




Prediction and validation of HIV-1 gp41 ecto-transmembrane domain post-fusion trimeric structure using molecular modeling

Biswajit Gorai, Satyabrata Das & Prabal K. Maiti

To cite this article: Biswajit Gorai, Satyabrata Das & Prabal K. Maiti (2019): Prediction and validation of HIV-1 gp41 ecto-transmembrane domain post-fusion trimeric structure using molecular modeling, Journal of Biomolecular Structure and Dynamics, DOI: [10.1080/07391102.2019.1635916](https://doi.org/10.1080/07391102.2019.1635916)

To link to this article: <https://doi.org/10.1080/07391102.2019.1635916>

 View supplementary material [↗](#)

 Accepted author version posted online: 24 Jun 2019.
Published online: 02 Jul 2019.

 Submit your article to this journal [↗](#)

 Article views: 9

 View Crossmark data [↗](#)



Prediction and validation of HIV-1 gp41 ecto-transmembrane domain post-fusion trimeric structure using molecular modeling

Biswajit Gorai, Satyabrata Das and Prabal K. Maiti

Department of Physics, Indian Institute of Science, Bangalore, Karnataka, India

Communicated by Ramaswamy H. Sarma

ABSTRACT

The glycoproteins on the surface of human immunodeficiency virus (HIV) undergoes cascade of conformational transitions to evade the human immune system. The virus replicates inside the host and infects the T-cells instigating acquired immunodeficiency syndrome (AIDS). The glycoprotein 41 (gp41) of HIV helps to mediate the fusion of virus and host membranes. The detailed mechanism of host cell invasion by virus remains obscure due to the unavailability of experimental structure of complete gp41. In the current study, the post-fusion (PoF) trimeric structure of ecto-domain including transmembrane domain of gp41 was modeled using multiple homologous templates of Simian immunodeficiency virus (SIV) and HIV-1. In order to validate the gp41 model, interactions of three peptide inhibitors: T20, C37 and C34; were studied using all-atom molecular dynamics (MD) simulations, binding free-energy calculation and per-residue energy decomposition analysis. The binding free energy calculated using MM-PBSA (Molecular Mechanics Poisson-Boltzmann surface area) method predicts maximum affinity for C34 and minimum by T20 for gp41, which is in good agreement with the available computational and experimental studies. The van der Waals interaction is a dominant contributor for the peptide-gp41 complexes. The per-residue decomposition of energy confirmed the role of Trp117, Trp120 and Ile124, present in C34 and C37, for the strong hydrophobic interactions with the deep pocket localized around the N-terminal of gp41, which is lacking in T20. The HIV-1 gp41 structure developed in this work can be used in future study to gain insight into the mechanism of virus invasion and probing potent inhibitor to eliminate AIDS.

ARTICLE HISTORY

Received 2 December 2018
Accepted 19 June 2019

KEYWORDS

Virus; protein modeling; peptide inhibitor; binding free energy; MM-PBSA

1. Introduction

HIV type 1 (HIV-1) is an enveloped retrovirus (Barré-Sinoussi, Ross, & Delfraissy, 2013; Goto, Nakai, & Ikuta, 1998) which instigate in human *via* cross-species (zoonotic) transmission of non-human primate lentiviruses (Gao et al., 1999; Hemelaar, 2012; Sharp & Hahn, 2011; Smyth, Davenport, & Mak, 2012). Infected individuals in the long run develop an inexorably fatal disease, AIDS (Maartens, Celum, & Lewin, 2014) in which progressive failure of immune system leads to clinical collapse. AIDS remains a major global health problem. Phenomenal success has been achieved by extensive research in last few decades to control the HIV-1 infection (Cohen et al., 2011; DeGruttola, Smith, Little, & Miller, 2010; Esté & Cihlar, 2010). A combinatorial therapy based on three drugs known as Highly Active Anti-Retroviral Therapy, (HAART) is proved to be highly effective (Arts & Hazuda, 2012) and is known to increase the life expectancy of infected individuals, and reduce the risk of transmission. However complete cure from HIV-1 infection remains a major challenge (Archin, Sung, Garrido, Soriano-Sarabia, & Margolis, 2014; Jessen, Allen, & Streeck, 2014). Error prone reverse transcriptase causes high rate of genomic mutation in HIV which often leads to drug resistant virions (Ruelas & Greene,

2013). The design of effective anti-HIV vaccine proved to be an exorbitantly difficult task (Fauci & Marston, 2015; Mascola, 2015) and a foremost priority to halt the global HIV-1 pandemic (Kallings, 2008; Sharp & Hahn, 2010, 2011). Efforts are underway to search and develop a more effective drug against new targets of HIV (Arhel & Kirchhoff, 2010; Arts & Hazuda, 2012; Flexner, 2007; Menéndez-Arias, 2013).

HIV-1 virus consists of approximately 15 proteins (Briggs & Kräusslich, 2011; Kuzembayeva, Dilley, Sardo, & Hu, 2014; Russell, Liang, & Wainberg, 2004; Turner & Summers, 1999) and two copies of genomic RNA (Lu, Heng, & Summers, 2011; Paillart, Shehu-Xhilaga, Marquet, & Mak, 2004) which remains encased within a cone-shaped capsid core (Schur et al., 2015; Zhao et al., 2013). The schematic of HIV-1 virus structure with major components is shown in Figure 1. HIV-1 capsid is surrounded by matrix protein p17 and a lipid bilayer (Brügger et al., 2006) studded with the virus encoded envelope glycoproteins (Env), gp120 and gp41 (Checkley, Luttge, & Freed, 2011; Merk & Subramaniam, 2013; Munro & Mothes, 2015; Sundquist & Kräusslich, 2012). In the first step of HIV entry, surface glycoprotein gp120 bind to human CD4⁺ T-cell surface receptor protein (Kwong et al., 1998). This specific and tight binding triggers series of

conformational changes in gp120 (Guttman et al., 2012; Kwon et al., 2012; Pancera et al., 2014) essential for its further binding to the co-receptor CCR5 or CXCR4 (Berger, Murphy, & Farber, 1999; Flanagan, 2014). At this stage, the hydrophobic N-terminal region of gp41 become exposed and tether to the host cell membrane. Gp41 takes charge of membrane fusion upon dissociation from the gp120 (Cai, Gochin, & Liu, 2011; Harrison, 2005, 2008, 2015; Melikyan, 2008).

The membranotropic glycoprotein gp41 consists of about 345 aa (amino acids) and popularly known as the fusion engine (Cai et al., 2011). It has 4–8 glycosylation sites in the N-terminal region and the N-linked glycans are proved to be essential for viral infectivity (Mathys & Balzarini, 2015; Perrin, Fenouillet, & Jones, 1998). The transmembrane protein, gp41 possess three plausible regions: N-terminal ecto-domain (~172 residues); a helical transmembrane domain, TMD (22 residues) and a long cytoplasmic tail, CT (~151 residues). N-terminal ecto-domain and TMD domain plays an active role in membrane fusion and undergoes large conformational change during different stages of the fusion process. Different segments of ecto-domain are named based on their functional role as shown in Figure 2: a 16 residue N-terminal glycine-rich fusion peptide (FP) followed by FP proximal region (FPPR), long N-terminal helical region with heptad repeat (NHR or HR1), immunodominant loop (IL) region with conserved disulfide bond, C-terminal helical region with heptad repeat (CHR or HR2), aromatic rich membrane proximal

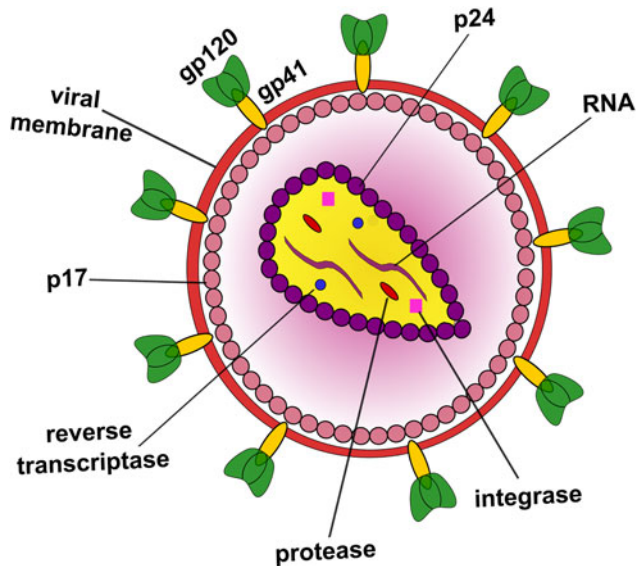


Figure 1. Schematic diagram of the mature HIV-1 virion. The two RNA strands along with protease, integrase, and reverse transcriptase is encased inside a cone-shaped capsid composed of p24 proteins. The core capsid is subsequently surrounded by p17 matrix and outer membrane. The outer surface of virion contains gp120 (green) and gp41 (yellow) glycoproteins.

extra-cellular region (MPER), TMD connecting CT and MPER of gp41. Fusion domain and MPER are highly membranotropic, part of the NHR helix and IL are also partially membranotropic (Moreno, Giudici, & Villalaín, 2006; Moreno, Pascual, & Villalaín, 2004). The notable exception is the CHR domain, which is rich in negatively charged residues, lacks membranotropic property (Moreno et al., 2006). Studies on sequence-to-structure-to-function correlations of each of the functional domains of HIV-1 gp41 fetch substantial understanding, however, even more, need to be explored.

