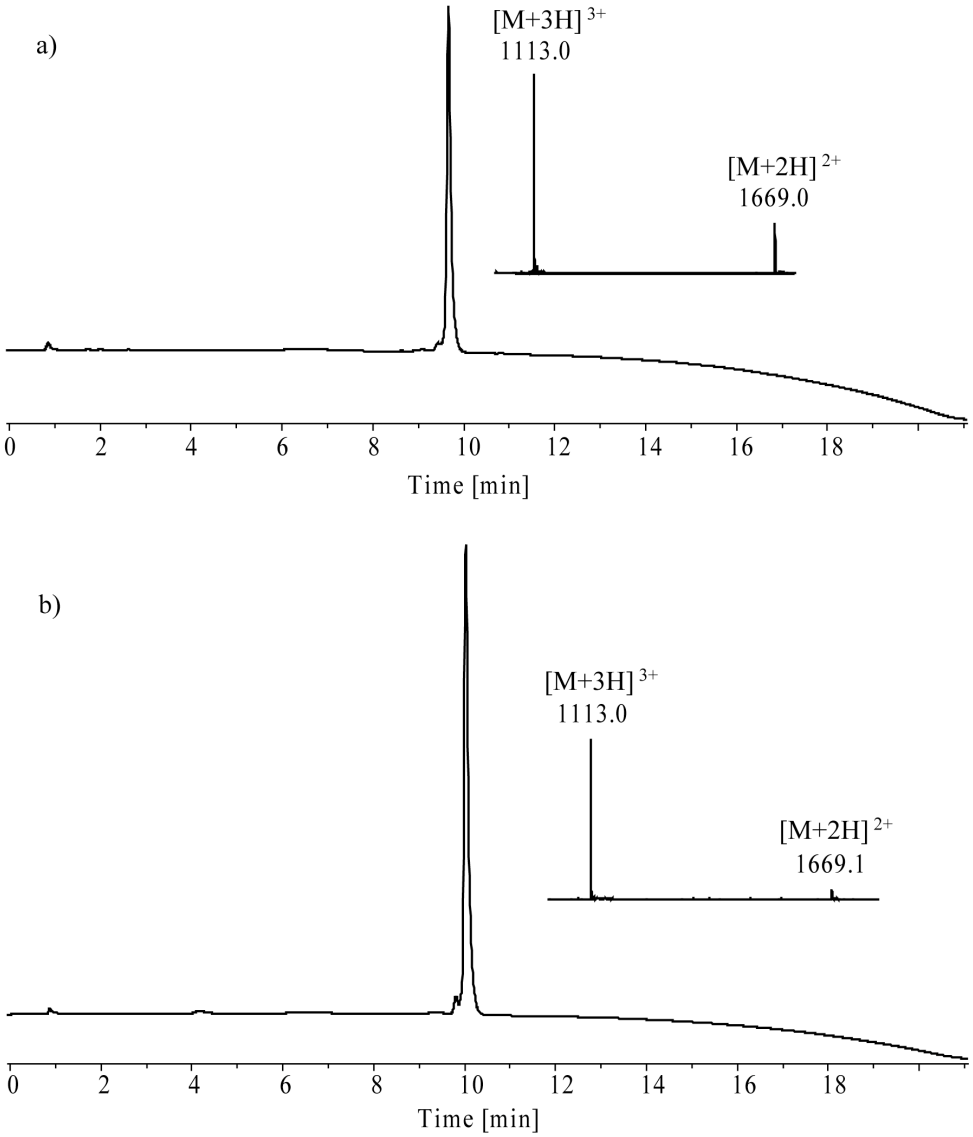

Figure S1. The LC-MS profiles of the folded and purified synthetic BmBKTx1 enantiomers. (a) D-BmBKTx1 (ob = 3336.0 Da, ca = 3336.0 (av isotopes)). (b) L- BmBKTx1 (ob = 3336.0 ± 0.1 Da, ca = 3336.0); the minor, earlier eluting peak corresponds to product containing an oxidized methionine (+16 Da). Analytical chromatography was performed using a linear gradient (1 - 61%) of buffer B in buffer A over 15 min (buffer A = 0.1% trifluoroacetic acid (TFA) in water; buffer B = 0.08% TFA in acetonitrile), with detection at 214nm and on-line ion trap electrospray MS.

