

# Computational Modeling of pH-dependent gp120-CD4 Interactions in Founder and Chronic HIV Strains

Jonathan Howton  
jh6w@mtmail.mtsu.edu  
Middle Tennessee State University  
Department of Computer Science  
Murfreesboro, TN 37130, USA

Joshua L. Phillips\*  
Joshua.Phillips@mtsu.edu  
Middle Tennessee State University  
Department of Computer Science  
Murfreesboro, TN 37130, USA

## ABSTRACT

Human Immunodeficiency Virus has been studied for several decades, but a consistently effective vaccine has not yet been produced. While most experimental and computational work in this area has been performed under slightly basic conditions (eg. blood/plasma), viral transmission events often occur under highly acidic conditions (eg. vaginal mucosa). Environmental pH can greatly affect protein structure and epitope exposure to either inhibit or facilitate transmission. We develop a computational pipeline for analyzing the pH sensitivity of protein-protein interactions, and use the pipeline to analyze the transmission-critical interaction between the HIV gp120 and host CD4 proteins. The interaction between gp120 and CD4 is found to be stronger at low pH for all strains tested, which is consistent with previous work and supports the accuracy of the introduced pipeline. Also, early transmitted founder (TF) strains are found to generally bind CD4 better at low pH and are more pH sensitive than systemically circulating chronic control (CC) strains.

## CCS CONCEPTS

• **Applied computing** → **Molecular structural biology**;

## KEYWORDS

HIV; protein-protein interaction; environmental factors; vaccines; antibodies; molecular modeling; gp120; CD4

## 1 INTRODUCTION

More than thirty years after the discovery of Acquired Immune Deficiency Syndrome (AIDS), there is still no vaccine against the Human Immunodeficiency Virus (HIV) that causes the disease [17]. Since HIV has a very high mutation rate, antigenic regions which are targeted by host antibodies vary greatly across HIV virions even within a single host. Therefore, most vaccine research has focused on inducing so-called broadly neutralizing antibodies (bnAbs). The bnAbs are able to target regions of the virus that must be conserved due to functional requirements [4], most of which are found on

\*The corresponding author

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from [permissions@acm.org](mailto:permissions@acm.org).

ACM-BCB'17, August 20-23, 2017, Boston, MA, USA.

© 2017 Copyright held by the owner/author(s). Publication rights licensed to ACM. 978-1-4503-4722-8/17/08...\$15.00

DOI: <http://dx.doi.org/10.1145/3107411.3107506>

the gp120 extracellular subunit of the viral envelope protein (Env) that is responsible for binding CD4 on the surface of host T-Cells to begin infection [31]. Vaccines have been produced from Env fragments that have been computationally optimized to (potentially) invoke the production of bnAbs [11]. Results from these vaccines have varied from successful [3] to unsuccessful [19]. A possible explanation for this inconsistency is that the bnAbs are isolated from the blood, which has a slightly basic pH, while HIV is often transmitted at the mucosa, which are highly acidic. Since protein structure and protein-protein interactions are typically affected by pH, it is likely that the structure of Env and its affinity for other proteins, such as CD4 and bnAbs, is altered. In addition, experimental and computational studies have shown that pH impacts both Env conformation and CD4 binding [29]. In particular, trimeric Env proteins were found to bind CD4 better under acidic conditions, Env surface charge was more complimentary (negative) to the host CD4 surface (positive), and the unbound to bound conformational change was electrostatically more preferable at mucosal pH levels.

Since HIV mutates rapidly within the host, we anticipate that strains in a chronic infection, so called chronic control (CC) strains, will likely have adapted to the systemic pH, and will be less efficient at binding CD4 under acidic conditions when compared to transmitted founder (TF) strains. Consequently, bnAbs produced in chronically infected individuals might be less likely to neutralize HIV transmission at the mucosa. Therefore, it is important to study gp120-CD4 binding under mucosal pH because this conserved interaction is an important target for vaccine production. However, the large variation in gp120 sequence across HIV strains makes experimental studies prohibitive, but computational modeling can help fill this gap in a predictive capacity.

We describe the implementation of a computational modeling pipeline capable of quickly creating and analyzing models of unbound gp120 and the gp120-CD4 interaction for a large number of Env sequences across a wide pH range. A dataset of TF and CC pairs, spanning HIV clades B and C, is used to validate the pipeline and explore several contingent hypotheses: 1) that the Env-CD4 interaction would be strongest at low pH, 2) that the Env protein from TF strains would bind CD4 better under low pH compared to CC strains, and 3) that the Env-CD4 interaction would be more pH sensitive in TF strains. Data obtained from the pipeline is also used to elucidate potential mechanisms responsible for differences between HIV classes.

## 2 BACKGROUND

### 2.1 HIV Variant Subclasses

In most clinical HIV infections, a single TF virion is responsible for the transmission event [16]. TF viruses share common traits that distinguish them from chronic control (CC) strains, and these traits likely enhance TF virus fitness for crossing the mucosal barrier and promoting productive initial infection [23]. Since TF virions are often transmitted at acidic mucosa, it is expected that TF strains are better adapted for transmission at low pH relative to CC strains.

Clades B and C are subgroups of the HIV-1 group major [30]. Clade C is the most prevalent clade of HIV globally, and is the most common in China, India, and Africa, while clade B is most prevalent in America and Europe [1]. Clade C is less virulent [1], which causes a slower progression and a longer asymptomatic period and increases opportunities for transmission [25].

