Infectious Disease Modelling 10 (2025) 287–301

Contents lists available at ScienceDirect

Infectious Disease Modelling

journal homepage: www.keaipublishing.com/idm

Stronger binding affinities of *gp120/CD4* in *Catarrhini* provide insights into HIV/host interactions

Vladimir Li [a, 1], Chul Lee [b, 1], TaeHyun Park [c], Erich D. Jarvis [b, d, **], Heebal Kim [a, e, f, *]

[a] *Interdisciplinary Program in Bioinformatics, Seoul National University, Seoul, Republic of Korea*
[b] *Laboratory of Neurogenetics of Language, The Rockefeller University, New York, NY, USA*
[c] *Department of Anesthesiology, Weill Cornell Medical College, New York, NY 10065, USA*
[d] *Department of Neurobiology, Howard Hughes Medical Institute (HHMI), and Duke University Medical Center, Durham, NC, 27710, USA*
[e] *Department of Agricultural Biotechnology and Research Institute for Agriculture and Life Sciences, Seoul National University, Seoul, Republic of Korea*
[f] *eGnome, Seoul, Republic of Korea*

A R T I C L E I N F O

Article history:
Received 5 June 2024
Received in revised form 29 September 2024
Accepted 18 October 2024
Available online 19 October 2024
Handling Editor: Dr. Raluca Eftimie

Keywords:
HIV
Pandemic zoonotic infection
Virus-host interaction
Computational binding affinities
Positive selection

A B S T R A C T

Human immunodeficiency virus-1 (HIV-1) exploits the viral *gp120* protein and host *CD4*/*CCR5* receptors for the pandemic infection to humans. The host co-receptors of not only humans but also several primates and HIV-model mice can interact with the HIV receptor. However, the molecular mechanisms of these interactions remain unclear. Using Shaik et al. (2019)'s *gp120/CD4/CCR5* structure of HIV-1B and human, here, we investigate the molecular dynamics between HIV sub-lineages (B, C, N, and O) and potential hosts in *Euarchontoglires* (primates and rodents). Although both host genes show similar protein structures conserved in all animals, *CD4* gene demonstrates significantly stronger binding affinities in *Catarrhini* (apes and Old-World monkeys). Its known candidate residues interacted with gp120 fail to explain these affinity variations. Therefore, we identified novel candidate sites under positive selection on the *Catarrhini* lineage. Among four positively selected sites, residue R58 in humans is located within an antigen-antibody binding domain, exhibiting apomorphic amino acid substitutions as Arginine (R) in *Catarrhini*, which are mutually exclusive to the other animals where Lysine (K) is prevalent. Applying for artificial mutation test, we validated that K to R substitutions can lead stronger binding affinities of *Catarrhini*. Ecologically, these dynamics may relate to shared equatorial habitats in Africa and Asia. Our findings suggest a new candidate site R58 driven by the lineage-specific evolution as a molecular foundation on HIV infection.

* Corresponding author. Interdisciplinary Program in Bioinformatics, Seoul National University, Seoul, Republic of Korea.
** Corresponding author. Laboratory of Neurogenetics of Language, The Rockefeller University, New York, NY, USA.
E-mail addresses: ejarvis@rockefeller.edu (E.D. Jarvis), heebal@snu.ac.kr (H. Kim).
Peer review under responsibility of KeAi Communications Co., Ltd.
[1] Equally contributed.

https://doi.org/10.1016/j.idm.2024.10.003

1. Introduction

Human immunodeficiency virus (HIV), a member of the Retroviridae family, is the cause of acquired immunodeficiency syndrome (AIDS) on a global scale (Bbosa et al., 2019; Sharp and Hahn). By the close of 2022, around 39 million individuals had been affected by the virus, marking its pandemic prevalence (Sharp and Hahn; Number of people, 2021; HIV data and statistics, 2023). The classification of HIV-1 into M (main), N (non-M, non-O), O (outlier), and P groups highlights its genetic diversity, exhibiting varying genetic distances among subtypes and circulating recombinant forms within the main group (Bbosa et al., 2019).

Critical to HIV infection is the formation of a ternary complex (Berger et al., 1999; Bour et al., 1995; Tamamis & Floudas, 2014). The viral *gp120*, derived from the cleavage of *gp160* into *gp120* and *gp41*, intricately interacts with host *CD4* and *CCR5* co-receptors, initiating the infection process (Berger et al., 1999; Bour et al., 1995; Tamamis & Floudas, 2014). A pivotal advancement in understanding this process was achieved by Shaik et al., who unveiled the 3D structure of the HIV *gp120/CD4/CCR5* co-receptor complex using cryo-electron microscopy [30]. This complex serves as a crucial template for investigating molecular dynamics, particularly the impact of amino acid evolution on these interactions (Shaik et al., 2019). HIV originated from the simian immunodeficiency virus (SIV) found in chimpanzees and evolved to become a virus specific to humans (Sharp and Hahn; Gao et al., 1999; Hirsch et al., 1989).

HIV, while originating from Simian Immunodeficiency Virus (SIV), is largely specific to humans, and infection rarely progresses in non-human organisms. However, past research has shown that HIV can interact with certain primate species, such as chimpanzees, gibbons, and pig-tailed macaques (Agy et al., 1992; Lusso et al., 1988; Mwaengo & Novembre, 1998; Novembre et al., 1997). For instance, one well-documented case involved a chimpanzee (C499) that developed AIDS approximately 10 years after initial HIV exposure, illustrating the potential for disease progression in primates despite its rarity (Novembre et al., 1997). Although HIV typically does not cause disease in non-human primates, the initial interaction between the virus's *gp120* protein and host co-receptors, such as *CD4* and *CCR5*, can offer valuable insights into the evolution of viral infectivity across species.

Studies by Warren et al. further demonstrated that the *CD4* receptors of species such as white-cheeked gibbons, olive baboons, geladas, black snub-nosed monkeys, colobus, and owl monkeys exhibit interactions with HIV-1 *gp120* that are similar to those observed in humans (Warren et al., 2019). These findings suggest that some non-human primates, which are phylogenetically close to human may possess co-receptor structures conducive to HIV binding. Conversely, rodents exhibit minimal interaction with HIV, which has led to the development of humanized rodent models for HIV research (Gorantla et al., 2012; Maanen & Sutton, 2003).

Many previous studies have been limited to *in vitro* analyses or small sample sizes, leaving gaps in our understanding of how evolutionary pressures shape HIV susceptibility across a broader range of primate species. This study aims to address these gaps by using computational modeling to predict the evolutionary boundaries of HIV infectivity across the *Catarrhini* lineage, utilizing homology modeling and protein docking simulations to explore interactions between *gp120*, *CD4,* and *CCR5*. In doing so, this research also examines the taxonomic boundaries of HIV-host interactions across *Euarchontoglires*, including both primates and rodents. Although rodents are not naturally infected with HIV, including them as negative controls provides valuable insights into virus-host interactions. By applying comparative proteomic approaches and analyzing the overall protein structures and binding affinities of the *gp120/CD4* and *gp120/CCR5* complexes based on expanded genome assemblies, this study aims to enhance our understanding of the molecular mechanisms driving stronger binding affinities in certain species. Additionally, novel candidate sites under episodic adaptation have been identified in functional protein domains, while ecological factors are explored to better understand clade-specific evolution in relation to viral infection.

2. Materials and methods

An outline of the workflow of this study is presented in Supplementary Fig. 1.

