Integrated Analysis of Residue Coevolution and Protein Structures Capture Key Protein Sectors in HIV-1 Proteins

Yuqi Zhao[1,2]*[‡], Yanjie Wang[3‡], Yuedong Gao[4], Gonghua Li[1], Jingfei Huang[1,5]*

1 State Key Laboratory of Genetic Resources and Evolution, Kunming Institute of Zoology, Chinese Academy of Sciences, No.32 Jiaochang Donglu Kunming, 650223 Yunnan, China, 2 Department of Integrative Biology and Physiology, University of California, Los Angeles, 610 Charles E. Young Drive East, Los Angeles, California, United States of America, 3 Key Laboratory of Animal Models and Human Disease Mechanisms of Chinese Academy of Sciences and Yunnan Province, Kunming Institute of Zoology, Chinese Academy of Sciences, Kunming, Yunnan 650223, China, 4 Kunming Biological Diversity Regional Center of Instruments, Kunming Institute of Zoology, Chinese Academy of Sciences, Kunming 650223, China, 5 Collaborative Innovation Center for Natural Products and Biological Drugs of Yunnan, Kunming, Yunnan 650223, China

‡ These authors contributed equally to this work.
* zhaoyq@mail.kiz.ac.cn (YZ); huangjf@mail.kiz.ac.cn (JH)

Abstract

HIV type 1 (HIV-1) is characterized by its rapid genetic evolution, leading to challenges in anti-HIV therapy. However, the sequence variations in HIV-1 proteins are not randomly distributed due to a combination of functional constraints and genetic drift. In this study, we examined patterns of sequence variability for evidence of linked sequence changes (termed as coevolution or covariation) in 15 HIV-1 proteins. It shows that the percentage of charged residues in the coevolving residues is significantly higher than that in all the HIV-1 proteins. Most of the coevolving residues are spatially proximal in the protein structures and tend to form relatively compact and independent units in the tertiary structures, termed as "protein sectors". These protein sectors are closely associated with anti-HIV drug resistance, T cell epitopes, and antibody binding sites. Finally, we explored candidate peptide inhibitors based on the protein sectors. Our results can establish an association between the coevolving residues and molecular functions of HIV-1 proteins, and then provide us with valuable knowledge of pathology of HIV-1 and therapeutics development.

Introduction

It has been over 30 years since human immunodeficiency virus (HIV) was first identified as the causative virus of Acquired immune deficiency syndrome (AIDS) [1]. HIV has two types, HIV-1 and HIV-2, which share many features, such as modes of transmission, intracellular replication pathways and clinical consequences [2]. However, HIV-1 is characterized by higher transmissibility and increased likelihood of progression to AIDS [3,4]. Morbidity and mortality rates due to HIV/AIDS are probably the highest in the world, with over 25 million deaths

recorded globally while at least 10,000 youths infected every month [5]. Many efforts have been made to prevent or cure HIV infection. In the recent 20 years, diverse antiretroviral drugs were developed in the treatment of HIV infection [6]. Furthermore, devising an effective vaccine to prevent HIV infection or curtail its progression is considered a promising therapeutic approach [7,8]. However, finding an effective, safe HIV vaccine or drug compound is still an ongoing struggle for HIV-1, which is mainly caused by its rapid genetic evolution. In fact, the evolution rate of HIV-1 proceeds is about 1 million times faster than that of the human genome [9], which is well evidenced from the large number of different HIV-1 strains isolated worldwide. Consequently, the high genetic variation leads to the high adaptation of HIV-1 and poses serious challenges for chemotherapy and vaccine development for HIV-1 infection [10,11]. For example, it shows that drug resistance-associated mutations are present in at least 15% to 25% of the HIV population [12]. Besides, mutations within epitopes in HIV-1 have been studied to affect host-virus interaction, with possible implications for immune recognition [13].

Despite the high degree of mutations in the HIV-1 proteins in the setting of antiretroviral therapy, the spectrum of possible virus variants seems to be limited by patterns of amino acid covariation [14]. The amino acid covariation, also known as coevolution, is conceptualized as correlated mutational behavior between columns of a multiple sequence alignment of protein sequences [15]. The structure and function of proteins need to be maintained throughout correlated substitution patterns between intra- and inter-protein residues. Such correlated mutations are suggestive of compensatory changes that occur between entangled residues to maintain protein function. For HIV proteins, the coevolution events should be more important in maintaining their functions or structures or else the high point mutations might result in severe functional inactivity at any time. Understanding what determines the phenotypical impact of these compensatory mutations is important both for planning targeted mutation experiments in the laboratory and for analyzing naturally occurring mutations found in patients. During the recent years, software and method development for assessing amino acid coevolution have made great advances. Using Statistical Coupling Analysis (SCA), Ranganathan et al. detected correlation rules in the WW domain, which describe aspects of the fold architecture going beyond simple protein contacts [16]. Onuchic et al. applied direct coupling analysis (DCA) to genomics-aided structure prediction [17]. With the increase of sequenced HIV protein sequences, we believe the covariation analysis of HIV-1 proteins will be valuable for studying the functions of HIV-1 proteins and anti-HIV therapies.

In this study, we explored all potential coevolution events in HIV-1 proteins. In addition, we applied molecular dynamic simulations to determine the structural features of the coevolving residue pairs. These resides are organized into physically contiguous networks, termed as 'protein sectors'. We further estimated the association between protein sectors and the functional sites in HIV proteins, such as drug-binding regions, catalytic sites, and epitopes. Our results can establish association between the coevolving residues and molecular functions of HIV-1 proteins.

Results

Coevolution events in HIV-1 proteins

After multiple sequence alignment and gap filtering, we detected the coevolving residues in the 15 HIV-1 proteins using DCA (Materials and Methods). It shows that the coevolution events exist in HIV-1 proteins (Fig. 1), with count from two (Fig. 1B) to 407 (Fig. 1N). The accessory proteins (P6, NEF, REV, TAT, VIF, VPR, and VPU) have significantly higher mean DI values than the other two groups (Wilcoxon rank sum test, p = $3.11{\times}10^{-4}$), including viral enzymes and structural proteins. Although the coevolving residues show different patterns among 15
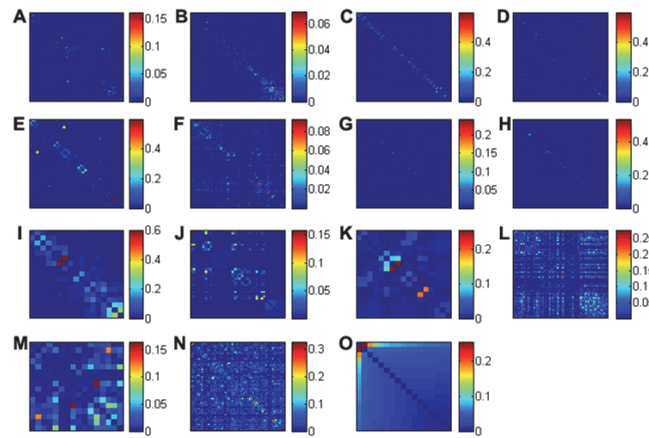
Fig 1. Coevolution patterns in 15 HIV-1 proteins. The panels (A-L) are heat maps of the direct information (DI) values of residue pairs in multiple sequence alignments of GP120, GP41, MA, CA, NC, PR, RT, IN, P6, NEF, REV, TAT, VIF, VPR, and VPU, respectively. The x- and y-axes represent the positions of amino acid residues in the multiple sequence alignments with gap filtering (see Methods).

doi:10.1371/journal.pone.0117506.g001

HIV-1 proteins, the majority of them are more proximal in protein sequences compared with random residue pairs (One-sided Two-sample Kolmogorov-Smirnov test followed by 10000 permutations, all $p < \times 10^{-2}$). In addition, we performed mutual information analysis and SCA on the multiple sequence alignments and found that the coevolving residues detected through DCA tended to show significantly higher Z-scores than the random residue pairs (two-sample student t-test, $p = 4.71 \times 10^{-5}$ and $p = 1.10 \times 10^{-4}$ separately), suggesting that our results were robust to different methods. The percentage of four charged amino acids (including Glu, Asp, Lys, and Arg) in coevolving residues was significantly higher than that in all the proteins (Fisher's exact test, $p = 2.1 \times 10^{-3}$).

Furthermore, we explored the frequency of the coevolving residue pairs and found that some residue pairs were most abundant, for example Gln-Arg and Arg-Glu. More interestingly, the frequency of the amino acid composition is quite different among different protein categories (Table 1). For the structural proteins, Lys-Glu, and Cys-Cys are the most two abundant pairs, indicating that the presence of salt bridges and disulphide bridges in the HIV-1 structural proteins are critical for their functions.

The protein structural features of coevolving residues

Coevolving residues in several protein families have been proven to work together to enable protein-protein interactions [18], promote folding [16], or contribute to an enzymatic activity [19]. As a result, we explored whether the coevolving residues in HIV-1 proteins showed specific patterns in tertiary structures.

We mapped the coevolving residues to the selected protein structures (Table 2) and then reconstructed coevolution networks for all the HIV-1 proteins (Fig. 2). Some of the residues in

Table 1. The top frequent residue pairs in HIV proteins.

