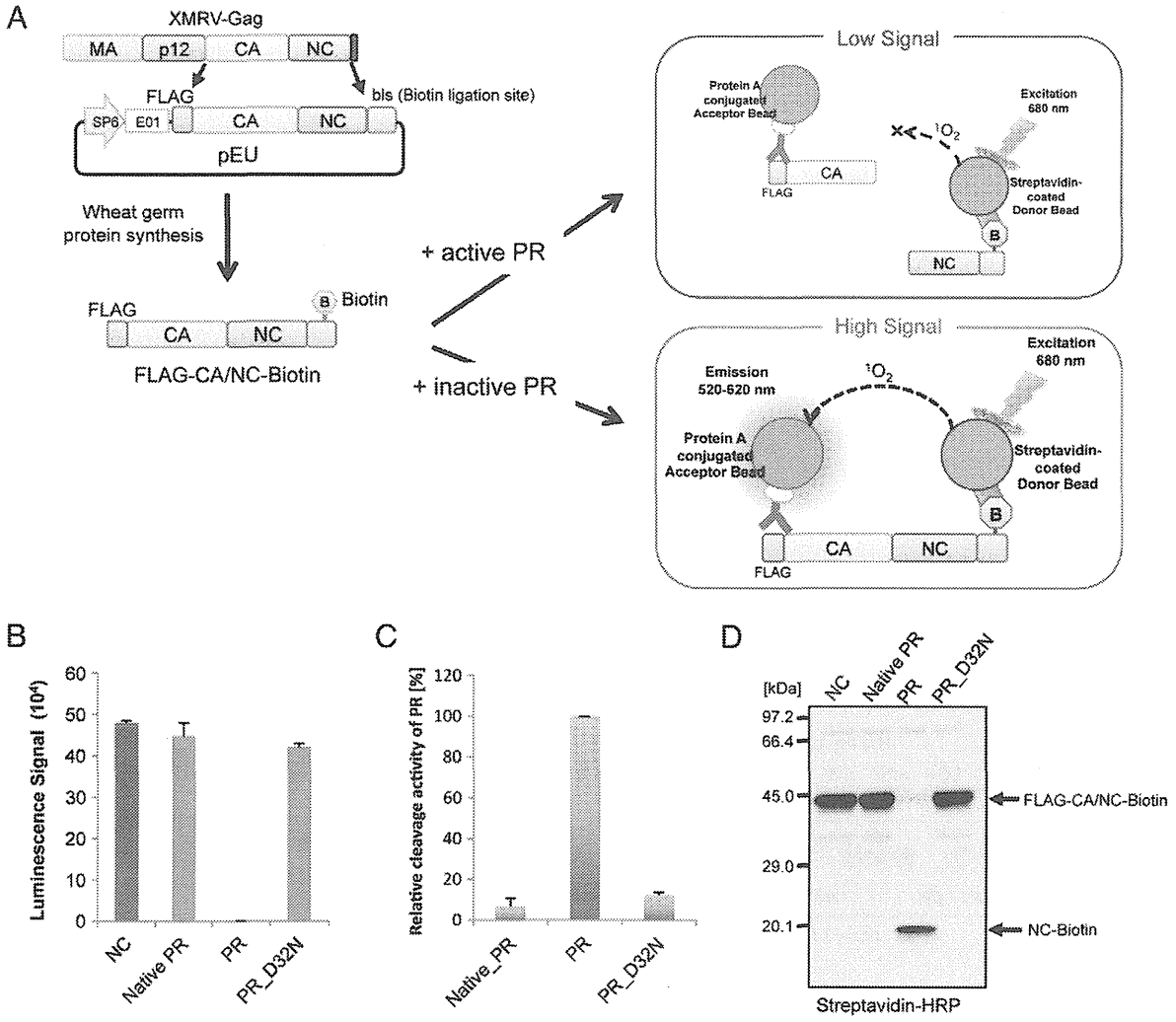


Fig. 2 – Development of a cleavage activity assay for XMRV PR using the luminescent assay AlphaScreen. A. Schematic diagram of the substrate construction of XMRV PR and detection system for the cleavage activity of XMRV PR by luminescent analysis. Substrate was designed as XMRV Gag capside (CA) and nucleocapsid (NC) flanked by N-terminal FLAG and C-terminal biotin (FLAG_CA-NC_biotin). PR was incubated with the substrate, for 1 h at 37 °C. Subsequently, protein A-conjugated acceptor beads with anti-FLAG antibody and streptavidin coated donor beads were added and bound to the tagged substrate. Upon laser excitation, Donor beads convert ambient oxygen to a singlet oxygen. In the case of non-activity PR, singlet oxygen transfers across to activate Acceptor beads and subsequently emit light at 520–620 nm. In the case of active PR, no light is produced because the singlet oxygen can not transfer from Donor beads to Acceptor beads due to the distance (>200 nm). B,C,D. Cleavage activity of XMRV PR was quantitated by the luminescent assay (Fig.2B). Actual cleavage of XMRV Gag substrate was also confirmed by immunoblotting with streptavidin-HRP (Fig.2D). The arrow indicates the band for the non-cleaved substrates (FLAG_CA/NC_biotin).

remains controversial [31], XMRV can however proliferate in other human prostate cancer cells such as LNCaP or PC3 without severe cytopathic effects [32]. Such conditions of persistent infection without cell death could conceivably lead to prolonged exposure of host cell proteins to XMRV PR, increasing their susceptibility to cleavage with oncogenic consequences. The important question remains, however, as to whether this virus has indeed tumorigenic capability. Previous reports have indicated that XMRV integration is characterized by a strong preference for transcriptional start sites, CpG islands, and DNase-hypersensitive regions, all features that are frequently associated with structurally-open transcription regulatory regions of the chromosome in prostate cancer cells [33]. Integration of XMRV occurs preferentially in actively-transcribed genes and gene-dense regions within the chromosome [33]. Oncogenic properties of XMRV have been investigated in cell culture models. Although XMRV has been reported to lack direct transforming activity, the virus is able to induce low rates of transformation in cultured fibroblast cells [34]. Therefore, the molecular link between XMRV infection and cell transformation merits further investigation.

Our current data demonstrates that APV is a potent antagonist of XMRV PR. During the preparation of this manuscript, Li et al. reported the crystal structure of complexes

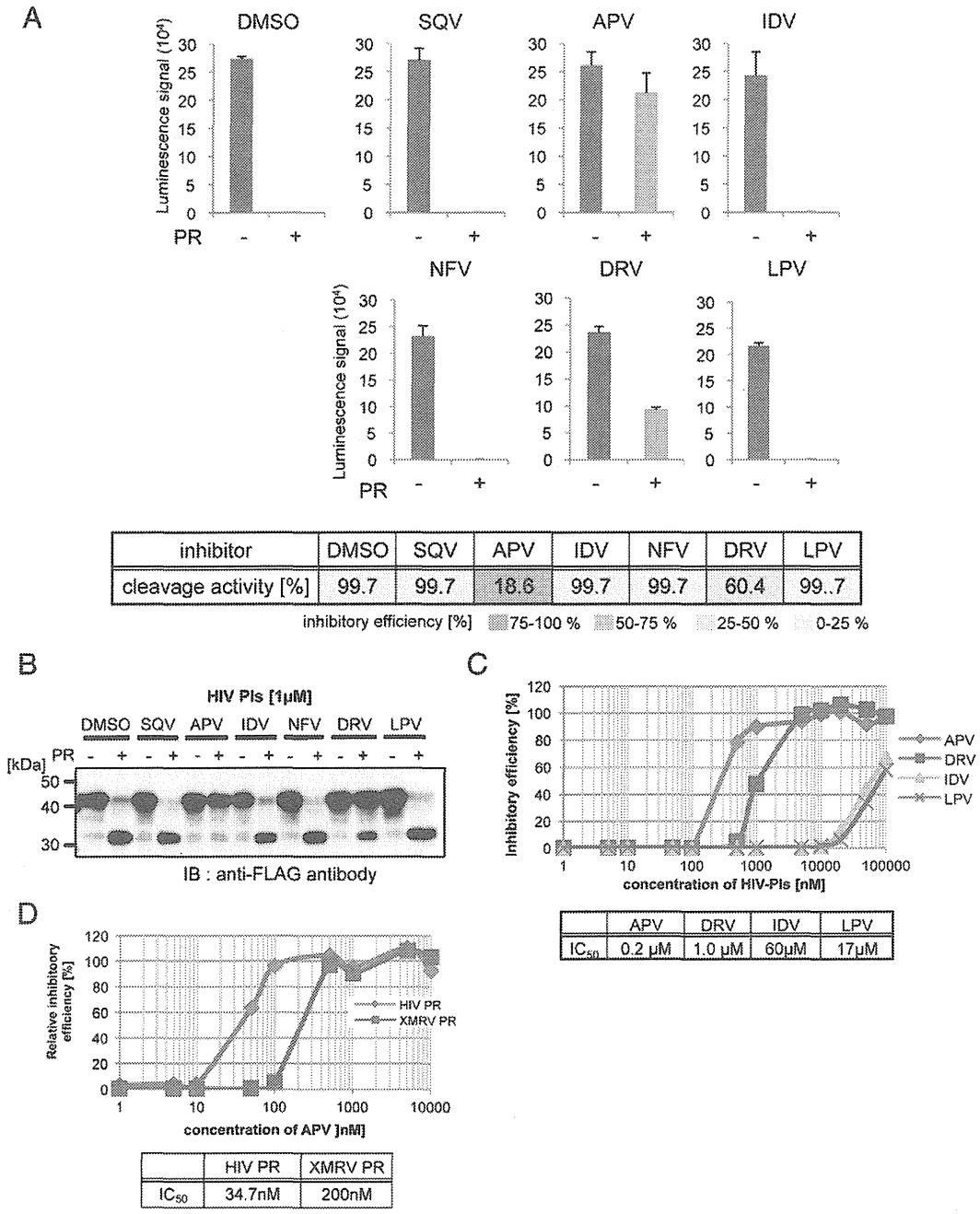
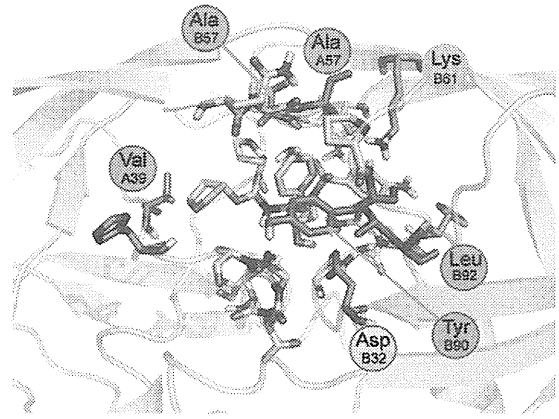


Fig. 3 – Drug screening for XMRV PR based on the cleavage activity. A,B. XMRV protease (+) or DHFR (–) was pre-incubated with indicated HIV PIs (SQV, saquinavir; APV, amprenavir; IDV, indinavir; NFV, nelfinavir ; DRV, darunavir; LPV, lopinavir; 1 μM each) and then subjected to AlphaScreen. Luminiscent AlphaScreen signal (upper panel) and relative enzymatic activity (lower panel) were listed. C. Conformation of the cleavage of the tester polypeptide by immunoblot analysis with anti-FLAG antibody. D. Dose–response curve of XMRV PR with HIV PIs using AlphaScreen (upper panel). IC_{50} values were calculated for each inhibitor (lower panel). E. Dose–response curve of XMRV PR and HIV-1 PR with APV using AlphaScreen (upper panel). IC_{50} values were calculated for each protease (lower panel).

formed between XRMV PR and several protease inhibitors, including APV [24,35]. In the current study we moved a step closer to clarifying the molecular interactions between XMRV PR and APV during drug-resistance, by developing an effective cell-free in vitro protease assay for XMRV PR. This assay revealed that an Ala57Val substitution induced significant drug-resistance to APV regardless of the integrity of the protease activity. The data indicates that this cell-free assay is useful for analyzing the drug-resistance properties of retroviral proteases.

Proteases often modify the activities of their target substrates [36]. Identification of the specific substrates cleaved by viral PR is of great significance for understanding the molecular etiology of virus infection. Proteomic studies with mass spectrometry could, theoretically, exhaustively identify the cellular proteins cleaved

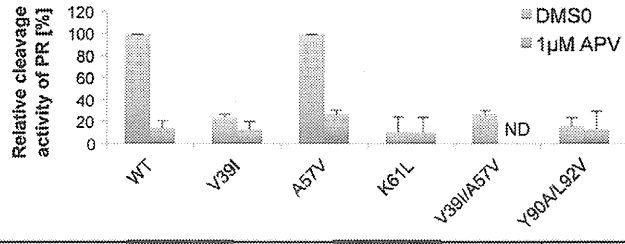
A



B

```
                10          20          30          40          50
XMRV  TLGDQGGQGQ EPPPEPRITL KVGGQPVTFL VDTGAQHSVL TQNPGPLSDK
HIV-1 ------PQITL W--QRPLVTI KIGGQLKEAL LDTGADDTVL EEMSLPGRWK
                5          13          23          33          43
                60          70          80          90          100
XMRV  SAWVQGATGG KRYRWTTDRK VHLATGKVTH SFLHVPDCPY PLLGRDLLTK
HIV-1 PKMIGGIGGF IKVRQYDQIL IEICGHKAIG TVL-VGPTPV NIIGRNLLTQ
                53          63          73          82          92
                110         120    125
XMRV  LKAQIHFEGS GAQVVGPMGQ PLQVL
HIV-1 IGCTLNF--- ---------- -----
          99
```

C

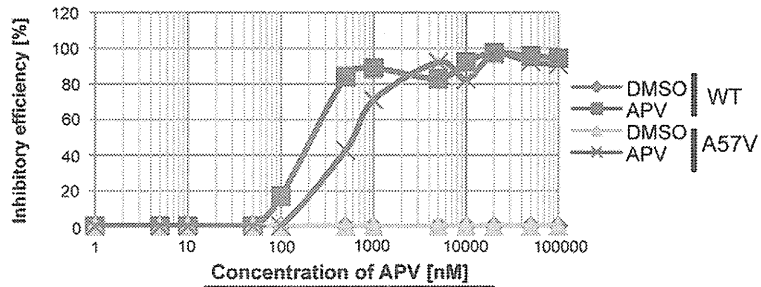


	WT	V39I	A57V	K61L	V39I/A57V	Y90A/L92V
cleavage activity [%]	99.7	23.5	99.7	10.3	27.0	15.6
inhibitory efficiency [%]	13.9	53.0	26.6	95.0	0.0	81.7

cleavage activity [%] 75-100 % 50-75 % 25-50 % 0-25 %

inhibitory efficiency [%] 75-100 % 50-75 % 25-50 % 0-25 %

D



	WT	A57V
IC_{50}	200 nM	596 nM

Fig. 4 – Prediction of the amino acid residues of XMRV PR interacting with APV. A. The predicted 3D-structure for the interaction between XMRV PR and APV. This homology modeling was based on the HIV-1 PR and APV complex as a template. B. Sequence alignment of XMRV PR and HIV-1 PR. The amino acids related to interaction of APV with HIV-1 PR and the corresponding amino acids in XMRV PR are highlighted with red letters. C. Cleavage activity of XMRV PR-WT and its mutants in the presence of 1 μM APV or equivalent amount of DMSO (control). Lower panel is cleavage activity and inhibitory efficiency (APV value/DMSO value) for each XMRV PR. D. Dose–response curve of the inhibitory rate of PR-WT or PR-A57V by APV.

