JOURNAL OF
EMERGING INVESTIGATORS

Article

The analysis of the viral transmission and structural interactions between the HIV-1 envelope glycoprotein and the lymphocyte receptor integrin α4β7

Harrish Ganesh, Kathleen M. Morrow
Thomas Jefferson High School for Science and Technology, Alexandria, Virginia

SUMMARY

The Human Immunodeficiency Virus (HIV) infects approximately 40 million people globally, and one million people die every year from Acquired Immune Deficiency Syndrome (AIDS)-related illnesses. The HIV-1 virus proliferates, infects people, and kills those who host the virus. Unfortunately, there is little research on the structural interactions between HIV-1 proteins and human cell receptors. This study examined the interactions between the HIV-1 envelope glycoprotein gp120 and the human lymphocyte receptor integrin α4β7, the putative first long-range receptor for the envelope glycoprotein of the virus in mucosal tissues. The specific site of binding between the glycoprotein and the receptor has not been determined. However, discovering the binding site can open the field towards the research of molecules that could potentially block this interaction and prevent the initial binding of the HIV-1 virus. We hypothesized that the V1 and V2 loops of the envelope glycoprotein of HIV-1 are involved in the binding between the HIV-1 virus and the human lymphocyte receptor α4β7. Using structural analysis software, we analyzed the electrostatics and structural interactions between the glycoprotein and receptor. We report structural insights into the interactions between α4β7 and envelope glycoprotein gp120. Our data support the claim that the V1 loop is involved in the binding between α4β7 and the HIV-1 envelope glycoprotein through molecular dockings.

INTRODUCTION

Globally, Human Immunodeficiency Virus-1 (HIV-1) is the most common of the different strains of HIV (1-3). It takes millions of lives each year, and to date, there is no cure for the virus (4). When HIV-1 enters a host organism, it interacts with many other surface molecules that can assist with the entry of the virus into a lymphocyte (5). Examples of these surface molecules include C-type lectin receptors such as dendritic cell- specific intercellular adhesion molecule-3-grabbing non-integrin (DC-SIGN), heparin sulfate proteoglycan (HSPG), and the α4β7 integrin (6-8).

Within the HIV-1 infection process, the primary HIV-1 receptor and co-receptor are CD4 and CCR5 (or CXCR4), respectively (9). Both receptors on the lymphocyte interact with envelope glycoprotein (i.e., gp120) on the surface of CD4+ T-cells in the mucosa (10-11). The gp120 HIV trimer envelope glycoprotein interacts with the CD4 receptor, which causes a structural change that exposes CCR5 (12). Upon binding to CCR5, a series of conformational changes occur in the protomer, resulting in the insertion of the gp41 fusion peptide into the host cell membrane (13). This allows the host lymphocyte to fuse with the envelope glycoprotein via the α4β7 receptor protein.

The α4β7 receptor protein is a highly specialized integrin, part of the α4 integrin family, and is found in the lymphocyte cell, a type of white blood cell. The α4β7 protein allows the transmission of the viral genome into the host lymphocyte through the CD4 receptor and CCR5 co-receptor, eventually leading to lysis and release of virions (14). The release of the virions will infect other nearby lymphocytes in the host organism, ultimately leading to the near-complete infection of host lymphocytes (14). The natural ligand of α4β7 is a mucosal vascular addressin cell adhesion molecule 1 (MadCAM-1). MadCAM-1 directs lymphocyte traffic into the mucosal tissues, generally controlling the contents of the lymphocyte (15). Thus, α4β7 uses this natural ligand to regulate the internal conditions of the cell. The HIV-1 proteins involved in the initial binding to the host cell also include the viral envelope glycoprotein, gp120, in addition to the human lymphocyte receptor protein α4β7 described above (14). There are five amino acid loops on the gp120 HIV trimer envelope glycoprotein known as the V1, V2, V3, V4, and V5 loops (16). However, the particular loops of the gp120 trimer involved in the binding of α4β7 are not known.

