



# Coevolution Analysis of HIV-1 Envelope Glycoprotein Complex

The Harvard community has made this  
article openly available. [Please share](#) how  
this access benefits you. Your story matters

Citation	Rawi, Reda, Khalid Kunji, Abdelali Haoudi, and Halima Bensmail. 2015. "Coevolution Analysis of HIV-1 Envelope Glycoprotein Complex." PLoS ONE 10 (11): e0143245. doi:10.1371/journal.pone.0143245. <a href="http://dx.doi.org/10.1371/journal.pone.0143245">http://dx.doi.org/10.1371/journal.pone.0143245</a> .
Published Version	<a href="https://doi.org/10.1371/journal.pone.0143245">doi:10.1371/journal.pone.0143245</a>
Citable link	<a href="http://nrs.harvard.edu/urn-3:HUL.InstRepos:23845255">http://nrs.harvard.edu/urn-3:HUL.InstRepos:23845255</a>
Terms of Use	This article was downloaded from Harvard University's DASH repository, and is made available under the terms and conditions applicable to Other Posted Material, as set forth at <a href="http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA">http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA</a>

RESEARCH ARTICLE

# Coevolution Analysis of HIV-1 Envelope Glycoprotein Complex

Reda Rawi<sup>1\*</sup>, Khalid Kunji<sup>1</sup>, Abdelali Haoudi<sup>2,3</sup>, Halima Bensmail<sup>1</sup>

**1** Computational Sciences and Engineering Center, Qatar Computing Research Institute, Hamad Bin Khalifa University, Doha, Qatar, **2** Division of Genetics and Genomics, Boston Children's Hospital, Harvard Medical School, Boston, MA, United States of America, **3** King Abdullah International Medical Research Center, King Abdulaziz Medical City, Riyadh, Saudi Arabia

\* [rawi@qf.org.qa](mailto:rawi@qf.org.qa)

## Abstract

The HIV-1 Env spike is the main protein complex that facilitates HIV-1 entry into CD4<sup>+</sup> host cells. HIV-1 entry is a multistep process that is not yet completely understood. This process involves several protein-protein interactions between HIV-1 Env and a variety of host cell receptors along with many conformational changes within the spike. HIV-1 Env developed due to high mutation rates and plasticity escape strategies from immense immune pressure and entry inhibitors. We applied a coevolution and residue-residue contact detecting method to identify coevolution patterns within HIV-1 Env protein sequences representing all group M subtypes. We identified 424 coevolving residue pairs within HIV-1 Env. The majority of predicted pairs are residue-residue contacts and are proximal in 3D structure. Furthermore, many of the detected pairs have functional implications due to contributions in either CD4 or coreceptor binding, or variable loop, gp120-gp41, and interdomain interactions. This study provides a new dimension of information in HIV research. The identified residue couplings may not only be important in assisting gp120 and gp41 coordinate structure prediction, but also in designing new and effective entry inhibitors that incorporate mutation patterns of HIV-1 Env.



## OPEN ACCESS

**Citation:** Rawi R, Kunji K, Haoudi A, Bensmail H (2015) Coevolution Analysis of HIV-1 Envelope Glycoprotein Complex. PLoS ONE 10(11): e0143245. doi:10.1371/journal.pone.0143245

**Editor:** Yuxian He, Chinese Academy of Medical Sciences, CHINA

**Received:** August 25, 2015

**Accepted:** November 2, 2015

**Published:** November 18, 2015

**Copyright:** © 2015 Rawi et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** All relevant data are within the paper and its Supporting Information files.

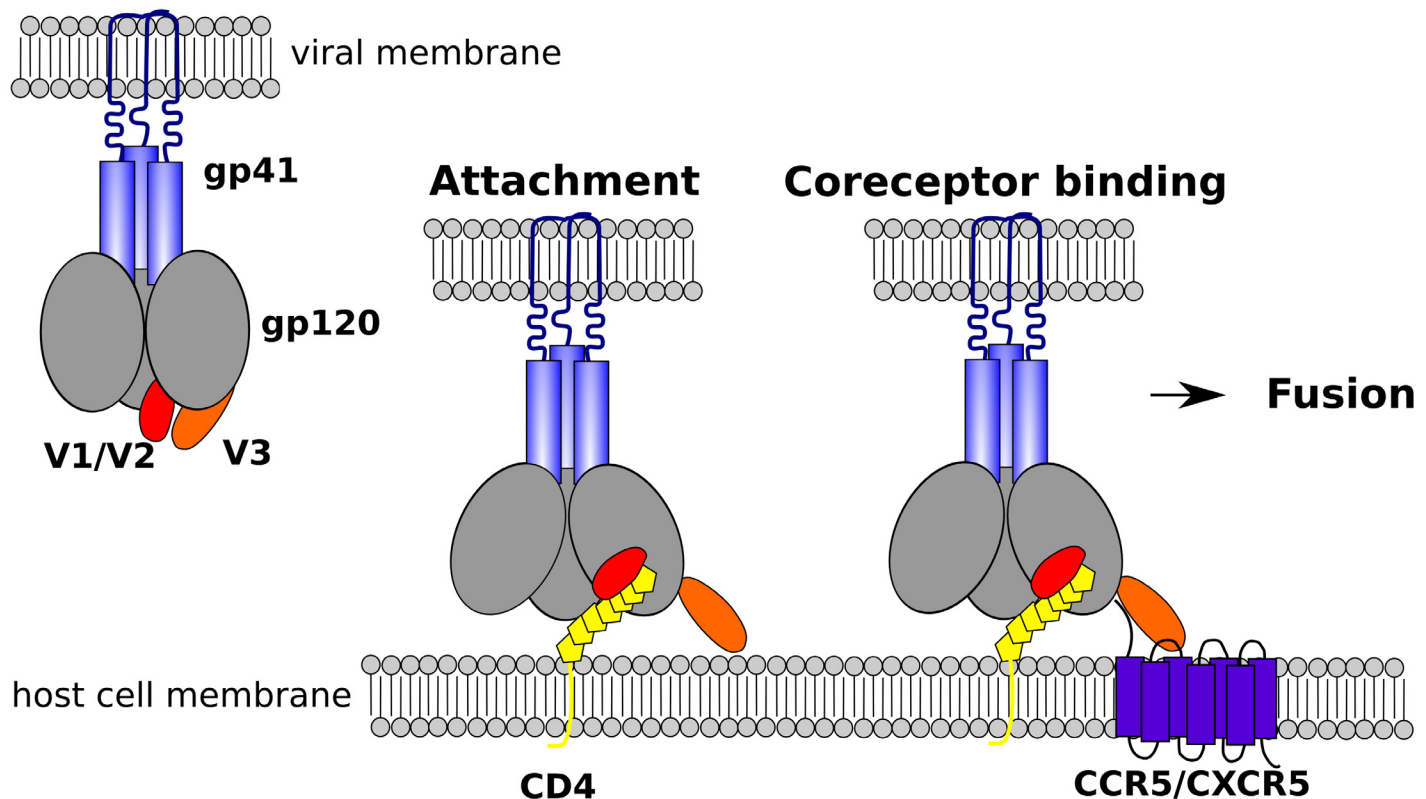
**Funding:** The authors have no support or funding to report.

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

## Introduction

Human immunodeficiency virus type 1 (HIV-1) envelope (Env) glycoprotein complex mediates binding and entry into human host cells. It is a heterodimer composed of a non-covalently bound exterior surface glycoprotein 120 (gp120) and transmembrane glycoprotein 41 (gp41) located as trimers at the surface of the viral membrane. The surface of the protein complex is highly glycosylated, enabling evasion of immune pressure. The entry process involves three main steps (see [Fig 1](#)). The attachment, initiated by the interaction of gp120 and the Cluster of Differentiation 4 Receptor (CD4), which triggers major conformational changes in gp120, including the formation of the bridging sheet (BS), spatial approach of inner (ID) and outer domain (OD) (as defined by Kwong et al. [1]) and the detachment of the variable loop 3 (V3), resulting in formation and exposure of the chemokine coreceptor binding site [1–5]. Next, the

## Native Env Trimer



**Fig 1. HIV cell entry.** Schematic illustration of HIV-1 entry steps attachment and coreceptor binding.

doi:10.1371/journal.pone.0143245.g001

coreceptor binding, where gp120 binds in general either C-C Chemokine Receptor 5 (CCR5) or C-X-C Chemokine Receptor 4 (CXCR4), causing further conformational changes that lead to re-arrangements of the previously inaccessible gp41 into an intermediate state in which the fusion peptide of gp41 is embedded into the host cell membrane. The final step is the fusion of the viral and host cell membranes. Despite that several crystal and cryo-electron microscopy/tomography structures of gp120 in unliganded state exist [6–25] (as well as in complex with CD4, CD4 mimics, or various antibodies, and of gp41 in intermediate and post-fusion state), a comprehensive understanding of structural arrangements and communication within gp120 and gp41 domains during entry is far from complete. Interestingly, even though HIV-1 Env is target of immense immune pressure, revealed through extensive sequence diversity in the Env gene, it still maintains the protein complex structure and entry functionality. Hence, detection of coevolution of important sites in Env sequences may not only point out interesting biological interactions, but also highlight functional constraints of protein structure that could help in decrypting the complexity of function and communication during HIV entry.

The extraction of coevolution patterns out of a multiple sequence alignment (MSA) has been targeted by numerous studies during the past decades [26–31] (a recent review is provided by de Juan et al. [32]). For many years such methods required large numbers of homologous and variable protein sequences, and were not able to distinguish between real direct couplings and indirect correlations that arise from phylogenetic relationships within the sequences. Recent methodological improvements, incorporated in methods such as PSICOV [33], DCA

[34, 35], plmDCA [36] or GREMLIN [37, 38] have overcome the drawbacks and demonstrated enormous accuracy in predicting real couplings and coevolution.

