Supporting Information

Total chemical synthesis and X-ray structure of kaliotoxin by racemic crystallography

Brad L. Pentelute[1], Kalyaneswar Mandal[1], Zachary P. Gates[1], Michael R. Sawaya[2,3], Todd O. Yeates[2,3], Stephen B. H. Kent[1]*

[a]*Department of Chemistry, Department of Biochemistry and Molecular Biology, Institute for Biophysical Dynamics, The University of Chicago, Chicago, Illinois 60637;* [b]*Molecular Biology Institute, Department of Chemistry and Biochemistry, University of California, Los Angeles;* [c]*University of California Los Angeles-United States Department of Energy Institute for Genomics and Proteomics*
Email: skent@uchicago.edu

Contents

1. **Experimental procedures**
2. **X-ray data collection and structure determination**
3. **Figure S1: [Gly1-Gln13]α-thioester peptide**
4. **Figure S2: CD spectra of kaltiotoxin enantiomers.**
5. **Table S1: X-ray data collection and refinement statistics for racemic kaliotoxin**

EXPERIMENTIAL PROCEDURES

Materials

2-(1H-Benzotriazol-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate (HBTU) was purchased from Peptides International, Louisville, KY. Nα-Boc-L-amino acids and Nα-Boc-D-amino acids (manufactured at the Peptide Institute, Osaka, Japan) were obtained from Peptides International. N,N-Diisopropylethylamine (DIEA) was purchased from Applied Biosystems. N,N-Dimethylformamide (DMF), dichloromethane (DCM), diethyl ether, HPLC-grade acetonitrile, and guanidine hydrochloride were purchased from Fisher or VWR. Trifluoroacetic acid (TFA) was from Halocarbon Products, New Jersey. HF was purchased from Matheson Tri-gas. Aminomethyl-resin was prepared from Bio-Beads SX-1 (Bio-Rad Laboratories) by published procedures or purchased from Rapp Polymere, Tubingen. All other reagents were purchased from Sigma-Aldrich.

Peptide Segment Synthesis (peptide-$^{\alpha}$carboxylate or peptide-$^{\alpha}$thioester) D-peptides and L-peptides were prepared manually by 'in situ neutralization' Boc chemistry on a 0.4 mmol scale by stepwise solid phase peptide synthesis (SPPS), on Xaa-OCH_2-Pam-resins (free $^{\alpha}$carboxyl peptides) or on $HSCH_2CH_2CO$-Xaa-OCH_2-Pam-resin ($^{\alpha}$thioester peptides).[1, 2] Side-chain protection for amino acids was as follows: Arg(Tos), Asn(Xan), Asp(OcHex), Cys(4-CH3Bzl), His(Bom), Glu(OHex, Lys(2Cl-Z), Ser(Bzl), and Thr(Bzl). After completion of the chain assembly, peptides were deprotected and cleaved from the resin support by treatment with anhydrous HF containing p-cresol (90:10, vol/vol) for 1 h at 0 °C. After complete evaporation of the HF under reduced pressure, crude peptide products were precipitated and triturated with chilled diethyl ether, and the peptide products were dissolved in 50% aqueous

acetonitrile containing 0.1% TFA and lyophilized. Crude peptides were purified by preparative reversed phase HPLC.

The L-peptide building blocks (and corresponding masses) prepared and used in the synthesis were as follows: Gly^{1}-Val-Glu-Ile-Asn-Val-Lys-Cys-Ser-Gly-Ser-Pro-Gln^{13}-CO-S-CH_2-CH_2-CO-Ala-COOH (ob = 1476.1 ±0.5 Da, ca = 1476.78 Da (average isotopes)), Thz^{14}-Leu-Lys-Pro-Cys-Lys-Asp-Ala-Gly-Met-Arg-Phe-Gly-Lys^{27}-CO-S-CH_2-CH_2-CO-Ala-COOH (ob =1725.0 ±0.5 Da, ca = 1724.92 Da), and Cys^{28}-Met-Asn-Arg-Lys-Cys-His-Cys-Thr-Pro-Lys^{38}-COOH (ob = 1319.8 ±0.7 Da, ca = 1320.63 Da). Cys^{28}-Met-Asn-Arg-Lys-Cys-His-Cys-Thr-Pro-Lys^{38}-COOH was generated directly from the crude peptide synthesized as Thz^{28}-Met-Asn-Arg-Lys-Cys-His-Cys-Thr-Pro-Lys^{38}-COOH by treatment for 2 hours with buffer containing 0.2 M methoxyamine•HCl, pH = 4, 0.2 M phosphate, and 6 M Gu•HCl.