Therefore, this study investigated which amino acids are involved in the initial binding and transmission of the HIV-1 virus, as well as structural insight into the binding between HIV-1 envelope glycoproteins and α4β7. Previous research has not studied the structural basis for the interactions between α4β7 and the HIV-1 glycoprotein and the amino acids involved in the direct infection of the virus. With this in mind, we hypothesized that the V1 and V2 loops are involved in the binding between the HIV-1 virus and the human lymphocyte receptor α4β7, as has been previously reported (12). Through molecular dockings between α4β7 and HIV-

1 envelope glycoprotein 120, we found that the V1 loop of gp120 is primarily involved in HIV-1 viral transmission.

RESULTS

We aimed to identify which loops of gp120 (V1, V2, V3, V4, V5) primarily interact with α4β7 integrin through a computational approach. Through the UCSF Chimera software, we imported PDB files containing both α4β7 and the HIV-1 glycoprotein into the Rosetta Server (17). The Rosetta Server is the software used to conduct molecular docking experiments. Molecular docking experiments involve a computational simulation analyzing hundreds of possible conformations between two proteins (e.g., α4β7 and gp120). These conformations are evaluated by a docking score that accounts for binding energy, molecular distances, and intermolecular/intramolecular forces between proteins (18). The highest-scoring docking conformations are the most probable interaction between the inputted proteins (18). By analyzing these high-scoring binding conformations, one can gain more insight into the binding interaction of the HIV-1 glycoprotein upon the lymphocyte receptor integrin α4β7.

Structural model of the founder virus gp120

The use of the founder virus F100 structure was essential because founder viruses are the first viruses that establish an infection in a host and do not have accumulated escape mutations that occur after the initial contact (19). This structural model (**Figure 1**) contains the missing loops that the original structure published in January 2019 does not contain (20). The goal was to create a structure to utilize in docking experiments. Using UCSF Chimera to superimpose the structures, we created a model of an HIV trimer with the monomer model, and this HIV trimer permitted us to conduct docking experiments (**Figure 1**) (17). This model accounts for updated differences in founder viruses while also creating a suitable framework for structural interactions and analysis.

Structural analysis of integrin binding site

The α4β7 integrin has been studied in detail, and a structure of its head with a small antagonist (RO0505376) has been determined (PDB ID: 3v4v) (21). An antagonist is a

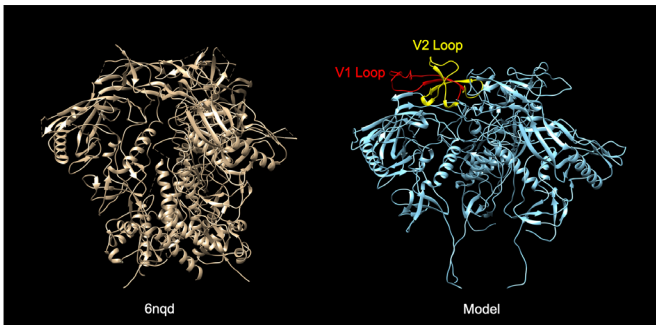

Figure 1. The SwissModel of F100 trimer gp120. On the left, is the original F100 trimer gp120. On the right, which is the model, is the same as the original structure (left, 6nqd) plus the loops.

molecule that prevents the binding or activation of a protein. The goal was to learn more about the binding site and how small antagonists in the structures disrupt the binding site. Therefore, more integrins were studied and superimposed to the α4β7 head. By analyzing the interactions between the α4β7 receptor head and the envelope glycoprotein of F100, one may draw critical conclusions of the HIV-1 infection process. We located the binding site between the integrin and the V1 and V2 subunits of gp120 in a groove (given its gap-like structure), in addition to three positive ions (given its large structure relative to the surrounding amino acids) that are essential for functionality (**Figure 3**).