### 2.2 Previous gp120-CD4 Electrostatic Modeling

The gp120 protein was previously modeled using several solved gp120 fragment crystal structures [29]. Partial atomic charges and protonation states were calculated across pH and salinity ranges using the PDB2PQR [7] and the PROPKA3.0 [28] framework. The APBS [2] tool was used to determine surface charge density. Similar approaches have been used to examine protein-protein interactions in available crystal structures [20]. The surface potentials of the CD4 protein and the CD4 binding site of gp120 were found to be complementary at low pH. Also, comparisons between bound and unbound conformations predicted that gp120 exhibits a lower solvation energy for the bound conformation at low pH, effectively priming Env for binding to CD4. Binding assays and electrophoretic measurements using trimeric Env confirmed this hypothesis, suggesting complementary charge density modulated by environmental pH as the primary mechanism. These calculations were facilitated by crystal structures, so only a limited number of sequences could be compared. Also, electrostatic contributions to the binding energy could not be quantified since CD4 is not present in most structures.

Two tools exist for extending the approach above for modeling unsolved gp120 structures. The first is MODELLER: software for homology modeling of protein structures [8, 26]. The second is FRODAN: software for computationally efficient geometric simulation of protein structures which can adapt modeled protein structures to target conformations while preserving stereochemical constraints [10]. There are several solved gp120 structures, such as 1RZK [14] and 2B4C [13], but all models have CD4, an antibody, or both bound. The only unbound model available is 2BF1, which is the simian immunodeficiency virus (SIV) gp120 subunit [5]. Despite HIV/SIV sequence differences, 2BF1 provides a reasonable template for unbound structural alignments with HIV gp120 as evidenced in the study above [29]. APBS [2] can then be used to directly calculate the electrostatic contribution to the binding energy,  $\Delta G_e$ , of the gp120-CD4 complex across a wide pH range.

## 3 METHODS

### 3.1 gp120 Sequences

One TF sequence and one CC sequence were analyzed from each of 24 individuals. Of these 24 pairs of sequences, 18 pairs were B

clade sequences, and 6 were C clade sequences. TF sequences were defined as sequences collected within the first 6 months of infection, while CC sequences were collected after this initial period. Additional sequence details including accession numbers and sequence alignments can be found in [12].

### 3.2 Pipeline Configuration and Automation

Bash 4.1.2, Python 3.4, and R 3.4.0 scripts were used to automate the modeling and analysis of all sequences within the dataset. The pipeline is initialized using sequences and target structures and proceeds in the stepwise fashion described below.

**3.2.1 Structural Modeling.** The analyzed sequences do not have solved structures, so MODELLER [26] was used with a preconstructed set of seven template gp120 structures to produce a set of homology models. The template structures used were 1G9M [18], 1RZK [14], 2B4C [13], 2NY7 [32], 3JWD [22], 3JWO [22], and 3LQA [6]. Ten homology models were produced for each tested sequence to account for natural structural variations in the flexible, variable-loop regions. Ten models were sufficient to assess statistical significance between the different Env classes and sensitivity conditions in this study. However, attempts to discriminate differences between individual Env sequences may require an order of magnitude or more increase in the number of models.

To determine the binding energy of the complex, the electrostatic energy of the complex and the individual components of the complex must be determined. FRODAN is used to adjust the core structure of the models to the three conformations (unbound, bound, and CD4-complex) while preserving stereochemical constraints similar to the approach used by [24]. 2BF1 was the target used to produce the unbound conformations. The bound and CD4-complex conformations were produced using target 1RZK.

**3.2.2 Electrostatic Binding Energy and Charge Density.** Charge and protonation data were determined for all models across a wide range of pH using PDB2PQR 2.0.0 [7] and PROPKA 3.0 [21, 28]. The tested pH range was from 3 to 9 in increments of 0.1. PQR files were generated over the pH range for all model conformations.

Electrostatic energy for each structure was calculated for all of the PQR files using APBS 1.4 [2] to solve the full non-linear Poisson-Boltzmann equation. For each set, the number of grid points, coarse mesh lengths, fine mesh lengths, and known center were calculated using the APBS [2] psize tool. The counter ion (e.g. NaCl) concentration was set to 0.155M for ions with a +1 charge and for ions with a -1 charge. The calculations were carried out with 310K as the system temperature. Surface potential data were saved in DX format for each molecule within a set at whole number pH values.

Binding energies were calculated in two ways. First, the standard bound form binding energy was calculated by subtracting the electrostatic energies of both the CD4 molecule and the bound conformation of gp120 from the electrostatic energy of the gp120-CD4 complex at a given pH. Second, the unbound form binding energy was calculated by subtracting the electrostatic energies of both the CD4 molecule and the unbound conformation of gp120 from the electrostatic energy of the gp120-CD4 complex at a given pH. The solvation energy difference was found by subtracting the

electrostatic energy of the bound gp120 conformation from the electrostatic energy of the unbound gp120 conformation.

