Available online at www.sciencedirect.com
SCIENCE DIRECT
Bioorganic & Medicinal Chemistry Letters 16 (2006) 3761–3764
Bioorganic & Medicinal Chemistry Letters

Evaluations of substrate specificity and inhibition at PR/p3 cleavage site of HTLV-1 protease

Hiromi Naka,[a] Kenta Teruya,[a] Jeong Kyu Bang,[a] Saburo Aimoto,[a] Tadashi Tatsumi,[b] Hiroyuki Konno,[b] Kazuto Nosaka[b] and Kenichi Akaji[b,*]

[a]*Institute for Protein Research, Osaka University, Osaka 565-0871, Japan*
[b]*Kyoto Prefectural University of Medicine, Kyoto 603-8334, Japan*

Received 22 March 2006; revised 8 April 2006; accepted 18 April 2006
Available online 8 May 2006

Abstract—Core sequences necessary for substrate recognition and its inhibition at the PR/p3 site of HTLV-1 protease were clarified for the first time. From the cleavage rates of peptides containing a part of the PR/p3 site, a heptapeptide was found to be the minimal sequence required for substrate recognition. The use of synthetic inhibitors containing hydroxyethylamine dipeptide isostere indicated that a tetrapeptide sequence was necessary to achieve potent inhibition.

Human T-cell leukemia virus type I (HTLV-1) is a retrovirus that is etiologically associated with human adult T-cell leukemia and a number of chronic diseases.[1] As in other retroviruses, HTLV-1 proteins are initially translated as large precursor polyproteins that undergo proteolytic processing by HTLV-1 protease. The protease is an aspartic protease and is itself auto-processed from a precursor protein.[2] Thus, HTLV-1 protease is a crucial factor for successful virus replication. In a previous paper, we reported stereo selective syntheses of HTLV-1 protease inhibitors containing hydroxyethylamine dipeptide isostere, and found for the first time that the configuration at the hydroxyl- and side chain-bearing asymmetric centers shows effects on the inhibitory activity by 2 orders of magnitude.[3] In this paper, we report a core sequence for protease inhibition at the PR/p3 site as well as the evaluation of a minimum substrate sequence necessary for protease recognition.

Identification of the minimum sequence required for substrate recognition is basic information for the design of protease inhibitors. HTLV-1 protease cleaves the precursor protein at several processing sites such as MA/CA (matrix/capsid) site, CA/NC (capsid/nucleocapsid) site, GAG/PR (Gag protein/protease) site, and PR/p3 (protease C-terminal) site.[4–6] Among them, we investigated the substrate sequence necessary for recognition at the PR/p3 site (KGPPVILPIQAP, in which LP is the scissile site), since no detailed examination at this sequence was achieved.

Twelve peptide amides (**S-1** to **S-12**) containing a portion of the PR/p3 processing site were synthesized using stepwise Fmoc-based solid-phase peptide synthesis (SPPS) starting from Rink amide resin[7] (4-(2′,4′-dimethoxyphenyl-Fmoc-aminomethyl)-phenoxy resin). SPPS was achieved by the combination of Fmoc deprotection using 20% piperidine/DMF with a coupling reaction using a standard diisopropylcarbodiimide (DIPCDI)/HOBt protocol. Each peptide was purified by preparative HPLC after cleavage from the resin by treatment with TFA–triisopropylsilane (TIS)–H_2O (95:2.5:2.5) at 25 °C for 1 h. Homogeneity was further confirmed by MALDI-TOF MS and amino acid analysis.[8]

The digestion of each synthetic peptide was conducted using a chemically synthesized mutant of HTLV-1 protease (C2A HTLV-1 PR), since the kinetic properties of this autodigestion-resistant mutant have already been evaluated.[9] The cleavage reaction was conducted according to the published procedure[9] and was monitored by analytical HPLC. Typical HPLC profiles are shown in Figure 1. Each hydrolysis product was identified by MALDI-TOF MS analysis.[10] After incubation with the protease for 2 h at 37 °C, the cleavage rate of