The majority of previous work, that studied coevolution within HIV-1 Env focused on the third variable loop (V3) [39–41], applying different sets of sequence subtypes with widely different prediction outcomes. The first coevolution study that considered the complete Env gene was performed by Travers and co-authors [42], where they included several HIV-1 group M subtypes (A,B,C,D,F,G,H,I,K) to identify coevolving pairs present among all subtypes. A recent study by Garimalla et al. [43] applied the coevolution detecting method DCA [35] on clade B HIV-1 gp120 protein sequences. Two other recent studies by Zhao et al. [44] and Li et al. [45] applied DCA and an ensemble of coevolution detecting techniques on a set of HIV-1 proteins.

In this study, we used the GREMLIN (Generative REgularized ModeLs of proteINs) approach, the most accurate method currently available for detecting coevolving residue pairs out of MSAs, and predicted 424 coevolving residue pairs within Env. The majority are real residue-residue contacts and are proximal in one of the gp120 or gp41 coordinate structures. Furthermore, we detected many coevolving pairs that have functional implications, such as CD4 or coreceptor binding, or variable loop, gp120-gp41, and interdomain interactions.

This new information should be considered in future coordinate structure predictions, but also when designing new and effective entry inhibitors to account for possible resistance mutations. To date, only two inhibitors have been approved; Maraviroc, a CCR5 antagonist that prevents the interaction between gp120 and CCR5 by blocking the transmembrane coreceptor cavity within the coreceptor, and T-20, a fusion inhibitor that prevents the fusion of the viral and host cell membranes by binding to gp41.

## Materials and Methods

### Dataset and Alignment

The input MSA was obtained from the HIV sequence database (<http://www.hiv.lanl.gov/>). We downloaded the filtered web alignment consisting of all group M subtype sequences including recombinants from the year 2013. The filtered web alignment represents a pre-cleaned alignment, excluding sequences with large insertions, high content of ambiguity codes, and multiple frame shifts. We subsequently applied several filtering steps. Initially we removed all sequences that contain non-standard amino acids or a gap. Next, we applied the pre-processing protocol suggested by the GREMLIN developers, which is composed of three additional steps. In the first step, we extracted all sequences from the MSA that have more than 25% gaps, followed by the removal of all columns with more than 25% gaps. The final filtering step was processed using HHfilter, a part of HHsuite (version: 2.0.15) [46], to generate a non-redundant MSA at 90% sequence identity. The final input MSA is available in (S1 File).

### Protein coordinate structures

The Protein Data Bank [47] (<http://www.rcsb.org>) was accessed to obtain seven HIV-1 Env crystal coordinate structures to evaluate the residue-residue contact predictions. We applied crystal structures representing gp120 in complex with CD4 and neutralising antibodies (PDB ID: 1GC1 [1], PDB ID: 2B4C [11], PDB ID: 2QAD [12]), gp120 in complex with antibody VRC01 (PDB ID: 3NGB [17]), gp120 including a gp41-interactive region (PDB ID: 3JWD [16]), stabilised HIV-1 Env in complex with antibodies PGT122 and 35O22 (PDB ID: 4TVP [48]), and the first stabilised trimeric structure of HIV-1 Env in complex with PGT122 (PDB ID: 4NCO [49]). A residue-residue contact prediction was considered true if the two coevolving amino acids are proximal in one of the seven 3D coordinate structures, in particular, if their  $C_{\beta}$ - $C_{\beta}$  ( $C_{\alpha}$ - $C_{\alpha}$  in the case of glycine) distance is less than 8 Ångström (Å) or their

minimum atomic distance is less than 6 Å. This approach has been applied by Jones and coauthors [33].

## GREMLIN

GREMLIN [37, 38] is a method to learn a statistical model that simultaneously captures conservation and coevolution in a MSA applying a pseudo-likelihood approach. It constructs a global statistical model of the paired alignment, assigning a probability to every amino acid sequence by optimising a regularised pseudo-likelihood objective fitness function in a statistically consistent method to estimate two parameters: position-specific amino acid propensities and amino acid coupling between positions. Previous approaches estimated those two parameters using an approximate moment matching approach by inverting a generalised covariance matrix [33, 35]. These rely on a Gaussian-like approximation to the global partition function. Unlike these approaches, estimation via the pseudo-likelihood avoids this approximation relying instead on local partition functions [36, 37]. The resulting general regularised structure learning is equivalent to an optimisation problem that is efficiently solved using standard convex optimisation techniques and provides estimates for both parameters.

## Results

The first and most critical step during coevolution analysis is the construction of the protein MSA. Hence, we obtained the filtered pre-made web HIV-1 Env alignment from the HIV sequence database to ensure the quality of the alignment. We restricted the analysis to the top  $L/2$  predictions (in our case 424), with  $L$  as the number of columns in the MSA (in our case  $L = 847$ ). This number of top predictions has previously been applied by many research groups to benchmark their coevolution and residue-residue contact detecting methods, including the GREMLIN [38] developers. Further, Michel et al. [50] showed in their structure prediction application, applying Rosettas ab initio folding tool [51], that the consideration of  $L/2$  predicted couplings as distance constraints, showed the best performances in the case of PSICOV [33] and plmDCA [36]. We identified coevolving pairs of amino acids in all gp120 and gp41 domains (see Table 1 and S1 Table). It is striking that a large number of coevolving residue pairs (in detail 54) are predicted within the first variable loop (V1), considering that the loop is composed of only 24 amino acids. In general, it was noteworthy that the variable loops in gp120 account for more than 30% of the coevolving pairs, despite that the fraction of amino acids is only around 17% of the total HIV-1 Env length. Furthermore, they identified more interdomain coevolving pairs.

To evaluate the performance of the coevolution predictions we applied seven gp120 coordinate structures (see Materials and Methods). A prediction was considered true if the coevolving residues had a  $C_{\beta}$ - $C_{\beta}$  ( $C_{\alpha}$ - $C_{\alpha}$  in the case of glycine) distance less than 8 Å or a minimum atomic distance less than 6 Å in at least one of the seven structures. The structural analysis revealed that the majority of predicted crystallised coevolving pairs are in contact; 84% of the predictions are true positive (TP). However, we also identified long-distance coevolving residue pairs that may play important roles as interdomain, alternative conformation or binding-partner contacts.

## Predicted coevolving pairs in V3

V3, a highly sequence- and structure-variable loop within gp120, is of essential functional, immunological and structural importance during the entry of HIV into human host cells. Previous coevolution studies in HIV-1 Env mainly focused on V3 and identified several coevolving amino acid pairs [39–42]. In our study, we identified 24 coevolving pairs within V3 (see Fig 2B

**Table 1. Count of coevolving residue pairs within and between specific HIV-1 Env regions.**

	SP	C1	V1	V2	C2	V3	C3	V4	C4	V5	C5	FP	Ecto	TM	Endo
SP	29	2	0	0	0	0	0	0	0	0	0	0	0	0	0
C1		6	1	0	7	0	1	0	1	0	0	0	1	0	0
V1			54	0	0	1	0	0	0	0	0	0	0	0	0
V2				24	1	2	0	0	2	0	0	0	0	0	0
C2					34	2	10	0	4	1	4	0	1	0	0
V3						24	2	1	3	0	0	0	0	0	0
C3							34	4	0	2	0	0	0	0	0
V4								37	0	0	0	0	0	0	0
C4									8	3	0	0	0	0	0
V5										6	1	0	0	0	0
C5											2	0	2	0	0
FP												5	0	0	0
Ecto													40	0	0
TM														0	2
Endo															58

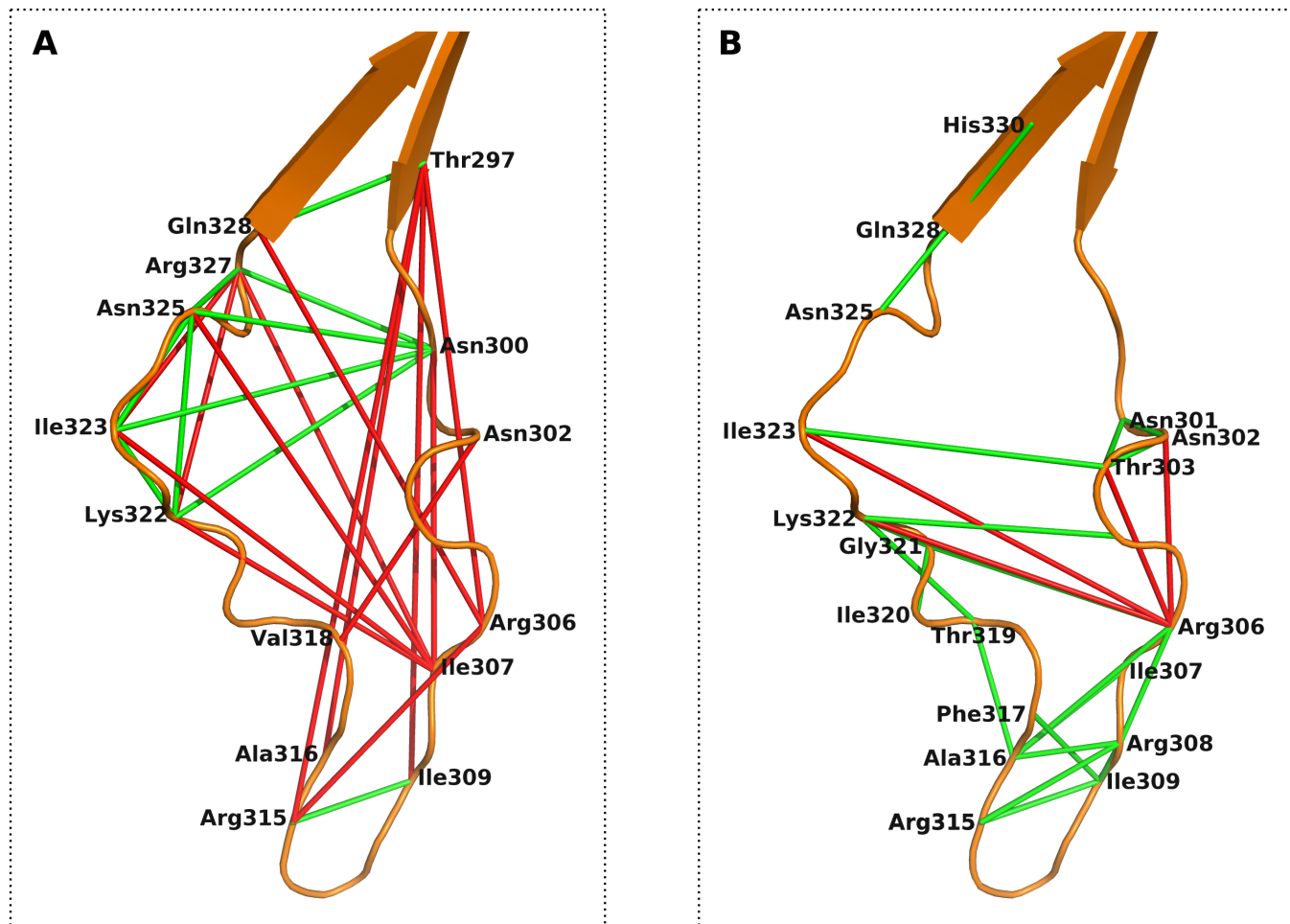
HIV-1 Env regions: signal peptide (SP), conserved regions (C1–C5), variable loops (V1–V5), fusion peptide (FP), ectodomain (Ecto), transmembrane domain (TM) and endodomain (Endo).