At different stages of HIV entry to host cell, gp41 undergoes large conformational changes. Three distinct conformations of gp41 are native gp120 bound state, pre-hairpin fusion (PrF) intermediate and thermodynamically stable post-fusion (PoF) six-helix bundles (6HB) or 'trimer-of-hairpins'. HIV-gp41 anchor the host cell by inserting FP into the host cell membrane at an early stage of the fusion process, probably immediately after its dissociation from gp120. At this stage, gp41 adopts the extended PrF conformation between virus and host cell. FP's are inserted into the host cell membrane, while the NHR and CHR helices remain separated at the initial stage of viral invasion. This extended conformation subsequently undergoes large conformational transitions to form stable 6HB or PoF conformer, which leads to membrane apposition followed by fusion. It is generally believed that trimer-of-hairpin conformer is adapted by the gp41 at the end of the fusion process. During the formation of 6HB, negatively charged CHR helix packed with the positively charged NHR domain in anti-parallel fashion, subsequently positioning the FP/FPPR regions in close proximity to MPER/TMD of gp41.

The formation of native 'trimer-of-hairpins' structure is extremely important for successful viral entry. In general, about 14 Env spikes are present on the virion surface and are the principal target available on the HIV surface for vaccine design. Rapid sequence variations of the surface residues and structural dynamism are responsible for its successful immune evasion. Any kind of perturbation either by sequence mutation or inhibitor which inhibits the formation of native trimer-of-hairpins structure hinders the entry of virus to host cell. As a result, the design of inhibitors targeted to coiled-coil NHR trimer or CHR helix or trimer-of-hairpins is an active area of research. It has been observed that several long stretches of the peptide, partly or fully, belonging to the NHR and CHR regions act as an effective inhibitor. One of such peptide inhibitor is T20 (also known as DP178, Enfuvirtide or Fuzeon) which is the first fusion inhibitor peptide approved by FDA and used to arrest the HIV-1 replication for the treatment of HIV-1 infected patients. It is a 36 aa peptide (residues 127–162 of gp41) of which 26 aa overlap with the C-terminal regions of CHR helix and remaining 10 aa belongs to the sequence of MPER region. The mechanism



Figure 2. Bar diagram depicting the functional domains of the HIV-1 gp41. Commonly used name and position of the sub-domains (represented with distinct colors) are given at the top and bottom of the bar, respectively.

of T20 inhibition is yet to understand fully. Most probably the primary target of this drug is the NHR trimer. Moreover, the experimental results suggest the ability of T20 to hinder the interaction between CXCR4 co-receptor with the gp120-CD4 complex during the intermediate stage of virus entry (Yuan, Craig, Si, Farzan, & Sodroski, 2004). However, HIV has the ability to successfully develop mutated virion resistant to T20. As a result, several attempts have been made to develop next generation peptide inhibitors to understand the exact mechanism and inhibit the activity of HIV. C34 (a 34 residue peptide derived from sequence 117–150 of gp41) and C37 (a 37 residue peptide derived from sequence 114–150 of gp41) are peptide inhibitors also derived from CHR region of gp41 hinder the six-helix bundle formation effectively by docking to a hydrophobic pocket on N36 (residues S35–L70) of NHR domain (Allen, Yi, Gochin, Jacobs, & Rizzo, 2015; Chan, Chutkowski, & Kim, 1998; Champagne, Shishido, & Root, 2009).

The structure-function correlation of all the functional regions of gp41 has been studied separately using various techniques such as X-ray crystallography, NMR, and IR. However, the structure of full ecto-domain of HIV-1 is still not available. It is a sheer challenge to obtain a stable structure of the HIV-1 ecto-domain at physiological pH. The labile Gly/Ala rich hydrophobic FP region, intrinsic flexibility of the IL loop with several hydrophobic and tryptophan residues as well as the high flexibility of membrane loving aromatic rich MPER regions make the full ecto-domain construct highly non-favorable for the structural studies through experimental techniques. Here, we have made an attempt to build a 3D model of the full HIV-1 gp41 ecto-domain along with the TMD using homology modeling method. The mode of binding and affinity of CHR based peptide inhibitors viz. T20, C34 and C37 are studied using MM-PBSA method (Genheden & Ryde, 2015; Gilson & Zhou, 2007; Reddy et al., 2014; Wang, Greene, Xiao, Qi, & Luo, 2018) to validate the modeled structure. The gp41 model can be used in future studies to understand the HIV invasion mechanism to the human cell with molecular details using classical MD simulations. Such studies are currently underway in our laboratory.

2. Materials and methods

2.1. Homology model of gp41 using multiple templates

The gp41 amino-acid sequence of HIV-1 was derived from full envelope glycoprotein gp160 isolated from HXB2 deposited at UniProtKB (ID: P04578). Modeller version 9.14 (Fiser & Sali, 2003; Sali & Blundell, 1993) has been employed to generate the homology model of 194 residues long HIV-1 ecto-TMD domain of gp41 (512–705 with respect to gp160 sequence) trimer. Multiple HIV-1 templates were collected from the Protein Data Bank (PDB) archive, namely, 1AIK, 1ENV, 2ARI, 2LP7, 2M7W, 2MG1, 2X7R, 3UIA and 3VGX; to generate the full model of HIV-1 gp41 trimer. Trimeric truncated-gp41 (NHR-IL-CHR) structures (PDB ID: 2EZO, 1QCE and 1QBZ; having sequence identity around 49% with HIV-1 gp41) from SIV were also used to model the IL domain of the model. Details of the templates used to model distinct

domains of HIV-1 gp41 trimer in the current study are available in [Supplementary Table S1](#). Altogether 582 residues were modeled to obtain the HIV-1 gp41 trimer. Three fold symmetry restraints were applied to C-alpha atoms of the homology model to maintain the trimeric structure. Disulfide bond (C87–C93) in the IL-loop region was maintained. The secondary structure of the TMD region is also restrained to the alpha helix. We choose the best model among 200 Modeller generated models for further simulation studies based on the low value of Modeller objective function (molpdf), discrete optimized protein energy (DOPE) score and GA341 score (Shen & Sali, 2006). It observed that due to near identical sequence identity, the GA341 score remains 1.0 for most of the successful models. Therefore, the selection of the best gp41 model is based on the low value of molpdf and high negative value of DOPE score.

2.2. Gp41 sequence analysis

Close homologous full length HIV-1 gp41 sequences were analyzed using the ConSurf (Glaser et al., 2003) web server. Homologs of HXB2 gp41 (P04578) were searched against UniRef-90 database using CSI-BLAST search algorithm (three iteration and *E* value cut-off 0.0001) and all the 497 sequences were subjected to multiple sequence alignment by MAFFT algorithm. Few sequences with rare insertion were manually deleted and rest 484 sequences were realigned and N-terminal region in all the sequences are made continuous manually. The rate of evolution at each site was calculated using the Bayesian method. Also, the graphical summary of MSA is generated by WebLogo server (Crooks, Hon, Chandonia, & Brenner, 2004).

2.3. Molecular dynamics simulations of trimeric PrF and PoF gp41 conformers and free-energy landscape (FEL) analysis

Fully stretched hypothetical PrF conformer of gp41 was also derived from the modeled PoF structure after flipping outward the C-terminal fragment from the center of flexible loop region (T95) maintaining the three fold symmetry using Coot (Emsley & Cowtan, 2004; Emsley, Lohkamp, Scott, & Cowtan, 2010). Molecular dynamics simulations of both conformers of HIV-1 gp41, PrF and PoF conformations, were performed using GROMACS version 5.1.4 (Berendsen, van der Spoel, & van Drunen, 1995) simulation package in Cray XC40 machine at SERC, IISc, Bangalore, India using Amber ff99sb-ildn force-field (Lindorff-Larsen et al., 2010). Initially, the homology built protein structures were placed at the center of cubic boxes maintaining a minimum distance of 2 nm from the wall. The periodic boxes with protein were solvated with TIP3P water and systems were neutralized with the requisite number of Na⁺ and Cl⁻ ions to maintain the overall salt

Table 1. Details of the system size (atom number of each constituent) of both the conformers.

Sl. No.	Gp41 conformer	Protein	Water	Ions	Total atoms
1	PrF	9399	2,621,658	3257	2,634,314
2	PoF	9399	435,339	555	445,293

concentration of 0.15 M. The PoF system contains 145,113, 279 and 276 number of water, Na^+ and Cl^- ions, respectively, with a total system size of 445,293 atoms. The PrF system contains 873,886, 1630 and 1627 number of water, Na^+ and Cl^- ions, respectively, with a total system size of 2,634,314 atoms. The detailed components of the simulated systems are given in Table 1. The charge neutralized systems were subjected to energy minimization using 3000 steps of steepest descent (Vrahatis, Androulakis, Lambrinos, & Magoulas, 2000) algorithm followed by 3000 steps of conjugate-gradient algorithm imposing a restraining force of $2.0 \text{ kcal/mol/\AA}^2$ on the heavy atoms of the proteins to remove steric clashes and bad geometry. The minimized systems were gradually heated from 0 K to 310 K within 1 ns in NVT ensemble in a V-rescale thermostat with a collision frequency of 2.0 ps^{-1} followed by equilibration for 2 ns at 1 atm pressure under NPT ensemble using Parrinello-Rahman barostat. Bonds involving hydrogen atoms were constrained using SHAKE algorithm (Ryckaert, Ciccotti, & Berendsen, 1977) to achieve the integration time-step of 2 fs. Long-range electrostatic interactions were calculated using Particle Mesh Ewald (PME) algorithm (Darden, York, & Pedersen, 1993). After equilibration, solute restraints were removed and production runs were performed for 100 ns for both the PoF and PrF systems using NPT ensemble. Coulomb interactions and van der Waals interactions were truncated at 10 \AA and the non-bonded pair lists were updated every 10 cycles. Trajectories were saved every 10 ps for further analysis. The FEL for PrF and PoF were calculated from their last 80 ns trajectories to understand the conformational stabilities during the unrestrained simulation. The FEL at 310 K was obtained by projecting energy landscapes onto the three-dimensional space defined by radius of gyration (R_g), root mean square deviation (RMSD) and Gibbs free energy (kcal/mol) generated using the GROMACS utility 'gmx sham'. The third variable, free energy, which was estimated from populations (probability distributions) of the system with respect to R_g and RMSD values.

