Supplementary Information

Synthesis of the full-length hepatitis B virus core protein and its capsid formation

Keisuke Aoki,[a,b] Shugo Tsuda,[c] Naoko Ogata,[b] Michiyo Kataoka,[d] Jumpei Sasaki,[a] Shinsuke Inuki,[a] Hiroaki Ohno,[a] Koichi Watashi,[e] Taku Yoshiya[c,f] and Shinya Oishi[a,b,*]

[a] *Graduate School of Pharmaceutical Sciences, Kyoto University, Sakyo-ku, Kyoto 606-8501, Japan*
[b] *Laboratory of Medicinal Chemistry, Kyoto Pharmaceutical University, Yamashina-ku, Kyoto 607-8412, Japan*
[c] *Peptide Institute, Inc. Ibaraki, Osaka 567-0085, Japan*
[d] *Department of Pathology, National Institute of Infectious Disease, Shinjuku-ku, Tokyo 162-8640, Japan*
[e] *Research Center for Drug and Vaccine Development, National Institute of Infectious Disease, Shinjuku-ku, Tokyo 162-8640, Japan*
[f] *Institute for Protein Research, Osaka University, Suita, Osaka 565-0871, Japan*

Contents

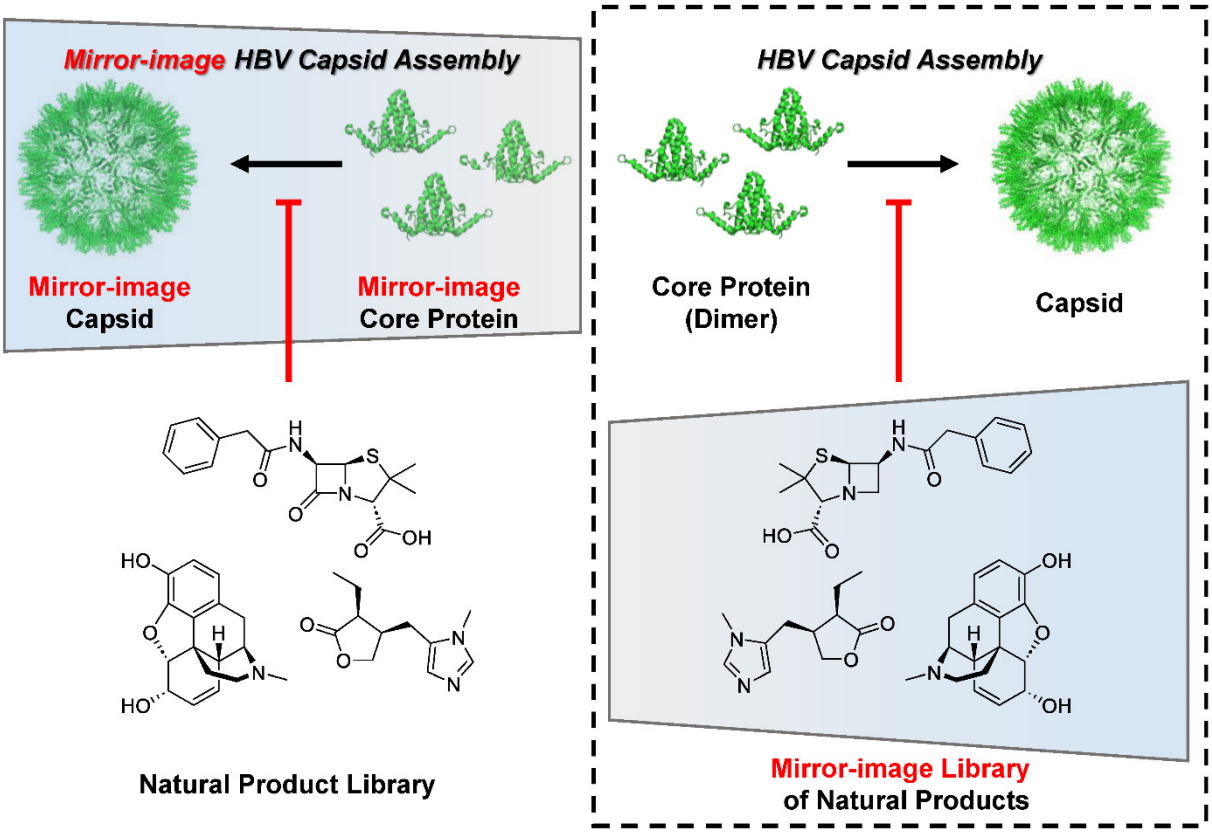


Fig. S1 Screening concept for HBV capsid assembly modulators from a virtual mirror-image library of natural products. Screening of the natural product library using the mirror-image HBV core protein corresponds to the mirror-image library of natural products using the natural HBV core protein in a mirror. The HBV core protein structure was obtained from the Protein Data Bank (PDB ID: 6HTX).