doi:10.1371/journal.pone.0143245.t001

and [S2 Table](#)), of which only four are false positive (FP). We mapped the predicted residue pairs on the HIV-1 gp120 coordinate structure solved by Huang et al. [11] and highlighted them as connected coloured bonds (TP shown in green and FP in red) in [Fig 2B](#) (all figures were generated using PyMOL software [52]). To compare our results with previous work [42], we mapped their predictions on the same coordinate structure shown in [Fig 2A](#) and detected that only nine out of 24 coevolution predictions are TP. Interestingly, all FP predicted contacts in our study involve residues that are either coreceptor binding (Thr303, Arg306, Ile323) or coreceptor specific sites (Asn302, Thr303, Arg306, Asp322, Ile323), according to Korber and Gnanakaran [53] (residue numbering is according to the HXB2 reference sequence, Uniprot [54] ID: P04578). The predicted coevolution between these residues is most likely mediated through their interaction with one of the chemokine receptors, either CCR5 or CXCR4, and, hence, a typical example for an interaction partner mediated coevolution.

Next to coevolving pairs within V3, we also identified eleven coevolving residues located in V3 and other structural regions in the HIV-1 Env glycoprotein complex (see [Fig 3A](#) and [S3 Table](#)), with two of the residue pairs predicted as FP, in particular the pairs Glu293-Thr297 and His330-Ser334. The interaction between these two coevolving pairs is also mediated through a binding partner, N-linked glycans (see [Fig 3B](#)). Among the eleven predictions we identified three coevolving residue pairs located in V1 and the second variable loop (V2), and V3 respectively, in particular Ile154-Asn300, Glu172-Lys305 and Tyr173-Lys305 (see [Fig 3C](#)). The three missing coevolving pairs (out of the eleven) are Thr303-Ser440, Asn325-Arg419 and His330-Thr415. The first two pairs are binding mediated coevolving pairs. The third pair, His330-Thr415, might represent an interesting coevolution pair, with His330 reported as coreceptor binding [53] and Thr415, located at the end of variable loop 4 (V4), adjacent to critical residues that maintain gp160 processing and maturation [55].

The plentitude and composition of intra- and inter-coevolution of V3 residues reflects the functional and structural importance of V3 during the entry into host cells. Further, this coevolution suggests extensive communication across the whole protein complex.

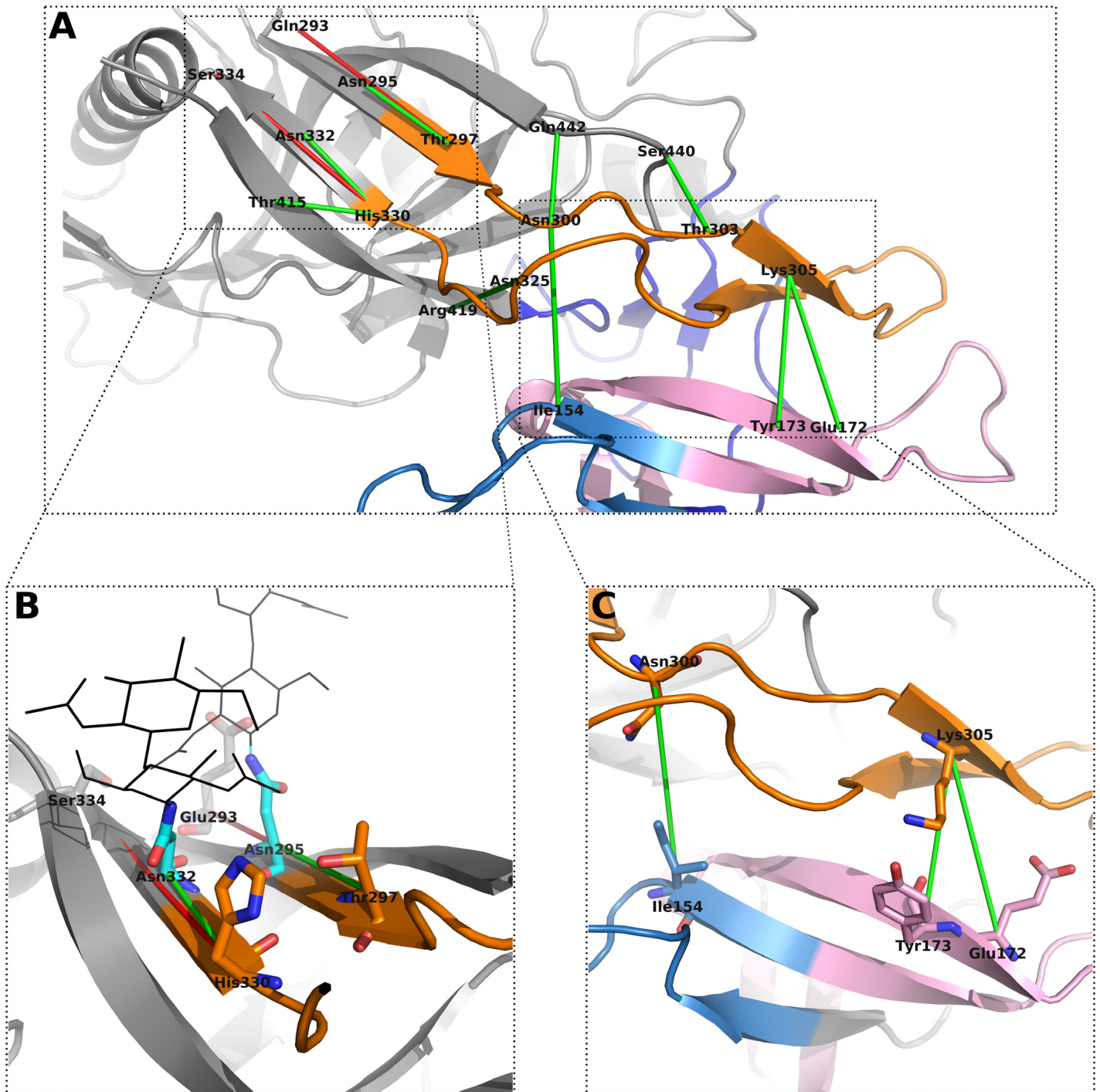


**Fig 2. Predicted coevolving residue pairs within V3.** TP predicted coevolving pairs are connected with a green dash and the FP ones are shown as red bonds. Amino acid numbering is according to the HXB2 reference sequence and the V3 coordinate structure solved by Huang et al. [11] is applied for visualisation. (A) Travers and co-authors [42] identified 24 coevolving pairs of which the majority is FP. (B) Coevolution predictions made by GREMLIN identified almost exclusively TP.