Structural docking

We performed docking experiments using the Rosetta server with the gp120 trimer model, including the V1 loop and the V2 loop, and with the α4β7 integrin head. The Rosetta server performs a series of molecular dockings and evaluates thousands of conformations using a docking score, which accounts for factors such as binding energy and molecular distance (18). We used this server to evaluate the various binding conformations between the α4β7 head and the HIV-1 envelope glycoprotein. Our analysis found that the V1-V2 amino acid region comprises approximately 80 amino acids, from amino acids 95 to 174, the putative region of binding between envelope glycoprotein and α4β7 integrin. The goal was to analyze the various HIV-1 amino acid loop interactions

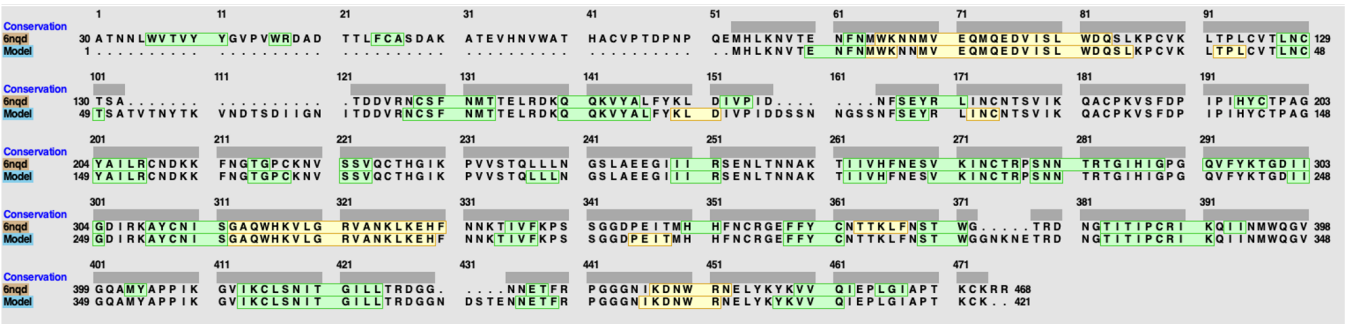

Figure 2. The alignment between the PDB and the model. Conservation of residues is on the top, which shows the modeled four loops (gaps). In the left-hand column, the "6nqd" F100 structure is colored brown, and the model is colored blue. Highlighted in yellow are the alpha-helices and highlighted in green are the beta strands.

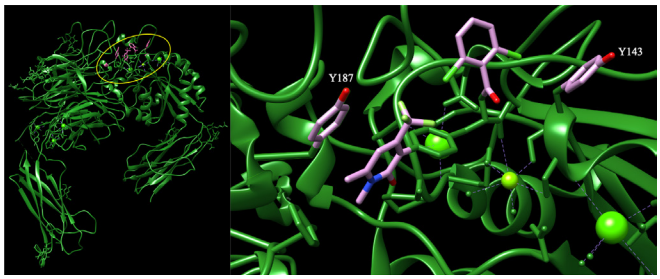

Figure 3. On the left, α4β7 head (PDB ID: 3v4v) showing the binding site in the groove with three ions (yellow circle) essential for functionality. The ion in the center is the MIDAS (metal-ion dependent adhesion site). This integrin's features are described, such as its head and its structure and functions (16). On the right is the α4β7 binding site with the small antagonist, RO0505376. There are aromatic-aromatic interactions (red circles) and a disruption of the MIDAS ion (yellow circle). Aromatic-aromatic interactions are vital for ligand binding and are defined as a collection of physicochemical properties determining specific features of cyclic or polycyclic π-electron molecules (17-19).

with α4β7 to identify which loops are primarily involved in the viral infection. The V1 and V2 loops are critical because they can insert themselves into the binding groove of the α4β7 head. Without them, the binding groove is not reachable. The structural analysis shows that the V1 loop is an ideal choice, as reported previously (12). The SwissModel program was used to synthesize a complete structure of the glycoprotein head for structural analysis (22). We concluded that the V1 loop in the SwissModel trimer is pointing outward more and therefore has a greater chance to insert itself in the binding groove of the α4β7 head, compared to the V2 loop (**Table 1**). The Maestro Schrödinger software was used to analyze electrostatic interactions between the α4β7 and the HIV-1 envelope glycoprotein (23). According to our electrostatic

Table 1. Comparison of the V1 and V2 loops of gp120.

