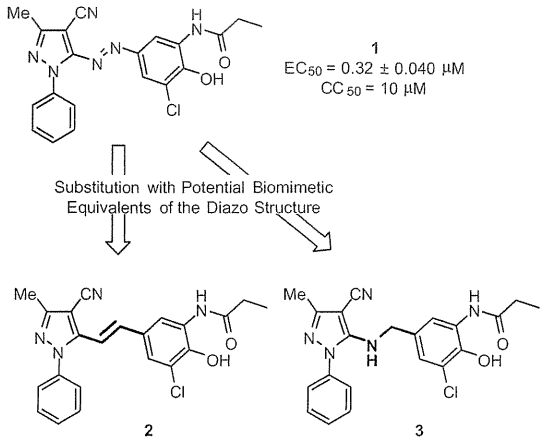
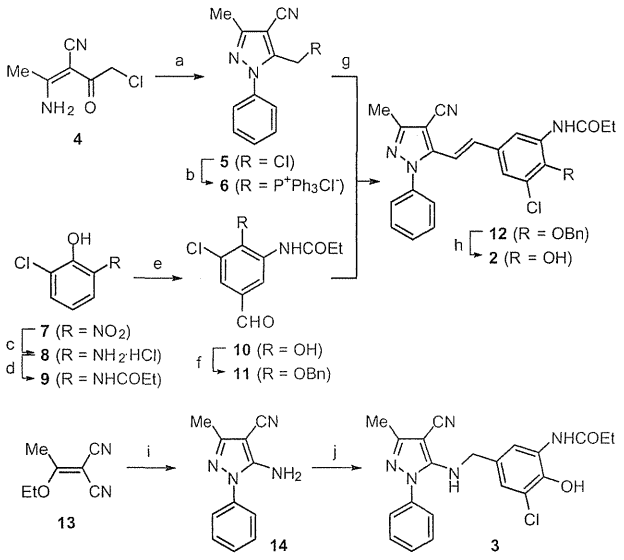


Figure 1. Structures of lead compound **1** and the designed derivatives **2** and **3**.



Scheme 1. Synthesis of compounds **2** and **3**. Reagents and conditions: (a) Ph-$NHNH_2$, AcOH, $MeOCH_2CH_2OH$, 105 °C; (b) PPh_3, toluene, 110 °C; (c) $FeCl_3$, $NH_2NH_2{\cdot}H_2O$, activated carbon, *i*-PrOH, 80 °C, then 4 M-HCl/1,4-dioxane; (d) EtCOCl, pyridine, THF, rt; (e) hexamethylenetetramine, AcOH, 130 °C; (f) BnBr, K_2CO_3, $CHCl_3$/MeOH, 60 °C; (g) *n*-BuLi, THF, 0 °C, then **11**, THF, rt; (h) TMSI, MeCN, 80 °C; (i) Ph-$NHNH_2$, EtOH, rt; (j) **10**, PPTS, toluene, reflux, then $LiBH_4$, THF, 0 °C to rt.

Compound **3** was also synthesized by using benzaldehyde **10**. The treatment of phenylhydrazine with acrylonitrile **13** gave 5-aminopyrazole derivative **14**. The reductive amination of aldehydes **10** with **14** provided the desired aminomethylene derivative **3**.

With two derivatives having potentially equivalent linkers available, we then began our investigation for the structure–activity relationships of lead compound **1** for anti-HIV activity (Table 1). The bioevaluation of compounds **2** and **3** by the NCK assay[10] revealed that styrene derivative **2** and aminomethylene derivative **3** had comparable anti-HIV activity to the original compound **1** (EC_{50} = 0.44 and 0.36 μM, respectively). These results encouraged us to extend the structural optimization from the potent aminomethylene compound **3**.

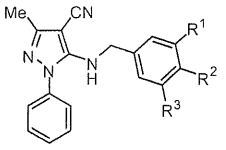
First, to investigate the structural requirements of the right 3-chloro-4-hydroxy-5-propionamidobenzylamine part in compound **3**, a variety of derivatives **24a–h** with modification on the benzene ring were designed and synthesized (Table 2). The synthesis of

Table 1
Structure–activity relationship of the linkage part

Compound	EC_{50}[a] (μM)
1	0.32 ± 0.04
2	0.44 ± 0.12
3	0.36 ± 0.11

[a] EC_{50} values represent the concentration of compound required to inhibit the HIV infection by 50%. The data were obtained from three independent experiments.

Table 2
Structure–activity relationships of the benzylamine moiety

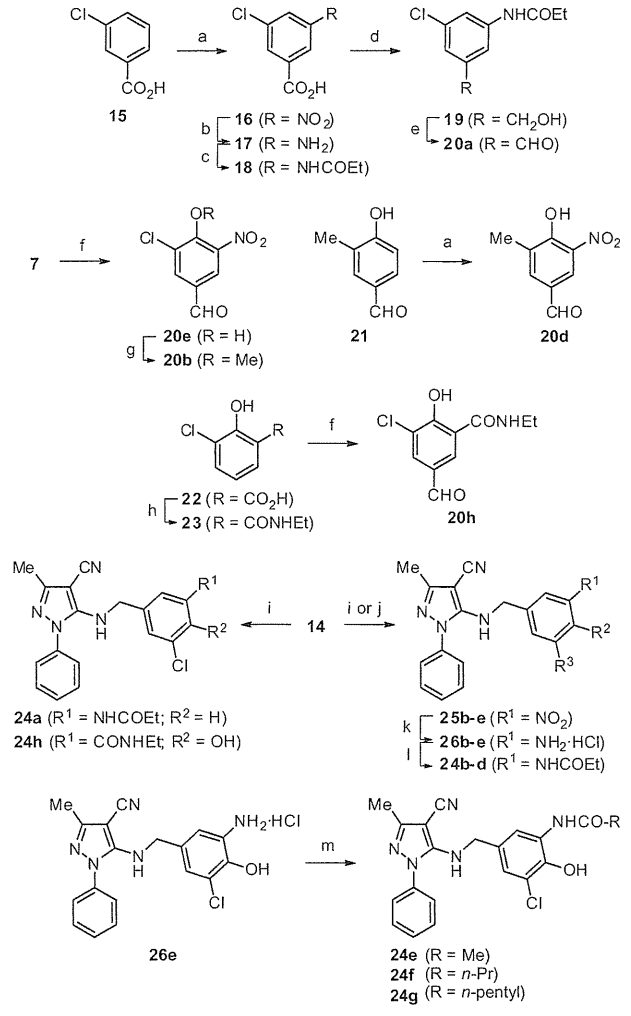


Compound	R^1	R^2	R^3	EC_{50}[a] (μM)
3	NHCOEt	OH	Cl	0.36 ± 0.11
24a	NHCOEt	H	Cl	>10
24b	NHCOEt	OMe	Cl	>10
24c	NHCOEt	OH	H	>10
24d	NHCOEt	OH	Me	>10
24e	NHCOMe	OH	Cl	>1.0[b]
24f	NHCO-*n*-Pr	OH	Cl	3.3 ± 0.89
24g	NHCO-*n*-pentyl	OH	Cl	>10
24h	CONHEt	OH	Cl	>1.0[b]

[a] EC_{50} values represent the concentration of the compound required to inhibit the HIV-1 infection by 50%. The data were obtained from three independent experiments.
[b] Cytotoxicity was observed at 10 μM.

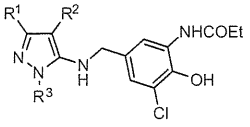
these derivatives **24a–h** started with the synthesis of various benzaldehydes **20** (Scheme 2). Briefly, nitration of 3-chlorobenzoic acid **15** followed by subsequent functional group transformations including Fe-mediated reduction, acylation, reduction of the carboxylic group and PCC oxidation gave aldehyde **20a**. Formylation of **7** using hexamethylenetetramine gave a phenol **20e**, which was subjected to subsequent O-methylation to give a methoxybenzaldehyde **20b**. Nitration of 4-hydroxy-3-methyl-benzaldehyde **21** gave compound **20d**. Condensation of a salicylic acid **22** with ethylamine followed by formylation furnished aldehyde **20h**. Reactions of benzaldehydes **20a–e**[11] and **20h** with 5-aminopyrazole **14** gave benzylamine derivatives **24a,h** and **25b–e**. $FeCl_3$-mediated reduction of benzylamines **25b–e** followed by acylation afforded derivatives **24b–g**.

Removal of the 4-hydroxy or 3-chloro group on the benzene ring resulted in a loss of anti-HIV activity (**24a** and **24c**, EC_{50} >10 - μM) (Table 2). Substitution with either a 4-methoxy or 3-methyl group also led to loss of anti-HIV activity (**24b** and **24d**, EC_{50} >10 - μM), indicating that the 4-hydroxy group and the 3-chloro group are essential functional groups for anti-HIV activity. In the study of the *N*-acyl group at the 5-propionamido moiety, a significant decrease in anti-HIV activity was observed by substitution of the 5-propionamido group with the shorter acetamido group (**24e**, EC_{50} >1.0 μM), and the longer butyramido (**24f**, EC_{50} = 3.30 μM) and hexanamido groups (**24g**) (EC_{50} >10 μM). The retroamide transformation of the propionamido group with the ethylcarbamoyl group did not improve anti-HIV activity (**24h**, EC_{50} >1.0 μM). Of note, compounds **24e** and **24h** showed cytotoxicity at 10 μM.



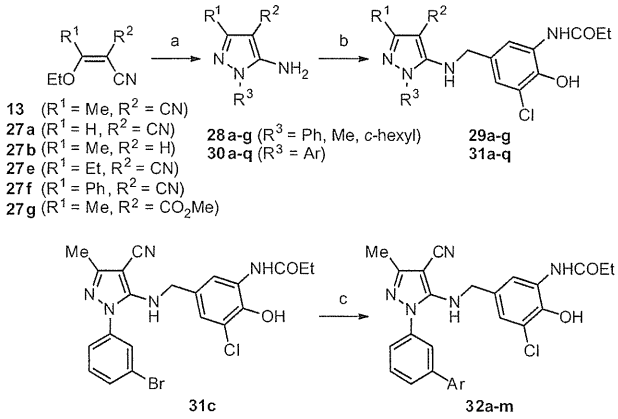
Scheme 2. Synthesis of compounds **24a–h** with a variety of substituted benzylamine moieties. Reagents and conditions: (a) H_2SO_4, HNO_3, 0 °C to rt; (b) Fe, NH_4Cl, *i*-PrOH, H_2O, 90 °C; (c) EtCOCl, Et_3N, THF, 0 °C to rt; (d) EtOCOCl, Et_3N, THF, −10 °C, then $NaBH_4$, MeOH, 0 °C to rt; (e) PCC, CH_2Cl_2, rt; (f) hexamethylenetetramine, AcOH, 130 °C; (g) MeI, K_2CO_3, DMF, rt; (h) CBr_4, Et_3N, $EtNH_2$, PPh_3, CH_2Cl_2, rt; (i) **20**, PPTS, toluene, reflux, then $LiBH_4$, THF, 0 °C to rt; (j) **20**, TsOH, toluene, reflux, then $NaBH_4$, EtOH, 0 °C to rt; (k) $FeCl_3$, $NH_2NH_2{\cdot}H_2O$, activated carbon, *i*-PrOH, 80 °C, then 4 M-HCl/1,4-dioxane; (l) $EtCO_2H$, EDC, HOBt, $(i\text{-}Pr)_2NEt$, DMF, 45 °C; (m) RCO_2H, EDC, HOBt, $(i\text{-}Pr)_2NEt$, DMF, 45 °C.

Table 3
Structure–activity relationship of the pyrazole moiety



Compound	R^1	R^2	R^3	EC_{50}[a] (μM)
3	Me	CN	Ph	0.36 ± 0.11
29a	H	CN	Ph	>1.0[b]
29b	Me	H	Ph	>10
29c	Me	CN	Me	>10
29d	Me	CN	*c*-hexyl	>1.0[b]
29e	Et	CN	Ph	>10
29f	Ph	CN	Ph	>10
29g	Me	CO_2Me	Ph	>1.0[b]

[a] EC_{50} values represent the concentration of the compound required to inhibit the HIV-1 infection by 50%. The data were obtained from three independent experiments.
[b] Cytotoxicity was observed at 10 μM.



Scheme 3. Synthesis of compounds **29a–g**, **31a–q**, and **32a–m** with a variety of *N*-arylpyrazole and *N*-alkylpyrazole moieties. Reagents and conditions: (a) R^3-$NHNH_2$, EtOH, rt; (b) **10**, PPTS, toluene, reflux, then $LiBH_4$, THF, 0 °C to rt; (c) Ar-$B(OH)_2$ or Ar-Bpin, $Pd(dba)_2$, PCy_3, K_3PO_4, 1,4-dioxane, H_2O, 80 °C.

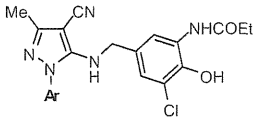
We next examined substituent effects on the left pyrazole part of compound **3** (Tables 3–5). A variety of 5-aminopyrazoles **28a–g** and **30a–q** were prepared by reaction of 3-ethoxyacrylonitrile derivatives (**13** and **27**) with aryl- or alkyl-hydrazines (Scheme 3). Reductive amination of **28a–g** and **30a–q** with benzaldehyde **10** provided the series of pyrazole derivatives **29a–g** and **31a–q**. 1-Biaryl-5-aminopyrazole derivatives **32a–m** were synthesized using a Suzuki–Miyaura cross-coupling reaction of compound **31c** with aryl boronic acid or the pinacol ester.

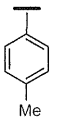
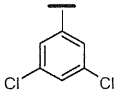
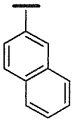
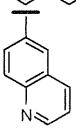
Removal of the 3-methyl or 4-cyano group on the pyrazole **3** reduced anti-HIV activity (**29a** and **29b**, EC_{50} >1.0 μM), as shown in Table 3. In addition, no potent derivatives were obtained by all of our attempts including substitutions of the *N*-phenyl group with alkyl groups (**29c** and **29d**), substitutions of the 3-methyl group with either an ethyl (**29e**) or phenyl group (**29f**), and substitution of the 4-cyano group with a methoxycarbonyl group (**29g**).

We turned to examine the structure–activity relationship of the *N*-aryl moiety on the pyrazole (Table 4). The modification at the *meta*-position of the *N*-phenyl ring with methyl (**31b**), fluoro (**31d**), or phenyl groups (**31g**) resulted in comparable or more potent anti-HIV activity (EC_{50} = 0.37, 0.33 and 0.13 μM, respectively) to that of compound **3**, whereas a significant decrease in anti-HIV activity was observed following the introduction of a methoxy (**31a**), nitro (**31f**), or acetamido group (**31h**). Anti-HIV activity of bromo (**31c**) or trifluoromethyl (**31e**) derivatives could not be determined because of their significant cytotoxicity. The methyl modification at the *ortho*- and *para*-positions (**31i** and **31j**) afforded reduced anti-HIV activities (EC_{50} >1.0 μM). Compound **31k** with a 3,5-dichlorophenyl group showed no antiviral activity (EC_{50} >1.0 - μM). In addition, replacement of the phenyl group with naphthyl (**31l** and **31m**), 3-pyridyl (**31o**), or quinolinyl group (**31p** and **31q**) failed to improve anti-HIV activity (EC_{50} >1.0 μM). 2-Pyridyl derivative **31n** showed unexpected levels of cytotoxicity.