**3.2.3 Statistical Analysis.** Binding energy sensitivity was determined as the binding energy at low-pH (3.5, 3.6, 3.7, 3.8, 3.9, 4.0, 4.1, 4.2, 4.3, 4.4, and 4.5) subtracted from the binding energy at high-pH (7.0, 7.1, 7.2, 7.3, 7.4, 7.5, 7.6, 7.7, 7.8, 7.9, and 8.0), respectively. This produces 11 binding energy sensitivity values for each of the 10 models within each sequence. Individual sequence sensitivity was determined by pooling the 11 sensitivities from all 10 models within a sequence and creating a boxplot from these values. Group sensitivity was determined by finding the median sensitivity across the 10 models within each sequence. The 11 median sensitivity values for each sequence were then pooled with the sensitivity values from all sequences within a corresponding group to produce group sensitivity boxplots. The pH sensitivity of solvation energy differences between the bound and unbound conformations was calculated similarly, but solvation energy differences at the indicated pH values were used in place of binding energies. Significant differences between median values for these and all subsequent results were determined using the Mann-Whitney test.

Whole molecule charge density was calculated as the sum of all charges determined by APBS divided by the total solvent accessible surface area, which was determined using VMD [15]. The median charge density was determined within each group at pH 4, 5, 7, and 8. Residue-specific charge density was calculated as the sum of all charges determined by APBS that were on the surface of the residue, divided by the solvent accessible surface area of the residue. VMD [15] was used to assign electrostatic charge coordinates to corresponding residues and to determine the solvent accessible surface area of each residue.

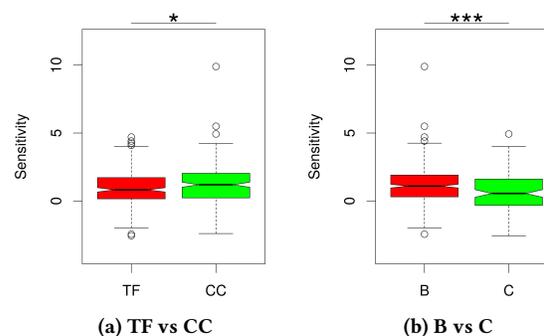
To determine pH sensitivity of whole molecule charge density, first charge density differences were determined by subtracting the unbound charge density from the bound charge density at each pH value. These charge density differences were directly compared. The sensitivity was calculated as this value at pH 4 subtracted from this value at pH 7, this value at pH 5 subtracted from this value at pH 8, or the average of these two sensitivities.

**3.2.4 Sequence Alignment-Based Model Fitting.** Next, identification of sequence motifs which correlate with differences between residue-specific surface charge density was performed. Clustal Omega [27] was used to align all analyzed sequences. Within each group tested, charges were assigned to alignment positions in two ways as described in [12]. Groups were compared by subtracting each assigned value from the corresponding value in the other group. The top 1% of residues were identified as residues that had an absolute sensitivity difference larger than 1% of the absolute sum of residue sensitivity differences in a given comparison.

## 4 RESULTS

The pH sensitivity of gp120 charge density was found to be significantly more sensitive in the CC strains when compared to TF strains (Figure 1a). Additionally, B clade strains were found to be significantly more sensitive than C clade strains (Figure 1b).

When the unbound structure was not considered, the gp120-CD4 interaction was significantly more favorable (lower binding



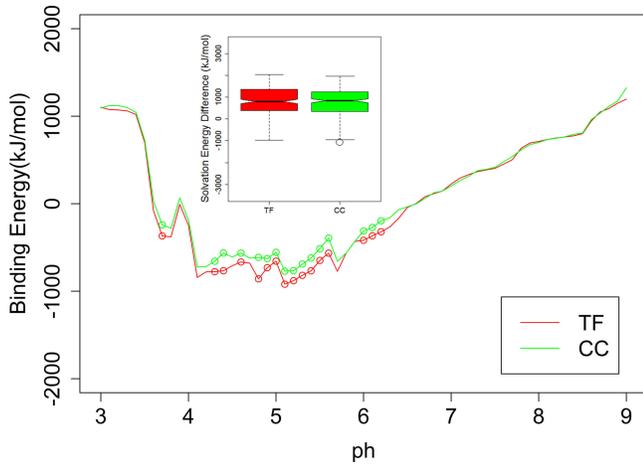
**Figure 1: Charge density sensitivity comparisons between (a) TF and CC classes and (b) B and C clades. (\*  $p < 0.05$ , \*\*\*  $p < 0.001$ )**

energy values) for TF strains than CC strains at many pH values between pH 4 and 6.5 (Figure 2a). However, there was no significant difference when comparing the pH sensitivity of this interaction (Figure 2a inset boxplot). Considering the energy contribution of the unbound to bound conformational change altered the shape of the binding energy curve, this calculation suggests that CC strains bind CD4 better at pH values between 6.5 and 8.5. However, this was only significant at pH 7.9, 8, and 8.1 (Figure 2b). The binding energy in TF strains was found to be significantly more sensitive to pH than CC strains when considering this conformational change (Figure 2b inset boxplot). These results appear to be most representative within the B clade (data not shown; reported in [12]).