2.4. Calculation of relative binding free energy of gp41 trimer with peptide inhibitors

The gp41 truncated NHR trimer, NT89 (1–89 residues of all chains) and three peptide inhibitors: T20 (127–162 residues), C34 (117–150 residues) and C37 (114–150 residues) were derived from the homology built gp41 model predicted in the current study. In order to obtain the complex of gp41 with peptide inhibitors, we have considered the PoF conformer of gp41 and truncated the structure keeping NT89 residues of three chains and residues belonging to respective peptide inhibitors of a single chain from the complete trimeric structure. Specifically, to obtain the N89–T20 complex, the residues 1–89 of three chains and residues 127–162 from a single chain were retained and rest residues were removed from the full PoF conformer. Three systems were built with NT89 complexed with either T20, C34 or C37 peptide inhibitors and simulated at physiological conditions. The N-terminal and C-terminal of peptides were capped with acetyl (ACE) and amide

(NME) groups, respectively, to stabilize the native helical structure of short peptides (Higo et al., 2001). Initially, the complexes were placed at the center of cubic boxes and were solvated using TIP3P water molecules. The requisite number of Na^+ ions were added to neutralize the system. The above systems were minimized, equilibrated and simulated following similar protocol as mentioned in the previous section, except the heavy atoms of FP domain (residues 1–16) and truncated loop (residues 82–89) of NT89 were position restrained during 100 ns production run, to basically imitate the PrF conformation of gp41 during HIV invasion.

The relative binding free energies between NT89 (truncated gp41) modeled structure and peptide inhibitors (T20, C34 and C37) were estimated using MM-PBSA module of AMBER (Miller et al., 2012). Thousand frames in the interval 50 ps were extracted for MM-PBSA calculations from the last 50 ns of 100 ns production run (refer the previous step). G_{polar} calculations were performed at 0.15 M salt concentration with atomic radii of 0.95 \AA and 1.81 \AA for sodium and chloride ions, respectively. The dielectric constants were fixed to 1, 2 and 80 for vacuum, solute and solvent, respectively. The Adaptive Poisson-Boltzmann Solver (APBS) (Baker, Sept, Joseph, Holst, & McCammon, 2001) was used to solve the equations of continuum electrostatics with the probe radius of 1.4 \AA . We have used single trajectory approach and the conformational entropy change of complex was ignored during the calculation. The binding energy contribution of each residue of the peptide inhibitors were further decomposed to identify the residues crucial for the interaction with gp41.

3. Results and discussion

3.1. Homology modeling and sequence analysis

The homology model of HIV-1 gp41 ecto-TMD domain (1–194 residues) in 'trimer-of-hairpins' form was generated by the multi-template approach using Modeller v9.14. The extended PrF and six-helix bundled PoF conformers of trimer of HIV-1 gp41 are shown in Figure 3. HXB2 gp41 sequence was used for model building and templates used either belong to SIV or HIV-1. The gp41 sequence of SIV structures was aligned continuously with the corresponding region of HIV-1 except for four residue deletions in the immunodominant loop region. The conserved disulfide bond of gp41 at IL-loop region was maintained during model building.

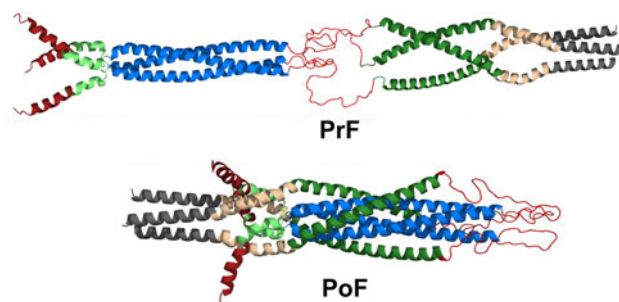


Figure 3. The models generated for HIV-1 gp41 trimer pre-fusion (PrF) and post-fusion (PoF) intermediates are depicted in the figure. Functional domains of gp41 are represented in different colors (for color code of functional regions please refer Figure 2).

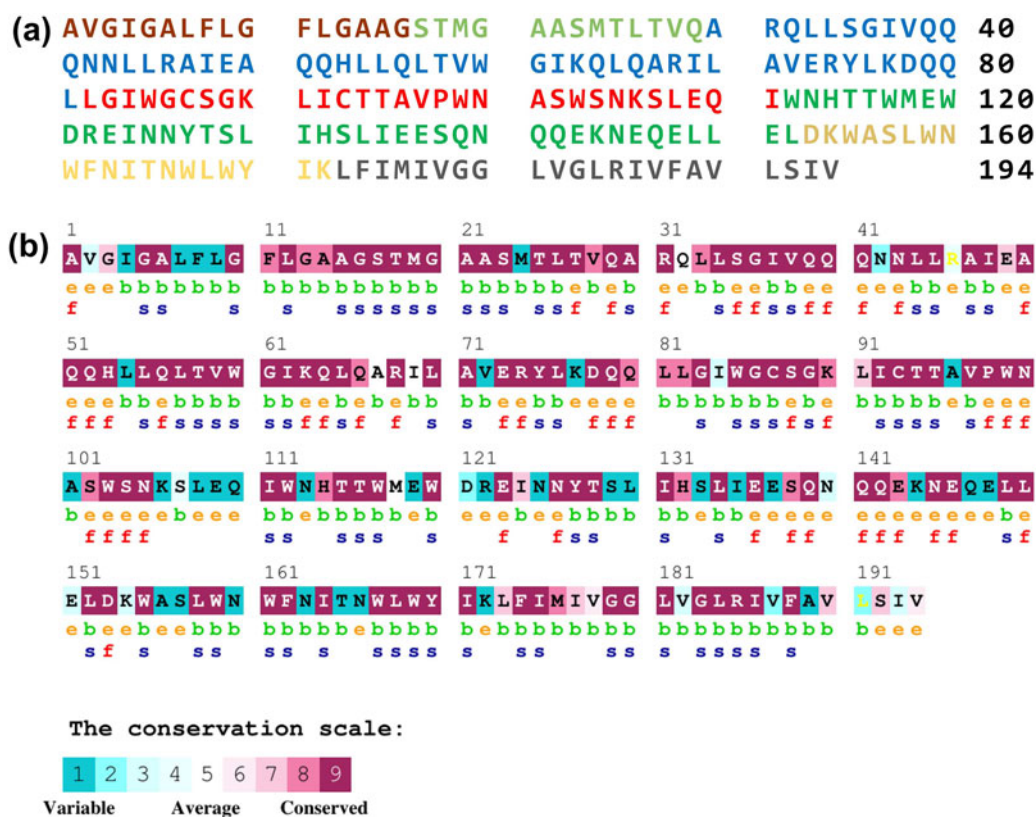


Figure 4. (a) Protein sequence of HIV-1 (HXB2) gp41 ecto-TMD domain (1-194) colored based on the location of functional domains. (b) Evolutionary conservation pattern of HIV-1 gp41 sequences analyzed by ConSurf server using 484 homologs. The color coding of the conservation scale, from turquoise to maroon represents variable to conserved amino-acids at the particular position. The characters 'b' and 'e' listed in the first row below the sequence infers buried and exposed status of the residues, respectively. The 's' and 'f' characters in the second row refers buried and exposed status of structurally and functionally conserved residues, respectively.

Incidentally, these conserved cysteine residues were mutated to alanine in SIV structure. It has been experimentally demonstrated that the FP region is capable of adopting both helix and β -strand secondary structure, which might vary with the experimental conditions used for structure determination (Lai & Freed, 2014). In the present study, only the helical conformation of FP is considered during gp41 modeling. The native helical geometry has been also imposed for TMD domain (172–194 residues). The connection of C-terminal MPER and N-terminal TMD constitute a continuous helical region which is in full consonance with the reported experimental result (Apellániz et al., 2015). The conserved Arg185 is situated almost at the middle of the TMD helix and their side-chains are oriented towards the trimeric helical core i.e. away from the hydrophobic lipid layer. The inter hydrogen bonds within the Arg185 from three chains provide stability to the trimeric TM domain of gp41. A molecular dynamics simulation study of HIV-1 gp41 TMD helix trimer shows the similar orientation of this conserved arginine (Kim, Hartley, Curran, & Engelman, 2009). As expected the modeled region of gp41 is mostly helical in nature. We preferred the model with N-terminal FP helices protruding out of the 'hairpin-of-trimer' structure which is probably essential for its interaction with the host cell membrane. Experimental evidence supports the interaction of FP helix with that of the TMD in post-fusion 'trimer-of-hairpins' conformation. In several graphical depictions, the FP helix is shown to be projected outside from the trimeric axis (Tamm, Lee, & Liang, 2014).