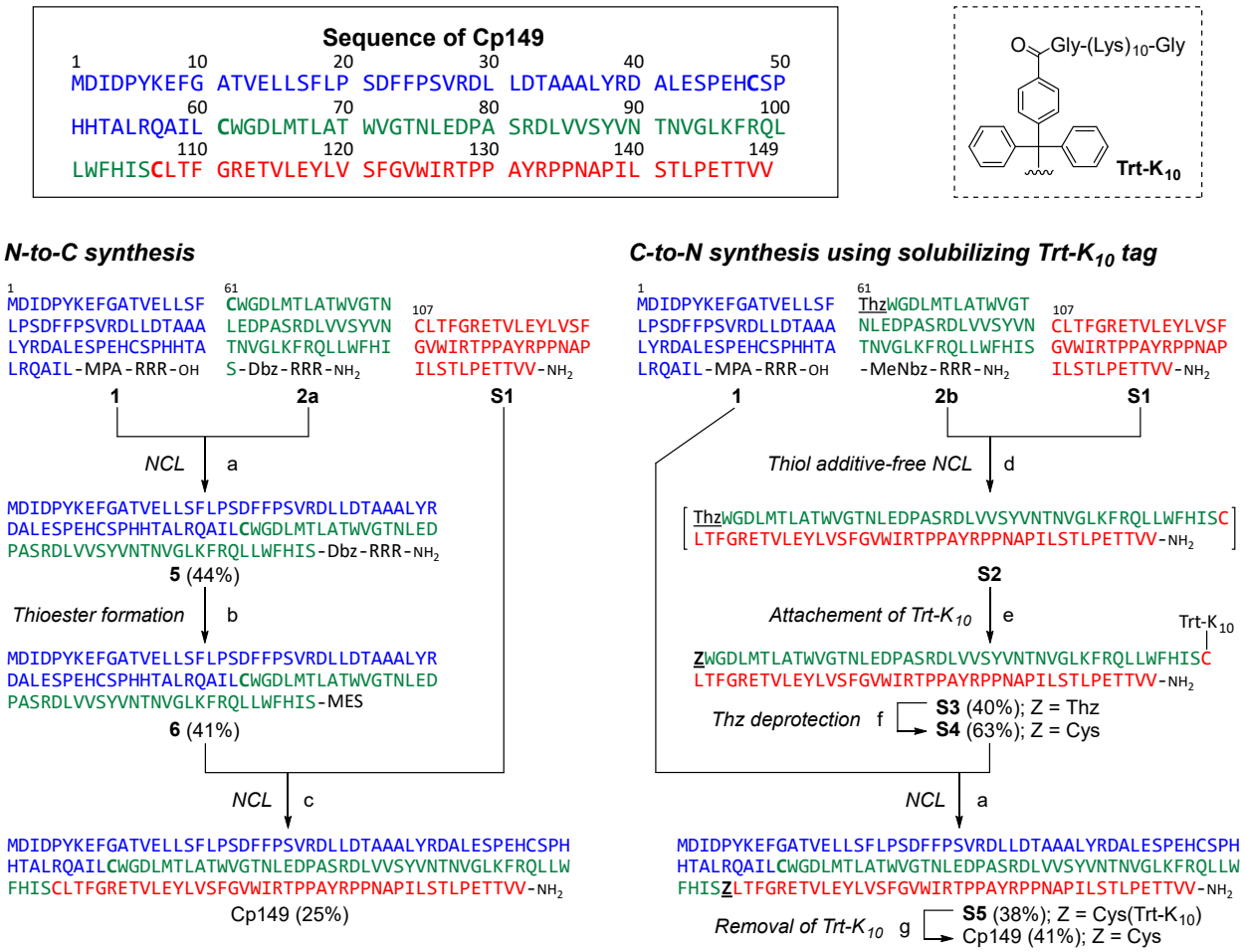


Fig. S2 Previous synthetic studies on Cp149.[S1] *Reagents and conditions*: (a) MPAA, TCEP, 6 M guanidine·HCl, and 200 mM phosphate buffer (pH 7.0); (b) $NaNO_2$, 6 M guanidine·HCl, 200 mM phosphate buffer, then MESNa and TCEP; (c) MPAA, TCEP, 6 M guanidine·HCl, and 200 mM phosphate buffer (pH 7.0) containing 25% NMP; (d) 1,2,4-triazole, TCEP, 6 M guanidine·HCl, and 100 mM phosphate buffer (pH 7.1); (e) Trt(OH)-K_{10} and TFA; (f) methoxyamine, 6 M guanidine·HCl, and 200 mM phosphate buffer (pH 4.0); (g) TFA/TIS (95:5). *Abbreviations*: Dbz, 3,4-diaminobenzoic acid; MeNbz, *N*-acyl-*N*′-methylacylurea; MES, 2-mercaptethanesulfonate; MPA, 3-mercaptopropionic acid; MPAA, 4-mercaptophenylacetic acid; TCEP, tris(2-carboxyethyl)phosphine; Thz, thiazolidine carboxylic acid.

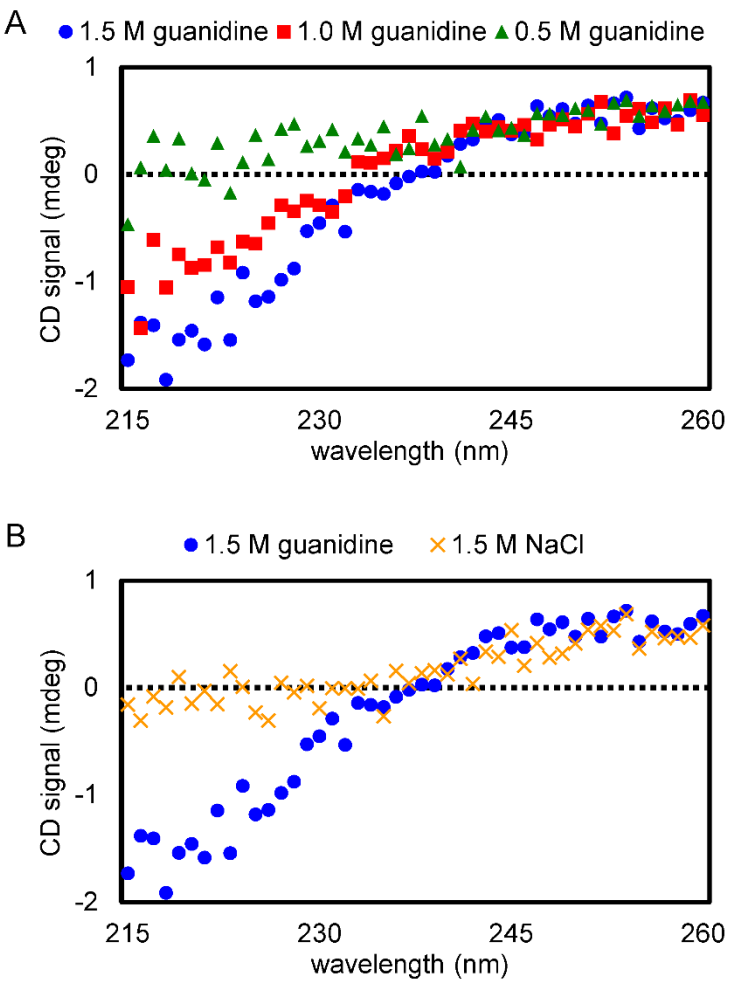


Fig. S3 CD analysis of the synthetic HBV core protein under various conditions. (A) CD spectra of synthetic Cp183(C183A) at various concentrations of guanidine. (B) Comparison of synthetic Cp183(C183A) CD spectra in a 1.5M guanidine or NaCl solution.

