Supporting Information

Native Chemical Ligation at Asx-Cys, Glx-Cys: Chemical Synthesis and High Resolution X-ray Structure of ShK Toxin by Racemic Protein Crystallography

Bobo Dang, Tomoya Kubota, Kalyaneswar Mandal, Francisco Bezanilla, Stephen B. H. Kent

Department of Chemistry, Department of Biochemistry & Molecular Biology, Institute for Biophysical Dynamics, University of Chicago, Chicago, IL 60637

Reagents

2-(1*H*-Benzotriazol-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate (HBTU) and Nα-Boc protected L-amino acids (Peptide Institute, Osaka, Japan) were obtained from Peptides International (Louisville, Kentucky). Nα-Boc protected D-amino acids were obtained from Peptide Institute (Osaka, Japan) and from Peptides International (Louisville, Kentucky). Side-chain protecting groups used were Arg(Tos), Asp(OcHex), Asn(Xan), Cys(4-CH_3Bzl), Glu(OcHex), Lys(2Cl-Z), Tyr(2Br-Z). Boc-L-aspartic acid α-cyclohexyl ester dicyclohexylammonium salt, Boc-L-aspartic acid α-amide, Boc-L-glutamic acid α-cyclohexyl ester, Boc-L-glutamic acid α-amide were purchased from CHEM-IMPEX INT`L INC. Boc-(N'-xanthyl)-L-aspartic acid α-amide was chemically prepared.[1] Aminomethyl-resin (1.07 mmol/gram) was prepared from Biobeads S-X1 (BioRad, California).[2] Boc-L-Phe-OCH_2-phenylacetic acid was purchased from NeoMPS, Strasbourg, France. N,N-Diisopropylethylamine (DIEA) was obtained from Applied Biosystems. Diethyl ether, dichloromethane, N, N-

Dimethylformamide (DMF), 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 and were of the purest grade available.

Peptide synthesis

Peptide segments were synthesized using manual 'in situ neutralization' Boc chemistry protocols for stepwise SPPS.[3] All the peptides were synthesized on Boc-*Phe*-OCH_2-Pam-resin[4] at a 0.3 mmol scale. The peptide-thioesters were synthesized on trityl-SCH_2CH_2CO-*Phe*-OCH_2-Pam-resin[4] at a 0.3 mmol scale. After removal of the N-αBoc group, peptides were cleaved from the resin and simultaneously deprotected by treatment at 0 °C for 1 h with anhydrous HF containing 5%v/v p-cresol as scavenger. After removal of HF by evaporation under reduced pressure, each crude peptide was precipitated and washed with diethyl ether, then dissolved in 50% aqueous acetonitrile containing 0.1% TFA and lyophilized.

Native Chemical Ligation of LYRAX-$^{\alpha}$*COSR*` and CFYANF:

All ligation reactions were carried out by dissolving LYRAX-$^{\alpha}$*COSR*` (~0.35 mg, ~2mM) and CFYANF (~0.37 mg, ~2.4 mM) in ligation buffer (0.2 mL) containing 20 mM MPAA, 6 M Gu·HCl, 200 mM Na_2HPO_4, 10 mM TCEP·HCl. The pH was adjusted to 7.0 unless otherwise indicated. Ligation buffer was purged with helium for 15 min before use. All the reactions were monitored by LC-MS.

LYRAD-$^{\alpha}$*COSR*`` stability test:

LYRAD-$^{\alpha}$*COSR*` 2mM was kept at RT in ligation buffer 6 M Gu·HCl, 200 mM $Na_2HPO_4$, 10 mM TCEP·HCl, pH 6.8. About 30% hydrolyzed product was formed from the peptide-thioester LYRAD-$^{\alpha}$*COSR*` after 19 h (Figure S1). Formation of ~30% of a product with mass 18 Da lower than expected for simple hydrolysis of the thioester was also observed under these conditions; this presumably corresponds to formation of a carboxylic anhydride at the C-terminal of the peptide upon loss of thioester. Initial formation of the anydride from the thioester would explain the large amount of hydrolysis observed for the peptide with a C-terminal Asp-thioester.