The stereo-chemical quality of the modeled structure was evaluated by means of the Ramachandran plot (Ramachandran, Ramakrishnan, & Sasisekharan, 1963). The distribution of dihedral angles (Φ and Ψ) for the non-proline and non-glycine residues of the modeled structure were found to be either in the most favored or allowed regions. The superposition of 3D structure of modeled gp41 to the principal templates *viz.* 1ENV, 1QBZ, 1QCE and 2EZ0 showed RMSD of 0.48 Å, 1.25 Å, 1.11 Å and 2.07 Å, respectively. ConSurf analysis of homologous gp41 sequence (Figure 4) implies that the FP region is constituted mostly by conserved glycine (four), small aliphatic hydrophobic residues (Ala, Leu) as well as phenylalanine residues. FPPR region also contains small aliphatic hydrophobic residues, a conserved methionine residue and have a prevalence for small polar residues (five positions are occupied in FPPR). There is a helix break between FP and NHR domain. Overall FPPR region shows a high degree of sequence conservation with small polar residues like S or T. The FP is mostly constituted by hydrophobic residues (A, G, L, F), which are crucial for its membrane fusion activity. We found that NHR helix is the most conserved region of gp41 ecto-domain. It has 22 hydrophobic residues and eleven conserved glutamine residues with more than 97% occurrence, it also contains one conserved asparagine residue. Six positions have conserved positively charged residues (R/K) with >90% occurrence and three positions with conserved negatively charged residues (D/E). There are only one conserved tryptophan and tyrosine residue. NHR

domain has two conserved glycine and one conserved serine, threonine and histidine residues each. The only disulfide bond, between C87 and C93, in HIV-1 gp41 is found in the IL domain. The residues within the disulfide bonds are mainly hydrophobic however a conserved positively charged residue (lysine or arginine) is present at the center which is known to contribute in the membrane fusion mechanism of the virus (Ashkenazi, Merklinger, & Shai, 2012; Ashkenazi, Viard, Wexler-Cohen, Blumenthal, & Shai, 2011; Pascual, Moreno, & Villalaín, 2005; Peisajovich, Blank, Eband, Eband, & Shai, 2003). Sequence conservation is comparatively lesser in CHR domain than that of the NHR. Based on the alignment profile, it has 18 residues which are conserved more than 94%. There are single methionine, tryptophan, and tyrosine conserved residues constituting CHR domain. Negatively charged residues are slightly more prevalent in CHR helix than that of NHR, the opposite is true for the positively charged residues (R/K). The CHR domain is mainly composed of polar uncharged residues (S, T, N, Q), which may be the reason for its less membranotropic activity (Moreno et al., 2006). The MPER domain of gp41 plays a crucial role in fusion and is rich in conserved tryptophan residues (Muñoz-Barroso, Salzwedel, Hunter, & Blumenthal, 1999; Salzwedel, West, & Hunter, 1999). The 17 residues long stretch of hydrophobic and uncharged amino-acids (W155–I171) have a decisive role in the membrane fusion. The N-terminal region (L173–I186) of TMD is well conserved with overall >90% occurrence. This region contains a conserved GXXXG motif and R185 which is known to contribute to membrane fusion and viral infectivity (Gangupomu & Abrams, 2010). The buried and exposed residues predicted by ConSurf is based solely on the sequence

and leads to certain anomalies. It can be noticed that the conserved buried residues (G10–N100) are concealed by ecto C-terminal exposed segment (A101–K172) of PoF six-helix bundle conformer. In summary, the C-terminal of FP, FPPR, NHR, and N-terminal residues of IL are relatively preserved than that of C-terminal of IL, CHR, MPER and TMD residues within homologous of gp41.

3.2. Comparison of structural stability of PoF and PrF of gp41

To elucidate the relative stability of gp41 trimer of PrF and PoF conformations, MD simulations at near physiological conditions were performed for 100 ns. The RMSD, root mean square fluctuation (RMSF), secondary structure, and FEL were evaluated from the MD trajectory saved at an interval of every 10 ps. The PoF conformer remained close to its native-like structure throughout the simulation with backbone RMSD fluctuating between 7 and 9 Å (Figure 5a). However, the PrF structure is unable to maintain a stable structure and the RMSD ranges from 30 to 50 Å during the 100 ns long unrestrained MD simulation. Significantly low RMSD of PoF conformer highlights its structural rigidity over the PrF conformer. RMSF of the backbone atoms of individual residues of PoF and PrF gp41 trimers have been depicted in Figure 5b. The plot clearly indicates the structural rigidity of PoF over PrF. The RMSF plot implies higher flexibility of residues at the terminals and IL domain of both the conformers. The flexibility of PrF conformer is relatively higher than its PoF conformer which further accounts for the higher RMSD of PrF conformer. The FP (N-terminal) and TMD (C-terminal) of

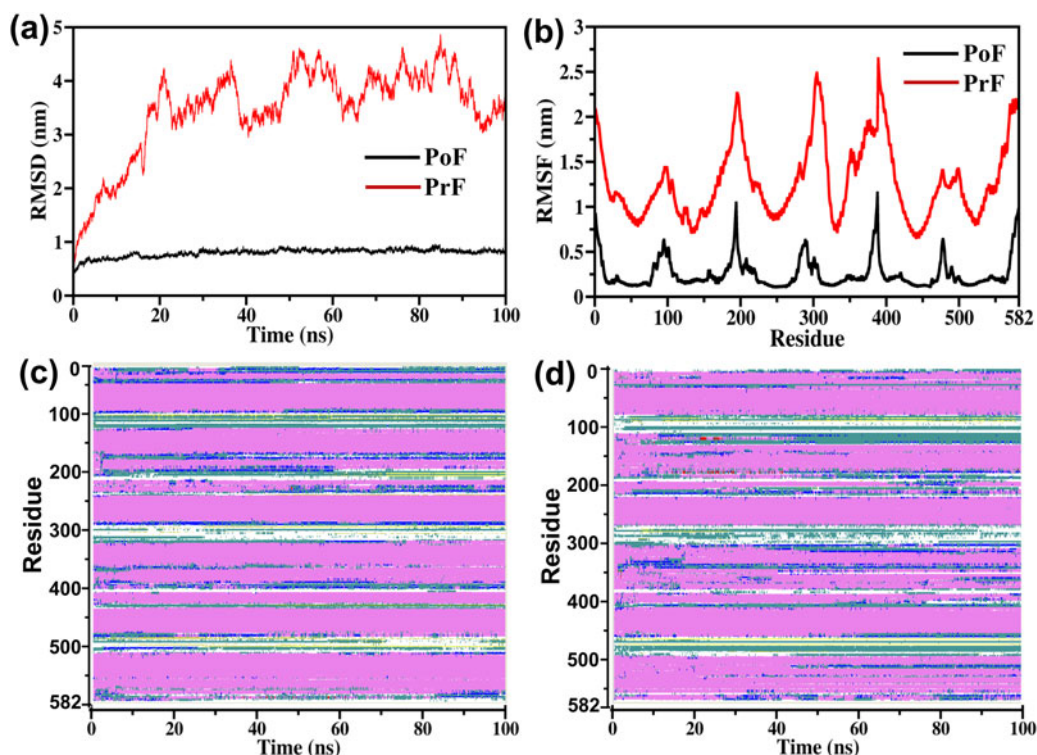


Figure 5. RMSD, RMSF and secondary structure content plots. (a) Time evolution of RMSD and (b) RMSF of backbone atoms for PoF (red line) and PrF (black line). Secondary structural contents of (c) PoF and (d) PrF conformers as estimated from their MD trajectories stored at every 10 ps are shown as a function of the simulation time. Secondary structural segments such as strands, turns, 3_{10} -helices, and random coils are represented by yellow, green, blue and white colors, respectively.

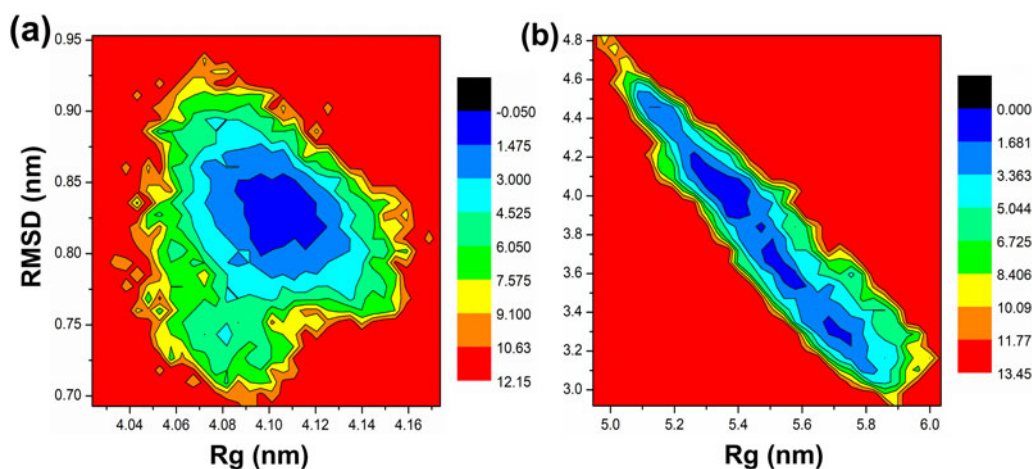


Figure 6. Contour plots representing the 2D free-energy landscape (FEL) of (a) PoF conformer and (b) PrF conformer. RMSD (in nm) and R_g (in nm) are selected as reaction coordinates to obtain the relative Gibbs free-energy change (in kJ/mol) at 300 K. The magnitude of the free energy is represented by different colors, as quantified by the color scale on the right of the respective figures.

both the conformers are mostly fragile probably due to likeness to get stabilized in a membrane environment. The local dynamics of N-terminal, IL, and C-terminal of PoF trimer are relatively low implying that this may be the stable conformation adopted by gp41. We have also calculated the angle between the C-alpha atoms of first residue (A1), an IL residue (L108) and last residue (V194) with respect to simulation time for PrF (Supplementary Figure S2). The PrF structure undergoes drastic conformational change during the simulation and the angle drops gradually from 180° to 79° at 85 ns. The bending of gp41 trimer during simulation signifies its inherent tendency to acquire the PoF conformer. We observed that the interactions between the IL domain residues of first and third chain of gp41 trimer persist throughout simulation whereas the second chain unwinded marginally to loose affinity from the rest two chains. The inter chain hydrogen bonds (HB) crucial for the interactions between IL domain of first and third chains of gp41 are identified within I84–L496, C87–I480, T95–E497, N100–N501, N100–W500, W112–N493 and N113–S492. The unwinding of the second chain also results in higher RMSF of the IL region residues as shown in Figure 5b. Probably, the dynamics of labile second chain and cooperative HB interactions between the first chain and third chain at IL induce the bending of PrF conformer of gp41. Further, the secondary structure analysis was performed using the Timeline plugin of VMD for both the PrF and PoF (Figure 5c,d, respectively) conformers. The secondary structure profile of each residue during the simulation for both the conformers seems to be very similar. Gp41 is mainly α -helix protein with fleeting turns and random coils around the IL region. Transient β -sheet structures are also observed in the IL region for both PrF and PoF conformers during explicit solvent MD simulation. The non-native strands observed during MD simulation is due to the formation of intermittent intra-HB at the highly flexible IL region.