A

No.	Symbol	Gene bank accession No.	aa	Da	Cleavage by XMRV PR
1	ANXA1	BC001275	346	38714	No
2	GJB2	BC017048	226	26215	No
3	ARL11	BC013150	196	21391	Yes
4	RND3	BC012513	244	27368	No
5	FHIT	BC032336	147	16858	No
6	KCNRG	BC032343	272	31048	No
7	SFRP1	BC036503	314	35386	No
8	WNT5A	BC064694	380	42339	No
9	SMAD3	BC050743	425	48081	No
10	p21	BC000312	164	18119	No
11	NME1	BC000293	152	17149	No
12	MAL	BC000458	153	16714	No
13	SPINT2	BC001668	252	28228	No
14	SLC39A1	BC003152	324	34250	No
15	CHEK2	BC004207	543	60915	No
16	DKK3	BC007660	350	38291	Yes
17	TIMP3	BC014277	211	24145	No
18	BAX	BC014175	192	21184	Yes
19	ALPL	BC021289	524	57305	No
20	CST6	BC031334	149	16510	No
21	KLF5	BC042131	457	50792	No
22	CAV1	BC082246	178	20472	No
23	PTEN	BC005821	403	47166	Yes
24	TP53	BC003596	393	43653	No

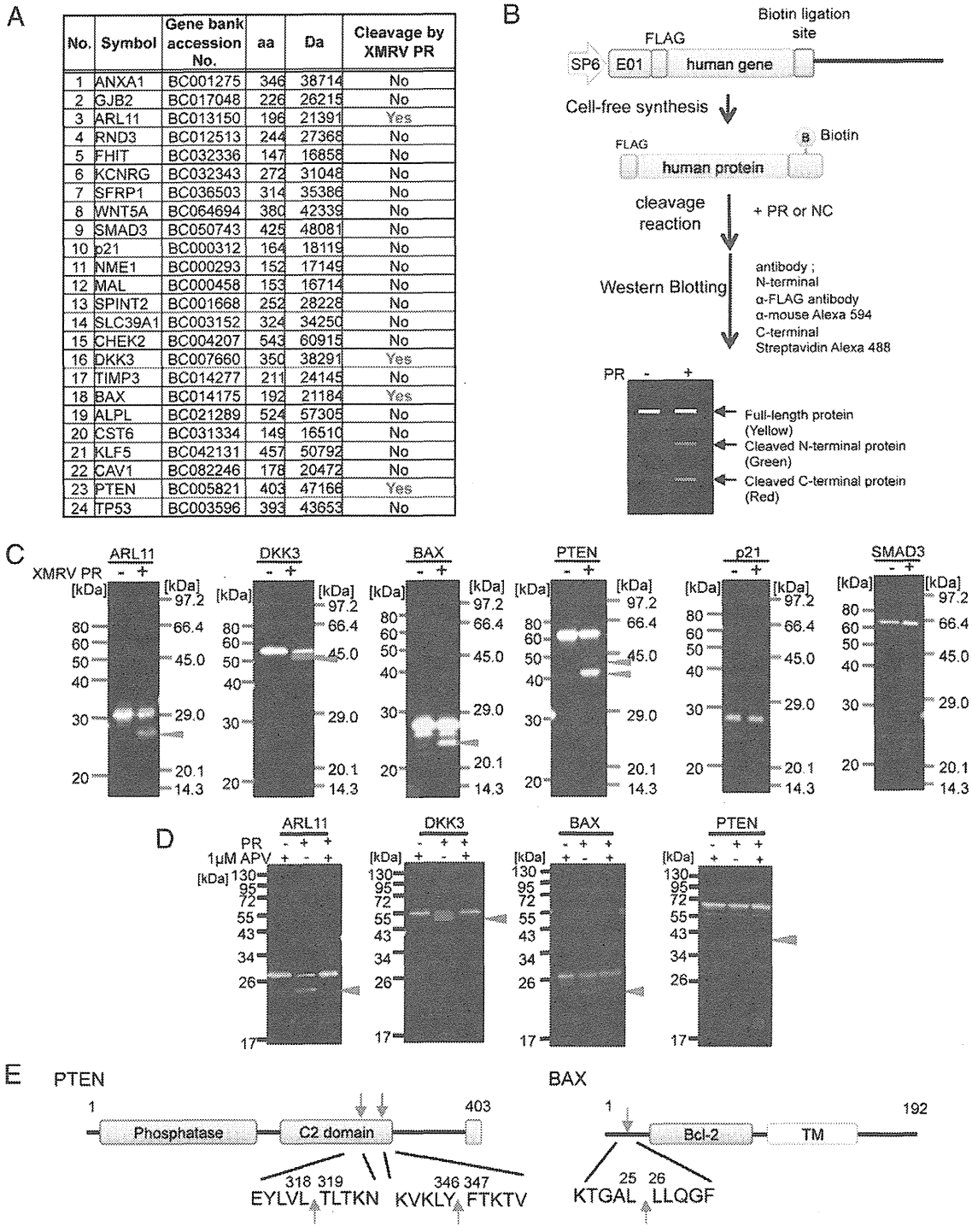


Fig. 5 – Screening of host proteins cleaved by XMRV PR in vitro. A. The list of human tumor suppressor proteins tested in this study. B. Scheme of the tester proteins construction and the cleavage assay system by immunoblotting. The genes were amplified by PCR with primer sets containing either FLAG or biotin ligation site (bls) in the flanking sequence, respectively. The recombinant host proteins flanking FLAG and biotin (FLAG-X-biotin) were incubated with XRMV PR at 37 °C for 2 h followed by SDS-PAGE. The proteins were detected using anti-FLAG-Alexa592 antibody (green) and Alexa488-conjugated streptavidin (red). Full-length protein is seen as a yellow band. C. Tester proteins were treated with XMRV PR or carrier. 2-color immunoblot analysis was performed as in Materials and methods. D. Tester proteins were treated with XMRV PR in the absence or presence of amprenavir. Immunoblot analysis was performed as in C. E. Identification of the cleavage site in the XMRV PR amino acid sequence.

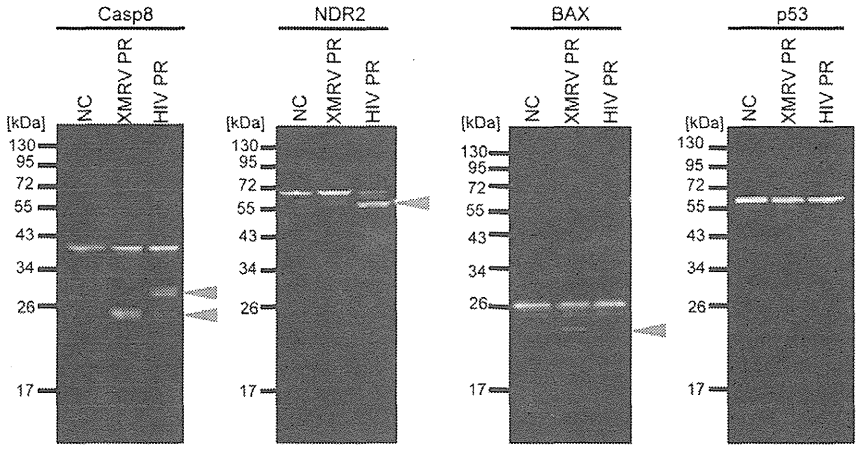


Fig. 6 – Comparative analysis of host proteins cleaved by XMRV PR and HIV-1 PR. The recombinant host proteins flanking FLAG and biotin (FLAG-X-biotin) were incubated with either XRMV PR or HIV-1 PR at 37 °C for 2 h followed by SDS-PAGE. The proteins were detected using anti-FLAG-Alexa592 antibody (green) and Alexa488-conjugated streptavidin (red). Full-length protein is seen as a yellow band. Arrows depict the cleavage products.

by retroviral proteases in infected cells. However, this cell-based will run into difficulty identifying individual substrates if several host proteases act simultaneously on the substrate. To circumvent this potential problem we developed the cell-free in vitro method for the identification of substrates cleavable by XMRV PR. Wheat extracts purified rarely include endogenous proteases that can interfere with the proteolytic reaction, making them suitable for the cell-free protease assay.

Tumor suppressor proteins play a major role in preventing tumor initiation. Our current results demonstrate that XMRV PR can cleave PTEN and BAX tumor suppressors as well as the intrinsic substrate XMRV Gag. It has been reported that the C-terminal region of PTEN is important for the protein's stability, and the C-terminal deletion mutant is degraded rapidly in cells [37]. Since XMRV cleaves within the C-terminal region, the native function and stability of PTEN might be abrogated by XMRV infection. The N-terminal region of BAX has been demonstrated to mediate its activity in apoptosis [38]. We demonstrated in the present study that XMRV PR can cleave the N-terminal region of BAX, suggesting that XMRV infection might affect the activity of BAX protein.

A biochemical approach to the evaluation of PR-inhibitor susceptibility has been attempted previously using several related methods [39,40]. The essence of each of these procedures is the synthesis of catalytically-active PR and substrate peptide and inhibitor in vitro, and measurement of the amount of substrate cleavage. The advantage of this approach is that it can directly detect the catalytic activity of PR. However, it is often difficult to produce sufficient quantities of enzymatically active viral PR in conventional cell-based protein expression systems such as *E. coli* or insect cells. In our current study, we successfully created catalytically-active XMRV PR in a cell-free system that, when mixed with a reporter substrate flanked with N- and C-terminal fluorophores, substrate cleavage could be assayed by AlphaScreen or 2-color IB. This approach directly evaluates the cleavage activity of the PR and, in addition, cleavage sites can be estimated by the size of cleavage products. The current availability of full-length cDNA libraries, derived from higher eukaryotes, will facilitate the in vitro synthesis of full-length proteins, making this cell-free system approach could further be applicable to the assay of a broad range of, not only viral, but also host proteases.

5. Conclusion

We have delineated the molecular and enzymatic characteristics of XMRV PR by utilizing wheat-germ cell-free protein synthesis and AlphaScreen. Furthermore, we have developed an in vitro cleavage assay for drug screening based on the enzymatic activity. Our results suggest that XMRV-protease cleavage of certain host proteins and inhibited by APV. Further in vivo studies with XMRV-infected cells will be necessary to confirm a molecular link between XMRV and human diseases.

Acknowledgments

We thank Drs. G. Quinn, Y. Kojima and A. Kudo for the discussion and comments. This work was supported in part by grants from the Ministry of Education, Culture, Sports, Science and Technology of Japan and Research Grants on HIV/AIDS Health Labour Sciences Research Grant from The Ministry of Health Labour and Welfare of Japan to A.R. MK was supported by grants from MEXT, JST, Sumitomo-Denko and Iwatani.

REFERENCES

[1] Urisman A, Molinaro RJ, Fischer N, Plummer SJ, Casey G, Klein EA, et al. Identification of a novel Gammaretrovirus in prostate tumors of patients homozygous for R462Q RNASEL variant. PLoS Pathog 2006;2:e25.
[2] Schlaberg R, Choe DJ, Brown KR, Thaker HM, Singh IR. XMRV is present in malignant prostatic epithelium and is

associated with prostate cancer, especially high-grade tumors. Proc Natl Acad Sci U S A 2009;106:16351–6.
[3] Knox K, Carrigan D, Simmons G, Teque F, Zhou Y, Hackett Jr J, et al. No evidence of murine-like gammaretroviruses in CFS patients previously identified as XMRV-infected. Science 2011;333:94–7.
[4] van Kuppeveld FJ, van der Meer JW. XMRV and CFS–the sad end of a story. Lancet 2012;379:e27–8.
[5] Simmons G, Glynn SA, Komaroff AL, Mikovits JA, Tobler LH, Hackett Jr J, et al. Failure to confirm XMRV/MLVs in the blood of patients with chronic fatigue syndrome: a multi-laboratory study. Science 2011;334:814–7.
[6] Paprotka T, Delviks-Frankenberry KA, Cingoz O, Martinez A, Kung HJ, Tepper CG, et al. Recombinant origin of the retrovirus XMRV. Science 2011;333:97–101.
[7] Groom HC, Yap MW, Galao RP, Neil SJ, Bishop KN. Susceptibility of xenotropic murine leukemia virus-related virus (XMRV) to retroviral restriction factors. Proc Natl Acad Sci U S A 2010;107:5166–71.
[8] Dong B, Silverman RH. Androgen stimulates transcription and replication of xenotropic murine leukemia virus-related virus. J Virol 2010;84:1648–51.
[9] Abudu A, Takaori-Kondo A, Izumi T, Shirakawa K, Kobayashi M, Sasada A, et al. Murine retrovirus escapes from murine APOBEC3 via two distinct novel mechanisms. Curr Biol 2006;16:1565–70.
[10] Ventoso I, Blanco R, Perales C, Carrasco L. HIV-1 protease cleaves eukaryotic initiation factor 4G and inhibits cap-dependent translation. Proc Natl Acad Sci U S A 2001;98:12966–71.
[11] Zaragoza C, Saura M, Padalko EY, Lopez-Rivera E, Lizarbe TR, Lamas S, et al. Viral protease cleavage of inhibitor of kappaBalpha triggers host cell apoptosis. Proc Natl Acad Sci U S A 2006;103:19051–6.
[12] Takai K, Sawasaki T, Endo Y. Practical cell-free protein synthesis system using purified wheat embryos. Nat Protoc 2010;5:227–38.
[13] Kamura N, Sawasaki T, Kasahara Y, Takai K, Endo Y. Selection of 5′-untranslated sequences that enhance initiation of translation in a cell-free protein synthesis system from wheat embryos. Bioorg Med Chem Lett 2005;15:5402–6.
[14] Tadokoro D, Takahama S, Shimizu K, Hayashi S, Endo Y, Sawasaki T. Characterization of a caspase-3-substrate kinome using an N- and C-terminally tagged protein kinase library produced by a cell-free system. Cell Death Dis 2010;1:e89.
[15] Akagi T, Shimizu K, Takahama S, Iwasaki T, Sakamaki K, Endo Y, et al. Caspase-8 cleavage of the interleukin-21 (IL-21) receptor is a negative feedback regulator of IL-21 signaling. FEBS Lett 2011;585:1835–40.
[16] Sakuma R, Sakuma T, Ohmine S, Silverman RH, Ikeda Y. Xenotropic murine leukemia virus-related virus is susceptible to AZT. Virology 2010;397:1–6.
[17] Sawasaki T, Kamura N, Matsunaga S, Saeki M, Tsuchimochi M, Morishita R, et al. Arabidopsis HY5 protein functions as a DNA-binding tag for purification and functional immobilization of proteins on agarose/DNA microplate. FEBS Lett 2008;582:221–8.
[18] Takahashi H, Nozawa A, Seki M, Shinozaki K, Endo Y, Sawasaki T. A simple and high-sensitivity method for analysis of ubiquitination and polyubiquitination based on wheat cell-free protein synthesis. BMC Plant Biol 2009;9:39.
[19] Sawasaki T, Gouda MD, Kawasaki T, Tsuboi T, Tozawa Y, Takai K, et al. The wheat germ cell-free expression system: methods for high-throughput materialization of genetic information. Methods Mol Biol 2005;310:131–44.
[20] Sawasaki T, Hasegawa Y, Tsuchimochi M, Kamura N, Ogasawara T, Kuroita T, et al. A bilayer cell-free protein synthesis system for high-throughput screening of gene products. FEBS Lett 2002;514:102–5.
[21] Matsuoka K, Komori H, Nose M, Endo Y, Sawasaki T. Simple screening method for autoantigen proteins using the N-terminal biotinylated protein library produced by wheat cell-free synthesis. J Proteome Res 2010;9:4264–73.
[22] Baker D, Sali A. Protein structure prediction and structural genomics. Science 2001;294:93–6.
[23] Labute P. The generalized Born/volume integral implicit solvent model: estimation of the free energy of hydration using London dispersion instead of atomic surface area. J Comput Chem 2008;29:1693–8.
[24] Li M, Dimaio F, Zhou D, Gustchina A, Lubkowski J, Dauter Z, et al. Crystal structure of XMRV protease differs from the structures of other retropepsins. Nat Struct Mol Biol 2011;18:227–9.
[25] Johnson VA, Brun-Vezinet F, Clotet B, Gunthard HF, Kuritzkes DR, Pillay D, et al. Update of the drug resistance mutations in HIV-1. Top HIV Med 2008;16:138–45.
[26] Alvarez E, Castello A, Menendez-Arias L, Carrasco L. HIV protease cleaves poly(A)-binding protein. Biochem J 2006;396:219–26.
[27] Bellecave P, Sarasin-Filipowicz M, Donze O, Kennel A, Gouttenoire J, Meylan E, et al. Cleavage of mitochondrial antiviral signaling protein in the liver of patients with chronic hepatitis C correlates with a reduced activation of the endogenous interferon system. Hepatology 2010;51:1127–36.
[28] Nie Z, Phenix BN, Lum JJ, Alam A, Lynch DH, Beckett B, et al. HIV-1 protease processes procaspase 8 to cause mitochondrial release of cytochrome c, caspase cleavage and nuclear fragmentation. Cell Death Differ 2002;9:1172–84.
[29] Devroe E, Silver PA, Engelman A. HIV-1 incorporates and proteolytically processes human NDR1 and NDR2 serine-threonine kinases. Virology 2005;331:181–9.
[30] Knouf EC, Metzger MJ, Mitchell PS, Arroyo JD, Chevillet JR, Tewari M, et al. Multiple integrated copies and high-level production of the human retrovirus XMRV (xenotropic murine leukemia virus-related virus) from 22Rv1 prostate carcinoma cells. J Virol 2009;83:7353–6.
[31] Stieler K, Schindler S, Schlomm T, Hohn O, Bannert N, Simon R, et al. No detection of XMRV in blood samples and tissue sections from prostate cancer patients in Northern Europe. PLoS One 2011;6:e25592.
[32] Rodriguez JJ, Goff SP. Xenotropic murine leukemia virus-related virus establishes an efficient spreading infection and exhibits enhanced transcriptional activity in prostate carcinoma cells. J Virol 2010;84:2556–62.
[33] Kim S, Kim N, Dong B, Boren D, Lee SA. Das Gupta J, et al. Integration site preference of xenotropic murine leukemia virus-related virus, a new human retrovirus associated with prostate cancer. J Virol 2008;82:9964–77.
[34] Metzger MJ, Holguin CJ, Mendoza R, Miller AD. The prostate cancer-associated human retrovirus XMRV lacks direct transforming activity but can induce low rates of transformation in cultured cells. J Virol 2010;84:1874–80.
[35] Li M, Gustchina A, Matuz K, Tozser J, Namwong S, Goldfarb NE, et al. Structural and biochemical characterization of the inhibitor complexes of xenotropic murine leukemia virus-related virus protease. FEBS J 2011;278:4413–24.
[36] Etlinger JD, Gu M, Li X, Weitman D, Rieder RF. Protease/inhibitor mechanisms involved in ATP-dependent proteolysis. Rev Biol Cell 1989;20:197–216.
[37] Georgescu MM, Kirsch KH, Akagi T, Shishido T, Hanafusa H. The tumor-suppressor activity of PTEN is regulated by its carboxyl-terminal region. Proc Natl Acad Sci U S A 1999;96:10182–7.
[38] Toyota H, Yanase N, Yoshimoto T, Moriyama M, Sudo T, Mizuguchi J. Calpain-induced Bax-cleavage product is a more potent inducer of apoptotic cell death than wild-type Bax. Cancer Lett 2003;189:221–30.
[39] Dreyer GB, Metcalf BW, Tomaszek Jr TA, Carr TJ, Chandler 3rd AC, Hyland L, et al. Inhibition of human immunodeficiency virus 1 protease in vitro: rational design of substrate analogue inhibitors. Proc Natl Acad Sci U S A 1989;86:9752–6.
[40] Hoffmann D, Buchberger B, Nemetz C. In vitro synthesis of enzymatically active HIV-1 protease for rapid phenotypic resistance profiling. J Clin Virol 2005;32:294–9.