The D-peptide building blocks (and corresponding masses) prepared and used in the synthesis were as follows: Gly^{1}-Val-Glu-Ile-Asn-Val-Lys-Cys-Ser-Gly-Ser-Pro-Gln^{13}-CO-S-CH_2-CH_2-CO-Ala-COOH (ob = 1476.7 ±0.5 Da, ca = 1476.78 Da (average isotopes)), Thz^{14}-Leu-Lys-Pro-Cys-Lys-Asp-Ala-Gly-Met-Arg-Phe-Gly-Lys^{27}-CO-S-CH_2-CH_2-CO-Ala-COOH (ob =1725.1 ±0.5 Da, ca = 1724.92 Da), and Cys^{28}-Met-Asn-Arg-Lys-Cys-His-Cys-Thr-Pro-Lys^{38}-COOH (ob = 1320.9 ±0.7 Da, ca = 1320.63 Da). Cys^{28}-Met-Asn-Arg-Lys-Cys-His-Cys-Thr-Pro-Lys^{38}-COOH was generated directly from the crude peptide synthesized as Thz^{28}-Met-Asn-Arg-Lys-Cys-His-Cys-Thr-Pro-Lys^{38}-COOH by treatment for 2 hours with buffer containing 0.2 M methoxyamine•HCl, pH = 4, 0.2 M phosphate, and 6 M Gu•HCl.

Analytical HPLC Peptide compositions were confirmed by analytical reverse phase LC-MS using a gradient of acetonitrile versus {0.1% trifluoroacetic acid (TFA) in water}. For the work reported in this paper, analytical HPLC was carried out as follows: Vydac C4 3 micron, 2.1 X 150 mm column, using a linear gradient of 1-61 % (5-65 %) buffer B over 15 min at 40 °C with a flow rate of 0.5 mL/min

(buffer A= 0.1% TFA in H_2O; buffer B = 0.08% TFA in acetonitrile) or Hypersil C18 5 micron, 4.6 X 50 mm column using a linear gradient of 1-61 % buffer B over 15 min at 40 °C with a flow rate of 1.0 mL/min (buffer A= 0.1% TFA in H_2O; buffer B = 0.08% TFA in acetonitrile). The eluent was monitored at 214 nm, and by on-line electrospray mass spectrometry (MS).

Preparative HPLC Peptides were purified on 10 micron C4, C8, or C18 silica with columns of dimension 22 X 250 mm, 10 X 250 mm, or 10 X 100 mm. The silica used was TP Vydac or self-packed Varian Microsorb or Agilent Zorbax. Crude peptides were loaded onto the prep column in ~10% acetonitrile/90% {0.1%TFA in water}, and eluted at a flow rate of 10 mL per minute with a shallow gradient (e.g. 20%B-40%B over 60 minutes) of increasing concentrations of solvent B (0.1% TFA in acetonitrile) in solvent A (0.1% TFA in water). Fractions containing the pure target peptide were identified by analytical LC and MALDI MS, and were combined and lyophilized.

Native Chemical Ligation Ligation reactions were carried out on a scale of 10 micromoles (~15 milligrams) at a concentration of 10 mM of each peptide (total volume 1 mL), under previously published conditions[3]: 200 mM sodium phosphate buffer containing 6 M Gu•HCl, 20 mM TCEP•HCl, pH = 7, 100 mM mercaptophenylacetic acid (MPAA) as catalyst, purged and sealed under argon. After the first ligation reaction, 0.2 M methoxyamine•HCl was added to convert the Thz-peptide product to Cys-peptide and left overnight at pH 4. The pH was adjusted to 7 and the next peptide-thioester (10 micromoles) was added in equimolar amount and the ligation reaction allowed to proceed at room temperature for 7 hours, to give the desired [Gly1-Lys38]-COOH polypeptide (ob = 4156.1 $\pm$ 1.0 Da, ca = 4156.0 Da (av. isotopes)).