To gain further insight on the conformational changes of both PoF and PrF conformers, we have plotted the 2D FEL as a function of R_g and RMSD in Figure 6. As shown in Figure 6, PrF structure traverses several transient conformational states

during the simulation trajectory. PrF conformer also lacks structural convergence and remains unstable throughout the 100 ns simulation. In contrast, the FEL of PoF structure is localized within the RMSD range of 7.1–9.2 Å and R_g of 40.4–41.5 Å. The minimum-energy conformer of PoF has RMSD and R_g value of around 8.4 Å and 41.1 Å, respectively. The energy value of PoF conformer is marginally lower than PrF conformer probably due to reaching a stable low energy structure during MD simulation. The free-energy profile of the PoF revealed a protein fold with a stable global minimum. In contrast, for PrF we observe three distinct basins of similar free-energy minima distributed over broader ranges of RMSD and R_g , representing multiple transient conformations of PrF. The RMSD, RMSF, bending angle and FEL analyses elucidated that PoF is the most rigid and ordered conformer preferred by gp41 than PrF during fusion pathway. We note that a quantitative comparison of the stabilities of PrF and PoF requires accurate free-energy estimations by using advanced sampling methods, such as Umbrella Sampling or Metadynamics. However, performing this calculation for a system size of 2,621,658 atoms, at all-atom level requires enormous computational resources and is clearly beyond the scope of this work. We are also aware that protein folding is a stochastic phenomenon and a single simulation trajectory is insufficient to explore all conformational space. In the current study, due to limitations of computational facility, we have performed single all-atom simulation of both PoF and PrF conformers. Multiple simulations of the system with different initial conditions can sample broad regions of phase space and similar will be attempted in the near future with the availability of larger computational facilities.

3.3. Validation of ecto-TMD gp41 model

In order to validate the homology built model, MM-PBSA free-energy calculation was performed and the free-energy values were compared with the available experimental results (Chan et al., 1998; Champagne et al., 2009; Hamburger, Kim, Welch, & Kay, 2005; McGillick, Balius, Mukherjee, & Rizzo, 2010; Strockbine & Rizzo, 2007). For the

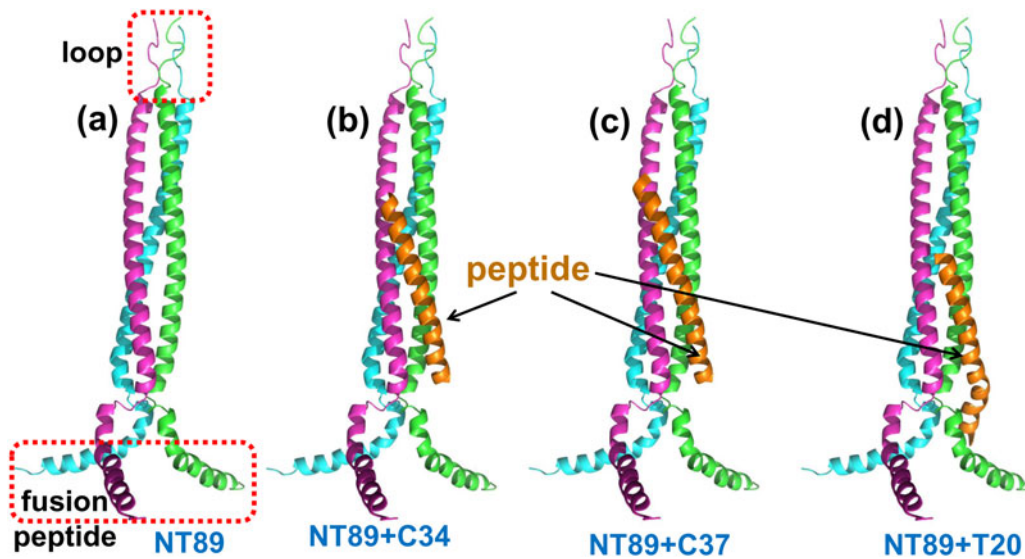


Figure 7. HIV-1 gp41 and peptide complex. The trimer of truncated gp41 (residues 1–89) complexed with (b) C34, (c) C37 and (d) T20 is depicted in the figure. The regions of gp41 position restrained during simulation are shown within (a) red dotted box. Peptide inhibitors are represented in the orange helix.

Table 2. The binding free-energy components of peptides with truncated gp41 trimer calculated using MM-PBSA algorithm and available from experiments is compared.

Energy	T20 (kcal/mol)	C37 (kcal/mol)	C34 (kcal/mol)
ΔE_{vdw}^a	-147.21 ± 8.27	-149.48 ± 8.73	-169.00 ± 10.00
ΔE_{coul}^b	-818.26 ± 35.76	-962.93 ± 53.48	-1120.73 ± 56.27
ΔG_{polar}^c	855.70 ± 34.01	994.28 ± 49.45	1164.49 ± 53.54
$\Delta G_{non-polar}^d$	-15.39 ± 0.57	-15.09 ± 0.52	-16.86 ± 0.81
$\Delta G_{binding}^e$	-125.17 ± 9.45	-133.22 ± 10.64	-142.09 ± 11.71
IC ₅₀	26	2.8	1.5

^aVan der Waals contribution.

^bCoulombic contribution.

^cPolar solvation contribution.

^dNon-polar solvation contribution.

^ePredicted binding energy; IC₅₀ values for T20, C37 and C34 (in nano-molar concentration) are taken from: He et al. (2008), Hamburger, Kim, Welch, and Kay (2005), and Chan, Chutkowski, and Kim (1998), respectively.

free-energy calculations, we have generated the complexes by truncating the homology model of ecto-TMD PoF trimer structure from residues 1 to 89 (represented as N89 in this manuscript) and three well-studied peptide inhibitors derived from the C-heptad repeat (CHR) domain of gp41 residues: C34 (residues 117–150), C37 (residues 114–150) and T20 (residues 127–162) as inhibitors as shown in Figure 7. The binding free-energy values of each peptide inhibitor are given in Table 2. We have also listed various component of the free energy in Table 2.

From Table 2, we observed that the Coulombic interactions are the dominant contributor to the binding free energy because of the charged or polar residues of peptide inhibitors. However, the overall electrostatic interactions ($\Delta G_{elec} = \Delta E_{coul} + \Delta G_{polar}$) was compensated by higher desolvation penalty (ΔG_{polar}). The favorable contribution is mainly attributed to the van der Waals interactions between the peptides and gp41. The calculated binding affinity of the peptides follows the following trend: T20 < C37 < C34, which is in agreement with the available experimental data based on the IC₅₀ values as in Table 2. Note that we have not taken into account the entropy effect and get an overestimated free-energy value. The configurational entropy change of

peptide at bound state neglected during the MM-PBSA calculation gives an enhanced binding energy value and is well documented in the literature (Gilson & Zhou, 2007; Stoica, Sadiq, & Coveney, 2008; Rastelli, Del Rio, Degliesposti, & Sgobba, 2010). It is worth mentioning that for T20, our calculated binding energy value (-125.17 kcal/mol) is comparable to the value (-120.83 kcal/mol) reported by McGillick et al. (2010) estimated using MMGBSA method.

In order to identify the crucial residues responsible for stabilizing the interactions between peptide inhibitors and gp41 receptor, binding free-energy contribution for each residue calculated using MM-PBSA were decomposed (Figure 8). The pivotal residues from the free-energy binding spectra with an arbitrary cut-off of -2.0 kcal/mol were identified. Accordingly, for T20: Tyr127, Ile131, Leu134, Ile135, Ser138, Gln139, Gln142, Leu149, Leu152, Asp153 and Trp159 are identified as crucial; for C37: Trp117, Trp120, Ile124, Tyr127, Ile131, Leu134, Ile135, Glu137, Gln139, Gln141, Gln142, Asn145, Glu148, Leu149 and Leu150 are significant and for C34: Trp117, Trp120, Ile124, Tyr127, Ile131, Leu134, Ile135, Glu137, Gln139, Gln141, Gln142, Leu149, Leu150 are having major contribution for the interaction. The side-chains of three key residues from the N-terminal of C37 and C34 (Trp117, Trp120 and Ile124) are oriented toward the gp41 core (Supplementary Figure S3) to form strong hydrophobic interactions into the deep pocket localized around the N-terminal (Leu57 to Trp60), N36 cavity. Specifically, Trp120 is having the highest interaction within the pocket due to hydrophobic interactions and a stable hydrogen bond with Leu57 of gp41 NHR, which is in agreement with available experimental data (Chan et al., 1998). Peptide C37 and C34 showed higher anti-HIV activity than T20 due to the hydrophobic interactions of three C-terminal residues with the cavity at the NHR region (Champagne et al., 2009). The conserved HIV-1 gp41 N36 hydrophobic pocket is one of the favorable targets to design effective inhibitor (Chan, Fass, Berger, & Kim, 1997). Experimental studies showed that the inhibition potency of CHR based peptides decreases with