The optimization studies indicated that the introduction of a biphenyl-3-yl group on the pyrazole moiety (**31g**) effectively improved the anti-HIV activity. To develop more potent anti-HIV agents, we extended the structure–activity relationship investigations to the modified aryl groups onto the *N*-phenylpyrazole moiety (Table 5). We designed and synthesized a variety of methoxyphenyl, tolyl or chlorophenyl derivatives (**32a–i**). Among

Table 4
Structure–activity relationship of the *N*-arylpyrazole moiety

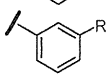
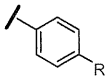
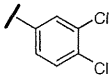
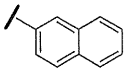
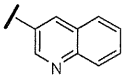


Compound	Ar		EC_{50}[a] (μM)
3		R = H	0.36 ± 0.11
31a		R = OMe	>10
31b		R = Me	0.37 ± 0.11[b]
31c		R = Br	>0.1[c]
31d		R = F	0.33 ± 0.12[b]
31e		R = CF_3	>0.1[c]
31f		R = NO_2	>1.0[b]
31g		R = Ph	0.13 ± 0.03[c]
31h		R = $NHCOCH_3$	>10
31i			>1.0[b]
31j			>1.0[b]
31k			>1.0[b]
31l			>1.0[b]
31m			>1.0[b]
31n			>0.1[c]
31o			>10
31p			>1.0[b]
31q			>10

[a] EC_{50} values represent the concentration of the compound required to inhibit the HIV-1 infection by 50%. The data were obtained from three independent experiments.
[b] Cytotoxicity was observed at 10 μM.
[c] Cytotoxicity was observed at 1.0 μM.

these, significantly improved anti-HIV activity was observed in 3-tolyl (**32e**) and 4-tolyl (**32h**) derivatives (EC_{50} = 0.058 and 0.097 μM, respectively). Additionally, the 3,4-dichloro derivative (**32j**) also exhibited more potent anti-HIV activity when compared with compound **31g** (EC_{50} = 0.047 μM). In contrast, naphthyl (**32k**), pyridyl (**32l**), and quinolinyl (**32m**) groups showed lower anti-HIV activity. Of note, three potent derivatives (**32e**, **32h** and **32j**) exhibited moderate cytotoxicity at 1.0 μM; therefore, improvement in the selectivity index is needed by further optimizations.

Table 5
Structure–activity relationship of the *N*-biarylpyrazole moiety

Compound	Ar		EC_{50}[a] (μM)
32a		R = OMe	>0.1[c]
32b		R = Me	>0.1[c]
32c		R = Cl	0.25 ± 0.13[b]
32d		R = OMe	>0.1[c]
32e		R = Me	0.058 ± 0.007[c]
32f		R = Cl	>0.1[c]
32g		R = OMe	>0.1[c]
32h		R = Me	0.097 ± 0.040[c]
32i		R = Cl	>0.1[c]
32j			0.047 ± 0.011[c]
32k			>0.1[c]
32l			>1.0[b]
32m			>0.1[c]

[a] EC_{50} values represent the concentration of the compound required to inhibit the HIV-1 infection by 50%. The data were obtained from three independent experiments.
[b] Cytotoxicity was observed at 10 μM.
[c] Cytotoxicity was observed at 1.0 μM.

In conclusion, we have designed and synthesized a series of phenylpyrazole derivatives for the development of novel anti-HIV agents. The structure–activity relationship study demonstrated that the aminomethylene group can be used as an appropriate isosteric unit of the diazo linkage in the lead compound **1**. The substituents on the right phenol part and left pyrazole ring are essential for the potent anti-HIV activity. The optimization study on *N*-phenylpyrazoles led to the identification of a potent 3′,4′-dichloro-(1,1′-biphenyl)-3-yl derivative **32j**[12] which exhibited six-fold more potent anti-HIV activity when compared with compound **1**. Although the target molecules remain unclear, further optimization of these compounds to improve the potency and selectivity should provide a therapeutic approach to HIV infection.

Acknowledgements

This work was supported by Health and Labor Science Research Grants (Research on HIV/AIDS, Japan); Grants-in-Aid for Scientific Research and Platform for Drug Discovery, Informatics, and Structural Life Science from MEXT, Japan. We thank Ms. Kumiko Hiyama and Dr. Kazunobu Takahashi for their technical supports. T.M. and T.T. are grateful for JSPS Research Fellowships for Young Scientists and Strategic Young Researcher Overseas Visits Program for Accelerating Brain Circulation.

References and notes

1. Weiss, R. A. *Science* **1993**, *260*, 1273.
2. UNAIDS Report on the Global AIDS Epidemic 2012: http://www.unaids.org/en/media/unaids/contentassets/documents/epidemiology/2012/gr2012/20121120_UNAIDS_Global_Report_2012_en.pdf.
3. (a) Esté, J. A.; Cihlar, T. *Antiviral Res.* **2010**, *85*, 25; (b) Mehellou, Y.; De Clercq, E. *J. Med. Chem.* **2010**, *53*, 521. and references cited therein.

4. (a) Thompson, M. A.; Aberg, J. A.; Cahn, P.; Montaner, J. S. G.; Rizzardini, G.; Telenti, A.; Gatell, J. M.; Günthard, H. F.; Hammer, S. M.; Hirsch, M. S.; Jacobsen, D. M.; Reiss, P.; Richman, D. D.; Volberding, P. A.; Yeni, P.; Schooley, R. T. *J. Am. Med. Assoc.* **2010**, *304*, 321; (b) Rathbun, R. C.; Lockhart, S. M.; Stephens, J. R. *Curr. Pharm. Des.* **2006**, *1045*, 12.
5. Johnson, V. A.; Calvez, V.; Günthard, H. F.; Paredes, R.; Pillay, D.; Shafer, R.; Weinsing, A. M.; Richman, D. D. *Top. Antivir. Med.* **2011**, *19*, 156.
6. Hawkins, T. *Antiviral Res.* **2010**, *85*, 201.
7. (a) Fujii, N.; Oishi, S.; Hiramatsu, K.; Araki, T.; Ueda, S.; Tamamura, H.; Otaka, A.; Kusano, S.; Terakubo, S.; Nakashima, H.; Broach, J. A.; Trent, J. O.; Wang, Z.-X.; Peiper, S. C. *Angew. Chem., Int. Ed.* **2003**, *42*, 3251; (b) Ueda, S.; Oishi, S.; Wang, Z.-X.; Araki, T.; Tamamura, H.; Cluzeau, J.; Ohno, H.; Kusano, S.; Nakashima, H.; Trent, J. O.; Peiper, S. C.; Fujii, N. *J. Med. Chem.* **2007**, *50*, 192; (c) Inokuchi, E.; Oishi, S.; Kubo, T.; Ohno, H.; Shimura, K.; Matsuoka, M.; Fujii, N. *ACS Med. Chem. Lett.* **2011**, *2*, 477; (d) Mizuhara, T.; Oishi, S.; Ohno, H.; Shimura, K.; Matsuoka, M.; Fujii, N. *Org. Biomol. Chem.* **2012**, *10*, 6792; (e) Mizuhara, T.; Oishi, S.; Ohno, H.; Shimura, K.; Matsuoka, M.; Fujii, N. *Bioorg. Med. Chem.* **2012**, *20*, 6434.
8. (a) Peppercorn, M. A.; Goldman, P. *J. Pharmacol. Exp. Ther.* **1972**, *181*, 555; (b) Scheline, R. R. *Pharmacol. Rev.* **1973**, *25*, 451.
9. (a) Duff, J. C.; Bills, E. J. *J. Chem. Soc.* **1932**, 1987; (b) Ferguson, L. N. *Chem. Rev.* **1946**, *38*, 227.
10. (a) Kajiwara, K.; Kodama, E.; Matsuoka, M. *Antivir. Chem. Chemother.* **2006**, *17*, 215; (b) Kajiwara, K.; Kodama, E.; Sakagami, Y.; Naito, T.; Matsuoka, M. *J. Clin. Microbiol.* **2008**, *46*, 792.
11. 4-Hydroxy-3-nitrobenzaldehyde **20c** is commercially available.
12. *Compound **32j**:* ^{1}H NMR (400 MHz, $CDCl_3$) δ: 1.27–1.24 (m, 4H), 2.33 (s, 3H), 2.49–2.42 (m, 2H), 4.55 (d, J = 5.9 Hz, 2H), 4.64 (t, J = 5.9 Hz, 1H), 7.08 (d, J = 2.0 Hz, 1H), 7.74–7.35 (m, 9H). HRMS (FAB): m/z calcd for $C_{27}H_{23}Cl_3N_5O_2$ $[M+H]^+$ 554.0917; found: 554.0916.

Journal of General Virology (2013), 94, 933–943 DOI 10.1099/vir.0.047167-0

Impact of antiretroviral pressure on selection of primary human immunodeficiency virus type 1 envelope sequences *in vitro*

Shigeyoshi Harada,[1,2] Kazuhisa Yoshimura,[1,2] Aki Yamaguchi,[1] Samatchaya Boonchawalit,[1,2] Keisuke Yusa[3] and Shuzo Matsushita[1]

Correspondence
Kazuhisa Yoshimura
ykazu@nih.go.jp

[1]Center for AIDS Research, Kumamoto University, 2-2-1 Honjo, Chuo-ku, Kumamoto 860-0811, Japan

[2]AIDS Research Center, National Institute of Infectious Diseases, 1-23-1 Toyama, Shinjuku-ku, Tokyo 162-8640, Japan

[3]Division of Biological Chemistry and Biologicals, National Institute of Health Sciences, 1-18-1 Kami-youga, Setagaya-ku, Tokyo 158-8501, Japan

The initiation of drug therapy results in a reduction in the human immunodeficiency virus type 1 (HIV-1) population, which represents a potential genetic bottleneck. The effect of this drug-induced genetic bottleneck on the population dynamics of the envelope (Env) regions has been addressed in several *in vivo* studies. However, it is difficult to investigate the effect on the *env* gene of the genetic bottleneck induced not only by entry inhibitors but also by non-entry inhibitors, particularly *in vivo*. Therefore, this study used an *in vitro* selection system using unique bulk primary isolates established in the laboratory to observe the effects of the antiretroviral drug-induced bottleneck on the integrase and *env* genes. Env diversity was decreased significantly in one primary isolate [KP-1, harbouring both CXCR4 (X4)- and CCR5 (R5)-tropic variants] when passaged in the presence or absence of raltegravir (RAL) during *in vitro* selection. Furthermore, the RAL-selected KP-1 variant had a completely different Env sequence from that in the passage control (particularly evident in the gp120, V1/V2 and V4-loop regions), and a different number of potential *N*-glycosylation sites. A similar pattern was also observed in other primary isolates when using different classes of drugs. This is the first study to explore the influence of anti-HIV drugs on bottlenecks in bulk primary HIV isolates with highly diverse Env sequences using *in vitro* selection.

Received 15 August 2012
Accepted 20 December 2012

INTRODUCTION

Human immunodeficiency virus type 1 (HIV-1) shows a high degree of genetic diversity owing to its high rates of replication and recombination and the high mutation rate of the HIV-1 reverse transcriptase (Nájera *et al.*, 2002). Even in a single infected individual, the virus can best be described as a population of distinct, but closely related, genetic variants or 'quasi-species' (Eigen, 1993; Nijhuis *et al.*, 1998). The quasi-species behaviour of viruses is recognized as a key element in our understanding and modelling of viral evolution and disease control (Vignuzzi *et al.*, 2006).

Combination antiretroviral (ARV) therapy results in a contraction of the viral population, which represents a potential genetic bottleneck (Charpentier *et al.*, 2006; Delwart *et al.*, 1998; Ibáñez *et al.*, 2000; Kitrinos *et al.*, 2005; Nijhuis *et al.*, 1998; Nora *et al.*, 2007; Sheehy *et al.*, 1996; Zhang *et al.*, 1994). Whilst this bottleneck has a direct effect on the region that is being targeted by the drugs (e.g. protease or reverse transcriptase), it also affects other regions of the viral genome. Indeed, the effect of the drug-induced genetic bottleneck on the population dynamics of the envelope (Env) regions has been addressed in several *in vivo* studies (Charpentier *et al.*, 2006; Delwart *et al.*, 1998; Ibáñez *et al.*, 2000; Kitrinos *et al.*, 2005; Nijhuis *et al.*, 1998; Nora *et al.*, 2007; Sheehy *et al.*, 1996; Zhang *et al.*, 1994).

Virus bottleneck evolution of the HIV-1 *env* gene might be important when choosing the optimal drugs to treat a particular patient. Indeed, a CCR5 antagonist (maraviroc, MVC) and a fusion inhibitor (enfuvirtide, T-20) have now

The GenBank/EMBL/DDBJ accession numbers for the *env* sequences of HIV-1 KP-1, KP-2 and KP-4, are AB640872–AB640881, AB641341–AB641351 and AB641335–AB641340, respectively.

Two supplementary figures are available with the online version of this paper.

been approved for use as HIV-1 entry inhibitors. Analysing the dynamics of drug-induced genetic bottlenecks and studying drug-resistant mutation profiles in response to HIV-1-specific ARV drugs are both important if we are to understand fully HIV-1 drug resistance and pathogenesis.

The aim of the present study was to understand better the effect of *in vivo* drug-induced genetic bottlenecks. *In vitro* selection of different primary HIV-1 isolates was performed using the recently approved HIV integrase inhibitor raltegravir (RAL) (Steigbigel *et al.*, 2008). Two R5-, one X4-, one dual- and one mixed R5/X4-tropic isolates were passaged through a RAL-induced genetic bottleneck. We also performed *in vitro* selection of the R5/X4 isolate using lamivudine (3TC), saquinavir (SQV) and MVC, and compared the results with those from the RAL-selected isolate.