Keywords: HTLV-1 protease; Inhibitor; Hydroxyethylamine dipeptide isostere; Substrate specificity.
* Corresponding author. E-mail: akaji@koto.kpu-m.ac.jp

0960-894X/$ - see front matter © 2006 Elsevier Ltd. All rights reserved.
doi:10.1016/j.bmcl.2006.04.056

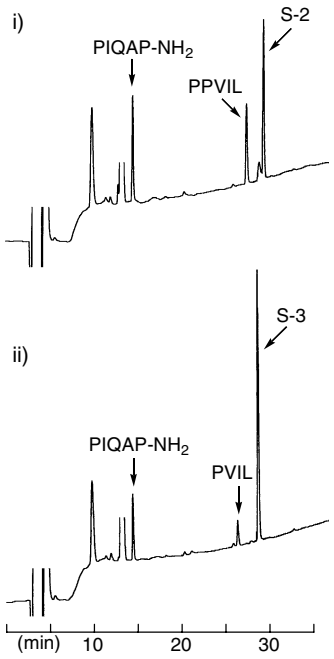


Figure 1. HPLC profiles of hydrolysis products obtained after digestion with HTLV-1 protease: (i) digestion of **S-2** containing P_5–P'_5 sequence; (ii) digestion of **S-3** containing P_4–P'_5 sequence. [HPLC: Cosmosil 5C18-AR column (4.6 × 250 mm); 1.0 ml/min; CH_3CN (5–35% in 40 min) in 0.1% aqueous TFA].

each synthetic peptide was estimated from the amount of the corresponding hydrolysis product. As summarized in Figure 2, a 9-residue peptide covering the P_5–P'_4 sequence (**S-5**) was cleaved to yield approximately 85% of the hydrolysis product compared with the 12-residue peptide (**S-1**) used as the standard substrate.[9] Five peptides (**S-4**, **S-7**, and **S-10**–**12**) were not cleaved even after 24 h hydrolysis with the protease. The minimal sequence required for substrate recognition was a 7-residue peptide containing the P_4–P'_3 sequence (**S-9**).

We then synthesized inhibitors containing a hydroxyethylamine dipeptide isostere, a typical transition-state isostere,[3,11] to examine the minimal sequence required for inhibition of the HTLV-1 protease. Among the possible isostere configurations, (*R*)-hydroxy- and (*S*)-side chains at the scissile site were adopted as suitable configurations according to our previous report.[3] Based on the above results regarding substrate recognition, six inhibitors (**I-1** to **I-6**) containing a portion of the P_5–P'_2 sequence of the PR/p3 site were synthesized on a solid support.

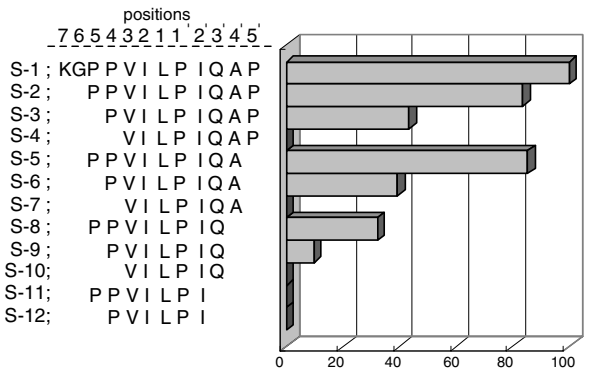
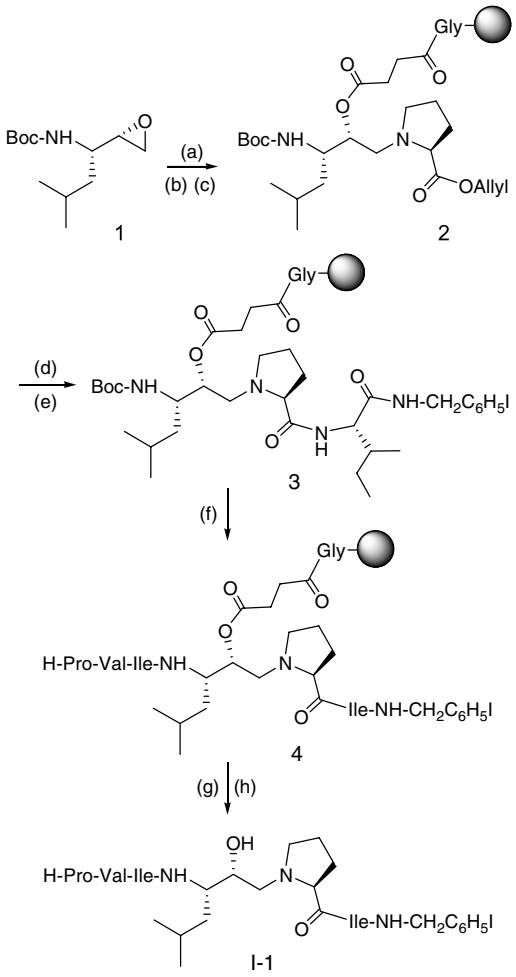