doi:10.1371/journal.pone.0143245.g002

### Predicted coevolving pairs in V1V2

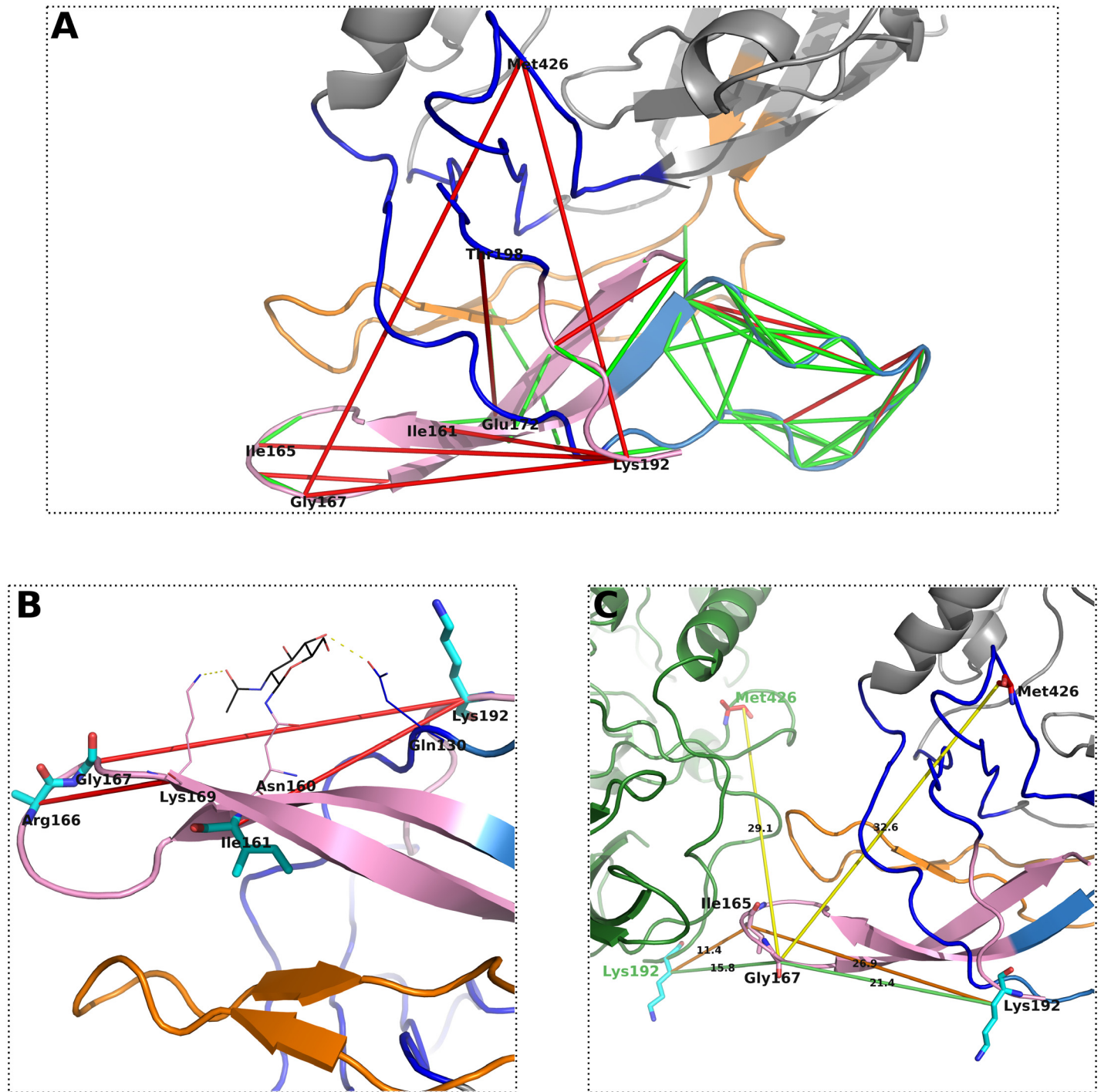
As previously mentioned, it has been reported that the V1V2 domain is important in shielding the coreceptor binding site and in conformational control of gp120 structure [22]. In our study, we identified 85 coevolving residue pairs that include at least one residue from the V1V2 region (see S1 Table). Out of the 59 intra-domain pairs, 47 are TP and 12 FP. Interestingly, we identified only coevolving pairs between residues either within V1 or V2, but no residue coevolution between the two loops. In Fig 4A we highlighted the 59 predicted residue pairs as green and red bonds. V1 is coloured skyblue and V2 pink, while V3 is indicated in the background in orange and the BS is shown in dark blue. The FP predicted residue pairs within V1 have minimum atomic distances between 6.29 Å for amino acid pair Glu150-Ile154 and 11.42 Å for Met149-Ile154. We presume that the FP predicted pairs may be TP in other conformations, since previous studies reported that the V1V2 region is in motion upon interaction with CD4 and the coreceptors. In fact, the recently published work by Munro et al. [56] showed that the unliganded HIV-1 Env is intrinsically dynamic, by transitioning between three distinct conformations. Hence, the predicted residue pairs may be TP in one of the three characterised



**Fig 3. Predicted coevolving pairs between amino acids located in V3 and other structural regions of HIV-1 Env.** Gp120 is shown in cartoon representation, with V1 coloured in blue, V2 in pink and V3 in orange. (A) All inter V3 coevolving pairs are highlighted with green (TP) or red (FP) coloured dashes. (B) Coevolving amino acid pairs Glu293-Asn295, Glu293-Thr297, His330-Asn332 and His330-Ser334 (shown in sticks representation) are mediated by N-linked glycans (shown as black lines). (C) Predicted contacts between amino acids located in V1/V2 and V3. The involved amino acids are highlighted as coloured sticks.

doi:10.1371/journal.pone.0143245.g003





**Fig 4. Coevolving pairs between amino acids in V1V2 and other structural HIV-1 Env domains.** V1, V2 and V3 are shown in skyblue, pink and orange coloured cartoon illustration. (A) 59 predicted and in [49] crystallised coevolving residue pairs; with TP illustrated as green and FP as red dashes. (B) Two long-distance coevolving amino acids are quite likely mediated by a N-linked glycan. The involved amino acids are shown in stick representation. (C) Three long-distance residue pairs (Ile165-Lys192, Gly167-Lys192, and Gly167-Met426) are presumably inter gp120 contacts. The intra- and inter-gp120 distances are shown as coloured (orange, light green and yellow) bonds. The inter-gp120 distances are in all cases smaller than the intra-gp120 ones.

doi:10.1371/journal.pone.0143245.g004

conformations. Furthermore, we identified N-linked glycan mediated long-distance coevolving pairs within V2, similar to V3. The involved pairs are Ile161-Lys192 and Gly167-Lys192, with Ile161 adjacent to a glycan binding asparagine amino acid, Asn160, which was recently shown to be among the essential N-linked glycosylation sites that interfere in the interaction with monoclonal antibodies such as 2G12 [57].

The coevolving pair Arg166-Lys169 may also have an effect on the glycan binding by contributing with Lys169 as direct binding partner of the glycan (see Fig 4B). The coevolving pair Gly167-Lys192 might also be an inter-gp120 contact within the Env trimeric complex (see Fig 4C), with a smaller atomic distance to the neighbouring gp120 than the intra-gp120. The same applies for two other pairs, Ile165-Lys192 and Gly167-Met426. In particular, the two long-distance coevolving pairs, Gly167-Lys192 and Gly167-Met426, might represent interesting communication sites between functionally important regions, such as Met426 as a CD4 binding residue located adjacent to the Phe43 cavity, and Gly167 as adjacent to coreceptor specific and N-linked glycan binding site.

### Predicted coevolving pairs including CD4 binding residues

HIV entry into host immune cells is initiated by the interaction of gp120 and CD4, which triggers conformational change in the Env protein complex. We investigated coevolving residue pairs, including residues that directly bind CD4 and residues that coevolve, but are not direct binding partners of CD4 (see Fig 5). Among the 27 coevolving pairs, only seven are FP. Four of these FP coevolving pairs are present in a subnetwork located above the Phe43 cavity of gp120, at the nexus of the bridging sheet (BS), inner domain (ID) and outer domain (OD). The remaining three long-distance coevolving pairs are located in the BS and V2 and might play key roles in inter-gp120 domain interaction, intra-gp120 communication connecting important CD4 binding residues located at the BS with residues adjacent to N-linked glycan binding and coreceptor specific sites, or different conformational arrangements of gp120 since it is well documented in previous work that conformational change is triggered following CD4 binding. Furthermore, it is worth mentioning that the residues within this CD4 coevolution network are located in different regions of gp120, in particular the BS, OD, V2, V4, as well as V5.

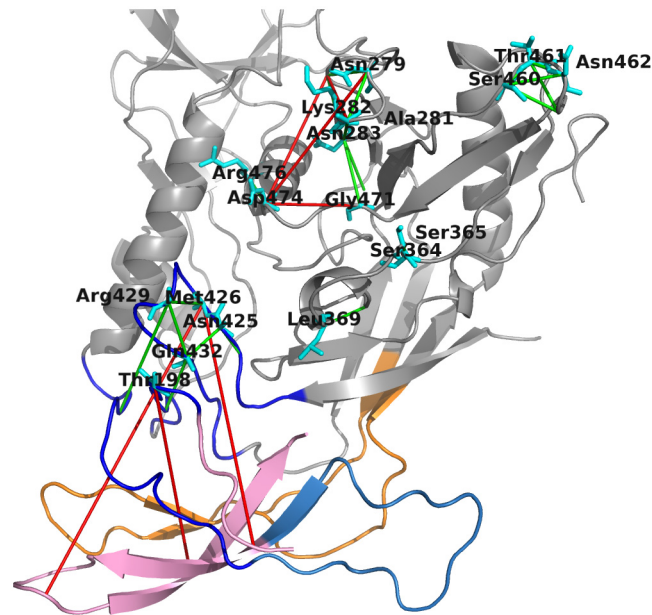
### Inter gp120-gp41 coevolving residue pairs

We identified four coevolving pairs between residues located in gp120 and gp41 (see Table 2). Two of the pairs are proximal in the coordinate structure solved by Pancera et al. [48], although separated by more than 100 amino acids in the sequence. The coevolving pair Val84-Ala578, although predicted as FP, involves two important residues, with Val84 adjacent to Val85, which has been previously reported as gp41 interacting [16], and Ala578, recently showed [58], that when mutated, influences the sensitivity of HI viruses to fusion/entry inhibitors T-20 and C34, by reduced anti-HIV-1 activity and decreased  $\alpha$ -helicity of the gp41 N-terminal heptad-repeat.

