Table 2 Number of polymorphic sites and non-synonymous sites in ULBP2 genes in macaques and human

Locus	Number of alleles	Exon 2		Intron 2	Exon 3	
		Polymorphic sites	Non-synonymous sites (%)	Polymorphic sites	Polymorphic sites	Non-synonymous sites (%)
Mamu ULBP2.1	15	10	10 (100)	14	7	4 (57.1)
Mafa ULBP2.1	11	8	7 (87.5)	12	6	3 (50.0)
Mamu ULBP2.2	11	7	6 (85.7)	9	14	9 (64.3)
Mafa ULBP2.2	10	17	7 (41.2)	13	12	9 (75.0)
Human ULBP2	NC	17	14 (82.3)	0	15	11 (74.0)
Human ULBP6	NC	14	13 (92.9)	1	17	11 (64.7)

NC not counted because the polymorphisms in the public databases are indicated for each site and not for full sequence

Identification of alleles for another ULBP2 gene, *ULBP2.2*

PCR products for *ULBP2.2* could be obtained from genomic DNAs of both rhesus and cynomolgus macaques. Sequencing data from the cloned PCR products of 1,080–1,085 bp were classified into 11 different alleles, *Mamu-ULBP2.2*1* to *-ULBP2.2*11*, from the rhesus macaques and 10 alleles, *Mafa-ULBP2.2*1* to *-ULBP2.2*10*, from the cynomolgus

a

```
Mamu-ULBP2.1*1     1:PRWSQTGGDD PHSLCYDITI IPKFRPGPQW CAVQGQVDKK TFLHYDCGNK TVTSVSPLGK KLNVTKAWKA QNPVLREVVD MLTEQLLDIQ LENYTPREPL 100
Mamu-ULBP2.1*2       :********** ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
Mamu-ULBP2.1*3       :********** ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------E ----------
Mamu-ULBP2.1*4       :********** ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
Mamu-ULBP2.1*5       :********** ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
Mamu-ULBP2.1*6       :********** L--------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
Mamu-ULBP2.1*7       :********** ---------- --------R- ---------- ---------- ---------- ---------- --T------- ---------- ----------
Mamu-ULBP2.1*8       :********** ---------- ---------- ---------- ---------- -----T---- ---------- ---------- ---------- ----------
Mamu-ULBP2.1*9       :********** L--------- ---------- ---------- ---------- ----I-T--- ---------- ---------- ---------- ----------
Mamu-ULBP2.1*10      :********** L--------- ---------- Y--------- ---------- -----T---- ---------- ---------- ---------- ----------
Mamu-ULBP2.1*11      :********** ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
Mamu-ULBP2.1*12      :********** ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
Mamu-ULBP2.1*13      :********** ---------- ---------- ---------- ---------- ---P------ ---------- ---------- ---------- ----------
Mamu-ULBP2.1*14      :********** ---------- ---------- ---------- ---------- ---------- Q--------- ---------- ---------- ----------
Mamu-ULBP2.1*15      :********** ---------- ---------- ---------- ---------- ---------- Q--------- ---------- ---------- ----------
Mafa-ULBP2.1*1       :********** ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
Mafa-ULBP2.1*2       :********** ---------- ---------- ----D----- ---------- ---------- ---------- ---------- ---------E ----------
Mafa-ULBP2.1*3       :********** ---------- ---------- ----D----- ---------- ---------- ---------- ---------- ---------E ----------
Mafa-ULBP2.1*4       :********** ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
Mafa-ULBP2.1*5       :********** L--------- --------R- ---------- ---------- ---P------ ---------- ---------- ---------- ----------
Mafa-ULBP2.1*6       :********** L--------- ---------- ---------- ---------- ------T--- ---------- ---------- ---------- ----------
Mafa-ULBP2.1*7       :********** L--------- ---------- ---------- ---------- ------T--- ---------- ---------- ---------- ----------
Mafa-ULBP2.1*8       :********** L--------- ---------- ---------- ---------- ------T--- ---------- ---------- ---------- ----------
Mafa-ULBP2.1*9       :********** L-------I- ---------- ---------- ---------- ------T--- ---------- ---------- ---------- ----------
Mafa-ULBP2.1*10      :********** ---------- --------R- ----D----- ---------- ------T--- ---------- ---------- ---------- ----------
Mafa-ULBP2.1*11      :********** ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------E ----------
Mamu-ULBP2.2*1       :SL--RA-RA- L-----E--- --------RR- ---------- ---------- I--P------ --S------- ---------- ---------- ----------
Mamu-ULBP2.2*2       :********** L-----E--- --------RR- ---------- ---------- I--P----V- -FS------- ---------- ---------- ----------
Mamu-ULBP2.2*3       :********** L-----E--- -------RR- ---------- ---------- I--P------ -FS------- ---------- ---------- ----------
Mamu-ULBP2.2*4       :********** L-----E--- -------RR- ---------- ---------- I--------- -FS------- ---------- ---------- ----------
Mamu-ULBP2.2*5       :********** L-----E--- -------RR- ---------- ---------- I--PI----- -FS------- ---------- ---------- ----------
Mamu-ULBP2.2*6       :********** L-----E--- -------RR- ---------- ---------- I--PI---V- -FS------- ---------- ---------- ----------
Mamu-ULBP2.2*7       :********** L-----E--- -------RR- ---------- ---------- I--PI----- -FS------- ---------- ---------- ----------
Mamu-ULBP2.2*8       :********** L-----E--- --------RR- ---------- ---------- I--P------ -FS------- ---------- ---------- ----------
Mamu-ULBP2.2*9       :********** L-----E--- --------RR- ---------- ---------- ---P------ -FS------- ---------- ---------- ----------
Mamu-ULBP2.2*10      :********** L-----E--- -------RR- ---------- ------Y--- I--P------ -LS------- ---------- ---------- ----------
Mamu-ULBP2.2*11      :********** L-----E--- -------RR- ---------- ---------- I--P------ -LS------- ---------- ---------- ----------
Mafa-ULBP2.2*1       :********** L-----E--- -------RR- ---------- ---------- I--P------ -LS------- ---------- ---------- ----------
Mafa-ULBP2.2*2       :********** L-----E--- -------RR- ---------- ---------- I--P------ -FS------- ---------- ---------- ----------
Mafa-ULBP2.2*3       :********** L-----E--- -------RR- ---------- ---------- I--P------ -FS------- ---------- ---------- ----------
Mafa-ULBP2.2*4       :********** L-----E--- -------RR- ---------- ---------- I--P------ -FS------- ---------- ---------- ----------
Mafa-ULBP2.2*5       :********** L-----E--- --------RR- ---------- ---------- I--P------ -FS------- ---------- ---------- ----------
Mafa-ULBP2.2*6       :********** L-----E--- -------RR- ---------- ---------- I--P------ --S------- ---------- ---------- ----------
Mafa-ULBP2.2*7       :********** L-----E--- -------RR- ---------- ---------- I--P------ --S------- ---------- ---------- ----------
Mafa-ULBP2.2*8       :********** ---------- -------PR- ---------- ---------- ---P--T--- ---------- ---------- ---------- ----------
Mafa-ULBP2.2*9       :********** L-----E--- -------RR- ---------- ---------- I--P------ -FS------- ---------- ---------- ----------
Mafa-ULBP2.2*10      :********** L-----E--- -------RR- ---------- ---------- I--P------ -FS------- ---------- ---------- ----------
Human ULBP2          :SG--RA-RA- ------D--V -------PR- --------E- ---------- ---P------ -----T---- ---------- I-----R--- ------K---
Human ULBP6          :FG--RARR-- ------D--V -------PR- --------E- ---------- ---P------ -----M---- ---------- I--------- ------K---
```

Fig. 2 Alignment of deduced amino acid sequences of α1 and α2 domains of ULBP2-related molecules from rhesus and cynomolgus macaques and human. Amino acid sequences deduced from nucleotide sequences for *Mamu-ULBP2.1*1* and alleles of *Mamu-ULBP2.1*, *Mafa-ULBP2.1*, *Mamu-ULBP2.2*, and *Mafa-ULBP2.2*, were aligned with human *ULBP2* (AY026825), and *ULBP6* (AL355497). *Numbers* represent the amino acid positions in mature protein. Sequences for the predicted α helix structure are *underlined* and *contact sites* with NKG2D are *boxed* in the *Mamu-ULBP2.1*1* sequences. *Dashes* indicate identity to the *Mamu-ULBP2.1*1* sequences. *Asterisks* represent not sequenced regions. Amino acid sequences are shown **a** from 1 to 100 and **b** from 101 to 191

b

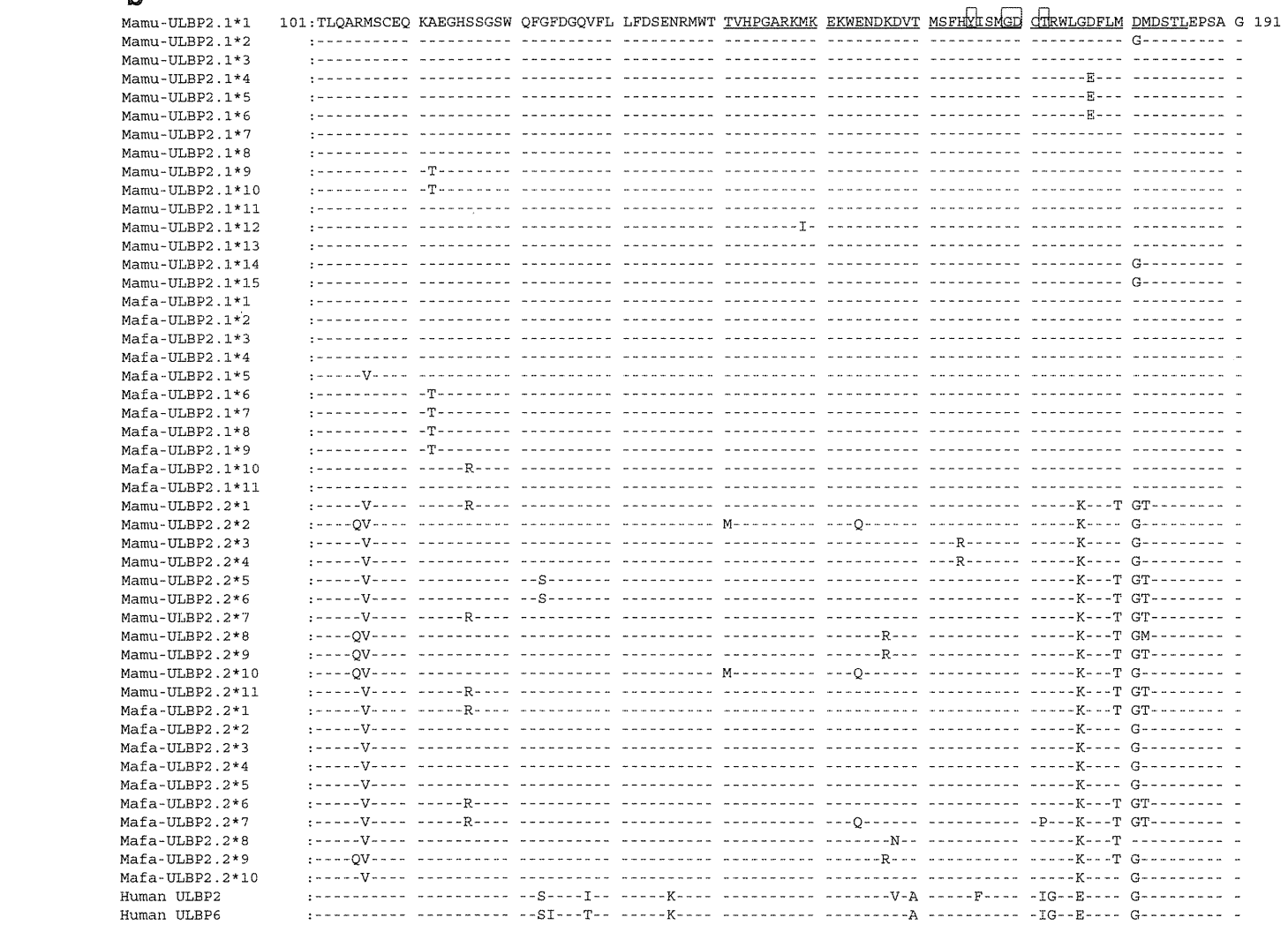


Fig. 2 continued.

macaques (Table 1). In this study, we found a repeat number polymorphism in the A stretch in intron 2 in both *ULBP2.1* and *ULBP2.2* from both rhesus and cynomolgus macaques. Polymorphisms in exons 2 and 3 and intron 2 including the repeat polymorphism were included in the allele designation. Nucleotide sequences of *Mamu-ULBP2.2*1* and *Mafa-ULBP2.2*1* were identical to those of LOC694600. That the identical *ULBP2* alleles were shared in part by both rhesus and cynomolgus macaques was consistent with a cross-breeding between these macaques as suggested by the studies of diversity in the MHC class I genes (Saito et al. 2012).

Divergence of *ULBP2* genes in the higher primates

Nucleotide sequence homologies among the alleles of *ULBP2.1* and *ULBP2.2* in rhesus and cynomolgus macaques were 94.3 and 98.7 %, respectively, suggesting that *ULBP2.2* is less diverged than *ULBP2.1*. To figure out the evolutional divergence and diversity of *ULBP2* genes in the higher primates, we conducted a neighbor-joining analysis by using nucleotide sequences of exons 2 and 3 from *ULBP2.1* and *ULBP2.2* in macaques along with sequences of corresponding region from *ULBP2* and/or *ULBP6* reported for human, chimpanzee, gorilla, and another Old World monkey, olive baboon. *ULBP5* sequences from human and rhesus were also included in the analysis. As shown in Fig. 1, alleles of *ULBP2.1* and those of *ULBP2.2* in macaques were separately clustered, but both rhesus and cynomolgus alleles were found in the same cluster. Quite interestingly, *ULBP2.2* of macaques including olive baboon *ULBP2* and rhesus *ULBP5* appeared to be diverged from an ancestral *ULBP2.1*. In addition, *ULBP2* sequences from human, gorilla and chimpanzee as well as *ULBP5* and *ULBP6* sequences from human were clustered as a branch of *ULBP2.1* (Fig. 1). Furthermore, when other *ULBP* genes, *ULBP1*, *ULBP3*, and *ULBP4*, were included in the phylogenetic analysis, these genes were also clustered as

another branch of *ULBP2.1*, implying that these genes were diverged from the ancestral *ULBP2* in the primates (Supplementary Figure S1).

Diversity of *ULBP2* genes in the Old World Monkey

As for the diversity of *ULBP2* in macaques, non-synonymous substitutions were found at 14 sites in *Mamu-ULBP2.1*, 10 sites in *Mafa-ULBP2.1*, 15 sites in *Mamu-ULBP2.2*, and 16 sites in *Mafa-ULBP2.2* (Table 2). Amino acid sequences were deduced from the nucleotide sequences and alignment of the ULBP2 alleles showed that the ULBP2/RAET1H molecules in rhesus and cynomolgus macaques were homologous by more than 90 % to the human ULBP2 molecule in the α1 and α2 domains (Fig. 2). Among the polymorphic amino acid residues, six and eight residues in the ULBP2.1 and ULBP2.2 molecules, respectively, were observed in both rhesus and cynomolgus macaques (Fig. 2).

To investigate a possible role of the polymorphic residues, we created 3D structure models for ULBP2 molecules by referring the crystallographic data for human ULBP3 in complex with NKG2D, where ULBP2 residues at positions 22, 23, 70, 73, 77, 80, 83, 94, 165, 169, 170, and 172 composed of interacting surface with NKG2D (Supplementary Figure S2). It was found that three polymorphic residues at positions 73, 177, and 181 of rhesus ULBP2.1 were on the upper surface of α helix structures and pointed to the NKG2D receptor (Fig. 3a). Interestingly, the residue at position 73 could be a contacts site with Ser195 of NKG2D receptor, as deduced from the equivalent structure of human ULBP3. In contrast, none of the polymorphic residues were mapped on the surface of α helix structures in cynomolgus ULBP2.1 (Supplementary Figure S3).

On the other hand, two polymorphic residues at positions 158 and 181 of cynomolgus ULBP2.2 were on the upper surface of the α helix structure, while another polymorphic residue at position 172 was a possible interface site with Glu183 and Met184 of NKG2D in the equivalent human ULBP3, although it was not pointed up on the α helix of ULBP (Fig. 3b). In clear contrast and quite interestingly, none of the polymorphic residues were mapped on the surface of α helix in rhesus ULBP2.2 (Supplementary Figure S4). It should be noted here that the residue at position 20 is polymorphic in human ULBP2, while the residues at positions 80 and 172 are polymorphic in human ULBP6, indicating that both ULBP2 and ULBP6 might carry the allelic differences in the interaction with NKG2D in humans (Supplementary Figure S5).