RESULTS

Genotypic profiles of the HIV-1 primary isolates

Four genetically heterogeneous HIV-1 primary isolates (KP-1–4) from Japanese drug-naïve patients were used to assess the extent to which RAL affected the selection of bulk primary viruses *in vitro*. A laboratory isolate, strain 89.6, was also used in the study (rather than a molecular clone) to allow escape mutants to be selected from each quasi-species pool and to be generated *de novo*. First, the sequences of the integrase (IN) regions of the four primary isolates were determined. Table 1 shows the detailed evaluation of the R5/X4 mixture subtype B (KP-1), R5-CRF08_BC (KP-2), R5 subtype B (KP-3) and X4-CRF01_AE (KP-4) primary isolates, and the dual-tropic subtype B laboratory virus (89.6). Although some naturally occurring polymorphisms were observed within the IN regions of these isolates compared with the subtype B consensus sequence available from the Los Alamos National Laboratory HIV sequence database, we did not identify any primary resistant mutations to RAL. Three baseline viruses (KP-1, KP-4 and 89.6) were sensitive to RAL, with IC_{50} values ranging from 1.2 to 4 nM, which are comparable with those reported previously (Kobayashi *et al.*, 2008). However, KP-2 and KP-3 showed minor resistance to RAL, with IC_{50} values of 16 and 32 nM, respectively. These two isolates contained amino acid mutations at positions 72, 125 and 201 within the IN region [previously reported as L-870,810 and S-1360 resistance mutations (Hombrouck *et al.*, 2008; Rhee *et al.*, 2008), but not as RAL-resistance mutations]. KP-2 also contained a unique insertion at position 288 (NQDME) at the C-terminal end of the IN region.

In vitro selection of variants of the primary isolates and 89.6 using RAL

To induce RAL-selected HIV-1 variants *in vitro*, PM1/CCR5 cells, a T-cell line expressing high levels of CCR5, were exposed to the four primary isolates and strain 89.6. The viruses were then serially passaged in the presence of RAL. As a control, each isolate was passaged under the same conditions, but without RAL, to allow monitoring of spontaneous changes occurring in the viruses during prolonged PM1/CCR5 cell passage (the passage control). The selected viruses were initially propagated at a RAL concentration equal to each IC_{50} value. The RAL concentrations were then increased from 20 to 85 nM during the course of the selection procedure (Table 1).

Only small shifts in the IC_{50} to RAL were observed in four of the five isolates (KP-1, KP-2, KP-4 and 89.6), with fold changes in IC_{50} values of 3.4, 6.5, 16 and 9.2, respectively. KP-3 did not show resistance to RAL. IC_{50} values in all the passage controls were comparable with those of the baseline viruses (Table 1).

IN region sequences in RAL-selected variants

The full-length *IN* genes were amplified and cloned to determine the genetic basis of selection in the presence or absence of RAL. Ten to 12 clones from each sample were sequenced.

Substitutions within IN were observed at passages 30 (G189R) and 29 (T210I) in two RAL-selected isolates (KP-2 and KP-4, respectively). Neither of these has been reported as IN inhibitor-resistant mutations. No substitutions in the IN regions of KP-3 and 89.6 were found. However, A125T and V180I substitutions were observed in the KP-3 and 89.6 control variants at the last passage. No previously reported mutations were identified in the IN region of KP-1 (an R5/X4 mixture isolate) after 17 passages. However, four amino acids (K7/K111/H216/D278) were selected by RAL from the baseline quasi-species, whereas different amino acids (R7/R111/Q216/N278) were selected in the control-passage variants (Table 1).

Taken together, these findings showed that RAL-induced selection pressure causes adaptation within the IN regions of bulk primary viruses during *in vitro* passage in the target cells, and confirmed that this system can be used to analyse drug-selected variants *in vitro*.

Comparison of *env* gene sequences in RAL-selected and passage-control isolates

A highly diverse gp120 region was observed in the baseline R5/X4 mixture isolate, KP-1; however, the viral diversity of variants passaged in the presence or absence of RAL decreased significantly during *in vitro* selection (overall mean distance after RAL selection of 0.056 at baseline to 0.007 after passage 17; mean overall distance in the passage control of 0.01 after 20 passages, Table 2). Moreover, the RAL-selected and control variants utilized CCR5 to enter the target cell; neither variant used CXCR4 (Table 3).

Interestingly, the low-diversity RAL-selected variant contained a completely different Env sequence from that of the passage-control variant (Fig. 1a). Different regions spanning

Table 1. Susceptibility of HIV-1 isolates to RAL and distinct differences in IN region sequences between RAL-selected and control-passaged viruses

Isolate	Subtype	Tropism	Passage no.	Concn (nM)	RAL-selected variant*		Passage control	
					IN sequence	RAL IC_{50} (nM)	IN sequence	RAL IC_{50} (nM)
KP-1	B	Mix	0	0	*K/R7, K/R111, Q/H216, D/N278*	4	*K/R7, K/R111, Q/H216, D/N278*	4
			8	20	**K**111, **H**216, **D**278	31 (7.8)	**R**7, **R**111, **Q**216, **N**278	4.5 (1.2)
			17†	20	**K**7, **K**111, **H**216, **D**278	26 (6.5)	**R**7, **R**111, **Q**216, **N**278	0.4 (0.1)
KP-2	CRF08_BC	R5	0	0	*I201, ins289NQDME*	16	*I201, ins289NQDME*	16
			18	40	G189G/R, *I201, ins289NQDME*	32 (2)	*I201, ins289NQDME*	16 (1)
			30	85	G189R, *I201, ins289NQDME*	55 (3.4)	*I201, ins289NQDME*	25 (1.6)
KP-3	B	R5	0	0	*V72, A125*	32	*V72, A125*	32
			11	25	*V72, A125*	25 (0.78)	*V72, A125*	33 (1)
			22	27.5	*V72, A125*	37 (1.2)	*V72,* A125T	13 (0.41)
KP-4	CRF01_AE	X4	0	0	–	2.1	–	2.1
			8	40	–	33 (16)	R166R/K, D279N	4.4 (2.1)
			29	40	T210I	22 (10)	G163E, R166R/K, D279N/S	4.1 (2)
89.6	B	R5X4	0	0	–	1.2	–	1.2
			8	15	–	34 (28)	–	4.4 (3.7)
			34	20	–	11 (9.2)	V180I	1.2 (1)

*Amino acid changes in each passage variant are shown. Italicized letters represent mutations relative to the consensus subtype BC or B present in the baseline isolates. Bold letters represent amino acids selected out of the quasi-species cloud. The fold increase in RAL IC_{50} values is shown in parentheses for *in vitro*-selected variants compared with those in the baseline isolates.
†The RAL variant selected after 17 passages was compared with the control selected after 20 passages.

Table 2. Comparison of amino acid length and number of PNGs between RAL-selected and control-passage KP-1 variants

	Passage no.	Genetic diversity*	Mean ENV 1–474 length (range)†	Mean V1/V2 length (range)	Mean V3 length (range)	Mean V4 length (range)	Mean PNGs (range)
Baseline	0	0.056	472 (461–480)	69 (60–74)	34 (33–34)	30 (29–31)	24 (22–28)
RAL-selected virus	2	0.038	479 (472–480)‡	74 (71–74)‡	34 (33–34)‡	31 (29–31)‡	27 (25–28) ‡
	8	0.0070	480	74	34	31	28 (26–29)
	17	0.0070	480	74	34	31	27 (26–27)§
Passage control	2	0.045	464 (461–466)‡	64 (60–74)‡	34 (33–34)‡	29 (29–31)‡	24 (22–27)‡
	8	0.0070	463 (462–463)	62	34	29	23 (22–23)
	10	0.0080	462 (459–463)	62	34	29	23 (22–23)
	20	0.010	463	62	34	29	23 (22–23)§
P value			<0.0001‡	<0.0001‡	0.91‡	0.0048‡	0.0019‡ <0.0001§

*Overall mean distance.
†Sequence from gp120 SP to the V5 region (aa 1–474).
‡, § *P* values were calculated using the homoscedastic *t*-test between the RAL-selected and the passage-control variants indicated by the same symbols above.

the whole envelope sequence [from the signal peptide (SP) to V5] were compared in the RAL-selected and passage-control viruses. The results showed that, after only two passages, the gp120, V1/V2 and V4-loop regions within RAL-selected variants were longer than those in the control variants, and the number of putative *N*-linked glycosylation sites (PNGs) was significantly higher than that in the control-passage viruses (Table 2). This phenomenon was seen consistently in two independent experiments.

We also analysed the gp120 sequences in the other four isolates. Although the number of positional differences between the RAL-selected and passage-control variants for these four isolates was lower than that in KP-1 (between three and nine, compared with >40), there was a similar pattern of separation between the Env sequences (Fig. 1). In three of the four isolates (KP-2, KP-3 and KP-4), positional differences were observed in SP, C1 and all the variable regions of gp120 (Fig. 1b–d). In strain 89.6, differences were observed in the C2, C3 and V4 regions (Fig. 1e).

These results suggested that RAL treatment of target cells causes a decrease in viral diversification within quasi-species Env regions via a route different from that in untreated target cells.

In vitro induction of RAL-selected V3-loop library virus variants

To investigate further the effects of RAL on viral Env sequences, we used the V3-loop library virus (JR-FL-V3Lib) developed by Yusa *et al.* (2005), which carries a set of random combinations from zero to ten substitutions (27 648 possibilities) in the V3 loop (residues 305, 306, 307, 308, 309, 317, 319, 322, 323 and 326; V3 loop from Cys^{296} to Cys^{331}). The variants contained in the library were polymorphic mutations derived from 31 R5 clinical isolates (Yusa *et al.*, 2005). PM1/CCR5 cells were exposed to the JR-FL-V3Lib and serially passaged in the presence of RAL. After two passages, the V3 sequence within the RAL-selected variant was completely different from that in the passage control (Fig. 1f). This suggested that, under pressure from RAL, the infectious clone harbouring different V3 region sequence from the passage control had adapted to the target cells, despite containing the same IN sequences.

Phylogenetic analysis of the Env regions after passage with or without RAL

To confirm the temporal and spatial differences observed in each of the RAL-selected and passage-control viruses, phylogenetic analyses were conducted using complete SP–V5 sequences. The neighbour-joining phylogenetic tree showed a clear and distinct branching between RAL-selected and passage-control KP-1 viruses (Fig. 2a). We also identified a similar pattern in all the other isolates tested (Fig. 2b–e).

Table 3. Comparison of amino acid length, number of potential *N*-linked glycosylation sites, V3 sequences and co-receptor usage between anti-retroviral drug-selected and control-passaged KP-1 variants

	Passage no.	Genetic diversity*	Mean ENV_{1-474} length (range)†	Mean V1/V2 length (range)	Mean V3 length (range)	Mean V4 length (range)	Mean PNGs (range)	V3 region		Geno2 pheno (%)§
								Prevalence (%)	Sequence‡	
Baseline	0	0.056	472 (461–480)	69 (60–74)	34 (33–34)	30 (29–31)	24 (22–28)	41.9	CTRPNNNTRKGIHIGPGKFYATGAIIGDIRQAHC	41.2
								22.6	.....V.....	41.2
								16.1	....-..I.....T.R..T.RD...N..K...	1.7
								13.0	....-..I.....T.R..T.KT...N.KK...	2.9
								3.2	....-..I.....	7.4
								3.2	.....D.....	55.3
Passage control	8	0.0070	463 (462–463)	62	34	29	23 (22–23)	100.0	.....V.....	41.2
RAL-selected virus	8	0.0070	480	74	34	31	28 (26–29)	100.0	.....	41.2
3TC-selected virus	6	0.020	478 (475–480)	74	34	31 (29–31)	27 (25–28)	83.3	.....	41.2
SQV-selected virus	11	0.0040	474	71	34	31	26	100.0	.....	41.2
MVC-selected virus	7	0.0080	469 (468–469)	69	33	29	24 (23–24)	100.0	....-..I....R..T.R..T.KT...N.KK...	1.7

*Overall mean distance.

†Sequence from gp120 SP to the V5 region (aa 1–474).

‡V3 sequences of each variant are shown. Dots denote sequence identity and dashes indicate a deletion mutation.

§Prediction of viral co-receptor tropism using Geno2pheno based on a selectable 'false positive rate'.

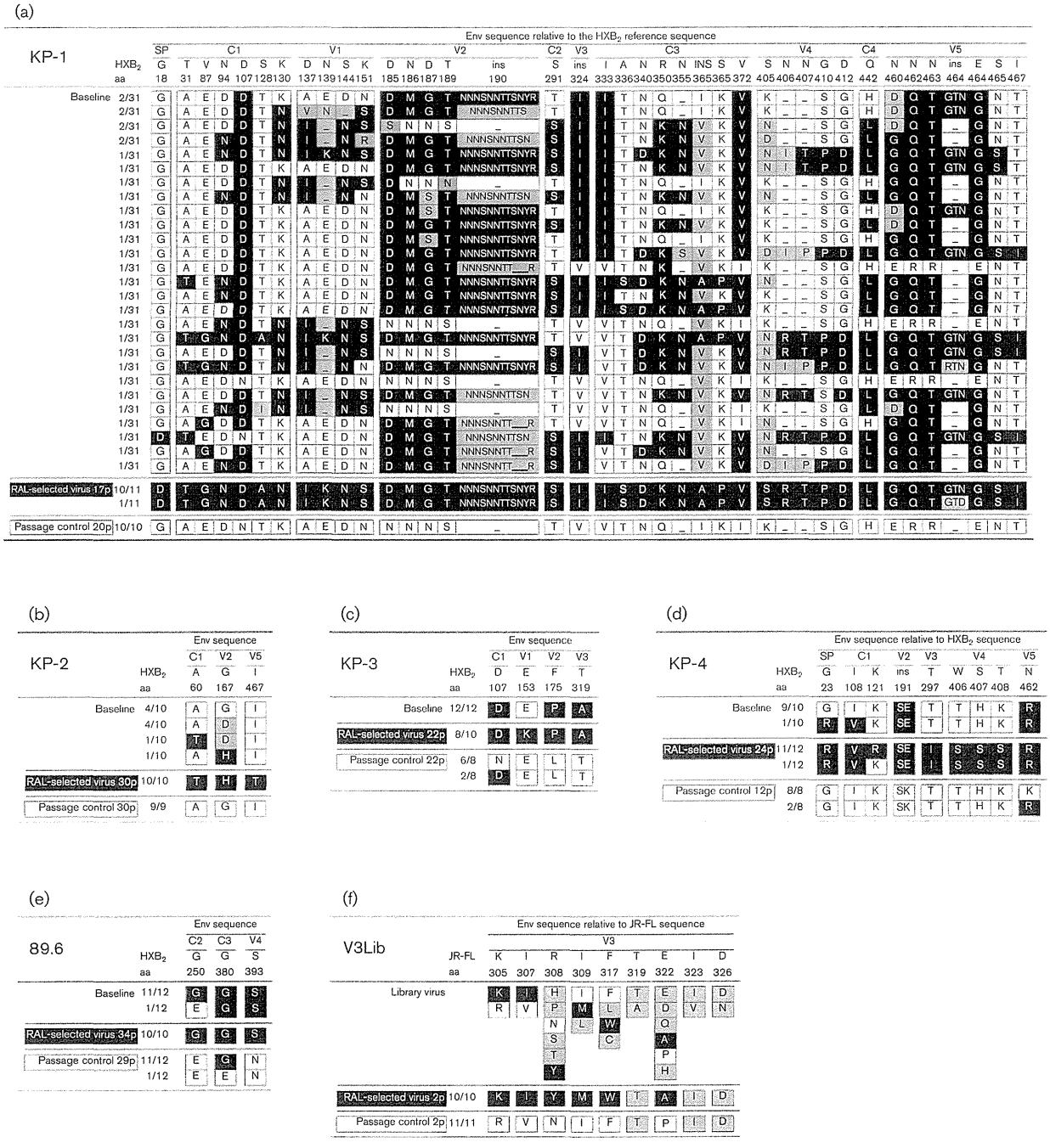


Fig. 1. Comparison of the gp120 sequences between RAL-selected and control-passaged viruses. The gp120 sequences of baseline, RAL-selected and the passage-control viruses were aligned for KP-1 (a), KP-2 (b), KP-3 (c), KP-4 (d) and strain 89.6 (e). Each amino acid in (a)–(e) is numbered relative to the HIV-1 HXB_2 reference sequence. The V3 sequences from the JR-FL-V3Lib baseline library, RAL-selected and passage-control viruses were aligned (f). Filled cells denote the most dominant amino acids observed in RAL-selected variants at the latest passage, open cells denote the most dominant amino acids observed in the passage-control variants at the latest passage and shaded cells show amino acids deleted by the end of both passages, whilst '–' indicates a deletion mutation. The number of passages is indicated, e.g. 17p for passage 17.