Comparison between V1 and V2 Loops		
Characteristic	**V1 loop**	**V2 loop**
Length	27 amino acids	12 amino acids
Protrusion	Prominent	Extremely small
Location	Side of trimer	Top of trimer

analysis within Maestro Schrödinger, the binding groove is more negatively charged than a more even spread of negative and positive charges. The binding groove is where the positive lysine of the V1 loop inserts itself.

To have a successful, high-feasibility docking, the curvature of the ten best results from Rosetta, based on the docking score, should have a precise funnel shape (**Figure 4**). A lower docking score represents a higher probability of docking. The RMSD (Root Mean Square Deviation) is used to analyze the similarity between two superimposed molecules' location quantitatively. A smaller RMSD indicates a higher probability docking result. The Rosetta server provided data, statistical tests, and feasibilities of binding between the α4β7 head and the integrin. In our highest-scoring result, the binding conformation between α4β7 and the HIV-1 envelope glycoprotein contained a score of -1543.835, supporting the prominent role of the V1 loop in HIV-1 infection (**Figure 5**). This score takes into account molecular distances, binding energies, and intermolecular and intramolecular forces. The score's extremely negative value indicates that this specific conformation (the V1 loop prominently interacting within the groove) is highly probable. The α4β7 head was positioned close to the V1 loop and was used in many docking experiments (**Figures 5-6**). The three amino acids

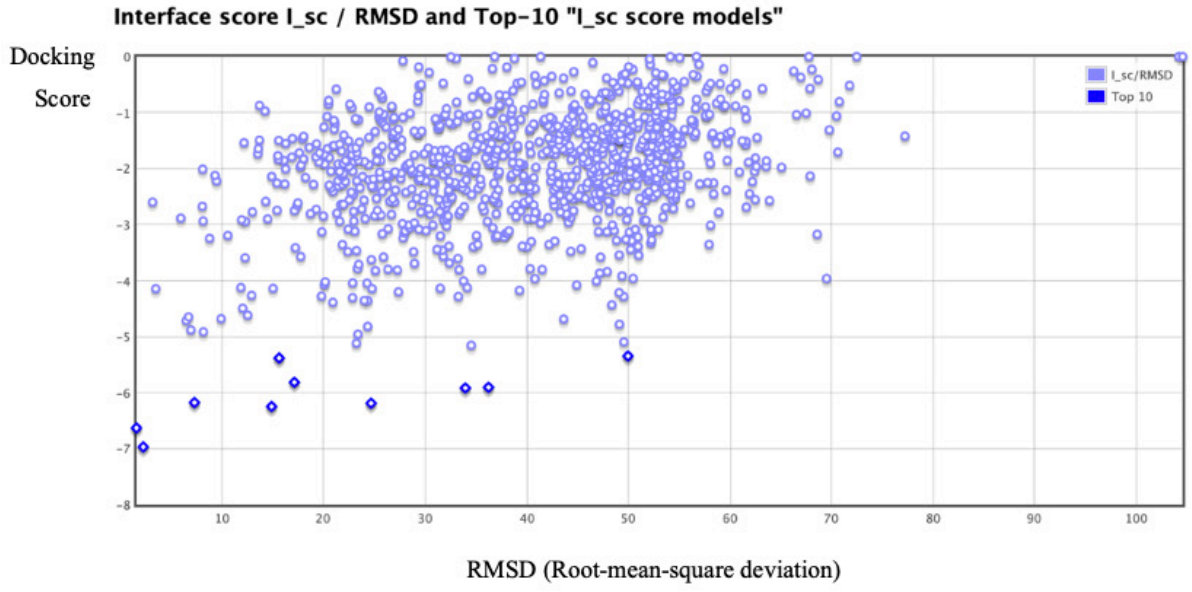

Figure 4. Interface docking score. The ten best results (dark blue) point to the best result on the left corner of the chart. Each dot represents a separate docking simulation between the α4β7 head and the HIV-1 gp120.

indicated in **Figure 6** are the most prominent for interaction with the binding site. Lysine 141 disrupts the MIDAS ion as well as asparagine 143. Tyrosine 139 stabilizes the docking interacting with proline and tyrosine of β7.