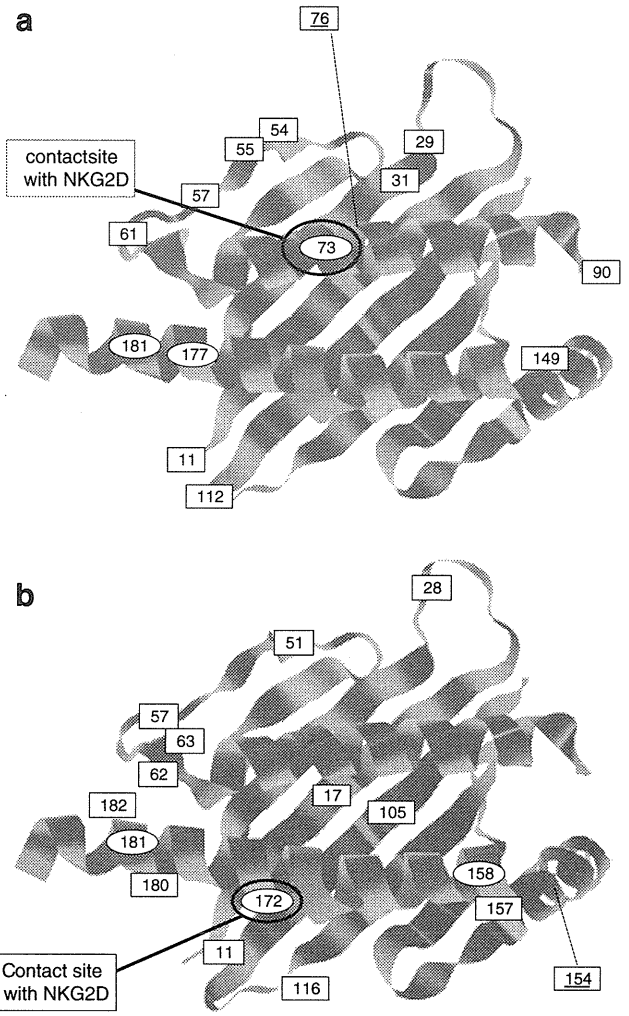


Fig. 3 Mapping of polymorphic sites on the 3D-structure model of macaque ULBP2 molecule. Polymorphic sites were mapped on the 3D structure model of ULBP2/RAET1H. Positions of polymorphic amino acid resides in the rhesus ULBP2.1 molecule (**a**) and cynomolgus ULBP2.2 molecule (**b**). Residues on the upper side of α helix are indicated by *circles*, while those on the outer side or on the β seat are indicated by *squares*. The residues mapped behind or beneath the α helix are underlined and represented by *dotted lines*. Possible contact sites with NKG2D (Radaev et al. 2001) are indicated

Discussion

In this study, we investigated the polymorphic nature of *ULBP2/RAET1H* in the Old World monkey. We previously reported that each member of the *ULBP/RAET1* gene family, except for *ULBP6/RAET1L*, was duplicated in the rhesus genome (Naruse et al., 2011). As expected, we obtained *ULBP2.1* and *ULBP2.2* sequences from both rhesus and cynomolgus macaques. On the other hand, any orthologous genes to human *ULBP6* were not detected in the macaques, even though *ULBP6* showed 96 % homology to *ULBP2* in humans (Radosavljevic et al. 2001). It was considered that *ULBP2.1* or *ULBP2.2* might be orthologous to human *ULBP6*. However, our phylogenetic analysis indicated that both human *ULBP2* and *ULBP6* were clustered with *ULBP2*s from chimpanzee and gorilla, as a branch of *ULBP2.1*. In addition, intron 2 sequences of ULBP2.2 in

macaques were relatively well conserved among *ULBP2*s from olive baboon, chimpanzee, western gorilla, and human as well as in human *ULBP6* than exon sequences. Furthermore, a Blast search showed that there was no *ULBP6*-like sequence in the genomes of chimpanzee and gorilla. These observations suggested that human *ULBP2* and *ULBP6* were diverged from an ancestral *ULBP2* after the diversification of human and other higher primates. The phylogenetic analysis also indicated that *ULBP2.2* might be diverged from *ULBP2.1* and the clustering of *ULBP2.1* alleles and *ULBP2.2* alleles did not depend on the species, supporting a trans-species evolution. Our observations were consistent with that ULBP/REAT molecule of placental mammals was originally diverged and duplicated in each species after an emigration from the *MHC* region (Kondo et al. 2010).

On the other hand, *Mafa-ULBP2.1*5* was placed at the diverging point of *ULBP2.2* in the phylogenetic tree (Fig. 1 and Supplementary Figure S1). As shown in Fig. 2, among the *ULBP2.1* alleles, only the *Mafa-ULBP2.1*5* has a replacement of Arg with Val at the position 105, which is a common feature of *Mamu-ULBP2.2* and *Mafa-ULBP2.2*. In addition, *Mafa-ULBP2.1*5* carries *ULBP2.2*-like sequences at the positions of 29 and 54. These characteristic features may eventually position the *Mafa-ULBP2.1*5* at the diverging point of *ULBP2.2*. Alternatively, gene conversion-like events from *ULBP2.2* had occurred in *ULBP2.1* to generate the *Mafa-ULBP2.1*5*.

In the present study, we denoted 15 and 11 *ULBP2.1* alleles and 11 and 10 *ULBP2.2* alleles in rhesus and cynomolgus macaques, respectively, of which more than 70 % of polymorphisms were non-synonymous. Induced expression of human ULBP2 molecule is involved in the recognition of virus-infected cell by NKG2D (Ward et al. 2009), although the functional significance of the polymorphisms in the extracellular domain of ULBP2 molecules remains to be deciphered. We demonstrated that several polymorphisms of ULBP2 molecules were located at the presumed contact sites with NKG2D or on the upper surface of α helix, which might have functional impacts. On the other hand, it has been reported that the sequence identities are less than 60 % among the ULBP molecules, while those between the ULBP and MIC molecules are about 25 % (Cosman et al. 2001). The interface residues appeared to be less conserved than the overall sequence among the ULBP molecules, and it was predicted that NKG2D could recognize the diversities of ULBP1 and ULBP2 molecules through induced-fit mechanisms in a similar manner as that of ULBP3 (Radaev et al. 2001). However, other previous studies revealed that the structural differences among the α2 domains of ULBP and MIC molecules affect the binding affinity to NKG2D and UL16, respectively (Wittenbrink et al. 2009; Spreu et al. 2006), suggesting that the ULBP polymorphisms demonstrated in this study might influence the efficacy of recognition by NKG2D. The functional impact of the polymorphisms should be investigated in future studies to decipher the evolutional and biological significance of the ULBP2 polymorphisms in the Old World monkeys.

Acknowledgments We thank Yukiko Ueda, Yasuko Saida, and Nana Ohkubo for their technical assistances. This work was supported in part by research grants from the Ministry of Health, Labor and Welfare, Japan and a Grant-in-Aid for scientific research from the Ministry of Education, Culture, Sports, Science, and Technology (MEXT), Japan. This work was also supported by a supporting program for women researchers from the Tokyo Medical and Dental University.

References

Antoun A, Jobson S, Cook M, O'Callaghan CA, Moss P, Briggs DC (2010) Single nucleotide polymorphism analysis of the NKG2D ligand cluster on the long arm of chromosome 6: extensive polymorphisms and evidence of diversity between human populations. Hum Immunol 71:610–620

Bacon L, Eagle RA, Meyer M, Easom N, Young NT, Trowsdale J (2004) Two human ULBP/RAET1 molecules with transmembrane region are ligands for NKG2D. J Immunol 173:1078–1084

Bauer S, Groh V, Wu J, Steinle A, Phillips JH, Lanier LL, Spies T (1999) Activation of NK cells and T cells by NKG2D, a receptor for stress-inducible MICA. Science 285:727–729

Chalupny NJ, Sutherland CL, Lawrence WA, Rein-Weston A, Cosman D (2003) ULBP 4 is a novel ligand for human NKG2D. Biochem Biophys Res Commun 305:129–135

Cosman D, Mullberg J, Sutherland CL, Chin W, Armitage R, Fanslow R, Kubin M, Chalupny NJ (2001) ULBPs, novel MHC class I-related molecules, bind to CMV glycoprotein UL16 and stimulate NK cytotoxicity through the NKG2D receptor. Immunity 14:123–133

Eagle RA, Traherne JA, Ashiru O, Wills MR, Trowsdale J (2006) Regulation of NKG2D ligand gene expression. Hum Immunol 67: 1159–1169

Eagle RA, Flack G, Warford A, Martinez-Borra J, Jafferji I, Traherne JA, Ohashi M, Boyle LH, Barrow AD, Caillat-Zucman S, Young NT, Trowsdale J (2009a) Celllar expression, trafficking, and function of two isoforms of human ULBP5/RAET1G. PLoS ONE 4:e4503

Eagle RA, Traherne JA, Hair JR, Jafferji I, Trowsdale J (2009b) ULBP6/RAET1L is an additional human NKG2D ligand. Eur J Immunol 39: 3207–3216

Gibbs RA, Rogers J, Katze MG et al (2007) Evolutionary and biomedical insights from the rhesus macaque genome. Science 316:222–234

Kondo M, Maruoka T, Otsuka N, Kasamatsu J, Fugo K, Hanzawa N, Kasahara M (2010) Comparative genomic analysis of mammalian NKG2D ligand family genes provides insights into their origin and evolution. Immunogenetics 62:441–450

Kulski JK, Anzai T, Shiina T, Inoko H (2004) Rhesus macaque class I duplicon structures, organization, and evolution within the alpha block of the major histocompatibility complex. Mol Biol Evol 21: 2079–2091

Matusali G, Tchigjou HK, Pontrelli G, Bernardi S, D'Ettorre G, Vullo V, Buonomini AR, Andreoni M, Santoni A, Cerboni C, Doria M (2013) Soluble ligands for the NKG2D receptor are released during HIV-1 infection and impair NKG2D expression and cytotoxicity of NK cells. FASEB J 27:2440–2450

Naruse TK, Chen Z, Yanagida R, Yamashita T, Saito Y, Mori K, Akari H, Yasutomi Y, Miyazawa M, Matano T, Kimura A (2010) Diversity of MHC class I genes in Burmese-origin rhesus macaque. Immunogenetics 62:601–611

Naruse TK, Okuda Y, Mori K, Akari H, Matano T, Kimura A (2011) ULBP4/RAET1E is highly polymorphic in the Old World Monkey. Immunogenetics 63:501–509

Otting N, Otting N, de Vos-Rouweler AJM, Heijmans CMC, de Groot NG, Doxiadis GGM, Bontrop RE (2007) MHC class I A region diversity and polymorphism in macaque species. Immunogentics 59:367–375

Pappworth IY, Wang EC, Rowe M (2007) The switch from latent to productive infection in Epstein-Barr virus-infected B cell is associated with sensitization to NK cell killing. J Virol 81:474–482

Pende D, Rivera P, Marcenaro S, Chang CC, Biassoni R, Conte R, Kubin M, Cosman D, Ferrone S, Moretta L, Moretta A (2002) Major histocompatibility complex class I-related chain A and UL16-binding protein expression on tumor cell lines of different histotypes: analysis of tumor susceptibility to NKG2D-dependent natural killer cell cytotoxicity. Cancer Res 62:6178–6186

Radaev S, Rostro B, Brooks AG, Colonna M, Sun PD (2001) Conformational plasticity revealed by the cocrystal structure of NKG2D and its class I MHC-like ligand ULBP3. Immunity 5: 1039–1049

Radosavljevic M, Cuillerier B, Wilson MJ, Clement O, Wicker S, Gilfillan S, Beck S, Trowsdale J, Bahram S (2001) A cluster of ten novel MHC class I related genes on human chromosome 6q24.2-q25.3. Genomics 79:114–123

Raulet DH (2003) Roles of the NKG2D immunoreceptor and its ligands. Nat Rev Immunol 3:781–790

Richard J, Phan TNQ, Ishizuka Y, Cohen EA (2013) Viral protein R upregulates expression of ULBP2 on uninfected bystander cells during HIV-1 infection of primary CD4+ T lymphocytes. Virology 443:248–256

Romphruk AV, Romphruk A, Naruse TK, Raroengjai S, Puapairoj C, Inoko H, Leelayuwat C (2009) Polymorphisms of NKG2Dligands: Diverse RAET1/ULBP genes in northeastern Thais. Immunogenetics 61:611–617

Saito Y, Naruse TK, Akari H, Matano T, Kimura A (2012) Diversity of MHC class I haplotypes in cynomolgus macaques. Immunogenetics 64:131–141

Spreu J, Stehle T, Steinle A (2006) Human cytomegalovirus-encoded UL16 discriminates MIC molecules by their α2 domains. J Immunol 177:3143–3149

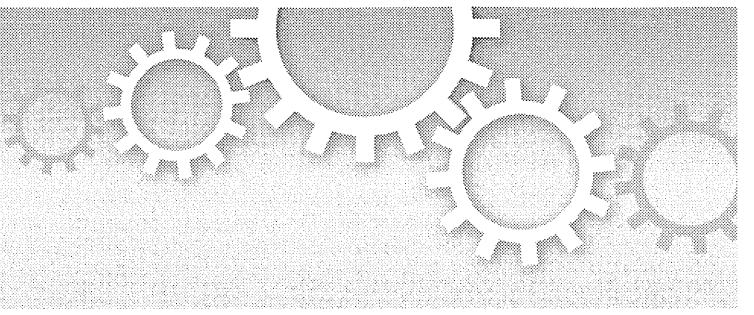
Ward J, Bonaparte M, Sacks J, Guterman J, Fogli M, Mavilio D, Barker E (2007) HIV modulates the expression of ligands important in triggering natural killer cell cytotoxic responses on infected primary T-cell blasts. Blood 110:1207–1214

Ward J, Davis Z, DeHart J, Zimmerman E, Bosque A, Brunette E, Mavilio D, Placeless V, Barker E (2009) HIV-1 PR triggers natural killer cell-mediated lysis of infected cells through activation of the ATR-mediated DNA damage response. PLoS Pathogen 5:e1000613

Wittenbrink M, Spreu J, Steinle A (2009) Differential NKG2D binding to highly related human NKG2D ligands ULBP2 and RAET1G is determined by a single amino acid in the α2 domain. Eur J Immunol 39:1642–1651

Wu J, Song Y, Bakker ABH, Bauer S, Spies T, Lanier LL, Phillips JH (1999) An activating immunoreceptor complex formed by NKG2D and DAP 10. Science 285:730–732

SCIENTIFIC REPORTS



OPEN

SUBJECT AREAS:
VIRAL HOST RESPONSE
HIV INFECTIONS
ADAPTIVE IMMUNITY

Received
15 July 2013

Accepted
15 October 2013

Published
6 November 2013

Correspondence and requests for materials should be addressed to A.I. (aikichi@ra3.so-net.ne.jp; aikichi@ims.u-tokyo.ac.jp)

Structure of TCR and antigen complexes at an immunodominant CTL epitope in HIV-1 infection

Akihisa Shimizu[1], Ai Kawana-Tachikawa[1], Atsushi Yamagata[2,3], Chungyong Han[1], Dayong Zhu[4], Yusuke Sato[2,3], Hitomi Nakamura[4,5], Tomohiko Koibuchi[5], Jonathan Carlson[6], Eric Martin[7], Chanson J. Brumme[8], Yi Shi[9], George F. Gao[9], Zabrina L. Brumme[7,8], Shuya Fukai[2,3,10] & Aikichi Iwamoto[1,4,5]

[1]Division of Infectious Diseases, Advanced Clinical Research Center, the Institute of Medical Science, the University of Tokyo. 4-6-1 Shirokanedai, Minato-ku, Tokyo 108-8639, Japan, [2]Structural Biology Laboratory, Life Science Division, Synchrotron Radiation Research Organization and Institute of Molecular and Cellular Biosciences, the University of Tokyo. 1-1-1, Yayoi, Bunkyo-ku, Tokyo 113-0032, Japan, [3]Department of Medical Genome Sciences, Graduate School of Frontier Sciences, The University of Tokyo. 5-1-5 Kashiwanoha, Kashiwa-shi, Chiba 277-8561, Japan, [4]Department of Infectious Disease Control, the International Research Center for Infectious Diseases, the Institute of Medical Science, the University of Tokyo. 4-6-1 Shirokanedai, Minato-ku, Tokyo 108-8639, Japan, [5]Department of Infectious Diseases and Applied Immunology, Hospital, the Institute of Medical Science, the University of Tokyo. 4-6-1 Shirokanedai, Minato-ku, Tokyo 108-8639, Japan, [6]Microsoft Research, Los Angeles, CA, USA. 1100 Glendon Ave PH-1. Los Angeles, CA 90024, USA, [7]Faculty of Health Sciences, Simon Fraser University. 8888 University Drive, Burnaby, BC, Canada V5A 1S6, [8]British Columbia Centre for Excellence in HIV/AIDS. 608–1081 Burrard Street, Vancouver, BC, Canada V6Z 1Y6, [9]CAS Key Laboratory of Pathogenic Microbiology and Immunology, Institute of Microbiology, Chinese Academy of Sciences. No. A3 Datun Road, Chaoyang District, Beijing 100101, China, [10]JST, CREST, 1-1-1, Yayoi, Bunkyo-ku, Tokyo 113-0032, Japan.