In vitro selection of KP-1 variants by 3TC, SQV and MVC

To determine whether other HIV drugs also changed the route of adaptation to the target cells, we attempted to select KP-1 variants using a reverse transcriptase inhibitor (3TC), a protease inhibitor (SQV) and a CCR5 inhibitor (MVC). As shown in Fig. 2(f), the pattern of clustering at distinct positions between the selected isolates and the passage-control variants was similar to that observed for the RAL-selected variants. The selected variants showed decreased diversity in the gp120 sequences; however, the length of the gp120, V1/V2 and V4 sequences increased (apart from in the MVC-selected variants). In addition, the number of PNGs within gp120 was higher than that in the control (Table 3). We also compared the V3 sequences between the passage-control and each of the drug-selected variants. The V3 sequences in all the SQV-selected variants and 83.3 % of those in the 3TC-selected variants, were comparable with those in the RAL-selected variants. This was not the case for the passage controls. Comparison of variants passaged with RAL and 3TC showed that the length of the V1/V2 and V4 regions and the number of PNGs was similar; however, these parameters were different in the SQV-selected variants (Table 3). This indicated that the time at which a drug acts (e.g. during the early or late phase of the HIV life cycle) influences the selection of Env sequences. During selection with MVC, CXCR4-tropic variants were selected from the baseline mixture after seven passages.

Taken together, these results suggested that, in treated cells, different classes of anti-HIV drugs may suppress the variability of quasi-species during *in vitro* selection via a route different from that in untreated cells.

DISCUSSION

This study evaluated the impact of anti-HIV drugs on the Env bottleneck in bulk HIV-1 primary isolates during selection *in vitro*. RAL-, 3TC- and SQV-selected variants of the unique viral isolate, KP-1, harbouring both X4 and R5 variants and with a very high level of baseline viral diversity, were used to study the final destination (genetic bottleneck) of a large variety of Env sequences. Interestingly, the phylogenetic clustering of RAL-selected KP-1 variants was completely different from that of non-drug-treated controls (Fig. 2). Our results also confirmed differences in the length of the gp120, V1/V2 and V4-loop regions and in the number of PNGs (Tables 2 and 3).

It is not clear why viruses cultured under pressure from the non-Env-directed drug RAL result in different *env* genotypes compared with those without the drug. Thus, we cloned the *IN–env* region of the proviral genome from passaged viruses and sequenced the *env* and *IN* regions on the same cloned plasmid, and compared them among the baseline and passages 1, 2, 8 and 17 of the KP-1 virus. Under low concentrations of the IN inhibitor RAL, K7 was selected for at a late passage after accumulation of the other three amino acids, K111, D278 and H216, in IN. During the sequential accumulation of these four amino acids (K111, D278, H216 and K7), the RAL-selected Env sequences at passage 17 (the Env sequences shown as filled boxes in Fig. 1) sequentially accumulated mutations in the same proviral genome (Fig. S1, available in JGV Online). However, we did not find a clone including both the RAL-selected Env at passage 17 and RAL-selected IN at passage 17 in the baseline or each passaged virus, except for in the last passage. We also examined the gp120 and IN sequences of the 3TC- and SQV-selected KP-1 variants. Compared with the RAL-selected region, the variable regions of gp120 in these selected variants were very similar to each other, except for the V1/V2 region (Fig. S2). However, the passage-control variant was very different from the drug-selected variants (Fig. 1a). Furthermore, the IN sequences were different in each passaged virus: K111/D278/H216/K7 in RAL-selected, R111/D278/Q216/R7 in 3TC-selected, K111/D278/H216/R7 in SQV-selected and R111/N278/Q216/R7 in virus without drug treatment (underlined residues indicate amino acids different from those in viruses without drug treatment). To explain these results, we believe that, under pressure from anti-HIV drugs (non-entry ARVs), the virus might show a primitive reaction to select for the Env sequence and recombine from quasi-species to gain advantage for entry and/or enhance replication in target cells. Meanwhile, IN was selected from quasi-species by a direct and/or indirect effect of RAL-induced pressure. The combination of both selective pressures may affect the selection for Env and IN during adaptation in drug-treated conditions (Figs 1a and S2). These results suggest that non-entry inhibitors, such as RAL, 3TC and SQV, might also affect cell adaptation to PM1/CCR5 cells.

Many *in vivo* studies have reported the effects of the anti-HIV drug-induced bottleneck on the *env* gene (Charpentier *et al.*, 2006; Delwart *et al.*, 1998; Ibáñez *et al.*, 2000; Kitrinos *et al.*, 2005; Nijhuis *et al.*, 1998; Nora *et al.*, 2007; Sheehy *et al.*, 1996; Zhang *et al.*, 1994). However, these studies had several limitations. Because viruses were placed under *in vivo* selective pressure using at least two anti-HIV drugs and by the host immune response, it is difficult to separate the different effects and to draw clear conclusions, particularly *in vivo*. Delwart *et al.* (1998) and Kitrinos *et al.* (2005) avoided some of these limitations by employing a heteroduplex tracking assay, although *in vivo* peculiarities still remained. Therefore, we used an *in vitro* selection system using unique bulk primary isolates established in our laboratory (Hatada *et al.*, 2010; Shibata *et al.*, 2007; Yoshimura *et al.*, 2006, 2010b) to observe the effects of the anti-retroviral drug-induced bottleneck on the *IN* and *env* genes.

This selection provides a sensitive approach for analysing virus population dynamics. The effectiveness of ARV drugs can be examined during the *in vitro* passage of a single variant or mixture of variants without being affected by many of the factors encountered *in vivo*. In addition,

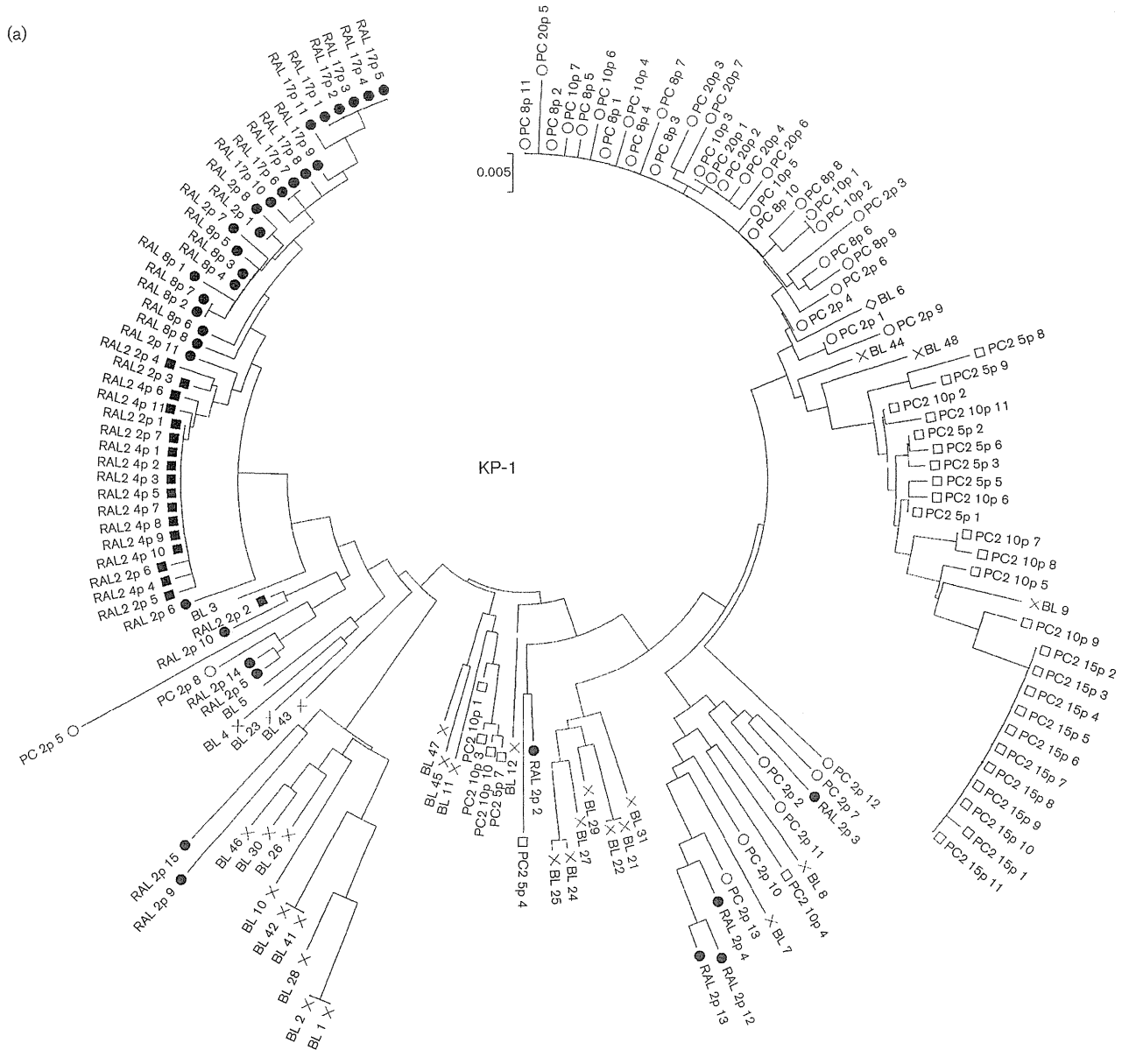
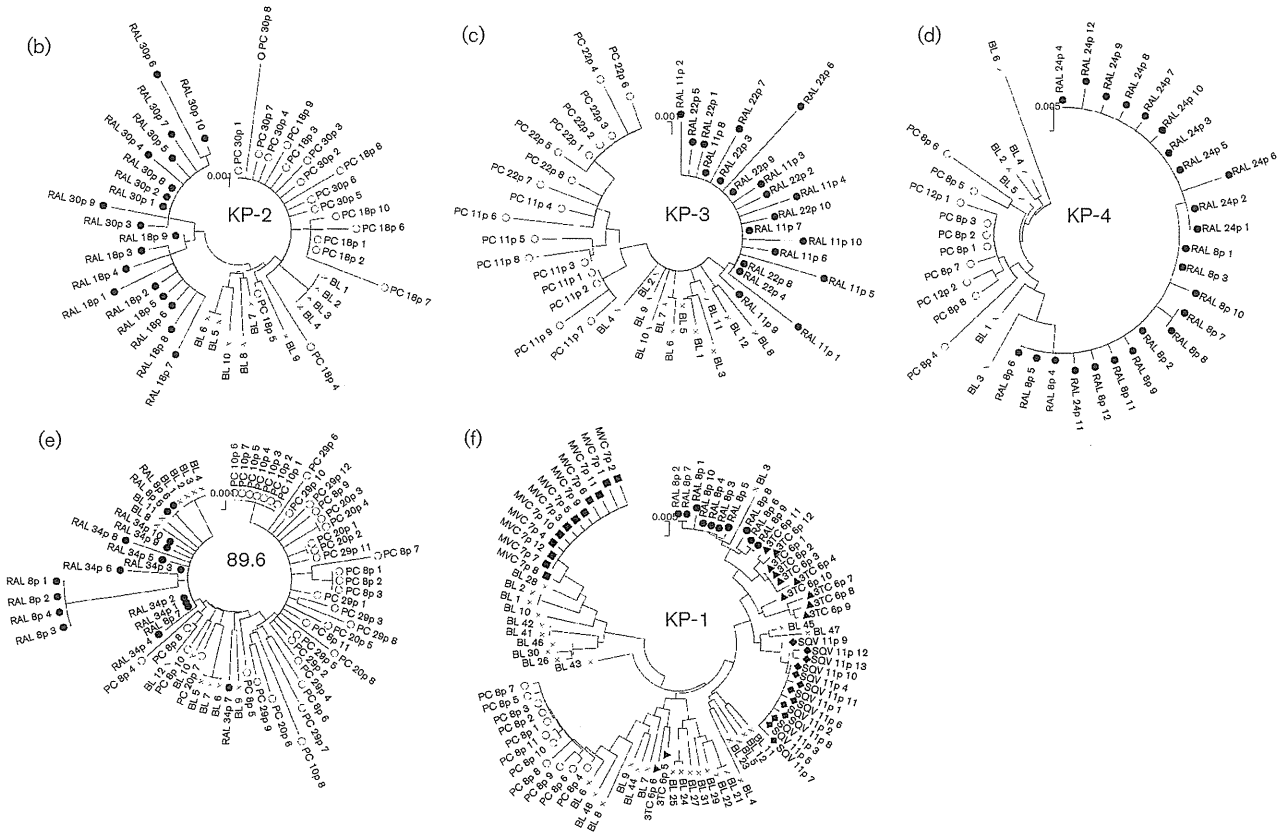


Fig. 2. Phylogenetic analyses of the Env regions from *in vitro*-passaged viruses selected with or without ARV drugs. (a–e) Phylogenetic trees were constructed using gp120 SP–V5 sequences from RAL-selected and passage-control variants of KP-1 (a), KP-2 (b), KP-3 (c), KP-4 (d) and strain 89.6 (e). An '×' represents baseline (BL) variants, and closed and open symbols represent RAL-selected (RAL) and passage-control (PC) variants, respectively. In (a), the results of the second experiment are indicated as RAL2 and PC2, respectively. (f) A phylogenetic tree was constructed using gp120 SP–V5 sequences from RAL-, 3TC-, SQV-, MVC-selected and control-passaged variants of KP-1. ○, Control variants after eight passages; ●, RAL-selected variants after eight passages; ▲, 3TC-selected variants after six passages; ◆, SQV-selected variants after 11 passages; ■, MVC-selected variants after seven passages. The trees were constructed using the neighbour-joining algorithm embedded within the MEGA software.

differences in the Env sequences between the baseline and selected variants can be compared after any number of passages. The results of the present study provide important information that will enhance our understanding of the drug-induced genetic bottleneck. This phenomenon can be examined *in vitro* using bulk primary isolates treated with or without drugs.