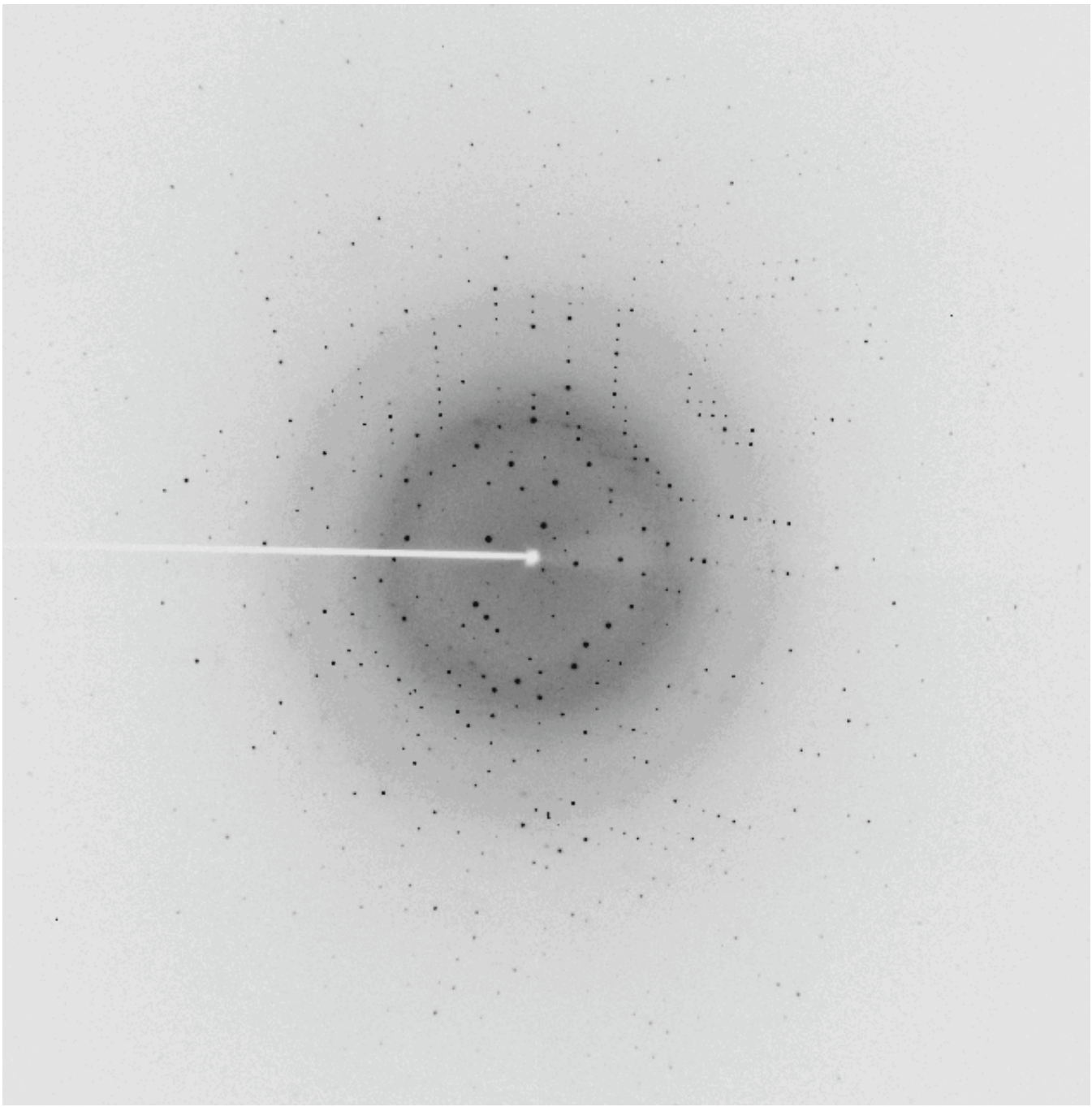

Figure S2. X-ray diffraction pattern with oscillation angle 40 degree and detector distance 122.15 mm. Data were collected to 1.1 Å resolution using 0.97911 Å synchrotron radiation.