HIV proteins	Structural Proteins	Viral Enzymes	Accessory Proteins
Residue pairs	LYS-GLU, CYS-CYS, ASN-THR, LEU-SER, GLU-LEU	ILE-ILE, ASN-LYS, GLY-ALA, ASP-LYS, LYS-GLN	GLN-ARG, ARG-ARG, LEU-ARG, GLN-THR, ARG-GLU

doi:10.1371/journal.pone.0117506.t001

Table 2. HIV-1 protein structures used in the study.

| Categories	Proteins	Structures	Chains	Length	References
Structural Proteins	GP120	1G9M	G	321	[25]
	GP41	2CMR	A	226	[59]
	MA	1HIW	A	137	[60]
	CA	3H47	A	231	[61]
	NC	1A1T	A	55	[62]
Viral Enzymes	PR	7HVP	A, B	99	[63]
	RT	1HYS	A	553	[24]
	IN	1EX4	A	212	[64]
Accessory Proteins	P6	2C55	A	52	[65]
	NEF	1AVV	A	130	[66]
	REV	1ETF	B	23	[67]
	TAT	1JFW	A	86	[68]
	VIF	3DCG	E,F	39	[69]
	VPR	1ESX	A	96	[70]
	VPU	1VPU	A	45	[71]

doi:10.1371/journal.pone.0117506.t002

the networks interact with more than one residue, for example, Thr81 in MA (Fig. 2A) and Asp67 in RT (Fig. 2B). Some of the coevolving residues tend to form closely connected modules, for example, residues in VIF (Fig. 2C). In addition, it shows that accessory proteins VPR and TAT cover more resides than other proteins when we set the same criteria for all the proteins. There are two reasons: these two proteins are of short lengths (96 amino acids for VPR while 86~101 for TAT) and they have relatively longer conserved sequences [20].

The average distances between the coevolving residues are significantly shorter than that between the random residue pairs (One-sided Two-sample Kolmogorov-Smirnov test followed by 10000 permutations, all $p < \times 10^{-6}$). It was observed that the coevolving residues in the proteins structures tended to be located proximal to each other, forming relatively independent units. According to previous studies, the similar units in other proteins were termed as "protein sectors" that underlie conserved, independently varying biological activities [21,22]. We found that most of the detected protein sectors in HIV-1 proteins are typically built around protein active sites (Fig. 3, S1 Fig.). For reverse transcriptase (Fig. 3A), there are three opposite charged coevolving residue pairs in the proteins sectors, including Asp67-Lys70, Glu28-Lys32, and Asp67-Lys219, which was located near (within 10 Å) the three catalytically essential amino acid residues (Asp110, Asp185, and Asp186) for polymerase catalysis [23,24]. For gp120 (Fig. 3B), the protein sector was located near the protein-protein interface between gp120 and CD40 [25], especially for Glu267, Glu268, Thr278 and Asp279, suggesting that the protein sector was involved in the HIV entry. For VPU protein, 6 out of 12 coevolving residues in the protein sector are charged amino acids (Fig. 3C). In addition, the mesh surfaces of the coevolving residues suggest that the protein sectors are relatively compact and independent in the HIV-1 protein structures (Fig. 3, S1 Fig.).

The dynamic behaviors of protein sectors in molecular dynamic simulations

Molecular dynamics (MD) simulations are becoming a standard part of workflows in structural biology and enable us study the dynamical properties of a system in full atomic details [26]. Here we applied MD simulations to all the HIV-1 proteins and explored the dynamics
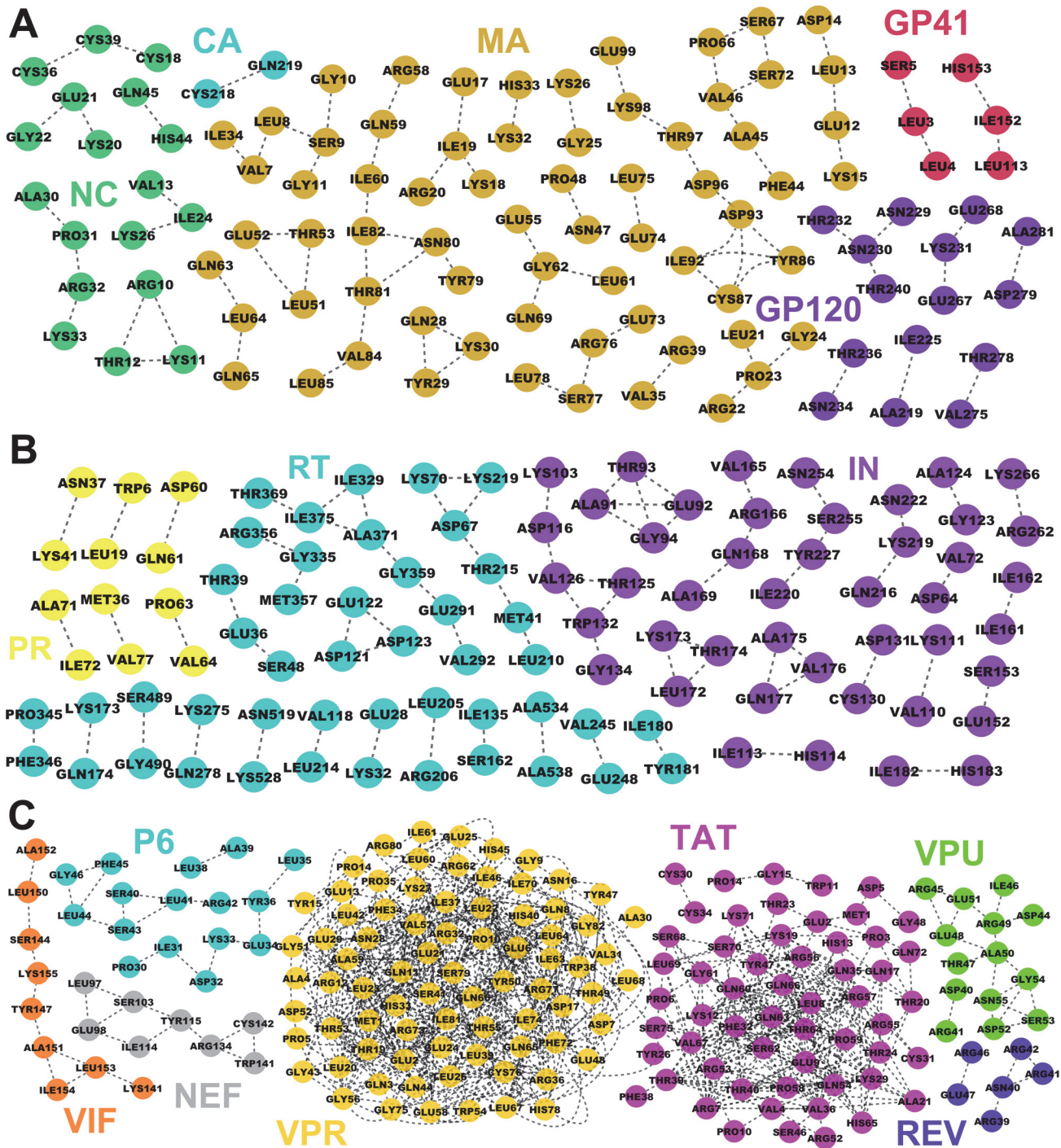
Fig 2. Coevolution networks of HIV-1 proteins. The nodes represent the amino acid residues in HIV-1 proteins while the edges are the coevolving relationships among the residues. The amino acid labels come from the protein tertiary structures (Table 2). The proteins are classified into three categories, including structural proteins (A), viral enzymes (B), and accessory proteins (C).

doi:10.1371/journal.pone.0117506.g002

behaviors of the proteins sectors in the tertiary structures. The average backbone root-mean square fluctuation (RMSF) for coevolving residues in proteins sectors is significantly smaller than the average RMSF for residues outside protein sectors (Fig. 4A-4C, S2 Fig.; two-sample
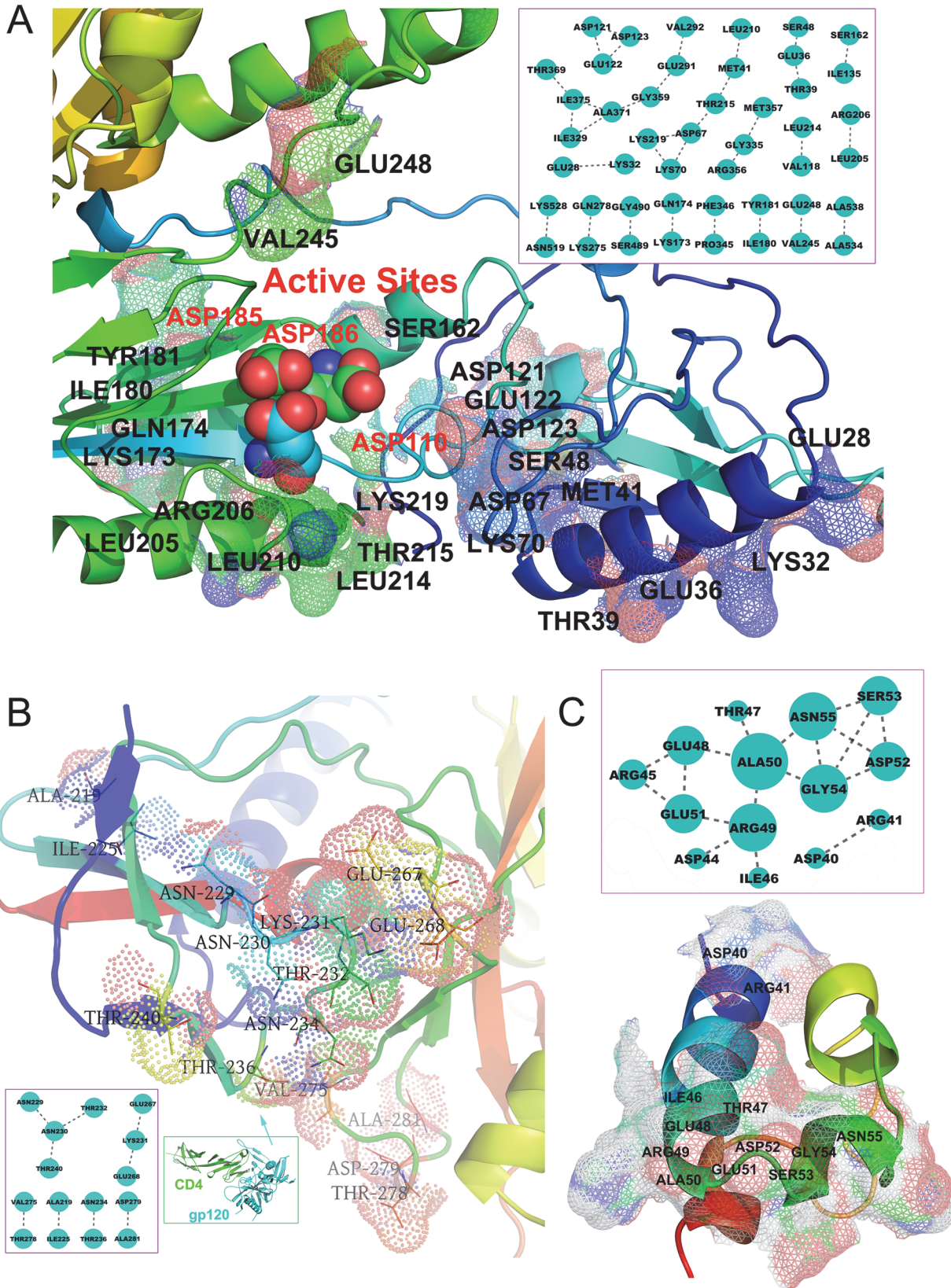
Fig 3. HIV protein sectors underlying conserved, independently varying biological activities. (A) The coevolving residues in RT enzyme were located near the three catalytically essential amino acid residues (Asp110, Asp185, and Asp186) for polymerase catalysis. (B) For gp120, the coevolving residues were located near the protein-protein interface between gp120 and CD40, especially for Glu267, Glu268, Thr278 and Asp279. (C) For VPU protein, 6 out of 12 coevolving residues in the protein sector are charged amino acids. The figures were generated using PyMol (http://www.pymol.Org). The protein structures were colored with a default spectrum of rainbow colors in Pymol. The mesh surfaces of the coevolving residues were added while the different colors correspond to different amino acid residues.

