Angewandte Chemie
Eine Zeitschrift der Gesellschaft Deutscher Chemiker

Supporting Information

69451 Weinheim, Germany

Convergent Chemical Synthesis and Crystal Structure of a 203 Amino Acid 'Covalent Dimer' HIV-1 Protease Enzyme Molecule.

Vladimir Yu. Torbeev, Stephen B. H. Kent

Ref.[3a]: G. G. Kochendoerfer, S.-Y. Chen, F. Mao, S. Cressman, S. Traviglia, H. Shao, C. L. Hunter, D. W. Low, L. N. Cagle, M. Carnevali, V. Guerigtiian, P. J. Keogh, H. Porter, S. M. Stratton, M. C. Wiedeke, J. Wilken, J. Tang, J. J. Levy, L. P. Miranda, M. M. Crnogorac, S. Kalbag, P. Botti, J. Schindler-Horvat, L. Savatski, J. W. Adamson, A. Kung, S. B. H. Kent, J. A. Bradburne, *Science* **2003**, *299*, 884-887.

Reagents. Boc-amino acids, 4-Me-benzylhydrylamine-resins and $-OCH_2$-Pam-resins (free $^{\alpha}$carboxyl peptides) were purchased from Peptides International, Louisville, Kentucky; Boc-*L*-thiazolidine-4-carboxylic acid from NovaBiochem, San Diego; diethyl ether from Fisher; HF from Matheson Tri-Gas; methoxylamine hydrochloride, *p*-cresol, sodium 2-mercaptosulphonate (MESNA), triisopropylsilane from Sigma-Aldrich; 4-mercaptophenylacetic acid was obtained from Sigma-Aldrich and purified by HPLC before use; tris(2-carboxyethyl)phosphine hydrochloride (TCEP) from Fluka; trifluoroacetic acid (TFA) from Halocarbon Products, New Jersey.

Peptide Synthesis. Peptides were prepared manually by "in situ neutralization" Boc-chemistry stepwise solid phase peptide synthesis (SPPS).[1] Side-chain protection for amino acids was as follows: Arg(Tos), Asn(Xan), Asp(OcHex), Cys(4-CH_3Bzl), Glu(OcHex), Lys(2Cl-Z), Ser(Bzl), Thr(Bzl), Tyr(Br-Z), Trp(CHO), His(Bom). The 1,3-thiazolidine-4-carboxyl (Thz) group was introduced to protect the N-terminal Cys of the (Cys-Gly_4-B1-B40)-($^{\alpha}$thioalkylester) peptide segment, and Boc-*L*-thiazolidine-4-carboxylic acid was used for peptide synthesis. After chain assembly was complete, peptides were deprotected and simultaneously cleaved from the resin support by treatment with anhydrous HF containing 10% (v/v) *p*-cresol for 1 h at 0 ºC. After evaporation of the HF under reduced pressure, crude products were precipitated and

triturated with chilled diethyl ether, and the peptide products were dissolved in 50% aqueous acetonitrile containing 0.1% TFA.

LC and LC-MS analysis. Analytical RP-HPLC was performed on an Agilent 1100 system with in-house packed C-4 and C-18 silica columns (2.1 × 50 mm, Microsorb 300 Å, 3 μm) at flow rate of 0.5 mL/min. Peptides were eluted from the column using a gradient of acetonitrile/0.08% TFA (solvent A) versus water/0.1% TFA (solvent B). Peptide masses were obtained using on-line electrospray MS detection with an Agilent 1100 LC/MSD ion trap. Masses of synthetic peptides were: (A1-A40)-$^{\alpha}COSCH_2CH_2Arg_4$.amide, obs. 5125.2 ± 0.3 Da, calc. 5125.0 Da; (Cys^{A41}-A99)-$^{\alpha}COSCH_2CH_2Arg_4$.amide, obs. 7066.3 ± 0.5 Da, calc. 7065.5 Da; Thz^{201}-Gly_4-(B1-B40)-$^{\alpha}COSCH_2CH_2Arg_4$.amide, obs. 5468.5 ± 0.4 Da, calc. 5468.4 Da; (Cys^{B41}-B99)-COOH, obs. 6354.1 ± 0.6 Da, calc. 6353.6 Da. Calculated masses were of average isotope compositions.