One-Pot Folding reaction After confirming the completion of the ligation reaction, the reaction mixture (~1mL) was diluted with 6 mL of pH = 7.8 buffer

containing 6 M Gu•HCl and 0.1 M TRIS•HCl. Then 35 mL of redox buffer was added containing 100 mM TRIS•base, 8 mM cysteine, 1 mM cystine•2HCl, at pH 7.8. The folding and disulfide bond formation was complete after overnight reaction, as indicated by a mass change of minus 6 Da (ob. 4150.1 ± 0.7 Da, ca. 4150.0 Da (average isotopes)). The product was then purified by reverse phase HPLC and lyophilized, and 3.3 micromole of synthetic kaliotoxin was isolated (13.7 milligrams; 33% yield based on starting peptide segments).

CD spectroscopy CD spectra were recorded using a Jasco 715 instrument at 23 ˚C by dissolving 0.02 mg (prepared from a stock solution) of D-kaliotoxin or L-kaliotoxin in 200 µL of 50 mM phosphate buffer pH = 6.9. The path length of the cell was 1 mm.

X-ray 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 X-ray diffraction data were collected at 100K at the Argonne National Laboratory (Advanced Photon Source, beamline 23ID equipped with a MARCCD 300 detector).

X-ray structure determination Crystal diffraction images were processed and scaled with HKL2000.[4] The diffraction data were indexed in space group P1. Matthews cell content analysis showed that there were two molecules in the asymmetric unit.[5] The structure of racemic kaliotoxin was solved in space group P1<bar> by direct methods using SHELXD.[6] The Patterson seeding algorithm was used with instructions to find 12 sulfur atoms participating in disulfide bonds. All data to 0.95 Å resolution were used and 1600 trial structures were generated. The 974th trial was the best with a correlation coefficient of 0.577. Coordinates for 412 out of non-hydrogen 769 protein atoms (including multiple conformations) were obtained directly by SHELXD. The kaliotoxin model was refined with REFMAC5.[7] After each refinement step, the model was visually inspected in COOT, using both 2Fo-Fc and Fo-Fc difference maps.[8] All hydrogen atoms

connected to carbon atoms and backbone nitrogen atoms were included at their geometrically calculated positions and refined using a riding model. All models were validated with the following structure validation tools: PROCHECK,35 ERRAT,36 and VERIFY3D.37 PROCHECK reported that 88% of the residues are in the most favoured region of the Ramachandran plot, 10% of the residues are in additionally allowed regions. Phe25 of one kaliotoxin molecule was found in the generously allowed region. There were no residues in the disallowed regions of the Ramachandran plot. ERRAT reported that 100% of the residues were below the 95% certainty of rejection limit. The coordinates of the final model and the merged structure factors will be deposited in the RCSB Protein Data Bank.

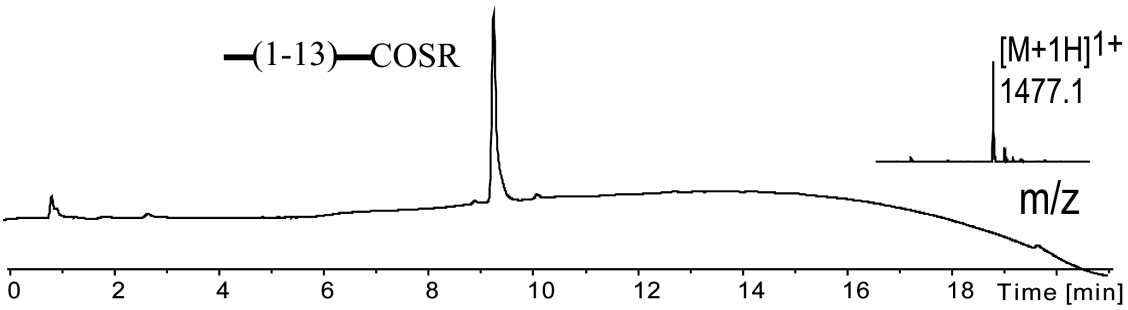

FIGURE S1 The purified [Gly1-Gln13]α-thioester peptide (mass: ob = 1476.1 $\pm$ 0.5 Da, ca = 1476.8 Da (average isotopes)) as used for the second ligation shown in Figure 2.