student's t-test, p = 7.16×10^{-7}), indicating that the protein sectors are significantly stable during the molecular dynamic simulations. However, we observed some exceptions, for example, Leu3, Leu4, and Ser5 in gp41. Recent studies reported that the hydrophobic fusion peptide (FP), where the coevolving residues were located, played important roles in gp41 fusion conformations but did not add stability [27]. In addition, we determined the interactions between coevolving residues during the simulations. To probe the key interactions between coevolving residues in the protein sectors, the contact map was analyzed over the 10 ns molecular dynamics simulations. We take gp120 as an example. It indicates that residues in protein sectors tend to form densely packed substructures (Fig. 5A). Snapshots of molecular dynamics simulation of protein sector in gp120 indicate that the interactions between coevolving residues are stable (Fig. 5B-5F). In addition, it shows that most of the coevolving residues were located in disordered loop structures. For other HIV-1 proteins, the protein sectors also tended to form densely packed substructures in molecular dynamic simulations (S3 Fig.).

The drug resistance mutations and epitope regions in protein sectors

Drug resistance is a common cause of treatment failure for HIV infection. We explored whether the polymorphisms in the detected protein sectors might be associated with drug resistance. It shows that for almost all the types of HIV antiretroviral drugs, the polymorphisms leading to drug resistance are involved in the coevolving proteins sectors (Table 3). For example, the coevolving residue pairs, ASP67-LYS70 and MET41-THR215 were reported to be the most common mutation patterns for nucleoside RT Inhibitors, including azidothymidine (AZT), Stavudine (d4T), Tenofovir Disoproxil Fumarate (TDF), Abacavir (ABC), Didanosine (DDI), and lamivudine (3TC) [28]. It also shows that the predominant polymorphisms of residue 36 and 77 in protease are the branched chain amino acids (Ile, Val, and Leu), but the transitions among these amino acids resulted in the resistance of protease inhibitors [29]. In addition, it was observed that the many coevolving residue pairs in protein sectors had not been studied in the HIV drug resistance, suggesting that these regions could be served as potential target sites for HIV drugs. For example, two of the six coevolving residues in gp41 have been proven important for the interactions between the C-terminal heptad repeats (CHR) and N-terminal heptad repeat (NHR) domains [30] (S4 Fig.). Moreover, the residues are located in or nearby the peptide HIV Fusion Inhibitors, such as T20 and N36 (S4 Fig.). However, the interactions between coevolving residues are still functionally unknown, which might need site-directed mutagenesis approach to identify the associations between these coevolution events HIV fusion.

We also investigated the relationships between HIV-1 T cell epitopes and protein sectors. We took gp120 as an example. It shows that protein sectors have overlaps with multiple T cell epitopes (Fig. 6A-6B). We also observed that the sequence from 229 to 236 (NNKTFNGT) was associated with helper T lymphocytes (T-helper/CD4+, Fig. 6A) while not overlapped with cytotoxic T lymphocytes (CTL/CD8+, Fig. 6B). In addition, we searched all the antibodies against gp120 in HIV Molecular Immunology Database (http://www.hiv.lanl.gov/content/immunology). It shows that the protein sectors have more overlaps with antibody epitopes than the other sites (Fisher exact test, p = 3.07×10^{-5}).
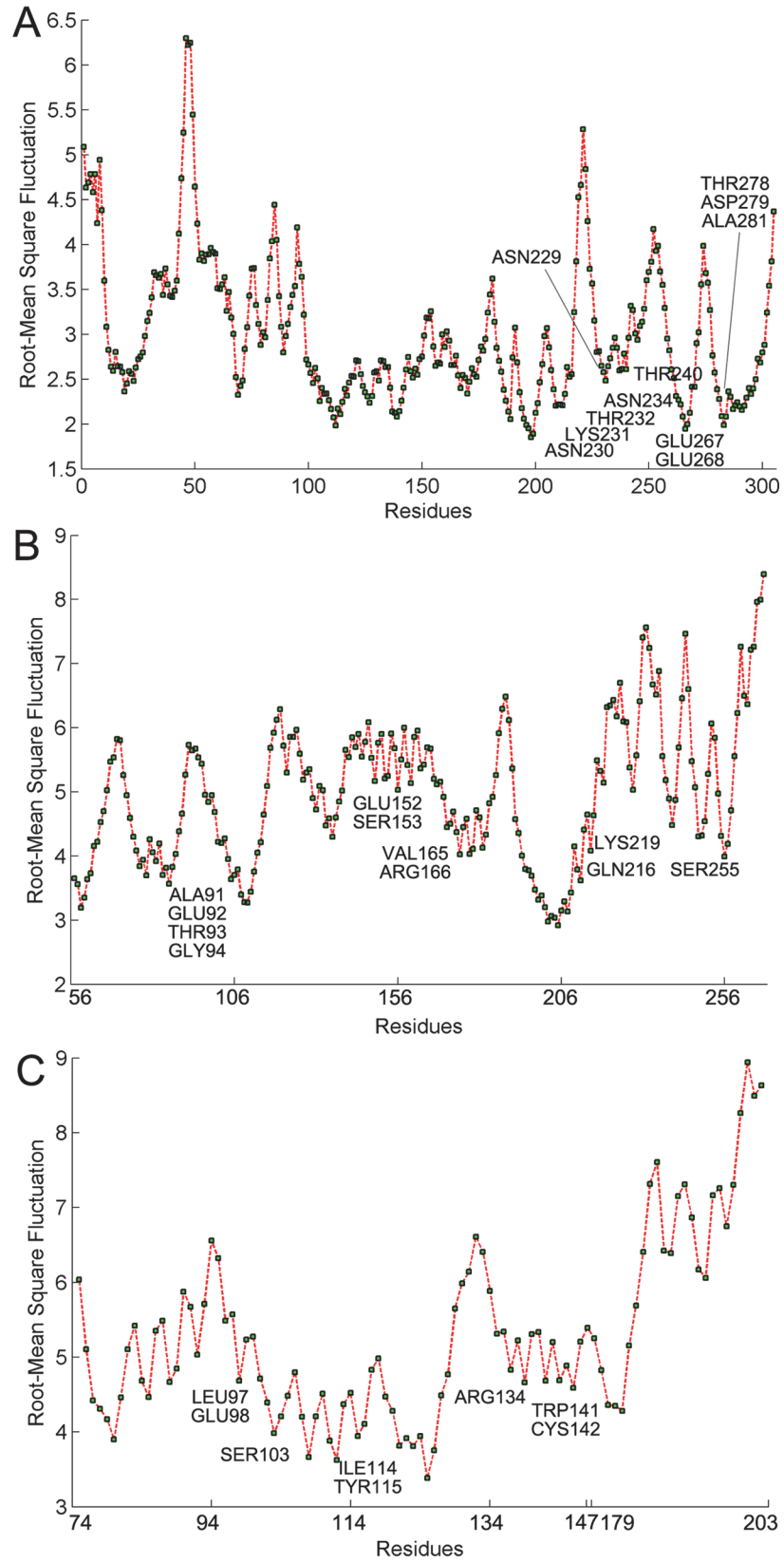
Fig 4. RMSF plot during molecular dynamic simulations. The figure shows backbone RMSF of GP120 (A), IN (B), and NEF (C) in molecular dynamics simulations of 10 ns. The x-axis represents protein sequences while the y-axis is average RMSF values.

doi:10.1371/journal.pone.0117506.g004

The candidate peptide inhibitors of HIV-1 proteins

We detected 33 candidate peptide inhibitors for 12 HIV-1 proteins (Table 4). It shows that most of the peptides (25/33) are of less than 20 amino acids. Most of the interactions between peptide inhibitors and HIV proteins agree with the previous studies [31]. However, the peptide SLLSSPQ (ID: HIP100) was reported to be an integrase inhibitor [32] while was also predicted to interact with gp41. Interestingly, the coevolving residue pair alone can act as effective HIV inhibitors. For example, the peptide DQ (ID: HIP3) was the strongest inhibitor with inhibition constants (Ki) of >1000-fold increase [33]. We can observe that the peptide DQ binds to PR near the protein sector, suggesting that the peptide inhibitor might mimic the ASP60-GLN61 residue pair and then perturbed the functions of the protein sector (S5 Fig.).