Structural Dynamics of HIV-1 Envelope Gp120 Outer Domain with V3 Loop

Masaru Yokoyama[1]*, Satoshi Naganawa[2], Kazuhisa Yoshimura[3], Shuzo Matsushita[3], Hironori Sato[1]*

1 Laboratory of Viral Genomics, Pathogen Genomics Center, National Institute of Infectious Diseases, 4-7-1 Gakuen, Musashi Murayama-shi, Tokyo, Japan, **2** Department of Microbiology and Cell Biology, Tokyo Metropolitan Institute of Medical Science, 2-1-6 Kamikitazawa, Setagaya-ku, Tokyo, Japan, **3** Division of Clinical Retrovirology and Infectious Diseases, Center for AIDS Research, Kumamoto University, 2-2-1 Honjo, Kumamoto, Japan

Abstract

Background: The net charge of the hypervariable V3 loop on the HIV-1 envelope gp120 outer domain plays a key role in modulating viral phenotype. However, the molecular mechanisms underlying the modulation remain poorly understood.

Methodology/Principal Findings: By combining computational and experimental approaches, we examined how V3 net charge could influence the phenotype of the gp120 interaction surface. Molecular dynamics simulations of the identical gp120 outer domain, carrying a V3 loop with net charge of +3 or +7, showed that the V3 change alone could induce global changes in fluctuation and conformation of the loops involved in binding to CD4, coreceptor and antibodies. A neutralization study using the V3 recombinant HIV-1 infectious clones showed that the virus carrying the gp120 with +3 V3, but not with +7 V3, was resistant to neutralization by anti-CD4 binding site monoclonal antibodies. An information entropy study shows that otherwise variable surface of the gp120 outer domain, such as V3 and a region around the CD4 binding loop, are less heterogeneous in the gp120 subpopulation with +3 V3.

Conclusions/Significance: These results suggest that the HIV-1 gp120 V3 loop acts as an electrostatic modulator that influences the global structure and diversity of the interaction surface of the gp120 outer domain. Our findings will provide a novel structural basis to understand how HIV-1 adjusts relative replication fitness by V3 mutations.

Citation: Yokoyama M, Naganawa S, Yoshimura K, Matsushita S, Sato H (2012) Structural Dynamics of HIV-1 Envelope Gp120 Outer Domain with V3 Loop. PLoS ONE 7(5): e37530. doi:10.1371/journal.pone.0037530

Editor: John J. Rossi, Beckman Research Institute of the City of Hope, United States of America

Received February 21, 2012; **Accepted** April 20, 2012; **Published** May 18, 2012

Funding: This work was supported by grants-in-aid from the Ministry of Health, Labor and Welfare, Japan. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: yokoyama@nih.go.jp (MY); hirosato@nih.go.jp (HS)

Introduction

The third variable (V3) element of the human immunodeficiency virus type 1 (HIV-1) envelope gp120 protein is usually composed of 35 amino acids. The element forms a protruding loop-like structure on the gp120 outer domain [1], is rich in basic amino acids, and has aromatic amino acids for the aromatic stacking interaction with proteins. The V3 loop participates in direct binding to the entry coreceptor [2] and constitutes the most critical determinant for the coreceptor use of HIV-1 [3,4,5,6]. In addition, the tip of V3 is highly immunogenic and contains neutralization epitopes for antibodies [7,8,9], although the epitopes can be inaccesible in the gp120 trimer on a virion of the HIV-1 primary isolates [10,11] or HIV-1 recombinants with less positively charged V3 [12,13]. Moreover, the V3 is reported to be the major determinant of HIV-1 sensitivity to neutralization by the soluble form of CD4 [14,15,16], a recombinant protein that binds to the cleft of the gp120 core [17]. Thus, the V3 loop plays a key role in modulating biological and immunological phenotypes of HIV-1. However, the molecular mechanisms underlying these modulations remain poorly understood.

It has been reported that the net charge of the V3 loop is tightly linked to the phenotype of HIV-1. The V3 loops of CCR5 tropic HIV-1s are usually less positively charged than those of CXCR4 tropic HIV-1s [18,19,20,21]. An increase in the V3 net charge can convert CCR5 tropic viruses into CXCR4 tropic viruses [4,22,23,24], and antibody resistant viruses into sensitive viruses [12,13]. Thus the V3 loop may be viewed as an electrostatic modulator of the structure of the gp120 interaction surface, an assumption that is largely unexamined.

Increasing evidence has indicated that the dynamics property of molecules in solution is critical for protein function and thus for many biological processes [25,26,27]. Molecular dynamic (MD) simulation is a powerful method that predicts the structural dynamics of biological molecules in solution, which is often difficult to analyze by experiments alone [28,29,30]. Recent advances in biomolecular simulation have rapidly improved the precision and application performance of this technique [28,29,30]. We have previously applied this technique to investigating the structural factors that regulate biological phenotype of viruses [13,31,32]. In this study, by combining MD simulations with antibody neutralization experiments and diversity analysis of the viral protein sequences, we studied a structural basis for the regulation of HIV-1 phenotype by V3 loop.

Results

Molecular dynamics simulation study

To address the potential role of the V3 net charge in modulating the structure and dynamics of the gp120 surface, we performed MD simulations of the identical gp120 outer domains carrying different V3 loops with net charges of +7 or +3 (Fig. 1A). The initial structures for the simulations were constructed by homology modeling using the crystal structure of HIV-1 gp120 containing an entire V3 loop as the template. Due to the perfect identity of the outer domain sequences of the V3 recombinant gp120s, the outer domain structures of the initial models for the MD simulations were identical before the simulations. The modeling targets in this study belong to HIV-1 subtype B and had a sequence similarity of about 87.3% to the modeling template. This similarity was high enough to construct high-accuracy models with an RMSD of ~1.5 Å for the main chain between the predicted and actual structures in the tested cases with homology models and x-ray crystal structures [33]. These initial models were lacking in V1/V2 loops and glycans on the gp120. The recombinant models are therefore suitable for exploring the potency of the structural regulation that is intrinsic to the V3 loop.

Using these models as the initial structures, we analyzed the structural dynamics of the gp120 outer domains in the absence of soluble CD4 by MD simulation. It was expected that the MD simulations would eliminate initial distortions in the template crystal structure, which could be generated during crystallization, and search for the most stable structures of unliganded gp120 outer domains at 1 atm at 310 K in water. The simulations showed that the same gp120 outer domains, carrying different V3 loops with net charges of +7 or +3, exhibited marked changes in conformations and fluctuations at several functional loops at 1 atm at 310 K in water (Figs. 1 and 2).

To quantitatively monitor the overall structural dynamics of the outer domain during MD simulation, the RMSDs between the initial model and models at given times of MD simulation were measured. The RMSD sharply increased soon after heating of the initial model and then gradually reached a near plateau after 10 ns of the MD simulations (Fig. 1B). The results suggested that most of the backbone heavy atoms of the outer domain reached a thermodynamic equilibrium after 10 ns of the simulation under the conditions employed. However, fluctuations of the RMSDs were still detectable even at around 30 ns of the simulations, suggesting that some regions of the outer domains continued to fluctuate.

To map the heavily fluctuating sites in the gp120 outer domain, we calculated the RMSF of the main chains of individual amino acids during the MD simulations. The RMSFs, which provide information about the atomic fluctuations during MD simulations [34], were found to be much greater in the amino acids constituting loops than those of the structured regions, such as helixes and β-sheets (Figs. 1C and 1D). These results are consistent with the general observations of proteins in solution, and indicate that the loops of the gp120 outer domain intrinsically possess structural flexibility in water. Notably, the RMSFs in some loops were markedly different between the two V3 recombinant gp120s. For example, the RMSF in the β20–β21 loop was much greater in the $Gp120_{LAI\text{-}TH09V3}$ (Fig. 1C). Conversely, those in the D loop were greater in the $Gp120_{LAI\text{-}NH1V3}$.

HIV-1 gp120 V3 loop often has a motif for the N-linked glycosylation that is usually preferentially conserved in R5 viruses (Fig. 1A). To address potential impacts of the glycan on the MD simulations, we performed MD simulation in the presence of a high mannose oligosaccharide in the V3 loop. We observed any significant differences in the structure and dynamics of gp120 outer domain in the presence or absence of the glycan (data not shown). This is reasonable because the glycosylation site is exposed toward an opposite direction from the gp120 core (Fig. 1D).

To clarify structural differences between the $Gp120_{LAI\text{-}NH1V3}$ and $Gp120_{LAI\text{-}TH09V3}$, we constructed their averaged structures using the 40,000 snapshots obtained from 10–30 ns of MD simulations using ptraj module in Amber 9. Superposition of the averaged structures showed that the relative configuration of the V3 loops and β20–β21 was markedly different between the two outer domains: the V3 tip protruded a greater distance from the β20–β21 loop in the $Gp120_{LAI\text{-}TH09V3}$ than in the $Gp120_{LAI\text{-}NH1V3}$ (Fig. 2A). The superposed structures also revealed differences in a region around the CD4 binding site (Fig. 2A, right panel with enlarged CD4 binding site). The relative configuration of the CD4 binding loop to the exit loop is critical for the gp120 binding to the CD4, a primary infection receptor of HIV-1 [17]. Therefore, we analyzed the distance between the CD4 binding and exit loops by measuring the distance ($D_{115\text{–}221}$) between the Cα of Gly115 and the Cα of Gly221 as an indicator (Fig. 2B). As expected from the fluctuations of the CD4 binding loop, the $D_{115\text{–}221}$ fluctuated during the MD simulations (Fig. 2C). However, the $D_{115\text{–}221}$ was significantly smaller in the $Gp120_{LAI\text{-}TH09V3}$ than in the $Gp120_{LAI\text{-}NH1V3}$ (Fig. 2D; $p<0.001$, Student's *t*-test): the $D_{115\text{–}221}$ ranged from 4–15 Å with an average of ~8 Å for the $Gp120_{LAI\text{-}TH09V3}$ and from 7–17 Å with an average of ~10 Å for the $Gp120_{LAI\text{-}NH1V3}$. These data suggest that the CD4 binding loop tended to be positioned more closely to the exit loop and thus tended to be sterically less exposed in the $Gp120_{LAI\text{-}TH09V3}$ than the $Gp120_{LAI\text{-}NH1V3}$.

Neutralization study

The above structural data raised the possibility that the reduction in the V3 net charge might reduce HIV-1 neutralization sensitivity by the anti-CD4 binding site antibodies. To address this possibility, we performed a neutralization assay using the two isogenic HIV-1 recombinant viruses, $HIV\text{-}1_{LAI\text{-}NH1V3}$ and $HIV\text{-}1_{LAI\text{-}TH09V3}$ [35], which carry the $Gp120_{LAI\text{-}NH1V3}$ and $Gp120_{LAI\text{-}TH09V3}$, respectively. These viruses were pre-incubated with various human MAbs against the CD4 binding site, and the reductions in viral infectious titers were measured using a HeLa-cell-based single-round viral infectivity assay system [36].

Table 1 summarizes the results of the neutralization assay. As expected, the two viruses exhibited markedly distinct neutralization sensitivities to the three human MAbs against the CD4 binding site. $HIV\text{-}1_{LAI\text{-}NH1V3}$ was consistently neutralized with all three MAbs against the CD4 binding site (49G2, 42F6, and 0.5δ), with ND_{50} values ranging between 0.224 and 0.934 μg/ml. In marked contrast, $HIV\text{-}1_{LAI\text{-}TH09V3}$ was highly resistant to neutralization by these MAbs, and 10 μg/ml of antibodies failed to block the viral infections. The two viruses were equally resistant to an anti-Gp120 antibody (4C11) that recognizes the Gp120 structure after CD4 binding. The result indicates that the CD4-induced gp120 epitope of the 4C11 are not preserved in the V3 recombinant viruses used in the present study. Conversely, they were equally sensitive to another ant-Gp120 antibody (4301 [37]) whose epitope is located outside of the CD4 binding site. A human MAb 8D11 used as a negative control had no effect on the viral infectivity in this assay.