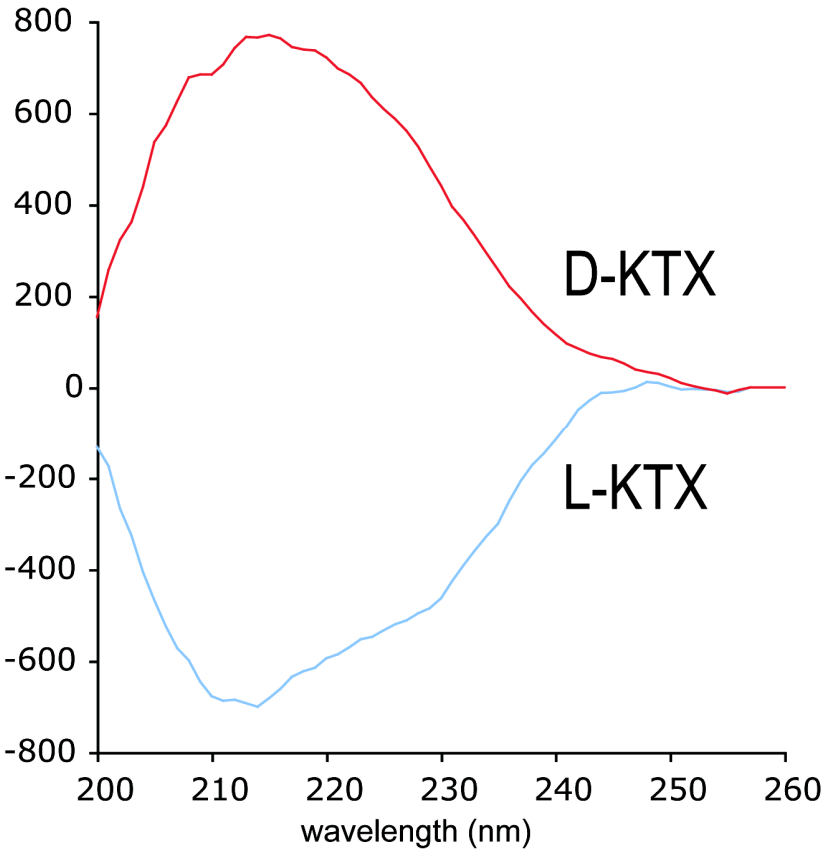

FIGURE S2 CD spectra of kaltiotoxin (KTX) enantiomers.

TABLE 1 X-ray data collection and refinement statistics for racemic kaliotoxin.

Data collection*	
Beamline	23ID (Advanced photon source)
Space group	P1<bar>
Wavelength (Å)	0.95373
Number of crystals	1
Cell dimensions	
a, *b*, *c* (Å)	25.2, 30.5, 41.1
$\alpha\beta\gamma$(□)	109.4, 97.4, 97.1
Mol/asymmetric unit	2
Mol/unit cell	4
Solvent content (%)	29.7
Resolution (Å)	50.0 – 0.95 (0.98 – 0.95)
R_{merge}**	0.057 (0.673)
$I/\sigma I$	19.9 (1.8)
Completeness (%)	88.2 (50.9)
Redundancy	4.3 (3.5)
Refinement	
Resolution (Å)	38.13-0.95 (0.98-0.95)
No. reflections	59134
R_{work}/ R_{free}	0.188/0.203 (0.494/0.553)
Completeness (working + test, %)	87.92 (45.08)
No. atoms	
Protein	769
Water	115
Others	40
B-factors (Å^2)	
Protein	10.56
R.m.s deviations	
Bond lengths (Å)	0.011
Bond angles (º)	1.6

*Highest resolution shell is shown in parenthesis.

** R_{merge} = SUM (ABS(I - <I>)) / SUM (I).

References

1. T. M. Hackeng, J. H. Griffin and P. E. Dawson, *Proceedings of the National Academy of Sciences of the United States of America*, 1999, **96**, 10068-10073.
2. M. Schnolzer, P. Alewood, A. Jones, D. Alewood and S. B. H. Kent, *International Journal of Peptide Research and Therapeutics*, 2007, **13**, 31-44.
3. E. C. Johnson and S. B. Kent, *J Am Chem Soc*, 2006, **128**, 6640-6646.
4. Z. Otwinowski and W. Minor, *Methods in Enzymology*, 1997, **276**, 307-326.
5. K. A. Kantardjieff and B. Rupp**,** *Protein Sci,* 2003, **12**, 1865–1871.
6. T. R. Schneider and G. M. Sheldrick, *Acta Crystallographica Section D-Biological Crystallography*, 2002, **58**, 1772-1779.
7. G. N. Murshudov, A. A. Vagin and E. J. Dodson, *Acta Crystallographica Section D-Biological Crystallography*, 1997, **53**, 240-255.
8. P. Emsley, B. Lohkamp, W. G. Scott and K. Cowtan, *Acta Crystallogr D Biol Crystallogr*, 2010, **66**, 486-501.