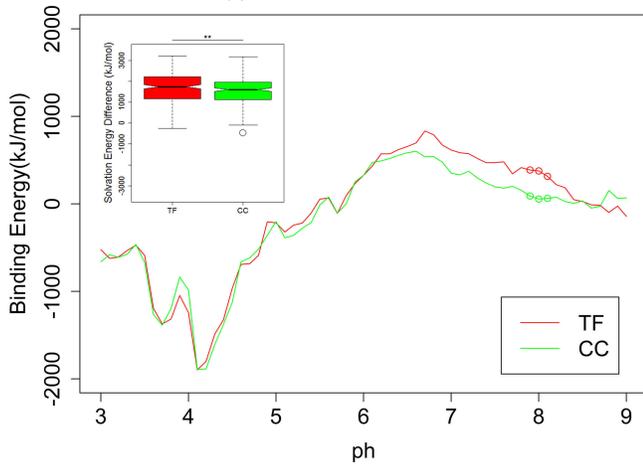
The unbound binding energy calculation was used to compare B and C clades within each class (Figure 3). Within the TF class, the only significant difference was at pH 9, though there was a general trend of C binding CD4 better at pH values above 6.5 (Figure 3a). There was no significant difference in sensitivity within TF (Figure 3a inset boxplot). Within the CC class, the only significant difference found was at pH 3.7 (Figure 3b). However, CD4 binding was found to be significantly more pH sensitive in C clade when compared to B clade (Figure 3b inset boxplot).

The difference in solvation energy between the bound and unbound conformations was used to determine the conformation that was most energetically favorable at each pH. A difference above zero indicates that the bound conformation is more favorable, while a value below zero indicates that the unbound conformation is more favorable. Within B clade, the CC class had a significantly more positive solvation energy difference than the TF class at many pH values above 6, as well as pH 4.7 and 4.8 (Figure 4a). This conformation preference was found to be significantly more sensitive to pH in the TF class (Figure 4a inset boxplot). Within the CC class, there were no significant differences in solvation energy difference at any pH value between B and C clades, though there was a trend of B clade being more positive at pH values from 5 to 8 (Figure 4b). The conformation preference was found to be significantly more pH sensitive in C clade (Figure 4b inset boxplot).

Residue-specific charge density pH sensitivity was determined and the difference in sensitivity between corresponding groups was calculated as described above. The most significant difference in binding energy between TF and CC strains was found within B



(a) Bound Calculation



(b) Unbound Calculation

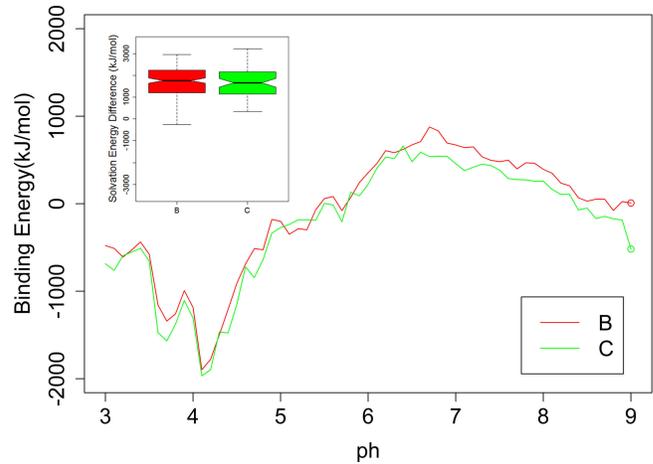
**Figure 2: Median binding energy comparison between TF and CC strains across pH range using (a) bound calculation and (b) unbound calculation. Significantly different binding energies are identified by corresponding points. Boxplots compare pH sensitivity of binding energy between TF (left) and CC (right). (\*\*  $p < 0.01$ )**

clade, so the residue sensitivity was compared within this clade. Differences were found at residues 343 and 402. However, for 402 these differences only occur in a single TF and a single CC strain, and is not representative of the groups overall (see [12]).

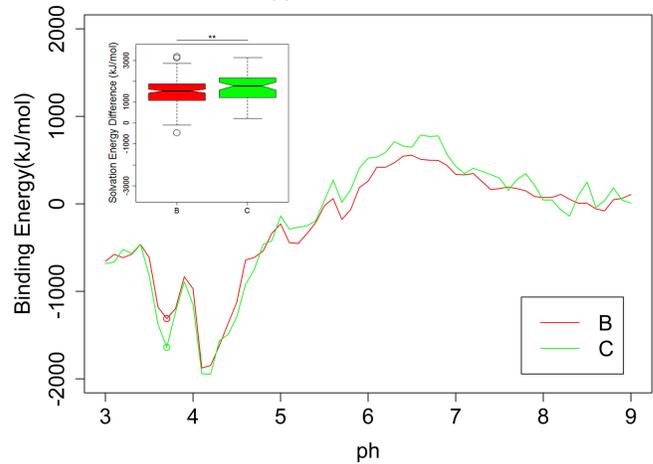
Residues with the largest sensitivity difference were also identified by comparing B and C clades within the CC class. Only a single residue was identified (474), and it was also poorly representative of the overall sequences as indicated by low bit scores (see [12]).