Recently, several new ARV drugs have been licensed for use in HIV-1-infected patients. MVC, approved in 2006, is the



first CCR5 inhibitor (Gulick *et al.*, 2008). One important advantage associated with this drug is the absence of cross-resistance with previously available ARV compounds (Gulick *et al.*, 2008; Steigbigel *et al.*, 2008). However, as is usual with anti-HIV drugs, resistant variants with mutations in the Env, gp120 and gp41 sequences are induced both *in vivo* and *in vitro* (Anastassopoulou *et al.*, 2009; Berro *et al.*, 2009; Tilton *et al.*, 2010; Yoshimura *et al.*, 2009, 2010a). As shown in the present study, distinct Env sequences from each quasispecies might be selected by the different anti-HIV drugs (e.g. length of the V1/2 and/or V4 regions, V3 region depletion and the number of PNGs). Moreover, many of the novel anti-retroviral drugs in pre-clinical trials are viral entry inhibitors (e.g. PRO140, ibalizumab, BMS-663068 and PF-232798; Jacobson *et al.*, 2010; McNicholas *et al.*, 2010; Nettles *et al.*, 2011; Stupple *et al.*, 2011; Toma *et al.*, 2011). Therefore, it is necessary to examine whether such entry inhibitors are effective when used alongside conventional drugs.

In conclusion, we studied the genetic bottleneck in bulk primary HIV-1 isolates from untreated patients and drugs targeting the Env (and other) regions. The results showed, for the first time, the presence of drug-selected Env sequences in these isolates. Although our observations were based on a limited number of HIV-1 isolates and need to be confirmed by independent studies, we believe that they provide a new paradigm for HIV-1 evolution in the new combination ARV therapy era.

METHODS

Patients and isolates. Primary HIV-1 isolates were isolated from four drug-naïve patients in our laboratory (KP-1–4) and passaged in phytohaemagglutinin-activated PBMCs. Infected PBMCs were then co-cultured for 5 days with PM1/CCR5 cells (a kind gift from Dr Y. Maeda; Maeda *et al.*, 2008; Yusa *et al.*, 2005) and the culture supernatants were stored at −150 °C (Hatada *et al.*, 2010; Shibata *et al.*, 2007; Yoshimura *et al.*, 2006, 2010b).

After isolation of the primary viruses, we checked the sensitivity of each primary isolate to MVC. The KP-1 isolate was relatively MVC-resistant compared with KP-2 and KP-3 (54 vs 5.9 and 8.7 nM, respectively). KP-1 became MVC sensitive after eight passages in PM1/CCR5 cells [IC_{50}, 3.4 nM; Geno2pheno value (see below), 41.2 %], whilst under the pressure of MVC, KP-1 became highly resistant to MVC after eight passages (IC_{50}, >1000 nM; Geno2pheno value, 1.7 %). These results indicated that the bulk KP-1 isolate used in this study harboured primarily R5 viruses with X4- or dual-tropic viruses as a minor population.

Cells, culture conditions and reagents. PM1/CCR5 cells were maintained in RPMI 1640 (Sigma) supplemented with 10 % heat-inactivated FCS (HyClone Laboratories), 50 U penicillin ml^{-1}, 50 μg streptomycin ml^{-1} and 0.1 mg G418 (Nacalai Tesque) ml^{-1}. MVC, RAL and SQV were kindly provided by Pfizer, Merck & Co. and Roche Products, respectively. 3TC was purchased from Wako Pure Chemical Industries.

The laboratory-adapted HIV-1 strain 89.6, which was obtained through the NIH AIDS Research and Reference Reagent Program, was propagated in phytohaemagglutinin-activated PBMCs. The viral-competent library pJR-FL-V3Lib, which contains 176 bp V3–loop DNA fragments with 0–10 random combinations of amino acid substitutions, was introduced into pJR-FL, as described previously (Yusa *et al.*, 2005).

***In vitro* selection of HIV-1 variants using anti-HIV drugs.** The four primary HIV isolates (KP-1–4), strain 89.6 and JR-FL-V3Lib were treated with various concentrations of RAL and used to infect PM1/CCR5 cells to induce the production of RAL-selected HIV-1 variants, as described previously, with minor modifications (Hatada *et al.*, 2010; Shibata *et al.*, 2007; Yoshimura *et al.*, 2006, 2010b). Briefly, PM1/CCR5 cells (4×10^4 cells) were exposed to 500 $TCID_{50}$ HIV-1 isolates and cultured in the presence of RAL. Virus replication in PM1/CCR5 cells was monitored by observing the cytopathic effects. The culture supernatant was harvested on day 7 and used to infect fresh PM1/CCR5 cells for the next round of culture in the presence of increasing concentrations of RAL. When the virus began to propagate in the presence of the drug, the compound concentration was increased further. Proviral DNA was extracted from lysates of infected cells at different passages using a QIAamp DNA Blood Mini kit (Qiagen). The proviral DNAs obtained were then subjected to nucleotide sequencing. *In vitro* selection of the KP-1 isolate using SQV, 3TC and MVC was also performed using the procedure described above.

Amplification of proviral DNA and nucleotide sequencing. Proviral DNA was subjected to PCR amplification using PrimeSTAR GXL DNA polymerase and Ex-*Taq* polymerase (Takara), as described previously (Hatada *et al.*, 2010; Shibata *et al.*, 2007; Yoshimura *et al.*, 2006, 2010b). The primers used were 1B and H for the gp120 region (Hatada *et al.*, 2010; Shibata *et al.*, 2007; Yoshimura *et al.*, 2006, 2010b), IN 1F (5′-CAGACTCACAATATGCATTAGG-3′) and IN 1R (5′-CCTGTATGCAGACCCCAATATG-3′) for the *IN* region, and IN 1F and H for the *IN*–gp120 region. The first-round PCR products were used directly in a second round of PCR using primers 2B and F (Hatada *et al.*, 2010; Shibata *et al.*, 2007; Yoshimura *et al.*, 2006, 2010b) for gp120, IN 2F (5′-CTGGCATGGGTACCAGCACACAA-3′) and IN 2R (3′-CCTAGTGGGATGTGTACTTCTGAACTTA-3′) for *IN*, and IN 2F and F for *IN*–gp120. The PCR conditions used were as described above. The second-round PCR products were purified and cloned into a pGEM-T Easy Vector (Promega) or pCR-XL-TOPO Vector (Invitrogen), and the *env* and *IN* regions in both the passaged and selected viruses were sequenced using an Applied Biosystems 3500xL Genetic Analyzer and a BigDye Terminator v3.1 Cycle Sequencing kit (Applied Biosystems). Phylogenetic reconstructions were generated using the neighbour-joining method embedded in the MEGA software (http://www.megasoftware.net) (Tamura *et al.*, 2007). Overall, mean distances for viral diversity were also calculated using MEGA software. The number and location of putative PNGs were estimated using *N*-GlycoSite (http://www.hiv.lanl.gov/content/sequence/GLYCOSITE/glycosite.html) from the Los Alamos National Laboratory database.

Susceptibility assay. The sensitivity of the passaged viruses to various drugs was determined as described previously with minor modifications (Hatada *et al.*, 2010; Shibata *et al.*, 2007; Yoshimura *et al.*, 2006, 2010b). Briefly, PM1/CCR5 cells (2×10^3 cells per well) in 96-well round-bottomed plates were exposed to 100 $TCID_{50}$ of the viruses in the presence of various concentrations of drugs and incubated at 37 °C for 7 days. The IC_{50} values were then determined using a Cell Counting Kit-8 assay (Dojindo Laboratories). All assays were performed in duplicate or triplicate.

Predicting co-receptor usage by the V3 sequence. HIV-1 tropism was inferred using Geno2pheno [coreceptor] program, with a false rate positive (FPR) value of 5.0 %, which is freely available (http://coreceptor.bioinf.mpi-inf.mpg.de/index.php). This genotyping tool more accurately predicts virological responses to the CCR5 antagonist MVC in ARV-naïve patients than a reference phenotypic tropism test (Sing *et al.*, 2007).

Statistical analyses. Pairwise comparisons of the different parameters between variants in the two groups was calculated using the homoscedastic *t*-test. A P value of <0.05 was considered statistically significant.

ACKNOWLEDGEMENTS

We are grateful to Dr Yosuke Maeda for providing the PM1/CCR5 cells. We also thank Syoko Yamashita, Yoko Kawanami, Noriko Shirai and Akiko Shibata for technical assistance. This study was supported in part by the Ministry of Education, Culture, Sports, Science and Technology, Japan, by a Grant-in-Aid for Young Scientists (B-22790163); grants from the Ministry of Health, Labour and Welfare; the Program of Founding Research Centers for Emerging and Re-emerging Infectious Diseases; and the Global COE program Global Education and Research Center Aiming at the Control of AIDS.

REFERENCES

Anastassopoulou, C. G., Ketas, T. J., Klasse, P. J. & Moore, J. P. (2009). Resistance to CCR5 inhibitors caused by sequence changes in the fusion peptide of HIV-1 gp41. *Proc Natl Acad Sci U S A* **106**, 5318–5323.

Berro, R., Sanders, R. W., Lu, M., Klasse, P. J. & Moore, J. P. (2009). Two HIV-1 variants resistant to small molecule CCR5 inhibitors differ in how they use CCR5 for entry. *PLoS Pathog* **5**, e1000548.

Charpentier, C., Nora, T., Tenaillon, O., Clavel, F. & Hance, A. J. (2006). Extensive recombination among human immunodeficiency virus type 1 quasispecies makes an important contribution to viral diversity in individual patients. *J Virol* **80**, 2472–2482.

Delwart, E. L., Pan, H., Neumann, A. & Markowitz, M. (1998). Rapid, transient changes at the *env* locus of plasma human immunodeficiency virus type 1 populations during the emergence of protease inhibitor resistance. *J Virol* **72**, 2416–2421.

Eigen, M. (1993). The origin of genetic information: viruses as models. *Gene* **135**, 37–47.

Gulick, R. M., Lalezari, J., Goodrich, J., Clumeck, N., DeJesus, E., Horban, A., Nadler, J., Clotet, B., Karlsson, A. & other authors (2008). Maraviroc for previously treated patients with R5 HIV-1 infection. *N Engl J Med* **359**, 1429–1441.

Hatada, M., Yoshimura, K., Harada, S., Kawanami, Y., Shibata, J. & Matsushita, S. (2010). Human immunodeficiency virus type 1 evasion of a neutralizing anti-V3 antibody involves acquisition of a potential glycosylation site in V2. *J Gen Virol* **91**, 1335–1345.

Hombrouck, A., Voet, A., Van Remoortel, B., Desadeleer, C., De Maeyer, M., Debyser, Z. & Witvrouw, M. (2008). Mutations in human immunodeficiency virus type 1 integrase confer resistance to the naphthyridine L-870,810 and cross-resistance to the clinical trial drug GS-9137. *Antimicrob Agents Chemother* **52**, 2069–2078.

Ibáñez, A., Clotet, B. & Martinez, M. A. (2000). Human immunodeficiency virus type 1 population bottleneck during indinavir therapy causes a genetic drift in the env quasispecies. *J Gen Virol* **81**, 85–95.

Jacobson, J. M., Thompson, M. A., Lalezari, J. P., Saag, M. S., Zingman, B. S., D'Ambrosio, P., Stambler, N., Rotshteyn, Y., Marozsan, A. J. & other authors (2010). Anti-HIV-1 activity of weekly or biweekly treatment with subcutaneous PRO 140, a CCR5 monoclonal antibody. *J Infect Dis* **201**, 1481–1487.

Kitrinos, K. M., Nelson, J. A., Resch, W. & Swanstrom, R. (2005). Effect of a protease inhibitor-induced genetic bottleneck on human immunodeficiency virus type 1 *env* gene populations. *J Virol* **79**, 10627–10637.

Kobayashi, M., Nakahara, K., Seki, T., Miki, S., Kawauchi, S., Suyama, A., Wakasa-Morimoto, C., Kodama, M., Endoh, T. & Oosugi, E. (2008). Selection of diverse and clinically relevant integrase inhibitor-resistant human immunodeficiency virus type 1 mutants. *Antiviral Res* **80**, 213–222.

Maeda, Y., Yusa, K. & Harada, S. (2008). Altered sensitivity of an R5X4 HIV-1 strain 89.6 to coreceptor inhibitors by a single amino acid substitution in the V3 region of gp120. *Antiviral Res* **77**, 128–135.

McNicholas, P., Wei, Y., Whitcomb, J., Greaves, W., Black, T. A., Tremblay, C. L. & Strizki, J. M. (2010). Characterization of emergent HIV resistance in treatment-naive subjects enrolled in a vicriviroc phase 2 trial. *J Infect Dis* **201**, 1470–1480.

Nájera, R., Delgado, E., Pérez-Alvarez, L. & Thomson, M. M. (2002). Genetic recombination and its role in the development of the HIV-1 pandemic. *AIDS* **16** (Suppl. 4), S3–S16.

Nettles, R., Schurmann, D., Zhu, L., Stonier, M., Huang, S. P., Chien, C., Krystal, M., Wind-Rotolo, M., Bertz, R. & Grasela, D. (2011). Pharmacodynamics, safety, and pharmacokinetics of BMS-663068: a potentially first-in-class oral HIV attachment inhibitor. In *18th Conference on Retroviruses and Opportunistic Infections*, abstract 49. Boston, MA.

Nijhuis, M., Boucher, C. A., Schipper, P., Leitner, T., Schuurman, R. & Albert, J. (1998). Stochastic processes strongly influence HIV-1 evolution during suboptimal protease-inhibitor therapy. *Proc Natl Acad Sci U S A* **95**, 14441–14446.

Nora, T., Charpentier, C., Tenaillon, O., Hoede, C., Clavel, F. & Hance, A. J. (2007). Contribution of recombination to the evolution of human immunodeficiency viruses expressing resistance to antiretroviral treatment. *J Virol* **81**, 7620–7628.

Rhee, S.-Y., Liu, T. F., Kiuchi, M., Zioni, R., Gifford, R. J., Holmes, S. P. & Shafer, R. W. (2008). Natural variation of HIV-1 group M integrase: implications for a new class of antiretroviral inhibitors. *Retrovirology* **5**, 74.

Sheehy, N., Desselberger, U., Whitwell, H. & Ball, J. K. (1996). Concurrent evolution of regions of the envelope and polymerase genes of human immunodeficiency virus type 1 observed during zidovudine (AZT) therapy. *J Gen Virol* **77**, 1071–1081.