Preparative HPLC of crude peptides after SPPS was performed on Agilent 1100 prep system on a Agilent C-3 (250 × 22 mm, 300Å, 7μm) column or on in-house packed columns C4 (250 × 7 mm, Microsorb, 300Å, 5μm) and C8 (250 × 10 mm, Microsorb, 300Å, 10μm). Peptides from SPPS and from ligations were eluted from the column using an appropriate shallow gradient of acetonitrile/0.08% TFA versus water/0.1% TFA. Fractions containing the desired purified peptide product were identified by analytical LC and LC-MS, combined and lyophilized. Segments (Cys^{A41}-A99)-$^{\alpha}COSCH_2CH_2Arg_4$.amide and ($Cys^{B41}$-B99)-COOH were purified two times.

Conversion of (A1-A40)-($^{\alpha}$thioalkylester) to (A1-A40)-($^{\alpha}$thioarylester). 31.2 mg (6.1 μmol) of peptide (A1-A40)-$^{\alpha}COSCH_2CH_2Arg_4$.amide was dissolved in 5 mL buffer (6 M Gn·HCl, 0.1 M Na_2HPO_4) containing 109 mg (0.65 mmol) 4-mercaptophenylacetic acid and 17 mg (0.06 mmol) TCEP. The pH was adjusted to 5.8 and kept for 4 hours, then purified by RP-HPLC on C8 column. LC-MS: obs. 4563.3 Da, calc. 4563.3 Da). Isolated yield 8.2 mg or 1.8 μmol (29.5 %).

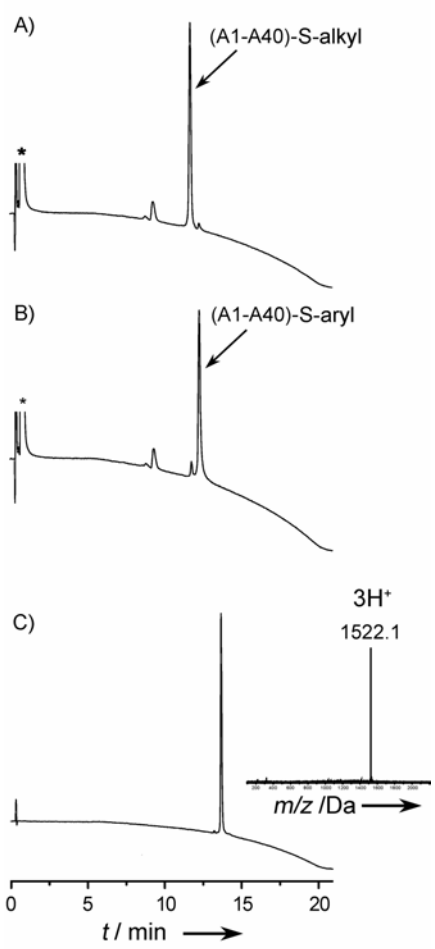


Figure S1. Analytical HPLC traces (λ = 214 nm) of transthioesterification reaction (C4 column). A) t < 1min. B) t = 4 h. C) After HPLC purification (C18 column). Insert is mass-spectrum of (A1-A40)-$^{\alpha}COSC_6H_4CH_2COOH$ integrated over elution time of target peak. Asterisk shows elution of 4-mercaptophenylacetic acid and other buffer components.

Native chemical ligation of Thz-Gly$_4$-(B1-B40)-$^{\alpha}$COSCH$_2$CH$_2$Arg$_4$ and (Cys^{B41}-B99). 7.6 mg (1.39 μmol) Thz^{201}-Gly$_4$-(A1-A40)-$^{\alpha}$COSCH$_2$CH$_2$Arg$_4$.amide and 8 mg (1.3 μmol) (Cys^{B41}-B99)-COOH were mixed in 1 mL of buffer containing 8 M Gn·HCl, 0.2 M Na_2HPO_4 and 22 mM TCEP. The 4-mercaptophenylacetic acid was added at 70 mM concentration. After 3 h, the reaction mixture was deluted to 3.5 mL with buffer and 50 mg (0.36 mmol) of 2-bromoacetamide was added (pH 6.7). After 15 min 62 mg (0.38 mmol) of MESNA were added to quench excess of 2-bromoacetamide. 58.5 mg (0.7 mmol) of MeONH$_2$·HCl (*c* 0.2 M) was added to deprotect (Thz)Cys. After 8 h of reaction, the product was purified by RP-HPLC. LC-MS: obs. 11137.5 ± 0.7 Da, calc. 11137.1 Da). Isolated yield 5.5 mg or 0.5 μmol (38%).