We investigated the crystal structure of an HLA-A*2402-restricted CTL epitope in the HIV-1 *nef* gene (Nef134-10) before (pHLA) or after TCR docking. The wild type epitope and two escape mutants were included in the study. Y135F was an early-appearing major mutation, while F139L was a late-appearing mutation which was selected in the patients without Y135F. F139 was an eminent feature of the Nef134-10 epitope. Wild type-specific TCR was less fit to F139L mutant suggesting that F139L is an escape from the CTL against the wild type epitope. Although Y135F mutation disrupted the hydrogen bond to HLA-A*2402 His70, newly formed hydrogen bond between T138 and His70 kept the conformation of the epitope in the reconstituted pMHC. TCR from Y135F- or dually-specific CTL had unique mode of binding to the mutant epitope. Y135F has been reported as a processing mutant but CTL carrying structurally adequate TCR can be found in the patients.

Cytotoxic T lymphocytes (CTL) can exert an efficient control on HIV-1 replication if HIV-1 antigens are presented on the surface of infected cells and properly recognized by the heterodimeric T cell receptors (TCR) attached to the CTL[1–4]. For this to occur, viral proteins synthesized in the infected cells must first be digested to fragments, transported to the rough endoplasmic reticulum (ER), and bound within a groove formed by two α-helices of the major histocompatibility complex (HLA in human) class I molecules as a 8–10-mer peptide[5–7]. The peptide-bound HLA class I molecules (pHLA) are then exported to the cell surface for recognition by TCR on the CTL cell membranes[8,9]. The encounter between TCR and pHLA is the fundamental requirement and initial step that allows CTL to deliver their cytotoxicity.

The genetic hypermutability of HIV-1 can interfere with immune surveillance by CTL by introducing amino acid changes that allow the virus to "escape" immune recognition. Escape phenomena may occur at different steps during the process of antigen presentation and recognition. Mutations can result in changes in processing the viral proteins or in the way such that the processed peptides cannot bind to the HLA molecules or the pHLA cannot interact properly with TCR. For example, a CTL epitope in the HIV-1 *nef* gene, Nef134-10 (RYPLTFGWCF), is highly immunogenic in HLA-A*2402 (HLA-A24)-positive patients[10,11]. At a very early phase of primary HIV-1 infection, a Tyr-to-Phe mutation at the 2nd position of the Nef134-10 epitope (Y135F; Nef134-10(2F)) is frequently selected. HLA-A24 is the most prevalent HLA class I allele among Japanese[12].

In an earlier study of pHLA/TCR interactions, we used double staining with Sendaivirus-derived pHLA-tetramers to differentiate three classes of CD8+ T cells in the peripheral mononuclear cells (PBMC) of HLA-A24-positive patients with chronic HIV-1 infection: Nef134-10(wt)-specific, Nef134-10(2F)-specific and dual-specific (reacting to both wt and 2F epitopes)[13,14]. Since the encounter between TCR and pHLA is critical for CTL activation, structural studies examining the interactions between pHLA and TCR should be highly relevant in understanding the immune response to HIV-1 infection, including viral escape from immune surveillance. However, very few crystal structures relevant to escape mutations have been solved to date. To gain insights into the battle between HIV-1 and cellular immune responses we established CTL clones representing each of the three classes of antigen specificity targeting the Nef134-10 epitope, reconstituted the pHLA/TCR interactions *in vitro*, and solved the crystal structures. In the small number of patients without Y135F mutation, F139L mutation was selected. Since a wild type specific TCR showed a substantial affinity to the epitope with F139L mutation (Nef134-10(6L)), we also solved the crystal structure of the pHLA/TCR interaction between them.

Results

Characterization of escape mutations and specific T cell responses by chromium release and sequencing. Before initiating studies on the interactions between the epitope and T cell receptors, we examined the prevalence of HLA-A24-related mutations in the Nef134-10 epitope in two independent HIV-1 cohorts: the IMSUT cohort (188 Japanese HIV-1-positive individuals in Tokyo) and the HOMER cohort (1018 HIV-1-positive individuals in British Columbia, Asian prevalence of 5%). The results confirmed the association between HLA-A24 and the Y135F mutation (Supplementary Fig. 1a, b). In both cohorts Y135F and F139L mutations were each significantly associated with the presence of HLA-A24 ($p < 0.0001$ and $p = 0.0017$, respectively). In a multicenter longitudinal acute/early infection cohort of 16 individuals expressing HLA-A24, we reconfirmed the previous finding that Y135F is selected very early after the estimated date of infection, while F139L appeared late ($p = 0.0046$) (Supplementary Fig. 1c).

Next we established cellular clones representing dual-specific, Nef134-10(wt)-specific, Nef134-10(2F)-specific CTL from HLA-A24-positive patients in the IMSUT cohort. We showed previously that TCR repertoire of the dual-specific CD8+ T cells was highly restricted[14]. TRBV4-1 and TRAV8-3 gene segments were used almost exclusively as TCR β and α chains, respectively. C1-28, which used TRBV4-1 and TRAV8-3 was established and analyzed further as a CTL clone representing the dual-specific population. The TCR repertoire of the wild type specific CD8+ T cells was more diverse than the dual-specific population, but TRBV7-9 was used in more than 25% of the population. CTL clone H27-14 which used TRBV7-9 was established and represented the wild type-specific population in this study. Nef134-10(2F)-specific CD8+ T cell population was rare even after ex vivo stimulation with the cognate peptide, however, we could establish CTL clone T36-5 which represented Nef134-10(2F)-specific population. Specificities of the clones were examined by conventional chromium release assay (Fig. 1a–c).

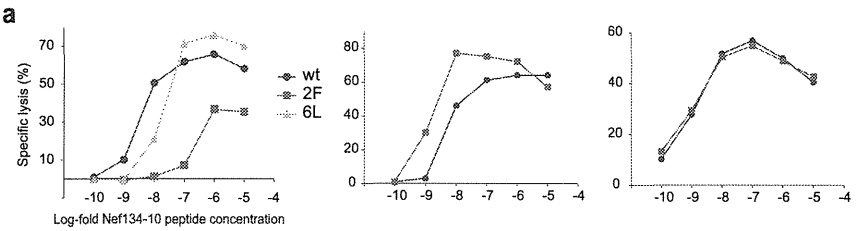
Binding characteristics of purified TCR and pHLA. The heavy chain and β2-microglobulin were expressed in *E. coli* and synthesized cognate peptides were refolded and pHLA was purified as previously described[15]. TCR derived from each CTL clone were cloned, expressed and purified from inclusion bodies of *E. coli*[16]. Refolded TCR and pHLA complexes were subjected to the surface plasmon resonance (SPR) assay (Biacore™).

Binding affinities observed between the pHLAs and TCRs are consistent with the results of the cellular killing assay (Table S3 and Supplementary Fig. 2), with the dissociation constant (KD) of the H27-14 TCR/pHLA complex increasing in the following order: A24/Nef134-10(wt) < A24/Nef134-10(6L) < A24/Nef134-10(2F). Functional TCRs against cognate epitopes have an estimated affinity ranging from 1–100 μM[17]. A CTL population with H27-14 TCR could recognize HIV-1-infected cells expressing the wild type A24/Nef134-10 epitope very well and could also recognize cells expressing A24/Nef134-10(6L), although less efficiently. However, H27-14 TCR affinity was too low to recognize A24/Nef134-10(2F). T36-5 TCR was highly specific against A24/Nef134-10(2F) but the TCR could recognize also with A24/Nef134-10(wt). C1-28 TCR bound almost equally with Nef134-10(wt) and Nef134-10(2F).

Kinetic analysis of H27-14 and T36-5 TCR with cognate epitopes (Table S3) yielded results that were compatible with results from equilibrium analysis. We were unable to perform kinetic analysis on C1-28 TCR due to low refolding yields.

General structures of free pHLA and TCR-pHLA. To reveal the structural differences posed by amino acid substitutions in the epitope, we determined and compared structures of A24/Nef134-10(wt), A24/Nef134-10(2F) and A24/Nef134-10(6L) before ligation to TCR (Fig. 2 and Supplementary Fig. 3a–c). The electron densities for three peptides that bound to HLA-A24 were unambiguous (Supplementary Fig. 4a–c). Large conformational changes were not seen when the Nef134-10(2F) and Nef134-10(6L) peptides in the groove of HLA-A24 were superimposed onto Nef134-10(wt) (Fig. 2a). Root mean square deviation (RMSD) values were 0.260Å

a



b

TCR	H27-14	T36-5	C1-28
Vα	21	12-2	8-3
CDR3α	AVRMDSSYKLIF	GTYNQGGKLI	AVGAPSGAGSYQLT
Vβ	7-9	27	4-1
CDR3β	ASSSWDTGELF	ASSGASHEQY	ASSPTSGIYEQY

Figure 1 | Specificities and TCR usage of the CTL clones specific to Nef134-10 epitopes. (a) Cytolytic activities of HLA-A24-restricted Nef134-10-specific CTL clones derived from HLA-A24-positive HIV-1-infected patients. Clones H27-14 (left), T36-5 (middle) and C1-28 (right) killed HLA-A24-positive target cells pulsed with log-fold dilutions of Nef134-10 peptides, Nef134-10(wt); blue, Nef134-10(2F); red, Nef134-10(6L); orange. The effector-versus-target ratio was 5 : 1. (b) Vα and Vβ gene segments and CDR3 sequences of each CTL clone.

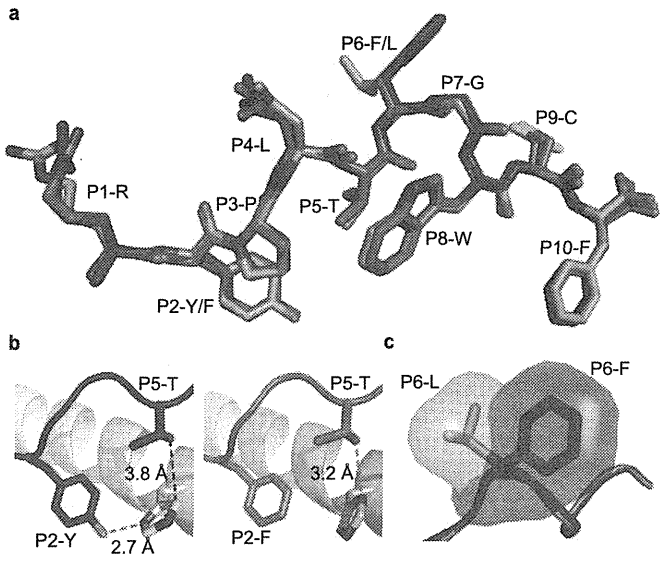


Figure 2 | Conformation of the three Nef134-10 peptides in the complex with HLA-A24. (a) Overlay of three Nef134-10 peptides (Nef134-10(wt), Nef134-10(2F) and Nef134-10(6L)) bound to the HLA-A24. Nef134-10(wt), Nef134-10(2F) and Nef134-10(6L) peptides are shown blue, red and orange stick representations, respectively. HLA-A24 binding groove are not shown for clarity. (b) Interaction between the Nef134-10 peptide and HLA-A24. P2-Tyr in the Nef134-10(wt) forms a hydrogen bond with His at position 70 in the HLA-A24 (**left**), while P5-Thr in the Nef134-10(2F) peptide forms a hydrogen bond with His70 (**right**). Red dashed lines indicate hydrogen bonds. A black dashed line indicates the shortest distance between P5-Thr in the Nef134-10(wt) and His70 in the HLA-A24. (c) The side chains of P6 in the Nef134-10(wt) and Nef134-10(6L) peptides, which protrude out of the HLA-A24 binding cleft, are shown as stick and surface representations. Nef134-10(wt) (blue) and Nef134-10(6L) (orange).

and 0.259Å, respectively. Y135(P2-Y) of the Nef134-10 (wt) peptide formed a hydrogen bond with His at position 70 of the HLA-A24 (His70), and T138(P5-T) of the Nef134-10(2F) peptide formed a potential hydrogen bond with His70 (Fig. 2b).

Side chains of L137(P4-L) and F139(P6-F) protruded from the A24 binding groove in the A24/Nef134-10, which appeared to be the eminent feature of this epitope. The solvent-accessible surface area of the side chain of Phe(F) of the Nef134-10(wt) epitope is larger than that of Leu(L) of the Nef134-10(6L) epitope by approximately 40Å^2 (188.2Å^2 for P6-F, 151.9Å^2 for P6-L) (Fig. 2c).

To understand how three TCRs (H27-14, T36-5 and C1-28) interact with their cognate pHLAs, we determined four TCR/pHLA complexes, including the H27-14 complexed to the agonist ligand, A24/Nef134-10(6L) (Supplementary Fig. 3d–g). The electron densities for CDR3 loops and peptide at the individual TCR/pHLA complex interfaces were unambiguous (Supplementary Fig. 4d–g).

Wild type-specific H27-14 TCR interacted with Nef134-10(wt), surrounding the F139(P6-F) residue by CDR loops (Fig. 3a, d). This complex exhibited a shape complementarity (Sc) of 0.75, which is at the higher end of the range of published TCR/pMHC[17]. R93α, Y98α and R31β of the H27-14 CDR rearranged to accommodate the aromatic residue of F139(P6-F) upon ligation, compared to their unligated state (Supplementary Fig. 5a, b). Y98α and R31β formed hydrogen bonds with T138(P5-T) (Supplementary Fig. 5b and Supplementary Table 3). The side chain of R93α rotated approximately 90° and formed a hydrogen bond with G100β, helping to stabilize CDRα and CDRβ at the TCR/pHLA interface.

In the T36-5-A24/Nef134-10(2F) complex, the A96 residue of CDR3β formed a hydrogen bond with the W141(P8-W) residue of Nef134-10(2F) upon ligation (Fig. 3b). The T36-5 TCR interacted with A24/Nef134-10(2F) over a wide range, with a BSA approximately 200 Å^2 higher than that of the other two complexes solved in this study (Fig. 3e and Supplementary Table 4) and with the highest affinity of the three TCRs in this study. The side chains of Y92α and H98β of the T36-5 TCR were packed against each other in the unligated state, but interacted with Nef134-10(2F) upon ligation (Supplementary Fig. 5c). Upon ligation the side chain of Q94α of the CDR3 loops moved 3.7 Å toward HLA-A24-α1 helix and contacted with the 135F(P2-F) carbonyl via a water molecule, while CDRβ loops (E30, A96 and H98) formed five hydrogen bonds with the amino acids (F6, G7, W8 and C9) located at the C-terminus of Nef134-10(2F) peptide.

In the C1-28-A24/Nef134-10(2F) complex, the CDR1α and 3β loops were located mainly above the Nef134-10(2F) peptide (Fig. 3c), whereas the CDR3α and 3β loops were located above the peptide in both H27-14-TCR-A24/Nef134-10(wt) and T36-5-TCR-A24/Nef134-10(2F) complexes (Fig. 3a, b). The side chains of Y32α and Y102α and the CDR3β loop surrounded the F139(P6-F) residue of the Nef134-10(2F) peptide. Water molecules acted as a "molecular glue" to stabilize the interface of the C1-28 TCR-A24/Nef134-10(2F) complex, but with only two hydrogen bonds between TCR and peptide in the C1-28-A24/Nef134-10(2F) complex, compared with four hydrogen bonds in the H27-14-A24/Nef134-10(wt) complex and five hydrogen bonds in the T36-5-A24/Nef134-10(2F) complex (Supplementary Table 3).