2.1. Data collections of HIV subtypes and host species

We obtained four pertinent HIV-1 subtypes (B, C, N, and O) from the Los Alamos National Laboratory's Pathogen Research Database (Los Alamos National Laboratory, 2021). Detailed information on these viral strains is available in the Supplementary Table S1. Organisms were meticulously chosen for analysis – 28 species for *CD4* and 24 for *CCR5* – based on their representative positions within the ENSEMBL species tree (Howe et al., 2021). The comprehensive list of these selected species is provided in the Supplementary Table S2. Key sites in *CD4* and *CCR5* crucial for interactions with *gp120* were thoughtfully selected, guided by their pivotal role in protein-protein docking, as elucidated in previous studies (Fontenot et al., 2007; Krykbaev et al., 1997; Moebius et al., 1992; Prévost et al., 2020; Shaik et al., 2019). These include positions 10, 14, and 15 for *CCR5*, and positions 43, 46, and 59 for *CD4* (Fontenot et al., 2007; Krykbaev et al., 1997; Moebius et al., 1992; Prévost et al., 2020; Shaik et al., 2019).

2.2. Sequence alignment

CCR5 and *CD4* coding sequences were aligned using the webPRANK software with subsequent manual refinement (Löytynoja & Goldman, 2010). Notably, raw *CD4* sequences encompassed a region outside the protein's 3D structure; however, residue numbering adhered to the raw sequence files. For the orangutan *CD4*, the ENSEMBL sequence was incomplete and thus truncated, lacking 2/3 of the key residues, whereas the NCBI sequence was more complete (Chaudhury & Gray, 2008; Federhen, 2012; Howe et al., 2021). We thus used the NCBI sequence (Supplementary Table S2). Conservation scores of primate and rodent clades in *CD4* and *CCR5* were calculated using Jalview (Fig. 1d, Supplementary Table S4) (Waterhouse et al., 2009).

2.3. Protein 3D structures estimation

The protein sequences were also obtained from ENSEMBL, and 3D structures were generated through the SWISS-MODEL homology modeling server (Waterhouse et al., 2018). The template structure (PDB entry: 6MET) was retrieved directly from the PDB (Shaik et al., 2019). Subsequent refinement of the resulting structures was conducted using the Rosetta Relax protocol (Bradley et al., 2005; Misura & Baker, 2005). We conducted thorough analysis and visualization of all protein models using PyMOL, and further validation of the models was performed using the ProSA-Web server (Supplementary Table S3) (Wiederstein & Sippl, 2007; The PyMOL Molecular Graphics System). ProSA-Web evaluates the quality of a protein model by comparing it with experimentally obtained protein structures (NMR, X-ray) deposited in the PDB. (Wiederstein & Sippl, 2007).

2.4. Protein-protein docking and binding affinity

The relaxed protein model pairs were initially oriented to face each other at their docking sites using PyMOL, followed by the application of the RosettaDock, a Monte Carlo-based multi-scale docking algorithm with two main stages: a low-resolution centroid-mode and a high-resolution all-atom refinement (The PyMOL Molecular Graphics System; Sircar et al., 2010; Chaudhury et al., 2011). Initially, proteins are represented coarsely, using a 500-step Monte Carlo search to optimize their orientation and side-chain conformation. After identifying the lowest energy structure, a high-resolution refinement follows, incorporating 50 Monte Carlo + Minimization (MCM) steps to further optimize rigid-body orientation and side-chain conformations. We employed Monte Carlo methods for local sampling of potential binding modes between the two proteins, incorporating random perturbations of 3 Å and 8° rotations before each simulation (Chaudhury et al., 2011). Each docking sampling comprised 500 cycles, from which the top-scoring models were selected. The chosen structures, exhibiting the best interface score (I_sc), underwent visual inspection and were compared to the reference 6MET structure. Subsequently, these top-scoring structures underwent refinement using the Rosetta Refine and Relax protocols, ensuring their accuracy and reliability (DiMaio et al., 2009; Sircar et al., 2010).

2.5. In silico mutagenesis and docking analysis

Amino acid residues at position 58 of *CD4* were reciprocally swapped between *Catarrhini* and other species, and protein 3D structures were generated following the procedures outlined above. Subsequently, the docking process between *CD4* and HIV-1 subtypes B, C, N, and O was conducted, employing the methods detailed in the previous section. To assess the significance of the observed outcomes, p-values were calculated using Welch's *t*-test (Welch, 1947).

2.6. Estimation of positive selection

To discern sites under positive selection contributing to higher *gp120/CD4* affinity scores (I_sc values) in apes and old-world monkeys (*Catarrhini* parvorder) compared to other primates and rodents, we conducted dN/dS ratio analysis using the branch-site model of CodeML in PAML (Dutrillaux et al., 1982; Yang, 2007). We prepared two inputs: (1) the alignment file of *CD4* sequences; and (2) an unrooted tree in Newick format marking the foreground branch at the most recent common ancestral branch of the *Catarrhini* lineage for both null and alternative hypotheses (Cayley, 1857; Dutrillaux et al., 1982). The ratio (dN/dS) of nonsynonymous substitutions on nonsynonymous sites (dN) per synonymous substitutions on synonymous sites (dS) was calculated with specific options in the control files: For both hypotheses, CodonFreq = 2 (F3X4), model = 2 (2 or more dN/dS ratios for branches), NSsites = 2 (selection), omega = 1 (as initial omega), cleandata = 0 (calculate sites with ambiguity data), and other options applied as default values in the manual; the only different option between hypotheses was fix_omega = 1 (fixed) for the null and 0 (estimate) for alternative hypotheses (Biswas & Akey, 2006; Yang, 2007). The statistical significance of the alternative hypothesis for positive selection and the posterior probability to classify positively selected sites were computed from the chi-squared test (degree of freedom = 1) with the likelihood ratio test ($D > 1$) and the empirical Bayes method (BEB value > 0.5), respectively (Casella, 1985; Murphy et al., 2001).