The last pair within this subset, Pro238-Glu630, might be coevolving within a subnetwork that affects gp120-gp41 interaction. Pro238 is further coevolving with residues Gln92 and Thr236, and Glu630 with Arg633 (see Fig 6). The coevolving partners of Pro238 (Gln92) and Glu630 (Arg633) have a minimum atomic distance of 7.83 Å. Also, Gln92 and Pro238 are reported to be gp41 interface contacts.

### Intra gp41 coevolving residue pairs

Among our 424 predictions, we identified 105 coevolving residue pairs within gp41 (see S1 Table). However, due to the lack of a complete gp41 coordinate structure that comprises all



**Fig 5. CD4 coevolution network.** The coevolution network is composed of residues that directly bind CD4, highlighted as cyan coloured sticks and labeled in black. Detected coevolving pairs are shown as green (TP) or red bonds (FP).

doi:10.1371/journal.pone.0143245.g005

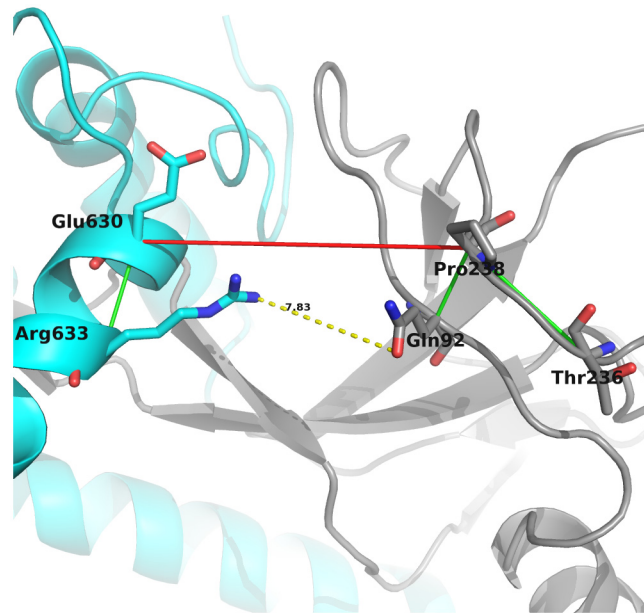
functional regions, we were not able to judge all coevolving pairs according to structural proximity. Nevertheless, we applied the 3D-structure solved by Pancera et al. [48] and evaluated the coevolving pairs whose residues are crystallised, by splitting the predicted intra-gp41 pairs into two subsets. In the first subset we included residue pairs adjacent in sequence with a maximum distance of five. The majority of predicted coevolving pairs, 76, are adjacent in sequence and within the first subset. The residues of 20 out of the 76 pairs are structurally solved and all of them are TP. We assume that the remaining pairs are also proximal in structure and TP due to the adjacency in sequence. More than half of the residue pairs, 44 out of 76, are located in the endodomain of gp41 with 7 pairs coevolving between residues located in the highly immunogenic region, known as the Kennedy epitope.

The second subset is composed of 29 coevolving pairs, with five residue pairs crystallised in the coordinate structure solved by Pancera and co-authors [48]. Only one out of five pairs is TP. However, the remaining four residue pairs are also structurally proximal with minimum distances between 8.22 Å and 11.8 Å. Out of 29 coevolving pairs, 16 are predicted between residues located in the endodomain of gp41, which is C-terminal to the viral membrane-spanning domain.

**Table 2. Predicted coevolving pairs between residues located in gp120 and gp41.**

	Pos i		Pos j		GREMLIN score	TP
17	502	C5	607	gp41	0.72	1
184	500	C5	619	gp41	0.33	1
317	84	C1	578	gp41	0.16	0
416	238	C2	630	gp41	0.23	0

doi:10.1371/journal.pone.0143245.t002



**Fig 6. Coevolution network of the inter gp120-gp41 coevolving pair Pro238-Glu630.** This pair may affect the gp120-gp41 interaction, although they are not proximal, through their intra-domain coevolving residue partners.

doi:10.1371/journal.pone.0143245.g006

## Discussion

In this study, we successfully predicted coevolving pairs of residues within Env across all HIV-1 group M subtypes. We also identified residues of high biological interest, whose evolution is under functional, structural and interactional constraints. Previous coevolution studies within Env, mainly focussed on V3 and detected subtype specific coevolution [39–41]. Travers et al. [42] were the first that considered the complete Env gene in their analysis applying a coevolution detecting method based on substitution correlations [59]. Recent methodological improvements that establish a global statistical model from the MSA and infer direct contacts that disentangle directly from indirectly coupled positions, are more suitable in this context. Therefore, we applied the GREMLIN approach, the most accurate method currently available, for detecting coevolving residue pairs out of MSAs.

Within the top  $L/2$  predicted pairs (424 pairs), we detected that the variable loops in gp120 account for more than 30% of the coevolving pairs. Such a concentration of coevolving pairs within the variable loops is not surprising, considering that coevolution detecting methods require variations at the sequence level. Travers and co-authors [42] identified more coevolving pairs in the conserved rather than in the variable regions of gp120. Remarkably, 54 coevolving pairs have been observed within V1, a small loop composed of only 24 amino acids. Despite this, V1 and V2 are highly sequence flexible due to immense immune pressure, but still maintain functionality in shielding the coreceptor binding site from antibodies and are involved in glycosylation.

HIV-1 V3 plays a crucial role in coreceptor binding and is the main determinant of coreceptor usage. Previous studies suggested a two-fold interaction of V3 with the coreceptor, pinpointing the interaction of the tip with the coreceptor's binding pocket and the base with the coreceptor's N-terminus. We predicted coevolving pairs within and between residues in V3 and other Env domains. The identified intra-V3 pairs turned out to be almost exclusively TP, applying a structural performance criterion that evaluates structural proximity. Applying the

same performance criteria on Travers et al. [42] intra-V3 predictions, we identified that the majority are FP (see Fig 2). Nevertheless, we also observed four FP within our intra-V3 subset that may hint at a binding-partner mediated coevolution between the residues, since the involved amino acids are known to be either coreceptor binding (Thr303, Arg306, Ile323) or coreceptor specific sites (Asn302, Thr303, Arg306, Asp322, Ile323). The FP-predicted coevolving pairs might also present a critical intra-V3 communication, since it has been shown that Arg306, among other residues located at the tip of V3, is an important amino acid involved in the interaction of V3 with the chemokine receptor binding pocket, whereas its coevolving residue partners (Asn302, Thr303, Asp322, Ile323) are required in the interaction with the N-terminal part of the receptors [12, 60]. Beyond that, we identified coevolving pairs between residues located in V3 and other Env domains, amongst others a binding-partner mediated coevolution, the N-glycan mediated coevolution between amino acid residue pairs Glu293-Thr297 and His330-Ser334, and a coevolution between residues in V1V2 and V3 (see Fig 3). The important interaction between V1V2 and V3 has already been reported and emphasised by several groups [61–67], describing it as a mechanism of HIV to shield the coreceptor binding site, located around the stem of V3, from antibodies. However, in most of the previous studies they inferred the interactions between V1V2 and V3 from low-resolution electron-microscopy structures. In this study, we pinpoint the interacting amino acid pairs, which are in particular Ile154-Asn300, Glu172-Lys305 and Tyr173-Lys305. The first coevolving pair Ile154-Asn300 is a critical V1V2–V3 communication, since this is the only coevolving residue pair including a residue located in V1 and another Env domain. In addition, Asn300, located next to a critical glycan binding site and involved in coreceptor binding, has the coevolving residue partner Gln442 (see Fig 3A), which also performs interaction with the coreceptor. The other two remaining coevolving pairs, Glu172-Lys305 and Tyr173-Lys305, include Lys305 located in V3, which according to Schnur et al. [60] is also involved in coreceptor binding. One of the two coevolving partners is Tyr173, which was recently highlighted as one of two tyrosines (sulfated form) in V2 that mediate and stabilise intramolecular interaction between V2 and V3 by mimicking the sulfated tyrosines in chemokine receptor CCR5 and antibodies such as 412d [65]. The second coevolving partner of Lys305 is the neighbouring Glu172. This residue has other coevolving partners (see S1 Table), such as the residue Tyr198, located in the BS. Tyr198 is an interesting residue within the BS, because it is not only adjacent to a glycan binding site, but also a CD4 contact residue and coreceptor specific [17, 53].

Furthermore, we have emphasised many coevolving pairs that are located in other Env regions, such as V1, V2 or the ID and OD, and that are also binding-partner mediated, either by N-glycans or CD4. We illustrated a CD4 network including residues that directly bind CD4 and their coevolving residue partners (see Fig 5). Within this network we identified coevolving pairs that might be involved in intra- or inter-protein communication, especially the pairs Asp167-Met426 and Arg192-Met426. Previous studies [1, 3] showed that upon CD4 binding major conformational re-arrangements take place, including the detachment of V3. The identified coevolving residues might be part of functionally important locations that maintain overall protein functionality effecting conformation and communication within the HIV-1 Env trimer.

In addition, we identified many coevolving residues within gp41. Most of the detected pairs are adjacent in sequence and, hence, most likely proximal in structure. Due to the lack of complete coordinate structures of gp41 in different states during HIV entry, we were not able to assign biological meanings to all pairs. Nevertheless, Travers et al. [42] identified coevolving pairs that support the model suggested by Hollier and Dimmock [68] that the C-terminal part of gp41 consists of 3 membrane-spanning domains and 2 ectodomains, a major and a minor. However, evidence against the suggested model has been presented by Postler and co-authors

[69]. Their experiments point to the conventional model composed of one membrane spanning domain without any extracellular loops. Within our identified endodomain set of coevolving residues, we were not able to identify coevolving residues that specifically support one of the two models.

Despite that we assigned biological explanations to the majority of identified coevolving pairs, some of the residue couplings might be due to intra- or inter-protein communication to conserve Env functionality during the process of entry into host cells. However, some might just be real FP, although the GREMLIN approach proved to be very sensitive, especially when considering only the top  $L/2$  predictions.