References

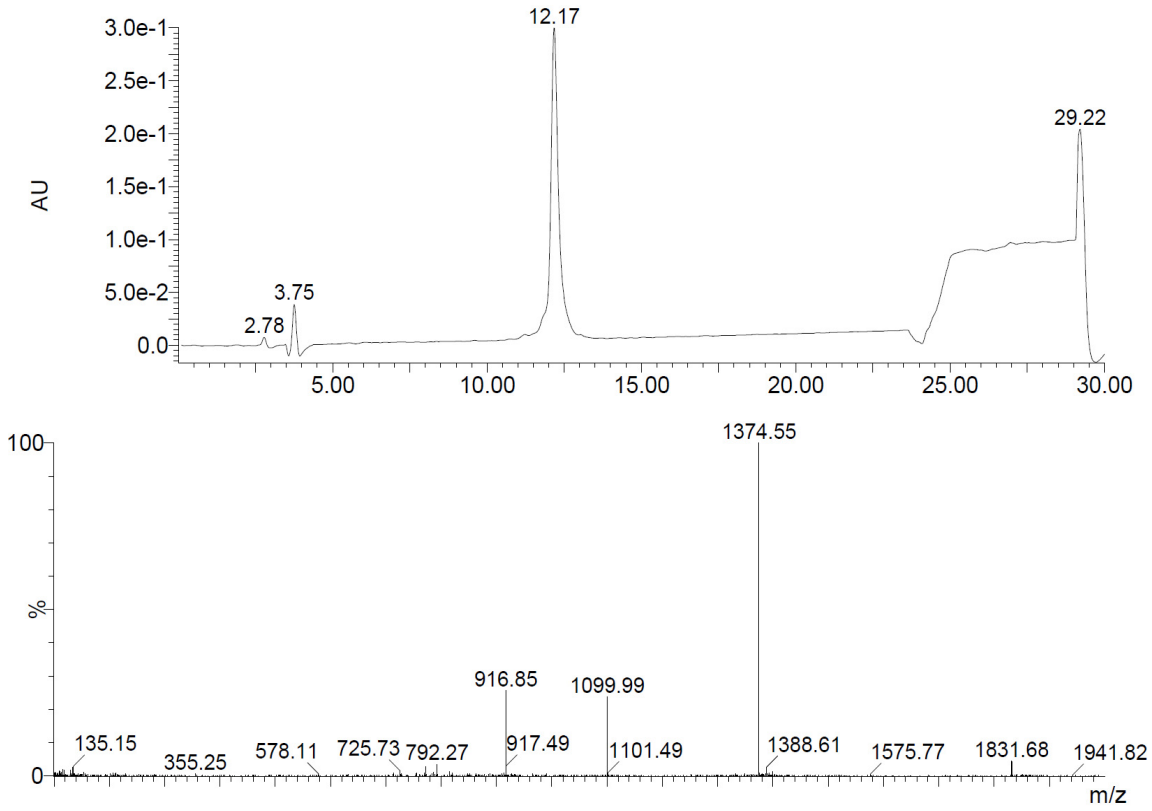
S1 S. Tsuda, M. Mochizuki, H. Ishiba, K. Yoshizawa-Kumagaye, H. Nishio, S. Oishi and T. Yoshiya, *Angew. Chem. Int. Ed.*, 2018, **57**, 2105–2109.

Analytical HPLC Chromatograms and Mass Spectrometry Data of Synthetic Peptides

[Thz^{107}]-Cp183$^{107–153}$-MPAA (3)