DISCUSSION

We hypothesized that the V1 and V2 loops of the gp120 envelope glycoprotein of HIV-1 are involved in the binding between the HIV-1 virus and the human lymphocyte receptor α4β7. The lymphocyte receptor, α4β7, interacts with the envelope glycoprotein of the HIV-1 trimer. However, previous research has not studied the structural basis for the interactions between these proteins. In addition, the amino acids involved in the direct binding of the virus have not been confirmed. HIV-1 interacts with many other surface molecules besides the canonical entry receptor CD4 that can assist with the entry of the virus into a lymphocyte (5). A vital example of these surface molecules includes the α4β7 integrin (21). The α4β7 is a receptor for a lymphocyte cell (i.e., white blood cell) and has a natural ligand of MadCAM-1, which directs lymphocyte traffic into the mucosal tissues (24-25). With this information, the present study tested whether both the V1 and V2 loops are involved in the binding between the HIV-1 virus and the human lymphocyte receptor, α4β7.

Based on structural analysis, docking tests, and electrostatic comparisons, we found that the V1 loop of the envelope glycoprotein is likely to be involved in its interactions with α4β7. Many amino acids of the V1 loop interact with α4β7, according to the electrostatic figures, docking tests, and an enlargement of the interacting amino acids at the docking site. The V1 loop protrudes far enough to interact with the α4β7 and is also the loop closest to the site of interaction with the integrin.

Therefore, we proposed that the V1 loop is directly involved in the interaction between the two proteins. Previously, researchers believed that the V2 loop was involved in the interaction between the HIV-1 virus and α4β7 (12). We performed analyses on the V2 loop, but the loop proved to be too small and unobtrusive to insert itself efficiently into the α4β7 groove (**Figure 1** and **Table 1**). The V1 loop on the trimer side suggests that it can readily interact with the binding groove of the α4β7 head.

The V2 loop also does not protrude far enough to interact at the binding site sufficiently. However, a future goal is to analyze the V2 loop more extensively to test these assumptions. The other loops, which include V3, V4, and V5, are also unlikely to be involved in interactions with α4β7 because they are located on the opposite ends of gp120, away from the binding site. However, there is still much more research needed to be conducted. *In vitro* experiments in the future need to involve the physical interactions of the α4β7 lymphocyte receptor protein and the HIV-1 virus directly.

More docking tests could be done between α4β7 and the HIV-1 envelope glycoprotein to support our findings further. X-ray crystallography could be used to visualize the interactions between the two proteins. Additionally, based on the conditions the HIV-1 envelope glycoprotein exists in, interactions may change (possibly due to pH, temperature, pressure, and other factors). These environmental factors could alter which amino acids are involved in the binding between HIV-1 envelope glycoprotein and α4β7. Additionally, glycans associated with the glycoprotein *in vivo* may make it more challenging to get accurate binding interaction models. Thus, further experiments could control the mentioned variables, such that the binding between the two proteins can be analyzed. Further, the findings can be extended to experiments that involve probing the interactions between HIV-1 envelope glycoproteins and host integrins as a means for viral entry.