## 5 DISCUSSION

Though AIDS and HIV have been studied for several decades, a viable vaccine has yet to be produced. Since the significance of the acidic pH of the typical mucosal transmission site has been broadly



(a) Within TF



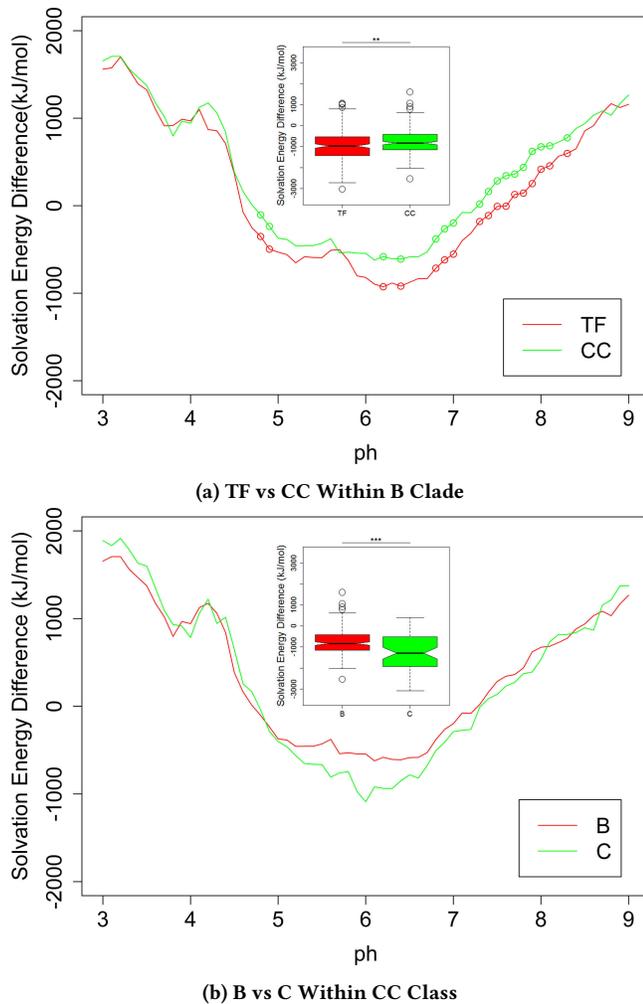
(b) Within CC

**Figure 3: Median binding energy comparison between B and C clades across pH range using the unbound calculation (a) within the TF class and (b) within the CC class. Significantly different binding energies are identified by corresponding points. Boxplots compare pH sensitivity of binding energy between B (left) and C (right). (\*\*  $p < 0.01$ )**

overlooked, we constructed a pipeline to analyze the pH sensitivity of the gp120-CD4 interaction in TF and CC strains.

The charge density of CC strains was found to be more pH sensitive than TF strains (Figure 1a). Also, the surface charge of B clade sequences was significantly more sensitive to pH than that of C clade sequences (Figure 1b). While no difference between CC and TF strains nor B and C clades was reported in previous work [29], these difference are most likely due to the increased number of sequences analyzed here.

Calculations using the bound conformation found TF strains to bind CD4 significantly better than CC strains at low pH (Figure 2a). At high pH values, CC was found to bind CD4 better in the calculation using the unbound gp120 conformation (Figure 2b). Additional analysis (data not shown, see [12]) showed that this was



**Figure 4: Median solvation energy difference between bound and unbound conformations across pH range compared between (a) TF and CC classes within B clade and (b) B and C clades within the CC class. Significantly different binding energies are identified by corresponding points. Boxplots compare pH sensitivity of binding energy between B (left) and C (right). (\*\*  $p < 0.01$ , \*\*\*  $p < 0.001$ )**

found to be significant within B clade while the CC class within B clade also significantly prefers the unbound conformation at both high and low (but not intermediate) pH. These results suggest that the increased CD4 binding at low pH in TF strains is not due to increased pressure to assume the bound conformation, but is more likely due to a more favorable interaction between gp120 and CD4. Conversely, the increased binding ability of CC strains at higher pH values appears to be influenced by an increased preference to assume the bound conformation.

Using the unbound calculation, there was a trend within class TF in which C clade bound CD4 better at pH values above 6.5 (Figure 3a). Additional analysis (see [12]) revealed that B clade preferred the unbound conformation over C clade from pH 5 to 7 within

the overall group and within the CC class, but C clade prefers the unbound conformation within class TF from pH 6 to 8. The trend within class TF suggests that C clade may bind better at higher pH due to a preference for the bound conformation. However, this is based upon an observed trend, so a larger number of C clade samples would be needed to evaluate its significance. Regardless, across all binding energy calculations, gp120 bound CD4 better at low pH, consistent with previous experimental results [29].

When calculating sensitivity with the bound gp120 conformations, no significant differences could be found between TF and CC (Figure 2). However, when using the unbound conformation for the calculation, the TF gp120-CD4 interaction was significantly more sensitive to pH (Figure 2b), particularly within B clade [12]. This was consistent with previous experimental results [29]. Additionally, the preference for the bound conformation was significantly more pH sensitive for TF strains within the B clade (Figure 4a). This suggests that the increased binding energy pH sensitivity of the TF strains within B clade is due to the effect of pH on the gp120 conformation, and that modeling of the unbound conformation was integral to these results.

Within the CC class, CD4 binding was found to be significantly more sensitive in C clade when using the unbound conformation (Figure 3b). Also within CC, the preference for the bound conformation was found to be significantly more sensitive in C clade (Figure 4b). These results suggest that within the CC class, pH affects CD4 binding through the conformational shift in C clade.

Efforts to understand a mechanism of binding sensitivity identified a few residues that may contribute to the observed differences (see results and additional data reported in [12]). Unfortunately, sequence comparisons did not indicate any clear sequence difference that could contribute to the observed sensitivity differences.