This coevolution study adds a new dimension of information to consider in HIV research. The most interesting coevolving residue pairs, for instance those located in the variable loops, may be evaluated for their importance in future mutagenesis studies. Newly-designed entry inhibitors or antibodies, including attachment inhibitors targeting gp120, coreceptor antagonists, or fusion inhibitors targeting gp41 should account for coevolution information to anticipate possible resistance mutations that may emerge within coevolving networks of the targeted residues.

## Supporting Information

**S1 File. HIV-1 Env MSA.** Filtered web alignment of all HIV-1 Env group M subtype protein sequences from the year 2013 in FASTA format.  
(FASTA)

**S1 Table. Coevolving residue pairs in HIV-1 Env.**  
(PDF)

**S2 Table. Coevolving residues within V3.**  
(PDF)

**S3 Table. Coevolving pairs between residues in V3 and other structural regions in the HIV-1 Env.**  
(PDF)

## Acknowledgments

The authors gratefully acknowledge Anthony Fauci, Carl Dieffenbach and Peter Kwong from the National Institute of Allergy and Infectious Diseases at the National Institute of Health, Bethesda, for their insightful comments.

## Author Contributions

Conceived and designed the experiments: RR HB. Performed the experiments: RR. Analyzed the data: RR KK. Contributed reagents/materials/analysis tools: RR. Wrote the paper: RR KK AH HB.

## References

1. Kwong PD, Wyatt R, Robinson J, Sweet RW, Sodroski J, Hendrickson WA. Structure of an HIV gp120 envelope glycoprotein in complex with the CD4 receptor and a neutralizing human antibody. *Nature*. 1998 Jun; 393(6686):648–59. doi: [10.1038/31405](https://doi.org/10.1038/31405) PMID: [9641677](https://pubmed.ncbi.nlm.nih.gov/9641677/)
2. Rizzuto CD, Wyatt R, Hernández-Ramos N, Sun Y, Kwong PD, Hendrickson WA, et al. A conserved HIV gp120 glycoprotein structure involved in chemokine receptor binding. *Science (New York, NY)*. 1998 Jun; 280(5371):1949–53. doi: [10.1126/science.280.5371.1949](https://doi.org/10.1126/science.280.5371.1949)

3. Sattentau QJ, Moore JP. Conformational changes induced in the human immunodeficiency virus envelope glycoprotein by soluble CD4 binding. *The Journal of experimental medicine*. 1991 Aug; 174(2):407–15. doi: [10.1084/jem.174.2.407](https://doi.org/10.1084/jem.174.2.407) PMID: [1713252](https://pubmed.ncbi.nlm.nih.gov/1713252/)
4. Chen B, Vogan EM, Gong H, Skehel JJ, Wiley DC, Harrison SC. Determining the structure of an unliganded and fully glycosylated SIV gp120 envelope glycoprotein. *Structure (London, England: 1993)*. 2005 Feb; 13(2):197–211. doi: [10.1016/j.str.2004.12.004](https://doi.org/10.1016/j.str.2004.12.004)
5. Wyatt R, Moore J, Accola M, Desjardin E, Robinson J, Sodroski J. Involvement of the V1/V2 variable loop structure in the exposure of human immunodeficiency virus type 1 gp120 epitopes induced by receptor binding. *Journal of virology*. 1995 Sep; 69(9):5723–33. PMID: [7543586](https://pubmed.ncbi.nlm.nih.gov/7543586/)
6. Wild C, Greenwell T, Matthews T. A synthetic peptide from HIV-1 gp41 is a potent inhibitor of virus-mediated cell-cell fusion. *AIDS research and human retroviruses*. 1993 Nov; 9(11):1051–3. doi: [10.1089/aid.1993.9.1051](https://doi.org/10.1089/aid.1993.9.1051) PMID: [8312047](https://pubmed.ncbi.nlm.nih.gov/8312047/)
7. Chan DC, Fass D, Berger JM, Kim PS. Core structure of gp41 from the HIV envelope glycoprotein. *Cell*. 1997 Apr; 89(2):263–73. doi: [10.1016/S0092-8674\(00\)80205-6](https://doi.org/10.1016/S0092-8674(00)80205-6) PMID: [9108481](https://pubmed.ncbi.nlm.nih.gov/9108481/)
8. Tan K, Liu J, Wang J, Shen S, Lu M. Atomic structure of a thermostable subdomain of HIV-1 gp41. *Proceedings of the National Academy of Sciences of the United States of America*. 1997 Nov; 94(23):12303–8. doi: [10.1073/pnas.94.23.12303](https://doi.org/10.1073/pnas.94.23.12303) PMID: [9356444](https://pubmed.ncbi.nlm.nih.gov/9356444/)
9. Weissenhorn W, Dessen A, Harrison SC, Skehel JJ, Wiley DC. Atomic structure of the ectodomain from HIV-1 gp41. *Nature*. 1997 May; 387(6631):426–30. doi: [10.1038/387426a0](https://doi.org/10.1038/387426a0) PMID: [9163431](https://pubmed.ncbi.nlm.nih.gov/9163431/)
10. Kwong PD, Wyatt R, Majeed S, Robinson J, Sweet RW, Sodroski J, et al. Structures of HIV-1 gp120 Envelope Glycoproteins from Laboratory-Adapted and Primary Isolates. *Structure*. 2000 Dec; 8(12):1329–1339. doi: [10.1016/S0969-2126\(00\)00547-5](https://doi.org/10.1016/S0969-2126(00)00547-5) PMID: [11188697](https://pubmed.ncbi.nlm.nih.gov/11188697/)
11. Huang Cc, Tang M, Zhang MY, Majeed S, Montabana E, Stanfield RL, et al. Structure of a V3-containing HIV-1 gp120 core. *Science (New York, NY)*. 2005 Nov; 310(5750):1025–8. doi: [10.1126/science.1118398](https://doi.org/10.1126/science.1118398)
12. Huang CC, Lam SN, Acharya P, Tang M, Xiang SH, Hussan SSU, et al. Structures of the CCR5 N terminus and of a tyrosine-sulfated antibody with HIV-1 gp120 and CD4. *Science (New York, NY)*. 2007 Sep; 317(5846):1930–4. doi: [10.1126/science.1145373](https://doi.org/10.1126/science.1145373)
13. Chen L, Kwon YD, Zhou T, Wu X, O'Dell S, Cavacini L, et al. Structural basis of immune evasion at the site of CD4 attachment on HIV-1 gp120. *Science (New York, NY)*. 2009 Nov; 326(5956):1123–7. doi: [10.1126/science.1175868](https://doi.org/10.1126/science.1175868)
14. Zhou T, Xu L, Dey B, Hessel AJ, Van Ryk D, Xiang SH, et al. Structural definition of a conserved neutralization epitope on HIV-1 gp120. *Nature*. 2007 Feb; 445(7129):732–7. doi: [10.1038/nature05580](https://doi.org/10.1038/nature05580) PMID: [17301785](https://pubmed.ncbi.nlm.nih.gov/17301785/)
15. Diskin R, Marcovecchio PM, Bjorkman PJ. Structure of a clade C HIV-1 gp120 bound to CD4 and CD4-induced antibody reveals anti-CD4 polyreactivity. *Nature structural & molecular biology*. 2010 May; 17(5):608–13. doi: [10.1038/nsmb.1796](https://doi.org/10.1038/nsmb.1796)
16. Pancera M, Majeed S, Ban YEA, Chen L, Huang Cc, Kong L, et al. Structure of HIV-1 gp120 with gp41-interactive region reveals layered envelope architecture and basis of conformational mobility. *Proceedings of the National Academy of Sciences of the United States of America*. 2010 Jan; 107(3):1166–71. doi: [10.1073/pnas.0911004107](https://doi.org/10.1073/pnas.0911004107) PMID: [20080564](https://pubmed.ncbi.nlm.nih.gov/20080564/)
17. Zhou T, Georgiev I, Wu X, Yang ZY, Dai K, Finzi A, et al. Structural basis for broad and potent neutralization of HIV-1 by antibody VRC01. *Science (New York, NY)*. 2010 Aug; 329(5993):811–7. doi: [10.1126/science.1192819](https://doi.org/10.1126/science.1192819)
18. Diskin R, Scheid JF, Marcovecchio PM, West AP, Klein F, Gao H, et al. Increasing the potency and breadth of an HIV antibody by using structure-based rational design. *Science (New York, NY)*. 2011 Dec; 334(6060):1289–93. doi: [10.1126/science.1213782](https://doi.org/10.1126/science.1213782)
19. Pejchal R, Doores KJ, Walker LM, Khayat R, Huang PS, Wang SK, et al. A potent and broad neutralizing antibody recognizes and penetrates the HIV glycan shield. *Science (New York, NY)*. 2011 Nov; 334(6059):1097–103. doi: [10.1126/science.1213256](https://doi.org/10.1126/science.1213256)
20. Wu X, Zhou T, Zhu J, Zhang B, Georgiev I, Wang C, et al. Focused evolution of HIV-1 neutralizing antibodies revealed by structures and deep sequencing. *Science (New York, NY)*. 2011 Sep; 333(6049):1593–602. doi: [10.1126/science.1207532](https://doi.org/10.1126/science.1207532)
21. Tran EEH, Borgnia MJ, Kuybeda O, Schauder DM, Bartesaghi A, Frank GA, et al. Structural mechanism of trimeric HIV-1 envelope glycoprotein activation. *PLoS pathogens*. 2012 Jan; 8(7):e1002797. doi: [10.1371/journal.ppat.1002797](https://doi.org/10.1371/journal.ppat.1002797) PMID: [22807678](https://pubmed.ncbi.nlm.nih.gov/22807678/)
22. Kwon YD, Finzi A, Wu X, Dogo-Isonagie C, Lee LK, Moore LR, et al. Unliganded HIV-1 gp120 core structures assume the CD4-bound conformation with regulation by quaternary interactions and variable