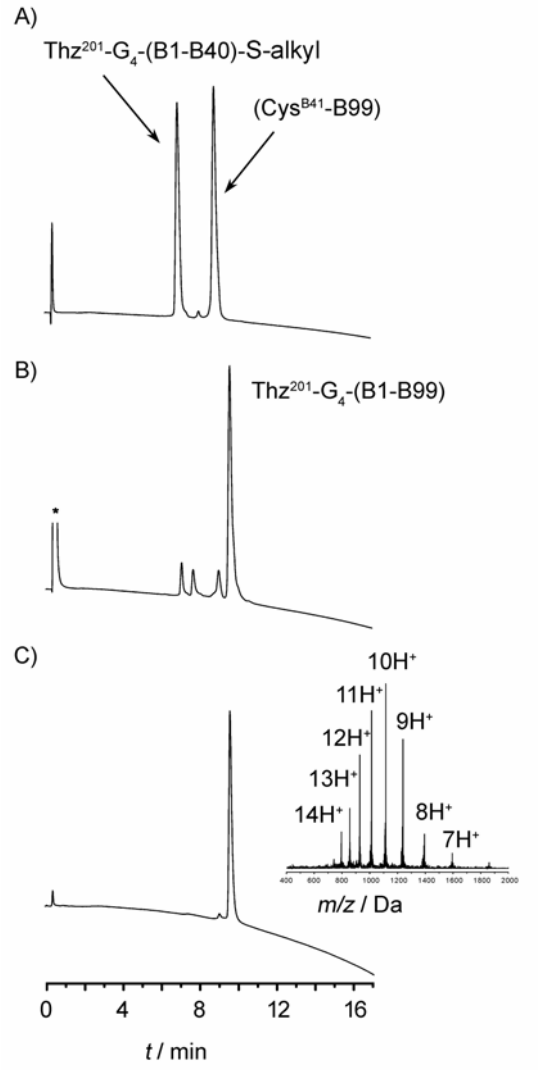


Figure S2. Analytical HPLC traces (λ = 214 nm) of native chemical ligation of Thz^{201}-Gly$_4$-(B1-B40)-$^{\alpha}$COSCH$_2$CH$_2$Arg$_4$.amide and (Cys^{B41}-B99)-COOH. A) *t* < 1min. B) *t* = 3 h. C) After HPLC purification. Insert is mass-spectrum of Cys^{201}-Gly$_4$-(B1-B99)-COOH (obs. 11137.5 ± 0.7 Da, calc. 11137.06 Da) integrated over elution time of target peak. Asterisk shows elution of 4-mercaptophenylacetic acid and other buffer components.

Folding. 0.5 mg of tethered construct of HIV-1 protease was dissolved in 3 mL of 6 M Gn·HCl, 50 mM NaOAc, pH 6.0 buffer and dialyzed first against 2 M Gn·HCl, 50 mM NaOAc, 10% (v/v) glycerol, pH 5.6 and then against 50 mM NaOAc, 10% (v/v) glycerol, pH 5.6. Yield is 29.1% based on final concentration of protein determined by absorbance at 280 nm using a calculated extinction coefficient of 25120 $M^{-1}cm^{-1}$.

Enzymology. Kinetic parameters were determined through assay with fluorogenic substrate Abz-Thr-Ile-Nle-Phe(*p*-NO_2)-Gln-Arg.amide (Abz = 2-aminobenzoyl),[2] which was added at various concentrations. Fluorimeter Jobin Yvon FluoroMax-3 (cell 1×1 cm) was used. Assay was performed in 50 mM NaOAc, 0.2 M NaCl buffer with 1 % (v/v) DMSO at pH 5.6 and 37 ºC. Values of k_{cat} 10.3 ± 0.2 s^{-1}, K_m 27 ± 1.4 μM were determined by nonlinear least square fitting of the initial rates to the Michaelis-Menten kinetic model with the help of SigmaPlot 9.0 software.