More experimental evidence for the functions of HIV-1 protein sectors

Beside the *In silico* evidence described above, recent studies now provide experimental evidence for the functions of the detected sectors in HIV-1 proteins (Table 5). For example, the protein sector in GP120 (residue 275–281in Loop D) was proven to be involved in a loop-based mechanism of CD4-binding-site recognition [34]. The protein sectors in IN enzyme were involved in late-stage event in HIV replication, the disruption of which will lead to the reverse transcription block [35]. Putting all the evidence together, we concluded that the detected protein sectors in HIV-1 proteins are essential during different steps of the HIV life cycle.

Discussion

Evolution of HIV-1 proceeds about 1 million times faster than that of the human genome, with approximately one error incorporated into the viral genome each time the virus is replicated [9]. This rapid mutation rate of HIV-1 proteins is widely considered a major stumbling block in the development of therapies to combat acquired immunodeficiency syndrome. To overcome the limitations, we determined the coevolving events in all the HIV-1 proteins and studied their structural features in the study. We found that coevolution showed quite different characteristics among different classes of HIV-1 proteins. The charged amino acids are overrepresented in the coevolving residues. The coevolving residues tend to form protein sectors in tertiary structures, in which interactions between coevolving residues show stable behaviors in the dynamic environment. These protein sectors are closely associated with HIV-1 drug resistance and epitopes. The findings will be helpful in understanding the pathogenesis and developing potential antiviral compounds.

The charged amino acids were enriched in the coevolving residues, suggesting that the interactions mediated by these charged residues are of importance to the functions of HIV-1 proteins. It is universal accepted that the salt bridges in proteins most often arise from the anionic carboxylate (COO-) of negatively charged amino acids (Asp or Glu) and the cationic ammonium/guanidinium (NH3+ or NHC(NH2)2+) from positively charged amino acids (Lys or Arg) [36]. The salt bridges in HIV proteins mediate the critical activities of the virus, such as entry to host cells [37], replication [38,39], and assembly [40]. Salt bridges are of critical importance for host-virus interactions. Wu et al. found that salt bridges formed between HIV entry inhibitors and CCR5 chemokine receptor, which acts as a co-receptor for HIV-1 viral entry, potentially locked the receptor in an inactive conformation [41]. Therefore, the salt bridges mediated
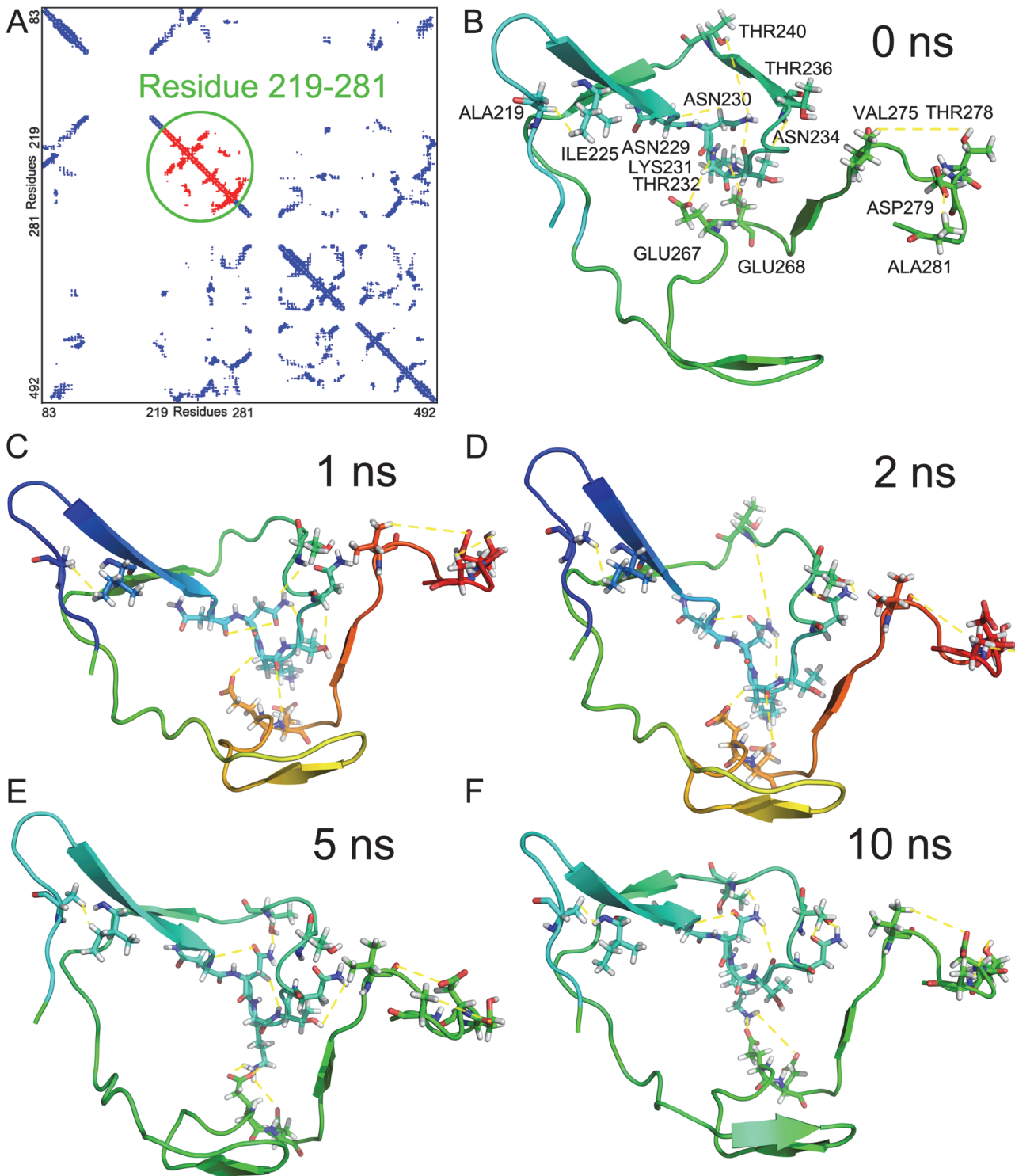
Fig 5. Interactions between coevolving residues in GP120. (A) The contact (hydrogen bonds) maps from molecular dynamics simulation of GP120; (B-F) Snapshots of the interactions between coevolving residues in protein sector of gp120 during molecular dynamics simulations of 0 ns, 1 ns, 2 ns, 5 ns, and 10 ns separately.

Table 3. Polymorphisms in protein sectors of HIV-1 proteins leading to drug resistance.

Coevolving Residues [a]	Top mutation patterns	Affected drugs	References
	Reverse transcriptase		
41, 67, 70, 210, 215	41L,67N,210W,215Y	Nucleoside reverse transcriptase inhibitors	[46]
	41L,210W,215Y		
	41L		
	67N,70R		
	70R		
	41L,215Y		
	41L,184V		
	67N,70R,215F		
	215Y		
	41L,210W		
	67N		
181	181C	non-nucleoside reverse transcriptase inhibitors	[46]
	181I		
	181V		
	Integrase		
92, 153	92Q	Integrase inhibitors	[72]
	153Y		
	Protease		
36, 63, 71, 77	36I	Protease Inhibitors	[29]
	36L		
	36V		
	63P		
	71V		
	71T		
	77I		
	gp120		
91, 92, 93, 94, 172, 173, 174, 175, 176	Changes within gp120 surrounding the Phe 43 cavity	Entry inhibitors	[73]
	Changes within gp120 (V3, C2, V2 and C4)		
	Changes within gp120 (V3, V1, V2 and V4)		

Note: [a] The residue numbering is based on the protein structures in Table 2.