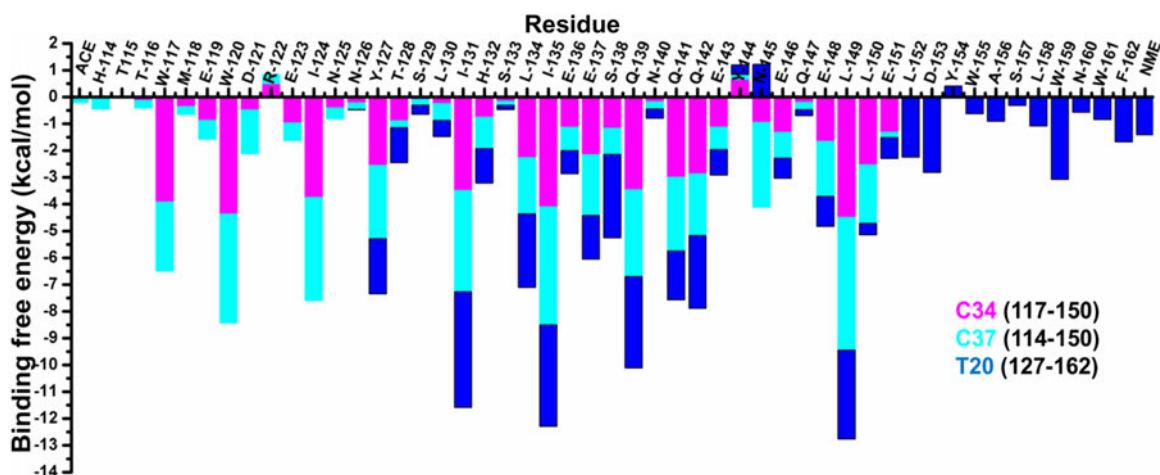


Figure 8. Per-residue binding free-energy decomposition of C34 (magenta), C37 (cyan) and T20 (blue) with truncated gp41 represented in a stacked bar chart.

increase in size, as we also observed a similar trend that the affinity of C34 is higher than C37 (Hamburger et al., 2005). The energy contribution by Met118 and Arg122 are less as their side-chains are projected outwards from the N36 cavity (Chan et al., 1998). Despite the common residues, Tyr127–Leu150, between three peptides considered in the current study, C34 and C37 have characteristic pocket binding domain (residues Trp117–Ile124) which interacts strongly with the hydrophobic groove of NHR, however T20 known to have lipid binding domain (residues Trp159–Phe162) which shows significant affinity for the lipids of host (Champagne et al., 2009; Hildinger et al., 2001; Lawless et al., 1996; Liu et al., 2007; Peisajovich, Gallo, Blumenthal, & Shai, 2003; Wexler-Cohen, Johnson, Puri, Blumenthal, & Shai, 2006). The side-chains of crucial residues at N-terminal of T20 (Tyr127, Ile131, Leu134 and Ile135) also oriented in a similar fashion and contributes to strong hydrophobic interactions between gp120-T20 complex (Supplementary Figure S3). The sequence of T20 is moved by 12 residues toward the C-terminal with respect to C34 and C37. The outstretched 12 C-terminal residues of T20 are composed of eight hydrophobic residues. Among the hydrophobic C-terminal residues of T20, Trp159 of our model has the highest affinity for gp41. The indole group of Trp159 played a crucial role in forming hydrophobic interactions with a cavity formed by residues Met19 to Met24 of gp41. The good packing of Trp159 is also discussed in a similar MD study conducted by McGillick et al. (2010).

4. Conclusion

The structure of the full length trimer of gp41 with the MPER and the TMD region has not been solved at the atomic level and remains one of the major challenges in the HIV research. In the present study, the full length HIV-1 gp41 trimer was modeled using multi-template approach for the first time. We have evaluated the structural parameters of the generated model using several bioinformatics tools. The relative binding affinity of available peptide inhibitors: T20, C34, and C37; were estimated using MM-PBSA technique to

further validate the model. The binding affinity of the peptides for gp41 follow the trend: T20 < C37 < C34; which is in good agreement with the available experimental data. The per-residue energy decomposition elucidate that the higher affinity is due to the strong hydrophobic interactions by residues Trp117, Trp120 and Ile124 of C34 and C37 with N-terminal cavity of gp41, which is lacking in T20. From the stability comparison of PrF and PoF conformers, we observed that later has high structural rigidity, which explains the tendency of PrF to acquire the energetically stable PoF conformer. The PrF conformer significantly bends near the IL region and showed an intrinsic propensity to acquire the PoF conformer during unrestrained MD at near physiological conditions. FEL plot constructed from RMSD and R_g elucidated the diverse conformational space explored by unstable PrF, however, the energy surface of PoF was funneled near the most stable conformer. The reliability of the HIV-1 gp41 model predicted herein is validated on structural and functional aspects. Further gp41 mediated fusion mechanism of HIV and human membrane models using MD studies will provide detailed insight into the invasion mechanism of the virus and can pave ways to design potent small molecule inhibitors to eradicate AIDS. Such studies are currently underway in our laboratory and will be published elsewhere.

Acknowledgements

We acknowledge Sahasrat of Supercomputer Education and Research Center (SERC) and Thematic Unit of Excellence on Computational Materials Science (TUE-CMS) of Solid State and Structural Chemistry Unit (SSCU) at IISc, Bangalore for providing access to the high performance supercomputer facilities.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

BG thanks to Dr. D. S. Kothari Postdoctoral Fellowship scheme by University Grants Commission (UGC), India.

References

- Allen, W. J., Yi, H. A., Gochin, M., Jacobs, A., & Rizzo, R. C. (2015). Small molecule inhibitors of HIVgp41 N-heptad repeat trimer formation. *Bioorganic & Medicinal Chemistry Letters*, *25*, 2853–2859. doi:10.1016/j.bmcl.2015.04.067
- Apellániz, B., Rujas, E., Serrano, S., Morante, K., Tsumoto, K., Caaveiro, J. M., ... Nieva, J. L. (2015). The atomic structure of the HIV-1 gp41 transmembrane domain and its connection to the immunogenic membrane-proximal external region. *Journal of Biological Chemistry*, *290*, 12999–13015. doi:10.1074/jbc.M115.644351
- Archin, N. M., Sung, J. M., Garrido, C., Soriano-Sarabia, N., & Margolis, D. M. (2014). Eradicating HIV-1 infection: Seeking to clear a persistent pathogen. *Nature Reviews Microbiology*, *12*, 750–764. doi:10.1038/nrmicro3352
- Arhel, N., & Kirchhoff, F. (2010). Host proteins involved in HIV infection: New therapeutic targets. *Biochimica et Biophysica Acta*, *1802*, 313–321. doi:10.1016/j.bbadis.2009.12.003
- Arts, E. J., & Hazuda, D. J. (2012). HIV-1 antiretroviral drug therapy. *Cold Spring Harbor Perspectives in Medicine*, *2*(4), a007161. doi:10.1101/cshperspect.a007161
- Ashkenazi, A., Merklinger, E., & Shai, Y. (2012). Intramolecular interactions within the human immunodeficiency virus-1 gp41 loop region and their involvement in lipid merging. *Biochemistry*, *51*, 6981–6989. doi:10.1021/bi300868f
- Ashkenazi, A., Viard, M., Wexler-Cohen, Y., Blumenthal, R., & Shai, Y. (2011). Viral envelope protein folding and membrane hemifusion are enhanced by the conserved loop region of HIV-1 gp41. *The FASEB Journal*, *25*, 2156–2166. doi:10.1096/fj.10-175752
- Baker, N. A., Sept, D., Joseph, S., Holst, M. J., & McCammon, J. A. (2001). Electrostatics of nanosystems: Application to microtubules and the ribosome. *Proceedings of the National Academy of Sciences of the United States of America*, *98*, 10037–10041. doi:10.1073/pnas.181342398
- Barré-Sinoussi, F., Ross, A. L., & Delfraissy, J. F. (2013). Past, present and future: 30 years of HIV research. *Nature Reviews Microbiology*, *11*, 877–883. doi:10.1038/nrmicro3132
- Berendsen, H. J. C., van der Spoel, D., & van Drunen, R. (1995). GROMACS: A message-passing parallel molecular dynamics implementation. *Computer Physics Communications*, *91*, 43–56. doi:10.1016/0010-4655(95)00042-E
- Berger, E. A., Murphy, P. M., & Farber, J. M. (1999). Chemokine receptors as HIV-1 coreceptors: Roles in viral entry, tropism, and disease. *Annual Review of Immunology*, *17*, 657–700. doi:10.1146/annurev.immunol.17.1.657
- Briggs, J. A., & Kräusslich, H. G. (2011). The molecular architecture of HIV. *Journal of Molecular Biology*, *410*, 491–500. doi:10.1016/j.jmb.2011.04.021
- Brügger, B., Glass, B., Haberkant, P., Leibrecht, I., Wieland, F. T., & Kräusslich, H. G. (2006). The HIV lipidome: A raft with an unusual composition. *Proceedings of the National Academy of Sciences of the United States of America*, *103*, 2641–2646. doi:10.1073/pnas.0511136103
- Cai, L., Gochin, M., & Liu, K. (2011). Biochemistry and biophysics of HIV-1 gp41 – Membrane interactions and implications for HIV-1 envelope protein mediated viral-cell fusion and fusion inhibitor design. *Current Topics in Medicinal Chemistry*, *11*, 2959–2984. doi:10.2174/156802611798808497
- Champagne, K., Shishido, A., & Root, M. J. (2009). Interactions of HIV-1 inhibitory peptide T20 with the gp41 N-HR coiled coil. *The Journal of Biological Chemistry*, *284*, 3619–3627. doi:10.1074/jbc.M809269200
- Chan, D. C., Chutkowski, C. T., & Kim, P. S. (1998). Evidence that a prominent cavity in the coiled coil of HIV type 1 gp41 is an attractive drug target. *Proceedings of the National Academy of Sciences of the United States of America*, *95*, 15613–15617. doi:10.1073/pnas.95.26.15613
- Chan, D. C., Fass, D., Berger, J. M., & Kim, P. S. (1997). Core structure of gp41 from the HIV envelope glycoprotein. *Cell*, *89*, 263–273.
- Checkley, M. A., Luttgé, B. G., & Freed, E. O. (2011). HIV-1 envelope glycoprotein biosynthesis, trafficking, and incorporation. *Journal of Molecular Biology*, *410*, 582–608. doi:10.1016/j.jmb.2011.04.042
- Cohen, M. S., Chen, Y. Q., McCauley, M., Gamble, T., Hosseini, M. C., Kumarasamy, N., ... Fleming, T. R. (2011). Prevention of HIV-1 infection with early antiretroviral therapy. *New England Journal of Medicine*, *365*, 493–505. doi:10.1056/NEJMoa1105243
- Crooks, G. E., Hon, G., Chandonia, J. M., & Brenner, S. E. (2004). WebLogo: A sequence logo generator. *Genome Research*, *14*, 1188–1190. doi:10.1101/gr.849004
- Darden, T., York, D., & Pedersen, L. (1993). Particle mesh Ewald: An N-log(N) method for Ewald sums in large systems. *The Journal of Chemical Physics*, *98*, 10089–10092. doi:10.1063/1.464397
- DeGruttola, V., Smith, D. M., Little, S. J., & Miller, V. (2010). Developing and evaluating comprehensive HIV infection control strategies: Issues and challenges. *Clinical Infectious Diseases*, *50*, S102–S107. doi:10.1086/651480
- Emsley, P., & Cowtan, K. (2004). Coot: Model-building tools for molecular graphics. *Acta Crystallographica. Section D, Biological Crystallography*, *60*, 2126–2132. doi:10.1107/S0907444904019158
- Emsley, P., Lohkamp, B., Scott, W. G., & Cowtan, K. (2010). Features and development of coot. *Acta Crystallographica. Section D, Biological Crystallography*, *66*, 486–501. doi:10.1107/S0907444910007493
- Esté, J. A., & Cihlar, T. (2010). Current status and challenges of antiretroviral research and therapy. *Antiviral Research*, *85*, 25–33. doi:10.1016/j.antiviral.2009.10.007
- Fauci, A. S., & Marston, H. D. (2015). PUBLIC HEALTH. Toward an HIV vaccine: A scientific journey. *Science (New York, N.Y.)*, *349*, 386–387. doi:10.1126/science.aac6300
- Fiser, A., & Sali, A. (2003). Modeller: Generation and refinement of homology-based protein structure models. *Methods in Enzymology*, *374*, 461–491. doi:10.1016/S0076-6879(03)74020-8
- Flanagan, C. A. (2014). Receptor conformation and constitutive activity in CCR5 chemokine receptor function and HIV infection. *Advances in Pharmacology*, *70*, 215–263.
- Flexner, C. (2007). HIV drug development: The next 25 years. *Nature Reviews Drug Discovery*, *6*, 959–966. doi:10.1038/nrd2336
- Gangupomu, V. K., & Abrams, C. F. (2010). All-atom models of the membrane-spanning domain of HIV-1 gp41 from metadynamics. *Biophysical Journal*, *99*, 3438–3444. doi:10.1016/j.bpj.2010.09.054
- Gao, F., Bailes, E., Robertson, D. L., Chen, Y., Rodenburg, C. M., Michael, S. F., ... Hahn, B. H. (1999). Origin of HIV-1 in the chimpanzee Pan troglodytes troglodytes. *Nature*, *397*, 436–441. doi:10.1038/17130
- Genheden, S., & Ryde, U. (2015). The MM/PBSA and MM/GBSA methods to estimate ligand-binding affinities. *Expert Opinion on Drug Discovery*, *10*, 449–461. doi:10.1517/17460441.2015.1032936
- Gilson, M. K., & Zhou, H. X. (2007). Calculation of protein-ligand binding affinities. *Annual Review of Biophysics and Biomolecular Structure*, *36*, 21–42. doi:10.1146/annurev.biophys.36.040306.132550
- Glaser, F., Pupko, T., Paz, I., Bell, R. E., Bechor-Shental, D., Martz, E., & Ben-Tal, N. (2003). ConSurf: Identification of functional regions in proteins by surface-mapping of phylogenetic information. *Bioinformatics*, *19*, 163–164. doi:10.1093/bioinformatics/19.1.163
- Goto, T., Nakai, M., & Ikuta, K. (1998). The life-cycle of human immunodeficiency virus type 1. *Micron (Oxford, England : 1993)*, *29*, 123–138.
- Guttman, M., Kahn, M., Garcia, N. K., Hu, S. L., & Lee, K. K. (2012). Solution structure, conformational dynamics, and CD4-induced activation in full-length, glycosylated, monomeric HIV gp120. *Journal of Virology*, *86*, 8750–8764. doi:10.1128/JVI.07224-11
- Hamburger, A. E., Kim, S., Welch, B. D., & Kay, M. S. (2005). Steric accessibility of the HIV-1 gp41 N-trimer region. *Journal of Biological Chemistry*, *280*, 12567–12572. doi:10.1074/jbc.M412770200
- Harrison, S. C. (2005). Mechanism of membrane fusion by viral envelope proteins. *Advances in Virus Research*, *64*, 231–261. doi:10.1016/S0065-3527(05)64007-9
- Harrison, S. C. (2008). Viral membrane fusion. *Nature Structural & Molecular Biology*, *15*, 690–698. doi:10.1038/nsmb.1456
- Harrison, S. C. (2015). Viral membrane fusion. *Virology*, *479–480*, 498–507. doi:10.1016/j.virol.2015.03.043
- He, Y., Cheng, J., Lu, H., Li, J., Hu, J., Qi, Z., ... Dai, Q. (2008). Potent HIV fusion inhibitors against Enfuvirtide-resistant HIV-1 strains. *Proceedings of the National Academy of Sciences of the United States of America*, *105*, 16332–16337. doi:10.1073/pnas.0807335105