Figure 2. Cleavage rate at the PR/p3 site of HTLV-1 protease: cleavage of **S-1** substrate was taken as 100%.

The inhibitor containing a P_4–P'_2 sequence (**I-1**) was synthesized on a solid support using the acid-stable succinate ester linker (Scheme 1). The (*S*,*R*)-aminoalkyl epoxide **1**, a key intermediate in our previous synthesis,[3] was reacted with H-Pro-OAllyl at 25 °C for 18 h to give the hydroxyethylamine product at 63% yield.[12] A succinic acid linker was introduced by reaction with succinic anhydride in the presence of DMAP at 25 °C for 16 h, and the product was anchored to H-Gly-MBHA (glycyl *p*-methylbenzhydrylamine) resin with DIPCDI/HOBt-mediated reaction (25 °C, 18 h). The resulting resin **2** contains an N-terminal Boc group and a C-terminal allyl group, and thus can be elongated selectively in



Scheme 1. Synthetic scheme for inhibitor **I-1** containing the P_4–P'_2 sequence. Reagents and conditions: (a) H-Pro-OAllyl in 2-propanol, 25 °C, 18 h; (b) succinic anhydride/DMAP, 25 °C, 16 h; (c) H-Gly-MBHA resin/DIPCDI/HOBt, 25 °C, 18 h; (d) $[(C_6H_5)_3P]_4Pd(0)$ under Ar, 25 °C, 2 h; (e) H-Ile-$NHCH_2C_6H_5I$/DIPCDI/HOBt, 25 °C, 18 h; (f) Boc-based solid-phase synthesis; (g) HF, 4 °C, 30 min; (h) aqueous ammonium acetate solution pH 10, 25 °C, 5 days.

either direction. The C-terminal allyl group of **2** was removed with $[(C_6H_5)_3P]_4Pd(0)$ (25 °C, 2 h),[13] and H-Ile-$NHCH_2C_6H_4I$ was coupled to the resulting resin using DIPCDI/HOBt (25 °C, 18 h). The Boc group of the product resin **3** was removed with 50% TFA/CH_2Cl_2, and Boc-Ile-OH was condensed by DIPCDI/HOBt. The same deprotection/condensation procedure was repeated for the successive introduction of Boc-Val-OH and Boc-Pro-OH. Part of the intermediate resin was separated for the parallel preparation of an inhibitor (**I-4**) containing the P_3–P'_2 sequence.

The product resin **4** was then treated with HF at 4 °C for 30 min to cleave the inhibitor precursor from the resin. The crude product showed a single major peak on HPLC and, without further purification, was treated with aqueous $AcONH_4$ at pH 10. After 5 days, the precursor disappeared on HPLC and the product was purified by preparative HPLC to produce **I-1** with 19% isolation yield (calculated from the starting MBHA resin). Other inhibitors containing different chain lengths (compound **I-2** covering the P_5–P'_1 sequence, compound **I-3** for the P_4–P'_1 sequence, compound **I-5** for the P_3–P'_1 sequence, and compound **I-6** for the P_2–P'_1 sequence) were similarly synthesized on MBHA resin.[14]