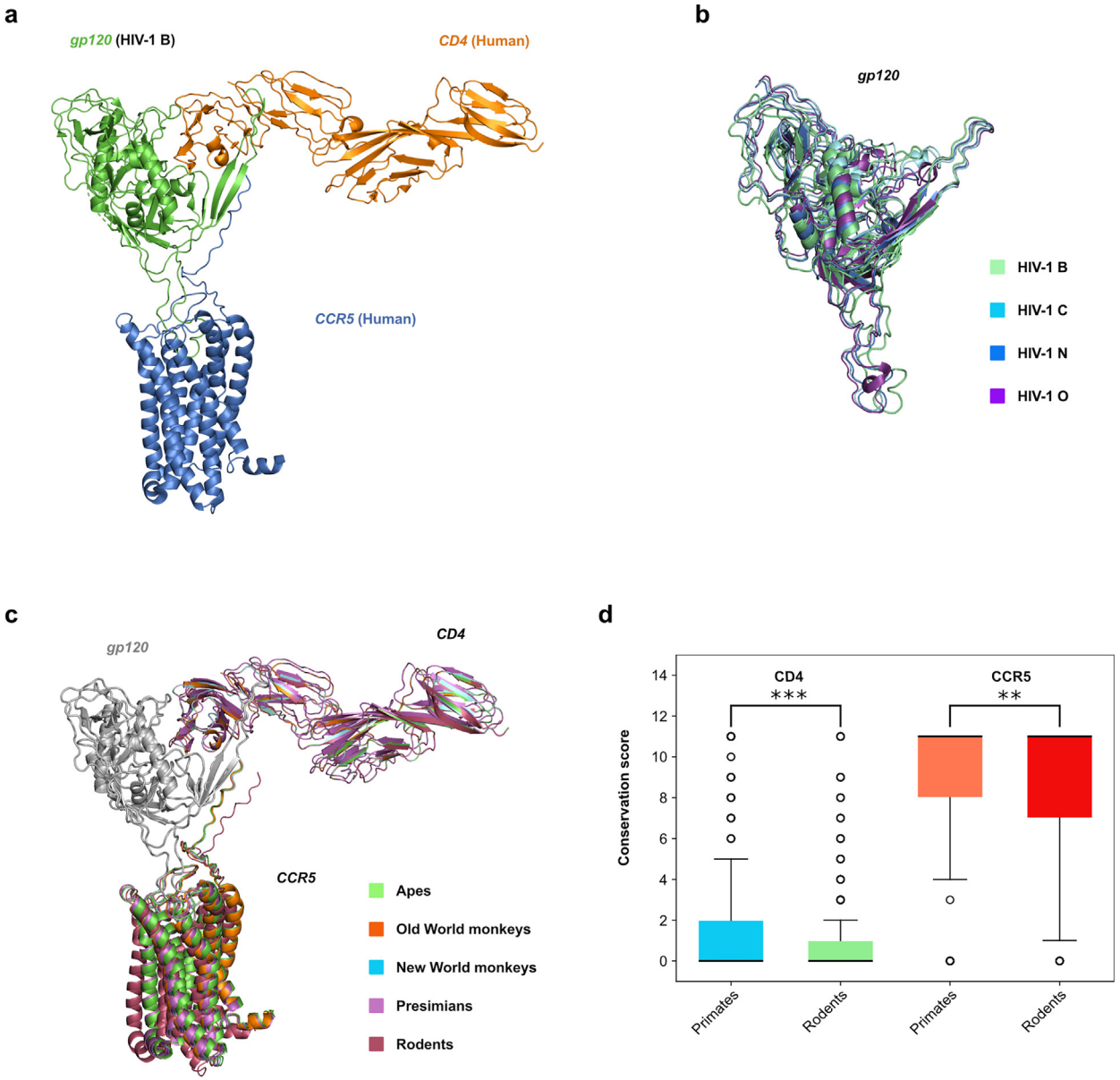

Fig. 1. 3D structures of *gp120*/*CD4*/*CCR5* tri-complexes across HIV-1 subtypes and animal species. (**a**) Template structure of *gp120* (HIV-1 B) and *CD4*/*CCR5* (Human) (PDB entry: 6MET) [48]. (**b**) Aligned tertiary structures of HIV-1 (group M (subtypes B and C), N, O) modelled through homology modelling methods. (**c**) Conservation scores of protein sequence alignments of *CD4* and *CCR5*. Primates *CD4*: sky blue, rodents *CD4*: light green; Primates *CCR5*: orange and rodents *CCR5*: red. Welch's *t*-test was employed to compare the conservation scores between primates and rodents in both *CD4* and *CCR5*, revealing significant differences (**p < 0.01, ***p < 0.001). (**d**) Docked structures of *gp120* (HIV-1 B) with *CD4* and *CCR5* co-receptors of representative species in each clade.

2.7. Ecological distribution of the primates

We obtained information on the natural habitats of the primates under investigation from the IUCN Red List of Threatened Species (https://www.iucnredlist.org/) (IUCN, 2024). The map template, sourced from OpenDataSoft (Explore) and titled "World Administrative Boundaries – Countries and Territories," served as the basis for illustrating primate habitats (Explore). Utilizing QGIS 3.18 software, we mapped and edited the distribution of primates on the template (QGIS. and QGIS). Visual enhancements were applied using Affinity Designer software (Affinity).

3. Results

3.1. Tri-complex protein structures of gp120 of HIV and CD4/CCR5 of animals

To investigate the evolutionary molecular dynamics of *gp120/CD4/CCR5* genes in HIV and animals, we collected their corresponding protein sequences. We retrieved representative viruses from groups M (subtypes B and C), N and O, from the HIV database (https://www.hiv.lanl.gov/), specifically targeting *gp120* protein sequences. The selection of these subtypes was informed by their prevalence in global infection rates, thus ensuring a comprehensive representation of HIV diversity (Supplementary Table S1) (Bbosa et al., 2019; Los Alamos National Laboratory, 2021). HIV-1 group P was excluded from the study due to its exceedingly rare occurrence and limited relevance to the research objectives (Bbosa et al., 2019). For animals, we obtained orthologous peptide sequences of *CD4* and *CCR5* genes of 18 primates and 12 rodents in Ensembl database (https://ensembl.org/, Supplementary Table S2) (Howe et al., 2021).

Building upon the established framework of the tri-complex *gp120/CD4/CCR5* structure of HIV-1 B and human (Fig. 1a), we computationally generated 3D protein structures using the peptide sequences of each viral sub-lineage and animal species. The modeling was performed using the SWISS-MODEL homology modeling server, leveraging ProMod3 module for structure generation (Moebius et al., 1992; Prévost et al., 2020; Shaik et al., 2019; Studer et al., 2021; Tamamis & Floudas, 2014). ProMod3 in its turn utilized the following modules to perform modeling tasks: First, the *loop model*, which introduced a BackboneList class for peptide backbone representation, allowing seamless conversion between Cartesian and internal coordinates (Studer et al., 2021); Second, the *scoring module* in ProMod3 incorporated both physics-based and knowledge-based methods to evaluate protein models (Studer et al., 2021); Third, the *sidechain* module in ProMod3 utilized rotamer libraries to efficiently model amino acid sidechains, implementing both rigid (RRM) and flexible (FRM) rotamer models (Studer et al., 2021); Last, the ProMod3 modelling module *integrated* loop, scoring, and sidechain functionalities using algorithms like Cyclic Coordinate Descent and Kinematic Closure (Studer et al., 2021).

Despite the considerable divergence observed among HIV-1 subtypes, with a maximum dissimilarity of 40%, their overall structural similarity remained consistent, as indicated by the analysis of total energy deviation as calculated with ProsaWeb server, yielding corresponding Z-scores of B = −6.88, C = −6.36, N = −6.62, and O = −6.23.

(Fig. 1b–Supplementary Table S3) (Wiederstein & Sippl, 2007). Despite variations in their protein sequences, animals exhibited similar structural features. Both groups displayed consistent structural patterns (Fig. 1c), suggesting that these similarities are not solely attributable to the use of a single template in homology modeling. This observation highlights the robustness of homology modeling in predicting protein structures across diverse species. Rodents demonstrated a broader spectrum of sequence diversity compared to primates (Fig. 1d–Supplementary Table S4).

3.2. In silico binding affinity of gp120/CD4 and gp120/CCR5

Based on the above protein structures of HIV subtypes and animals, we tested whether viral and host docking protein interactions are related to host phylogenetic distances known as one of key factors in viral infection. We estimated *in silico* binding affinities of each pair of *gp120/CD4* and *gp120/CCR5* utilizing the RosettaDock protocol. This protocol involves a multi-scale approach, including low-resolution centroid-mode and high-resolution refinement steps, which allowed us to accurately model protein-protein interactions (Chaudhury et al., 2011; DiMaio et al., 2009; Sircar et al., 2010). Out of these pairs, *gp120/CD4* demonstrated the strengths of protein-protein docking (interaction score, I_sc) was associated with phylogenetic relationships of animals (Table 1, Supplementary Table S5). Interestingly, we discovered that non-human apes and old-world monkeys from the *Catarrhini* clade have stronger binding affinities similar to human with lower averages of I_sc values as −44.164 to −48.711 than comparably distant animals such as new world monkeys and prosimians having lower binding affinities with higher averages of I_sc values as −39 to −23 (Fig. 2a–d). However, there was no significant difference of *gp120/CCR5* interactions among lineages (Fig. 2e–h). Among the top-scoring models, the human and Black snub-nosed monkey demonstrated notable I_sc scores of −37.562 and −37.221 for HIV-1 subtype B, respectively. Primates generally showed commendable scores within the range of −34.3 to −36.8 (Table 1, Supplementary Table S5). Similar inconsistencies were observed in the results for other HIV-1 strains (Supplementary Table S5).