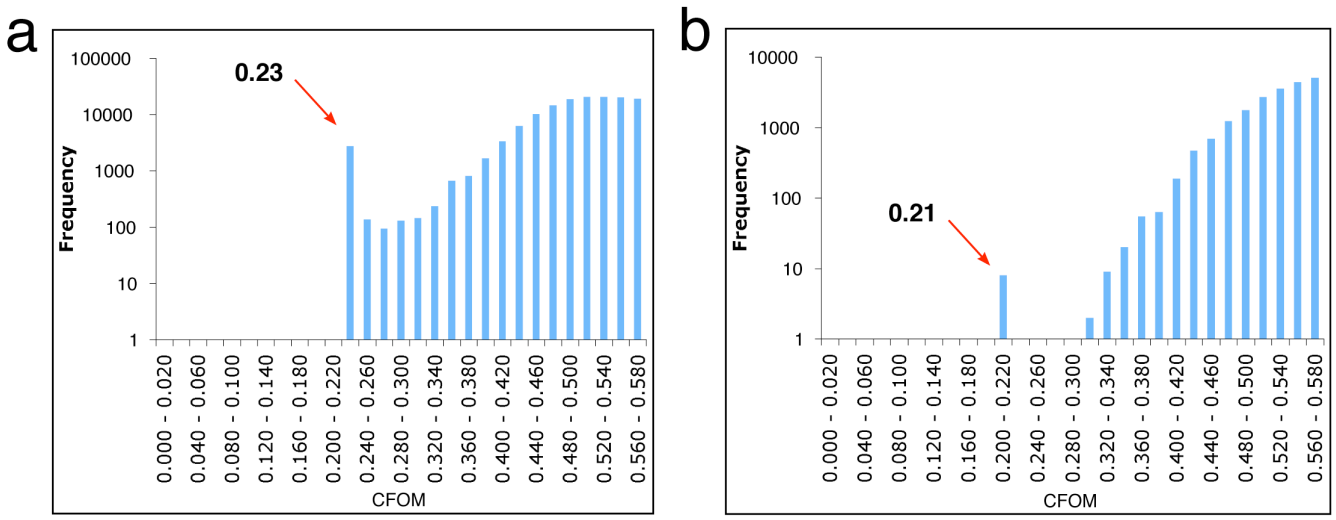

Figure S3. Histogram of combined figure-of-merit (CFOM) vs. frequency of solutions by direct methods obtained from the SHELXS run for BmBKTx1 (a) and plectasin (b) racemate crystals that belong to tetragonal *I*41/*a* and trigonal $P\bar{1}$ space groups, respectively. Note the unprecedented success rate in the high-symmetry *I*41/*a* space group. The best solution is indicated with red arrow in each panel. The BmBKTx1 structure is described in this work, and the plectasin structure will be published elsewhere. Note that the frequency is expressed in *logarithmic* form.

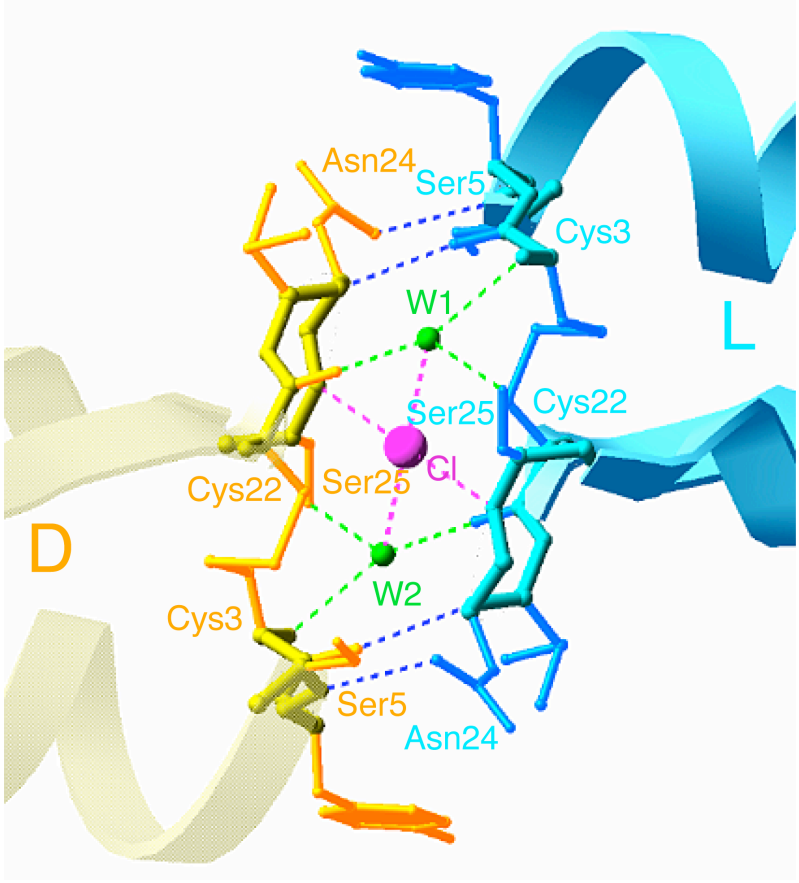

Figure S4. The close-up view of the chloride ion mediated packing interface between L- and D-enantiomers related by the second center of inversion symmetry. Chloride ion is shown as a magenta sphere, water molecules are shown as small green spheres and hydrogen bonds are shown as dashed lines. The hydrogen-bonded distances are listed in Table S3. The area buried at this interface between the D- and L-enantiomers is ~605 Å^2.