- loops. *Proceedings of the National Academy of Sciences of the United States of America*. 2012 Apr; 109(15):5663–8. doi: [10.1073/pnas.1112391109](https://doi.org/10.1073/pnas.1112391109) PMID: [22451932](https://pubmed.ncbi.nlm.nih.gov/22451932/)
23. Acharya P, Luongo TS, Louder MK, McKee K, Yang Y, Kwon YD, et al. Structural basis for highly effective HIV-1 neutralization by CD4-mimetic miniproteins revealed by 1.5 Å crystal structure of gp120 and M48U1. *Structure (London, England: 1993)*. 2013 Jun; 21(6):1018–29. doi: [10.1016/j.str.2013.04.015](https://doi.org/10.1016/j.str.2013.04.015)
  24. Jardine J, Julien JP, Menis S, Ota T, Kalyuzhnyi O, McGuire A, et al. Rational HIV immunogen design to target specific germline B cell receptors. *Science (New York, NY)*. 2013 May; 340(6133):711–6. doi: [10.1126/science.1234150](https://doi.org/10.1126/science.1234150)
  25. Morellato-Castillo L, Acharya P, Combes O, Michiels J, Descours A, Ramos OHP, et al. Interfacial cavity filling to optimize CD4-mimetic miniprotein interactions with HIV-1 surface glycoprotein. *Journal of medicinal chemistry*. 2013 Jun; 56(12):5033–47. doi: [10.1021/jm4002988](https://doi.org/10.1021/jm4002988) PMID: [23710622](https://pubmed.ncbi.nlm.nih.gov/23710622/)
  26. Göbel U, Sander C, Schneider R, Valencia A. Correlated mutations and residue contacts in proteins. *Proteins*. 1994 Apr; 18(4):309–17. doi: [10.1002/prot.340180402](https://doi.org/10.1002/prot.340180402) PMID: [8208723](https://pubmed.ncbi.nlm.nih.gov/8208723/)
  27. Casari G, Sander C, Valencia A. A method to predict functional residues in proteins. *Nature Structural Biology*. 1995 Feb; 2(2):171–178. doi: [10.1038/nsb0295-171](https://doi.org/10.1038/nsb0295-171) PMID: [7749921](https://pubmed.ncbi.nlm.nih.gov/7749921/)
  28. Lockless SW, Ranganathan R. Evolutionarily conserved pathways of energetic connectivity in protein families. *Science (New York, NY)*. 1999 Oct; 286(5438):295–9. doi: [10.1126/science.286.5438.295](https://doi.org/10.1126/science.286.5438.295)
  29. Fodor AA, Aldrich RW. Influence of conservation on calculations of amino acid covariance in multiple sequence alignments. *Proteins*. 2004 Aug; 56(2):211–21. doi: [10.1002/prot.20098](https://doi.org/10.1002/prot.20098) PMID: [15211506](https://pubmed.ncbi.nlm.nih.gov/15211506/)
  30. Martin LC, Gloor GB, Dunn SD, Wahl LM. Using information theory to search for co-evolving residues in proteins. *Bioinformatics (Oxford, England)*. 2005 Nov; 21(22):4116–24. doi: [10.1093/bioinformatics/bti671](https://doi.org/10.1093/bioinformatics/bti671)
  31. Dunn SD, Wahl LM, Gloor GB. Mutual information without the influence of phylogeny or entropy dramatically improves residue contact prediction. *Bioinformatics (Oxford, England)*. 2008 Feb; 24(3):333–40. doi: [10.1093/bioinformatics/btm604](https://doi.org/10.1093/bioinformatics/btm604)
  32. de Juan D, Pazos F, Valencia A. Emerging methods in protein co-evolution. *Nature reviews Genetics*. 2013 Apr; 14(4):249–61. doi: [10.1038/nrg3414](https://doi.org/10.1038/nrg3414) PMID: [23458856](https://pubmed.ncbi.nlm.nih.gov/23458856/)
  33. Jones DT, Buchan DWA, Cozzetto D, Pontil M. PSICOV: precise structural contact prediction using sparse inverse covariance estimation on large multiple sequence alignments. *Bioinformatics (Oxford, England)*. 2012 Jan; 28(2):184–90. doi: [10.1093/bioinformatics/btr638](https://doi.org/10.1093/bioinformatics/btr638)
  34. Weigt M, White RA, Szurmant H, Hoch JA, Hwa T. Identification of direct residue contacts in protein-protein interaction by message passing. *Proceedings of the National Academy of Sciences of the United States of America*. 2009 Jan; 106(1):67–72. doi: [10.1073/pnas.0805923106](https://doi.org/10.1073/pnas.0805923106) PMID: [19116270](https://pubmed.ncbi.nlm.nih.gov/19116270/)
  35. Morcos F, Pagnani A, Lunt B, Bertolino A, Marks DS, Sander C, et al. 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*. 2011 Dec; 108(49):E1293–301. doi: [10.1073/pnas.1111471108](https://doi.org/10.1073/pnas.1111471108) PMID: [22106262](https://pubmed.ncbi.nlm.nih.gov/22106262/)
  36. Ekeberg M, Lökvist C, Lan Y, Weigt M, Aurell E. Improved contact prediction in proteins: using pseudolikelihoods to infer Potts models. *Physical review E, Statistical, nonlinear, and soft matter physics*. 2013 Jan; 87(1):012707. doi: [10.1103/PhysRevE.87.012707](https://doi.org/10.1103/PhysRevE.87.012707) PMID: [23410359](https://pubmed.ncbi.nlm.nih.gov/23410359/)
  37. Balakrishnan S, Kamisetty H, Carbonell JG, Lee SI, Langmead CJ. Learning generative models for protein fold families. *Proteins*. 2011 Apr; 79(4):1061–78. doi: [10.1002/prot.22934](https://doi.org/10.1002/prot.22934) PMID: [21268112](https://pubmed.ncbi.nlm.nih.gov/21268112/)
  38. Kamisetty H, Ovchinnikov S, Baker D. Assessing the utility of coevolution-based residue-residue contact predictions in a sequence- and structure-rich era. *Proceedings of the National Academy of Sciences of the United States of America*. 2013 Sep; 110(39):15674–9. doi: [10.1073/pnas.1314045110](https://doi.org/10.1073/pnas.1314045110) PMID: [24009338](https://pubmed.ncbi.nlm.nih.gov/24009338/)
  39. Korber BT, Farber RM, Wolpert DH, Lapedes AS. Covariation of mutations in the V3 loop of human immunodeficiency virus type 1 envelope protein: an information theoretic analysis. *Proceedings of the National Academy of Sciences*. 1993 Aug; 90(15):7176–7180. doi: [10.1073/pnas.90.15.7176](https://doi.org/10.1073/pnas.90.15.7176)
  40. Bickel PJ, Cosman PC, Olshen RA, Spector PC, Rodrigo AG, Mullins JI. Covariability of V3 loop amino acids. *AIDS research and human retroviruses*. 1996 Oct; 12(15):1401–11. doi: [10.1089/aid.1996.12.1401](https://doi.org/10.1089/aid.1996.12.1401) PMID: [8893048](https://pubmed.ncbi.nlm.nih.gov/8893048/)
  41. Gilbert PB, Novitsky V, Essex M. Covariability of selected amino acid positions for HIV type 1 subtypes C and B. *AIDS research and human retroviruses*. 2005 Dec; 21(12):1016–30. doi: [10.1089/aid.2005.21.1016](https://doi.org/10.1089/aid.2005.21.1016) PMID: [16379605](https://pubmed.ncbi.nlm.nih.gov/16379605/)
  42. Travers SAA, Tully DC, McCormack GP, Fares MA. A study of the coevolutionary patterns operating within the env gene of the HIV-1 group M subtypes. *Molecular biology and evolution*. 2007 Dec; 24(12):2787–801. doi: [10.1093/molbev/msm213](https://doi.org/10.1093/molbev/msm213) PMID: [17921487](https://pubmed.ncbi.nlm.nih.gov/17921487/)