ThzLTFGRETVLEYLVSFGVWIRTPPAYRPPNAPILSTLPETTVVRRRG-MPAA

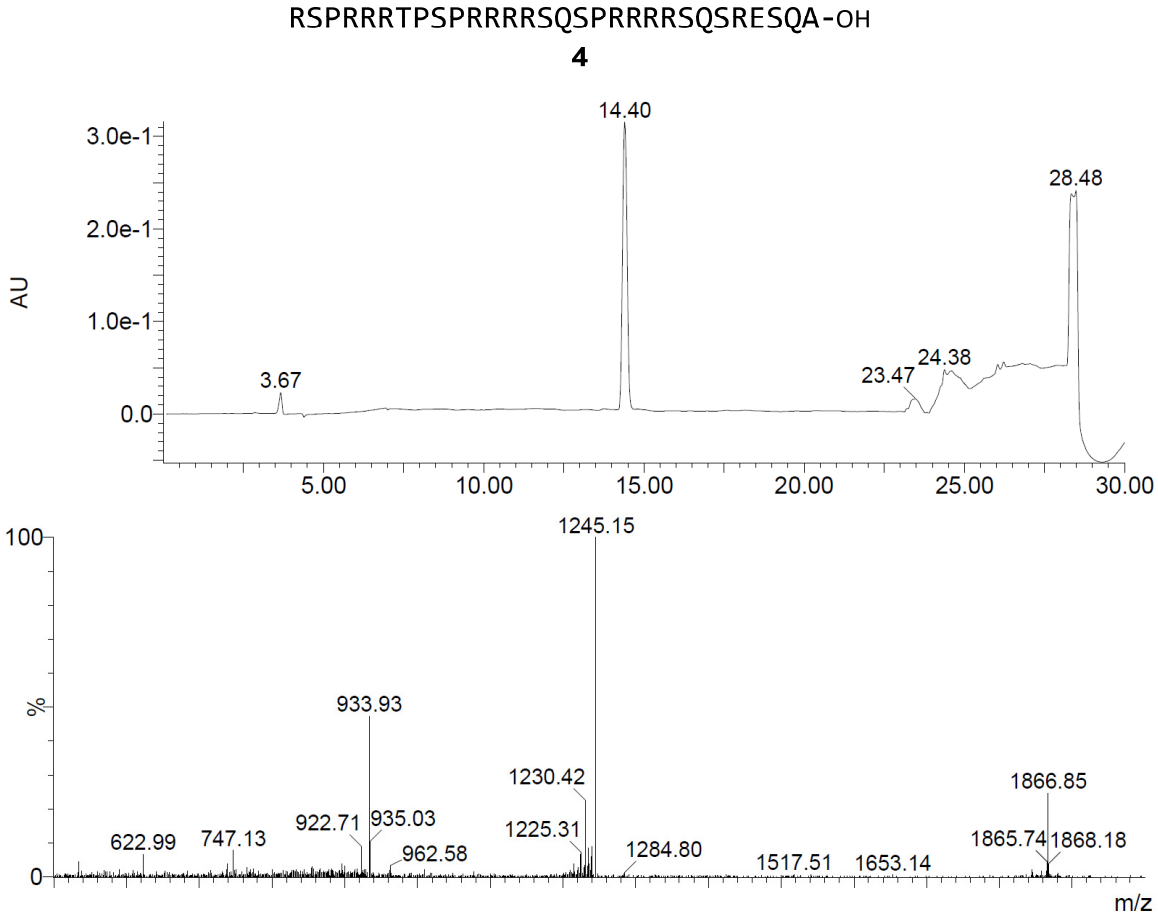
3



HPLC conditions: Cosmosil 5C18-AR300 column (Nacalai Tesque, 4.6 × 250 mm), linear gradient of 35–55% CH_3CN containing 0.1% TFA at a flow rate of 1 mL/min over 20 min.

MS analysis: Expected mass based on the sequence: 5493.45; Major observed ions: $[M+6H]^{6+}$ *m/z* = 916.85, $[M+5H]^{5+}$ *m/z* = 1099.99, $[M+4H]^{4+}$ *m/z* = 1374.55, $[M+3H]^{3+}$ *m/z* = 1831.68.

[Ala^{183}]-Cp183$^{154–183}$ (4)