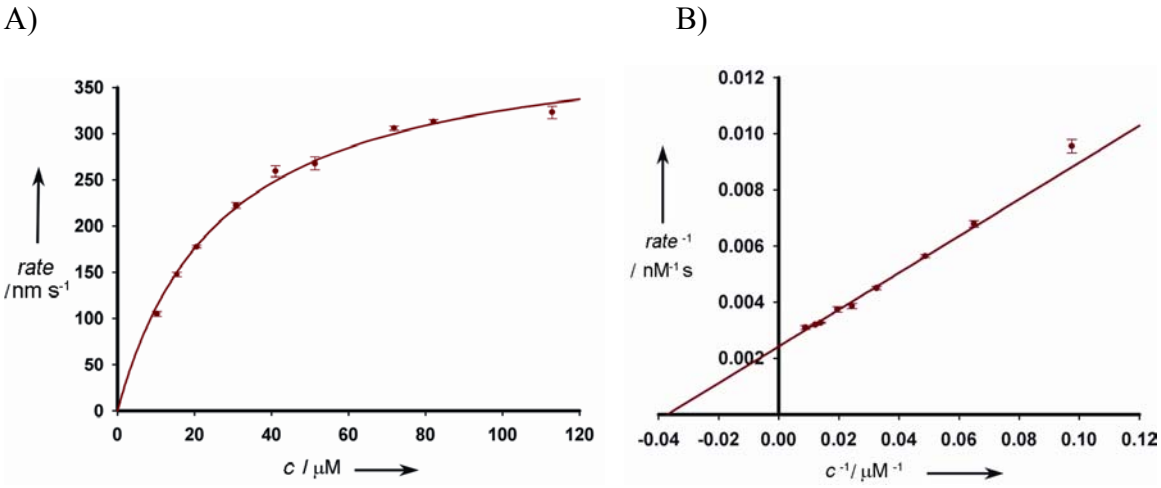
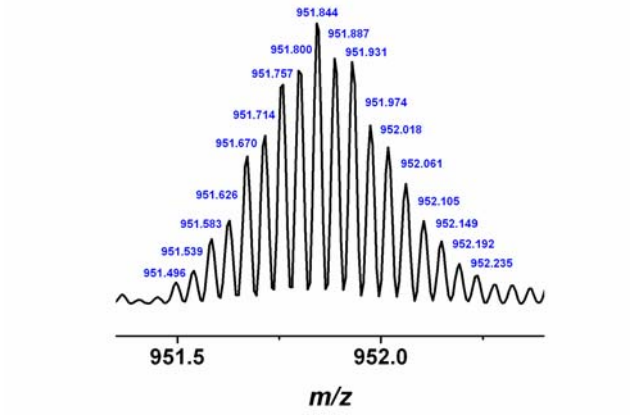


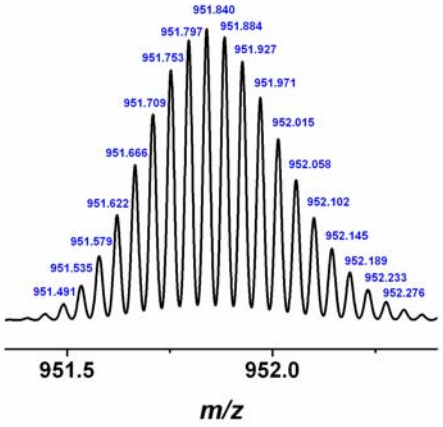
Figure S3. Enzymatic activity of the 'covalent dimer' HIV-1 PR. A) Initial rates versus concentration of fluorogenic substrate. B) Lineweaver-Burk plot.

FT-ICR experiments were performed at Chicago Biomedical Consortium proteomics facilities at University of Illinois at Chicago on Thermo Finnigan LTQ-FT LC/MS/MS instrument (Superconducting magnet 7.0 Tesla). Spectra were registered in LC-MS mode. Peak in total ion chromatogram was integrated and analyzed. Spectra were processed with Thermo Xcalibur v. 1.4 software. Simulation of isotopic pattern was done in IsoPro v. 3 and Gaussian fit and mathematic analysis was performed in Origin v. 7.0. Deconvolution to get average molecular weight was done with ProMass v. 2.5 software.