Overall, this work shows the importance of the pH sensitivity mechanism and its potential role in the gp120-CD4 interaction. Several subclass-specific mechanisms were identified using the pipeline even though residue-specific pH sensitivity provided no additional insights. This is particularly important for HIV vaccine research because the CD4 binding site is an important vaccine target, and pH has been shown to affect antibody binding at the mucosa [9]. Additionally, bnAbs typically target the CD4 binding site of gp120 [31]. Investigations into the pH sensitivity of gp120-bnAb interactions using the pipeline are currently ongoing.

Additionally, this work shows the effectiveness of the proposed pipeline for analyzing pH sensitivity for protein-protein interactions. The pipeline builds upon previous approaches for computing electrostatic potential across a range of environmental conditions [20, 29] by integrating high-throughput structural modeling, conformational search, and targeted docking for a large set of sequences [8, 10, 24]. Computed gp120-CD4 binding energy sensitivities were also consistent with previous work [29].

## REFERENCES

- [1] Kevin K Ariën, Guido Vanham, and Eric J Arts. 2007. Is HIV-1 evolving to a less virulent form in humans? *Nature Reviews. Microbiology* 5, 2 (2007), 141–51. DOI: <https://doi.org/10.1038/nrmicro1594>
- [2] N A Baker, D Sept, S Joseph, M J Holst, and J A McCammon. 2001. Electrostatics of nanosystems: application to microtubules and the ribosome. *Proceedings of the National Academy of Sciences of the United States of America* 98, 18 (2001), 10037–41. DOI: <https://doi.org/10.1073/pnas.181342398>