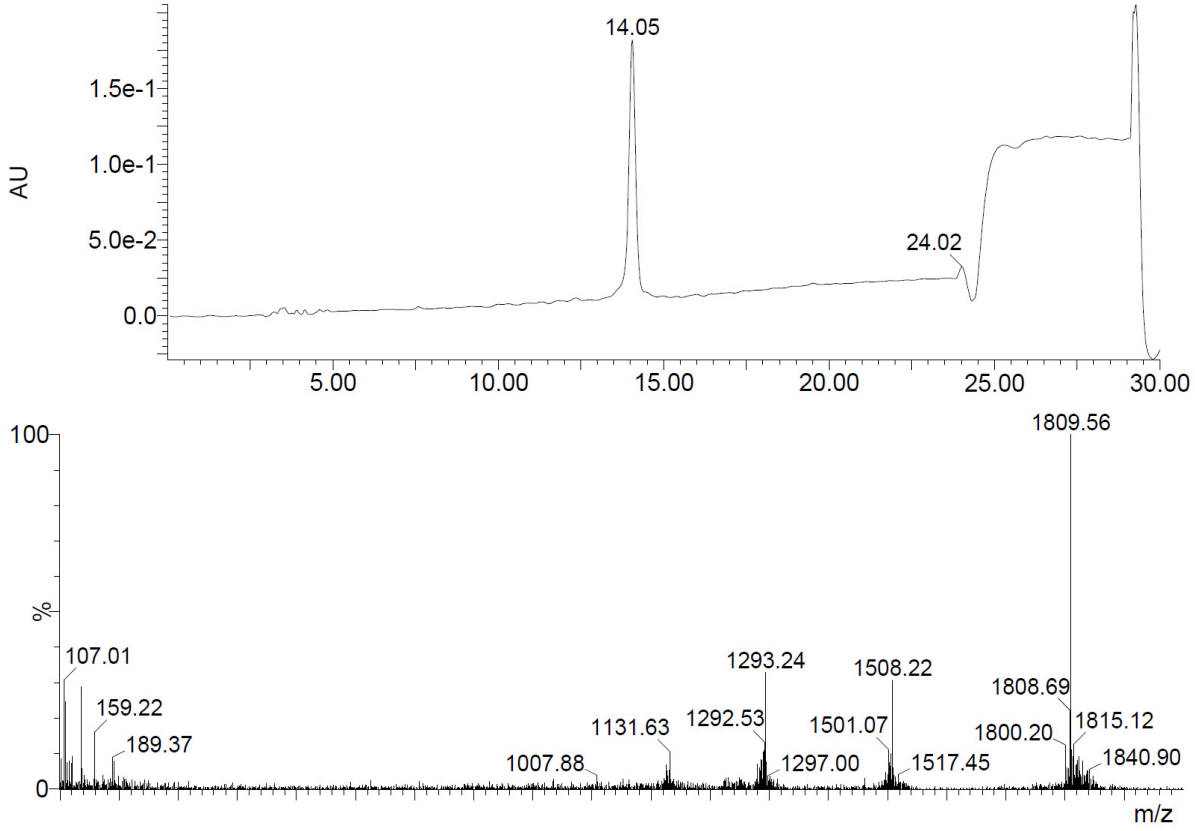
HPLC conditions: Cosmosil 5C18-AR300 column (Nacalai Tesque, 4.6 × 250 mm), linear gradient of 0–20% CH_3CN containing 0.05% TFA at a flow rate of 1 mL/min over 20 min.

MS analysis: Expected mass based on the sequence: 3732.18; Major observed ions: $[M+6H]^{6+}$ *m/z* = 622.99, $[M+5H]^{5+}$ *m/z* = 747.13, $[M+4H]^{4+}$ *m/z* = 933.93, $[M+3H]^{3+}$ *m/z* = 1245.15, $[M+2H]^{2+}$ *m/z* = 1866.85.

[Ala^{183}]-Cp183$^{107–183}$ (8)