doi:10.1371/journal.pone.0117506.t003

by the charged residues in the coevolving residue pairs will be suitable in designing potential anti-HIV drugs. When drugs disrupt a certain salt bridge in HIV-proteins, there will be four possible endings depending on the type of proteins, including structural collapse, inhibition of host-virus interactions, failure of virus assembly, and loss of catalytic activities. These potential mechanisms of the anti-HIV drugs can be validated by solid experimental evidence, which will be set forth as below. For envelope glycoprotein gp120, two detected coevolving residue pairs (LYS231-GLU267, and LYS231-GLU268) might be involved in the formation of salt bridges. The structural analysis validated that these two interactions were within 5 Å and they together with other coevolving residues formed a protein sector in the inner domain, which was recently proven critical for CD4-required conformational transitions in the HIV-1 Env trimer [42]. For Reverse transcriptase, there are three opposite charged coevolving residue pairs, including Asp67-Lys70, Glu28-Lys32, and Asp67-Lys219, which was located near (within 10 Å) the three
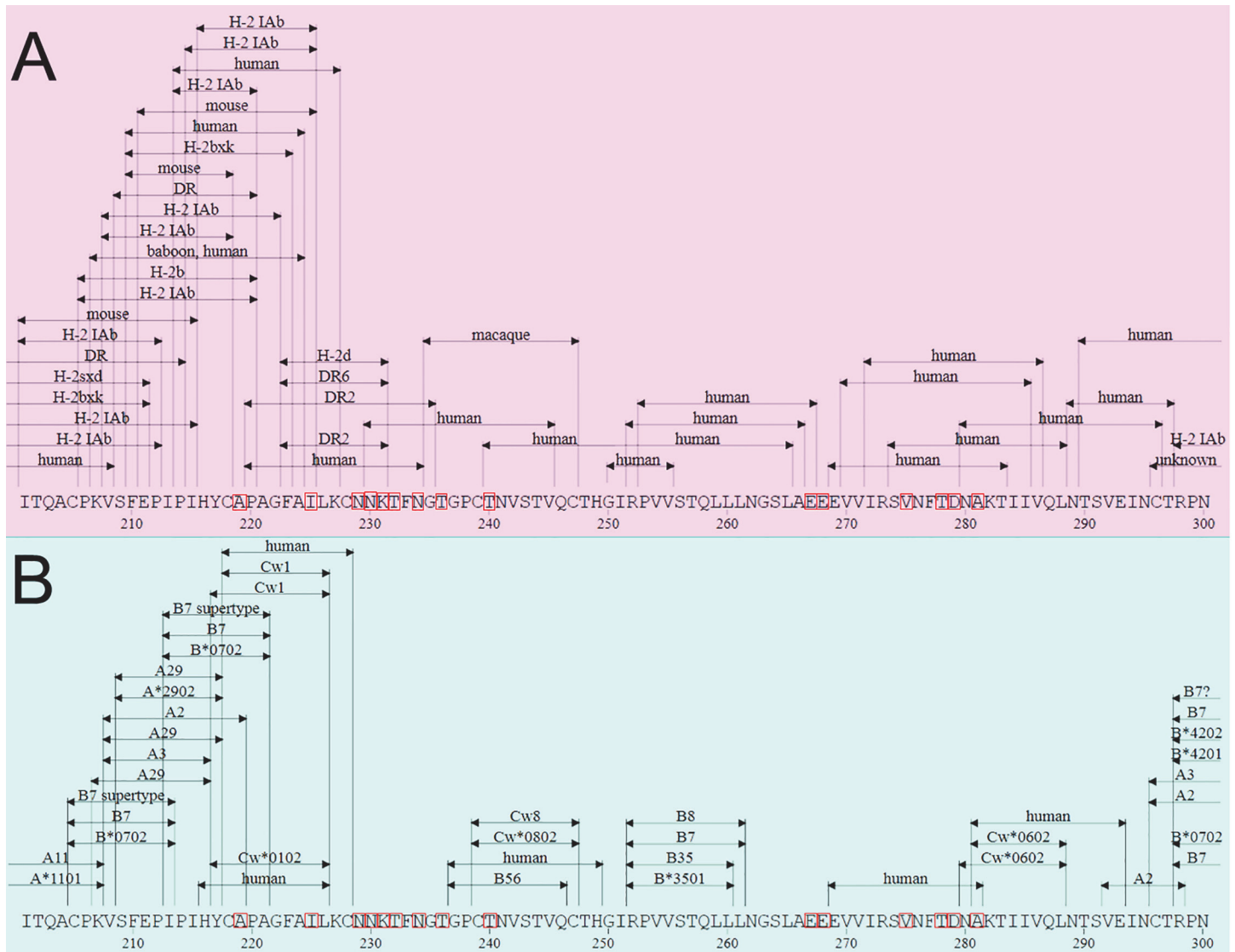
Fig 6. Epitopes of CD4+/CD8+ T lymphocytes for gp120 protein. The epitopes of CD4+ (A) and CD8+ (B) T lymphocytes in protein sector of gp120 protein.

doi:10.1371/journal.pone.0117506.g006

catalytically essential amino acid residues (Asp110, Asp185, and Asp186) for polymerase catalysis [23,24]. We presume that the salt bridges formed by these residue pairs play critical roles in maintaining the stability of the catalytic sites. A recent study found that Alizarine derivatives as new dual inhibitors of the HIV-1 reverse transcriptase-associated DNA polymerase and RNase H activities could block the salt bridges by occupying binding pockets near these co-evolving residues [43].

More interestingly, the coevolving residues in protein structures tended to form protein sectors, which are closely associated with critical features of HIV proteins, such as active sites, drug resistance, and epitopes. Protein sectors found in other protein families so far were related to conserved functional activities [15], which positively supported our findings. As a result, we will discuss according to the possible functions of the detected protein sectors. The residues in

Table 4. Candidate peptide inhibitors for HIV proteins.

Potential Targets	ID [a]	Sequence	Length	Mechanisms [b]
CA	HIP992	AAAPAATLEEHMTACQGV	18	Maturation
CA	HIP993	ALGAAATLEEMMTACQGV	18	Maturation
GP41	HIP766	RQLLSQIVQQQNNLLRAIEAQQHLLQLT	28	Fusion inhibitor
GP41	HIP763	MTLTVQARQLLSQIVQQQNNLLRAIEAQ	28	Fusion inhibitor
GP41	HIP100	SLLSSPQ	7	Integrase inhibitor
GP120	HIP962	EINCTRPNNNTRKSIRIQRGPGRAFVTIGKIGNMRQAHCNIS	42	Virus entry
GP120	HIP1127	IRKAHCNISRADWND	15	Fusion inhibitor
GP120	HIP1100	GIGDPVTCLKSGAIA	15	Fusion inhibitor
IN	HIP170	ACWWAGIKQEF	11	Integrase inhibitor
IN	HIP171	ACWWAGIRQEF	11	Integrase inhibitor
IN	HIP166	ACWWAGIKQAF	11	Integrase inhibitor
IN	HIP165	ACWWAGIKAEF	11	Integrase inhibitor
MA	HIP625	KRIVQRIKDFLRNLVPRTES	20	Multifunction
MA	HIP185	KRIVQRIKDFLR	12	Multifunction
NC	HIP1148	RSQKEGLHYTCSSHFPYSQYQFWK	24	Fusion inhibitor
NC	HIP1091	RSQKEGLHYTCSSHFPYSQYQFWK	24	Fusion inhibitor
PR	HIP3	DQ	2	Protease inhibition
PR	HIP82	LQITLW	6	Protease inhibition
PR	HIP22	SYEW	4	Protease inhibition
PR	HIP24	SYNL	4	Protease inhibition
PR	HIP20	SFNL	4	Protease inhibition
REV	HIP1168	TRQARRNRRRWRERQRAAAAC	21	Fusion inhibitor
REV	HIP1167	TRQARRNRRRWRERQR	16	Fusion inhibitor
RT	HIP276	ASCDKCQLKGEAMHG	15	Reverse Transcriptase
RT	HIP332	MHGQVDCSPGIWQLD	15	Reverse Transcriptase
RT	HIP616	KELKKIIGQVRDQAEHLKTA	20	Reverse Transcriptase
RT	HIP998	FKLPIQKETWETWWTEYWE	19	Reverse transcriptase
TAT	HIP944	GRKKRRQRRR	10	Virus entry
VIF	HIP360	SVTKLTEDRWNKPQK	15	Blocks multimerization of VIF
VIF	HIP355	RWNKPQKTKGHRGSH	15	Blocks multimerization of VIF
VIF	HIP328	LTEDRWNKPQKTKGH	15	Blocks multimerization of VIF
VPR	HIP253	KGLSGPSEWWVVWV	13	VPR
VPR	HIP256	KGLSGPTAWWVVV	13	VPR

Note: [a] The peptide IDs are HIPdb ID;

[b] The Mechanisms were the potential functions of the peptides reported in the previous studies [31].

doi:10.1371/journal.pone.0117506.t004

the protein sectors were located near the active sites, representing the structural basis for allosteric communication in proteins [44]. Especially for viral enzymes (IN, PR, and RT), one of the coupled positions is the active site and the other is the allosteric site. It can be easily inferred that binding of HIV inhibitors at the allosteric site will cause conformational changes, which consequently results in modified enzyme activity. In recent drug development studies, targeting allosteric sites of enzymes is becoming increasingly hot topic [45]. The protein sectors are available as potential target sites for allosteric activities. Second, the coevolving residues are observed closely associated with drug resistance. It was reported that over 50 percent of patients under anti-HIV therapies were infected with viruses that show resistance to antiretroviral

Table 5. Evidence for molecular functions of protein sectors in HIV-1 proteins.

Protein Sectors	Mechanisms	Experimental Approaches	References
GP120	Loop-based mechanism of CD4-binding-site recognition	Co-crystal structure with antibodies	[34]
GP41	Hydrophobic fusion peptide to infect host cell	Monotherapy	[74]
PR	Maintaining structural stability nearby active sites	Crystal structure of multidrug-resistant PR bearing 20 mutations	[75]
IN	Late-stage event in HIV replication	Treatment of virus-producing cell with non-catalytic site integrase inhibitors	[35]

doi:10.1371/journal.pone.0117506.t005

drugs [46]. The principle mechanisms for drug resistance in anti-HIV treatments are mutations (1) reduce affinity of the inhibitors for the proteins; (2) impair incorporation of nucleoside analogues; or (3) block protein-protein interactions. However, these mutations in protein sectors will not affect the basic functions of HIV proteins. For example, we found that the predominant polymorphisms of residue 36 and 77 in protease are the branched chain amino acids (Ile, Val, and Leu) and the mutations conferred high drug resistance. We presume that these compensatory mutations will have little effects on the interactions between coevolving residues while reduce the affinity between inhibitor and protein or block the access of the drug to protein cavity. At times, a single mutation in the genetic code can confer complete resistance to some antiviral drugs (Table 3), suggesting that the transitions of amino acid do not perturb the functions of proteins sectors. To avoid drug resistance, we can inactivate the whole protein sectors. In addition, we also observed that the sequences in protein sectors were associated with specific epitopes of T lymphocytes (Fig. 6). HIV-1 infection is characterized by CD4+ T cell depletion, CD8+ T cell expansion, and chronic immune activation that leads to immune dysfunction [47]. The HIV-specific CD4+ T cell response can be recovered after initiation of highly active antiretroviral therapy, which is inversely correlated with HIV viral load [48,49]. Therefore, it suggests that the HIV virus might be controlled by a vaccine incorporating improved CD4+ epitopes to induce a stronger CD4+ T cell response for helping HIV-specific CTL proliferation, together with similarly enhanced CTL epitopes [50]. The CD4+ epitopes improved by the integrated approaches of coevolution and structural analysis might be a component of a more effective second generation vaccine construct for HIV infection.