- [3] Dan H Barouch, Kathryn E Stephenson, Erica N Borducchi, Kaitlin Smith, Kelly Stanley, Anna G McNally, Jinyan Liu, Peter Abbink, Lori F Maxfield, Michael S Seaman, Anne-Sophie Dugast, Galit Alter, Melissa Ferguson, Wenjun Li, Patricia L Earl, Bernard Moss, Elena E Giorgi, James J Zsinger, Leigh Anne Eller, Erik A Billings, Mangala Rao, Sodsai Tovanabutra, Eric Sanders-Buell, Mo Weijjens, Maria G Pau, Hanneke Schuitemaker, Merlin L Robb, Jerome H Kim, Bette T Korber, and Nelson L Michael. 2013. Protective efficacy of a global HIV-1 mosaic vaccine against heterologous SHIV challenges in rhesus monkeys. *Cell* 155, 3 (Oct 2013), 531–9. DOI: <https://doi.org/10.1016/j.cell.2013.09.061>
- [4] Dennis R Burton, Pascal Poignard, Robyn L Stanfield, and Ian A Wilson. 2012. Broadly neutralizing antibodies present new prospects to counter highly antigenically diverse viruses. *Science (New York, N.Y.)* 337, 6091 (Jul 2012), 183–6. DOI: <https://doi.org/10.1126/science.1225416>
- [5] Bing Chen, Erik M Vogan, Haiyun Gong, John J Skehel, Don C Wiley, and Stephen C Harrison. 2005. Structure of an unliganded simian immunodeficiency virus gp120 core. (2005). DOI: <https://doi.org/10.1038/nature03327>
- [6] Ron Diskin, Paola M Marcovecchio, and Pamela J Bjorkman. 2010. Structure of a clade C HIV-1 gp120 bound to CD4 and CD4-induced antibody reveals anti-CD4 polyreactivity. *Nature Structural Molecular Biology* 17, 5 (May 2010), 608–13. DOI: <https://doi.org/10.1038/nsmb.1796>
- [7] Todd J. Dolinsky, Paul Czodrowski, Hui Li, Jens E. Nielsen, Jan H. Jensen, Gerhard Klebe, and Nathan A. Baker. 2007. PDB2PQR: Expanding and upgrading automated preparation of biomolecular structures for molecular simulations. *Nucleic Acids Research* 35, SUPPL.2 (2007), 522–525. DOI: <https://doi.org/10.1093/nar/gkm276>
- [8] Narayanan Eswar, Ben Webb, Marc A. Marti-Renom, M.S. Madhusudhan, David Eramian, Min-Yi Shen, Ursula Pieper, and Andrej Sali. 2002. *Comparative protein structure modeling using modeller*. John Wiley & Sons, Inc. DOI: <https://doi.org/10.1002/0471250953.bi0506s15>
- [9] Kelly M. Fahrback, Olga Malykhina, Daniel J. Stieh, and Thomas J. Hope. 2013. Differential Binding of IgG and IgA to Mucus of the Female Reproductive Tract. *PLOS ONE* 8, 10 (10 2013), 1–11. DOI: <https://doi.org/10.1371/journal.pone.0076176>
- [10] Daniel W. Farrell, Kirill Speranskiy, and M. F. Thorpe. 2010. Generating stereochemically acceptable protein pathways. *Proteins: Structure, Function and Bioinformatics* 78, 14 (2010), 2908–2921. DOI: <https://doi.org/10.1002/prot.22810>
- [11] Will Fischer, Simon Perkins, James Theiler, Tanmoy Bhattacharya, Karina Yusim, Robert Funkhouser, Carla Kuiken, Barton Haynes, Norman L Letvin, Bruce D Walker, Beatrice H Hahn, and Bette T Korber. 2007. Polyvalent vaccines for optimal coverage of potential T-cell epitopes in global HIV-1 variants. *Nature Medicine* 13, 1 (Jan 2007), 100–6. DOI: <https://doi.org/10.1038/nm1461>
- [12] Jonathan Howton. 2017. *A Computational Electrostatic Modeling Pipeline for Comparing pH-dependent gp120-CD4 Interactions in Founder and Chronic HIV Strains*. Master's thesis. Middle Tennessee State University, Murfreesboro, TN. <http://jewelscholar.mtsu.edu/xmlui/handle/mtsu/5324>
- [13] Chih-Chin Huang, Min Tang, Mei-Yun Zhang, Shahzad Majeed, Elizabeth Montabana, Robyn L Stanfield, Dimiter S Dimitrov, Bette Korber, Joseph Sodroski, Ian A Wilson, Richard Wyatt, and Peter D Kwong. 2005. Structure of a V3-containing HIV-1 gp120 core. *Science (New York, N.Y.)* 310, 5750 (Nov 2005), 1025–8. DOI: <https://doi.org/10.1126/science.1118398>
- [14] C C Huang, M Venturi, S Majeed, M J Moore, S Phogat, M Y Zhang, D S Dimitrov, W A Hendrickson, J Robinson, J Sodroski, R Wyatt, H Choe, M Farzan, and P D Kwong. 2004. Structural basis of tyrosine sulfation and VH-gene usage in antibodies that recognize the HIV type 1 coreceptor-binding site on gp120. *Proc Natl Acad Sci U S A* 101, 9 (2004), 2706–2711. DOI: <https://doi.org/10.1073/pnas.0308527100>
- [15] William Humphrey, Andrew Dalke, and Klaus Schulten. 1996. VMD – Visual Molecular Dynamics. *Journal of Molecular Graphics* 14 (1996), 33–38.
- [16] Brandon F Keele, Elena E Giorgi, Jesus F Salazar-Gonzalez, Julie M Decker, Kimmy T Pham, Maria G Salazar, Chuanxi Sun, Truman Grayson, Shuyi Wang, Hui Li, Xiping Wei, Chunlai Jiang, Jennifer L Kirchherr, Feng Gao, Jeffery A Anderson, Li-Hua Ping, Ronald Swanstrom, Georgia D Tomaras, William A Blattner, Paul A Goepfert, J Michael Kilby, Michael S Saag, Eric L Delwart, Michael P Busch, Myron S Cohen, David C Montefiori, Barton F Haynes, Brian Gaschen, Gayathri S Athreya, Ha Y Lee, Natasha Wood, Cathal Seoighe, Alan S Perelson, Tanmoy Bhattacharya, Bette T Korber, Beatrice H Hahn, and George M Shaw. 2008. Identification and characterization of transmitted and early founder virus envelopes in primary HIV-1 infection. *Proceedings of the National Academy of Sciences of the United States of America* 105, 21 (2008), 7552–7. DOI: <https://doi.org/10.1073/pnas.0802203105>
- [17] Bette Korber and S. Gnanakaran. 2011. Converging on an HIV Vaccine. *Science* 333, 6049 (2011), 1589–1590. DOI: <https://doi.org/10.1126/science.1211919>
- [18] P D Kwong, R Wyatt, S Majeed, J Robinson, R W Sweet, J Sodroski, and W A Hendrickson. 2000. Structures of HIV-1 gp120 envelope glycoproteins from laboratory-adapted and primary isolates. *Structure (London, England : 1993)* 8, 12 (Dec 2000), 1329–39. <http://www.ncbi.nlm.nih.gov/pubmed/11188697>