CLTFGRETVLEYLVSFGVWIRTPPAYRPPNAPILSTLPET
TVVRRRGRSPRRRTPSPRRRRSQSPRRRRSQSRESQA-OH

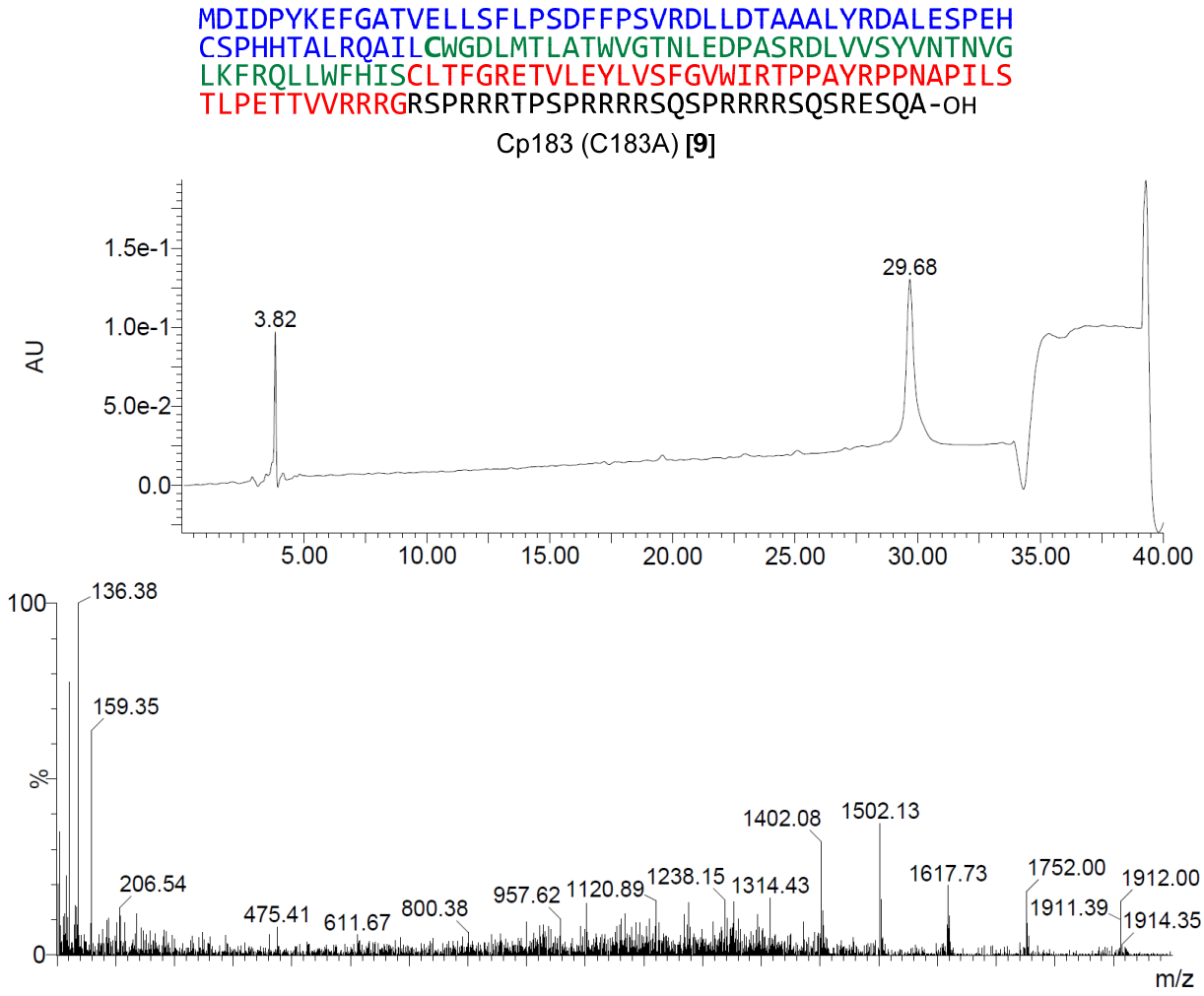
8



HPLC conditions: Cosmosil 5C4-AR300 column (Nacalai Tesque, 4.6 × 250 mm), linear gradient of 30–50% CH_3CN containing 0.1% TFA at a flow rate of 1 mL/min over 20 min.

MS analysis: Expected mass based on the sequence: 9045.41; Major observed ions: $[M+8H]^{8+}$ *m/z* = 1131.63, $[M+7H]^{7+}$ *m/z* = 1293.24, $[M+6H]^{6+}$ *m/z* = 1508.22, $[M+5H]^{5+}$ *m/z* = 1809.56.

[Ala^{183}]-Cp183$^{1–183}$ (9)