The inhibitory activity of each compound was examined using a synthetic dodecapeptide (**S-1**) as the substrate. Cleavage of the substrate by mutant HTLV-1 protease in the presence of various concentrations of the inhibitor was monitored by analytical HPLC, as described above. The inhibitory activity of each inhibitor was evaluated using the corresponding IC_{50} value obtained from the sigmoidal dose–response curve (Fig. 3). As summarized in Table 1, the P_3–P'_1 sequence (**I-5**) of the scissile site is the core sequence for the inhibition of HTLV-1 protease. The addition of P_4-Pro to the core sequence was more effective than P'_2-Ile (**I-3** vs **I-4**), and the P_4–P'_1 sequence was necessary to compare an inhibitor such as pepstatin, a standard inhibitor for aspartic acid proteases.

Thus, it has been shown for the first time that a 7-residue peptide containing the P_4–P'_3 sequence of the PR/p3 site is required for the substrate recognition of HTLV-1 protease. From the evaluation of synthetic inhibitors containing the hydroxyethylamine dipeptide isostere, the P_3–P'_1 sequence of the PR/p3 site has been shown to be a core sequence for achieving potent inhibition of the protease. Based on these findings, the construction of HTLV-1 protease inhibitor libraries is currently in progress.

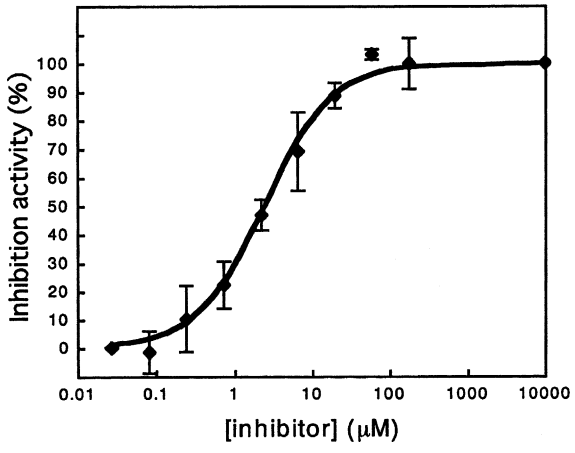


Figure 3. A typical sigmoidal dose–response curve used for estimation of the IC_{50} value of the inhibitor **I-2**.

Table 1. Inhibition of HTLV-1 protease by inhibitors containing the PR/p3 site sequence

	Compounds[a]	IC_{50}[b] (μM)
I-1	H-Pro-Val-Ile-[HEA]-Pro-Ile-$NHCH_2C_6H_5I$	0.13
I-2	H-Pro-Pro-Val-Ile-[HEA]-Pro-$NHCH_2C_6H_5I$	2.4
I-3	H-Pro-Val-Ile-[HEA]-Pro-$NHCH_2C_6H_5I$	5.8
I-4	H-Val-Ile-[HEA]-Pro-Ile-$NHCH_2C_6H_5I$	18.3
I-5	H-Val-Ile-[HEA]-Pro-$NHCH_2C_6H_5I$	19.6
I-6	H-Ile-[HEA]-Pro-$NHCH_2C_6H_5I$	na

[a] [HEA], $NHCH(CH_2CH(CH_3)_2)CH(OH)CH_2$–.
[b] Values are means of three experiments (na, not active).

Acknowledgments

This research was supported in part by a grant from Mitsubishi Pharma Research Foundation, and a grant from Japan Leukemia Research Fund.