G28α and G98β formed hydrogen bonds with the side chains of the R134(P1-R) and C142(P9-C) residues, respectively. Surprisingly, the Vα domains, compared to Vβ domains, contributed more (~80%) in the interaction with A24/Nef134-10(2F) and were critical in the recognition of primarily the N-terminal part of the A24/Nef134-10(2F) by C1-28 TCR (Fig. 3c, f).

Wild type-specific TCR against late/minor mutant. H27-14 was slightly weaker in recognizing A24/Nef134-10(6L) than in recognizing A24/Nef134-10(wt) (Fig. 1a and Table S3). To evaluate the differences in binding ability to A24/Nef134-10(wt) or A24/Nef134-10(6L), we compared the structures of H27-14-A24/Nef134-10(wt) and H27-14-A24/Nef134-10(6L) (Fig. 4a). Conformational differences were minimal, with RMSDs of 0.485Å, 0.123Å and 0.109Å for the TCRs, Nef134-10 peptides and A24 binding clefts, respectively.

In H27-14-A24/Nef134-10(wt), the aromatic side chain of F139 (P6-F) protruding out of HLA-A24 binding groove was accommodated by the TCR pocket formed by the CDR1 and CDR3 loops, especially with Y31α, R93α, Y98α, R31β (Fig. 4b). Interestingly, the same binding mode was found in H27-14 TCR-A24/Nef134-10(6L) (Fig. 4c). The side chain of F139(P6-F) in the Nef134-10(wt) peptide was bulky. The 139L(P6-L) side chain in Nef134-10(6L) was much smaller, resulting in a gap in the TCR pocket surrounding the 139L(P6-L) side chain in the H27-14 TCR-A24/Nef134-10(6L) complex.

The number of van der Waals contacts (<4Å) between the TCR and P6 were 27 and 13 in H27-14-A24/Nef134-10(wt) and H27-14-A24/Nef134-10(6L), respectively. Additionally, cation-pi interactions between R93α or R31β and F139(P6-F) in H27-14-A24/Nef134-10(wt) were deformed in the case of 139L(P6L). The H27-14 TCR pocket surrounding P6 was a better fit for F139(P6-F) than for 139L(P6-L). This subtle structural difference might have been reflected in the killing assay and SPR (Fig. 1a and Table S3). Taken together, these results suggested that F139L mutation could be selected weakly but substantially as a minor escape mutant under the CTL pressure against the Nef134-10 epitope.

We were able to confirm the transformation from F to L at 6th position of Nef134-10 epitope in an HLA-A24-positive patient (Supplementary Fig. 6a). F139L appeared between Days 306 to 735

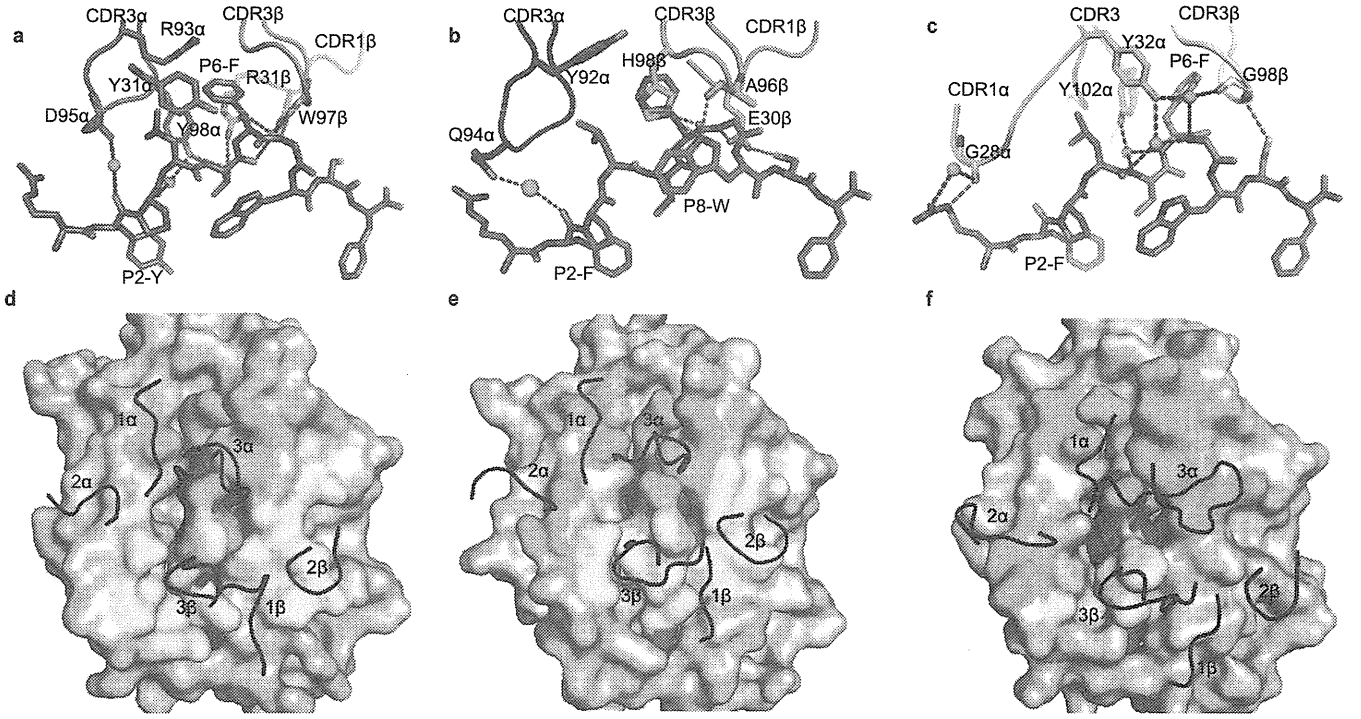


Figure 3 | Comparisons of interaction of TCRs with pHLAs. Binding interfaces: (a) the H27-14 TCR-A24/Nef134-10(wt), (b) the T36-5 TCR-A24/Nef134-10(2F) and (c) the C1-28 TCR-A24/Nef134-10(2F). TCRs are shown as loop representations and the residues important for binding with the peptide are represented as stick models. Nef134-10(wt), blue; Nef134-10(2F), red; water molecules, green spheres; hydrogen bonds, red dashed lines; water-mediated hydrogen bonds, blue dashed lines. pHLA surface representations with the binding footprint of TCRs: (d) H27-14 TCR with the A24/Nef134-10(wt), (e) T36-5 TCR with the A24/Nef134-10(2F) and (f) C1-28 TCR with the A24/Nef134-10(2F). The surfaces of pHLA interacted with TCR are shown in green for Vα and yellow for Vβ. The CDRs on the pHLAs are represented as black loops. HLA-A24, grey; Nef134-10(wt), blue; Nef134-10(2F), red.

and stayed until at least 2093 days after the initial visit. Notably, the 2nd position of Nef134-10 epitope was the wild type (Y) in all IMSUT cohort patients with F139L (Supplementary Fig. 6b). Although the number of patients with F139L was limited, the emergence of the mutation at position 139 suggests that the escape mechanism of the F139L mutant could be different from that of Y135F.

Interaction between mutant- or dual-specific TCRs and A24/Nef134-10(2F). Although Y135F has been described as an example of a CTL escape mutant in both epidemiological and immunological studies[10], A24/Nef134-10(2F) was recognized by T36-5 and by C1-28 (Fig. 1a and Table S3). Therefore, we decided to analyze the structure of TCR-A24/Nef134-10(2F) complexes in both mutant-specific T36-5 and dual-specific C1-28 TCRs.

Formation of T36-5-A24/Nef134-10(2F) introduced a large TCR-induced conformational change in the peptide, with insertion of H98β into a shallow pocket formed by L137(P4-L), F139(P6-F) and W141(P8-W) (Fig. 5a). Around this pocket, Y92α and Y31β were involved in the hydrophobic interactions with L137(P4-L) and F139(P6-F). In addition, H98β and S97β formed hydrogen bonds with the main chain of L137 (P4-L) and the side chain of W141(P8-W), respectively (Supplementary Table 3). As a consequence, the side chain of W141(P8-W), which was accommodated by HLA-A24 binding pocket in unligated state, moved about 5.6 Å shift toward TCR binding surface, forming a more exposed and featured peptide conformation. Thus, Nef134-10(2F) underwent large T36-5 TCR-induced conformational change affecting the main chain hydrogen bond networks within the peptide. In the unligated state, T138(P5-T) carbonyl formed two hydrogen bonds with the nitrogen atoms of G140(P7-G) and W141(P8-W); also the W141(P8-W) nitrogen formed a bond with the F139(P6-F) carbonyl (Fig. 5b). After the TCR ligation, the hydrogen bond between T138(P5-T) carbonyl and G140(P7-G) nitrogen was lost and T138 (P5-T) carbonyl reformed a single hydrogen bond with W141(P8-W) nitrogen (Fig. 5c).

C1-28 recognized Nef134-10(2F) and Nef134-10(wt) almost equally, while T36-5 recognized Nef134-10(2F) better than Nef134-10(wt). The extensive involvement of the Vα domain in the interaction of C1-28 TCR-A24/Nef134-10(2F) was unique and quite different from typical TCR-pHLA complexes (Fig. 3f). The extended tip of the non-germline-encoded CDR3α loop lay over the A24 α1-helix (65–69 residues); by contrast, the germline-encoded CDR1α loop interacted with the N-terminus of the Nef134-10(2F) peptide (Figs. 3f, 6a and Supplementary Fig. 7a). The framework residues of Vα8-3 had multiple interactions with HLA-A24: N-terminal A1α formed a hydrogen bond with E58, and R69α formed hydrogen bonds with A158, had VDW contacts with T163, and formed a water-mediated bond and salt bridges with D166 (Fig. 6a, c). In addition, CDR1α (Y27, G28, T30) and CDR2α (F51, S52) loops made multiple contacts with HLA-A24 residues (Fig. 6a, c and Supplementary Table 3). By contrast, Vβ4-1 made relatively few contacts with HLA-A24. S97 and I99 of the CDR3β interacted with HLA-A24 residues T73 and A150, respectively (Fig. 6b, d and Supplementary Table 3). In addition, S97β formed a water-mediated bond with the main chain of F139 (P6-F).

Combination bias in the TCR variable gene segments used in dual-specific CTL. We showed previously that Vα8-3 and Vβ4-1 were public TCRs used most frequently in the dual-specific CD8+ T cell population[14]. Many of the interactions with HLA-A24 described above could explain the bias for Vα8-3. We looked for the molecular clues for an exclusive use of Vβ4-1. Compared to the Vβ chains of H27-14 and T36-5 TCRs, Vβ4-1 sequences coded by the germline gene segments did not have much interaction with A24/Nef134-10(2F).

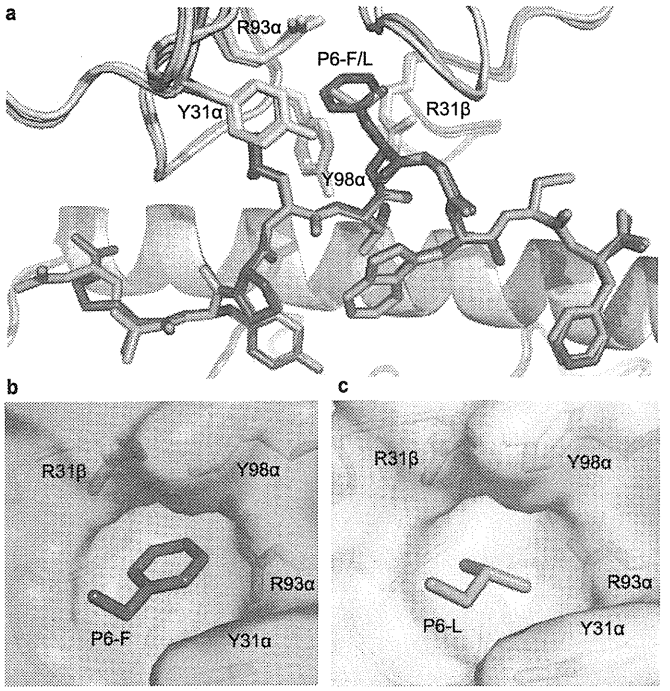


Figure 4 | The pocket of H27-14 TCR surrounding Phe/Leu at position 6 of the Nef134-10 peptide. (a) Overlay of the H27-14 TCR-A24/Nef134-10(wt) and H27-14 TCR-A24/Nef134-10(6L). The CDR loops in the H27-14 TCR in complex with A24/Nef134-10(wt) or A24/Nef134-10(6L) are shown in cyan or yellow, respectively. The Nef134-10(wt) or Nef134-10(6L) peptides are shown in blue or orange stick models, respectively. HLA-A24 α1 but not α2 helices are shown in grey cartoon representation for clarity. (b) The side chain of Phe at P6 (stick model in blue) is surrounded by the H27-14 TCR pocket formed by the CDR loops (cyan). (c) The side chain of Leu at P6 (stick model in orange) is surrounded by the H27-14 TCR pocket formed by the CDR loops (yellow).

The CDR2β loop in Vβ4-1 was displaced outside of HLA-A24 α1-helix by the CDR3α (Supplementary Fig. 7a, b). We reported previously that the GI (Gly-Ile) motif was present at position 98-99 of CDR3β loop in most dual-specific CD8+ T cell repertoire[14]. Interestingly, G98β formed a hydrogen bond with C142(P9-C) of the Nef134-10(2F) peptide and hydrophobic I99β interacted with hydrophobic A150 of the HLA-A24 (Fig. 6b, d and Supplementary Table 3).

HLA class I amino acid residues 65, 69, and 155, referred to as the restriction triad, are crucial for TCR recognition[18,19]. Although the types of contacts differed, all three TCRs solved in this study interact with the restriction triad (Supplementary Table 3). The tips of the CDR3α loop (residues 96-101) were sandwiched between HLA-A24 (G65 and A69) and Vβ (S55 and I56) in C1-28-A24/Nef134-10(2F); G98α and S55β formed hydrogen bonds with K68 of the HLA-A24 and S101α of the CDR3α, respectively (Fig. 6e). The presence of both S55β and I56β is unique to germline-encoded TRBV4-1 gene segment, and may contribute to stabilization between the CDR3α and the HLA-A24.

Although the CDR3α sequences (97–100 residues) of the public TCRs varied among dual-specific CTL clones[14], the sequences were composed of small amino acids such as Gly and Ser. For insertion into an interface between G65-A69 of the HLA-A24 and S55β-I56β and formation of stable interactions between the HLA-A24 complex and CDR3α residues, small size would be beneficial. Vβ4-1 may have been co-selected by geometric constraints imposed by Vα8-3 germline and hypervariable CDR3α sequences.

The plasticity of Nef134-10 peptides. In the interactions with HLA complexes, Nef134-10 peptides assumed an M-shaped conformation similar to the structure of the cancer-related telomerase peptide presented by HLA-A24 (Fig. 2a and 7a, b)[15]. The side chain of T138(P5-T) faced toward the groove of HLA-A24 and functioned as the secondary anchor residue (Fig. 2b). By switching the hydrogen bond from Y135(P2-Y)/H70 to T138(P5-T)/H70, the Y135(P2-Y) to 135F(P2-F) mutation maintained the conformation of the secondary anchor to HLA-A24 α1 helix. After TCR ligation, the C-terminus shifted more than the N-terminus in H27-14 TCR-A24/Nef134-10(wt) and T36-5 TCR/A24/Nef134-10(2F) (Fig. 7c). The main chain of F139(P6-F) shifted 2.81Å in H27-14 TCR-A24/Nef134-10(wt), and G140(P7-G) shifted 2.47Å in T36-5 TCR-A24/Nef134-10(2F). By contrast, C1-28 TCR-A24/Nef134-10(2F) had less shift in the main chain of the P5-P8 (0.32-0.74 Å) residues and greater shift of the P2-P4 residues (1.06–1.68 Å) in the N-terminus.

Discussion

Nef134-10 epitope (RYPLTFGWCF) was a highly immunogenic epitope restricted by HLA-A24[10,14,20]. Y135F was the major escape mutation which appeared early in the clinical course (Supplementary Fig. 1a–c). Our previous study suggesting that Y135F was a processing mutation was later confirmed by others[10,21]. HIV-1 with Y135F mutation has been accumulating in the population with high HLA-A24 prevalence such as Japanese. Hence, Y135F is the ultimate escape mutation which appears early in the clinical course. F139L was a minor but also HLA-A24-related mutation. Intriguingly, all 6 patients with F139L mutation kept the wild type residue at 135th position in the epitope.