Diversity study

Host immunity is a driving force behind the antigenic diversity of envelope proteins of the primate lentiviruses that establish persistent infection in hosts [23,38,39,40,41]. The above and

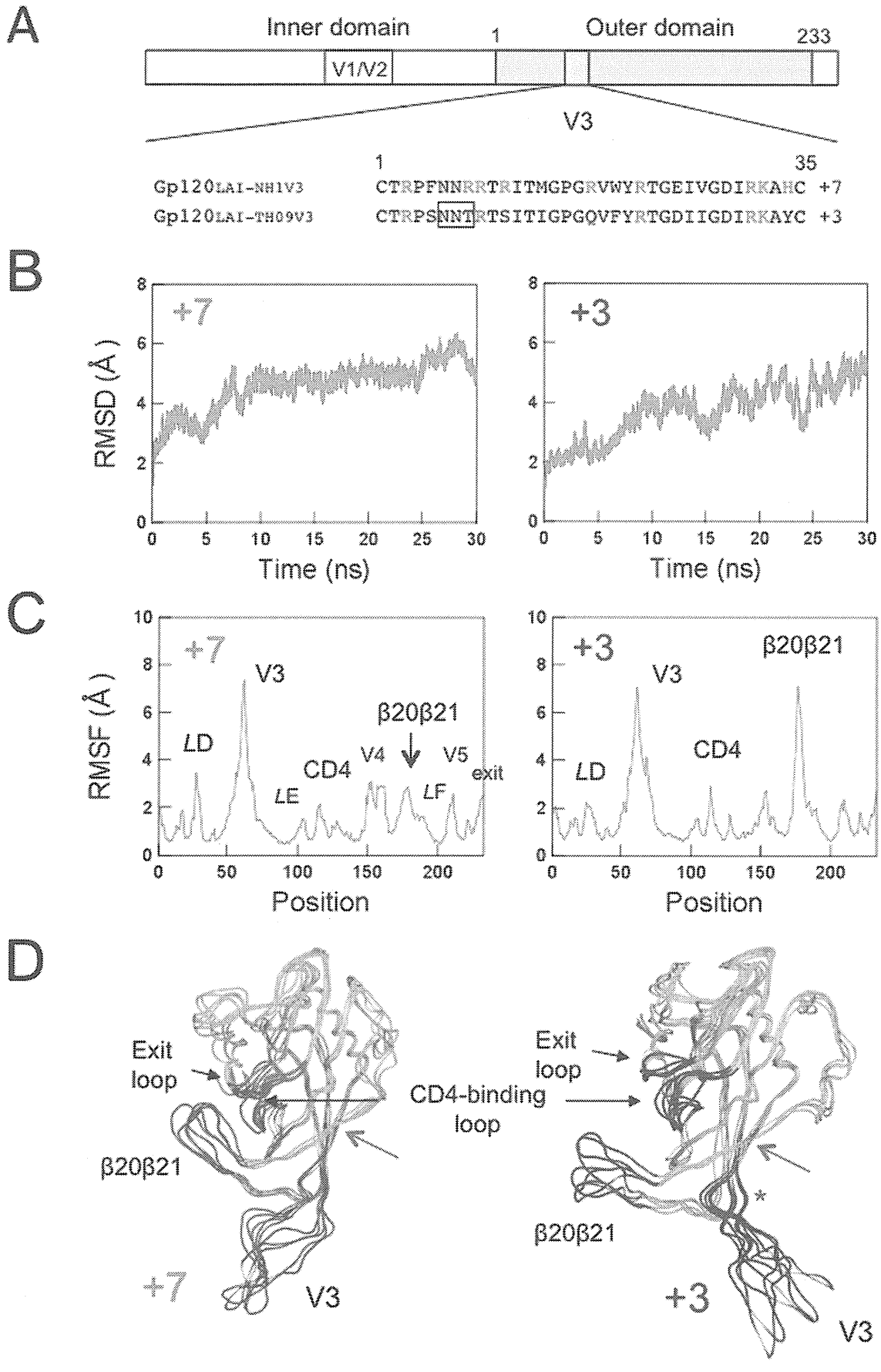

A
Inner domain
1
Outer domain
233
V1/V2
V3
1
35
Gp120LAI-NH1V3 CTRPFNNRRTRITMGPGRVWYRTGEIVGDIRKAHC +7
Gp120LAI-TH09V3 CTRPSNNTRTSITIGPGQVFYRTGDIIGDIRKAYC +3
B
RMSD (Å)
+7
+3
Time (ns)
C
RMSF (Å)
+7
+3
V3
β20β21
LD
CD4
V4
V5
exit
LE
LF
Position
D
Exit loop
CD4-binding loop
β20β21
+7
+3
V3
*

Figure 1. MD simulation of the identical gp120 outer domain carrying a V3 loop with net charge of +7 or +3. (A) Schematic representation of the gp120 open reading frame along with the amino acid sequences. The net charge indicates the number of positively charged amino acids (R, K, and H) minus the number of negatively charged amino acids (D and E) in the V3 loop. A light blue box indicates the outer domain used for the MD simulations. A pink box indicates the V3 loop. The numbers indicate amino acid positions at the outer domain (amino acids 1 to 233 in Figure 1A correspond to amino acids 256 to 489 in the gp120 of HIV-1_{LAI}) or the V3 loop. An open black box in the V3 loop sequence indicates a potential site for the N-linked glycosylation. (B–D) Left panels: $Gp120_{LAI\text{-}NH1V3}$; Right panels: $Gp120_{LAI\text{-}TH09V3}$. +7 and +3 indicate the net charges of V3 loops of the recombinant proteins. (B) Time course of RMSD during MD simulations. The RMSD values indicate the structural fluctuations of the outer domain in aqueous solution. The numbers in the horizontal axes indicate the time of MD simulation. (C) Distribution of RMSF in the gp120 outer domain. The RMSF values indicate the atomic fluctuations of the main chains of individual amino acids during 10–30 ns of MD simulations. The numbers in the horizontal axes indicate amino acid positions at the outer domain. (D) Superimposition of Gp120 models at 10, 15, 20, 25, and 30 ns of MD simulation. A green asterisk indicates approximate position of a potential N-glycosylation site at the V3 stem. A green arrow indicates the site of the disulfide bond at the V3 base.
doi:10.1371/journal.pone.0037530.g001

previous [12,13,14,15,16] neutralization studies raised the possibility that the gp120 surface might be less heterogeneous in gp120 subpopulations that have a less positively charged V3 loop, due to greater magnitudes of resistance to the antibody neutralization. To address this possibility, we performed an information entropy study using sequences in the public database. We extracted full-length gp120 amino acid sequences of HIV-1 subtype CRF01_AE that has the same evolutionary origin and is spread throughout southeast Asia [42]), and divided them into subgroups on the basis of the net charge of V3 loop (+2, +3, +4, +5, +6, +7, and +8). The sequences were used to calculate the Shannon entropy scores, *H(i)* [1], to denote the diversity of individual amino acids within each subpopulation.

Figure 3 shows the 3-D distribution of the *H(i)* scores of individual amino acids plotted on the HIV-1 gp120 crystal structure (PDB code: 2B4C [1]), where the green to orange regions were suggested to have more variable amino acids than the blue ones. In the gp120 subpopulation that has +7 V3 loop, the *H(i)* scores often exceeded 2.0 bits at many residues, reaching close to the maximum value of 4.4, i.e., the diversity was maximal, at the V5 region (Fig. 3A, left panel). Regions with high *H(i)* scores included the functional sites, such as the V3 loop and the regions around the CD4 binding site. In marked contrast, in the gp120 subpopulation carrying the +3 V3 loop, the *H(i)* scores were almost zero, i.e., the diversity was minimal, at many amino acids, but not at those in the V4, V5, and *L*E regions (Fig. 3A, right panel). Importantly, relatively high levels of conservation were also detected with amino acids in the otherwise highly variable V3 loop. Moreover, a region adjacent to the CD4 binding loop was also less heterogeneous compared with those of the gp120 subpopulation carrying +7 V3 loop (Figs. 3B). In the gp120 subpopulations carrying the +2, +3, 4, and +5 V3 loops, the *H(i)* scores were indistinguishable from each other: they were less heterogeneous than the subpopulations carrying the +6, +7, and +8 V3 loops. Similar results were obtained with HIV-1 subtype C that represents the most predominated HIV-1 in the world (data not shown).

Discussion

The ability of HIV-1 to rapidly change its phenotype greatly complicates our efforts to eradicate this virus. Elucidation of structural principles for the phenotypic change may provide a clue to control HIV-1. In this study, by combining MD simulations with antibody neutralization experiments and diversity analysis of the viral protein sequences, we studied a structural basis for the phenotypic change of HIV-1 by V3 mutations. To address this issue, we used a V3 recombinant system; we performed a computer-assisted structural study and an infection-based neutralization assay using gp120 proteins whose amino acid sequences are identical except for V3 loop. In combination with an informatics study, we obtained evidence that the HIV-1 V3 loop acts as an electrostatic modulator that influences the global structure and diversity of the interaction surface of the gp120 outer domain.

Using MD simulation, we first examined whether the V3 net charge could affect the structural dynamics of the HIV-1 gp120 outer domain surface. Initial structures of the outer domain of the two gp120s, $Gp120_{LAI\text{-}NH1V3}$ and $Gp120_{LAI\text{-}TH09V3}$, were identical before MD simulations, because the domains were both derived from HIV-1_{LAI} strain. Remarkably, however, the two molecules with distinct V3 loop exhibited markedly distinct structural dynamics following MD simulations (Figs. 1 and 2). These data strongly suggest that the V3 net charge can act as an intrinsic modulator that influences the structural dynamics of the interaction surface of the gp120 outer domain. Such a global effect on structure by a local electrostatic change has been reported with bacteriorhodopsin [43]. In general, the long-range effects of non-electrostatic contributions are negligible, whereas those of the electrostatic contributions are not [34]. Therefore, it is reasonable that the changes in overall charge of the V3 loop element caused the global effects on the gp120 structure via alteration of the electrostatic potentials of the gp120 surface.

We next studied biological impact of the structural changes predicted by MD simulations. The MD simulations suggested that the CD4 binding loop was less exposed in the $Gp120_{LAI\text{-}NH1V3}$ than the $Gp120_{LAI\text{-}NH1V3}$ (Fig. 2). The finding predicted that reduction in V3 net charge could cause reduction in neutralization sensitivity to the anti-CD4 binding site antibodies. This possibility was assessed by neutralization assay. We used infectious HIV-1_{LAI} clones having the $Gp120_{LAI\text{-}NH1V3}$ or the $Gp120_{LAI\text{-}TH09V3}$ to assess their neutralization sensitivities to the anti-CD4 binding site MAbs. Notably, we indeed observed marked reduction in the neutralization sensitivity in HIV-1_{LAI} having $Gp120_{LAI\text{-}TH09V3}$ (Table 1). The results are consistent with the structural changes predicted by MD simulations, as well as previous findings on neutralization sensitivity of HIV-1s to soluble CD4 [14,15,16].

We further studied evolutionary impact predicted by MD simulations and the neutralization studies. These studies predicted that reduction in V3 net charge could cause reduction in sequence diversity around the CD4 binding site due to reduced sensitivity to positive selection pressures of antibodies. Notably, we indeed observed marked reduction in the gp120 diversity: our Shannon entropy data show that otherwise variable surfaces of gp120, such as V3 and a region around the CD4 binding loop, are less heterogeneous in the gp120 subgroups carrying a V3 loop with a +3 charge (Fig. 3).

Previous cryo-electron microscopy studies have indicated that gp120 forms a trimer on an HIV-1 virion, where the CD4 binding sites are exposed on the outside surface in the solution [44,45,46]. Therefore, it is reasonable that gp120 with +3 V3 with less exposed CD4 binding loop is less sensitive to neutralization by anti-CD4 binding site antibodies (Table 1) and less heterogeneous around the CD4 binding site (Fig. 3). Collectively, our results

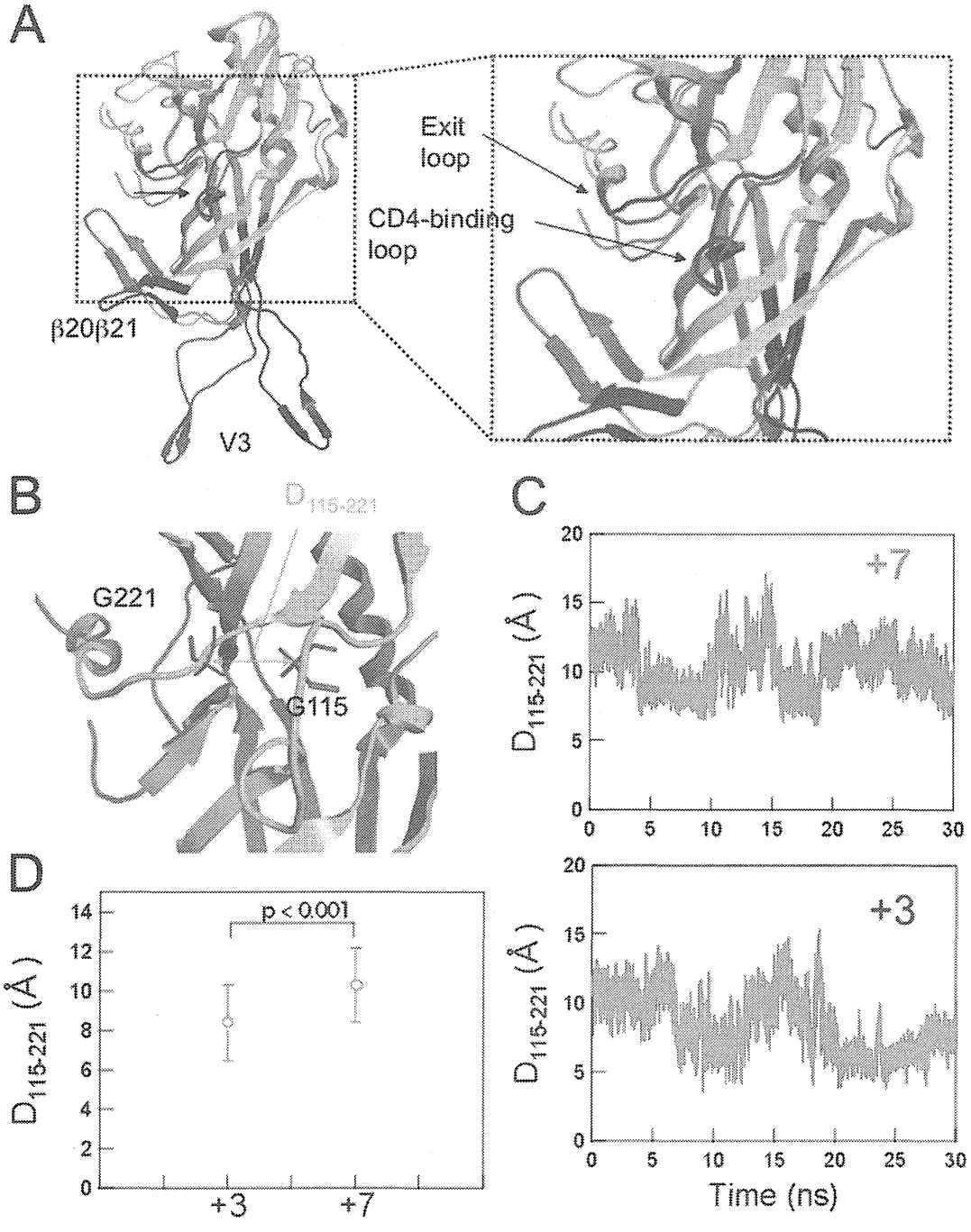


Figure 2. Comparison of the averaged 3-D models during MD simulation. (A) Superposition of the averaged structures obtained with the 40,000 snapshots obtained from 10–30 ns of MD simulations using ptraj module in Amber 9. Red and Blue ribbons: loops of $Gp120_{LAI\text{-}NH1V3}$ and $Gp120_{LAI\text{-}TH09V3}$ with V3 net charges of +7 and +3, respectively. (B–D) Configuration and structural dynamics of the CD4 binding loop. The distance between the Cα of Gly115 and the Cα of Gly221 in the CD4 binding loop was calculated to monitor configurational changes (B). The distance was monitored during the 10–30 ns of MD simulation (C) and the average distance with variance was plotted (D). +7: $Gp120_{LAI\text{-}NH1V3}$; +3: $Gp120_{LAI\text{-}TH09V3}$.
doi:10.1371/journal.pone.0037530.g002

obtained with all three approaches agree with each other and suggest that V3 net charge is an intrinsic factor that influences structural property, antibody sensitivity, and sequence diversity of CD4 binding site.