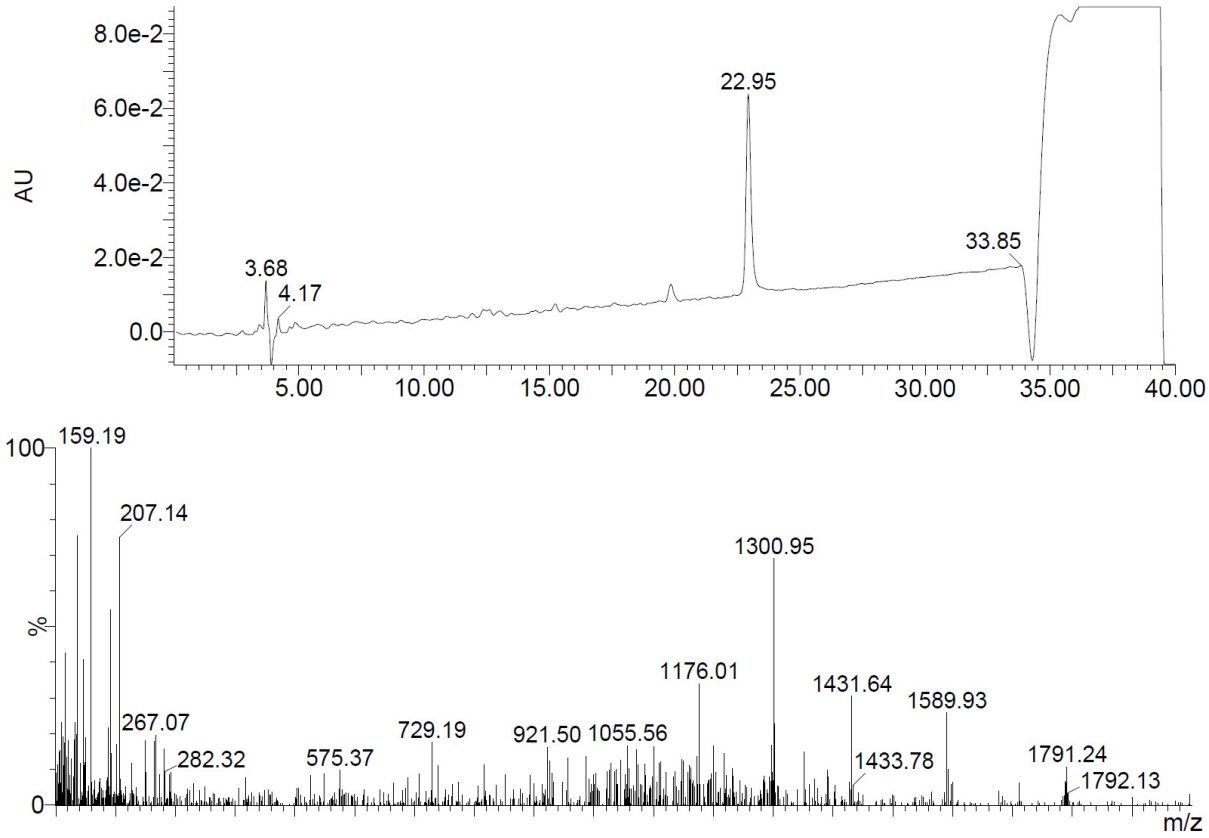
HPLC conditions: Cosmosil 5C4-AR300 column (Nacalai Tesque, 4.6 × 250 mm), linear gradient of 30–60% CH_3CN containing 0.1% TFA at a flow rate of 1 mL/min over 30 min.

MS analysis: Expected mass based on the sequence: 21010.06; Major observed ions: $[M+15H]^{15+}$ *m/z* = 1402.08, $[M+14H]^{14+}$ *m/z* = 1502.13, $[M+13H]^{13+}$ *m/z* = 1617.73, $[M+12H]^{12+}$ *m/z* = 1752.00, $[M+11H]^{11+}$ *m/z* = 1912.00.

[Ala^{183}]-Cp183$^{61–183}$ (11)

CWGDLMTLATWVGTNLEDPASRDLVVSYVNTNVGLKFRQLLW
FHISCLTFGRETVLEYLVSFGVWIRTPPAYRPPNAPILSTLP
ETTVVRRRGRSPRRRTPSPRRRRSQSPRRRRSQSRESQA-OH

11



HPLC conditions: Cosmosil 5C4-AR300 column (Nacalai Tesque, 4.6 × 250 mm), linear gradient of 30–60% CH_3CN containing 0.1% TFA at a flow rate of 1 mL/min over 30 min.

MS analysis: Expected mass based on the sequence: 14295.46; Major observed ions: $[M+11H]^{11+}$ *m/z* = 1300.95, $[M+10H]^{10+}$ *m/z* = 1431.64, $[M+9H]^{9+}$ *m/z* = 1589.93.