According to the crystal structure of the free wild type, A24/Nef134-10 epitope took M conformation epitope. L137(P4-L) and

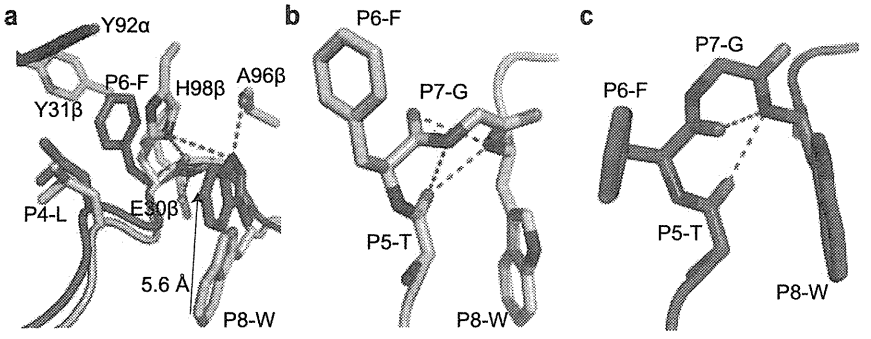


Figure 5 | Interaction of T36-5 TCR with A24/Nef134-10(2F). (a) pHLA unligated or ligated to T36-5 TCR. Nef134-10(2F) peptide of unligated pHLA, pink; the peptide ligated T36-5 TCR, red; CDRα, blue; CDRβ, green; hydrogen bonds, red dashed lines. (b) Peptide intramolecular hydrogen bonds in unligated pHLA. Color representation are the same as (a). (c) Peptide intramolecular hydrogen bonds in T36-5 TCR-A24/Nef134-10(2F) complex. Color representation are the same as (a).

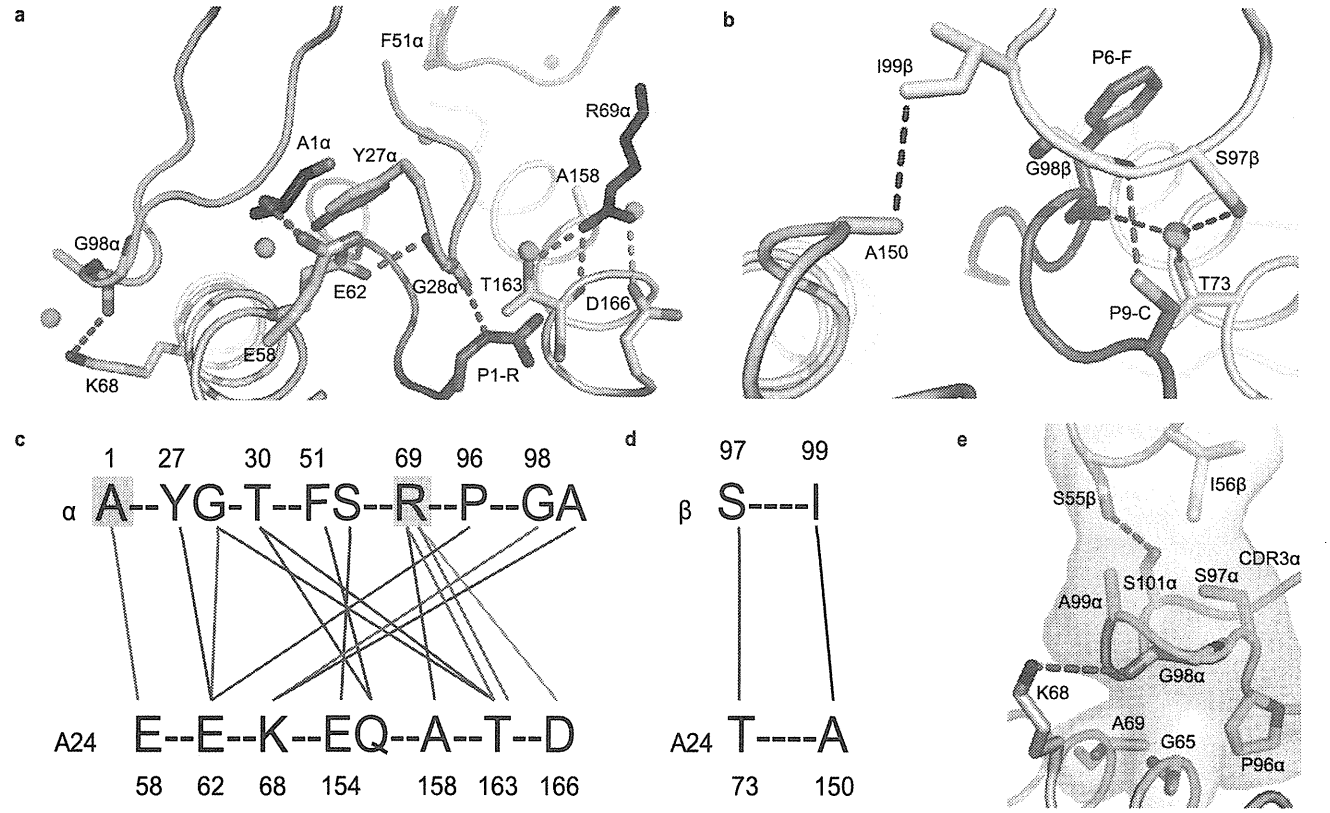


Figure 6 | Interaction of C1-28 TCR with A24/Nef134-10(2F). (a) Interaction of Vα and (b) Vβ with A24/Nef134-10(2F). CDRα, cyan; Fwα (Frame work region of Vα), black; CDR3β, yellow; Nef134-10(2F), red; HLA-A24, grey, water molecules, green spheres; red dashed lines, hydrogen bonds; a green dashed line, a salt bridge. Van der Waals contact (<4.0 Å), which is represented as a black dashed line, is shown in (b), but not in (a) for clarity. (c) Interactions of Vα and (d) Vβ with A24/Nef134-10(2F). Red, blue, green and black solid lines indicate hydrogen bonds, water-mediated hydrogen bonds, a salt bridge and van der Waals contact, respectively. Fwα residues (A1 and R69) are highlighted in grey. (e) For the C1-28 TCR, interactions of the tip of the CDR3α loop (residues 96-101, cyan) with S55β and I56β residues of the Vβ4-1 segments (yellow), and G65 and A69 residues of the HLA-A24 (grey). Red dashed lines indicate a hydrogen bond.

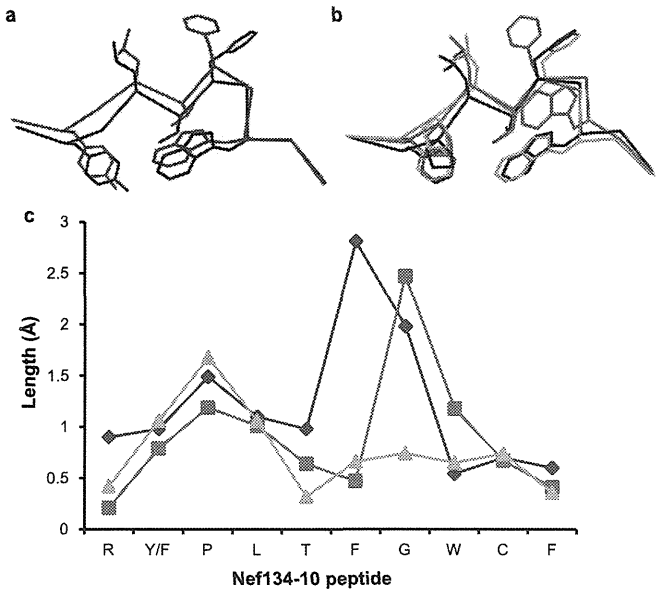


Figure 7 | Structural change of the Nef134-10 peptides in free and TCR-bound states. (a) Free Nef134-10(wt) (black) and Nef134-10(wt) bound to the H27-14 TCR (blue). (b) Free Nef134-10(2F) (black), the Nef134-10(2F) bound to T36-5 TCR (red) and the Nef134-10(2F) bound to the C1-28 TCR (green). (c) The migration length of the Nef134-10 main chain after TCR ligation, when HLA-A24 binding groove is superimposed.

F139(P6-F) were the eminent feature of this epitope and Y138(P5-T) was located in the valley of L137(P4-L) and F139(P6-F). N-terminal anchor, Y135(P2-Y), formed a hydrogen bond with His70 of HLA-A24 molecule. Although Y135F mutation interrupted this hydrogen bond, T138(P5-T) acted like a subanchor forming a hydrogen bond with HLA-A24 His70 keeping the M conformation very similar to the wild type. It should be noted that F is the 2nd best N-terminal anchor residue after Y for HLA-A24[22,23]. F139L mutation did not cause gross conformational change in the epitope, however, the solvent accessible surface area of the side chain became substantially smaller by the mutation. Namely, the epitope became slightly featureless by F139L mutation.

H27-14 TCR was highly specific to the wild type epitope (K_D = 9.7 ± 0.7). Crystal structure of the complex was typical TCR/pMHC interaction in which CDR loops of α and β chains contributed almost equally to interact with the eminent feature of the epitope, F139(P6-F) residue. The binding affinity against early/major Y135F mutation diminished to nonfunctional level (K_D = 291.5 ± 63.8). Although H27-14 TCR kept moderate affinity against F139L (K_D = 64.8 ± 3.4), its TCR pocket was a better fit for bulky side chain of F139(P6-F) than for small side chain of 139L(P6-L). These findings may suggest that CTL with H27-14 TCR clonotype could contribute to eliminate HIV-1 with the wild type epitope but might be responsible for selection of viruses with the late/minor F139L mutation *in vivo*.

Nef134-10(2F)-specific T36-5 TCR had a very high affinity to A24/Nef134-10(2F) and extensive interaction shown by the BSA (approximately 2200 Å^2). The side chain of 141W(P8-W) was accommodated by the HLA-A24 binding cleft in the undocked state;

however, it was apparently lifted up by T36-5 TCR. Fujiwara et al. described CTL clones which recognized A24/Nef134-10(2F) more efficiently than the wild type (A24/Nef134-10(wt)). T36-5 may be one of those clones[21].

In the great majority of HLA-A24-positive chronically infected patients, plasma viruses had Y135F mutation, while the patients harbored dual specific CD8+ T cell population with highly restricted TCR repertoire[14]. The dual-specific CD8+ T cell population expressed higher activation markers such as PD1 than the wild type-specific CD8+ T cell population (Kawana-Tachikawa, A., unpublished observation). Puzzled by the presence of dual-specific CD8+ T cell population stimulated by the epitope with the processing mutation, we wished to study the molecular interaction between the TCR and pMHC. C1-28 was chosen for the structural study as the representative of the dual-specific CD8+ T cell population. As far as we know, this is the first report of the crystal structure of Vα8-3.

According to the structural analysis of dominant public TCRs, germline residues either of the Vβ or Vα may play a major role[24,25]. In C1-28 TCR, Vα domains contributed predominantly (80%) in the interaction with A24/Nef134-10(2F). N-terminal A1α, R69α, residues in CDR1α (Y27, G28, T30) and CDR2α (F51, S52) had multiple interactions with HLA-A24. Nongermline CDR3α loop (residues 96-101) also contributed to interact with the two residues of restriction triad in HLA-A24, while S101α had a hydrogen bond with S55β, a germline residue of public Vβ4-1. Therefore, hypervariable CDR3α contributed to bridge the binding of HLA-A24 and Vβ4-1. In terms of the interaction with the Nef134-10(2F) peptide, germline-encoded CDR1α loop (G28α) interacted with the N-terminus of the peptide 134R(P1-R). Also, hydroxyl groups of Y32 of CDR1α and Y102 of non-germline-encoded CDR3α made water-mediated bonds with the main chain of the peptide, while both aromatic rings made hydrophobic interaction with the aromatic ring of F139 (P6-F). Thus, CDR1α and CDR3α provided a major role in interaction with the Nef134-10(2F) in addition to the interaction with HLA-A24.

Compared to the other two TCRs examined in this study, the role of Vβ4-1 of the C1-28 TCR in the recognition of HLA-A24 and the peptide was not impressive. The CDR2β loop was displaced outside of the HLA-A24 α1-helix (residues 65–69) by the CDR3α of Vα8-3. However, S55β and I56β, germline residues unique to Vβ4-1, could contribute to stabilize the tip of CDR3α, which had an important interaction with the restriction triad. Thus, the major cause of the co-selection of Vβ4-1 as a public TCR could be the bias brought by the extensive interaction of the germline Vα8-3 residues with HLA-A24. In addition, the interaction of the small residues such as Gly and Ser in CDR3α with restrictive elements in the HLA-A24 contributed to the selection of Vβ4-1. Non-germline CDR3β contributed to the interaction with HLA-A24 (T73 and A150) and C142(P9-C) of the peptide. Collectively, these structural analyses revealed that T36-5 and C1-28 TCRs have their own unique mode of binding to a mutated epitope. The three TCRs studied did not introduce a substantial conformational change in the N-terminal side of the peptide; however, two TCRs with antigen preference but not the dual-specific C1-28 TCR introduced a large conformational change in the C-terminal side.

Recently, TCR/pMHC structure of KK10 epitope restricted by HLA-B*2705 was reported[26]. Clone C12C was cross-reactive to both wild type and early appearing Leu268Met mutant. Although both C12C and C1-28 were dually specific to both wild type and early mutant, the character of early mutation and their TCR/pMHC structure were quite different. In the case of KK10 epitope restricted by HLA-B*2705, Leu268Met was an early mutation affecting TCR recognition and the ultimate mutations such as Arg264Lys disrupting antigen presentation follow later. However, in the case of Nef134-10 epitope restricted by HLA-A24, early Y135 mutation was the ultimate mutation. Although the number of patients are still limited, late mutation in Nef134-10 epitope, F139L, occurred only in the patients without early/ultimate Y135F mutation. It is interesting to note that HLA-B*27 is a protective but HLA-A24 is not for the disease progression after HIV-1 infection[27]. Although further study is needed, the different mode of mutant appearance between HLA-B*2705/KK10 and HLA-A24/Nef134-10 epitopes may be related to the role of the HLA alleles on the disease progression.

In the population with high HLA-A24 prevalence, HIV-1 with the wild type Nef134-10 epitope is replaced by the virus with Y135F mutation early after infection. Or HIV-1 with Y135F may infect HLA-A24-positive people. Whichever the case, HIV-1 with Y135F mutation replicate for many years in the presence of activated dual-specific CTL. It is tempting to speculate that the Y135F mutation could be an example of "stealth mutation." HIV-1-infected cells might not be detected by specific- or dual-specific CTL due to processing failure or excess processing. If there is a difference in antigen processing between professional antigen presenting cells and infected T cells, CTL with functional TCR might be kept activated through cross-priming by professional antigen-presenting cells[28,29]. If this is the case, the immune system cannot see the infected cells as targets but may be kept activated by the spurious epitope presented by the uninfected professional antigen presenting cells. Although this hypothesis must be proven by further studies, our findings support the possibility that drugs which alter virus-peptide processing may have potential as "therapeutic vaccines."

Methods

Approval of the study and recombinant DNA experiments in IMSUT. Plasma samples from HIV-1-positive patients attending the hospital affiliated with the Institute of Medical Science, the University of Tokyo (IMSUT) were collected and kept frozen until use. Patients provided written informed consent, and the study was approved by the Institutional Review Board of the University of Tokyo (approval number 20–31). Recombinant DNA experiments used in this work were approved by the Institutional Review Board (approval number 08–30).

British columbia HOMER cohort. Founded in 1996, the British Columbia HOMER cohort is an open cohort of antiretroviral-naïve, chronically HIV-1 infected individuals. The cohort is predominantly Caucasian. Plasma HIV-1 RNA sequencing and HLA class I sequence-based typing have been performed in the HOMER cohort as described[30]. Here, we investigated the relationship between HLA-A*24 expression and sequence variants at Nef codons 135 and 139 in 1018 HOMER participants with HIV-1 Nef and HLA-A data available.

Longitudinal acute/early infection cohort. A total of 16 HLA-A*24 expressing individuals from a longitudinal multicenter acute/early HIV-1 infection cohort were investigated to determine the time course of selection of sequence variants at Nef codons 135 and 139 using Kaplan-Meier methods as described in[11]. "Time to escape" was defined as the number of days elapsed between estimated infection date and first detection of the escape variant (as a full or partial amino acid change).