Shibata, J., Yoshimura, K., Honda, A., Koito, A., Murakami, T. & Matsushita, S. (2007). Impact of V2 mutations on escape from a potent neutralizing anti-V3 monoclonal antibody during in vitro selection of a primary human immunodeficiency virus type 1 isolate. *J Virol* **81**, 3757–3768.

Sing, T., Low, A. J., Beerenwinkel, N., Sander, O., Cheung, P. K., Domingues, F. S., Büch, J., Däumer, M., Kaiser, R. & other authors (2007). Predicting HIV coreceptor usage on the basis of genetic and clinical covariates. *Antivir Ther* **12**, 1097–1106.

Steigbigel, R. T., Cooper, D. A., Kumar, P. N., Eron, J. E., Schechter, M., Markowitz, M., Loutfy, M. R., Lennox, J. L., Gatell, J. M. & other authors (2008). Raltegravir with optimized background therapy for resistant HIV-1 infection. *N Engl J Med* **359**, 339–354.

Stupple, P. A., Batchelor, D. V., Corless, M., Dorr, P. K., Ellis, D., Fenwick, D. R., Galan, S. R., Jones, R. M., Mason, H. J. & other authors (2011). An imidazopiperidine series of CCR5 antagonists for the treatment of HIV: the discovery of N-(1S)-1-(3-fluorophenyl)-3-[(3-endo)-3-(5-isobutyryl-2-methyl-4,5,6,7-tetrahydro-1H-imidazo[4,5-c]pyridin-1-yl)-8-azabicyclo[3.2.1]oct-8-yl]propylacetamide (PF-232798). *J Med Chem* **54**, 67–77.

Tamura, K., Dudley, J., Nei, M. & Kumar, S. (2007). MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Mol Biol Evol* **24**, 1596–1599.

Tilton, J. C., Wilen, C. B., Didigu, C. A., Sinha, R., Harrison, J. E., Agrawal-Gamse, C., Henning, E. A., Bushman, F. D., Martin, J. N. & other authors (2010). A maraviroc-resistant HIV-1 with narrow cross-resistance to other CCR5 antagonists depends on both N-terminal and extracellular loop domains of drug-bound CCR5. *J Virol* **84**, 10863–10876.

Toma, J., Weinheimer, S. P., Stawiski, E., Whitcomb, J. M., Lewis, S. T., Petropoulos, C. J. & Huang, W. (2011). Loss of asparagine-linked glycosylation sites in variable region 5 of human immunodeficiency virus type 1 envelope is associated with resistance to CD4 antibody ibalizumab. *J Virol* **85**, 3872–3880.

Vignuzzi, M., Stone, J. K., Arnold, J. J., Cameron, C. E. & Andino, R. (2006). Quasispecies diversity determines pathogenesis through cooperative interactions in a viral population. *Nature* **439**, 344–348.

Yoshimura, K., Shibata, J., Kimura, T., Honda, A., Maeda, Y., Koito, A., Murakami, T., Mitsuya, H. & Matsushita, S. (2006). Resistance profile of a neutralizing anti-HIV monoclonal antibody, KD-247, that shows favourable synergism with anti-CCR5 inhibitors. *AIDS* **20**, 2065–2073.

Yoshimura, K., Harada, S., Hatada, M. & Matsushita, S. (2009). Mutations in V4 and C4 regions of the HIV-1 CRF08-BC envelope induced by the in vitro selection of Maraviroc Confer cross-resistance to other CCR5 inhibitors. In *16th Conference on Retroviruses and Opportunistic Infections*, p. 640. Montreal, Canada.

Yoshimura, K., Harada, S. & Matsushita, S. (2010a). Two step escape pathway of the HIV-1 subtype C primary isolate induced by the in vitro selection of Maraviroc. In *17th Conference on Retroviruses and Opportunistic Infections*, abstract 535. San Francisco, CA.

Yoshimura, K., Harada, S., Shibata, J., Hatada, M., Yamada, Y., Ochiai, C., Tamamura, H. & Matsushita, S. (2010b). Enhanced exposure of human immunodeficiency virus type 1 primary isolate neutralization epitopes through binding of CD4 mimetic compounds. *J Virol* **84**, 7558–7568.

Yusa, K., Maeda, Y., Fujioka, A., Monde, K. & Harada, S. (2005). Isolation of TAK-779-resistant HIV-1 from an R5 HIV-1 GP120 V3 loop library. *J Biol Chem* **280**, 30083–30090.

Zhang, Y. M., Dawson, S. C., Landsman, D., Lane, H. C. & Salzman, N. P. (1994). Persistence of four related human immunodeficiency virus subtypes during the course of zidovudine therapy: relationship between virion RNA and proviral DNA. *J Virol* **68**, 425–432.

Structural analysis for glycolipid recognition by the C-type lectins Mincle and MCL

Atsushi Furukawa[a,b,1], Jun Kamishikiryo[c,1], Daiki Mori[d,1], Kenji Toyonaga[d], Yuki Okabe[a,2], Aya Toji[a], Ryo Kanda[a], Yasunobu Miyake[d], Toyoyuki Ose[a], Sho Yamasaki[d,3], and Katsumi Maenaka[a,b,3]

[a]Laboratory of Biomolecular Science, Faculty of Pharmaceutical Sciences, Hokkaido University, Sapporo 060-0812, Japan; [b]Core Research for Evolutional Science and Technology (CREST), Japan Science and Technology Agency, Saitama 332-0012, Japan; [c]Faculty of Pharmacy and Pharmaceutical Sciences, Fukuyama University, Fukuyama 729-0292, Japan; and [d]Division of Molecular Immunology, Research Center for Infectious Diseases, Medical Institute of Bioregulation, Kyushu University, Fukuoka 812-8582, Japan

Edited* by Pamela J. Bjorkman, California Institute of Technology, Pasadena, CA, and approved September 9, 2013 (received for review July 7, 2013)

Mincle [macrophage inducible Ca^{2+}-dependent (C-type) lectin; CLEC4E] and MCL (macrophage C-type lectin; CLEC4D) are receptors for the cord factor TDM (trehalose-6,6′-dimycolate), a unique glycolipid of mycobacterial cell-surface components, and activate immune cells to confer adjuvant activity. Although it is known that receptor–TDM interactions require both sugar and lipid moieties of TDM, the mechanisms of glycolipid recognition by Mincle and MCL remain unclear. We here report the crystal structures of Mincle, MCL, and the Mincle–citric acid complex. The structures revealed that these receptors are capable of interacting with sugar in a Ca^{2+}-dependent manner, as observed in other C-type lectins. However, Mincle and MCL uniquely possess shallow hydrophobic regions found adjacent to their putative sugar binding sites, which reasonably locate for recognition of fatty acid moieties of glycolipids. Functional studies using mutant receptors as well as glycolipid ligands support this deduced binding mode. These results give insight into the molecular mechanism of glycolipid recognition through C-type lectin receptors, which may provide clues to rational design for effective adjuvants.

X-ray crystallography | innate immunity | mycobacteria | pattern-recognition receptors | myeloid cells

Pattern-recognition receptors (PRRs) play important roles in innate immunity. PRRs recognize nucleotides, sugars, lipopolysaccharides, other pathogen components, and self-ligands, and consequently trigger intracellular signaling cascades that initiate innate and adaptive immune responses (1). Among them, Toll-like receptors (TLRs) are well-characterized receptors, in terms of their ligand specificities, ligand-recognition mechanisms, and signaling pathways (2–4). C-type lectin receptors (CLRs) are also a large family of PRRs (5–7). The term "C-type lectin" was introduced to distinguish a group of Ca^{2+}-dependent lectins from other lectins. In the CLRs, two amino acids harboring long carbonyl side chains separated by a proline in a *cis* conformation coordinate a Ca^{2+} ion, which forms hydrogen bonds with monosaccharides and determines the binding specificity. CLRs have broad recognition abilities toward not only saccharides but also proteins (5, 7–9). For instance, human NKR-P1 interacts with Lectin-like transcript 1, and some members of the CD94/NKG2 family interact with HLA-E.

Macrophage inducible C-type lectin (Mincle; also called CLEC4E) is a type II transmembrane C-type lectin receptor that is expressed in macrophages, dendritic cells, and monocytes upon stimulation (10). We have reported that Mincle is an FcRγ-coupled activating receptor that recognizes pathogenic fungi and mycobacteria (11–13). Detailed investigations of the ligands of Mincle revealed that Mincle binds glycolipids, such as trehalose-6,6′-dimycolate (TDM) from *Mycobacterium tuberculosis*, and novel glyceroglycolipids from *Malassezia* fungus. *Malassezia* and *M. tuberculosis* ligands are recognized through the carbohydrate-recognition domain (CRD) in the extracellular region of Mincle (11, 12). The binding of TDM to Mincle leads to the phosphorylation of the immunoreceptor tyrosine-based activation motif (ITAM) in the FcRγ chain, which provides a binding site for the Syk tyrosine kinase. Syk activates the caspase recruitment domain family member 9-mediated NF-κB signaling pathway to promote the expression of TNF and IL-6. A recent report revealed that Mincle plays a nonredundant role in T-cell immune responses to infection by microbes and in the adjuvanticity of mycobacterial cord factor and its synthetic analog, trehalose-dibehenate (14, 15).

Macrophage C-type lectin (MCL; also called CLEC4D) is another C-type lectin receptor expressed in myeloid cells (16, 17). Recently, we found that MCL is also an FcRγ-coupled activating receptor that binds to TDM (15). MCL is distinct from Mincle in the following ways: (*i*) The expression of Mincle is inducible, whereas MCL is constitutively expressed in myeloid cells; (*ii*) MCL shows weaker binding affinity to TDM than does Mincle; and (*iii*) the glutamic acid-proline-asparagine (EPN) motif, a typical glucose/mannose-binding motif, is conserved in Mincle but not in MCL.

We now report the crystal structures of Mincle and MCL, as well as Mincle complexed with citric acid. They have similar overall structures to other typical CLRs, but exhibit characteristic conformations in the vicinity of the Ca^{2+}-binding motif. A patch of hydrophobic amino acids located adjacent to the carbohydrate binding site may likely contribute to the recognition of

Significance

Here we report the crystal structures of human C-type lectin receptors Mincle (macrophage inducible C-type "calcium-dependent" lectin; CLEC4E) and MCL (macrophage C-type lectin; CLEC4D), both of which are receptors for mycobacterial glycolipid adjuvant cord factor (also called trehalose-6,6′-dimycolate; TDM). Our structural and functional studies clearly reveal the simultaneous recognition of sugar and lipid moieties by Mincle and MCL, distinct from other C-type lectin receptors. Because better adjuvants are desired for enhancing vaccination effects of medical treatments for infectious diseases, cancer, and so forth, these structures provide a framework for the rational design of more effective adjuvants than TDM.

Author contributions: A.F., J.K., D.M., S.Y., and K.M. designed research; A.F., J.K., D.M., K.T., Y.O., A.T., R.K., Y.M., and T.O. performed research; A.F., J.K., D.M., K.T., Y.O., R.K., Y.M., T.O., S.Y., and K.M. analyzed data; and A.F., J.K., D.M., T.O., S.Y., and K.M. wrote the paper.

The authors declare no conflict of interest.

*This Direct Submission article had a prearranged editor.

Data deposition: The crystallography, atomic coordinates, and structure factors reported in this paper have been deposited in the Protein Data Bank, www.pdb.org (PDB ID codes 3WH3, 3WH2, and 3WHD).

[1]A.F., J.K., and D.M. contributed equally to this work.

[2]Present address: Department of Functional Biological Chemistry, Division of Science, Fukuoka University, Fukuoka 814-0180, Japan.

[3]To whom correspondence may be addressed. E-mail: yamasaki@bioreg.kyushu-u.ac.jp or maenaka@pharm.hokudai.ac.jp.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1312649110/-/DCSupplemental.

the fatty acid chain of TDM. The mutational analysis essentially supports this TDM-binding model, and may also explain the different affinities of MCL and Mincle.

Results

Preparation, Crystallization, and Structural Determination of MCL. The extracellular domain of human MCL (residues 61–215) (Fig. 1) was expressed in *Escherichia coli* as inclusion bodies, and was refolded in vitro by a dilution method. Ca^{2+} ions were required in the refolding procedure, and the crude, refolded MCL was purified by sequential gel-filtration chromatography steps (Fig. S1 *A* and *C*). The purified MCL was crystallized by the hanging-drop method with 0.1 M bis-Tris propane (pH 6.5), 0.2 M potassium thiocyanate, and 20% (wt/wt) PEG 3350. Crystals of the MCL protein [the space group was *I*-centered orthorhombic (*I*222), and the unit-cell parameters were $a = 85.19$ Å, $b = 96.06$ Å, $c = 104.53$ Å] were obtained, and the dataset was collected to the resolution limit of 2.2 Å at the BL32XU beamline at SPring-8 (Harima, Japan) (Table S1). The crystal structure of MCL has 2 α-helices (α1 and α2) and 11 β-strands (β1–β11) (Fig. 2*A*), which is a typical structural organization of CLRs and partly similar to the solution structure of MCL registered in the Protein Data Bank (PDB) (ID code 2LS8) (Fig. S1*E*). Two MCL molecules exist in the asymmetric unit. Gel-filtration analysis showed a mixture of the peaks (Fig. S1*A*), suggesting that MCL may have some conformational variation.

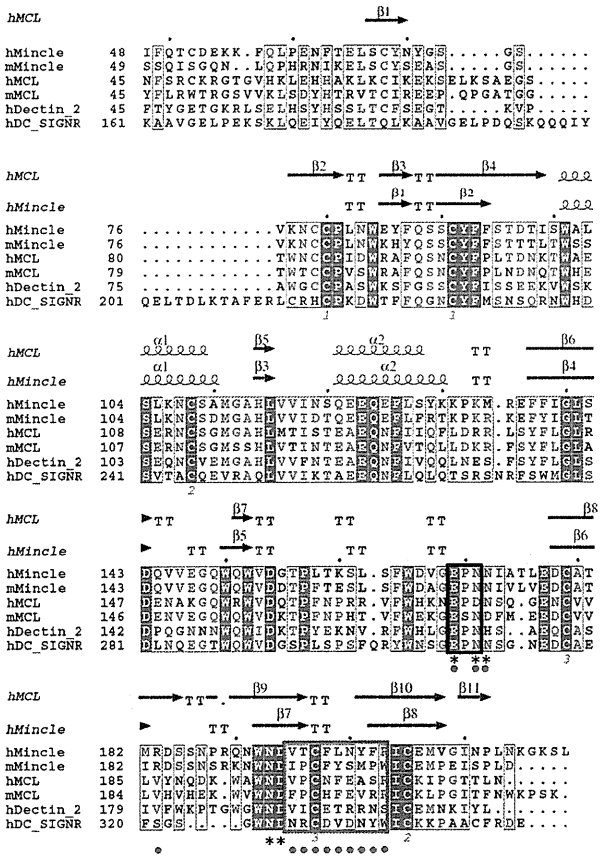


Fig. 1. Structure-based sequence alignment of Mincle, MCL, DC-SIGNR, and Dectin-2. The sequence alignment of the ectodomains of Mincle, MCL, DC-SIGNR, and Dectin-2 (h, human; m, mouse) is shown, as depicted with ESPript (44). Identical residues are highlighted in red, and similar residues are framed in blue. The secondary structure elements (α, α-helix; β, β-strand; T, turn) of Mincle and MCL are shown above the sequences. The box enclosed by the thick black line indicates EPN motifs, which are usually involved in carbohydrate recognition by C-type lectins. The box enclosed by the thick blue line indicates hydrophobic amino acid loops, and yellow-shaded amino acid residues are hydrophobic residues within Mincle and MCL. The asterisks below the sequences indicate the residues involved in calcium binding in Mincle and MCL. The red-filled circles below the sequences indicate the residues changed to other amino acid residues in the mutational studies. The numbers below the cysteine residues indicate disulfide-bond formation with the cysteine residue with the same number.