References and notes

1. Matsuoka, M. *Oncogene* **2003**, *22*, 5131.
2. Louis, J. M.; Oroszlan, S.; Tözser, J. *J. Biol. Chem.* **1999**, *274*, 6660.
3. Akaji, K.; Teruya, K.; Aimoto, S. *J. Org. Chem.* **2003**, *68*, 4755.
4. Ding, Y. S.; Rich, D. H.; Ikeda, R. A. *Biochemistry* **1998**, *37*, 17514.
5. Tözser, J.; Zahuczky, G.; Bagossi, P.; Louis, J. M.; Copeland, T. D.; Oroszlan, S.; Harrison, R. W.; Weber, I. T. *Eur. J. Biochem.* **2000**, *267*, 6287.
6. Kádas, J.; Weber, I. T.; Bagossi, P.; Miklóssy, G.; Boross, P.; Oroszland, S.; Tözser, J. *J. Biol. Chem.* **2004**, *279*, 27148.
7. Rink, H. *Tetrahedron Lett.* **1987**, *28*, 3787.
8. Compound **S-1**: *m/z* 1228.81 for $[M+H]^+$ (calcd 1228.78 for $C_{59}H_{102}N_{15}O_{13}$); Lys (1) 1.1, Gly (1) 1.1, Pro (4) 4.6, Val (1) 1.1, Ile (2) 2.1, Leu (1) 1.1, Gln (1) 1.0, Ala (1) 1.0. Compound **S-2**: *m/z* 1043.90 for $[M+H]^+$ (calcd 1043.66 for $C_{51}H_{87}N_{12}O_{11}$); Pro (4) 4.3, Val (1) 0.87, Ile (2) 2.0, Leu (1) 1.0, Gln (1) 1.1, Ala (1) 1.0. Compound **S-3**: *m/z* 946.28 for $[M+H]^+$ (calcd 946.61 for $C_{46}H_{80}N_{11}O_{10}$); Pro (3) 3.3, Val (1) 0.93, Ile (2) 1.8, Leu (1) 0.94, Gln (1) 1.0, Ala (1) 1.0. Compound **S-4**: *m/z* 871.45 for $[M+Na]^+$ (calcd 871.54 for $C_{41}H_{72}N_{10}O_9Na$); Pro (2) 2.6, Val (1) 1.0, Ile (2) 2.1, Leu (1) 1.1, Gln (1) 1.1, Ala (1) 1.0. Compound **S-5**: *m/z* 946.70 for $[M+H]^+$ (calcd 946.61 for $C_{46}H_{80}N_{11}O_{10}$); Pro (3) 3.6, Val (1) 1.0, Ile (2) 2.1, Leu (1) 1.1, Gln (1) 1.1, Ala (1) 1.0. Compound **S-6**: *m/z* 850.17 for $[M+H]^+$ (calcd 849.56 for $C_{41}H_{73}N_{10}O_9$); Pro (2) 2.5, Val (1) 0.95, Ile (2) 2.0, Leu (1) 1.1, Gln (1) 1.1, Ala (1) 1.0. Compound **S-7**: *m/z* 752.73 for $[M+H]^+$ (calcd 752.50 for $C_{36}H_{66}N_9O_8$); Pro (1) 1.5, Val (1) 1.0, Ile (2) 2.1, Leu (1) 1.0, Gln (1) 1.1, Ala (1) 1.0. Compound **S-8**: *m/z* 875.99 for

$[M+H]^+$ (calcd 875.57 for $C_{43}H_{75}N_{10}O_9$); Pro (3) 3.4, Val (1) 1.0, Ile (2) 2.0, Leu (1) 1.1, Gln (1) 1.1. Compound **S-9**: *m/z* 778.70 for $[M+H]^+$ (calcd 778.52 for $C_{38}H_{68}N_9O_8$); Pro (2) 2.3, Val (1) 1.0, Ile (2) 2.0, Leu (1) 1.0, Gln (1) 1.1. Compound **S-10**: *m/z* 681.89 for $[M+H]^+$ (calcd 681.47 for $C_{33}H_{61}N_8O_7$); Pro (1) 1.4, Val (1) 0.94, Ile (2) 2.0, Leu (1) 1.0, Gln (1) 1.1. Compound **S-11**: *m/z* 747.66 for $[M+H]^+$ (calcd 747.51 for $C_{38}H_{67}N_8O_7$); Pro (3) 3.2, Val (1) 0.94, Ile (2) 2.0, Leu (1) 1.0. Compound **S-12**: *m/z* 650.31 for $[M+H]^+$ (calcd 650.46 for $C_{33}H_{60}N_7O_6$); Pro (2) 2.1, Val (1) 0.97, Ile (2) 2.0, Leu (1) 1.1.