Sequencing of autologous viruses. Viral RNA from EDTA-treated plasma was isolated using the QIAamp viral RNA Mini kit (QIAGEN). For heparin-treated plasma, High Pure Viral Nucleic Acid Kit (Roche) was used for RNA isolation to remove the inhibitory effect of heparin on the PCR assay. HIV-1 *nef* region was amplified from the RNA using the Superscript III one-step RT-PCR system with Platinum Taq DNA polymerase with High Fidelity (Invitrogen) and *nef* specific primers. The second-round DNA-PCR was done with EX Taq DNA polymerase Hot Start enzyme (Takara). The sequences of primers for the above PCR reactions are available upon request. Purified PCR products were directly sequenced by using BigDye Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems) on an ABI 3130xl Genetic Analyzer.

Peptides. Synthetic peptides of Nef134-10(wt) [RYPLTFGWCF], Nef134-10(2F) [RFPLTFGWCF] and Nef134-10(6L) [RYPLTLGWCF] were purchased from Sigma-Genosys.

Generation of CTL clones. Nef134-10-specific CTL clones were established from peripheral mononuclear cells (PBMCs) derived from HIV-1 infected individuals carrying the HLA-A*2402, as previously described[13].

Sequencing of T-cell receptor α- and β-chains. Analysis of genes encoding TCR α- and β-chains from cloned CTL were done as previously described[14].

^{51}Cr release assay. Cytotoxicity was measured by a standard ^{51}Cr release assay as previously described[10].

Protein expression and purification. The H27-14, T36-5 and C1-28 TCRs were expressed, refolded and purified essentially as described[16]. For the H27-14 TCR refold, 24 mg of solubilized TCR α-chains and 20 mg of β-chains were injected into 1 L of a buffer containing 5 M urea, 100 mM Tris, pH 8.5, 400 mM L-arginine-HCl, 3.7 mM cystamine, 6.6 mM cysteamine, 0.2 mM PMSF (phenylmethylsulfonyl fluoride) at 4°C. The refolding solution was dialyzed twice for 24 h against 10 vol of milli Q water, and then 10 vol of 10 mM Tris, pH 8.5 at 4°C. The refolded TCR was then purified by Resource-Q column and Superdex 75 column (GE Healthcare). For the T36-5 TCR, 50 mg of α-chains and 40 mg of β-chains were refolded and purified as described above. For the C1-28 TCR refold, 20 mg of TCR α-chains and 35 mg of β-chains were injected twice into 2 L of a buffer containing 5 M urea, 100 mM Tris, pH 8.9, 400 mM L-arginine-HCl, 2 mM EDTA, 5 mM reduced glutathione, 0.5 mM oxidized glutathione, 0.2 mM PMSF at 4°C. After 24 hr incubation, the refolding solution was dialyzed twice for 24–36 h against 5 mM Tris, pH 8.5, 50 mM NaCl, and then 10 mM Tris, pH 8.5, 50 mM NaCl at 4°C. The resultant was purified by Resource-Q column, Mono-Q column and Superdex 75 column (GE Healthcare).

Preparation of pHLAs. The HLA-A*2402 heavy chain, HLA-A*2402 heavy chain-BSP (BirA substrate peptide) and β2 microglobulin (β2m) were also expressed separately in E. coli, as described[15]. For a 1 L refold, 45 mg of solubilized HLA-A*2402 heavy chains and 15 mg of β2m were injected into a refold buffer containing 100 mM Tris, pH 8.0, 400 mM L-arginine-HCl, 2 mM EDTA, 5 mM reduced glutathione, 0.5 mM oxidized glutathione, 0.2 mM PMSF in the presence of 10 mg of HIV-1 Nef 134-10 (wt), Nef134-10(2F) or Nef134-10 (6L) peptide. The refolded protein was purified by Superdex 75 column and MonoQ column. For SPR analysis, pHLA molecules were refolded in the same way as the above, using the HLA-A*2402-BSP instead of the HLA-A*2402 heavy chain. The pHLA-BSP was biotinylated as previously described[14].

Surface plasmon resonance. Surface plasmon resonance experiment was carried out at 25°C using BIAcore 2000 in a buffer containing 10 mM HEPES, pH 7.4, 150 mM NaCl, 3 mM EDTA, 0.005% Surfactant P20. Biotinylated pHLAs were immobilized to Sensor chip SA until the response was reached between 200–800 response units (RU). Each TCR was injected over the flow cells at a flow-rate of 20–30 μl/min with an indicated concentration range for equilibrium analysis. BIAevaluation software (version 4.1; GE healthcare) was used for data analysis. For kinetic analysis, 1 : 1 Langmuir binding model was used to calculate the K_{on} and K_{off} values.

Crystallization and data collection. All crystallizations were done by the sitting drop vapor diffusion method with a protein/reservoir drop ratio of 1 : 1 at 20°C. Crystallization conditions of grown crystals are shown in Supplementary Table S3. For cryoprotection, obtained crystals were soaked briefly and sequentially in reservoir solutions containing 10% and 20% ethylene glycol, and then flash-frozen in liquid nitrogen. Data were collected at the beamline BL41XU in SPring 8 (Hyogo, Japan) and BL-5A, NW12A and BL1A in the PF facility (Tsukuba, Japan), and processed with HKL2000[31] and the CCP4 program suite[32].

Structure determination and refinement. The structures were determined by molecular replacement using Molrep[33]. Search models used for molecular replacement were shown in Supplementary Table S3. Model building and refinement were conducted using Coot[34] and CNS 1.3[35], respectively. The further rounds of these refinements were performed using REFMAC implemented in CCP4. For building model of the T36-5 TCR-A24/Nef134-10(2F), diffraction intensities from the various crystals exhibited twin with an estimated twinning fraction of 0.45–0.49. Therefore, this structural model was refined using CNS 1.3 as a perfect twin.

The stereochemistry of the refined models was assessed with program Rampage[36]. All molecular graphics representations were created with the program PyMOL (DeLano Scientific; http://www.pymol.org). Data collection and refinement statistics are shown in Supplementary Table 2.

1. Kappler, J. *et al.* The major histocompatibility complex-restricted antigen receptor on T cells in mouse and man: identification of constant and variable peptides. *Cell* **35**, 295–302 (1983).
2. Marrack, P., Shimonkevitz, R., Hannum, C., Haskins, K. & Kappler, J. The major histocompatibility complex-restricted antigen receptor on T cells. IV. An antiidiotypic antibody predicts both antigen and I-specificity. *J Exp Med* **158**, 1635–1646 (1983).
3. McMichael, A. J. & Rowland-Jones, S. L. Cellular immune responses to HIV. *Nature* **410**, 980–987 (2001).
4. Walker, B. D. & Burton, D. R. Toward an AIDS vaccine. *Science* **320**, 760–764 (2008).
5. de la Salle, H. *et al.* Homozygous human TAP peptide transporter mutation in HLA class I deficiency. *Science* **265**, 237–241 (1994).
6. Fruh, K. *et al.* A viral inhibitor of peptide transporters for antigen presentation. *Nature* **375**, 415–418 (1995).
7. Maeurer, M. J. *et al.* Tumor escape from immune recognition: lethal recurrent melanoma in a patient associated with downregulation of the peptide transporter protein TAP-1 and loss of expression of the immunodominant MART-1/Melan-A antigen. *J Clin Invest* **98**, 1633–1641 (1996).
8. Bjorkman, P. J. *et al.* Structure of the human class I histocompatibility antigen, HLA-A2. *Nature* **329**, 506–512 (1987).
9. Davis, M. M. & Bjorkman, P. J. T-cell antigen receptor genes and T-cell recognition. *Nature* **334**, 395–402 (1988).
10. Furutsuki, T. *et al.* Frequent transmission of cytotoxic-T-lymphocyte escape mutants of human immunodeficiency virus type 1 in the highly HLA-A24-positive Japanese population. *J Virol* **78**, 8437–8445 (2004).
11. Brumme, Z. L. *et al.* Marked epitope- and allele-specific differences in rates of mutation in human immunodeficiency type 1 (HIV-1) Gag, Pol, and Nef cytotoxic T-lymphocyte epitopes in acute/early HIV-1 infection. *J Virol* **82**, 9216–9227 (2008).
12. Itoh, Y. *et al.* High-throughput DNA typing of HLA-A, -B, -C, and -DRB1 loci by a PCR-SSOP-Luminex method in the Japanese population. *Immunogenetics* **57**, 717–729 (2005).
13. Kawana-Tachikawa, A. *et al.* An efficient and versatile mammalian viral vector system for major histocompatibility complex class I/peptide complexes. *J Virol* **76**, 11982–11988 (2002).
14. Miyazaki, E. *et al.* Highly restricted T-cell receptor repertoire in the CD8+ T-cell response against an HIV-1 epitope with a stereotypic amino acid substitution. *AIDS* **23**, 651–660 (2009).
15. Cole, D. K. *et al.* Crystal structure of HLA-A*2402 complexed with a telomerase peptide. *Eur J Immunol* **36**, 170–179 (2006).
16. Boulter, J. M. *et al.* Stable, soluble T-cell receptor molecules for crystallization and therapeutics. *Protein Eng* **16**, 707–711 (2003).
17. Rudolph, M. G., Stanfield, R. L. & Wilson, I. A. How TCRs bind MHCs, peptides, and coreceptors. *Annu Rev Immunol* **24**, 419–466 (2006).
18. Tynan, F. E. *et al.* T cell receptor recognition of a 'super-bulged' major histocompatibility complex class I-bound peptide. *Nat Immunol* **6**, 1114–1122 (2005).
19. Burrows, S. R. *et al.* Hard wiring of T cell receptor specificity for the major histocompatibility complex is underpinned by TCR adaptability. *Proc Natl Acad Sci U S A* **107**, 10608–10613 (2010).
20. Ikeda-Moore, Y. *et al.* Identification and characterization of multiple HLA-A24-restricted HIV-1 CTL epitopes: strong epitopes are derived from V regions of HIV-1. *J Immunol* **159**, 6242–6252 (1997).
21. Fujiwara, M. *et al.* Different abilities of escape mutant-specific cytotoxic T cells to suppress replication of escape mutant and wild-type human immunodeficiency virus type 1 in new hosts. *J Virol* **82**, 138–147 (2008).
22. Ibe, M. *et al.* Role of strong anchor residues in the effective binding of 10-mer and 11-mer peptides to HLA-A*2402 molecules. *Immunogenetics* **44**, 233–241 (1996).
23. Sidney, J., Southwood, S. & Sette, A. Classification of A1- and A24-supertype molecules by analysis of their MHC-peptide binding repertoires. *Immunogenetics* **57**, 393–408 (2005).
24. Kjer-Nielsen, L. *et al.* A structural basis for the selection of dominant alphabeta T cell receptors in antiviral immunity. *Immunity* **18**, 53–64 (2003).
25. Stewart-Jones, G. B., McMichael, A. J., Bell, J. I., Stuart, D. I. & Jones, E. Y. A structural basis for immunodominant human T cell receptor recognition. *Nat Immunol* **4**, 657–663 (2003).
26. Ladell, K. *et al.* A molecular basis for the control of preimmune escape variants by HIV-specific CD8+ T cells. *Immunity* **38**, 425–436 (2013).
27. O'Brien, S. J., Gao, X. & Carrington, M. HLA and AIDS: a cautionary tale. *Trends Mol Med* **7**, 379–381 (2001).
28. Lehner, P. J. & Cresswell, P. Recent developments in MHC-class-I-mediated antigen presentation. *Curr Opin Immunol* **16**, 82–89 (2004).
29. Nakayama, K. *et al.* Imbalanced Production of Cytokines by T Cells Associates with the Activation/Exhaustion Status of Memory T Cells in Chronic HIV Type 1 Infection. *AIDS Res Hum Retroviruses* **28**, 702–714 (2011).
30. Brumme, Z. L. *et al.* Evidence of differential HLA class I-mediated viral evolution in functional and accessory/regulatory genes of HIV-1. *PLoS Pathog* **3**, e94 (2007).
31. Otwinowski, Z. & Minor, W. Processing of X-Ray Diffraction Data Collected in Oscillation Mode Macromolecular Crystallography. *Methods in Enzymology* **276**, 307–326 (1997).
32. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr D Biol Crystallogr* **50**, 760–763 (1994).
33. Vagin, A. & Teplyakov, A. MOLREP: an Automated Program for Molecular Replacement. *J Appl Cryst* **30**, 1022–1025 (1997).
34. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* **60**, 2126–2132 (2004).
35. Brunger, A. T. *et al.* Crystallography & NMR System: A New Software Suite for Macromolecular Structure Determination. *Acta Cryst.* **D54**, 905–921 (1998).
36. Lovell, S. C. *et al.* Structure validation by Calpha geometry: phi,psi and Cbeta deviation. *Proteins: Structure, Function, and Genetics* **50**, 437–450 (2003).

Acknowledgments

We thank the beam-line staffs at NW12A, BL1A and BL5A of Photon Factory (Tsukuba, Japan) and BL41XU of SPring8 (Hyogo, Japan) for technical help during data collection. We thank Drs. Richard Harrigan, Heiko Jessen, Anthony Kelleher, Martin Markowitz and Bruce Walker for specimen and/or data access. This work was supported in part by a contract research fund from the Ministry of Education, Culture, Sports, Science and Technology(MEXT)for Program of Japan Initiative for Global Research Network on Infectious Diseases (10005010)(AI); Global COE Program (Center of Education and Research for Advanced Genome-Based Medicine - For personalized medicine and the

control of worldwide infectious diseases-) of MEXT (F06)(AI); JSPS KAKENHI (25293226)(AKT); Grants for AIDS research from the Ministry of Health, Labor, and Welfare of Japan (H24-AIDS-IPPAN-008)(AKT); Research on international cooperation in medical science, Research on global health issues, Health and Labour Science Research Grants, the Ministry of Health, Labor, and Welfare of Japan (H25-KOKUI-SITEI-001)(AI); CJB is supported by a Vanier Canada Graduate Scholarship from the Canadian Institutes of Health Research (CIHR). EM is supported by a Master's Scholarship from the Canadian Association of HIV Research and Abbott Virology. ZLB is the recipient of a CIHR New Investigator Award and a Scholar Award from the Michael Smith Foundation for Health Research.

Author contributions

A.S. did all the experiments from protein synthesis, BIAcore assay to structural studies. A.K.-T. did all the immunological studies including establishment of CTL cell lines and characterization of CTLs. A.Y. and Y.S. did crystallographic studies. C.Y.H. and D.Z. contributed to sequencing of viruses and TCR genes and analyzed the relationship between HLA-A*2402 expression and sequence variants at Nef codons 135 and 139 in IMSUT cohort. H.N. provided the clinical data of IMSUT cohort. T.K. was responsible for the clinical management in the IMSUT hospital. J.C. analyzed the relationship between HLA-A*2402 expression and sequence variants at Nef codons 135 and 139 in longitudinal multicenter acute/early HIV-1 infection cohort. E.M. and C.J.B. analyzed the relationship between HLA-A*24 expression and sequence variants at Nef codons 135 and 139 in British Columbia HOMER cohort. Y.S. discussed the structural study described in this manuscript. G.F.G. has been collaborating with A.I. and joined the discussion for this study. Z.L.B. investigated the relationship between HLA-A*24 expression and sequence variants at Nef codons 135 and 139 in 1018 participants in British Columbia HOMER cohort and longitudinal multicenter acute/early HIV-1 infection cohort with HIV-1 Nef and HLA-A data available. S.F. managed and instructed the structural study group and provided the environment for the crystallographic study and analysis. A.I. managed all the work described and wrote the manuscript. All authors reviewed the manuscript.

Additional information

Accession codes C1-28 TCR/Nef134-10(2F), 3VXM; A24/Nef134-10(wt), 3VXN; A24/Nef134-10(2F), 3VXO; A24/Nef134-10(6L), 3VXP; H27-14 TCR, 3VXQ; H27-14 TCR-A24/Nef134-10(wt), 3VXR; H27-14 TCR-A24/Nef134-10(6L), 3VXS; T36-5 TCR, 3VXT; T36-5 TCR-A24/Nef134-10(2F), 3VXU and 3W0W.

Supplementary information accompanies this paper at http://www.nature.com/scientificreports

Competing financial interests: The authors declare no competing financial interests.

How to cite this article: Shimizu, A. *et al.* Structure of TCR and antigen complexes at an immunodominant CTL epitope in HIV-1 infection. *Sci. Rep.* 3, 3097; DOI:10.1038/srep03097 (2013).