- Hemelaar, J. (2012). The origin and diversity of the HIV-1 pandemic. *Trends in Molecular Medicine*, 18, 182–192. doi:10.1016/j.molmed.2011.12.001
- Higo, J., Ito, N., Kuroda, M., Ono, S., Nakajima, N., & Nakamura, H. (2001). Energy landscape of a peptide consisting of alpha-helix, 3(10)-helix, beta-turn, beta-hairpin, and other disordered conformations. *Protein Science*, 10, 1160–1171. doi:10.1110/ps.44901
- Hildinger, M., Dittmar, M. T., Schult-Dietrich, P., Fehse, B., Schnierle, B. S., Thaler, S., ... von Laer, D. (2001). Membrane-anchored peptide inhibits human immunodeficiency virus entry. *Journal of Virology*, 75, 3038–3042. doi:10.1128/JVI.75.6.3038-3042.2001
- Jessen, H., Allen, T. M., & Streeck, H. (2014). How a single patient influenced HIV research-15-year follow-up. *The New England Journal of Medicine*, 370, 682–683. doi:10.1056/NEJMc1308413
- Kallings, L. O. (2008). The first postmodern pandemic: 25 years of HIV/AIDS. *Journal of Internal Medicine*, 263, 218–243. doi:10.1111/j.1365-2796.2007.01910.x
- Kim, J. H., Hartley, T. L., Curran, A. R., & Engelman, D. M. (2009). Molecular dynamics studies of the transmembrane domain of gp41 from HIV-1. *Biochimica et Biophysica Acta*, 1788, 1804–1812. doi:10.1016/j.bbamem.2009.06.011
- Kuzembayeva, M., Dilley, K., Sardo, L., & Hu, W. S. (2014). Life of psi: How full-length HIV-1 RNAs become packaged genomes in the viral particles. *Virology*, 454–455, 362–370. doi:10.1016/j.virol.2014.01.019
- Kwon, Y. D., Finzi, A., Wu, X., Dogo-Isonagie, C., Lee, L. K., Moore, L. R., ... Kwong, P. D. (2012). 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*, 109, 5663–5668. doi:10.1073/pnas.1112391109
- Kwong, P. D., Wyatt, R., Robinson, J., Sweet, R. W., Sodroski, J., & Hendrickson, W. A. (1998). Structure of an HIV gp120 envelope glycoprotein in complex with the CD4 receptor and a neutralizing human antibody. *Nature*, 393, 648–659. doi:10.1038/31405
- Lai, A. L., & Freed, J. H. (2014). HIV gp41 fusion peptide increases membrane ordering in a cholesterol-dependent fashion. *Biophysical Journal*, 106, 172–181. doi:10.1016/j.bpj.2013.11.027
- Lawless, M. K., Barney, S., Guthrie, K. I., Bucy, T. B., Petteway, S. R., Jr., & Merutka, G. (1996). HIV-1 membrane fusion mechanism: Structural studies of the interactions between biologically-active peptides from gp41. *Biochemistry*, 35, 13697–13708. doi:10.1021/bi9606962
- Lindorff-Larsen, K., Piana, S., Palmo, K., Maragakis, P., Klepeis, J. L., Dror, R. O., & Shaw, D. E. (2010). Improved side-chain torsion potentials for the amber ff99SB protein force field. *Proteins: Structure, Function, and Bioinformatics*, 78, 1950–1958. doi:10.1002/prot.22711
- Liu, S., Jing, W., Cheung, B., Lu, H., Sun, J., Yan, X., ... Jiang, S. (2007). HIV gp41 C-terminal heptad repeat contains multifunctional domains. Relation to mechanisms of action of anti-HIV peptides. *Journal of Biological Chemistry*, 282, 9612–9620. doi:10.1074/jbc.M609148200
- Lu, K., Heng, X., & Summers, M. F. (2011). Structural determinants and mechanism of HIV-1 genome packaging. *Journal of Molecular Biology*, 410, 609–633. doi:10.1016/j.jmb.2011.04.029
- Maartens, G., Celum, C., & Lewin, S. R. (2014). HIV infection: Epidemiology, pathogenesis, treatment, and prevention. *Lancet*, 384, 258–271. doi:10.1016/S0140-6736(14)60164-1
- Mascola, J. R. (2015). HIV. The modern era of HIV-1 vaccine development. *Science (New York, N.Y.)*, 349, 139–140. doi:10.1126/science.aac7800
- Mathys, L., & Balzarini, J. (2015). Several N-glycans on the HIV envelope glycoprotein gp120 preferentially locate near disulphide bridges and are required for efficient infectivity and virus transmission. *PLoS One*, 10(6), e0130621. doi:10.1371/journal.pone.0130621
- McGillick, B. E., Balius, T. E., Mukherjee, S., & Rizzo, R. C. (2010). Origins of resistance to the HIVgp41 viral entry inhibitor T20. *Biochemistry*, 49, 3575–3592. doi:10.1021/bi901915g
- Melikyan, G. B. (2008). Common principles and intermediates of viral protein-mediated fusion: The HIV-1 paradigm. *Retrovirology*, 5, 111. doi:10.1186/1742-4690-5-111
- Menéndez-Arias, L. (2013). Molecular basis of human immunodeficiency virus type 1 drug resistance: Overview and recent developments. *Antiviral Research*, 98, 93–120. doi:10.1016/j.antiviral.2013.01.007
- Merk, A., & Subramaniam, S. (2013). HIV-1 envelope glycoprotein structure. *Current Opinion in Structural Biology*, 23, 268–276. doi:10.1016/j.sbi.2013.03.007
- Miller, B. R., 3rd, McGee, T. D., Jr., Swails, J. M., Homeyer, N., Gohlke, H., & Roitberg, A. E. (2012). MMPBSA.py: An efficient program for end-state free energy calculations. *Journal of Chemical Theory and Computation*, 8, 3314–3321. doi:10.1021/ct300418h
- Moreno, M. R., Giudici, M., & Villalain, J. (2006). The membranotropic regions of the endo and ecto domains of HIV gp41 envelope glycoprotein. *Biochimica et Biophysica Acta*, 1758, 111–123. doi:10.1016/j.bbamem.2006.01.007
- Moreno, M. R., Pascual, R., & Villalain, J. (2004). Identification of membrane-active regions of the HIV-1 envelope glycoprotein gp41 using a 15-mer gp41-peptide scan. *Biochimica et Biophysica Acta*, 1661, 105–197. doi:10.1016/j.bbamem.2003.12.003
- Muñoz-Barroso, I., Salzwedel, K., Hunter, E., & Blumenthal, R. (1999). Role of the membrane-proximal domain in the initial stages of human immunodeficiency virus type 1 envelope glycoprotein-mediated membrane fusion. *Journal of Virology*, 73, 6089–6092.
- Munro, J. B., & Mothes, W. (2015). Structure and Dynamics of the Native HIV-1 Env Trimer. *Journal of Virology*, 89, 5752–5755. doi:10.1128/JVI.03187-14
- Paillart, J. C., Shehu-Xhilaga, M., Marquet, R., & Mak, J. (2004). Dimerization of retroviral RNA genomes: An inseparable pair. *Nature Reviews Microbiology*, 2, 461–472. doi:10.1038/nrmicro903
- Pancera, M., Zhou, T., Druz, A., Georgiev, I. S., Soto, C., Gorman, J., ... Kwong, P. D. (2014). Structure and immune recognition of trimeric pre-fusion HIV-1 Env. *Nature*, 514, 455–461. doi:10.1038/nature13808
- Pascual, R., Moreno, M. R., & Villalain, J. (2005). A peptide pertaining to the loop segment of human immunodeficiency virus gp41 binds and interacts with model biomembranes: Implications for the fusion mechanism. *Journal of Virology*, 79, 5142–5152. doi:10.1128/JVI.79.8.5142-5152.2005
- Peisajovich, S. G., Blank, L., Eband, R. F., Eband, R. M., & Shai, Y. (2003). On the interaction between gp41 and membranes: The immunodominant loop stabilizes gp41 helical hairpin conformation. *Journal of Molecular Biology*, 326, 1489–1501. doi:10.1016/S0022-2836(03)00040-8
- Peisajovich, S. G., Gallo, S. A., Blumenthal, R., & Shai, Y. (2003). C-terminal octylation rescues an inactive T20 mutant: Implications for the mechanism of HIV/SIMIAN immunodeficiency virus-induced membrane fusion. *Journal of Biological Chemistry*, 278, 21012–21017. doi:10.1074/jbc.M212773200
- Perrin, C., Fenouillet, E., & Jones, I. M. (1998). Role of gp41 glycosylation sites in the biological activity of human immunodeficiency virus type 1 envelope glycoprotein. *Virology*, 242, 338–345. doi:10.1006/viro.1997.9016
- Ramachandran, G. N., Ramakrishnan, C., & Sasisekharan, V. (1963). Stereochemistry of polypeptide chain configurations. *Journal of Molecular Biology*, 7, 95–99. doi:10.1016/S0022-2836(63)80023-6
- Rastelli, G., Del Rio, A., Degliesposti, G., & Sgobba, M. (2010). Fast and accurate predictions of binding free energies using MM-PBSA and MM-GBSA. *Journal of Computational Chemistry*, 31, 797–810. doi:10.1002/jcc.21372
- Reddy, M. R., Reddy, C. R., Rathore, R. S., Erion, M. D., Aparoy, P., Reddy, R. N., & Reddanna, P. (2014). Free energy calculations to estimate ligand-binding affinities in structure-based drug design. *Current Pharmaceutical Design*, 20, 3323–3337. doi:10.2174/13816128113199990604
- Ruelas, D. S., & Greene, W. C. (2013). An integrated overview of HIV-1 latency. *Cell*, 155, 519–529. doi:10.1016/j.cell.2013.09.044
- Russell, R. S., Liang, C., & Wainberg, M. A. (2004). Is HIV-1 RNA dimerization a prerequisite for packaging? Yes, no, probably?. *Retrovirology*, 1, 23. doi:10.1186/1742-4690-1-23
- Ryckaert, J.-P., Ciccotti, G., & Berendsen, H. J. C. (1977). Numerical integration of the Cartesian equations of motion of a system with constraints: Molecular dynamics of n-alkanes. *Journal of Computational Physics*, 23, 327–341. doi:10.1016/0021-9991(77)90098-5
- Sali, A., & Blundell, T. L. (1993). Comparative protein modelling by satisfaction of spatial restraints. *Journal of Molecular Biology*, 234, 779–815. doi:10.1006/jmbi.1993.1626