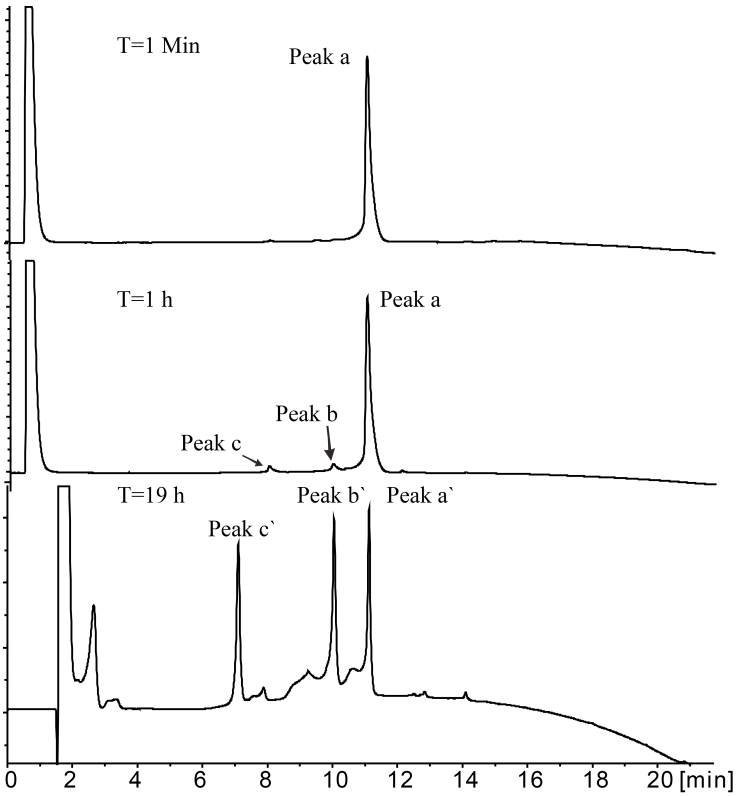

Figure S1. LYRAD-$^{\alpha}$*COSR*` stability test. Peak a and Peak a` are LYRAD-$^{\alpha}$*COSR*`; Peak b and Peak b` corresponds to formation of carboxylic anhydride at the C-terminus; Peak c and Peak c` are LYRAD-COOH (Hydrolyzed product). Note that different HPLC conditions were used for T= 19h. Products were identified by LC-MS.

Synthesized Peptides:

CFYANF	Observed Mass: 763.4±0.4Da	calc.: 763.8 Da
LYRAN-$^{\alpha}$*COSCH$_2$CH$_2$CO-Phe-COOH*	Observed Mass: 870.6±0.4 Da	calc.: 871.0 Da
LYRAD-$^{\alpha}$*COSCH$_2$CH$_2$CO-Phe-COOH*	Observed Mass: 871.9±0.4 Da	calc.: 872.0 Da
LYRAQ-$^{\alpha}$*COSCH$_2$CH$_2$CO-Phe-COOH*	Observed Mass: 884.6±0.4 Da	calc.: 885.0 Da
LYRAE-$^{\alpha}$*COSCH$_2$CH$_2$CO-Phe-COOH*	Observed Mass: 886.2±0.4 Da	calc.: 886.0 Da
LYRAN$^{\alpha}$CFYANF	Observed Mass: 1381.0±0.4 Da	calc.: 1381.5 Da
LYRAD$^{\alpha}$CFYANF	Observed Mass: 1383.0 ±0.4 Da	calc.: 1382.5 Da
LYRAQ$^{\alpha}$CFYANF	Observed Mass: 1395.0 ±0.4 Da	calc.: 1395.6 Da
LYRAE$^{\alpha}$CFYANF	Observed Mass: 1396.6±0.4 Da	calc.: 1396.6 Da
LYRAN$^{\beta}$CFYANF	Observed Mass: 1381.0±0.4 Da	calc.: 1381.5 Da
LYRAD$^{\beta}$CFYANF	Observed Mass: 1383.0±0.4 Da	calc.: 1382.5 Da
LYRAQ$^{\gamma}$CFYANF	Observed Mass: 1395.1 ±0.4 Da	calc.: 1395.6 Da
LYRAE$^{\gamma}$CFYANF	Observed Mass: 1396.6±0.4 Da	calc.: 1396.6 Da
(L-)ShK-[Cys17-Cys35]	Observed Mass: 2252.2±0.4 Da	calc.: 2252.7 Da
(L-) ShK-[Arg1-Gln16]-α-thioester	Observed Mass: 2061.0±0.4 Da	calc.: 2061.4 Da
(D-) ShK-[Cys17-Cys35]	Observed Mass:2252.4±0.4 Da	calc.: 2252.7 Da
(D-) ShK-[Arg1-Gln16]-α-thioester	Observed Mass: 2061.2±0.4 Da	calc.: 2061.4 Da