Teeranaipong P et al. *Journal of the International AIDS Society* 2013, **16**:18723
http://www.jiasociety.org/index.php/jias/article/view/18723 | http://dx.doi.org/10.7448/IAS.16.1.18723

JIAS Journal of the International AIDS Society

Research article

Development of a rapid cell-fusion-based phenotypic HIV-1 tropism assay

Phairote Teeranaipong*[1], Noriaki Hosoya*[2], Ai Kawana-Tachikawa[1], Takeshi Fujii[3], Tomohiko Koibuchi[3], Hitomi Nakamura[2,3], Michiko Koga[1,3], Naoyuki Kondo[4,5], George F Gao[6], Hiroo Hoshino[7], Zene Matsuda[4,5] and Aikichi Iwamoto[§,1,2,3,4]

[§]**Corresponding author:** Aikichi Iwamoto, Division of Infectious Diseases, Advanced Clinical Research Center, Institute of Medical Science, University of Tokyo, 4-6-1 Shirokanedai, Minato-Ku, Tokyo 108-8639, Japan. Tel: +81-3-6409-2202. Fax: +81-3-6409-2008. (aikichi@ra3.so-net.ne.jp, aikichi@ims.u-tokyo.ac.jp)
*These authors contributed equally to this work

Abstract

Introduction: A dual split reporter protein system (DSP), recombining *Renilla* luciferase (RL) and green fluorescent protein (GFP) split into two different constructs (DSP_{1-7} and DSP_{8-11}), was adapted to create a novel rapid phenotypic tropism assay (PTA) for HIV-1 infection (DSP-Pheno).

Methods: DSP_{1-7} was stably expressed in the glioma-derived NP-2 cell lines, which expressed CD4/CXCR4 (N4X4) or CD4/CCR5 (N4R5), respectively. An expression vector with DSP_{8-11} (pRE11) was constructed. The HIV-1 envelope genes were subcloned in pRE11 (pRE11-env) and transfected into 293FT cells. Transfected 293FT cells were incubated with the indicator cell lines independently. In developing the assay, we selected the DSP_{1-7}-positive clones that showed the highest GFP activity after complementation with DSP_{8-11}. These cell lines, designated N4R5-DSP_{1-7}, N4X4-DSP_{1-7} were used for subsequent assays.

Results: The env gene from the reference strains (BaL for R5 virus, NL4-3 for X4 virus, SF2 for dual tropic virus) subcloned in pRE11 and tested, was concordant with the expected co-receptor usage. Assay results were available in two ways (RL or GFP). The assay sensitivity by RL activity was comparable with those of the published phenotypic assays using pseudovirus. The shortest turnaround time was 5 days after obtaining the patient's plasma. All clinical samples gave positive RL signals on R5 indicator cells in the fusion assay. Median RLU value of the low CD4 group was significantly higher on X4 indicator cells and suggested the presence of more dual or X4 tropic viruses in this group of patients. Comparison of representative samples with Geno2Pheno [co-receptor] assay was concordant.

Conclusions: A new cell-fusion-based, high-throughput PTA for HIV-1, which would be suitable for in-house studies, was developed. Equipped with two-way reporter system, RL and GFP, DSP-Pheno is a sensitive test with short turnaround time. Although maintenance of cell lines and laboratory equipment is necessary, it provides a safe assay system without infectious viruses. With further validation against other conventional analyses, DSP-Pheno may prove to be a useful laboratory tool. The assay may be useful especially for the research on non-B subtype HIV-1 whose co-receptor usage has not been studied much.

Keywords: HIV-1; co-receptor; tropism; co-receptor usage; fusion; chemokine receptor.

Received 5 May 2013; **Revised** 10 August 2013; **Accepted** 19 August 2013; **Published** 18 September 2013

Introduction

A new class of drugs to combat HIV-1 infection emerged in 2007 with the marketing approval of maraviroc, a small molecule that binds specifically to the CCR5 co-receptor to block viral attachment and entry [1]. While entry inhibitors are a welcome addition to the antiretroviral arsenal, one problem with this new class of drugs is that treatment is effective only against viruses with the specified co-receptor usage. HIV-1 tropism is defined by the chemokine co-receptors used for viral attachment: R5-tropic viruses use CD4/CCR5, X4-tropic viruses use CD4/CXCR4 and R5X4- or dual-tropic viruses use both CD4/CCR5 and CD4/CXCR4 [2–4]. In clinical treatment with maraviroc, the presence of X4- or dual-tropic viruses is associated with treatment failure [5,6], and a tropism assay is mandatory before treatment initiation.

HIV-1 tropism may be examined genotypically or phenotypically. Genotypic tropism assay (GTA) is based on DNA amplification and sequencing of the third variable (V3) region of the envelope glycoprotein gp120, shown by genetic mapping to be the major determinant of HIV-1 tropism [7–10]. GTA has advantages of platform portability, low cost and rapid turnaround time [11]; however, the interpretation of the sequences is complicated because of the high variability [12]. The assay used in association with maraviroc treatment is the phenotypic tropism assay (PTA) Trofile™ (Monogram Biosciences Inc., CA, USA), a CD4 cell culture assay using replication-defective pseudoviruses [13]. Although Trofile™ is

considered the gold standard, a simpler and effective PTA would be useful.

Here, we describe a novel, cell-fusion-based PTA that uses a dual split reporter protein system (DSP) [14,15] to measure HIV-1 tropism by both Renilla luciferase (RL) activity and green fluorescent protein (GFP) activity. We validated the DSP-Pheno assay using HIV-1 reference strains and applied the assay to test clinical samples from patients with HIV-1 infection.

Methods

Approval of the study and recombinant DNA experiments

Plasma samples from HIV-1-positive patients attending the hospital affiliated with the Institute of Medical Science, the University of Tokyo (IMSUT) were collected and kept frozen until use. Patients provided written informed consent, and the study was approved by the Institutional Review Board of the University of Tokyo (approval number 20-31). Recombinant DNA experiments used in this work were approved by the Institutional Review Board (approval number 08-30), and by the review board in the Ministry of Education, Culture, Sports, Science and Technology (MEXT; approval number 23-1927).

Cell lines

Cell lines N4, N4X4 and N4R5 are derived from the human glioma NP-2 cell line and stably express CD4, CXCR4 and CCR5, respectively [16,17]. NP-2-derived cell lines were grown in M10+ medium (modified Eagle's medium (MEM; Sigma, St. Louis, MO, USA) supplemented with 10% heat-inactivated foetal bovine serum (FBS), 100 units/ml of penicillin and 0.1 mg/ml of streptomycin). 293FT cells (Invitrogen, Carlsbad, CA, USA) were grown in D10+ medium (Dulbecco's modified Eagle's medium (DMEM, Sigma) supplemented with 10% FBS, 100 units/ml of penicillin and 0.1 mg/ml of streptomycin). All cell cultures were maintained at 37°C in a humidified 5% CO_2 incubator.

Reference viral envelopes

The plasmids encoding reference HIV-1 envelopes with well-characterized co-receptor usage were obtained from NIH AIDS Research & Reference Reagent Program (NIH ARRRP: Germantown, MD, USA). NL4-3 and LAI represented X-4 tropic viruses; BaL represented R5-tropic viruses; and SF2 represented dual (R5X4)-tropic viruses. The co-receptor usage of these laboratory strains has been published [13,18–22].

Construction of DSP expression plasmids

The DSP system utilizes a pair of chimeric reporter proteins, DSP_{1-7} and DSP_{8-11}, each of which is a fusion of split green fluorescent protein (spGFP) and split Renilla luciferase (spRL) [15]. DSP_{1-7} fuses the N-terminal region of RL (amino acids 1–229) to the N-terminal region of GFP (amino acids 1–157), with a linker sequence separating the two regions. DSP_{8-11} has the complementary structure, with the C-terminal region of GFP (amino acids 158–231), fused to the C-terminal region of RL (amino acids 230–311), also separated by a linker sequence. When both reporter proteins are present in the same cell, they each recover full activity.

To generate pLenti-DSP_{1-7} plasmid, we first amplified an attB-flanked DSP_{1-7} fragment (1251 bp) using pDSP $_{1-7}$ as a template and attB-flanked primers [attB1-DSP1-1F (56-mer, 5′-GGGGACAAGTTTGTACAAAAAAGCAGGCTGGGCTAGCCACCA TGGCTTCCAAGGTG -3′) and attB2-DSP1-1R (51-mer, 5′-GG GGACCACTTTGTACAAGAAAGCTGGGTGCTCTAGATCACTTGT CGGCGG-3′)]. Sequential amplicons were transferred to pDONR-221 and pLenti6.3/V5-DEST (Invitrogen) using the ViraPower™ HiPerform™ Lentiviral Gateway® Expression System (Invitrogen) according to the manufacturer's protocol. Constructs were verified by sequencing.

An expression vector, pRE11 (Figure 1), was constructed for the co-expression of DSP_{8-11} and HIV-1 *env* by multiple rounds of PCR and subcloning. Source plasmids were pIRES2-AcGFP1 (Clontech), pmOrange (Clontech), $pDSP_{8-11}$ [15] and pmirGLO (Promega). pRE11 incorporated multiple cloning sites under the PGK promoter for the insertion of HIV-1 *env* (Shown as 5′-XbaI-XhoI-3′ in Figure 1b). Necessary restriction enzyme cleavage sites used for construction, including multiple cloning sites (XbaI-MluI-SwaI-AgeI-XhoI), were created using synthetic oligonucleotides and PCR. A CMV promoter drives $pDSP_{8-11}$ directly. The same CMV promoter expresses mOrange with a nuclear localization signal that serves as a marker for successful transfection via internal ribosomal entry site (IRES). All PCR fragments were confirmed by sequencing.

NP-2-derived fusion indicator cell lines

We used the ViraPower Packaging Mix with Lipofectamine 2000 (Invitrogen) to transfect 293FT cells with pLenti-DSP_{1-7} and create pseudoviruses containing the DSP_{1-7} expression cassette (Lenti-DSP_{1-7}). We next infected cell lines NP-2/CD4 (N4), CD4/CXCR4 (N4X4) and CD4/CCR5 (N4R5) with pseudoviruses containing LentiDSP_{1-7} for 2 hours. Cells were distributed in 96-well tissue culture plates at a density of 75 cells/plate (0.8 cell/well) and grown in the presence of 4 μg/ml blasticidin. Approximately 50 candidate clones from each cell line were randomly selected and tested for FITC intensity using FACS Calibur (BD Biosciences, Franklin Lakes, NJ, USA) 48 hours after transfection of $pDSP_{8-11}$. FACS data were analyzed by Flow Jo version 8.7.1 (Tree Star Inc., OR, USA). Clones with the highest median FITC intensity were expanded in M10+ supplemented with 4 μg/ml of blasticidin (M10 + 4) for further assays.

Generation of pRE11-env strains

Full-length HIV-1 *env* was prepared by PCR amplification from clinical plasma samples as described [23]. Viral RNA was extracted from 140 μl of patient's plasma by QIAamp Viral RNA Mini kit according to the manufacturer's recommendation (QIAGEN, Hilden, Germany). One-step RT-PCR using SuperScript III and High Fidelity Platinum® Taq DNA polymerase (Invitrogen) was carried out in five separate 15-μl reactions to minimize the bias created by PCR. The reaction mixture contained 2 μl of RNA template, 7.5 μl of 2 × reaction buffer, 0.3 μl of 5 mM $MgSO_4$, 0.3 μl of each 10 μM of forward primer (Env-1F, 25-mer, 5′-TAGAGCCCTG GAAGCATCCAGGAAG-3′) and reverse primer (Env-3Rmix, equimolar mixture of 30-mer, 5′ -TGCTGTATTGCTACTTGTGATTGCTCCATA-3′ and 30-mer, 5′ -TGCTGTATTGCTA CTTGTGATTGCTCCATG-3′), 0.6 μl of

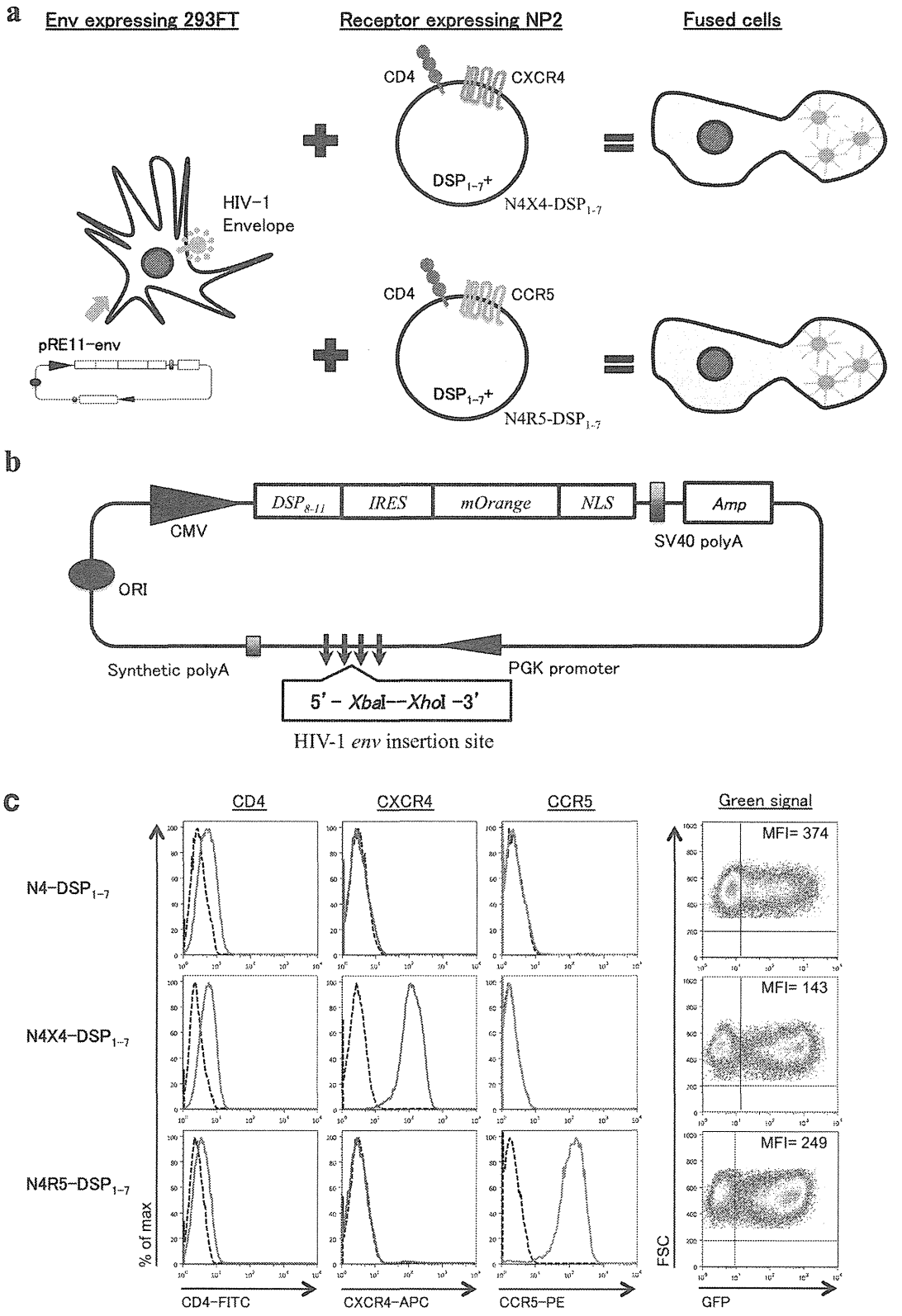


Figure 1. **A cell-fusion-based phenotypic tropism assay for HIV-1: DSP-Pheno. (a) Schematic representation of DSP-Pheno assay system. (b) Schematic representation of pRE11, an expression vector for HIV-1 env and DSP_{8-11}. pRE11 encodes also mOrange with a nuclear localization signal as an indicator of transfection. (c) NP-2-derived clones stably expressing DSP_{1-7} (N4-DSP_{1-7}, N4X4-DSP_{1-7} and N4R5-DSP_{1-7}) were selected by the high GFP expression after direct transfection of $pDSP_{8-11}$. The expression of CD4/co-receptors was reconfirmed by appropriate monoclonal antibodies and FACS analysis.**

SuperScript III and High Fidelity Platinum® Taq DNA polymerase, 0.25 μl of RNAse OUT and 3.75 μl of nuclease-free water with the final volume of 15 μl/reaction. The one-step RT-PCR condition was 55°C for 30 minutes, 94°C for 2 minutes followed by 30 cycles of 94°C for 20 seconds, 55°C for 30 seconds, 68°C for 4 minutes, then extension at 68°C for 5 minutes. The fragment by the first-round amplification extended from NL4-3 reference position of 5853–8936. Products from five independent reactions were combined. Four microliters of the mixed first-round PCR products were used as the template for each of five independent second-round PCR reactions employing EnvB-2F-Xba (41-mer, 5′-TAGCTCTAGAACGCGTCTTAGGCATCTCCTATGGCAG GAAG-3′) and EnvB-4R-Xho (41-mer, 5′ -TAGCCTCGAGACCGGT TACTTT TTGACCACTTGCCACCCAT-3′) as the forward and reverse primers, respectively. The second PCR was carried out according to the standard 50-μl PCR protocol of the Platinum® PCR SuperMix High Fidelity as described above. The fragments amplified by the second PCR extended from NL4-3 reference position of 5957–8817. After digestion with Xba I and Xho I, about 3-kb PCR products were purified by 1.2% agarose gel and QIAquick gel extraction kit (Qiagen). The purified products were inserted into pRE11 at XbaI and XhoI sites, resulting in HIV-1*env* expression plasmid (pRE11-envbulk) from each patient. pRE11-envbulk, representing a quasispecies of *env* population from each patient, was prepared by transfecting into *E. coli* JM109. For bulk analysis, transfected JM109 was expanded to 25 ml, followed by QIAGEN Plasmid Midi Kit (Qiagen) for DNA extraction.