A)



B)



C)

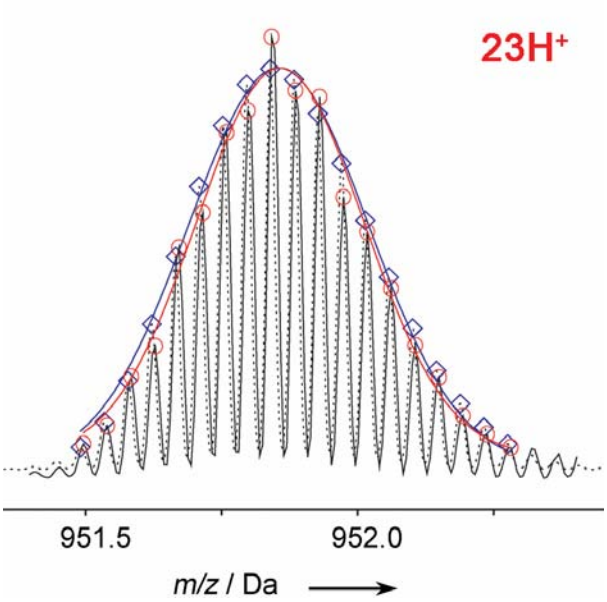


Figure S4. Experimental (A) and simulated (B) FT-ICR data for most abundant charge state $23H^+$. C) Superposition of isotope pattern for charge state $23H^+$ for experimental (solid line) and simulated (dashed line) spectra. Superposition is based on Gaussian best fit of maxima for experimental (red line, circles) and theoretical (blue line, rhombes) peaks.

Crystallography. Crystals were grown at 20 ºC by the hanging drop vapour diffusion method from a well solution consisting of 0.1 M citrate, 0.2 M NaH_2PO_4, 30% (w/v) saturated ammonium sulfate, 10% (v/v) DMSO, pH 6.0. Protein solution (~0.1 mM) was preincubated with a 30-fold molar excess of MVT-101, and was then mixed in different (v/v) ratios with well solution. Crystals grew within 1-4 days and were frozen in liquid nitrogen using mineral oil as cryoprotectant. Data were collected at the GM/CA-CAT (23ID) beamline at the Advanced Photon Source (Argonne National Laboratory). Crystal structure was solved by molecular replacement with the help of MolRep v. 8.1 and refined by Refmac v.5 (Collaborative Computational Project, Number 4. 1994. "The CCP4 Suite: Programs for Protein Crystallography." *Acta Cryst. D* **1994**, *50*, 760-763). Data were deposited to Protein Databank (PDB ID 2O40).

Table S1. Data collection and refinement statistics (highest resolution is in parenthesis).

Data collection	
Space group	$P2_12_12_1$
Cell dimensions	
a (Å)	51.326
b (Å)	58.306
c (Å)	61.699
$\alpha = \beta = \gamma$ (º)	90
Resolution (Å)	50 – 1.65
R_{merge}	5.5 (74.9)
$I/\sigma I$	14.6 (2.67)
Redundancy	4.8 (4.6)
Refinement	
Resolution (Å)	20 – 1.65
Completeness (%)	98.7
No. reflections	
Work/free set	21383 / 1163
R_{work}/R_{free}	19.1 / 23.7
No. atoms	
Protein	1597
Water	57
B-factors (Å^2)	
Protein	24.7
Inhibitor	27.9
Water	30.6
R.m.s. deviations	
Bond lengths (Å)	0.017
Bond angles (º)	1.599

Acknowledgements.

We acknowledge B. L. Pentelute (University of Chicago) for the help in the synthesis of MVT-101 inhibitor, Dr. E. C. B. Johnson (University of Chicago) for providing fluorogenic substrate Abz-Thr-Ile-Nle-Phe(*p*-NO_2)-Gln-Arg.amide and Dr. M. Senko (Thermo Finnigan) for providing IsoPro v.3 software.

[1] M. Schnölzer, P. Alewood, A. Jones, D. Alewood, S. B. H. Kent, *Int. J. Peptide Protein Res.* **1992**, *40*, 180-193.

[2] M. V. Toth, G. R. Marshall, *Int. J. Pept. Protein Res.* **1990**, *36*, 544-550.