Finally, we also detected several candidate HIV inhibitory peptides. The development of drugs for HIV infection began soon after the virus was discovered 30 years ago, during which peptide inhibitors had shown budding potential to exploit HIV proteins as targets for intervention. Peptide inhibitors possess several advantages over traditional anti-HIV drugs [51]: First, peptides have little toxic side effects for their specificity; second, peptides have more diverse targets (Table 4). The detected peptides mimic the interactions between coevolving residues and then disturb the stable substructures maintained by protein sectors. Furthermore, we observed that the peptides could target the key steps involved in virus attachment, fusion and replication etc, offering potentially attractive vaccine targets during immune response to HIV infection [52].

In sum, the integrated analysis captures several key protein sectors in HIV-1 proteins, providing us with valuable knowledge of pathology of HIV-1 and therapeutics development. Although our analysis covers all the HIV-1 proteins, there is still a lot more information to dig out from these results. The functions of these protein sectors should be further validated in the process of rational vaccine design and development of diagnostical tools. A greater understanding of the functions of HIV protein sectors may be critical in anti-HIV research.

Methods

HIV-1 protein sequences and alignment

The HIV genome has nine open reading frames (ORF, leading to nine primary translation products, including ENV, GAG, NEF, POL, REV, TAT, VIF, VPR, and VPU) but 15 proteins are made in all as a result of cleavage of three of the primary products. Among these ORFs, The primary ENV product is the protein GP160, which is cleaved to GP120 and GP41. The GAG protein is synthesized as a polyprotein in the cytosol of an infected cell, and contains four functional segments: MA, CA, NC, and p6. The three POL proteins, PR (protease), RT (reverse transcriptase), and IN (integrase), provide essential enzymatic functions and are encapsulated within the particle. The sequences of 15 HIV-1 proteins were retrieved from HIV Sequence Database (http://www.hiv.lanl.gov/), with each containing over 2000 non-redundant sequences (sequence identity was set to 99% as cutoff). The multiple sequences from the same patient or transmission chains were excluded. We applied MUSCLE program (http://www.drive5.com/muscle/) in the multiple sequence alignment (MSA). After the alignment, the columns with gap ratio >20% were removed. Furthermore, we separated the multiple sequences according to the genetically distinct subtypes in the "main" group M of HIV-1 strains. Based on the perturbations at sites in the MSA [44], we found the homologue sequences in each subtype were inadequate for representing the properties of the HIV proteins.

Detecting co-evolving residues

The MSA for each HIV-1 protein was analyzed using direct coupling analysis (DCA), is a statistical inference framework used to infer direct co-evolutionary couplings among residue pairs [15]. The main output, direct information (DI) values for all column pairs, is a measure of the direct coevolutionary coupling between residue positions. High DI was previously shown to be an accurate predictor for residue–residue contacts [15]. We determined the coevolving residues as the top residue pairs with DI>0.05. Simultaneously, the mutual information between amino acid residues were calculated using the DCA program.

Besides, the MSA for each HIV-1 protein were analyzed by using the statistical coupling analysis (SCA) method [53]. The SCA correlation matrix between amino acids was turned into Z-scores (also called Standard scores). If a Z-score was above a fixed threshold (cutoff = 4), two corresponding sites were linked by an edge, and each site was represented as a node.

Selection of protein structures

All the HIV-1 protein structures were obtained from RCSB PDB database (http://www.rcsb.org/pdb/), which stored more than 2000 HIV-1 protein structures (as to Jan 28, 2014). We set the following criteria to select the proper protein structures for the following analysis. First, the selected structure showed the highest sequence coverage. Then, we performed pairwise alignments between the query sequence (from the PDB file and chain ID) and every sequence in an MSA (alignment) to find the top hit sequence. Finally, a residue number list that relates alignment numbering to structure numbering was generated.

Molecular dynamic simulations

All simulations were performed using NAMD 2.8 [54] and the CHARMM31 force field with CMAP correction [55,56]. The ionized systems were minimized for 50,000 integration steps and constrained equilibrated for 10 ns with 2 fs time stepping and frames stored each picosecond. Constant temperature (T = 310 K) was enforced using Langevin dynamics with a damping time constant of 5 per picosecond. Constant pressure (p = 1 atm) was enforced through the

Nosé-Hoover Langevin piston method with a decay period of 100 fs and a damping time constant of 50 fs. Van der Waals interaction cutoff distances were set at 12 Å (smooth switching function beginning at 10 Å) and long-range electrostatic forces were computed using the particle-mesh Ewald (PME) with a grid size of less than 1.0 Å.

Data sets of HIV-1 drug resistance and epitopes

To examine the enrichment of the functional residues in HIV-1 proteins in the coevolving residues, we collected the information of HIV-1 drug resistance and epitopes for all the HIV-1 proteins. The drug resistance data sets were retrieved from HIV Drug Resistance Database (http://hivdb.stanford.edu/) and the epitopes information was obtained from HIV Molecular Immunology Database (http://www.hiv.lanl.gov/content/immunology). Because HIV Drug Resistance Database contains information for only several HIV proteins (PR and IN), we complemented the information for other proteins through text-mining approach from the PubMed database. In addition, we retrieved base-by-base details of the landmarks of HIV-1 proteome (http://www.hiv.lanl.gov/content/sequence/HIV/MAP/annotation.html).

Screening peptide library

We first collected the candidate HIV inhibitory peptides from HIPdb database [31]. We built up a structure library for these peptides using Open Babel toolbox (version 2.3.1) [57]. Then, we performed screening peptide library against all the 15 HIV proteins using LibDock [58]. In the molecular docking process, we set the coevolving residues in protein sectors as interaction sites with default parameters in LibDock. Then, the resulting peptides were further evaluated using coevolving residue pairs. For example, the resulting peptides for REV protein should contain patterns like Arg-Asn/ Asn- Arg, Arg-Arg, or Arg-Glu.

Statistical analysis

Two-sample student's t-test was performed to compare the coevolving residue pairs and random residue pairs in coevolution calculation and molecular dynamic simulations for each HIV-1 protein. We performed Wilcoxon rank-sum test to compare the coevolution patterns between HIV proteins. All the statistical analysis was done using R.

Supporting Information

S1 Fig. Protein sectors in HIV-1 proteins. The panels (A-L) represent CA, NEF, IN, GP41, PR, REV, MA, TAT, NC, VIF, VPR, and P6 respectively. Only part of the coevolving residue pairs was labeled. The detailed coevolution events were listed in Fig. 2.
(PDF)

S2 Fig. Root-mean square fluctuation (RMSF) of residues in HIV proteins during the molecular dynamics (MD) simulations. Panel A-L represent CA, GP41, MA, NC, P6, PR, REV, RT, TAT, VIF, VPR, and VPU respectively. For each protein, X axis is the sequence of the protein structure in Table 2 while Y axis is the average RMSF during the 10 ns molecular dynamic simulations.
(PDF)

S3 Fig. Contact map of HIV proteins during molecular dynamic simulations. Panel (A-N) represent CA, GP41, IN, MA, NC, NEF, P6, PR, REV, RT, TAT, VIF, VPR, and VPU respectively. The red regions are the protein sectors. For TAT and VPR, only the top ten coevolving

residues were marked.
(PDF)

S4 Fig. Schematic representation of the coevolving residues, functional domains, and peptide fusion inhibitors in gp41. The black dashed lines between NHR and CHR indicate interactions between the residues located at the e and g positions in the NHR and at the a and d positions in the CHR. The residues at the a and d sites in the CHR helical wheel are important for formation of the internal trimer by NHR domains while the residues at the e and g sites in the NHR helical wheel are involved in interactions between the NHR and CHR domains that result in the formation of six-helix bundle. The numbers of residues of peptides corresponding to T21, N36, T20, C34, and CP32M are shown. The red dashed lines represent the detected coevolution events in gp41. The pocket-forming sequence in the NHR domain, the pocket-binding domain (PBD), GIV-motif-binding domain (GBD), and lipid-binding domain (LBD) in the CHR domain are highlighted in purple, green, blue, and orange, respectively. The gp41 reference sequence in the figure was retrieved from [30].
(PDF)

S5 Fig. The docking results between HIV-1 protease and DQ peptide inhibitor. The stars represent conformation clusters of DQ peptide.
(PDF)

Acknowledgments

Author Contributions

Conceived and designed the experiments: YQZ YJW JFH. Analyzed the data: YQZ YJW YDG GHL. Wrote the paper: YQZ JFH.