METHODS

Structural analysis

University of California San Francisco (UCSF) Chimera, a protein structural analysis software, was used to structurally

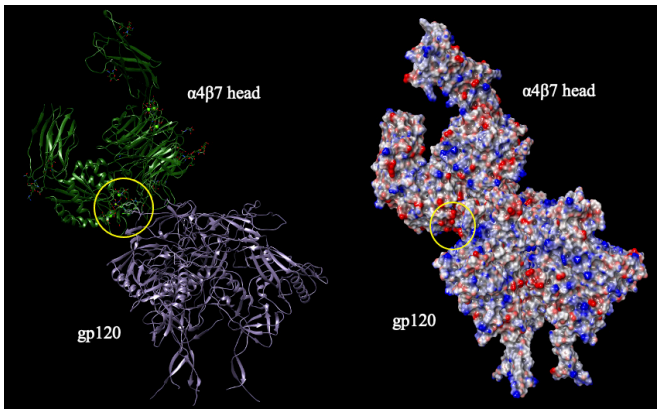

Figure 5. The docking result between α4β7 head and envelope glycoprotein. Red areas are negative, blue areas are positive, and gray areas are neutral.

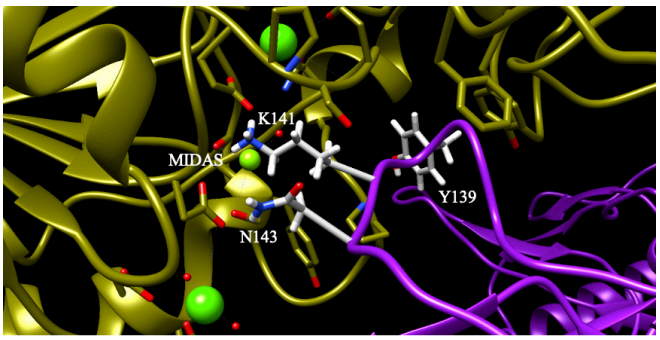

Figure 6. An enlargement of the docking site. The bronze/golden protein is the α4β7 head, and the magenta protein is the founder virus F100, representing the HIV-1 envelope glycoprotein infection process. The three amino acids indicated are the most prominent for interaction with the binding site. Lysine 141 disrupts the MIDAS ion as well as asparagine 143. Tyrosine 139 stabilizes the docking interacting with proline and tyrosine of β7.

analyze the proteins involved in the transmission of HIV-1 (17). The first step was to load the α4β7 protein into the program, which is done by selecting File->Fetch by ID->PDB->3v4v. The PDB is known as The Protein Data Bank. Proteins are labeled with a sequence of numbers and letters, allowing expeditious reference to any protein. The PDB identification for α4β7 is "3v4v". The software will load the structure of α4β7. Loading the glycoprotein of HIV-1, F100 (founder virus 100) gp120, was done by selecting File->Fetch by ID->PDB->6nqd. In this scenario, the PDB ID for F100 is "6nqd". UCSF Chimera will now enter F100 into the program along with α4β7. Additional manipulations were needed, such as the deletion of redundant chains and the positioning of the structures relative to each other. Furthermore, superimpositions with other similar integrins were made in pairs (the α4β7 head with other heads) to understand the integrins' structural similarities and differences and the binding pockets with their respective small-molecule antagonists (obtained from the PDB structure). Sequence alignments were made with UCSF Chimera and analyzed based on sequence and structure.

A sequence alignment between the original sequence F100 and the structure of F100 with missing loops was made and then was input into SwissModel to identify the monomer structure. Then, the monomer model of gp120 glycoprotein was superimposed in Chimera with the "6nqd" corresponding gp120 with an RMSD (root-mean-square deviation) of 0.174 (**Figure 4**), which is precise and accurate because they are practically identical in structure.

Docking experiments

The Rosetta docking server was essential for the many docking experiments between the α4β7 head and the envelope glycoprotein of F100 (PDB ID: 6nqd) (18, 26-27). The Rosetta server is freely accessible. The amino acid loops in this structure were unresolved. Therefore, a structural model containing the missing loops was created with SwissModel, and this trimer model of gp120 glycoprotein was used for docking experiments and the α4β7 head (PDB ID: 3v4v) (22). Furthermore, the server was given files that contained the α4β7 head and the integrin for docking probabilities, as well as the individual amino acid chains of each integrin. The server utilizes a probability algorithm to provide the most feasible docking conformations.