In addition, we expanded the investigation to include rodents, the next closest clade to primates (Supplementary Table S5). Docking scores of *gp120/CCR5* exhibited inconsistency and significant variation across all four HIV-1 subtypes. The *gp120/CD4* interaction consistently exhibited low binding scores across all four HIV-1 subtypes. However, a notable exception was observed with the squirrel model, which yielded a score of −42.941 when interacting with HIV-1 O (Supplementary Table S5). Despite obtaining interpretable results, the variability observed in the docking scores for rodents underscores the complexity of their interactions with *CD4* and *CCR5* receptors across different HIV-1 subtypes. This variability suggests that rodents may not exhibit consistent binding patterns.

3.3. Amino acid patterns in precedent candidate sites of HIV docking

To explain the phylogeny-dependent binding affinity variations, we investigated amino acid evolution in key residues within the D1 domain of *CD4* (43F, 46K, and 59R) crucial for *gp120* binding (Fig. 3a) (Fontenot et al., 2007; Krykbaev et al., 1997; Moebius et al., 1992; Prévost et al., 2020; Shaik et al., 2019). Additionally, the significance of Tyr residues in *CCR5*'s

Table 1
Docking scores between *gp120* (HIV-1 B) with *CCR5* and *CD4* co-receptors.

Species		Group	CCR5		CD4	
Common name	Scientific name		I_sc	Total score	I_sc	Total score
Human	*Homo sapiens*	Ape	−37.562	−357.551	−46.155	−188.112
Bonobo	*Pan paniscus*	Ape	−34.348	−365.356	−45.89	−181.203
Chimpanzee	*Pan troglodytes*	Ape	−34.73	−362.833	−44.597	−176.91
Gorilla	*Gorilla gorilla gorilla*	Ape	−36.54	57.546	−44.164	−177.789
Orangutan (ENSEMBLE)	*Pongo abelii*	Ape	−36.55	−356.386	−25.117	−93.027
Orangutan (NCBI)	*Pongo abelii*	Ape	–	–	−45.699	−170.933
Gibbon	*Nomascus leucogenys*	Small ape	–	–	−48.06	−236.956
Black snub-nosed monkey	*Rhinopithecus bieti*	Old World monkey	−37.221	−346.939	−46.773	−163.645
Crab-eating macaque	*Macaca fascicularis*	Old World monkey	−36.421	−356.104	−46.913	−153.422
Golden snub-nosed monkey	*Rhinopithecus roxellana*	Old World monkey	−34.821	−354.783	−46.647	−161.997
Olive baboon	*Papio anubis*	Old World monkey	−35.875	−364.892	−48.711	−183.248
Rhesus macaque	*Macaca mulata*	Old World monkey	−35.344	−357.347	−44.676	−165.855
Sooty mangabey	*Cercocebus atys*	Old World monkey	−31.419	−348.45	−44.912	−163.748
Pig-tailed macaque	*Macaca nemestrina*	Old World monkey	−34.356	−339.524	−48.729	−166.573
Marmoset	*Callithrix jacchus*	New World monkey	−35.25	−308.575	−39.277	−42.514
Ma's night monkey	*Aotus nancymaae*	New World monkey	–	–	−39.884	−61.887
Mouse lemur	*Microcebus murinus*	Lemur	−36.848	−352.16	−25.537	79.417
Tarsier	*Carlito syrichta*	Tarsier	–	–	−25.617	−39.455
Chinese hamster	*Cricetulus griseus*	Rodent	−29.846	−286.112	−31.79	29.137
House mouse	*Mus musculus*	Rodent	−29.43	−266.869	−23.842	88.016
Guinea pig	*Cavia porcellus*	Rodent	−45.457	−326.517	−28.172	154.8
Lesser Egyptian jerboa	Jaculus jaculus	Rodent	−36.63	−214.844	−30.102	140.808
Pika	*Ochotona princeps*	Rodent	−41.918	−284.013	−25.105	337.902
Prairie vole	*Microtus ochrogaster*	Rodent	−30.918	−286.346	−35.366	98.448
Rat	*Rattus norvegicus*	Rodent	−30.632	−299.972	−24.615	−185.62
Squirrel	*Ictidomys tridecemlineatus*	Rodent	−29.694	−333.682	−39.716	206.325
Golden hamster	Mesocricetus auratus	Rodent	−30.952	−311.605	−21.5	118.296
Naked mole rat	Heterocephalus glaber	Rodent	−39.82	−365.67	−24.801	400.259
Tree shrew	Tupaia belangeri	Shrew	–	–	−20.9	−172.49
Shrew mouse	Mus pahari	Shrew	−33.987	−276.137	−24.35	−12.252

N-terminal tail for *gp120* interaction was emphasized by Shaik et al. (Fig. 3b) (Shaik et al., 2019). Our evolutionary analysis aimed to identify signs of pressure on these residues and discover new positively selected sites. Subsequently, detailed sequence analysis revealed species-specific patterns in *CD4* and *CCR5* residues associated with *gp120/CD4/CCR5* docking (Fig. 3c and d). Ape *CD4* residues 43F, 46K, 59R were conserved, while Old World monkeys substituted the last position with lysine. New World monkeys and rodents exhibited variable substitutions at all positions, and *CCR5* substitutions generally correlated with *CD4*, with Tyr10, Tyr14, and Tyr15 conserved among apes and Old World monkeys. We constructed a phylogenetic tree for *CD4* and *CCR5*, conducting codon-wise multiple sequence alignments.

In the six candidate sites within these two key genes, human amino acids were highly conserved in Hominidae (great apes). For the three candidate sites in *CD4* (43F, 46K, and 59R), all apes exhibited identical amino acids (Fig. 3c). Old World monkeys replaced lysine with arginine at the 59R site, while New World monkeys and mouse lemurs demonstrated divergent substitutions at positions 46K and 59R (Fig. 3c). Notably, for the 43F site, primates displayed a conserved residue (phenylalanine), while rodents exhibited mutually exclusive substitutions (leucine, valine, and tyrosine).

In the *CCR5* alignment (Fig. 3d), two out of three key candidate residues (Y10 and Y14) were conserved across all primates and the majority of rodents. The only exceptions were the lesser Egyptian jerboa and squirrel, with position Y10 substituted by aspartate and serine. Human Y15 showed divergent substitutions in new world monkeys and the rodent clade with non-aromatic residues (aspartate, glutamate, glycine, and serine).

3.4. Novel candidate sites under positive selection related to catarrhini-specific gp120/CD4 affinities

In order to explain the boundary of higher *gp120/CD4* affinity scores (I_sc values ≥ -44.16) of *Catarrhini* parvorder than the other primates (I_sc values ≤ -39.88), we investigated whether the well-known key binding sites and novel candidate sites undergo positive selection, based on dN/dS analysis by using PAML program (Yang, 2007). Under the branch-site model A for *CD4* orthologous genes of 24 primates, we checked the higher likelihood value (-6999.60) for positive selection on *Catarrhini* than the value (-7000.02) of null hypothesis for neural selection (Likelihood ratio test, $D > 0$), but did not find the statistical significance (p-value $= 0.17$). We identified four positively selected sites (PSSs) with class 2 dN/dS on the foreground branch ($W = 20.73$), based on the posterior probability (BEB >0.5); these sites did not overlap with the previously reported three key sites on *gp120/CD4*. However, out of four PSSs (83R (58R), K324, N384, and T392 in human protein sequence of *CD4*), 58R is immediately adjacent to one of the key binding sites (59R) (Fig. 4a and b). By considering PSSs harboring protein functional domains, 83R (58R), K324 and N384 were located in Ig superfamily domains (Fig. 4c). Especially, 58R was located in Ig strand