The HIV-1 gp120 outer domain has several functional or immunogenic loops involved in binding to CD4, coreceptor and antibodies. Our MD simulations predicted that V3 net charge influences fluctuation and conformation of these loops (Figs. 1 and 2). The V3-based structural modulation of the gp120 surface loops may be an effective mechanism to alter effectively the phenotype and relative fitness of HIV-1. For example, a change in the V3 net charge by mutations may induce changes in V3 conformation (Figs. 1D and 2A) [13], which in turn may influence intra- or inter-molecular interactions among gp120 monomers and thus global structure of gp120 trimer on a virion. Generation of a swarm of structural variants by V3 mutations could help generating the best-

Table 1. Neutralization sensitivity of the isogenic V3 recombinant HIV-1 to anti-gp120 monoclonal antibodies.

Antibody ID	Ig subtype	Epitopes on Gp120	ND_{50} (μg/ml)[@]	
			$HIV\text{-}1_{LAI\text{-}NH1V3}$	$HIV\text{-}1_{LAI\text{-}TH09V3}$
49G2	human IgG1	CD4 binding site[#]	0.224	>10
42F9	human IgG1	CD4 binding site[#]	0.934	>10
0.5δ [59]	human IgG1	CD4 binding site[#]	0.444	>10
4C11 [59]	human IgG2	CD4 induced structure[$]	>20	>10
4301	mouse IgG	broadly reactive*	0.59	0.57
8D11	human IgG1	none	>20	>10

[#]Neutralization epitope in the Gp120 outer domain before CD4 binding.
[$]Neutralization epitope induced in Gp120 after CD4 binding.
*Epitopes outside of the CD4 binding site [37].
[@]The effect of each antibody on viral infectivity was tested in duplicate.
doi:10.1371/journal.pone.0037530.t001

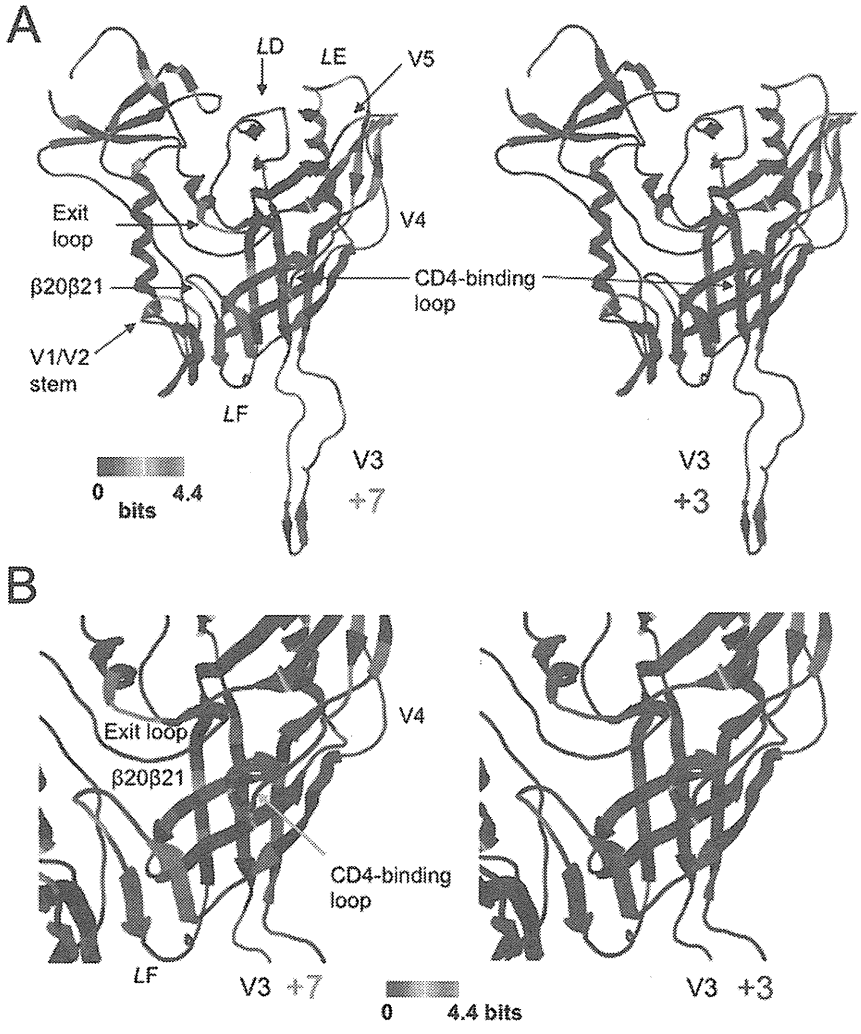


Figure 3. Diversity of the gp120 subpopulations carrying a V3 loop with net charge of +7 or +3. Full-length gp120 sequences of the HIV-1 CRF01_AE [42] were extracted from a public database, and divided into subgroups on the basis of the net charge of the V3 loop (+2~+8). The divided sequences were used to calculate the Shannon entropy, *H(i)* [63], within each subpopulation, and the *H(i)* values were plotted on the 3-D structure of gp120 (PDB code: 2B4C [1]). The results for the gp120 subgroups that have V3 loops with +7 (left panel) and +3 (right panel) charges are shown as representative. The numbers of sequences used to calculate the *H(i)* were 9 and 81 for +7 and +3, respectively. (A) Distribution of *H(i)* in the gp120 monomer. (B) Distribution of *H(i)* around the CD4 binding site.
doi:10.1371/journal.pone.0037530.g003

fit variants under changing environments during persistent infection of HIV-1 *in vivo*. Further study is necessary to address above issue.

Thus far fine structures of neither the intact gp120 monomer nor trimer are available. However, recent crystal structure study disclosed a structure of V1/V2 domain [47], which had been the major gp120 region lacking structural information. The V1/V2 domain is located on the outer surface of gp120, as is V3, and can participate in phenotypic changes of HIV-1 [48,49]. In this regard, Kwon et al [50] have found an intriguing role of gp120 variable loops; gp120 core has an intrinsic preference to form the CD4-bound conformation, whereas the variable loops, such as V1/V2 and V3 loops, play key roles in preventing conformational transitions into the CD4-bound state that is sensitive to neutralization. Thus it is conceivable that configurational changes of V3 loop by V3 mutations play roles in modulating structural dynamics of the unliganded gp120 core and neutralization sensitivity of HIV-1. Availability of the V1/V2 loop structure will promote structural study of the whole gp120 monomer containing V3 loop, V1/V2 domain, and glycans. Our findings will provide a structural basis to elucidate intra-molecular interactions of these elements, which in turn will allow the study of structure, function, and evolution of gp120 trimer. Incorporation of MD simulation into these studies will help understanding structural dynamics with which HIV-1 adjusts its relative replication fitness in nature.

Materials and Methods

Characteristics of the gp120 proteins and HIV-1s used

We used two isogenic recombinant gp120 proteins, termed $Gp120_{LAI\text{-}NH1V3}$ and $Gp120_{LAI\text{-}TH09V3}$ [35], for the present structural and neutralization studies. They differ only in their V3 loops. The $Gp120_{LAI\text{-}NH1V3}$ and $Gp120_{LAI\text{-}TH09V3}$ have the 35-amino-acid-length V3 loops from HIV-1-infected individuals in the gp120 backbone of the HIV-1_{LAI} strain [35]. The net charges of the V3 loops are +7 and +3 for the $Gp120_{LAI\text{-}NH1V3}$ and $Gp120_{LAI\text{-}TH09V3}$, respectively (the V3 net charge represents the number of positively charged amino acids (R, K, and H) minus the number of negatively charged amino acids (D and E) in the V3 loop). The HIV-1_{LAI} carrying the $Gp120_{LAI\text{-}NH1V3}$ (HIV-$1_{LAI\text{-}NH1V3}$) is the CXCR4 tropic virus, whereas that carrying the $Gp120_{LAI\text{-}TH09V3}$ (HIV-$1_{LAI\text{-}TH09V3}$) is the CCR5 monotropic virus [35]. The HIV-$1_{LAI\text{-}NH1V3}$ is sensitive to neutralization by antibodies with the ability to bind to the peptides containing the autologous V3 tip sequences, whereas HIV-$1_{LAI\text{-}TH09V3}$ is highly resistant to antibodies targeting the autologous V3 tip sequences [13].

MD simulation

As the initial structures for the MD simulation, we first constructed three-dimensional (3-D) models of the outer domains of the $Gp120_{LAI\text{-}NH1V3}$ and $Gp120_{LAI\text{-}TH09V3}$ by the comparative (homology) modeling method (reviewed in [33,51,52]), as described previously [13]. We used the crystal structure of HIV-1 gp120 containing an entire V3 region at a resolution of 3.30 Å (PDB code: 2B4C [1]) as the modeling template. The gp120 core is in complex with the CD4 receptor and the CD4 induced structure (CD4i) antibody X5 [1]: it represents the structure after the CD4 binding. We deleted the structures of the CD4 receptor and the X5 antibody from the 2B4C complex structure to construct the free gp120 outer domain models of HIV-1_{LAI} V3 recombinant viruses by homology modeling. Then the models were subjected to the MD simulations to analyze structure and dynamics of the gp120 outer domain in the absence of the CD4 receptor and the X5 antibody interactions. The homology modeling was performed using tools available in the Molecular Operating Environment (MOE) program (MOE 2008.10; Chemical Computing Group Inc., Montreal, Quebec, Canada). The 186 amino-terminal and 27 carboxyl-terminal residues were deleted to construct the gp120 outer domain structure. We optimized the 3-D structure thermodynamically via energy minimization using an MOE and an AMBER99 force field [53]. We further refined the physically unacceptable local structure of the models based on a Ramachandran plot evaluation using MOE. MD simulations were performed as described previously [13] using the Sander module in the Amber 9 program package [54] and the AMBER99SB force field [55] with the TIP3P water model [56]. Bond lengths involving hydrogen were constrained with the SHAKE algorithm [57], and the time step for all MD simulations was set to 2 fs. A nonbonded cutoff of 12 Å was used. After heating calculations for 20 ps to 310 K using the NVT ensemble, the simulations were executed using the NPT ensemble at 1 atm at 310 K for 30 ns. Superimpositions of the $Gp120_{LAI\text{-}NH1V3}$ and $Gp120_{LAI\text{-}TH09V3}$ structures were done by coordinating atoms of amino acids along the β-sheet at the gp120 outer domain. We performed two independent MD simulations with distinct MD codes and obtained similar results. Therefore, we present here the data set from one of the MD simulations as a representation.

Calculation of the root mean square deviation (RMSD) and root mean square fluctuation (RMSF)

The RMSD values between the heavy atoms of the two superposed proteins were used to measure the overall structural differences between the two proteins [34]. We also calculated the RMSF to provide information about the atomic fluctuations during MD simulations [34]. In this study, we calculated the RMSF of the main chains of individual amino acids using the 40,000 snapshots obtained from MD simulations of 10–30 ns. The average structures during the last 20 ns of MD simulations were used as reference structures for the calculation of the RMSF. Both the RMSD and RMSF were calculated using the ptraj module in Amber 9 [34].

Monoclonal antibodies (MAbs)

The 49G2, 42F9, 0.5δ and 4C11 antibodies used for the neutralization assay were the human MAbs established from an HIV-1-infected patient with long-term non-progressive illness. Human blood samples were collected after signed informed consent in accordance with study protocol and informed consent reviewed and approved by Ethics committee for clinical research & advanced medical technology at the Faculty of Life Science Kumamoto University. B cells from the patient's peripheral blood mononuclear cells were transformed by EBV, followed by cloning as described previously [58]. The culture supernatant from an individual clone was screened for the reactivity to $gp120_{SF2}$ by an enzyme-linked immunosorbent assay (ELISA). The specificity of antibodies was determined by gp120 capture ELISA and FACS analysis as described previously [59]. Briefly, reactivity of the mAbs against monomeric gp120 of HIV-1_{SF2} was measured with a gp120 capture assay in the absence or presence of soluble CD4 (0.5 μg/ml). Decrease in the binding activity was observed for the mAbs 0.5δ, 49G2, and 42F9 in the presence of soluble CD4, whereas enhancement in the reactivity was detected for the mAb 4C11. Reactivity of the mAbs against envelope protein on the cell surface was measured with a FACS analysis of PM1 cells chronically infected with JR-FL in the absence or presence of soluble CD4 (0.5 μg/ml). No significant difference was observed for the binding profiles of 0.5δ, 49G2, and 42F9 in the presence of

soluble CD4, whereas marked enhancement of binding was observed for the 4C11 in the presence of soluble CD4. Based on these binding data, we classified 49G2, 42F9, and 0.5δ as CD4 binding site Mabs, and 4C11 as a CD4-induced epitope. All MAbs used in this study were purified by affinity chromatography on Protein A Sepharose. A human MAb 8D11 was used as a negative control for the neutralization assay. Mouse MAb 4301 was purchased from Advanced BioScience Laboratories, Inc. (Kensington, MD). The 4301 was raised with a mixture of purified gp120 of HIV-1_{IIIB} and HIV-1_{MN} and broadly reactive with the gp120 of different HIV-1 isolates [37].

Neutralization assay

We used the two above-described V3 recombinant HIV-1s, HIV-$1_{LAI\text{-}NH1V3}$ and HIV-$1_{LAI\text{-}TH09V3}$ [35], for the neutralization study. The HIV-1 cell-free viruses were prepared by transfection of the plasmid DNAs into HeLa cells as described previously [24,35,60]. The neutralization activities of antibodies were measured in a single-round viral infectivity assay using $CD4^+CXCR4^+CCR5^+$ HeLa cells [36] as described previously [13]. Briefly, equal infectious titers of viruses (300 blue-cell-forming units) were incubated with serially diluted MAb preparations (0.03–10 μg/ml) for 1 hour at 37°C. The cells were infected with the virus-antibody mixture for 48 hours at 37°C, fixed, and stained with 5-bromo-4-chloro-3-indolyl-β-D-galacto-pyranoside. Each antibody dilution was tested in duplicate, and the means of the positive blue cell numbers were used to calculate the 50% inhibition dose of viral infectivity (ND_{50}).