43. Garimalla S, Kieber-Emmons T, Pashov AD. The Patterns of Coevolution in Clade B HIV Envelope's N-Glycosylation Sites. *PLOS ONE*. 2015 Jun; 10(6):e0128664. doi: [10.1371/journal.pone.0128664](https://doi.org/10.1371/journal.pone.0128664) PMID: [26110648](https://pubmed.ncbi.nlm.nih.gov/26110648/)
44. Zhao Y, Wang Y, Gao Y, Li G, Huang J. Integrated Analysis of Residue Coevolution and Protein Structures Capture Key Protein Sectors in HIV-1 Proteins. *PLOS ONE*. 2015 Feb; 10(2):e0117506. doi: [10.1371/journal.pone.0117506](https://doi.org/10.1371/journal.pone.0117506) PMID: [25671429](https://pubmed.ncbi.nlm.nih.gov/25671429/)
45. Li G, Theys K, Verheyen J, Pineda-Peña AC, Khouri R, Piampongsant S, et al. A new ensemble coevolution system for detecting HIV-1 protein coevolution. *Biology Direct*. 2015; 10(1):1. doi: [10.1186/s13062-014-0031-8](https://doi.org/10.1186/s13062-014-0031-8) PMID: [25564011](https://pubmed.ncbi.nlm.nih.gov/25564011/)
46. Söding J. Protein homology detection by HMM-HMM comparison. *Bioinformatics (Oxford, England)*. 2005 Apr; 21(7):951–60. doi: [10.1093/bioinformatics/bti125](https://doi.org/10.1093/bioinformatics/bti125)
47. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, et al. The Protein Data Bank. *Nucleic acids research*. 2000 Jan; 28(1):235–42. doi: [10.1093/nar/28.1.235](https://doi.org/10.1093/nar/28.1.235) PMID: [10592235](https://pubmed.ncbi.nlm.nih.gov/10592235/)
48. Pancera M, Zhou T, Druz A, Georgiev IS, Soto C, Gorman J, et al. Structure and immune recognition of trimeric pre-fusion HIV-1 Env. *Nature*. 2014 Oct.; doi: [10.1038/nature13808](https://doi.org/10.1038/nature13808)
49. Julien JP, Cupo A, Sok D, Stanfield RL, Lyumkis D, Deller MC, et al. Crystal structure of a soluble cleaved HIV-1 envelope trimer. *Science (New York, NY)*. 2013 Dec; 342(6165):1477–83. doi: [10.1126/science.1245625](https://doi.org/10.1126/science.1245625)
50. Michel M, Hayat S, Skwark MJ, Sander C, Marks DS, Elofsson A. PconsFold: improved contact predictions improve protein models. *Bioinformatics*. 2014 Sep; 30(17):i482–i488. doi: [10.1093/bioinformatics/btu458](https://doi.org/10.1093/bioinformatics/btu458) PMID: [25161237](https://pubmed.ncbi.nlm.nih.gov/25161237/)
51. Leaver-Fay A, Tyka M, Lewis SM, Lange OF, Thompson J, Jacak R, et al. ROSETTA3: an object-oriented software suite for the simulation and design of macromolecules. *Methods in enzymology*. 2011; 487:545–74. doi: [10.1016/B978-0-12-381270-4.00019-6](https://doi.org/10.1016/B978-0-12-381270-4.00019-6) PMID: [21187238](https://pubmed.ncbi.nlm.nih.gov/21187238/)
52. Schrödinger, LLC. The PyMOL Molecular Graphics System, Version 1.3r1; 2010.
53. Korber B, Gnanakaran S. The implications of patterns in HIV diversity for neutralizing antibody induction and susceptibility. *Current opinion in HIV and AIDS*. 2009 Sep; 4(5):408–17. doi: [10.1097/COH.0b013e32832f129e](https://doi.org/10.1097/COH.0b013e32832f129e) PMID: [20048705](https://pubmed.ncbi.nlm.nih.gov/20048705/)
54. The UniProt Consortium. Activities at the Universal Protein Resource (UniProt). *Nucleic acids research*. 2014 Jan; 42(Database issue):D191–8. doi: [10.1093/nar/gkt1140](https://doi.org/10.1093/nar/gkt1140) PMID: [24253303](https://pubmed.ncbi.nlm.nih.gov/24253303/)
55. Li Y, Yang D, Wang JY, Yao Y, Zhang WZ, Wang LJ, et al. Critical amino acids within the human immunodeficiency virus type 1 envelope glycoprotein V4 N- and C-terminals contribute to virus entry. *PLoS one*. 2014 Jan; 9(1):e86083. doi: [10.1371/journal.pone.0086083](https://doi.org/10.1371/journal.pone.0086083) PMID: [24465884](https://pubmed.ncbi.nlm.nih.gov/24465884/)
56. Munro JB, Gorman J, Ma X, Zhou Z, Arthos J, Burton DR, et al. Conformational dynamics of single HIV-1 envelope trimers on the surface of native virions. *Science*. 2014 Oct.; doi: [10.1126/science.1254426](https://doi.org/10.1126/science.1254426)
57. Davey NE, Satagopam VP, Santiago-Mozos S, Villacorta-Martin C, Bharat TAM, Schneider R, et al. The HIV Mutation Browser: A Resource for Human Immunodeficiency Virus Mutagenesis and Polymorphism Data. *PLoS computational biology*. 2014 Dec; 10(12):e1003951. doi: [10.1371/journal.pcbi.1003951](https://doi.org/10.1371/journal.pcbi.1003951) PMID: [25474213](https://pubmed.ncbi.nlm.nih.gov/25474213/)
58. Lu L, Tong P, Yu X, Pan C, Zou P, Chen YH, et al. HIV-1 variants with a single-point mutation in the gp41 pocket region exhibiting different susceptibility to HIV fusion inhibitors with pocket- or membrane-binding domain. *Biochimica et biophysica acta*. 2012 Dec; 1818(12):2950–7. doi: [10.1016/j.bbamem.2012.07.020](https://doi.org/10.1016/j.bbamem.2012.07.020) PMID: [22867851](https://pubmed.ncbi.nlm.nih.gov/22867851/)
59. Fares MA, Travers SAA. A novel method for detecting intramolecular coevolution: adding a further dimension to selective constraints analyses. *Genetics*. 2006 May; 173(1):9–23. doi: [10.1534/genetics.105.053249](https://doi.org/10.1534/genetics.105.053249) PMID: [16547113](https://pubmed.ncbi.nlm.nih.gov/16547113/)
60. Schnur E, Noah E, Ayzenshtat I, Sargsyan H, Inui T, Ding FX, et al. The conformation and orientation of a 27-residue CCR5 peptide in a ternary complex with HIV-1 gp120 and a CD4-mimic peptide. *Journal of molecular biology*. 2011 Jul; 410(5):778–97. doi: [10.1016/j.jmb.2011.04.023](https://doi.org/10.1016/j.jmb.2011.04.023) PMID: [21763489](https://pubmed.ncbi.nlm.nih.gov/21763489/)
61. Guttman M, Kahn M, Garcia NK, Hu SL, Lee KK. Solution structure, conformational dynamics, and CD4-induced activation in full-length, glycosylated, monomeric HIV gp120. *Journal of virology*. 2012 Aug; 86(16):8750–64. doi: [10.1128/JVI.07224-11](https://doi.org/10.1128/JVI.07224-11) PMID: [22674993](https://pubmed.ncbi.nlm.nih.gov/22674993/)
62. Hu G, Liu J, Taylor KA, Roux KH. Structural comparison of HIV-1 envelope spikes with and without the V1/V2 loop. *Journal of virology*. 2011 Mar; 85(6):2741–50. doi: [10.1128/JVI.01612-10](https://doi.org/10.1128/JVI.01612-10) PMID: [21191026](https://pubmed.ncbi.nlm.nih.gov/21191026/)
63. Liu L, Cimbri R, Lusso P, Berger EA. Intraprotomer masking of third variable loop (V3) epitopes by the first and second variable loops (V1V2) within the native HIV-1 envelope glycoprotein trimer. *Proceedings of the National Academy of Sciences of the United States of America*. 2011 Dec; 108(50):20148–53. doi: [10.1073/pnas.1104840108](https://doi.org/10.1073/pnas.1104840108) PMID: [22128330](https://pubmed.ncbi.nlm.nih.gov/22128330/)

64. Rusert P, Krarup A, Magnus C, Brandenberg OF, Weber J, Ehlert AK, et al. Interaction of the gp120 V1V2 loop with a neighboring gp120 unit shields the HIV envelope trimer against cross-neutralizing antibodies. *The Journal of experimental medicine*. 2011 Jul; 208(7):1419–33. doi: [10.1084/jem.20110196](https://doi.org/10.1084/jem.20110196) PMID: [21646396](https://pubmed.ncbi.nlm.nih.gov/21646396/)
65. Cimbri R, Gallant TR, Dolan MA, Guzzo C, Zhang P, Lin Y, et al. Tyrosine sulfation in the second variable loop (V2) of HIV-1 gp120 stabilizes V2–V3 interaction and modulates neutralization sensitivity. *Proceedings of the National Academy of Sciences of the United States of America*. 2014 Feb; 111(8):3152–7. doi: [10.1073/pnas.1314718111](https://doi.org/10.1073/pnas.1314718111) PMID: [24569807](https://pubmed.ncbi.nlm.nih.gov/24569807/)
66. Wang Y, Rawi R, Hoffmann D, Sun B, Yang R. Inference of global HIV-1 sequence patterns and preliminary feature analysis. *Virologica Sinica*. 2013 Aug; 28(4):228–38. doi: [10.1007/s12250-013-3348-z](https://doi.org/10.1007/s12250-013-3348-z) PMID: [23913180](https://pubmed.ncbi.nlm.nih.gov/23913180/)
67. Wang Y, Rawi R, Wilms C, Heider D, Yang R, Hoffmann D. A small set of succinct signature patterns distinguishes Chinese and non-Chinese HIV-1 genomes. *PloS one*. 2013 Jan; 8(3):e58804. doi: [10.1371/journal.pone.0058804](https://doi.org/10.1371/journal.pone.0058804) PMID: [23527028](https://pubmed.ncbi.nlm.nih.gov/23527028/)
68. Hollier MJ, Dimmock NJ. The C-terminal tail of the gp41 transmembrane envelope glycoprotein of HIV-1 clades A, B, C, and D may exist in two conformations: an analysis of sequence, structure, and function. *Virology*. 2005 Jul; 337(2):284–96. doi: [10.1016/j.virol.2005.04.015](https://doi.org/10.1016/j.virol.2005.04.015) PMID: [15913700](https://pubmed.ncbi.nlm.nih.gov/15913700/)
69. Postler TS, Martinez-Navio JM, Yuste E, Desrosiers RC. Evidence against extracellular exposure of a highly immunogenic region in the C-terminal domain of the simian immunodeficiency virus gp41 transmembrane protein. *Journal of virology*. 2012 Jan; 86(2):1145–57. doi: [10.1128/JVI.06463-11](https://doi.org/10.1128/JVI.06463-11) PMID: [22072749](https://pubmed.ncbi.nlm.nih.gov/22072749/)