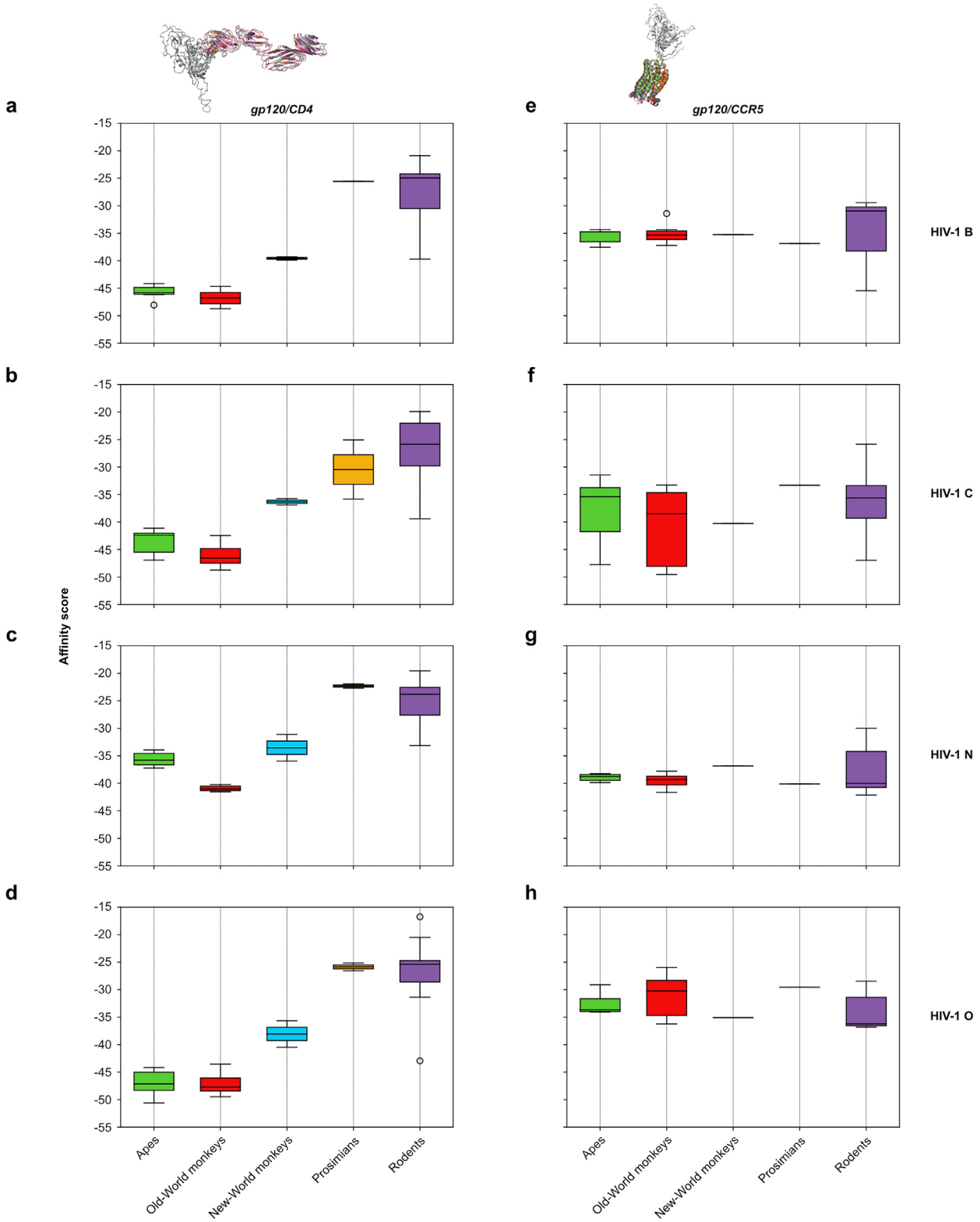

Fig. 2. Docking affinity distribution of gp120/CD4/CCR5 across animals and HIV-1 subtypes (B, C, N, O) in various clades. (**a-d**) Boxplots representing the affinity distribution of *gp120/CD4* binding scores of HIV-1 across multiple clades including apes (green), Old World monkeys (red), New World monkeys (sky blue), prosimians (orange) and rodents (purple). (**e-h**) Boxplots representing the affinity distribution of *gp120/CCR5* binding scores of HIV-1 across multiple clades including apes (green), Old World monkeys (red), New World monkeys (sky blue), prosimians (orange) and rodents (purple).

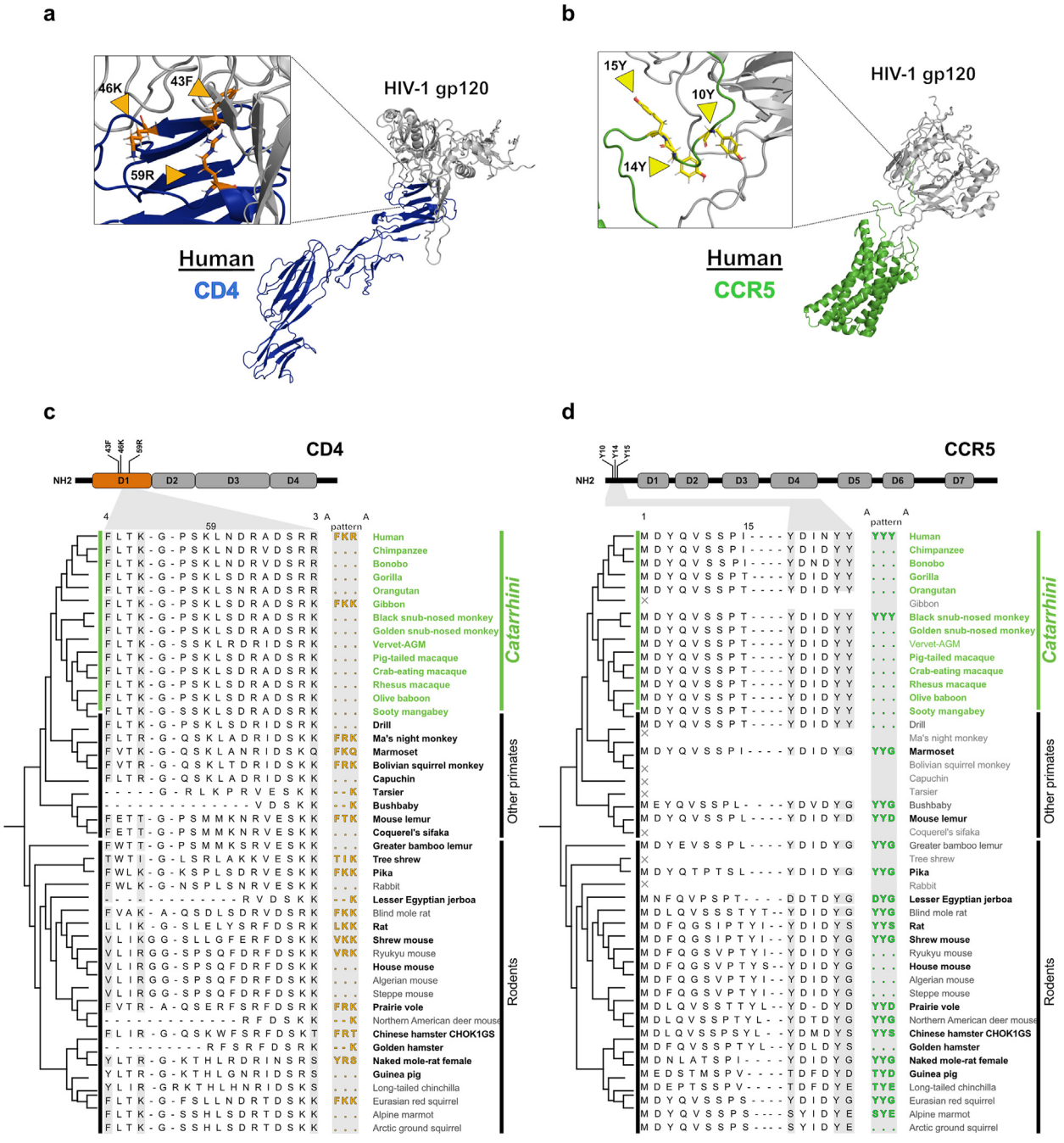

Fig. 3. Conserved human-like amino acid patterns in previously identified candidate sites. (**a, b**) Depiction of pre-reported candidate sites contributing to docking between *gp120* (light gray) and *CD4* (blue), as well as *gp120* (light gray) and *CCR5* (green). The corresponding key amino acids are highlighted in orange in *CD4* and in yellow in *CCR5*. (**c, d**) Illustration of the amino acid patterns at key candidate sites of *CD4* (orange) and *CCR5* (green) using the Ensembl tree (Howe et al., 2021; Stalker et al., 2004). As depicted in the figure, residues exhibit a notably higher level of conservation in *Catarrhini* compared to both *CD4* and *CCR5*.