- [19] Hua-Xin Liao, Mattia Bonsignori, S Munir Alam, Jason S McLellan, Georgia D Tomaras, M Anthony Moody, Daniel M Kozink, Kwan-Ki Hwang, Xi Chen, Chun-Yen Tsao, Pinghuang Liu, Xiaozhi Lu, Robert J Parks, David C Montefiori, Guido Ferrari, Justin Pollara, Mangala Rao, Kristina K Peachman, Sampa Santra, Norman L Letvin, Nicos Karasavvas, Zhi-Yong Yang, Kaifan Dai, Marie Pancera, Jason Gorman, Kevin Wiehe, Nathan I Nicely, Supachai Rerksgarm, Sorachai Nitayaphan, Janarit Kaewkungwal, Punnee Pitisuttithum, James Tartaglia, Faruk Sinangil, Jerome H Kim, Nelson L Michael, Thomas B Kepler, Peter D Kwong, John R Mascola, Gary J Nabel, Abraham Pinter, Susan Zolla-Pazner, and Barton F Haynes. 2013. Vaccine induction of antibodies against a structurally heterogeneous site of immune pressure within HIV-1 envelope protein variable regions 1 and 2. *Immunity* 38, 1 (Jan 2013), 176–86. DOI: <https://doi.org/10.1016/j.immuni.2012.11.011>
- [20] Aaron C. Mason and Jan H. Jensen. 2008. Protein-protein binding is often associated with changes in protonation state. *Proteins: Structure, Function and Genetics* 71, 1 (2008), 81–91. DOI: <https://doi.org/10.1002/prot.21657>
- [21] Mats H M Olsson, Chresten R Søndergaard, Michal Rostkowski, and Jan H Jensen. 2011. PROPKA3: Consistent Treatment of Internal and Surface Residues in Empirical pKa Predictions. *Journal of Chemical Theory and Computation* 7, 2 (2011), 525–537. DOI: <https://doi.org/10.1021/ct100578z>
- [22] Marie Pancera, Shahzad Majeed, Yih-En Andrew Ban, Lei Chen, Chih-chin Huang, Leopold Kong, Young Do Kwon, Jonathan Stuckey, Tongqing Zhou, James E Robinson, William R Schief, Joseph Sodroski, Amit Kumar, Richard Wyatt, and Peter D Kwong. 2010. Structure of HIV-1 gp120 with gp41-interactive region reveals layered envelope architecture and basis of conformational mobility. *Proceedings of the National Academy of Sciences of the United States of America* 107, 3 (Jan 2010), 1166–71. DOI: <https://doi.org/10.1073/pnas.0911004107>
- [23] Nicholas F Parrish, Feng Gao, Hui Li, Elena E Giorgi, Hannah J Barbian, Erica H Parrish, Lara Zajic, Shilpa S Iyer, Julie M Decker, Amit Kumar, Bhavna Hora, Anna Berg, Fangping Cai, Jennifer Hopper, Thomas N Denny, Haitao Ding, Christina Ochsenbauer, John C Kappes, Rachel P Galimidi, Anthony P West, Pamela J Bjorkman, Craig B Wilen, Robert W Doms, Meagan O'Brien, Nina Bhardwaj, Persephone Borrow, Barton F Haynes, Mark Muldoon, James P Theiler, Bette Korber, George M Shaw, and Beatrice H Hahn. 2013. Phenotypic properties of transmitted founder HIV-1. *Proceedings of the National Academy of Sciences of the United States of America* 110, 17 (2013), 6626–33. DOI: <https://doi.org/10.1073/pnas.1304288110>
- [24] Joshua L. Phillips and S. Gnanakaran. 2015. A data-driven approach to modeling the tripartite structure of multidrug resistance efflux pumps. *Proteins: Structure, Function and Bioinformatics* 83, 1 (2015), 46–65. DOI: <https://doi.org/10.1002/prot.24632>
- [25] Indianara Rotta and Sérgio Monteiro de Almeida. 2011. Genotypical diversity of HIV clades and central nervous system impairment. *Arquivos de Neuro-psiquiatria* 69, 6 (Dec 2011), 964–72. <http://www.ncbi.nlm.nih.gov/pubmed/22297889>
- [26] A Sali and T L Blundell. 1993. Comparative protein modelling by satisfaction of spatial restraints. *Journal of Molecular Biology* 234, 3 (1993), 779–815. DOI: <https://doi.org/10.1006/jmbi.1993.1262>
- [27] Fabian Sievers, Andreas Wilm, David Dineen, Toby J Gibson, Kevin Karplus, Weizhong Li, Rodrigo Lopez, Hamish McWilliam, Michael Remmert, Johannes Söding, Julie D Thompson, and Desmond G Higgins. 2011. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Molecular Systems Biology* 7, 1 (2011). DOI: <https://doi.org/10.1038/msb.2011.75>
- [28] Chresten R. Søndergaard, Mats H M Olsson, Michal Rostkowski, and Jan H. Jensen. 2011. Improved treatment of ligands and coupling effects in empirical calculation and rationalization of pKa values. *Journal of Chemical Theory and Computation* 7, 7 (2011), 2284–2295. DOI: <https://doi.org/10.1021/ct200133y>
- [29] Daniel J. Stieh, Joshua L. Phillips, Paul M. Rogers, Deborah F. King, Gianguido C. Cianci, Simon A. Jeffs, Sandrasegaram Gnanakaran, and Robin J. Shattock. 2013. Dynamic electrophoretic fingerprinting of the HIV-1 envelope glycoprotein. *Retrovirology* 10, 1 (2013), 33. DOI: <https://doi.org/10.1186/1742-4690-10-33>
- [30] Barbara S Taylor, Magdalena E Sobieszczyk, Francine E McCutchan, and Scott M Hammer. 2008. The challenge of HIV-1 subtype diversity. *The New England Journal of Medicine* 358, 15 (Apr 2008), 1590–602. DOI: <https://doi.org/10.1056/NEJMra0706737>
- [31] R Wyatt, P D Kwong, E Desjardins, R W Sweet, J Robinson, W A Hendrickson, and J G Sodroski. 1998. The antigenic structure of the HIV gp120 envelope glycoprotein. *Nature* 393, 6686 (Jun 1998), 705–11. DOI: <https://doi.org/10.1038/31514>
- [32] Tongqing Zhou, Ling Xu, Barna Dey, Ann J Hessel, Donald Van Ryk, Shi-Hua Xiang, Xinzheng Yang, Mei-Yun Zhang, Michael B Zwick, James Arthos, Dennis R Burton, Dimiter S Dimitrov, Joseph Sodroski, Richard Wyatt, Gary J Nabel, and Peter D Kwong. 2007. Structural definition of a conserved neutralization epitope on HIV-1 gp120. *Nature* 445, 7129 (Feb 2007), 732–7. DOI: <https://doi.org/10.1038/nature05580>