Cell-fusion assay

On the day before transfection, 500 μl aliquots of 293FT cells in DMEM supplemented with 10% FBS (D10) were seeded in 24-well tissue culture plates at a density of 2.8×10^5 cells/well and incubated overnight to 70–80% confluency. The cells were then transfected with pRE11-envstrain or pRE11-envbulk according to the manufacturer's protocol (Roche). On the same day, 100 μl aliquots of N4-DSP_{1-7}, N4X4-DSP_{1-7} and N4R5-DSP_{1-7} cells in MEM supplemented with 10% FBS (M10) were seeded in a 96-well tissue culture, optical bottom plate (NUNC, Thermo Fisher Scientific Inc., NY, USA) at a density of 1×10^4 cells/well and incubated at 37°C. Forty-eight hours after transfection, the medium of transfected 293FT cells was removed by aspiration and replaced with 1 ml of PBS (Sigma) at RT. Transfected 293FT cells were resuspended by gentle pipetting.

To start the cell-fusion assay, 150 μl/well of transfected cells were overlaid onto N4-DSP_{1-7}, N4X4-DSP_{1-7} and N4R5-DSP_{1-7} cells. The cells were incubated for fusion at 37°C in a humidified 5% CO_2 incubator for 6 hours, and then analyzed by automatic image capture using an In Cell Analyzer 1000 (GE Healthcare). Four fields/well of image were captured through red, green and bright field channels, and fused cells were identified by the presence of two or more red nuclei surrounded by a green area (cytoplasm). Immediately after image capturing, EnduRen™ Live Cell Substrate (Promega) was added to each well, and luciferase activity was measured three times using a Glomax 96 microplate luminometer (Promega), according to the manufacturer's instructions. The mean luciferase activity, recorded as relative light unit (RLU), was the average of three measurements per well. The experiments were conducted in triplicates and repeated independently at least three times.

To test the co-receptor specificity, 2 μM/well of the appropriate inhibitor was added to the cells 90 minutes prior to the cell-fusion assay (CXCR4 inhibitor AMD3100 (Sigma) to N4X4-DSP_{1-7} cells and CCR5 inhibitor maraviroc (Sigma) to N4R5-DSP_{1-7} cells).

Genotyping

pRE11-envbulk were sequenced in both the 5′ and 3′ directions using population-based sequencing on the ABI 3130xl genetic analyzer (Applied Biosystems, Foster City, CA, USA) using BigDye Terminator V3.1 (Applied Biosystems) with forward primer E110 (5′ -CTGTTAAATGGCAGTCTAGCAGAA-3′), and reverse primer Er115 (5′ -AGAAAAATTCCCCTCCACAATT AA-3′). The V3 nucleotide sequences were submitted to the Geno2Pheno (co-receptor) algorithm (http://coreceptor.bioinf.mpi-inf.mpg.de) setting the false positive rate (FPR) at 10%.

Results

Construction of DSP_{1-7} and DSP_{8-11} expression vectors

We inserted DSP_{1-7} or DSP_{8-11} sequences into blasticidin-resistant lentivirus vectors and then infected NP-2/CD4 (N4), NP-2/CD4/CXCR4 (N4X4) and NP-2/CD4/CCR5 (N4R5) cells with the recovered pseudoviruses, selecting for blasticidin-resistant clones. We screened 49, 51 and 43 lentivirus-infected and blasticidin-resistant clones from N4, N4X4 and N4X5, respectively, for high levels of DSP_{1-7} or DSP_{8-11} expression following super-transfection with the complementary plasmid (pDSP_{8-11} or pDSP_{1-7}, respectively). From each of the cell lines, we selected the blasticidin-resistant and DSP_{1-7}-positive clone that showed the highest GFP activity after complementation with DSP_{8-11}. These cell lines, designated N4-DSP_{1-7}, N4X4-DSP_{1-7} and N4R5-DSP_{1-7}, were re-evaluated for their expression of CD4, CXCR4 and CCR5 on the cell surface (Figure 1c).

Using this approach, we obtained N4- and N4R5-cells expressing high levels of DSP_{8-11}, but were unable to obtain a stable N4X4 cell line expressing DSP_{8-11} (data not shown). To circumvent this problem, we decided to generate 293FT cells transiently expressing both DSP_{8-11} and the HIV-1 *env* protein and develop a cell-fusion assay system using those cells together with the NP-2-derived cells stably expressing DSP_{1-7} (Figure 1a). Thus, we constructed the expression vector pRE11, containing the DSP_{8-11} expression cassette and cloning sites for insertion of HIV-1 *env* sequences under the control of the PGK promoter (Figure 1b).

Validation of the cell-fusion assay using the *env* gene from laboratory HIV-1 strains

We validated the DSP assay system (DSP-pheno) using pRE11 constructs engineered to contain *env* sequences from reference strains with known co-receptor usage. The *env* reference constructs, which also contained the DSP_{8-11} expression cassette, were the following: pRE11-HXB2, pRE11-LAI, and pRE11-NL4-3 (X4 strains); pRE11-BaL (R5 strain); and pRE11-SF2 (dual strain). The cell-fusion assays were performed with N4X4-DSP_{1-7} or N4R5-DSP_{1-7} cells in

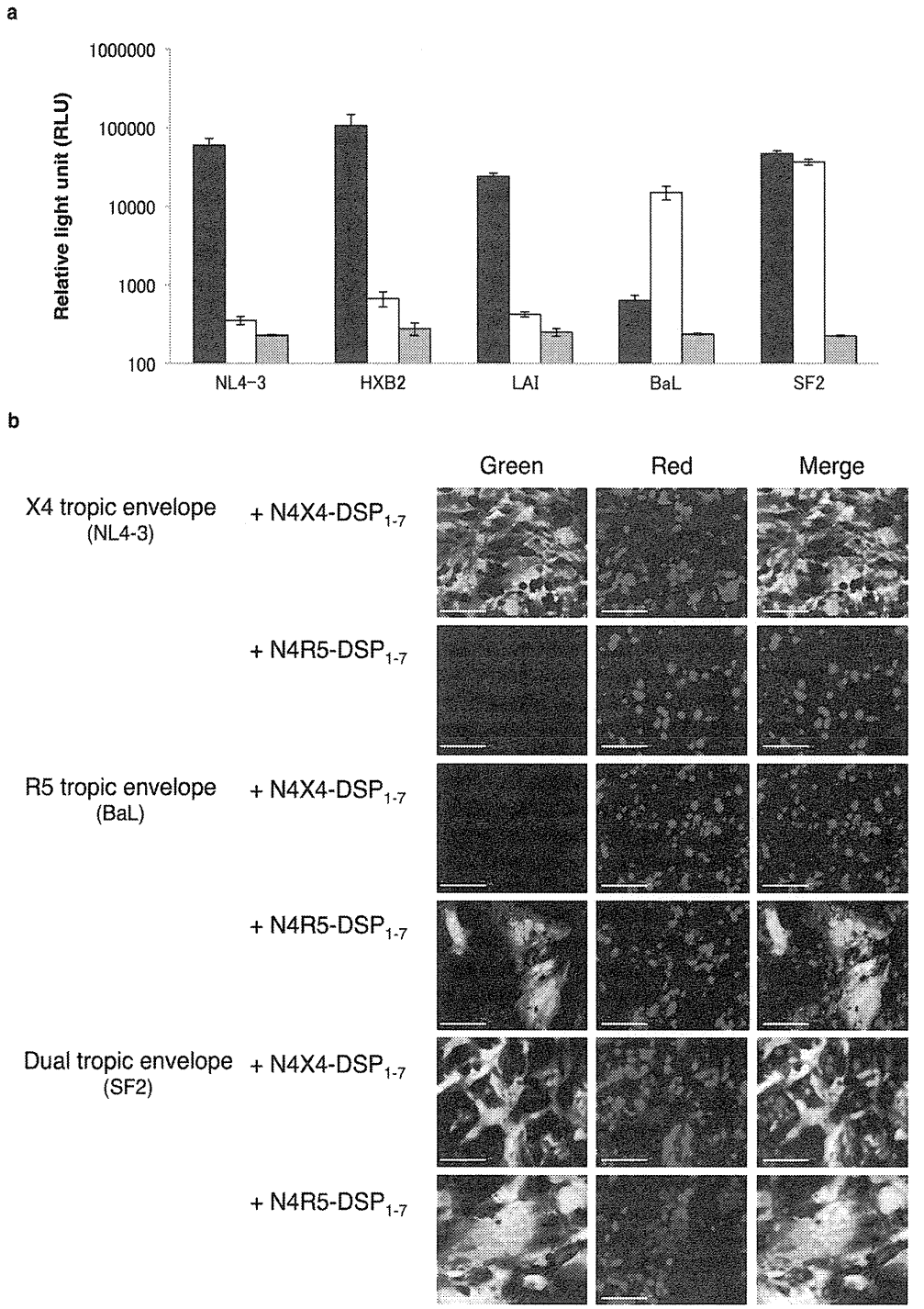


Figure 2. Validation of the cell-fusion assay using env genes from HIV-1 reference strains. (a) RL activities after cell fusion were measured using NL4-3, HXB2 and LAI as X4 reference strains while BaL or SF2 as R5 or R5X4 strains, respectively. Black columns: RL activities on N4X4-DSP$_{1-7}$. White columns: activities on N4R5-DSP$_{1-7}$. Grey columns: activities on N4-DSP$_{1-7}$. Small bars at the top of each column indicate the mean RLU $\pm$ SD from three independent experiments. (b) Successful cell fusion is indicated by the green fluorescence in the cytoplasm. Red spots were mOrange activity in the nuclei showing successful transfection. Merged images showed multinuclear cells with multiple yellow/orange nuclei surrounded by green cytoplasm.

combination with 293FT cells transiently expressing one of the pRE11-env constructs. In all the assays, both RL and GFP activities were restored only when cells expressing the appropriate *env* and co-receptor combinations were co-cultured (Figure 2a and b). Co-culture of N4X4-DSP$_{1-7}$ or N4R5-DSP$_{1-7}$ in combination with the 293FT cells transiently expressing the pRE11-env constructs of discordant tropism served as a negative control for expression of RL activities

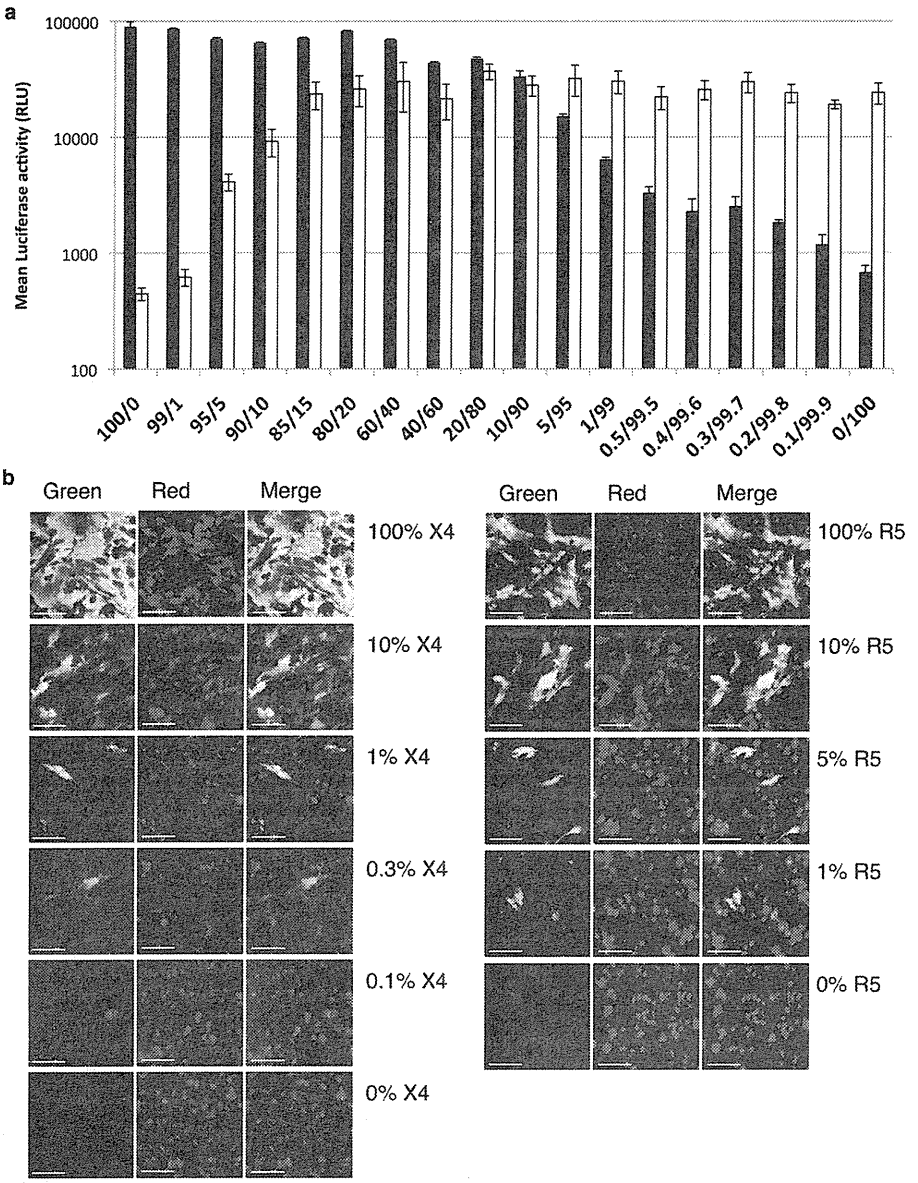


Figure 3. Assay sensitivities for minor populations. pRE11-NL4-3 (X4) and pRE11-BaL (R5) were mixed at indicated ratios. The total plasmid concentration in the mixture was adjusted to 100 ng/μl. The mixture was transfected to 293FT cells and fused to the indicator cells. (a) RL activities. Black columns show the RL activities on X4 indicator (N4X4-DSP$_{1-7}$), while white columns on R5-indicator (N4R5-DSP$_{1-7}$). Small bars at the top of each column indicate the mean RLU ± SD from three independent experiments. (b) GFP activities. Left panel shows activities of the mixture with indicated ratio on X4 indicator (N4X4-DSP$_{1-7}$), while right panel on R5-indicator (N4R5-DSP$_{1-7}$).

(Figure 2a) and GFP signals (Figure 2b). The absence of RL activities on N4-DSP$_{1-7}$ confirmed the importance of co-receptors for the fusion.

Assay detection thresholds and sensitivity for minor populations

To evaluate assay sensitivity in identifying minor variants within a single sample, we mixed pRE11-NL4-3 (X4) and pRE-BaL (R5) in varying ratios and measured RL activities and GFP signals (Figure 3a and b). Both methods of detection identified X4 viruses more readily than R5 viruses. Based on luciferase activity, the presence of approximately 0.3% X4 viruses gave values significantly higher than background (0% X4), while R5 viruses had to comprise approximately 5% of the mixture for the signal to be detectable over background (Figure 3a). Similarly, based on GFP signals, X4 viruses comprising as little as 0.1% of the mixture could be detected, while detection of R5 viruses had a minimum threshold of approximately 1% (Figure 3b).

Validation of the chemokine receptor specificity using the CXCR4 inhibitor AMD3100 and CCR5 inhibitor maraviroc

293FT cells expressing *env* from reference strains NL4-3 (X4) or BaL (R5) were co-cultured with N4X4-DSP$_{1-7}$ or N4R5-DSP$_{1-7}$ cells in the absence or presence of AMD3100 or maraviroc (Figure 4a and b). In the absence of inhibitors, RL activities of