D and framework region 3 (FR3) which is known as residue responsible for supporting the binding of the antigen to the antibody (Barclay, 1999; Su et al., 2017). Barclay, 1999Su et al., 2017To further investigate the findings from the dN/dS analysis, we conducted substitution analyses. Specifically, we introduced the R58K substitution in the *CD4* protein of *Catarhinni* species and the K58R substitution in other non-human primate species (Fig. 5). Subsequently, protein-protein docking predictions were performed. Statistical analysis was conducted to estimate the significance levels. For the subtypes B, C and N the values were not significant for both *Catarhinni* and other non-human primates (Fig. 5). However, the results of *Catarhinni* WT versus *Catarhinni* R58K for HIV-1 O were found to be significant (Welch's *t*-test, *p < 0.05) (Welch, 1947).

3.5. Investigating the relationship between primate ecological distribution and HIV binding affinities

To decipher the variations in *gp120/CD4* binding affinities among diverse primate species, we formulated a hypothesis proposing that their natural distribution significantly influences viral interactions. A discernible correlation emerges between habitat and binding affinities (Fig. 6). The majority of apes and Old World monkeys, sharing similar habitats due to their phylogenetic proximity, exhibit comparable binding affinity scores. This similarity arises from the high homology in their protein structures. In contrast, New World monkeys demonstrate a higher degree of disparity in affinity values. While some species share habitats, the broader territories inhabited by most primates lead to increased variation in the receptor docking regions and consequently result in lower binding affinities.

4. Discussion

In this study our objective was to examine various critical facets to gain a comprehensive understanding of the interaction between HIV-1 and its host receptors. This investigation involved *in silico* studies of virus-host docking, as well as analyses for candidate sites that could potentially undergo positive selection and their relevance to the identified key regions.

The findings presented by Warren et al. highlighted similar interaction rates among primate species, including gibbon, olive baboon, gelada, black snub-nosed monkey, owl monkey, and colobus, when compared to human *CD4* co-receptor, despite not being the closest relatives to human (Warren et al., 2019). Although, disease progression could not be observed in this study due to its nature, multiple non-primate *CD4* co-receptors were able to interact with the early HIV-1 isolates and resulted in viral entry to the cells (Warren et al., 2019). This might give a hint that there is a strong possibility that some primates closely related to human can be infected with the virus. Another, prime example, which confirms this theory is one of the early research conducted by Novembre et al. where a chimpanzee (C499) infected with HIV successfully developed AIDS (Novembre et al., 1997). Although, the disease progression took approximately 10 years to develop it indicates the high possibility of it in evolutionary progress. Gibbons and pig-tailed macaques are another example of experimental infection with HIV-1 [12,14]. While the infection rate was weaker, gibbons showed development of mild signs of the HIV infection (Lusso et al., 1988). Pig-tailed macaque was shown to be susceptible to HIV infection, as shown by Agy et al. (Agy et al., 1992). Although, the above studies are mostly limited to interaction between HIV and *CD4*, we believe that we have made an important step towards deeper understanding of the HIV infection.

We observed a consistent trend across species in this study: apes and Old World monkeys exhibited a higher propensity for interactions with either *CD4*, while New World monkeys yielded overall lower interaction scores. We included rodents in our study as the next closest clade to the primates out of curiosity about behavior of their receptors when compared to the primates (Howe et al., 2021). Here, we report that interaction patterns of rodents remained enigmatic, consistent with the fact that rodents are not known to be infected with HIV. Our *CD4* docking simulations supported our theory that some primate species exhibit strong binding affinities, corroborating past *in vitro* findings in species like the gibbon and Ma's night monkey (Warren et al., 2019). Interestingly, our study also revealed elevated docking scores for rhesus macaque, a discrepancy warranting further investigation to unravel its underlying implications. Our quest for polymorphism and pathogenic variants led us to intriguing revelations. Specifically, an examination of *CD4* key sites in the alignment yielded a lack of variants, hinting at the gene's highly conserved nature. This preservation might underlie the robust interaction observed between *gp120* and the *CD4* receptor. The results of our *gp120/CCR5* docking analysis, however, remained inconclusive, lacking a discernible trend in binding affinities among different species. These insights underscore the intricate nature of virus-host interactions, necessitating a comprehensive and holistic approach to understand the broader landscape of HIV pathogenesis. We believe that the reported key sites play crucial role in HIV infection, however, we were not able to fully confirm the same level of interaction in this study.

Despite strong and consistent results for protein-protein interaction across multiple species caution is warranted due to the potential bias introduced by using the same template for all species. Notably, an intriguing alignment of I_sc scores emerged among marmoset, Ma's night monkey, and squirrel, hinting at an evolutionary relatedness. While these results demonstrate robust binding probabilities, attributing HIV infection susceptibility solely to binding affinity is incorrect, as species relatedness and molecular factors play crucial roles. The use of a singular template for tertiary structures may introduce limitations, contributing to variations in docking outcomes among species. This nuanced consideration emphasizes the need for a comprehensive understanding of both binding affinity and structural integrity in assessing species susceptibility to HIV infection.

Although this study is purely computational and requires experimental validation to verify the findings, the data we've obtained could potentially inform new approaches for studying HIV interactions with non-human hosts. Our computational framework, combining protein modeling, protein-protein docking simulations, interaction scoring, and cross-species comparative analysis, might serve as a foundation for predicting potential animal models for HIV research. This approach could theoretically identify species with receptor binding profiles similar to humans, which may be worth investigating as new animal models. Furthermore, unexpected similarities or differences in computational interaction scores across species could potentially reveal evolutionary relationships or resistance mechanisms worth exploring. To enhance the predictive power of such computational models, future work could involve developing machine learning algorithms that integrate docking simulation data with known experimental results. Additionally, this type of modeling approach could potentially be adapted to create virtual screening platforms for HIV entry inhibitors, possibly accelerating the initial stages of drug discovery.