Analysis of amino acid diversity

Amino acid diversity at individual sites of the HIV-1 gp120 sequences was analyzed with Shannon entropy scores as described previously [13,61,62]. Full-length gp120 amino acid sequences of the HIV-1 subtypes CRF01_AE and C were obtained from the HIV Sequence Database (http://www.hiv.lanl.gov/content/sequence/HIV/mainpage.html). The sequences were divided into subgroups based on the net charge of V3 loop (+2~+8) using a software system, InforSense 5.0.1 (InforSense Ltd. http://www.inforsense.com/); arginine (R), lysine (K), and histidine (H) were counted as +1, aspartic acid (D) and glutamic acid (E) as -1, and other amino acids as 0. The numbers of sequences used for the analysis of CRF01_AE were 11, 81, 57, 28, 18, 9, and 4 for +2, +3, +4, +5, +6, +7, and +8, respectively. The amino acid diversity within each V3 subpopulation of the same HIV-1 subtype was calculated using Shannon's formula [63]:

$$H(i) = -\sum_{x_i} p(x_i)\log_2 p(x_i) \quad \{x_i = \text{G, A, I, V, ...}\},$$

where $H(i)$, $p(x_i)$, and i indicate the amino acid entropy score of a given position, the probability of occurrence of a given amino acid at the position, and the number of the position, respectively. An $H(i)$ score of zero indicates absolute conservation, whereas 4.4 bits indicates complete randomness. The $H(i)$ scores were displayed on the 3-D structure of an HIV-1 gp120 (PDB code: 2B4C [1]).

Acknowledgments

We thank Shingo Kiyoura (SGI Japan, Ltd.), and Kaori Sawada and Takashi Ikegami (Ryoka Systems Inc.) for their support with the computational analysis. We thank Hirotaka Ode of the Pathogen Genomics Center for his helpful comments on the manuscript.

Author Contributions

Conceived and designed the experiments: MY SN HS. Performed the experiments: MY SN HS. Analyzed the data: MY SN HS. Contributed reagents/materials/analysis tools: KY SM. Wrote the paper: MY HS.

References

1. Huang CC, Tang M, Zhang MY, Majeed S, Montabana E, et al. (2005) Structure of a V3-containing HIV-1 gp120 core. Science 310: 1025–1028.
2. Huang CC, Lam SN, Acharya P, Tang M, Xiang SH, et al. (2007) Structures of the CCR5 N terminus and of a tyrosine-sulfated antibody with HIV-1 gp120 and CD4. Science 317: 1930–1934.
3. Choe H, Farzan M, Sun Y, Sullivan N, Rollins B, et al. (1996) The beta-chemokine receptors CCR3 and CCR5 facilitate infection by primary HIV-1 isolates. Cell 85: 1135–1148.
4. Speck RF, Wehrly K, Platt EJ, Atchison RE, Charo IF, et al. (1997) Selective employment of chemokine receptors as human immunodeficiency virus type 1 coreceptors determined by individual amino acids within the envelope V3 loop. J Virol 71: 7136–7139.
5. Xiao L, Owen SM, Goldman I, Lal AA, deJong JJ, et al. (1998) CCR5 coreceptor usage of non-syncytium-inducing primary HIV-1 is independent of phylogenetically distinct global HIV-1 isolates: delineation of consensus motif in the V3 domain that predicts CCR-5 usage. Virology 240: 83–92.
6. Cho MW, Lee MK, Carney MC, Berson JF, Doms RW, et al. (1998) Identification of determinants on a dualtropic human immunodeficiency virus type 1 envelope glycoprotein that confer usage of CXCR4. J Virol 72: 2509–2515.
7. Goudsmit J, Debouck C, Meloen RH, Smit L, Bakker M, et al. (1988) Human immunodeficiency virus type 1 neutralization epitope with conserved architecture elicits early type-specific antibodies in experimentally infected chimpanzees. Proc Natl Acad Sci U S A 85: 4478–4482.
8. Rusche JR, Javaherian K, McDanal C, Petro J, Lynn DL, et al. (1988) Antibodies that inhibit fusion of human immunodeficiency virus-infected cells bind a 24-amino acid sequence of the viral envelope, gp120. Proc Natl Acad Sci U S A 85: 3198–3202.
9. Javaherian K, Langlois AJ, McDanal C, Ross KL, Eckler LI, et al. (1989) Principal neutralizing domain of the human immunodeficiency virus type 1 envelope protein. Proc Natl Acad Sci U S A 86: 6768–6772.
10. Cavacini LA, Duval M, Robinson J, Posner MR (2002) Interactions of human antibodies, epitope exposure, antibody binding and neutralization of primary isolate HIV-1 virions. Aids 16: 2409–2417.
11. Lusso P, Earl PL, Sironi F, Santoro F, Ripamonti C, et al. (2005) Cryptic nature of a conserved, CD4-inducible V3 loop neutralization epitope in the native envelope glycoprotein oligomer of CCR5-restricted, but not CXCR4-using, primary human immunodeficiency virus type 1 strains. J Virol 79: 6957–6968.
12. Bou-Habib DC, Roderiquez G, Oravecz T, Berman PW, Lusso P, et al. (1994) Cryptic nature of envelope V3 region epitopes protects primary monocytotropic human immunodeficiency virus type 1 from antibody neutralization. J Virol 68: 6006–6013.
13. Naganawa S, Yokoyama M, Shiino T, Suzuki T, Ishigatsubo Y, et al. (2008) Net positive charge of HIV-1 CRF01_AE V3 sequence regulates viral sensitivity to humoral immunity. PLoS One 3: e3206.
14. Hwang SS, Boyle TJ, Lyerly HK, Cullen BR (1992) Identification of envelope V3 loop as the major determinant of CD4 neutralization sensitivity of HIV-1. Science 257: 535–537.
15. Willey RL, Theodore TS, Martin MA (1994) Amino acid substitutions in the human immunodeficiency virus type 1 gp120 V3 loop that change viral tropism also alter physical and functional properties of the virion envelope. J Virol 68: 4409–4419.
16. Willey RL, Martin MA, Peden KW (1994) Increase in soluble CD4 binding to and CD4-induced dissociation of gp120 from virions correlates with infectivity of human immunodeficiency virus type 1. J Virol 68: 1029–1039.
17. Kwong PD, Wyatt R, Robinson J, Sweet RW, Sodroski J, et al. (1998) Structure of an HIV gp120 envelope glycoprotein in complex with the CD4 receptor and a neutralizing human antibody. Nature 393: 648–659.
18. Fouchier RA, Groenink M, Kootstra NA, Tersmette M, Huisman HG, et al. (1992) Phenotype-associated sequence variation in the third variable domain of the human immunodeficiency virus type 1 gp120 molecule. J Virol 66: 3183–3187.
19. Chesebro B, Wehrly K, Nishio J, Perryman S (1992) Macrophage-tropic human immunodeficiency virus isolates from different patients exhibit unusual V3 envelope sequence homogeneity in comparison with T-cell-tropic isolates: definition of critical amino acids involved in cell tropism. J Virol 66: 6547–6554.
20. Milich L, Margolin B, Swanstrom R (1993) V3 loop of the human immunodeficiency virus type 1 Env protein: interpreting sequence variability. J Virol 67: 5623–5634.
21. Milich L, Margolin BH, Swanstrom R (1997) Patterns of amino acid variability in NSI-like and SI-like V3 sequences and a linked change in the CD4-binding domain of the HIV-1 Env protein. Virology 239: 108–118.

22. de Jong J-J, de Ronde A, Keulen W, Tersmette M, Goudsmit J (1992) Minimal requirements for the human immunodeficiency virus type 1 V3 domain to support the syncytium-inducing phenotype: analysis by single amino acid substitution. J Virol 66: 6777–6780.
23. Shioda T, Oka S, Ida S, Nokihara K, Toriyoshi H, et al. (1994) A naturally occurring single basic amino acid substitution in the V3 region of the human immunodeficiency virus type 1 env protein alters the cellular host range and antigenic structure of the virus. J Virol 68: 7689–7696.
24. Kato K, Sato H, Takebe Y (1999) Role of naturally occurring basic amino acid substitutions in the human immunodeficiency virus type 1 subtype E envelope V3 loop on viral coreceptor usage and cell tropism. J Virol 73: 5520–5526.
25. Thorpe IF, Brooks CL, 3rd (2007) Molecular evolution of affinity and flexibility in the immune system. Proc Natl Acad Sci U S A 104: 8821–8826.
26. Lu HP, Xun L, Xie XS (1998) Single-molecule enzymatic dynamics. Science 282: 1877–1882.
27. Astumian RD (1997) Thermodynamics and kinetics of a Brownian motor. Science 276: 917–922.
28. Garcia-Viloca M, Gao J, Karplus M, Truhlar DG (2004) How enzymes work: analysis by modern rate theory and computer simulations. Science 303: 186–195.
29. Karplus M, Kuriyan J (2005) Molecular dynamics and protein function. Proc Natl Acad Sci U S A 102: 6679–6685.
30. Dodson GG, Lane DP, Verma CS (2008) Molecular simulations of protein dynamics: new windows on mechanisms in biology. EMBO Rep 9: 144–150.
31. Miyamoto T, Yokoyama M, Kono K, Shioda T, Sato H, et al. (2011) A single amino acid of human immunodeficiency virus type 2 capsid protein affects conformation of two external loops and viral sensitivity to TRIM5alpha. PLoS One 6: e22779.
32. Ode H, Yokoyama M, Kanda T, Sato H (2011) Identification of folding preferences of cleavage junctions of HIV-1 precursor proteins for regulation of cleavability. J Mol Model 17: 391–399.
33. Baker D, Sali A (2001) Protein structure prediction and structural genomics. Science 294: 93–96.
34. Case DA, Cheatham TE, 3rd, Darden T, Gohlke H, Luo R, et al. (2005) The Amber biomolecular simulation programs. J Comput Chem 26: 1668–1688.
35. Sato H, Kato K, Takebe Y (1999) Functional complementation of the envelope hypervariable V3 loop of human immunodeficiency virus type 1 subtype B by the subtype E V3 loop. Virology 257: 491–501.
36. Hachiya A, Aizawa-Matsuoka S, Tanaka M, Takahashi Y, Ida S, et al. (2001) Rapid and simple phenotypic assay for drug susceptibility of human immunodeficiency virus type 1 using CCR5-expressing HeLa/CD4(+) cell clone 1–10 (MAGIC-5). Antimicrob Agents Chemother 45: 495–501.
37. di Marzo Veronese F, Rahman R, Pal R, Boyer C, Romano J, et al. (1992) Delineation of immunoreactive, conserved regions in the external glycoprotein of the human immunodeficiency virus type 1. AIDS Res Hum Retroviruses 8: 1125–1132.
38. Simmonds P, Balfe P, Ludlam CA, Bishop JO, Brown AJ (1990) Analysis of sequence diversity in hypervariable regions of the external glycoprotein of human immunodeficiency virus type 1. J Virol 64: 5840–5850.
39. Burns DP, Desrosiers RC (1994) Envelope sequence variation, neutralizing antibodies, and primate lentivirus persistence. Curr Top Microbiol Immunol 188: 185–219.
40. Bonhoeffer S, Holmes SE, Nowak M (1995) Causes of HIV diversity. Nature 376: 125.
41. Lukashov VV, Kuiken CL, Goudsmit J (1995) Intrahost human immunodeficiency virus type 1 evolution is related to length of the immunocompetent period. J Virol 69: 6911–6916.
42. Buonaguro L, Tornesello ML, Buonaguro FM (2007) Human immunodeficiency virus type 1 subtype distribution in the worldwide epidemic: pathogenetic and therapeutic implications. J Virol 81: 10209–10219.
43. Brown LS, Kamikubo H, Zimanyi L, Kataoka M, Tokunaga F, et al. (1997) A local electrostatic change is the cause of the large-scale protein conformation shift in bacteriorhodopsin. Proc Natl Acad Sci U S A 94: 5040–5044.
44. Wu SR, Loving R, Lindqvist B, Hebert H, Koeck PJ, et al. (2010) Single-particle cryoelectron microscopy analysis reveals the HIV-1 spike as a tripod structure. Proc Natl Acad Sci U S A 107: 18844–18849.
45. White TA, Bartesaghi A, Borgnia MJ, Meyerson JR, de la Cruz MJ, et al. (2010) Molecular architectures of trimeric SIV and HIV-1 envelope glycoproteins on intact viruses: strain-dependent variation in quaternary structure. PLoS Pathog 6: e1001249.
46. Hu G, Liu J, Taylor KA, Roux KH (2011) Structural comparison of HIV-1 envelope spikes with and without the V1/V2 loop. J Virol 85: 2741–2750.
47. McLellan JS, Pancera M, Carrico C, Gorman J, Julien JP, et al. (2012) Structure of HIV-1 gp120 V1/V2 domain with broadly neutralizing antibody PG9. Nature 480: 336–343.
48. Shibata J, Yoshimura K, Honda A, Koito A, Murakami T, et al. (2007) Impact of V2 mutations on escape from a potent neutralizing anti-V3 monoclonal antibody during in vitro selection of a primary human immunodeficiency virus type 1 isolate. J Virol 81: 3757–3768.
49. Ogert RA, Lee MK, Ross W, Buckler-White A, Martin MA, et al. (2001) N-linked glycosylation sites adjacent to and within the V1/V2 and the V3 loops of dualtropic human immunodeficiency virus type 1 isolate DH12 gp120 affect coreceptor usage and cellular tropism. J Virol 75: 5998–6006.
50. Kwon YD, Finzi A, Wu X, Dogo-Isonagie C, Lee LK, et al. (2012) Unliganded HIV-1 gp120 core structures assume the CD4-bound conformation with regulation by quaternary interactions and variable loops. Proc Natl Acad Sci U S A 109: 5663–5668.
51. Sanchez R, Pieper U, Melo F, Eswar N, Marti-Renom MA, et al. (2000) Protein structure modeling for structural genomics. Nat Struct Biol 7 Suppl: 986–990.
52. Marti-Renom MA, Stuart AC, Fiser A, Sanchez R, Melo F, et al. (2000) Comparative protein structure modeling of genes and genomes. Annu Rev Biophys Biomol Struct 29: 291–325.
53. Ponder JW, Case DA (2003) Force fields for protein simulations. Adv Protein Chem 66: 27–85.
54. Case DA, Darden TA, Cheatham TE, Simmerling CL, Wang J, et al. (2006) AMBER 9. University of California: San Francisco.
55. Hornak V, Abel R, Okur A, Strockbine B, Roitberg A, et al. (2006) Comparison of multiple Amber force fields and development of improved protein backbone parameters. Proteins 65: 712–725.
56. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML (1983) Comparison of simple potential functions for simulating liquid water. J Chem Phys 79: 926–935.
57. Ryckaert JP, Ciccotti G, Berendsen HJC (1977) Numerical integration of the Cartesian equations of motion of a system with constraints: Molecular dynamics of n-alkanes. J Comput Phys 23: 327–341.
58. Matsushita S, Robert-Guroff M, Rusche J, Koito A, Hattori T, et al. (1988) Characterization of a human immunodeficiency virus neutralizing monoclonal antibody and mapping of the neutralizing epitope. J Virol 62: 2107–2114.
59. Yoshimura K, Harada S, Shibata J, Hatada M, Yamada Y, et al. (2010) Enhanced Exposure of Human Immunodeficiency Virus Type 1 Primary Isolate Neutralization Epitopes through Binding of CD4 Mimetic Compounds. J Virol.
60. Shiino T, Kato K, Kodaka N, Miyakuni T, Takebe Y, et al. (2000) A group of V3 sequences from human immunodeficiency virus type 1 subtype E non-syncytium-inducing, CCR5-using variants are resistant to positive selection pressure. J Virol 74: 1069–1078.
61. Motomura K, Oka T, Yokoyama M, Nakamura H, Mori H, et al. (2008) Identification of monomorphic and divergent haplotypes in the 2006–2007 norovirus GII/4 epidemic population by genomewide tracing of evolutionary history. J Virol 82: 11247–11262.
62. Oka T, Yokoyama M, Katayama K, Tsunemitsu H, Yamamoto M, et al. (2009) Structural and biological constraints on diversity of regions immediately upstream of cleavage sites in calicivirus precursor proteins. Virology 394: 119–129.
63. Shannon CE (1997) The mathematical theory of communication. 1963. MD Comput 14: 306–317.