The molecular docking process was repeated for approximately 100 trials, and the resulting docking data were then summarized. The results of the docking experiments were analyzed in UCSF Chimera (for structural analysis) and the Maestro Schrödinger software (for electrostatic analysis) (17, 23). The gp120 trimer model and the α4β7 head were loaded into Maestro, which added hydrogens and assigned partial charges. Then, a surface representation was created, and the option for electrostatics view was selected.

Received: December 13, 2020
Accepted: April 14, 2021
Published: April 27, 2021

REFERENCES

1. Rojas-Celis V, F Valiente-Echeverria, R Soto-Rifo and D Toro-Ascuy. "New Challenges of HIV-1 Infection: How HIV-1 Attacks and Resides in the Central Nervous System." *Cells*, vol. 8, no. 10, 2019, pp. 1245.
2. Kuyl, A C. and M Cornelissen. "Identifying HIV-1 dual infections." *Retrovirology*, 2007, pp. 67.
3. Gartner, Matthew J, *et al*. "Understanding the Mechanisms Driving the Spread of Subtype C HIV-1." *EBioMedicine*, Elsevier, 2020.
4. Brinkhof, M. "Mortality of HIV-infected patients starting antiretroviral therapy in sub-Saharan Africa: comparison with HIV-unrelated mortality." *PLoS medicine*, vol. 6, 2009.
5. Zhong P, L M. Agosto, J B. Munro, and W Mothes. "Cell-to-cell transmission of viruses." *Current opinion in virology*, vol. 3, 2013, pp. 50.
6. Geijtenbeek T B., R Torensma, S J. Vliet, G C. Duijnhoven, G J. Adema, Y Kooyk and C G. Figdor. "Identification of DC-SIGN, a novel dendritic cell-specific ICAM-3 receptor that supports primary immune responses." *Cells*, vol. 100, 2000, pp. 585.
7. Mondor I, S Ugolini and Q J. Sattentau. "Human immunodeficiency virus type 1 attachment to HeLa CD4 cells is CD4 independent and gp120 dependent and requires cell surface heparans." *Journal of virology*, vol. 72, 1998, pp. 3634.
8. Arthos J, C Cicala, E Martinelli, K Macleod, D Ryk, D Wei, Z Xiao, T D. Veenstra, T P. Conrad, R A. Lempicki, S McLaughlin, M Pascuccio, R Gopaul, J McNally, C. Cruz, N Censoplano, E Chung, K N. Reitano, S Kottilil and D Goode. "HIV-1 envelope protein binds to and signals through integrin alpha4beta7, the gut mucosal homing receptor for peripheral T cells." *Nature immunology*, vol. 9, 2008, pp. 309.
9. Basmaciogullari S, B Pacheco, S Bour and J Sodroski. "Specific interaction of CXCR4 with CD4 and CD8alpha: functional analysis of the CD4/CXCR4 interaction in the context of HIV-1 envelope glycoprotein-mediated membrane fusion." *Virology*, vol. 353, 2006, pp. 67.
10. Padilla-Sanchez, V. Macromolecular Assemblies Visualization with UCSF Chimera. *VIS*, 2019.
11. Cicala C, E Martinelli, J P. McNally, D J. Goode, R Gopaul, J Hiatt, K Jelicic, S Kottilil, K Macleod, A O'Shea, N Patel, D Ryk, D Wei, M Pascuccio, L Yi, L McKinnon, P Izulla, J Kimani, R Kaul and A Fauci. "The integrin alpha4beta7 forms a complex with cell-surface CD4 and defines a T-cell subset that is highly susceptible to infection by HIV-1." *Proceedings of the National Academy of Sciences of the United States of America*, vol. 106, 2009, pp. 20882.
12. Chand, S, E L. Messina, W AlSalmi, N Ananthaswamy,

G Gao, G Uritskiy, V Padilla-Sanchez, M Mahalingam, K K. Peachman, M L. Robb, M Rao, and V B. Rao. "Glycosylation and oligomeric state of envelope protein might influence HIV-1 virion capture by α4β7 integrin." *Virology*, 2017, pp. 212.