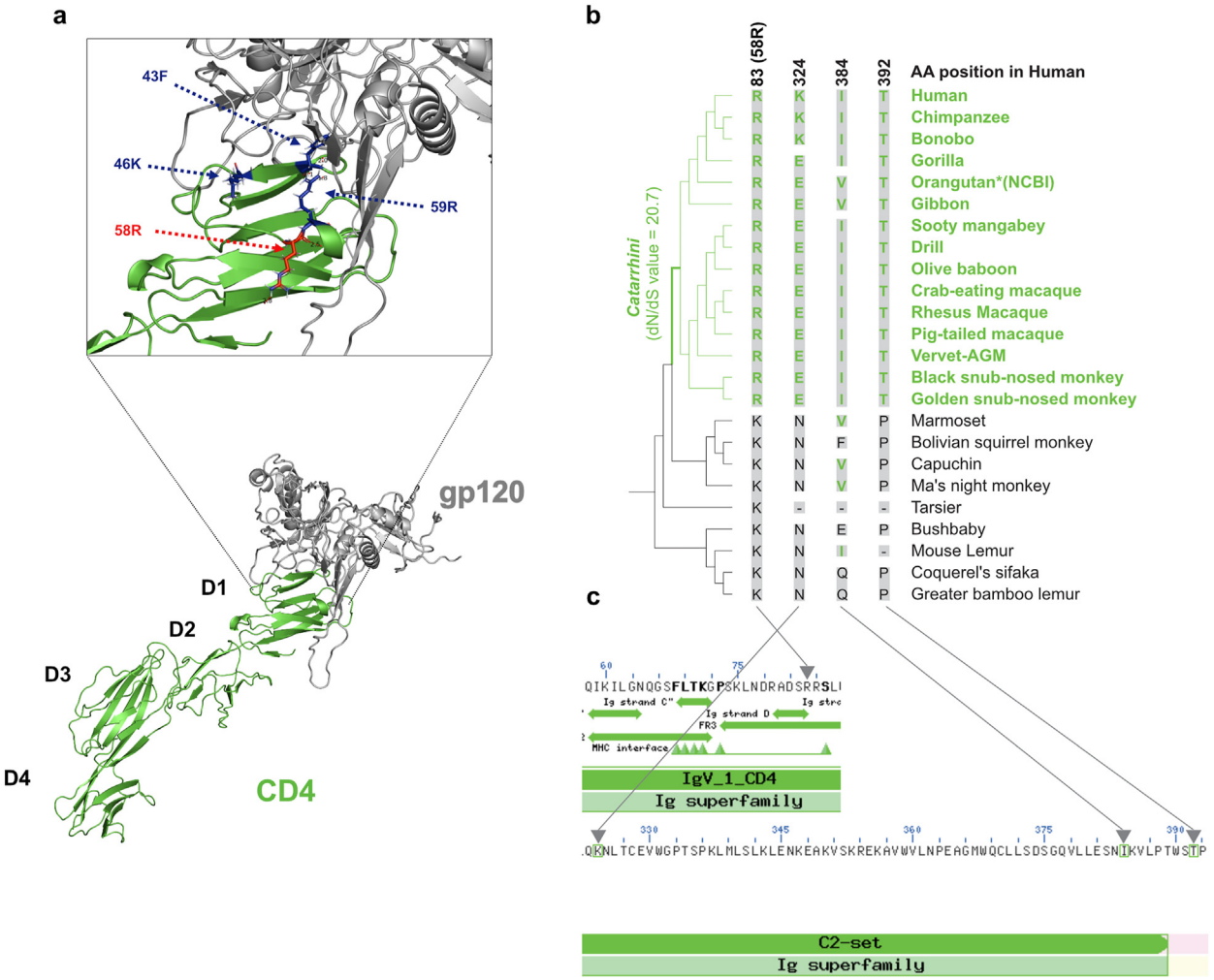

Fig. 4. Positively selected sites on *Catarrhini* lineage. (**a**) Amino acid positions of precedent candidate sites (blue) and a new candidate site under positive selection (58R) on *Catarrhini* (red) in *gp120/CD4* interaction. (**b**) Amino acid patterns at positively selected sites on *Catarrhini*. Residue 58R is conserved within the *Catarrhini* clade, whereas lysine is present in other primate species. (**c**) Localizations of the Positively Selected Sites located on functional domains visualized using UCSF Genome Browser (Karolchik et al., 2003).

However, it's crucial to emphasize that all these potential applications would require rigorous experimental validation to confirm their practical utility in HIV research.

A significant pioneering aspect of our investigation involved the identification of newly emerged candidate sites under positive selection. These sites, strategically positioned within the Ig domain renowned for antigen-receptor interactions, offered a fresh perspective on the intricate interplay between viral and host proteins (Biswas & Akey, 2006). This revelation underscored the integral role of mutational dynamics in molding protein-protein affinities. In pursuit of delineating the boundaries of our *gp120/CD4* docking scores, we performed a dN/dS analysis, unveiling four PSSs. Notably, although these PSSs did not align with previously reported key sites, the majority (three out of four) were situated within the Ig domain, notable for its involvement in antigen-antibody interactions (Barclay, 1999; Su et al., 2017). We sought to ascertain whether these positively selected key residues influenced the strength of protein-protein interactions, postulating that regions undergoing positive selection would display heightened affinity to *gp120*. Of particular interest was the discovery of position 58 as an identified PSS. Sequence analysis revealed its robust conservation across multiple species. Intriguingly, while arginine occupied this position within the *Catarrhini* lineage encompassing apes and Old World monkeys, lysine prevailed across the majority of other species. Notably, 58R appeared in various rodent species, including rat, prairie vole, golden hamster, Upper Galilee mountains blind mole rat, naked mole rat, and guinea pig. Initially, we speculated a potential correlation between the *gp120/CD4* interaction of these rodent species and the *Catarrhini* group. However, upon examination of the I_sc values, this hypothesis was invalidated. Subsequently, our scrutiny of the sequences revealed a conserved serine residue at position 60 within the *Catarrhini* lineage. While not initially identified as a predicted PSS, 60S resides adjacent to 59R within the Ig

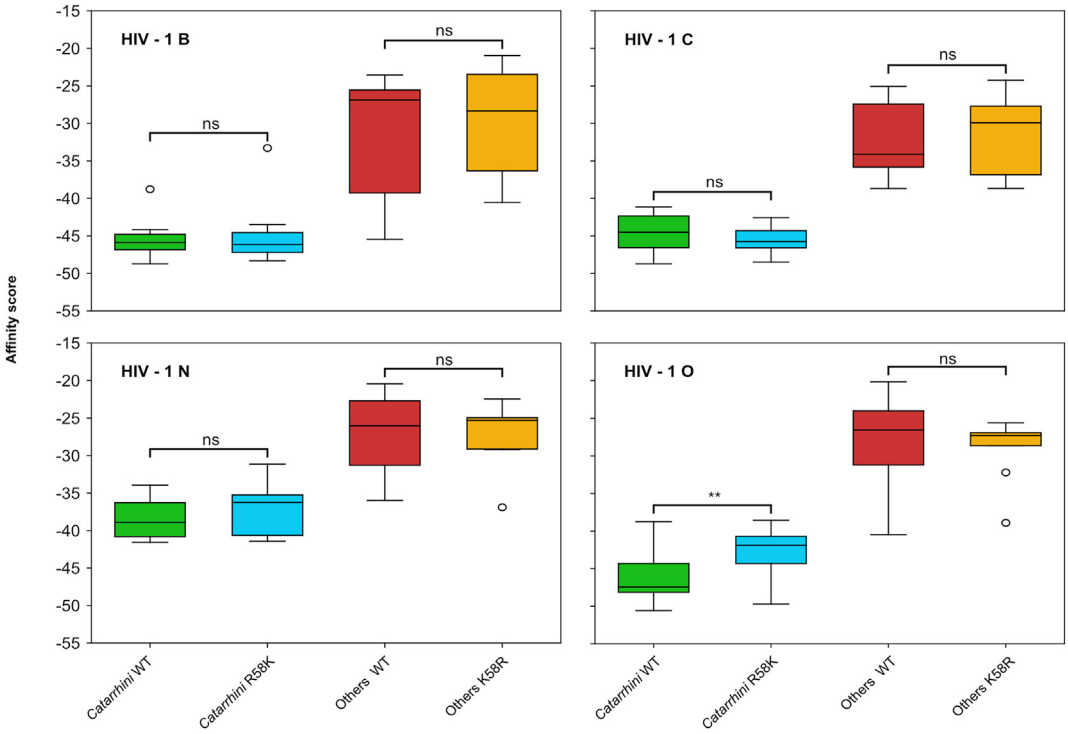

Fig. 5. Binding scores of *Catarrhini* and other animals with/without artificial mutation (K58R) in four HIV subtypes (B, C, N, O). (**a-d**) The impact of introducing a lysine residue to position 58 (R58K) in *Catarrhini* lineage species, and conversely, introducing an arginine residue to position 58 (K58R) in other primate species, on *gp120*/*CD4* affinity. Statistical significance was assessed using Welch's *t*-test, with 'ns' indicating not significant, and '**p < 0.01′ indicating statistical significance (Welch, 1947).