Preparation, Crystallization, and Structural Determination of Mincle. Using a similar refolding method to that for MCL, we also prepared the extracellular domain of Mincle (residues 74–219) (Fig. 1). The expression, refolding, and purification were successful. However, the crystallization was not successful, because the refolded Mincle was not sufficiently soluble at high concentrations. To improve the protein solubility, we performed site-directed mutagenesis and changed the hydrophobic amino acids presumably located on the surface of the Mincle protein to hydrophilic amino acids, as found in the corresponding residues of MCL. Among them, the mutant with the substitution of isoleucine to lysine at residue 99 (I99K mutant) formed good crystals by the hanging-drop method under two conditions. One is 1 M lithium chloride, 0.1 M citric acid (pH 4), and 20% (wt/vol) PEG 6000, and the other is 0.2 M NH_4SO_4, 0.1 M bis-Tris (pH 5.5), 25% (wt/vol) PEG 3350. These diffraction data were collected to the resolution limit of 1.3 and 1.35 Å at the BL5A and BL17A beamlines at High Energy Accelerator Research Organization (KEK) (Tsukuba, Japan), respectively. Both crystals have the same space group, primitive trigonal systems ($P3_1$), and similar unit-cell parameters (Table S1). Mincle exhibits a typical CLR fold, as shown in Fig. 2*B*. The asymmetric unit contained one molecule of Mincle, and no physiologically important packing was detected. This is consistent with the gel-filtration analysis showing that Mincle behaves as a monomer, although the eluted time is later than the expected one, likely due to the affinity of Mincle to the glucose-based dextran resin of the Superdex column (Fig. S1 *B* and *D*).

Structural Comparison Between Mincle, MCL, and Other C-Type Lectins. MCL and Mincle superimposed well on each other [root-mean-square deviation (rmsd) 1.5 Å for 124 Cα atoms] (Fig. 2 *A* and *B* and Fig. S2*B*). However, Mincle has two calcium ions, whereas MCL has only one. A DALI analysis (18) indicated that MCL and Mincle share high homology with mouse collectin (PDB ID code 2OX9) [rmsd of 1.3 Å for 120 Cα atoms, 28% identity (MCL); 1.07 Å for 118 Cα atoms, 35% identity (Mincle)] (19) and DC-SIGNR (PDB ID code 1K9J) [rmsd of 1.7 Å for 114 Cα atoms, 37% identity (MCL); 1.3 Å for 121 Cα atoms, 45% identity (Mincle)], which has been extensively studied as an entry receptor of HIV (20, 21). Because collectin recognizes fucose-based oligosaccharides, rather than glucose- or mannose-based ones, we chose to compare the structural features of MCL, Mincle, and DC-SIGNR (Fig. 2 *C* and *D*). The entire structures and positions of the amino acid residues in the putative CRD are similar. Specifically, the positions of the Ca^{2+} ions (site 1) are the same among the three proteins. The EPN motif (residues 169–171 in Mincle) is often observed in C-type lectins, and contributes to carbohydrate recognition via a Ca^{2+} ion-mediated binding network (Fig. 3 *A* and *B*). In contrast, the glutamic acid-proline-aspartic acid (EPD) motif of MCL (residues 173–175) is an unusual sequence among C-type lectins (6) (Fig. 1). However, the Ca^{2+} ion and other amino acids involved in carbohydrate recognition are located in this region, as in other C-type lectins (Figs. 2 and 3 *A*–*C*). These results indicated that Mincle and MCL recognize carbohydrates through these motifs in slightly different but similar ways.

The regions surrounding the Ca^{2+}-bound sites in MCL and Mincle are distinctly different from those in DC-SIGNR. In DC-SIGNR, two additional bound Ca^{2+} ions are observed close to the site (red, 2 and 3) (Fig. 2*C*), and they stabilize the typical protein conformation of the C-type lectins (6). The Ca^{2+} (site 2 and 3) ions push the loop (residues 312–317) close to the Ca^{2+} (site 1) ion (Fig. 2*D*, red dotted oval). In contrast, the corresponding

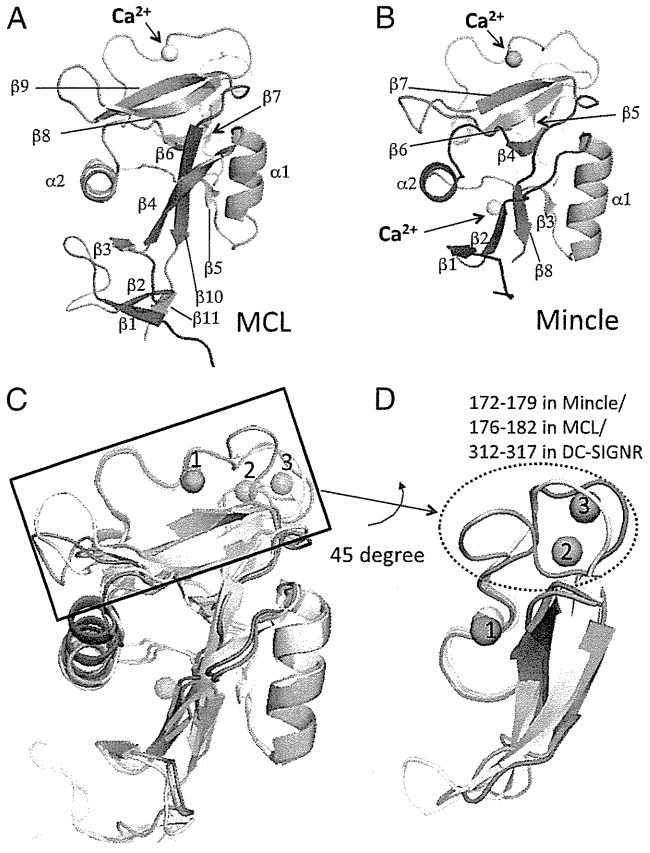


Fig. 2. Structures of MCL and Mincle, and structural comparison with DC-SIGNR. (*A* and *B*) Cartoon models of the overall structures of MCL (*A*) and Mincle (*B*). Secondary structure elements are shown. Gradient rainbow color from blue to red indicates N- to C-terminal. The yellow and cyan spheres represent Ca^{2+} ions in MCL and Mincle, respectively. (*C* and *D*) Overall structures (*C*) and putative ligand binding sites (*D*) of MCL (yellow), Mincle (cyan), and DC-SIGNR (pink) are shown. (*D*) Close-up view of the black box in *C*. Yellow, blue, and pink spheres represent Ca^{2+} ions in Mincle, MCL, and DC-SIGNR, respectively. The red dotted oval indicates the large structural differences of loops among these CLRs (*Results*).

loops in MCL and Mincle are located far from the Ca^{2+} (site 1) ion. The asparagine/aspartate residues just following the EPD/EPN sequences are conserved (Fig. 1). The directions of the asparagines in MCL (residue 176) and Mincle (residue 172) are different from those in other CLRs, such as DC-SIGNR (Fig. 3 *A–C*). The asparagine in DC-SIGNR is used to bind the Ca^{2+} (site 2) ion, and therefore the side chain faces the opposite direction of the Ca^{2+} (site 1) ion. In contrast, neither Mincle nor MCL coordinates Ca^{2+} (site 2 and 3) ions, and their asparagine side chains extend in different directions compared with other C-type lectins.

Calcium Binding and Ligand Recognition. In the crystals of Mincle grown in 1 M lithium chloride, 0.1 M citric acid (pH 4), and 20% (wt/vol) PEG 6000, a strong electron density in addition to that of the Mincle protein was observed close to the Ca^{2+} (site 1) ion and matched a citric acid molecule (Fig. 3*D*). The superimposition of the amino acids in the Ca^{2+} ion-binding regions of the ligand complex structures of Mincle and DC-SIGNR revealed the well-conserved locations of the oxygen atoms of the ligands citrate and mannose (equatorial 3- and 4-OH groups), respectively (Fig. 3*E*). Because the chemical property of the sugar moiety is different from citric acid, we cannot simply compare the recognition modes, but these data may support the idea that Mincle can use this Ca^{2+} ion to bind nucleophiles for the sugar moieties of TDM and *Malassezia* ligands, in essentially the same manner as generally observed in CLRs including DC-SIGNR.

The calcium binding site in human and mouse Mincle includes the EPN motif, well-conserved in the mannose-recognizing C-type lectins, as described above. We examined whether the EPN motif in Mincle is involved in direct TDM recognition using soluble Mincle protein (Mincle–Ig). Mincle–Ig ($Mincle^{WT}$), but not control Ig, selectively bound to plate-coated TDM, as previously reported (11, 12). This recognition was shown to require the EPN motif, as the binding was eliminated by introducing a mutation of EPN into glutamine-proline-aspartic acid (QPD), a putative galactose-recognition sequence ($Mincle^{QPD}$) (22). Substitution of the EPN motif into MCL-type EPD ($Mincle^{EPD}$) also impaired the binding capacity, although their reactivities to anti-hIgG were comparable (Fig. 3*F* and Fig. S3*A*). These data suggested that EPN in human Mincle is indispensable for TDM recognition, as previously shown in mouse Mincle (11, 12). In contrast, the direct binding of MCL to TDM was much weaker than that of Mincle (Fig. 3*G*), consistent a the previous report that MCL recognizes TDM with less affinity than Mincle (15). Mutation of the EPD sequence of MCL into QPD (MCL^{QPD}) did not have a large impact on TDM binding in higher concentrations. Unexpectedly, however, the EPD-to-EPN mutation in MCL, which was expected to coordinate the Ca^{2+} ion location well and facilitate carbohydrate binding, did not improve the affinity for TDM (Fig. 3*G*). These results suggested that the TDM binding site of MCL might be distinct from that used by Mincle. Furthermore, the side chain of Arg183 in Mincle is in a suitable position to interact with the hydroxyl groups of TDM, based on the crystal structure of Mincle (23) (Fig. 4*A*). This arginine residue of Mincle is well-conserved from fish to mammals. In contrast, the valine (Val186) at the corresponding position of human MCL is conserved among placentalia; however, its side chain cannot reach the putative carbohydrate recognition site (Fig. 4*B*). To verify the role of Arg183 in TDM recognition, we introduced the R183V mutation into Mincle and tested its function in an NFAT-GFP reporter assay. This mutation reduced NFAT-GFP activity in the reporter cell assay, suggesting that Arg183 of Mincle is crucially involved in ligand recognition (Fig. 4*C*).

Taken together, these results strongly suggested that the binding mode of the two OH groups of citric acid to Ca^{2+} reflects the equatorial 3- and 4-OH groups of mannose and glucose of Mincle/MCL ligands in a similar but slightly different manner from the CLRs (6).

Putative Lipid Recognition Sites. To determine whether Mincle and MCL use unique amino acids for their interactions with the lipid regions of glycolipids, we verified the characteristics of the surfaces surrounding the putative sites for Ca^{2+}-mediated sugar binding. A series of hydrophobic regions was specifically found in Mincle and MCL but not in other C-type lectins in the vicinity of the putative sugar binding sites (dotted circles in Fig. 4 *A* and *B*, yellow surfaces in Fig. 4 *D–F*, and yellow-shaded amino acid residues in the box enclosed in blue in Fig. 1). The regions are composed of Val195, Thr196, Phe198, Leu199, Tyr201, and Phe202 in Mincle, and Val197, Pro198, and Phe201 in MCL. Mincle has larger hydrophobic areas than MCL, whereas DC-SIGNR has only a much smaller one than both Mincle and MCL (Fig. 4 *D–F*). If the trehalose part of TDM is placed on the sugar binding site of Mincle, as in the binding mode of mannose to DC-SIGNR, then the mycolic acid attached to the 6-*O* of the glucose of TDM (Fig. 4 *A* and *B*, red arrows) is oriented toward the hydrophobic regions of Mincle and MCL, as described above. To investigate whether the hydrophobic region of Mincle contributes to the recognition of TDM, Ala substitutions of both Phe198 and Leu199 in this region were introduced in reporter cells expressing Mincle (Fig. 4*C*). The cells expressing the $Mincle^{F198A/L199A}$ mutant exhibited reduced NFAT activity in response to TDM. Moreover, we replaced the hydrophobic region of Mincle (residues 195–202) with the corresponding region of another CLR, Dectin-2 (residues 192–199), which lacks the hydrophobic residues (24, 25).

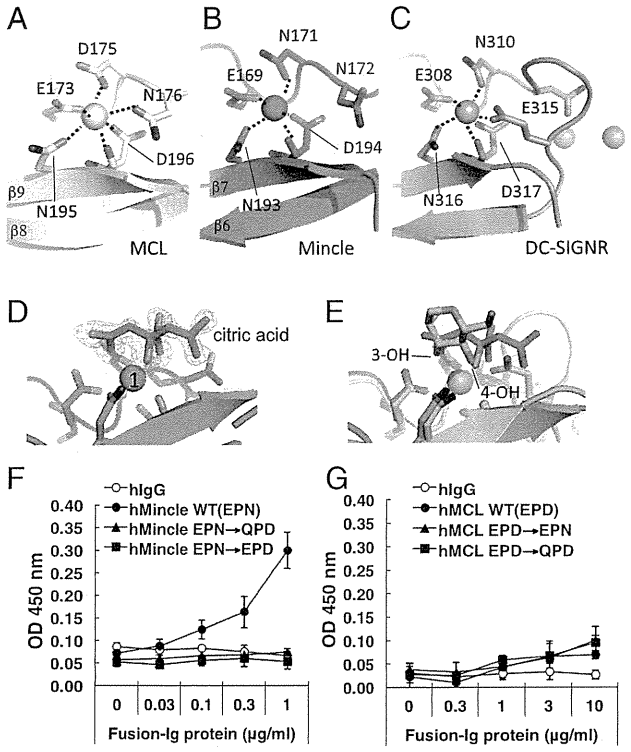


Fig. 3. Structural comparison of the putative ligand binding sites in MCL, Mincle, and DC-SIGNR, and in vitro binding assays of the Mincle and MCL mutants. (*A–C*) Close-up views of the putative ligand binding sites of MCL (yellow) (*A*), Mincle (cyan) (*B*), and DC-SIGNR (pink) (*C*) are shown. The amino acid residues involved in and close to Ca^{2+} ion binding are shown as stick models. Interactions with Ca^{2+} ions are shown as black dotted lines. (*D*) Composite omit map ($2F_o - F_c$) for citric acid in Mincle. The electron-density map is contoured at 1.0σ, and the resolution is 1.3 Å. The citric acid is shown with the O atoms colored red and the C atoms in green. Putative amino acids involved in Ca^{2+} binding are depicted by sticks. (*E*) The superimposed structures of Mincle (cyan; same as *D*) and DC-SIGNR (pink) are shown. The stick model indicates the mannose (the O atoms are colored red and the C atoms are in pink) in the DC-SIGNR complex. (*F* and *G*) Mincle–Ig, mutated Mincle–Ig, or hIgG (*F*) and MCL–Ig, mutated MCL–Igs, or hIgG (*G*) were incubated with plate-coated TDM. Bound proteins were detected by anti-hIgG-HRP. Error bars indicate ± standard deviation of three independent experiments.