Development of a Novel *In Silico* Docking Simulation Model for the Fine HIV-1 Cytotoxic T Lymphocyte Epitope Mapping

Masahiko Mori[1,2], Kei Matsuki[1], Tomoyuki Maekawa[1], Mari Tanaka[1], Busarawan Sriwanthana[3], Masaru Yokoyama[4], Koya Ariyoshi[1]*

1 Institute of Tropical Medicine, Nagasaki University, Nagasaki City, Nagasaki, Japan, **2** Department of Paediatrics, The Peter Medawar Building for Pathogen Research, University of Oxford, Oxford, United Kingdom, **3** Department of Medical Sciences, Ministry of Public Health, Nonthaburi, Thailand, **4** Pathogen Genomics Center, National Institute of Infectious Diseases, Shinjuku-ku, Tokyo, Japan

Abstract

Introduction: Class I HLA's polymorphism has hampered CTL epitope mapping with laborious experiments. Objectives are 1) to evaluate the novel *in silico* model in predicting previously reported epitopes in comparison with existing program, and 2) to apply the model to predict optimal epitopes with HLA using experimental results.

Materials and Methods: We have developed a novel *in silico* epitope prediction method, based on HLA crystal structure and a peptide docking simulation model, calculating the peptide-HLA binding affinity at four amino acid residues in each terminal. It was applied to predict 52 HIV best-defined CTL epitopes from 15-mer overlapping peptides, and its predictive ability was compared with the HLA binding motif-based program of HLArestrictor. It was then used to predict HIV-1 Gag optimal epitopes from previous ELISpot results.

Results: 43/52 (82.7%) epitopes were detected by the novel model, whereas 37 (71.2%) by HLArestrictor. We also found a significant reduction in epitope detection rates for longer epitopes in HLArestrictor ($p = 0.027$), but not in the novel model. Improved epitope prediction was also found by introducing both models, especially in specificity ($p<0.001$). Eight peptides were predicted as novel, immunodominant epitopes in both models.

Discussion: This novel model can predict optimal CTL epitopes, which were not detected by an existing program. This model is potentially useful not only for narrowing down optimal epitopes, but predicting rare HLA alleles with less information. By introducing different principal models, epitope prediction will be more precise.

Citation: Mori M, Matsuki K, Maekawa T, Tanaka M, Sriwanthana B, et al. (2012) Development of a Novel *In Silico* Docking Simulation Model for the Fine HIV-1 Cytotoxic T Lymphocyte Epitope Mapping. PLoS ONE 7(7): e41703. doi:10.1371/journal.pone.0041703

Editor: Jianming Tang, University of Alabama at Birmingham, United States of America

Received April 24, 2012; **Accepted** June 25, 2012; **Published** July 27, 2012

Funding: This study was financially supported by Japan foundation for AIDS prevention (JFAP), and the Ministry of Health, Labour and Welfare of Japan and the Grand-in-Aid for Scientific Research from the Japan Society for the Promotion of Science. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: kari@nagasaki-u.ac.jp

Introduction

Cytotoxic T lymphocytes (CTLs) play a crucial role in HIV replication control by eliminating virus-infected cells by recognizing class I Human Leukocyte Antigen (HLA) molecule-viral peptides (= epitope) complex. This response is thought to be a major determinant of the viral set point, and consequent disease progression [1]. However the efficacy of the CTL response is affected by the extent of polymorphisms in HLA loci and viral sequences. The HLA region is found on chromosome 6 and is the most polymorphic loci in the human genome [2]; each individual expresses up to six different class I alleles out of a vast pool of allelic variants, the reported number of which reaches 5,399 for class I HLA molecules (1,757 of HLA-A, 2,338 of HLA-B, and 1,304 of HLA-C alleles) [3]. In addition, the extensive diversity of HIV-1 owing to its extreme capacity to mutate has led to a reported 13 prototype clades and 43 circulating recombinant forms (CRFs) [4]. Despite such HLA polymorphism and HIV viral diversity environment, recent genome wide association study (GWAS) reported the best contribution of class I HLA for viral control, suggesting the importance of CTL epitope mapping with responsible HLA information [5]. Several major HIV-1 epitopes and their restricting HLA alleles have been defined through fine epitope mapping; 1,344 epitopes and their restricting HLA alleles have been reported as of February 2012 (CTL Epitopes. Los Alamos National Lab. http://www.hiv.lanl.gov/). The limitation of the dataset currently available however, is that the majority of these epitope/HLA combinations are derived from subtype B-infected Caucasians or C-infected Africans, and epitope information from other subtypes or ethnicities is rare.

The traditional, *in vitro* method of epitope detection involves using a matrix of overlapping peptides (OLPs) encoding viral proteins in Enzyme-Linked Immunospot (ELISpot) assays to

identify a single candidate peptide, from which the 8-11mer epitope is mapped down. This is typically followed by the confirmation of the restricting HLA alleles using tetramers or in a ^{51}Cr release assay using peptide-specific lines [6,7]. It is a difficult and labor-intensive process, particularly time-consuming in the case of epitopes restricted by rare HLA alleles because of the limited number of samples available.

Recently, alternative, *in silico* models for epitope prediction have been developed [8]. These can broadly be divided into two models; the first is an algorithm based on the peptide-binding motif, and the second is a structural algorithm model based on the crystal structure of HLA molecules. The former is characterized by the use of motif matrices deduced from refined motifs based on the pool sequence, enlisting optimal amino acid sequences at anchor positions in specific HLA alleles. An example of such an algorithm is the SYFPEITHI [9] database, which predicts the HLA-binding affinities of peptides by ranking them according to the presence of primary and secondary anchor amino acids. However these models are based on reported epitopes and their restricting HLA alleles, so their predictions are powerful in the context of well-published HLA alleles but not suitable against rare or novel alleles with little previous information. Another model of epitope prediction is the binding affinity model, which calculates the peptides' binding affinity and scores it using quantitative matrices (QMs), a well-known example being the NetMHC [10,11] or the HLArestrictor [12]. This model scores binding strength as binding affinity with thresholds to differentiate strong binding peptides and weak ones in each calculation.

On the other hand, the structural algorithm model does not require binding motif information, which is advantageous for the definition of epitopes restricted by HLA alleles with less published epitope information. Recently, a docking simulation model (DSM) which takes into consideration binding energy such as electrostatic interactions and van der Waals (vdw) interactions, together with the crystal structure of HLA alleles, has been developed [13–17].

Our objectives here are 1) to evaluate the novel *in silico* DSM in predicting previously reported best-defined epitopes in comparison with existing binding motif-based program, and 2) to apply the model to predict optimal size of the epitopes and restricting HLA alleles using results obtained from our previous study in a HIV-1 CRF01_AE-infected Thai cohort.

Materials and Methods

Ethic Statement

This study was approved by Thai Ministry of Public Health Ethics Committee. Written informed consent was obtained from all patients after explaining the purpose and expected consequences of the study.

Computational program and calculation

We used the commercial softwares Molecular Operating Environment® (MOE) (CCG Inc., Montreal, Canada) and MOE-ASEDock® (Ryoka System Inc., Tokyo, Japan) for the molecular binding affinity calculation [18]. HLA's 3D models were obtained from the X-ray crystallography database in MOE's library (1OGA for HLA-A*02:01, IQ94 for HLA-A*11:01, 2BCK for HLA-A*24:02, 1XR9 for HLA-B*15:01, 1JGE for HLA-B*27:05, 2CIK for HLA-B*35:01, 1E27 for HLA-B*52:01, 2RFX for HLA-B*57:01, and 1EFX for HLA-C*03:04). In cases where the original X-ray crystallography information was unavailable, we generated a 3D structural model using highly homologous HLA alleles as template, using rotamer explorer or homology modeling to reconstruct their structures by changing sequential difference sites, a method originally used in the point mutation program attached in MOE AMBER99 [19] for force field, calculations. For solvent effect energy calculation, a generalized Born model [20], were introduced. As an indicator of the affinity between epitope candidate peptides and the class I HLA allele, we measured the U_dock score [U_ele (electric energy)+U_vdw (van der Waals energy)+U_solv (Solvation energy)+U_strain (Strain energy)] (kcal/mol) [18]. We calculated the U_dock score of four residues at each N- and C-terminal, spanning the anchor position at each of the terminals, and scored the sum of them as binding affinity. A lower score indicates a higher affinity between the HLA molecule and peptides.

Evaluation of the novel DSM through an analysis of best-defined HIV CTL epitopes and their restricting HLA alleles

For the quality evaluation of this novel program, we first calculated the U_dock score for 52 best-defined HIV epitopes restricted by the alleles HLA-A*02:01, HLA-A*11:01, HLA-A*24:02, HLA-B*15:01, HLA-B*27:05, HLA-B*35:01 and HLA-B*57:01 as enlisted in Los Alamos database (CTL Epitopes. Los Alamos National Lab. http://www.hiv.lanl.gov/). We calculated the U_dock score between the restricting HLA alleles and the 8 to 11-mer peptides within 15-mer peptides of the viral strain HXB2, in which best-defined epitopes were included. 26 variants of 8 to 11-mer peptides were calculated in one HLA and 15-mer peptide combination, then the lowest U_dock score was ranked as the 1st and the highest score as the 26th in each calculation (Figure 1). Combinations that ranked within the top five were regarded as positive. In parallel with our DSM, we also performed epitope prediction using the latest artificial neural network (ANN) model, the HLArestrictor [12], using the affinity thresholds of Strong Binder (SB), Weak Binder (WB), Combined Binder (CB) and Non-binder (NB), according to their definitions.

We evaluated the sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV) for each best-defined epitope prediction using the DSM, HLArestrictor, as well as those defined as dual positive by both models.

Analysis of *in vitro* HIV-1 CRF01_AE Gag epitope candidates by using both *in silico* epitope prediction models

We then applied both the DSM and the HLArestrictor to predict the optimal size of epitopes, based on results obtained from our previous study [21], in which 31 candidate epitopes were detected by ELISpot assays using Gag 15-mer OLPs and their HLA associations detected by Fisher's exact test in a cohort of 137 (107 female and 30 male) HIV-1 CRF01_AE-infected Thais. All were chronically infected and treatment naïve, with median 461/ul CD4+T cell count (range 204–1,191) and 4.2 log copies/ml viral load (2.6–5.9).

Epitope prediction for the immunogenic Gag OLP $p24_{276–285}$ MYSPVSILDI using a ^{51}Cr release assay and both *in silico* models

In our previous study [21], the 15-mer peptides Gag $p24_{271–285}$ NKIVRMYSPVSILDI (NI15) and $p24_{276–290}$ MYSPVSILDIRQGPK (MK15) induced the largest responses in terms of both breadth and magnitude, and were statistically associated with the alleles HLA-A*02:07, HLA-B*46:01, and HLA-C*01:02, which were under linkage disequilibrium (LD) association [21]. Presuming that the optimal epitope resides in the overlapping amino acid sequence between NI15 and MK15, that is, $p24_{276–285}$

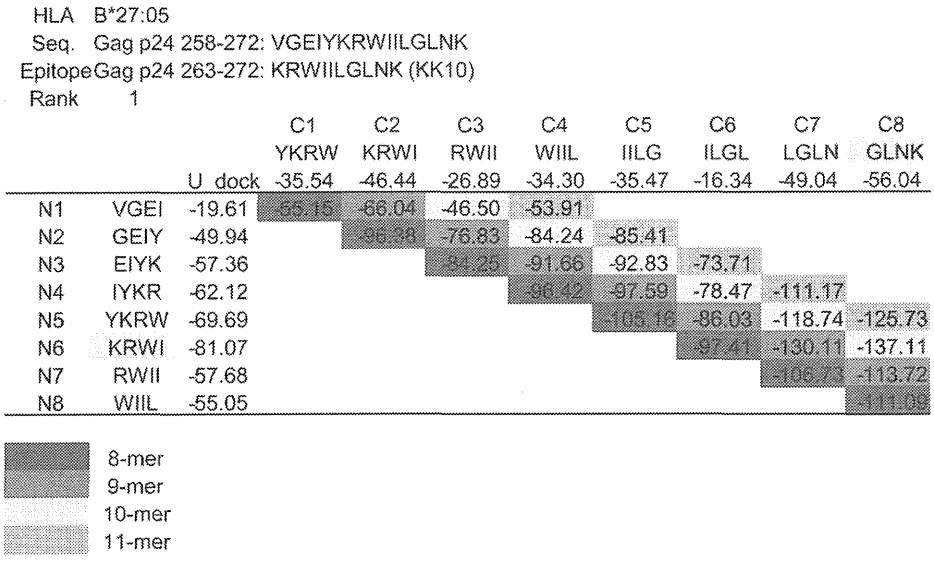


Figure 1. Example of epitope prediction using the novel *in silico* docking simulation model. U_dock scores of the N-terminal (Row N1–N8) and C-terminal (Column C1–C8) was calculated and their sum was scored as the U_dock score (kcal/mol) of each 8 to 11-mer peptide's. The lower score indicated stronger binding between the peptide and HLA. In this example, Gag $p24_{263-272}$ KRWIILGLNK (KK10), well-known as one of the best-defined epitopes, scored −137.11 kcal/mol against HLA-B*27:05 and was the lowest (ranked as the 1st) among 26 variants in 15-mer peptide of Gag $p24_{258-272}$ VGEIYKRWIILGLNK.
doi:10.1371/journal.pone.0041703.g001

MYSPVSILDI (MI10), we conducted a ^{51}Cr release assay as previously described [22].

Results

Prediction of best-defined epitopes by the DSM and the peptide binding motif model

We have evaluated the predictive power of our DSM by testing its ability to predict epitopes within 52 15-mer peptides spanning the epitopes for seven HLA alleles enlisted in the Los Alamos database as best-defined epitopes. Overall, DSM ranked 43/52 (82.7%) of the best-defined epitopes correctly within the top five candidates, within which 14 epitopes ranked as the 1st, 11 as the 2nd, 7 as the 3rd, 3 as the 4th, then 8 as the 5th (Table S1). This was comparable to the HLArestrictor, where 37/52 (71.2%, 43/52 vs 37/52, $p = 0.24$ by Fisher's exact test) best-defined epitopes scored within the threshold of binding affinity without having 4 or more other candidate epitopes: 20 as SB, 10 as WB and 7 as CB. Table 1 summarizes the performance on epitope prediction by each model and dual positives by both models, according to their sensitivity, specificity, PPV and NPV. The performance of the DSM is similar to that of HLArestrictor. Interestingly, by introducing both models, specificity increased with significance ($p<0.001$), and an additive effect was seen in the PPV. We believe this is the first study to report a structure-based epitope prediction model with comparable or greater predictive power than a peptide-binding motif based model.