9. Teruya, K.; Kawakami, T.; Akaji, K.; Aimoto, S. *Tetrahedron Lett.* **2002**, *43*, 1487.
10. H-Lys-Gly-Pro-Pro-Val-Ile-Leu-OH: *m/z* 723.39 for $[M+H]^+$ (calcd 723.48 for $C_{35}H_{63}N_8O_8$); Lys (1) 0.96, Gly (1) 1.3, Pro (2) 2.2, Val (1) 0.90, Ile (1) 0.89, Leu (1) 1.0. H-Pro-Ile-Gln-Ala-Pro-NH_2: *m/z* 546.18 for $[M+Na]^+$ (calcd 546.30 for $C_{24}H_{41}N_7O_6Na$); Pro (2) 2.3, Ile (1) 0.96, Gln (1) 1.32, Ala (1) 1.0. H-Pro-Pro-Val-Ile-Leu-OH: *m/z* 538.24 for $[M+H]^+$ (calcd 538.36 for $C_{27}H_{48}N_5O_6$); Pro (2) 2.2, Val (1) 0.92, Ile (1) 0.90, Leu (1) 1.0. H-Pro-Val-Ile-Leu-OH: *m/z* 463.18 for $[M+Na]^+$ (calcd 463.29 for $C_{22}H_{40}N_4O_5Na$); Pro (1) 0.97, Val (1) 0.96, Ile (1) 0.90, Leu (1) 1.0.
11. Maegawa, H.; Kimura, T.; Arii, Y.; Matsui, Y.; Kasai, S.; Hayashi, Y.; Kiso, Y. *Bioorg. Med. Chem. Lett.* **2004**, *14*, 5925.
12. The crude product was purified by flash chromatography using hexane/AcOEt (1:1) and identified by ^1H NMR (400 MHz, $CDCl_3$): δ 0.89 (d, $J = 6.6$ Hz, 3H), 0.95 (d, $J = 6.6$ Hz, 3H), 1.22–1.32 (m, 2H), 1.45 (s, 9H), 1.58–1.76 (br m, 2H), 1.80–2.04 (m, 3H), 2.12–2.22 (m, 1H), 2.60–2.73 (m, 3H), 3.09–3.19 (m, 1H), 3.29–3.35 (m, 1H), 3.40–3.52 (br m, 1H), 3.58 (br s, 1H), 3.65 (br s, 1H), 4.62–4.68 (d, $J = 5.7$ Hz, 2H), 4.73 (br d, $J = 9.0$ Hz, 1H), 5.20–5.30 (dd, $J = 9.3$ Hz, $J = 1.2$ Hz, 1H), 5.30–5.38 (dd, $J = 15.7$ Hz, $J = 1.5$ Hz, 1H), 5.88–5.98 (m, 1H). The other intermediates were similarly purified and characterized.
13. Bang, J. K.; Hasegawa, K.; Kawakami, T.; Aimoto, S.; Akaji, K. *Tetrahedron Lett.* **2004**, *45*, 99.
14. Compound **I-1**: *m/z* 883.03 for $[M+H]^+$ (calcd 882.44 for $C_{41}H_{69}N_7O_6I$). Compound **I-2**: *m/z* 866.52 for $[M+H]^+$ (calcd 866.40 for $C_{40}H_{65}N_7O_6I$). Compound **I-3**: *m/z* 769.09 for $[M+H]^+$ (calcd 769.35 for $C_{35}H_{58}N_6O_5I$). Compound **I-4**: *m/z* 785.33 for $[M+H]^+$ (calcd 785.38 for $C_{36}H_{62}N_6O_5I$). Compound **I-5**: *m/z* 672.01 for $[M+H]^+$ (calcd 672.30 for $C_{30}H_{51}N_5O_4I$). Compound **I-6**: *m/z* 573.58 for $[M+H]^+$ (calcd 573.23 for $C_{25}H_{42}N_4O_3I$).