References

1. Barre-Sinoussi F, Chermann JC, Rey F, Nugeyre MT, Chamaret S, et al. (1983) Isolation of a T-lymphotropic retrovirus from a patient at risk for acquired immune deficiency syndrome (AIDS). Science 220: 868–871. PMID: 6189183

2. Nyamweya S, Hegedus A, Jaye A, Rowland-Jones S, Flanagan KL, et al. (2013) Comparing HIV-1 and HIV-2 infection: Lessons for viral immunopathogenesis. Reviews in Medical Virology 23: 221–240. doi: 10.1002/rmv.1739 PMID: 23444290

3. De Cock KM, Adjorlolo G, Ekpini E, Sibailly T, Kouadio J, et al. (1993) Epidemiology and transmission of HIV-2. Why there is no HIV-2 pandemic. JAMA 270: 2083–2086. PMID: 8147962

4. De Cock KM, Jaffe HW, Curran JW (2012) The evolving epidemiology of HIV/AIDS. Aids 26: 1205–1213. doi: 10.1097/QAD.0b013e328354622a PMID: 22706007

5. Abubakar AA (2012) Distribution of CD4 Lymphocyte Cells Among Apparently Healthy HIV Seropositive and Seronegative Populations. N Am J Med Sci 4: 120–123. doi: 10.4103/1947-2714.93882 PMID: 22454823

6. Arts EJ, Hazuda DJ (2012) HIV-1 antiretroviral drug therapy. Cold Spring Harb Perspect Med 2: a007161. doi: 10.1101/cshperspect.a007161 PMID: 22474613

7. Burton DR, Desrosiers RC, Doms RW, Koff WC, Kwong PD, et al. (2004) HIV vaccine design and the neutralizing antibody problem. Nat Immunol 5: 233–236. PMID: 14985706

8. Streeck H, D'Souza MP, Littman DR, Crotty S (2013) Harnessing CD4(+) T cell responses in HIV vaccine development. Nat Med 19: 143–149. doi: 10.1038/nm.3054 PMID: 23389614

9. Coffin JM (1995) HIV population dynamics in vivo: implications for genetic variation, pathogenesis, and therapy. Science 267: 483–489. PMID: 7824947

10. Klein F, Diskin R, Scheid JF, Gaebler C, Mouquet H, et al. (2013) Somatic mutations of the immuno-globulin framework are generally required for broad and potent HIV-1 neutralization. Cell 153: 126–138. doi: 10.1016/j.cell.2013.03.018 PMID: 23540694

11. Hosseinipour MC, Gupta RK, Van Zyl G, Eron JJ, Nachega JB (2013) Emergence of HIV drug resistance during first- and second-line antiretroviral therapy in resource-limited settings. J Infect Dis 207 Suppl 2: S49–56. doi: 10.1093/infdis/jit107 PMID: 23687289

12. Li JZ, Paredes R, Ribaudo HJ, Svarovskaia ES, Metzner KJ, et al. (2011) Low-frequency HIV-1 drug resistance mutations and risk of NNRTI-based antiretroviral treatment failure: a systematic review and pooled analysis. JAMA 305: 1327–1335. doi: 10.1001/jama.2011.375 PMID: 21467286

13. Brackenridge S, Evans EJ, Toebes M, Goonetilleke N, Liu MK, et al. (2011) An early HIV mutation with-in an HLA-B*57-restricted T cell epitope abrogates binding to the killer inhibitory receptor 3DL1. Journal of Virology 85: 5415–5422. doi: 10.1128/JVI.00238-11 PMID: 21430058

14. Rhee SY, Liu TF, Holmes SP, Shafer RW (2007) HIV-1 subtype B protease and reverse transcriptase amino acid covariation. PLoS Comput Biol 3: e87. PMID: 17500586

15. Morcos F, Pagnani A, Lunt B, Bertolino A, Marks DS, et al. (2011) Direct-coupling analysis of residue coevolution captures native contacts across many protein families. Proceedings of the National Academy of Sciences of the United States of America 108: E1293–E1301. doi: 10.1073/pnas.1111471108 PMID: 22106262

16. Socolich M, Lockless SW, Russ WP, Lee H, Gardner KH, et al. (2005) Evolutionary information for specifying a protein fold. Nature 437: 512–518. PMID: 16177782

17. Sulkowska JI, Morcos F, Weigt M, Hwa T, Onuchic JN (2012) Genomics-aided structure prediction. Proc Natl Acad Sci U S A 109: 10340–10345. doi: 10.1073/pnas.1207864109 PMID: 22691493

18. Wang J, Zhao Y, Wang Y, Huang J (2013) Molecular dynamics simulations and statistical coupling analysis reveal functional coevolution network of oncogenic mutations in the CDKN2A-CDK6 complex. FEBS Lett 587: 136–141. doi: 10.1016/j.febslet.2012.11.001 PMID: 23178718

19. Ackerman SH, Gatti DL (2011) The Contribution of Coevolving Residues to the Stability of KDO8P Synthase. PloS one 6.

20. Frankel AD, Young JAT (1998) HIV-1: Fifteen proteins and an RNA. Annual Review of Biochemistry 67: 1–25. PMID: 9759480

21. Halabi N, Rivoire O, Leibler S, Ranganathan R (2009) Protein Sectors: Evolutionary Units of Three-Dimensional Structure. Cell 138: 774–786. doi: 10.1016/j.cell.2009.07.038 PMID: 19703402

22. McLaughlin RN, Poelwijk FJ, Raman A, Gosal WS, Ranganathan R (2012) The spatial architecture of protein function and adaptation. Nature 491: 138–U163. doi: 10.1038/nature11500 PMID: 23041932

23. Ding JP, Das K, Hsiou Y, Sarafianos SG, Clark AD, et al. (1998) Structure and functional implications of the polymerase active site region in a complex of HIV-1 RT with a double-stranded DNA template-primer and an antibody Fab fragment at 2.8 angstrom resolution. Journal of Molecular Biology 284: 1095–1111. PMID: 9837729

24. Sarafianos SG, Das K, Tantillo C, Clark AD, Ding J, et al. (2001) Crystal structure of HIV-1 reverse transcriptase in complex with a polypurine tract RNA: DNA. Embo Journal 20: 1449–1461. PMID: 11250910

25. Kwong PD, Wyatt R, Majeed S, Robinson J, Sweet RW, et al. (2000) Structures of HIV-1 gp120 envelope glycoproteins from laboratory-adapted and primary isolates. Structure 8: 1329–1339. PMID: 11188697

26. Phillips JC, Braun R, Wang W, Gumbart J, Tajkhorshid E, et al. (2005) Scalable molecular dynamics with NAMD. Journal of Computational Chemistry 26: 1781–1802. PMID: 16222654

27. Sackett K, Wexler-Cohen Y, Shai Y (2006) Characterization of the HIV N-terminal fusion peptide-containing region in context of key gp41 fusion conformations. Journal of Biological Chemistry 281: 21755–21762. PMID: 16751188

28. Shafer RW (2006) Rationale and uses of a public HIV drug-resistance database. Journal of Infectious Diseases 194: S51–S58. PMID: 16921473

29. Shafer RW, Rhee SY, Pillay D, Miller V, Sandstrom P, et al. (2007) HIV-1 protease and reverse transcriptase mutations for drug resistance surveillance. Aids 21: 215–223. PMID: 17197813

30. Yu XW, Lu L, Cai LF, Tong P, Tan SY, et al. (2012) Mutations of Gln64 in the HIV-1 gp41 N-Terminal Heptad Repeat Render Viruses Resistant to Peptide HIV Fusion Inhibitors Targeting the gp41 Pocket. Journal of Virology 86: 589–593. doi: 10.1128/JVI.05066-11 PMID: 22013063

31. Qureshi A, Thakur N, Kumar M (2013) HIPdb: A Database of Experimentally Validated HIV Inhibiting Peptides. Plos One 8. doi: 10.1371/journal.pone.0082806 PMID: 24482673

32. Desjobert C, de Soultrait VR, Faure A, Parissi V, Litvak S, et al. (2004) Identification by phage display selection of a short peptide able to inhibit only the strand transfer reaction catalyzed by human immunodeficiency virus type 1 integrase. Biochemistry 43: 13097–13105. PMID: 15476403

33. Louis JM, Dyda F, Nashed NT, Kimmel AR, Davies DR (1998) Hydrophilic peptides derived from the transframe region of Gag-Pol inhibit the HIV-1 protease. Biochemistry 37: 2105–2110. PMID: 9485357

34. Liao HX, Lynch R, Zhou TQ, Gao F, Alam SM, et al. (2013) Co-evolution of a broadly neutralizing HIV-1 antibody and founder virus. Nature 496: 469-+. doi: 10.1038/nature12053 PMID: 23552890

35. Balakrishnan M, Yant SR, Tsai L, O'Sullivan C, Bam RA, et al. (2013) Non-Catalytic Site HIV-1 Integrase Inhibitors Disrupt Core Maturation and Induce a Reverse Transcription Block in Target Cells. PloS one 8.