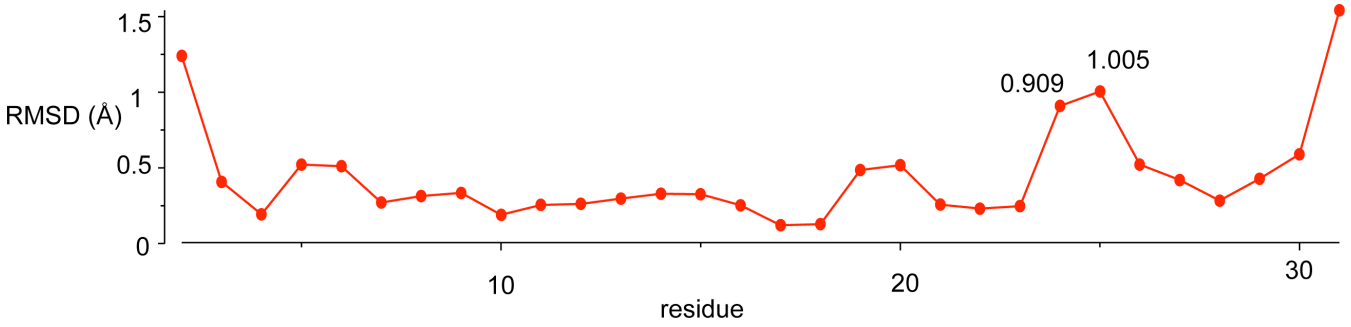

Figure S5. RMS deviations between C-α atoms in the X-ray structure of native BmBKTx1 reported here and the X-ray structure of the methylated variant (PDB 1R1G).

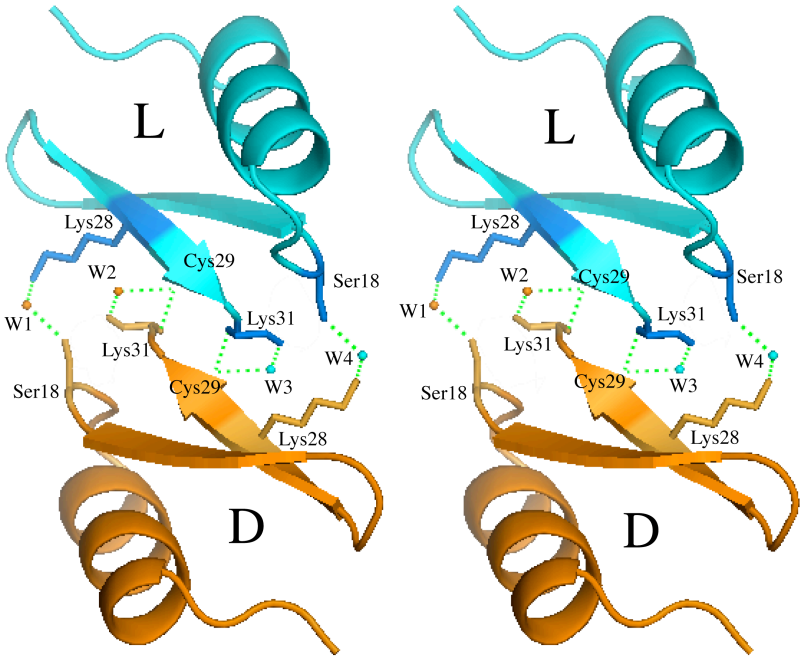

Figure S6 A stereopair showing the geometrical features of the intermolecular anti-parallel β-sheet interaction between the L- and D- enantiomers in the racemic crystal.

Table S1. The X-ray data collection and refinement statistics for the crystal structure of BmBKTx1.

	Data collection statistics
Space group	*I*41/*a*
Cell dimensions	
a, *b*, *c* (Å)	$a = 56.63$, $b = 56.63$, $c = 31.76$
α, β, γ (º)	$\alpha = 90$, $\beta = 90$, $\gamma = 90$
Mol/asymmetric unit	1
Mol/unit cell	16 (eight D- and eight L-molecule)
Wavelength	0.97911
Resolution (Å)	20 - 1.1 (1.129 - 1.10)
R_{merge}	0.045 (0.417)
$I/\sigma I$	41.6 (2.35)
Completeness (%)	99.78 (97.33)
Redundancy	5.9
	Refinement statistics
Resolution (Å)	20 - 1.1 (1.129 - 1.10)
No. reflections	17925
R_{work}/ R_{free}	0.200 (0.270)/ 0.217(0.262)
No. atoms	
Protein	268
Water	21
Avarage *B*-factor (Å^2)	10.1
R.m.s deviations	
Bond lengths (Å)	0.015
Bond angles (º)	1.92
	Ramachandran plot statistics
Most favored (%)	92.3 (L)
Additionally allowed (%)	3.8 (L)
Generously allowed (%)	3.8 (L)

*Highest resolution shell is shown in parenthesis.