Calculated masses are based on average isotope compositions.

Identification of Ligation Products:

When the ligation reactions were complete, a sample of authentic α-, β- or γ-linked peptide was added to the reaction solution as indicated below:

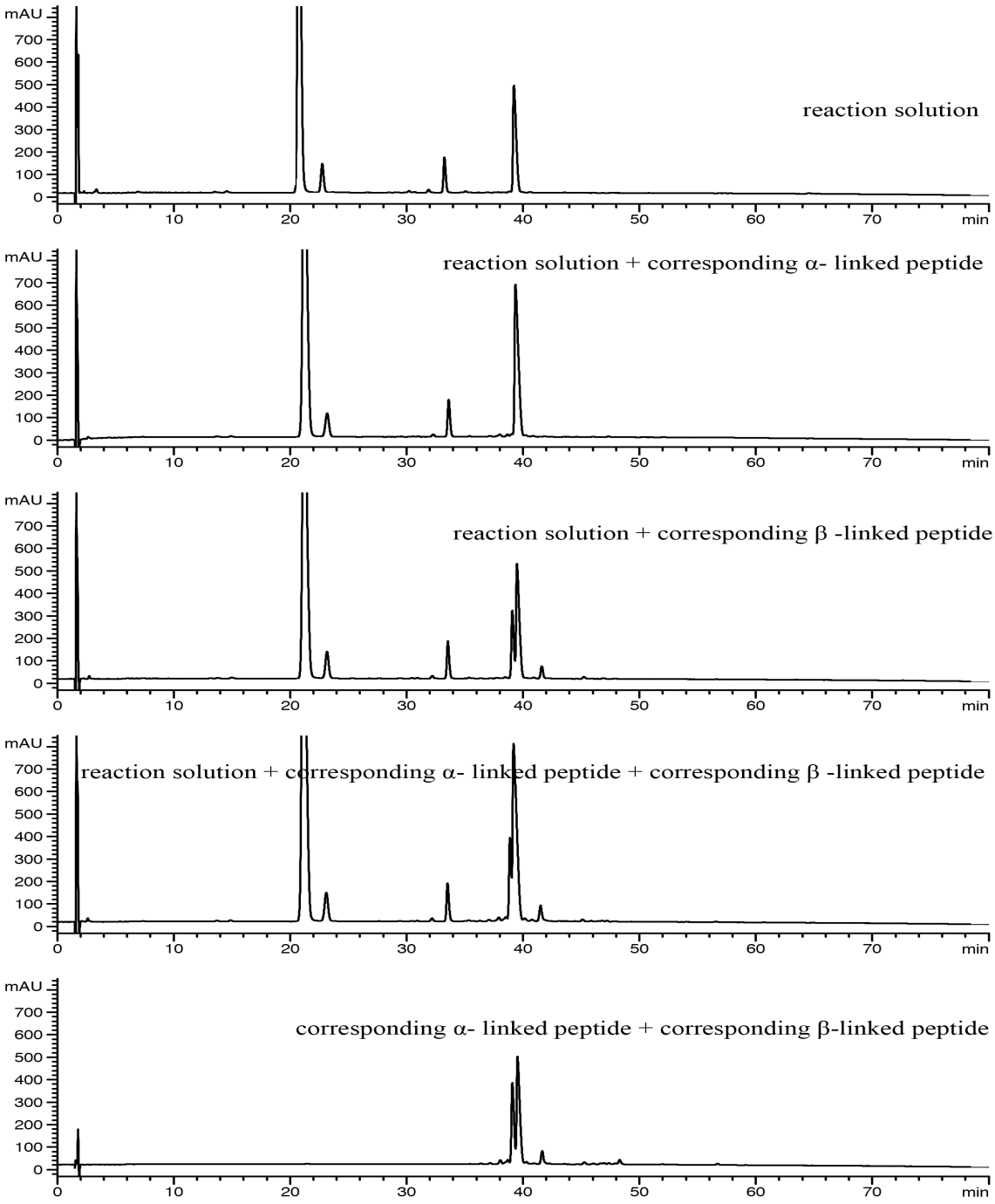

Figure S2. Products verification for ligation at –Asn-Cys- site

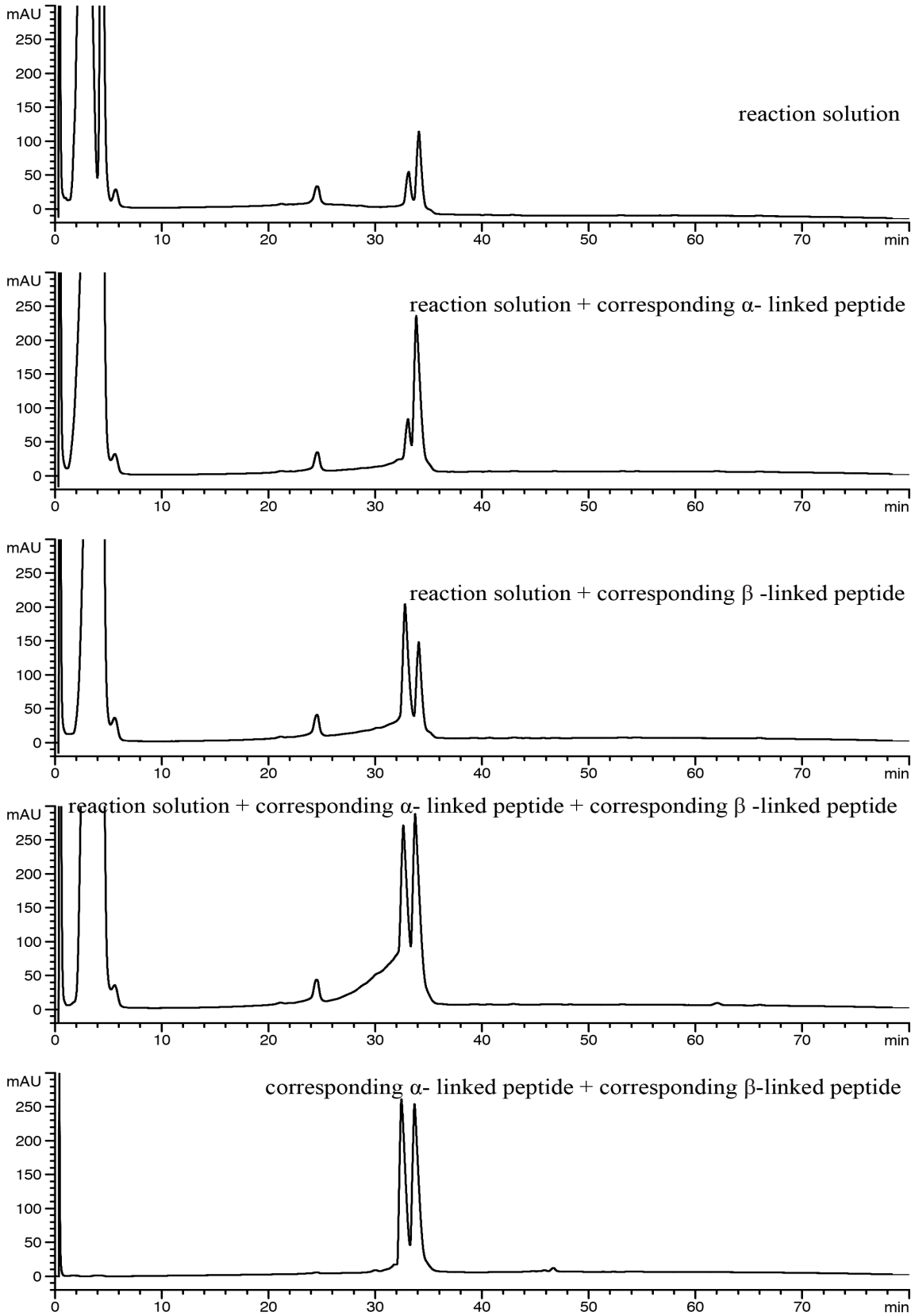