36. Echegoyen L (2011) Modern Physical Organic Chemistry. Journal of Physical Organic Chemistry 24: 743–743.

37. He YX, Liu SW, Jing WG, Lu H, Cai DM, et al. (2007) Conserved residue Lys(574) in the cavity of HIV-1 gp41 coiled-coil domain is critical for six-helix bundle stability and virus entry. Journal of Biological Chemistry 282: 25631–25639. PMID: 17616522

38. Hinkley T, Martins J, Chappey C, Haddad M, Stawiski E, et al. (2011) A systems analysis of mutational effects in HIV-1 protease and reverse transcriptase. Nature Genetics 43: 487-+. doi: 10.1038/ng.795 PMID: 21441930

39. Sluis-Cremer N, Arion D, Kaushik N, Lim H, Parniak MA (2000) Mutational analysis of Lys(65) of HIV-1 reverse transcriptase. Biochemical Journal 348: 77–82. PMID: 10794716

40. Cortines JR, Monroe EB, Kang S, Prevelige PE (2011) A Retroviral Chimeric Capsid Protein Reveals the Role of the N-Terminal beta-Hairpin in Mature Core Assembly. Journal of Molecular Biology 410: 641–652. doi: 10.1016/j.jmb.2011.03.052 PMID: 21762805

41. Tan QX, Zhu Y, Li J, Chen ZX, Han GW, et al. (2013) Structure of the CCR5 Chemokine Receptor-HIV Entry Inhibitor Maraviroc Complex. Science 341: 1387–1390. doi: 10.1126/science.1241475 PMID: 24030490

42. Desormeaux A, Coutu M, Medjahed H, Pacheco B, Herschhorn A, et al. (2013) The Highly Conserved Layer-3 Component of the HIV-1 gp120 Inner Domain Is Critical for CD4-Required Conformational Transitions. Journal of Virology 87: 2549–2562. doi: 10.1128/JVI.03104-12 PMID: 23255784

43. Esposito F, Kharlamova T, Distinto S, Zinzula L, Cheng YC, et al. (2011) Alizarine derivatives as new dual inhibitors of the HIV-1 reverse transcriptase-associated DNA polymerase and RNase H activities effective also on the RNase H activity of non-nucleoside resistant reverse transcriptases. Febs Journal 278: 1444–1457. doi: 10.1111/j.1742-4658.2011.08057.x PMID: 21348941

44. Suel GM, Lockless SW, Wall MA, Ranganathan R (2003) Evolutionarily conserved networks of residues mediate allosteric communication in proteins. Nature Structural Biology 10: 59–69. PMID: 12483203

45. Novinec M, Korenc M, Caflisch A, Ranganathan R, Lenarcic B, et al. (2014) A novel allosteric mechanism in the cysteine peptidase cathepsin K discovered by computational methods. Nat Commun 5: 3287. doi: 10.1038/ncomms4287 PMID: 24518821

46. Clavel F, Hance AJ (2004) Medical progress: HIV drug resistance. New England Journal of Medicine 350: 1023–1035. PMID: 14999114

47. Catalfamo M, Wilhelm C, Tcheung L, Proschan M, Friesen T, et al. (2011) CD4 and CD8 T Cell Immune Activation during Chronic HIV Infection: Roles of Homeostasis, HIV, Type I IFN, and IL-7. Journal of Immunology 186: 2106–2116. doi: 10.4049/jimmunol.1002000 PMID: 21257970

48. Rosenberg ES, Billingsley JM, Caliendo AM, Boswell SL, Sax PE, et al. (1997) Vigorous HIV-1-specific CD4(+) T cell responses associated with control of viremia. Science 278: 1447–1450. PMID: 9367954

49. Rosenberg ES, Altfeld M, Poon SH, Phillips MN, Wilkes BM, et al. (2000) Immune control of HIV-1 after early treatment of acute infection. Nature 407: 523–526. PMID: 11029005

50. Okazaki T, Pendleton CD, Sarobe P, Thomas EK, Iyengar S, et al. (2006) Epitope enhancement of a CD4 HIV epitope toward the development of the next generation HIV vaccine. Journal of Immunology 176: 3753–3759. PMID: 16517744

51. Castel G, Chteoui M, Heyd B, Tordo N (2011) Phage Display of Combinatorial Peptide Libraries: Application to Antiviral Research. Molecules 16: 3499–3518. doi: 10.3390/molecules16053499 PMID: 21522083

52. Pollard RB, Rockstroh JK, Pantaleo G, Asmuth DM, Peters B, et al. (2014) Safety and efficacy of the peptide-based therapeutic vaccine for HIV-1, Vacc-4x: a phase 2 randomised, double-blind, placebo-controlled trial. Lancet Infect Dis 14: 291–300. doi: 10.1016/S1473-3099(13)70343-8 PMID: 24525316

53. Lockless SW, Ranganathan R (1999) Evolutionarily conserved pathways of energetic connectivity in protein families. Science 286: 295–299. PMID: 10514373

54. Kale L, Skeel R, Bhandarkar M, Brunner R, Gursoy A, et al. (1999) NAMD2: Greater scalability for parallel molecular dynamics. Journal of Computational Physics 151: 283–312.

55. Mackerell AD, Feig M, Brooks CL (2004) Extending the treatment of backbone energetics in protein force fields: Limitations of gas-phase quantum mechanics in reproducing protein conformational distributions in molecular dynamics simulations. Journal of Computational Chemistry 25: 1400–1415. PMID: 15185334

56. Buck M, Bouguet-Bonnet S, Pastor RW, MacKerell AD (2006) Importance of the CMAP correction to the CHARMM22 protein force field: Dynamics of hen lysozyme. Biophys J 90: L36–L38. PMID: 16361340

57. O'Boyle NM, Banck M, James CA, Morley C, Vandermeersch T, et al. (2011) Open Babel: An open chemical toolbox. J Cheminform 3: 33. doi: 10.1186/1758-2946-3-33 PMID: 21982300

58. Diller DJ, Merz KM Jr. (2001) High throuput docking for library design and library prioritization. Proteins 43: 113–124. PMID: 11276081

59. Luftig MA, Mattu M, Di Giovine P, Geleziunas R, Hrin R, et al. (2006) Structural basis for HIV-1 neutralization by a gp41 fusion intermediate-directed antibody. Nature Structural &Molecular Biology 13: 740–747. PMID: 16862157

60. Hill CP, Worthylake D, Bancroft DP, Christensen AM, Sundquist WI (1996) Crystal structures of the trimeric human immunodeficiency virus type 1 matrix protein: Implications for membrane association and assembly. Proceedings of the National Academy of Sciences of the United States of America 93: 3099–3104. PMID: 8610175

61. Pornillos O, Ganser-Pornillos BK, Kelly BN, Hua YZ, Whitby FG, et al. (2009) X-Ray Structures of the Hexameric Building Block of the HIV Capsid. Cell 137: 1282–1292. doi: 10.1016/j.cell.2009.04.063 PMID: 19523676

62. De Guzman RN, Wu ZR, Stalling CC, Pappalardo L, Borer PN, et al. (1998) Structure of the HIV-1 nucleocapsid protein bound to the SL3 Psi-RNA recognition element. Science 279: 384–388. PMID: 9430589

63. Swain AL, Miller MM, Green J, Rich DH, Schneider J, et al. (1990) X-Ray Crystallographic Structure of a Complex between a Synthetic Protease of Human Immunodeficiency Virus-1 and a Substrate-Based Hydroxyethylamine Inhibitor. Proceedings of the National Academy of Sciences of the United States of America 87: 8805–8809. PMID: 2247451

64. Chen JCH, Krucinski J, Miercke LJW, Finer-Moore JS, Tang AH, et al. (2000) Crystal structure of the HIV-1 integrase catalytic core and C-terminal domains: A model for viral DNA binding. Proceedings of the National Academy of Sciences of the United States of America 97: 8233–8238. PMID: 10890912

65. Fossen T, Wray V, Bruns K, Rachmat J, Henklein P, et al. (2005) Solution structure of the human immunodeficiency virus type 1 p6 protein. Journal of Biological Chemistry 280: 42515–42527. PMID: 16234236

66. Arold S, Franken P, Strub MP, Hoh F, Benichou S, et al. (1997) The crystal structure of HIV-1 Nef protein bound to the Fyn kinase SH3 domain suggests a role for this complex in altered T cell receptor signaling. Structure 5: 1361–1372. PMID: 9351809

67. Battiste JL, Mao HY, Rao NS, Tan RY, Muhandiram DR, et al. (1996) alpha helix-RNA major groove recognition in an HIV-1 Rev peptide RRE RNA complex. Science 273: 1547–1551. PMID: 8703216

68. Peloponese JM, Gregoire C, Opi S, Esquieu D, Sturgis J, et al. (2000) H-1-C-13 nuclear magnetic resonance assignment and structural characterization of HIV-1 Tat protein. Comptes Rendus De L Academie Des Sciences Serie Iii-Sciences De La Vie-Life Sciences 323: 883–894.

69. Stanley BJ, Ehrlich ES, Short L, Yu YK, Xiao ZX, et al. (2008) Structural insight into the human immunodeficiency virus Vif SOCS box and its role in human E3 ubiquitin ligase assembly. Journal of Virology 82: 8656–8663. doi: 10.1128/JVI.00767-08 PMID: 18562529

70. Wecker K, Morellet N, Bouaziz S, Roques BP (2002) NMR structure of the HIV-1 regulatory protein Vpr in H2O/trifluoroethanol—Comparison with the Vpr N-terminal (1–51) and C-terminal (52–96) domains. European Journal of Biochemistry 269: 3779–3788. PMID: 12153575

71. Willbold D, Hoffmann S, Rosch P (1997) Secondary structure and tertiary fold of the human immunodeficiency virus protein U (Vpu) cytoplasmic domain in solution. European Journal of Biochemistry 245: 581–588. PMID: 9182993

72. Shafer RW, Schapiro JM (2008) HIV-1 drug resistance mutations: an updated framework for the second decade of HAART. Aids Reviews 10: 67–84. PMID: 18615118

73. Poveda E, Soriano V (2006) Resistance to entry inhibitors. In: Geretti AM, editor. Antiretroviral Resistance in Clinical Practice. London.

74. Forssmann WG, The YH, Stoll M, Adermann K, Albrecht U, et al. (2010) Short-Term Monotherapy in HIV-Infected Patients with a Virus Entry Inhibitor Against the gp41 Fusion Peptide. Science Translational Medicine 2. doi: 10.1126/scitranslmed.3001697 PMID: 21532938

75. Agniswamy J, Shen CH, Aniana A, Sayer JM, Louis JM, et al. (2012) HIV-1 Protease with 20 Mutations Exhibits Extreme Resistance to Clinical Inhibitors through Coordinated Structural Rearrangements. Biochemistry 51: 2819–2828. doi: 10.1021/bi2018317 PMID: 22404139