**Based on maximum likelihood.

Table S2 Hydrogen bond lengths *r* (Å) at the β-sheet interface between L- and D- enantiomers shown in Figure 2b.

Protein config.	Residue number	Residue detail	*r* (Å)			Residue detail	Residue number	Protein config.
L	31	LYS-N	2.90			CYS-O	29	D
D	31	LYS-N	2.90			CYS-O	29	L
Mediated by water molecules								
Protein config.	Residue number	Residue detail	*r* (Å)	Water	*r* (Å)	Residue detail	Residue number	Protein config.
L	28	LYS-NZ	2.61	w1	2.72	SER-OG	18	D
L	29	CYS-O	2.89	w2	2.74	LYS-O	31	D
L	29	CYS-O	2.89	w3	2.74	LYS-O	31	D
D	28	LYS-NZ	2.61	w4	2.72	SER-OG	18	L

Table S3 Hydrogen bond lengths *r* (Å) at the chloride mediated interface between L- and D-enantiomers shown in Figure S4.

Protein config.	Residue number	Residue detail	*r* (Å)			Residue detail	Residue number	Protein config.
L	3	CYS-O	2.87			ASN-N	24	D
L	5	SER-N	2.85			ASN-OD1	24	D
D	3	CYS-O	2.87			ASN-N	24	L
D	5	SER-N	2.85			ASN-OD1	24	L
Mediated by chloride and water molecule								
Protein config.	Residue number	Residue detail	*r* (Å)	Water or Cl	*r* (Å)	Residue detail	Residue number	Protein config.
L	25	SER-N	3.25	Cl	3.26	SER-N	25	D
		w1	3.36	Cl	2.96	w2		
D	22	CYS-O	2.76	w1	3.00	SER-OG	25	L
				w1	2.90	CYS-N	3	L
L	22	CYS-O	2.91	w2	3.30	SER-OG	25	D
				w2	2.90	CYS-N	3	D

Table S4 Hydrogen bond lengths *r* (Å) between two symmetry-related L- BmBKTx1 molecules.

Protein	Residue number	Residue detail	*r* (Å)			Residue detail	Residue number	Protein config.
L1	6	SER-OG	2.16			LYS-NZ	21	L2
			3.12			TYR-OH	30	L2
L1	7	ASP-OD1	2.63			TYR-OH	30	L2
Mediated by water molecules								
Protein config.	Residue number	Residue detail	*r* (Å)	Water	*r* (Å)	Residue detail	Residue number	Protein config.
L1	7	ASP-OD2	2.48	w1	2.81	LYS-NZ	21	L2

H-bonding interactions between the equivalent two symmetry-related D-molecules are identical to those given in Table S4.

(1) Tam, J. P.; Kent, S. B. H.; Wong, T. W.; Merrifield, R. B., *Synthesis* **1979**, 955-957.
(2) Mitchell, A. R.; Kent, S. B. H.; Engelhard, M.; Merrifield, R. B. *J. Org. Chem.* **1978**, *43*, 2845-2852.
(3) Schnolzer, M.; Alewood, P.; Jones, A.; Alewood, D.; Kent, S. B. H., *Int. J. Pept. Res. Therapeut.* **2007,** *13*, 31-44.
(4) Otwinowski, Z.; Minor, W., *Macromolecular Crystallography, Pt A* **1997,** 276, 307-326.
(5) Sheldrick, G. M.; Dauter, Z.; Wilson, K. S.; Hope, H.; Sieker, L. C., *Acta Crystallogr., Sect D: Biol. Crystallogr.* **1993,** *49*, 18-23.
(6) Cambillau, C. R., A *Turbo Frodo*, OpenGL.1; University Aix-Marseille II: Marseille, 1997.
(7) Murshudov, G. N.; Vagin, A. A.; Dodson, E. J., *Acta Crystallogr., Sect D: Biol. Crystallogr.* **1997,** *53*, 240-255.
(8) DeLano, W. L. *The PymolMolecular Graphics System*, DeLano Scientific: San Carlos, CA, USA, 2002.
(9) Carson, M., Ribbons 2.0. *J. Appl. Crystallogr.* **1994,** *24*, 958-961.