Figure S3. Products verification for ligation at -Asp-Cys- site

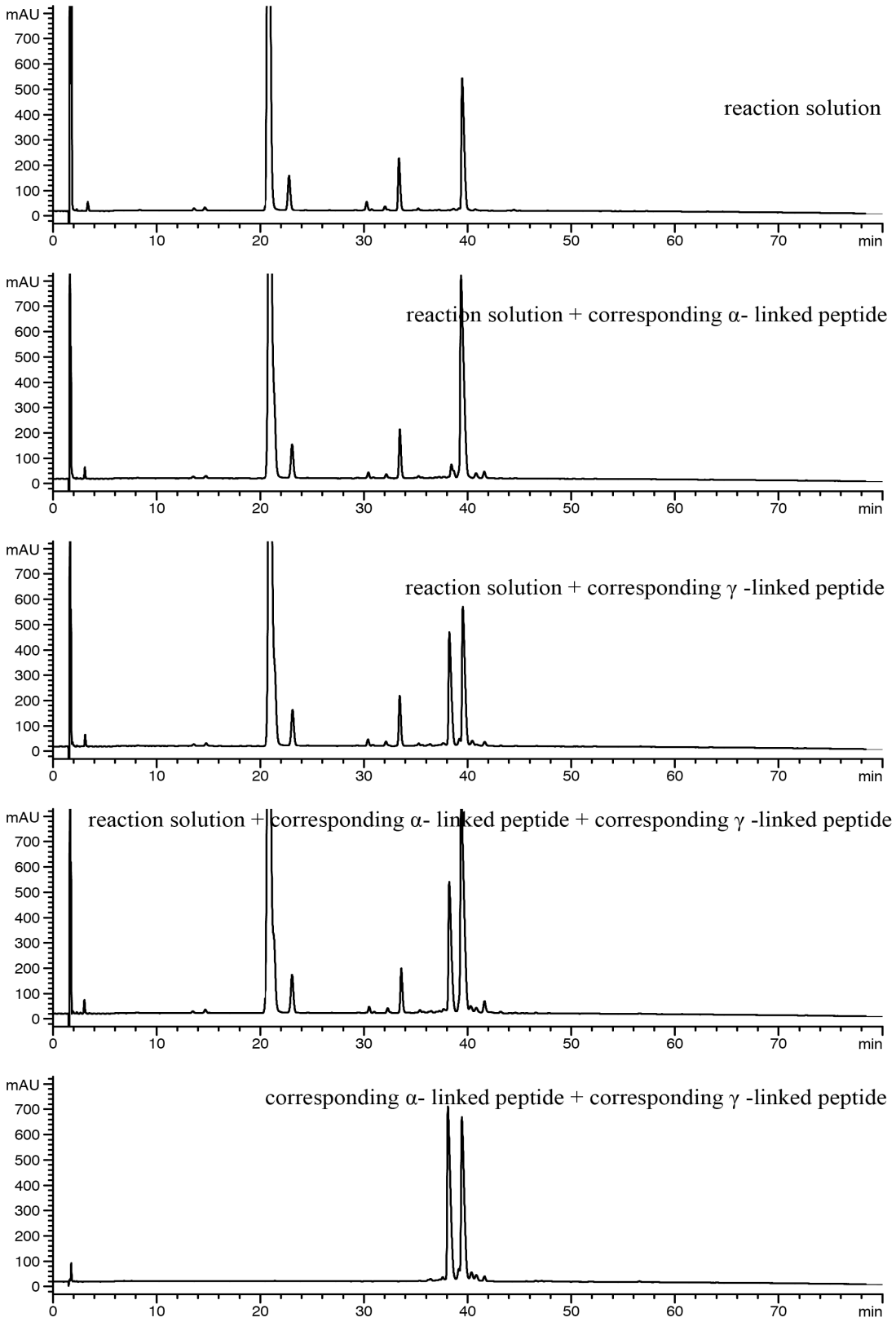

Figure S4. Products verification for ligation at -Gln-Cys- site:

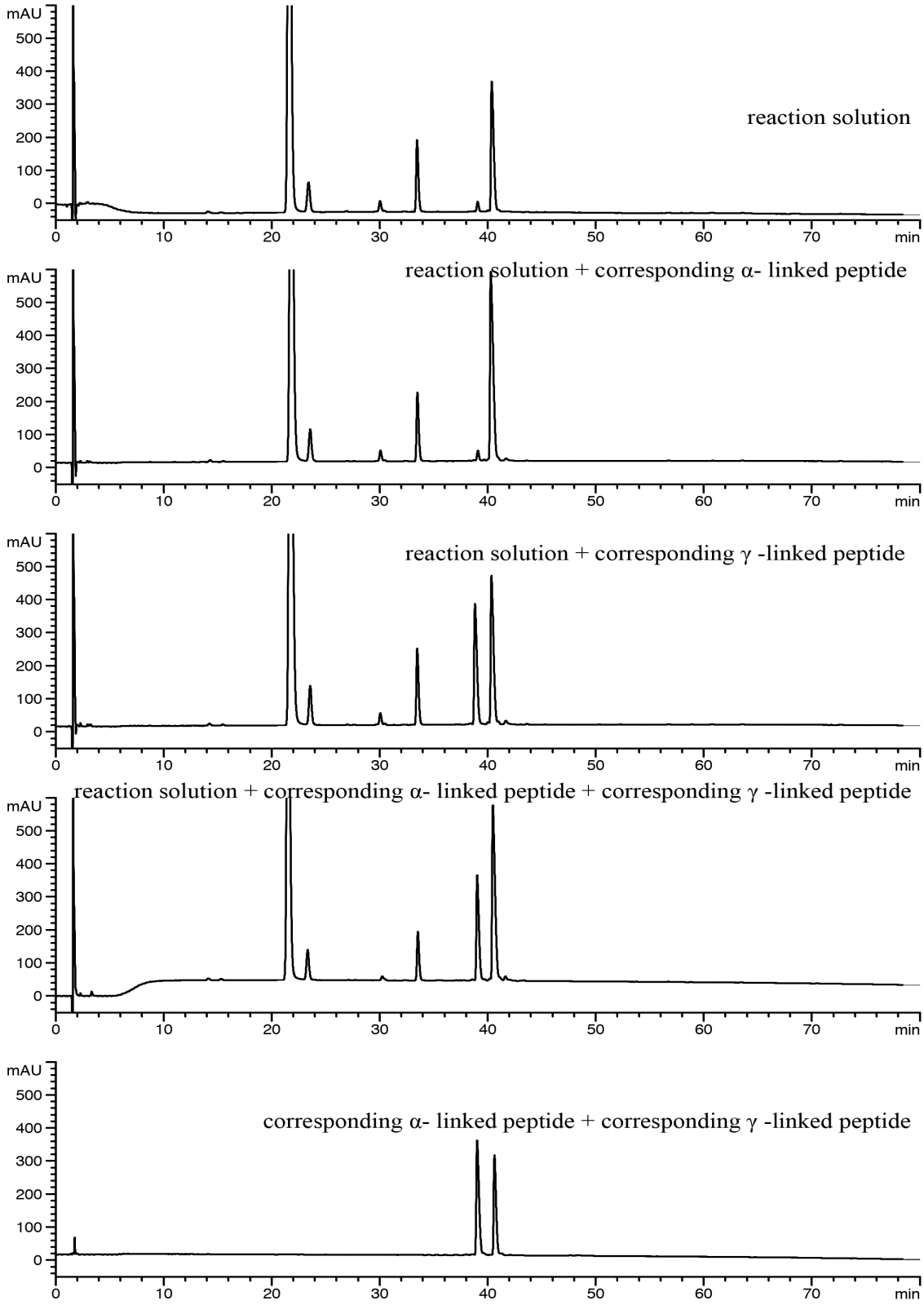

Figure S5. Products verification for ligation at –Glu-Cys- site:

X-ray data statistics for the racemic DL-ShK crystal structure

Wavelength (Å)	0.688
Resolution range (Å)	50.00 - 0.97 (0.99 - 0.97)
Space group	P 1 21/c 1
Cell dimensions	
a, b, c (Å)	30.912, 51.653, 21.065
α, β, γ (°)	90, 104.5, 90
Solvent content (%)	38.9
Mol/asymmetric	1
Mol/unit cell	4
Completeness (%)	94.60 (89.6)
Mean I/sigma(I)	13.37 (6.36)
R-merge	0.128 (0.339)
Average Redundancy	5.1 (4.9)
Unique reflections	35825
Refinement Statistics	
Resolution range (Å)	25.80 - 0.97 (0.99 - 0.97)
R-factor	0.153 (0.155)
R-free	0.165 (0.187)
Number of atoms	701
Number of water	84
Protein residues	35
RMS(bonds)	0.007
RMS(angles)	1.4
Ramachandran favored (%)	100
Ramachandran outliers (%)	0
Clashscore	0.00
Average B-factor	7.30

Highest resolution shell is shown in parentheses.
Coordinates and structure factors have been deposited in the Protein Data Bank with accession code 4LFS.

X-ray data statistics of the L-ShK crystal structure

Wavelength (Å)	0.688
Resolution range (Å)	50.00 - 1.06 (1.08 - 1.06)
Space group	C 2 2 2_1
Cell dimensions	
a, b, c (Å)	41.161, 49.491, 27.770