32/52 (61.5%) epitopes were detected as a significant epitope candidate by both models. 11/52 (21.2%) epitopes were detected only by the DSM, while 5/52 (9.6%) were detected only by HLArestrictor. 4/52 (7.7%) epitopes were not detected by either methods. Within the 14 epitopes not correctly predicted by HLArestrictor, incorrect epitopes were predicted in 7 epitopes. It is noteworthy that two epitopes, Nef_{75-82} PLRPMTYK (PK8) restricted by HLA-A*11:01 and $Nef_{117-127}$ TQGYFPDWQNY (TY11) restricted by HLA-B*15:01 were detected as a NB by HLArestrictor, whereas they were ranked as the 2nd in PK8 and the 1st in TY11 in the DSM. $Integrase_{179-188}$ AVFIHNFKRK (AK10) restricted by HLA-A*11:01 was predicted as a SB, but because there were 5 other SB candidates, 3 WB candidates and 1 CB candidate, this prediction was regarded as failure.

A striking feature of the DSM was that it had a high detection rate of best-defined epitopes independent of the peptide's length. The prediction rate of shorter epitopes (8 and 9-mer) was 27/31 (87.1%) while the rate for longer epitopes (10 and 11-mer) was 16/21 (76.2%), between which we found no significant difference by Fisher's exact test ($p = 0.46$). In contrast, the ability of HLArestrictor to accurately predict best-defined epitopes was highly dependent on epitope length, as the prediction rate of longer

Table 1. *Evaluation of best-defined epitope prediction among docking simulation model, HLArestrictor, and positives in dual models.*

	DSM	HLArestrictor	Dual positives	p (mxn Fisher's exact test)
Sensitivity	0.83	0.71	0.62	0.056
Specificity	0.83	0.94	0.97	<0.001
PPV	0.17	0.31	0.43	0.095
NPV	0.99	0.99	0.98	0.46

Evaluation of best-defined epitope prediction among each model and positives in dual models were statistically evaluated, according to their sensitivity, specificity, positive prediction value (PPV) and negative prediction value (NPV) by maximum Fisher's exact test. DSM: Docking simulation model.
doi:10.1371/journal.pone.0041703.t001

epitopes (11/21, 52.3%) was significantly lower than that of shorter ones (26/31, 83.9%) ($p = 0.027$).

Successful prediction with the DSM was dependent on the HLA allele and its peptides: in HLA-B*15:01, HLA-B*27:05 and HLA-B*35:01, all of the best-defined epitopes were ranked within the top 5th. However, four best-defined epitopes restricted by HLA-B*57:01 and HLA-A*02:01 scored within the worst 5th candidates: $Nef_{120-128}$ YFPDWQNYT, $p15_{433-442}$ FLGKIWPSYK, RT_{33-41} ALVEICTEM, and $p24_{161-172}$ KAFSPEVIPMF.

Optimal epitope prediction to analyze HIV-1 CRF01_AE Gag ELISpot assay data using two *in silico* models

We next applied the model to predict optimal epitopes against HIV-1 CRF01_AE Gag based on our previously obtained results in a Thai HIV cohort study [21]. In total, 31 peptide-HLA associations were analyzed: 5 in HLA-A, 13 in HLA-B, and 13 in HLA-C (Table S2). Among these, 10 overlapping peptides spanned previously reported epitopes (6 were best-defined epitopes and 4 were published but not enlisted as best-defined epitopes). In the DSM, 9/10 (90%) reported epitopes were successfully ranked within the 5th as significant epitope candidates, and all of the six best-defined epitopes ranked either the 1st or 2nd. In HLArestrictor, 8/10 (80%) epitopes were predicted as significant binders; 3 as SB, 4 as WB, and 1 as CB, but 2 epitopes (best-defined epitopes HLA-A*02:07-restircted YL9, and HLA-B15-restricted KL9) were not predicted as significant binders. HLArestrictor also predicted another 16 sequences as potential epitope candidates: 1 as SB, 12 as WB, and 2 as CB. Intriguingly only one WB candidate was ranked within the top five by the DSM, reflecting a considerable degree of discrepancy between the two prediction methods.

8 previously unreported peptides were predicted by both models: HLA-B*38:02-restricted $p24_{198-205}$ MQMLKETI (rank 1st in DSM and WB in HLArestrictor), HLA-B*40:01-restricted $p24_{311-321}$ QEVKNWMTETL (2nd and SB), HLA-B*46:01-restricted $p24_{275-283}$ RMYSPVVSIL (5th and SB), HLA-B*58:01-restircted $p17_{79-86}$ YNTVVTLW (1st and WB), HLA-B*58:01-restricted $p17_{77-86}$ SLYNTVVTLW (4th and WB), HLA-C*01:02-restricted $p24_{277-285}$ YSPVSILDI (2nd and WB in $p24_{271-285}$ and 3rd and WB in $p24_{276-290}$), HLA-C*01:02-restricted $p24_{276-285}$ MYSPVSILDI (4th and WB both in $p24_{271-285}$ and $p24_{276-290}$), and HLA-C*01:02-restricted $p24_{296-304}$ YVDRFYKTL (1st and WB).

Application of the *in silico* DSM to define the restricting HLA molecule

We conducted a ^{51}Cr release assay with a truncated peptide titration spanning the overlapping region between Gag $p24_{271-285}$ NKIVRMYSPVSILDI (NI15) and $p24_{276-290}$ MYSPVSILDIRQGPK (MK15). These induced the largest responses both in breadth and magnitude in our previous study, and were statistically associated with HLA-A*02:07, HLA-B*46:01, and HLA-C*01:02, which we calculated to be under LD association [21]. We found strong killing against HLA-B*46:01 and HLA-C*01:02-matched $p24_{276-285}$ MYSPVSILDI (MI10)- and $p24_{277-285}$ YSPVSILDI (YI9)-pulsed target cells but not in any other condition (Figure S1). However, we could not further specify the restricting HLA molecule because a single HLA-matched target cell was not available due to the strong LD between them. Therefore, we conducted *in silico* analysis in order to identify the responsible HLA. Table 2 shows the results of the DSM between these two peptides (MI10 and YI9) and three candidate HLA alleles (HLA-A*02:07, HLA-B*46:01 and HLA-C*01:02). Firstly, with the DSM, none of these two peptides were predicted within the top five candidate epitopes when binding to HLA-A*02:07 or HLA-B*46:01, and neither scored significant binding using the HLArestrictor, eliminating these as the restricting HLA molecules. However in the model with HLA-C*01:02, both two peptides ranked within the 5th; MI10 ranked as the 3rd in NI15 and the 4th in MK15, while YI9 was ranked as the 2nd in NI15 and the 3rd in MK15. Significant binding affinity of MI10 and YI9 to HLA-C*01:02 was also predicted by HLArestrictor. Secondly, in the binding motif of HLA-C*01:02 (x[AL][P]xxxxx[L]), both MI10 and YI9 encoded compatible or similar hydrophobic amino acids with the binding motif x[Y]xxxxxxx[I] in MI10 and xx[P]xxxxx[I] in YI9. Together, these results indicate that the optimal epitopes MI10 and YI9 are equally likely candidates recognized by HLA-C*01:02, with YI9 ranking slightly higher in the DSM.

Discussion

In this study, we demonstrated that the structure-based DSM can predict the peptide binding affinity with various HLA molecules, independently of peptide binding motif information. To our knowledge, this novel DSM is the first model of its kind that succeeded in predicting HIV-1 CTL best-defined epitopes, with better or at least equivalent accuracy to the latest binding motif-based program. We also found a high detection rate of best-defined epitopes independent of peptide size in the DSM, while the detection rate significantly decreased with longer epitopes in the other model.

Historically, comparisons of epitope prediction methods has generally shown that peptide-binding motif based methods outperform structure-based methods [23]. However, the increased availability of crystal structures of MHC-peptide complexes is enabling the development of prediction methods using such structural models and the calculation of free energy of binding [23,24]. In the review by Liao *et al* [23], their comprehensive comparison of structure-based models and peptide-binding motif models in epitope prediction showed that the structure-based model was able to outperform all other methods except the ANN model, which performed equally well. In our novel program, we use a measure of the binding affinity between the HLA molecule and the peptides at four residues spanning the N- and C-terminal.

Table 2. Prediction of the HLA restriction of Gag $p24_{276-285}$ MYSPVSILDI (MI10) and $p24_{277-285}$ YSPVSILDI (YI9) using *in silico* methods.

			U_dock rank		
HLA	**Binding motif**	**Peptide**	**NI15**	**MK15**	**HLArestrictor**
A*02:07	x[L][D]xxxxx[L]	MI10	13	13	
		YI9	14	16	
B*46:01	x[M(I)]xxxxxx[YF]	MI10	15	20	
		YI9	19	21	
C*01:02	x[AL][P]xxxxxx[L]	MI10	3	4	WB
		YI9	2	3	SB

HLA restriction prediction against two reactive Gag peptides, Gag $p24_{276-285}$ MYSPVSILDI (MI10) and $p24_{277-285}$ YSPVSILDI (YI9) was performed by the docking simulation model, and the binding motif HLArestrictor 1.2. The U_dock rank by the docking simulation model against MI10 and YI9 was analyzed in the original 15-mer peptides of Gag $p24_{271-285}$ NKIVRMYSPVSILDI (NI15) and $p24_{276-290}$ MYSPVSILDIRQGPK (MK15). SB: Strong Binder, WB: Weak Binder.
doi:10.1371/journal.pone.0041703.t002

This covers not only the anchor position sites but also their flanking sites, which have a considerable effect on peptide-HLA binding; this may also have led to the high detection rate of best-defined epitopes independent of epitope size. Together with precise HLA crystal structure information, we have also incorporated a fine calculation model for binding affinity [18], giving the DSM a high detection rate of best-defined epitopes equivalent to that of the latest binding motif-based program.

Intriguingly there was a considerable degree of discrepancy between the two methods: 21.2% of the 52 best-defined epitopes were detected as significant epitope candidates only by the DSM, while 9.6% was detected only by the HLArestrictor. Furthermore, two epitopes which ranked within the bottom five by DSM were successfully predicted as a single candidate by HLArestrictor, whereas five epitopes which were not detected by HLArestrictor, were successfully predicted as the best candidates by the DSM. This result highlights the importance of combining programs with different approaches, for example those based on peptide binding motif information and those that do not require peptide binding motif information, consistent with previous report in class II HLA peptide binding prediction model [25].

We therefore applied both models to predict optimal epitopes in HIV-1 CRF01_AE Gag and found 8 previously unreported optimal epitopes supported by both models. These potential epitopes need to be further confirmed *ex vivo* that they are true epitopes capable of stimulating T cell responses with either a ^{51}Cr release assay or tetramer assay. However, since the DSM alone predicted 11 other candidates that were not predicted by the HLArestrictor, combining both models would be important to reduce the cost of such experiments. Furthermore a substantial number of OLPs were recognized using an ELISpot assay but within the peptides that induced a response, no epitope was predicted by the HLArestrictor. This DSM would save the cost of experiments by reducing 26 potential candidate peptides to five.

The ability of the DSM model to accurately predict peptides was dependent on the HLA molecule in question, and our results suggest that this is due to variations in the C-terminal binding groove. Four best-defined epitopes restricted by the alleles HLA-A*02:01 and HLA-B*57:01 ranked among the worst from the 22nd up to the 26th in our program. In HLA-A*02:01, both FK10 and AM9 coded Leucine (L) at the 2nd position of sequence, compatible with the HLA-A*02:01 binding motif at the B pocket and scored a low and therefore strongly binding U_dock score at the N-terminal site [−47.8 kcal/mol in FK10 (5th in N1-N8 terminal) and −54.4 kcal/mol in AM9 (2nd)]. However, the sequences did not match with the HLA-A*02:01 binding motif at the C-terminal which contains a Valine (V) at the F pocket, and they scored the worst U_dock scores [−14.1 kcal/mol in KF10 (8th) and −48.5 kcal/mol in AM9 (8th)]. A similarly low score at the C-terminal was also found in HLA-B*57:01-restricted KF11 [−24.5 kcal/mol (8th)] and YT9 [−23.8 kcal/mol (8th)]. The importance of the C-terminal for peptide-binding stability has been previously reported [26], and with respect to structural differences between the B and F pockets, it is generally known that the B pocket has a rather round shape while the F pocket has a deep cleft-like shape, suggesting stricter peptide binding restriction at the F pocket compared to the B pocket among HLA-A*02:01 and HLA-B*57:01. In contrast, HLA-B*27:05 and HLA-B*35:01 had none or only one variant of their binding motif at C-terminal: x[R(K)]xxxxxxx or x[R]xxxxxx[LFYRHK(MI)] in HLA-B*27:05 and x[P(AV)]xxxxxxx or x[P(AVYRD)]xxxxxx[YFMLI] in HLA-B*35:01. In these two alleles, all of the best-defined epitopes ranked within the 5th. These results strongly suggest that the diversity of peptide binding at the F pocket defines the accuracy or difficulty of epitope prediction by DSM.

Recent studies have highlighted the importance of HLA-C alleles for HIV viral control, for instance in the population-based study from Africa [27], existence of dominant HLA-C*04-restricted epitopes [28], stimulation of NK cells through HLA-C and Killer-cell Immunoglobulin-like receptors (KIRs) [29,30], and HLA-C expression control by 35 kb upstream genotype of HLA-C allele and HIV viral control [31]. However, epitope mapping of HLA-C antigens has been held back for several reasons. Firstly, in *in vitro* studies it has been difficult to find target and effector cell combinations with singly matched HLA alleles which are not under LD association, as we found in our ^{51}Cr release assay. *In silico*, in contrast to HLA-A or B alleles, epitope prediction programs against HLA-C alleles have been sparse [9–11]. This can be attributed to the lack of reported epitopes information from HLA-C alleles, since binding motif-based models were originally programmed based on such reported data. Furthermore, LD of HLA-C alleles, especially with HLA-B alleles, hinders the confirmation of HLA-C alleles as the restricting alleles in statistical analyses. In our previous study, among 13 HLA-C-associated epitope candidates, nine were reported with HLA-A or B alleles which were under LD association [21]. Novel DSM could contribute to epitope detection by bypassing such obstacles to epitope prediction against HLA-C alleles.

This study had several limitations. First, we could not define the threshold of the U_dock score degree itself in novel program as defined in HLArestrictor. Related with this limitation, considering the HLA polymorphism, reported epitope number, and comparison between alleles with/without original crystal structure information, further calculations will be warranted for the quality evaluation of DSM. Second, this is a computational epitope prediction model whose algorithm is solely based on the binding between the peptide and the HLA molecule. Although peptide-HLA binding is the most selective event for epitope determination [32], CTL activation is a multi-step process involving the processing of viral peptides by proteasome [22,33,34] and the recognition of the peptide-HLA complex by T cell receptors (TCRs) [35], both of which are not accounted for in the model.

In conclusion, we have shown here a novel *in silico* DSM which can be used for epitope mapping, and combined with a binding motif-based model, this will significantly reduce the required experimental burden for epitope identification in the development of a CTL-based vaccine for HIV.

Supporting Information

Figure S1 Identification of HLA-B*46:01/C*01:02-restricted Gag p24$_{276–285}$ MI10 and p24$_{277–285}$ YI9 by a ^{51}Cr release assay. ^{51}Cr release assays under HLA-B*46:01/C*01:02-matched conditions were performed for each peptide. Significant % lysis was found in target cells pulsed with Gag p24$_{276–285}$ MI10: MYSPVSILDI and p24$_{277–285}$ YI9: YSPVSILDI.
(PPTX)

Table S1 Predicted best-defined epitopes using the docking simulation model and a comparison with HLArestrictor. The docking simulation model was applied to predict epitopes within 15-mer peptides spanning best-defined epitopes and compared with those predicted with the HLArestrictor. The U_dock score and their rank were calculated for each peptide in the docking simulation model, while with HLArestrictor the affinity thresholds of SB: Strong Binder, WB: Weak Binder, and CB: Combined Binder, and Non-binder were given, according to their definitions.
(XLS)