- Salzwedel, K., West, J. T., & Hunter, E. (1999). A conserved tryptophan-rich motif in the membrane-proximal region of the human immunodeficiency virus type 1 gp41 ectodomain is important for Env-mediated fusion and virus infectivity. *Journal of Virology*, *73*, 2469–2480.
- Schur, F. K., Hagen, W. J., Rumlová, M., Ruml, T., Müller, B., Kräusslich, H. G., & Briggs, J. A. (2015). Structure of the immature HIV-1 capsid in intact virus particles at 8.8 Å resolution. *Nature*, *517*, 505–508. doi:10.1038/nature13838
- Sharp, P. M., & Hahn, B. H. (2010). The evolution of HIV-1 and the origin of AIDS. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *365*, 2487–2494. doi:10.1098/rstb.2010.0031
- Sharp, P. M., & Hahn, B. H. (2011). Origins of HIV and the AIDS pandemic. *Cold Spring Harbor Perspectives in Medicine*, *1*(1), a006841. doi:10.1101/cshperspect.a006841
- Shen, M. Y., & Sali, A. (2006). Statistical potential for assessment and prediction of protein structures. *Protein Science*, *15*, 2507–2524. doi:10.1110/ps.062416606
- Smyth, R. P., Davenport, M. P., & Mak, J. (2012). The origin of genetic diversity in HIV-1. *Virus Research*, *169*, 415–429. doi:10.1016/j.virusres.2012.06.015
- Stoica, I., Sadiq, S. K., & Coveney, P. V. (2008). Rapid and accurate prediction of binding free energies for saquinavir-bound HIV-1 proteases. *Journal of the American Chemical Society*, *130*, 2639–2648. doi:10.1021/ja0779250
- Strockbine, B., & Rizzo, R. C. (2007). Binding of antifusion peptides with HIVgp41 from molecular dynamics simulations: Quantitative correlation with experiment. *Proteins: Structure, Function, and Bioinformatics*, *67*, 630–642. doi:10.1002/prot.21301
- Sundquist, W. I., & Kräusslich, H. G. (2012). HIV-1 assembly, budding, and maturation. *Cold Spring Harbor Perspectives in Medicine*, *2*(7), a006924. doi:10.1101/cshperspect.a006924
- Tamm, L. K., Lee, J., & Liang, B. (2014). Capturing glimpses of an elusive HIV gp41 prehairpin fusion intermediate. *Structure*, *22*, 1225–1226. doi:10.1016/j.str.2014.08.010
- Turner, B. G., & Summers, M. F. (1999). Structural biology of HIV. *Journal of Molecular Biology*, *285*(1), 1–32. doi:10.1006/jmbi.1998.2354
- Vrahatis, M. N., Androulakis, G. S., Lambrinos, J. N., & Magoulas, G. D. (2000). A class of gradient unconstrained minimization algorithms with adaptive stepsize. *Journal of Computational and Applied Mathematics*, *114*, 367–386. doi:10.1016/S0377-0427(99)00276-9
- Wang, C., Greene, D., Xiao, L., Qi, R., & Luo, R. (2018). Recent developments and applications of the MMPBSA method. *Frontiers in Molecular Biosciences*, *4*, 87. doi:10.3389/fmolb.2017.00087
- Wexler-Cohen, Y., Johnson, B. T., Puri, A., Blumenthal, R., & Shai, Y. (2006). Structurally altered peptides reveal an important role for N-terminal heptad repeat binding and stability in the inhibitory action of HIV-1 peptide DP178. *Journal of Biological Chemistry*, *281*, 9005–9010. doi:10.1074/jbc.M512475200
- Yuan, W., Craig, S., Si, Z., Farzan, M., & Sodroski, J. (2004). CD4-induced T-20 binding to human immunodeficiency virus type 1 gp120 blocks interaction with the CXCR4 coreceptor. *Journal of Virology*, *78*, 5448–5457. doi:10.1128/JVI.78.10.5448-5457.2004
- Zhao, G., Perilla, J. R., Yufenyuy, E. L., Meng, X., Chen, B., Ning, J., ... Zhang, P. (2013). Mature HIV-1 capsid structure by cryo-electron microscopy and all-atom molecular dynamics. *Nature*, *497*, 643–646. doi:10.1038/nature12162