The reporter cells expressing this Mincle–Dectin-2 chimeric molecule ($Mincle^{MDchimera}$) (Fig. S4) still retained activity against the anti-Mincle mAb 13D10-H11 (Fig. S3*C*), which recognizes the conformational epitope on Mincle (Fig. S3*D*), indicating that the mutation as well as other mutations in this study did not have remarkable effects on overall protein folding and stability on the cell surface. However, the TDM recognition of $Mincle^{MDchimera}$ was severely compromised (Fig. 4*C*). As described above, the Oδ atom of the corresponding residue Asn172 just following the EPN motif in Mincle does not face the Ca^{2+} ion, which is an unusual type of Ca coordination among C-type lectins (Fig. 3 *A* and *B*). Instead, the Nδ atom of Asn172 forms a hydrogen bond with the Oδ atom of Thr196 of the hydrophobic patch (Fig. 4*A*). The reporter cells expressing the mutant Mincle ($Mincle^{N172Q}$), which has only one additional methylene group, showed reduced NFAT-GFP activity (Fig. 4*C*). This result may suggest that the N172Q mutation indirectly influences the hydrophobicity of the putative lipid-binding patch via the side chain of Thr196.

To further examine the effect of a set of acyl chains, we performed surface plasmon resonance (SPR) binding assays using a set of trehalose-based glycolipids, which have a single acyl chain with different carbon lengths (C8, C10, C12). These glycolipids have a single and short tail, and thus are expected to be water-soluble while retaining ligand activity. The single acyl chains with trehalose (C10 and C12) bound to Mincle (Fig. 4*G* and Fig. S5). The affinity of C8 to Mincle is much lower than those of C10 and C12 (Fig. 4*G*). The crystal structure clearly indicated that the 10-carbon acyl chain with trehalose is reasonably accommodated within the hydrophobic portion in Mincle (Fig. S6).

Discussion

We have determined the crystal structures of the ectodomains of Mincle and MCL, which confirmed that the overall structures of Mincle and MCL are similar to those of other CLRs. Furthermore, we have also solved the crystal structure of Mincle complexed with citric acid, which revealed that the binding mode to citric acid essentially resembles that of glucose/mannose recognition by typical CLRs. We further performed competition binding of glycolipids with citric acid as in Fig. S7, clearly showing that citric acid inhibits glycolipid binding to Mincle, whereas acetic acid does not. Notably, another mannose-binding C-type lectin, codakine, bound similar positions of oxygens of glycerol and glycan in a Ca^{2+} ion-mediated manner (26). Citric acid is likely accommodated at this position to block the ligands, and hydroxyl groups are likely used following the coordination of Ca^{2+} ions generally observed in CLRs.

Glycolipids play pivotal roles in innate immunity, as exemplified by the functions of CD1-mediated natural killer T (NKT) cells (27, 28). CD1 family molecules display a variety of glycolipids toward semi-invariant NKT cell receptors to activate NKT cells. Structural analyses of CD1 family proteins have revealed that the lipid parts of glycolipids are deeply accommodated inside the hydrophobic cores of the proteins (29, 30), and thus only the sugar moieties are exposed for recognition by NKT cell receptors (Fig. S6). On the other hand, our present study showed that the putative TDM binding sites of Mincle and MCL include hydrophobic loops uniquely found in Mincle and MCL, which distinguish them from other C-type lectins (Fig. 4 *A*, *B*, and *D–F*). These loops form shallow hydrophobic patches extending from the corresponding position of the 6-OH of glucose in the structure of the mannose complex of DC-SIGNR, which is attached to mycolic acid in the case of TDM (Fig. S6). The mutational study suggested that these CLRs directly recognize the acyl groups of the glycolipid TDM using this shallow hydrophobic region, which is close to the Ca^{2+} binding site (Fig. 4 *D* and *E*). Notably, the SPR binding study using a set of glycolipids clearly showed that the single acyl chain is sufficient for Mincle binding. In addition, importantly, at least a C10 length of the lipid moiety is required (Fig. 4*G*). These observations might suggest that Mincle recognizes only the sugar-proximal part of the acyl chain of glycolipids. The hydrophobic patch branches out from the potential sugar binding site (downward and to right in Fig. 4*D* and Fig. S6). These might the sites accommodating the branched acyl chains in mycolic acids, such as TDM and trehalose monomycolate (TMM) (11). The recently discovered ligands of Mincle, which also have branched acyl chains, may interact similarly with TMM and TDM (13). Therefore, the recognition of glycolipids by Mincle and presumably MCL seems to be significantly distinct from those of lipid-recognition proteins, such as CD1 and the Toll-like receptor 4–MD2 complex, which have deep hydrophobic grooves to accommodate the acyl moieties of glycolipids (Fig. S6). Thus, a minimum acyl-chain length is required for glycolipid recognition by CLRs. The unique modes of CLR-glycolipid recognition would be advantageous for host defense responses, because they may allow receptors to recognize these bipolar ligands even within a microbial cell wall or in the micellar form in aqueous solution. Future study of cocrystallization with glycolipids harboring short branched acyl chains, which might have increased binding affinity, would elucidate the lipid-binding modes.

The production of NO and IL-6 by bone marrow-derived macrophages (BMMφ), which express Mincle and MCL, was reportedly changed by stimulation with several lengths of acyl chains, revealing the importance of acyl-chain length (31). Fungal

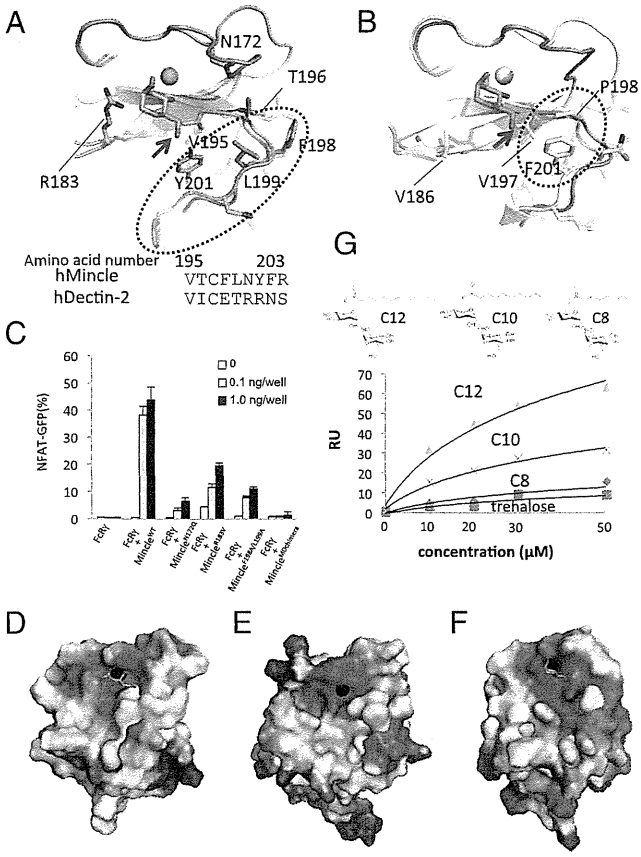


Fig. 4. Unique amino acid residues in MCL and Mincle, reporter assays of mutant Mincle, and SPR analysis. (*A* and *B*) The superimposed structures of Mincle (cyan) and DC-SIGNR (*A*) and MCL (yellow) and DC-SIGNR (pink) (*B*) are shown. Arrows indicate the oxygen atom connected with mycolic acid in TDM (mannose binding to DC-SIGNR is shown). Dotted circles indicate the hydrophobic loops found in Mincle and MCL. A sequence comparison between human Mincle and Dectin-2 is shown. (*C*) Analyses of Mincle and its mutants were performed. Reporter cells expressing human Mincle or its mutants were stimulated with TDM for 18 h. Error bar indicates ± standard deviation of three independent experiments. (*D–F*) Electrostatic potentials of Mincle (*D*), MCL (*E*), and DC-SIGNR (*F*) are shown. Electrostatic surface potentials were calculated using the program APBS (45) and are represented by PyMOL (www.pymol.org), with the color of the surface potentials in the scale ranging from negatively charged (–4.0 kbT/ec; red) to positively charged amino acids (4.0 kbT/ec; blue). Black spheres are Ca^{2+} ions. The yellow surface indicates the hydrophobic site. (*G*) SPR analysis of Mincle and several lengths of acyl chains with trehalose was performed. The C12, C10, and C8 glycolipids used in this experiment are shown. RU, response units.

glycolipids, recently identified as Mincle ligands, have more complicated and branched lipid moieties. The structural and functional data presented here showed that Mincle and MCL probably require an acyl chain longer than 10 carbons for glycolipid recognition, thus clearly providing important clues for the design of better adjuvants than TDM.

The present study indicated that Mincle has a higher affinity for TDM than MCL, which is consistent with our in vitro binding study (Fig. 3 *F* and *G*). The crystal structures of MCL and Mincle clearly revealed that MCL has a smaller hydrophobic patch next to the putative Ca^{2+}-mediated sugar binding site compared with that in Mincle. The different sizes of these hydrophobic sites might explain the affinity differences of the two CLRs observed in the binding data.

Typical CLRs that simply recognize sugars, such as DC-SIGNR and CEL-IV, exhibit remarkably low affinities (K_d ~mM) (32, 33). They require multiple valencies of sugar ligands to mediate signaling. However, SPR analysis revealed that Mincle seemed to show higher affinity, suggesting that it can detect small numbers of glycolipids on fungal surfaces. On the other hand, MCL showed much lower affinity than Mincle but essentially the same ligand specificity. It is plausible that MCL-mediated signaling requires multiple valencies of glycolipid ligands. Therefore, Mincle and MCL may play distinct roles in physiological events.

CLRs often form homodimers or heterodimers on the cell surface. As described above, the multivalent ligands on the bacterial surface likely induce the multimerization of CLRs (either monomeric or dimeric structures), which may mediate efficient signaling. The recent report by Lobato-Pascual et al. demonstrated that Mincle and MCL form a disulfide-linked heterodimer associated with the FcRγ chain (34). The heterodimeric complex formation between Mincle and MCL through the N-terminal β-strand and/or stalk regions, as previously reported for maltose-binding protein (35), for efficient recognition/signaling would be an intriguing issue to be addressed.

Materials and Methods

Plasmid Construction. *E. coli* expression plasmids encoding the partial extracellular domain of human Mincle (residues 74–219), pET22-Mincle, and the extracellular domain (residues 61–215) of human MCL, pET22-MCL, were constructed (details are available in *SI Materials and Methods*) (36).

To improve the solubility and crystallization of Mincle, we synthesized, purified, and crystallized several mutated Mincle proteins. Among them, the I99K mutant was produced with a high yield and generated good crystals.

Preparation of Recombinant Proteins. pET22-Mincle, pET22-Mincle I99K, and pET22-MCL plasmids were transformed into *E. coli* strain BL21(DE3) pLysS, and the protein was obtained as inclusion bodies. The protein was solubilized in a buffer containing 6 M guanidine·HCl, 50 mM MES (pH 6.5), 100 mM NaCl, and 10 mM EDTA for 12 h at 4 °C. One hour after the addition of DTT (10 mM), the solubilized proteins were slowly diluted into 1 L of buffer containing 0.1 M Tris·HCl (pH 8.5), 1 M L-arginine, 2 mM EDTA, 6.3 mM cystamine, 3.7 mM cysteamine, and 0.1 mM phenylmethylsulfonyl fluoride. The refolding mixture was purified by gel-filtration chromatography. The buffer was finally exchanged to 20 mM Tris·HCl (pH 8.0), with 5 mM $CaCl_2$ for crystallization.

Crystallization and Structure Determination. Crystals of purified Mincle I99K and MCL were grown at 20 °C [reservoir solutions: 1 M lithium chloride, 0.1 M citric acid, pH 4, 20% (wt/vol) PEG 6000, and 0.1 M bis-Tris propane, pH 6.5, 0.2 M potassium thiocyanate, 20% (wt/wt) PEG 3350, respectively] by the hanging-drop vapor-diffusion method. Crystals were equilibrated in a cryoprotectant consisting of reservoir solution supplemented with 16% (vol/vol) glycerol. X-ray data were collected on beamlines BL32XU in SPring-8 and BL5A in KEK. The data were processed with HKL2000 (37) or XDS (38). The structure was solved by molecular replacement with Phaser (39), using CD69 as the search model (PDB ID code 1FM5). Several rounds of model building in Coot (40) and refinement in PHENIX (41) were performed. The final refinement statistics are provided in Table S1. The coordinates for the refined Mincle, Mincle–citrate complex, and MCL structures have been deposited in the Protein Data Bank (ID codes 3WH3, 3WH2, and 3WHD, respectively).

Binding Assay Using Ig-Fusion Proteins. The MCL–Ig and Mincle–Ig fusion proteins were prepared as described previously. Briefly, the C terminus of the extracellular domain of human MCL (residues 42–215), human Mincle (residues 46–219), or their mutants was fused to the N terminus of the hIgG1 Fc region. The Ig-fusion proteins were incubated with 0.2 μg per well of plate-coated TDM or plate-coated anti-human IgG, and the bound proteins were detected by using HRP-labeled anti-human IgG.

Reporter Assay. Reporter cells were prepared as described previously (11, 36). Briefly, 2B4-NFAT-GFP reporter cells were transfected with FcRγ, together with Mincle and the mutants. The reporter cells were stimulated with various concentrations of plate-coated TDM or anti-human Mincle antibody (13D10-H11). The activation of NFAT-GFP was monitored by flow cytometry.