α, β, γ (°)	90, 90, 90
Solvent content (%)	30.2
Mol/asymmetric	1
Mol/unit cell	8
Unique reflections	13043
Average Redundancy	7.3 (6.3)
Completeness (%)	98.3 (98.8)
Mean I/sigma(I)	17.4 (2.9)
R-merge	0.093 (0.434)
Refinement statistics	
Resolution range (Å)	24.74 - 1.06 (1.14 - 1.06)
R-factor	0.133 (0.150)
R-free	0.157 (0.212)
Number of protein atoms	680
Number of water	62
Protein residues	35
RMS(bonds)	0.009
RMS(angles)	1.30
Ramachandran favored (%)	100
Ramachandran allowed (%)	0
Ramachandran outliers (%)	0
Clash score	3.42
Average B-factor	13.00

Highest resolution shell is shown in parentheses.
Coordinates and structure factors have been deposited in the Protein Data Bank with accession code 4LFQ.

References

(1) Atherton, E.; Caviezel, M.; Fox, H.; Harkiss, D.; Over, H.; Sheppard, R. C., *J. Chem. Soc. Perk. T. 1* **1983**, *1*, 65.

(2) Mitchell, A. R.; Kent, S. B. H.; Engelhard, M.; Merrifield, R. B., *J. Org. Chem.* **1978,** *43*, 14, 2845.

(3) Schnölzer, M.; Alewood, P.; Jones, A.; Alewood, D.; Kent, S. B. H. *Int. J. Pept. Protein Res.* **1992**, *40*, 180.

(4) Hackeng, T. M.; Griffin, J. H.; Dawson, P. E., *P. Natl. Acad. Sci. USA* **1999,** *96*, *18*, 10068.