13. Weissenhorn W, A Dessen, L J. Calder, S C. Harrison, J J. Skehel and D C. Wiley. "Structural basis for membrane fusion by enveloped viruses." *Molecular membrane biology*, vol. 16, 1999, pp. 9.
14. Wilen, Craig B, *et al*. "HIV: Cell Binding and Entry." *Cold Spring Harbor Perspectives in Medicine*, *Cold Spring Harbor Laboratory Press*, 2012.
15. Kuhbandner K, A Hammer, S Haase, E Terbrack, A Hoffmann, A Schippers, N Wagner, R Z. Hussain, W A. Miller-Little, A Y. Koh, J S. Stoolman, B M. Segal and R Linker. "MAdCAM-1-Mediated Intestinal Lymphocyte Homing Is Critical for the Development of Active Experimental Autoimmune Encephalomyelitis." *Frontiers in immunology*, 2019, pp. 903.
16. Merk, A and S Subramaniam. "HIV-1 envelope glycoprotein structure." *Current opinion in structural biology*, vol. 23, 2013, pp. 276.
17. Pettersen, E F., T D. Goddard, C C. Huang, G S. Couch, D M. Greenblatt, E C. Meng, and T E. Ferrin. "UCSF Chimera--a visualization system for exploratory research and analysis." *Journal of computational chemistry*, vol. 25, 2004, pp. 1612.
18. Lyskov S and J J. Gray. "The RosettaDock server for local protein-protein docking." *Nucleic acids research*, vol. 36, 2008, pp. 238.
19. Parker, Zahra F., *et al*. "Transmitted/Founder and Chronic HIV-1 Envelope Proteins Are Distinguished by Differential Utilization of CCR5." *Journal of Virology*, American Society for Microbiology Journals, 2013.
20. Ananthaswamy N, Q Fang, W AlSalmi, S Jain, Z Chen, T Klose, Y Sun, Y Liu, M Mahalingam, S Chand, S Tovanabutra, M L. Robb, M G. Rossmann and V B. Rao. "A sequestered fusion peptide in the structure of an HIV-1 transmitted founder envelope trimer." *Nature communications*, vol. 10, 2019, pp. 873.
21. Yu Y, J Zhu, L Z. Mi, T Walz, H Sun, J Chen, and T A. Springer. "Structural specializations of α(4)β(7), an integrin that mediates rolling adhesion." *The Journal of cell biology*, vol. 196, 2012, pp. 146.
22. Waterhouse A, M Bertoni, S Bienert, G Studer, G Tauriello, R Gumienny, F T. Heer, T Beer, C Rempfer, L Bordoli, R Lepore and T Schwede. "SWISS-MODEL: homology modelling of protein structures and complexes." *Nucleic acids research*, vol. 46, 2018, pp. 303.
23. Schrödinger LLC. "Schrödinger Release 2021-1: Maestro." Schrödinger, 2021.
24. Dando J, K W. Wilkinson, S Ortlepp, D J. King and R L. Brady. "A reassessment of the MAdCAM-1 structure and its role in integrin recognition." *Acta crystallographica*, vol. D, 2002, pp. 241.
25. Yu Y, J Zhu, P S. Huang, J H. Wang, N Pullen, and T A. Springer. "Domain 1 of mucosal addressin cell adhesion molecule has an I1-set fold and a flexible integrin-binding loop." *The Journal of biological chemistry*, vol. 288, 2013, pp. 6294.
26. Chaudhury S, M Berrondo, B D. Weitzner, P Muthu, H Bergman, and J J. Gray. "Benchmarking and analysis of protein docking performance in Rosetta v3.2." *PloS one*, vol. 6, 2011, pp. 22477.
27. Lyskov S and F Chou. "Serverification of molecular modeling applications: the Rosetta Online Server that Includes Everyone (ROSIE)." *PloS one*, vol. 8, 2013, pp. 63906.