superfamily, suggest a plausible influence on *gp120/CD4* interaction. We acknowledge that some of the retrieved sequences were missing amino acid residues, particularly in key regions. While our analysis suggests that these gaps did not significantly alter our main findings, we recognize that this incompleteness could potentially affect the comprehensiveness of our results. To address this limitation, future research will prioritize collecting and supplementing data for these species, especially focusing on the key residues. This more complete dataset will enhance the accuracy of our analyses and provide a more robust foundation for our conclusions. Additionally, *in vitro*/*in vivo* validation will be crucial to confirm our computational findings and to fully understand the impact of these sequence variations. In summary, our positive selection analysis introduced compelling prospects for understanding the evolutionary dynamics of protein-protein interactions. The identification of these new sites of interest, particularly within the context of the Ig domain, underscores the intricate nuances governing the relationship between viral and host proteins.

Our experimentation extended to the introduction of mutations in the *CD4* protein. Position 58 revealed a captivating pattern that spanned the array of studied species. Particularly noteworthy was the consistent presence of arginine at position 58 among all terminal taxa of the *Catarhinni* lineage. This unifying trait set them apart from other non-human primates, where lysine was observed at the same position. Despite the shared basic properties of both amino acids, the consistent cross-species pattern captured our attention, warranting further exploration. Driven by this insight, we executed a strategic manipulation, substituting arginine with lysine in *Catarhinni* and vice versa, followed by a series of *gp120*/*CD4* docking simulations using the methods described above. Remarkably, no significant differences materialized for subtypes B, C, and N upon introduction of the mutations. However, subtype O presented a statistically significant contrast between *Catarhinni* WT with R and artificial single point mutants R58K with K. While this hypothesis held promising potential, its applicability was constrained, yielding statistically insignificant outcomes for the majority of the scrutinized species. Despite the limitations encountered, this exploratory approach underscores the complexity of the virus-host interaction landscape and necessitates further inquiry to decipher the intricate influence of mutational dynamics. Although our computational analysis has provided strong initial assumptions about the interactions, we cannot be entirely certain of their validity. A major improvement for future studies will be to introduce the studied mutations using both *in vitro*/*in vivo* experimental validation to assess their effect on the binding affinity of *gp120*/*CD4* and to fully understand their biological significance. This will also help to verify the impact of these mutations on HIV infection.

It's also worth noting that Shaik et al. 's research illuminated the distinctive nature of Tyr15 [8,50,51]. Among the trio of residues-Tyr10, Tyr14, and Tyr15; Tyr15 uniquely lacks sulfation due to spatial limitations for accommodating a sulfate group

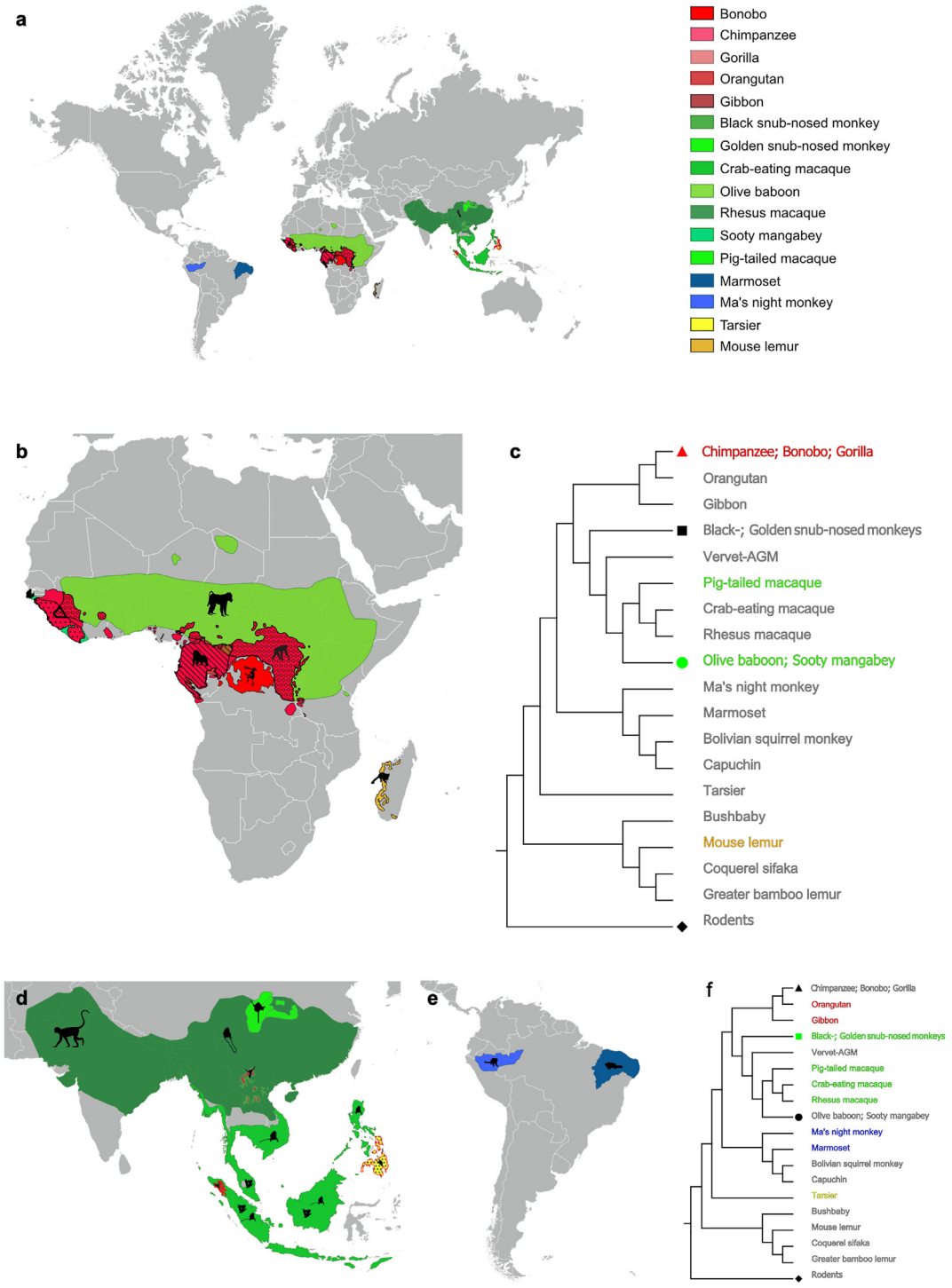

Fig. 6. Species distribution map. (**a**) A world map illustrating the global distribution of primates. Each species is represented by a distinct color code. (**b-f**) Individual species distribution maps with the corresponding phylogenetic trees. Phylogenetic trees were constructed using information retrieved from Ensembl tree (Howe et al., 2021; Stalker et al., 2004).

(Niehrs et al., 1993; Shaik et al., 2019; Farzan et al.). Tyrosine sulfation sites tend to exhibit conservation across mammalian species, and regions featuring multiple tyrosine residues in proximity often manifest lower conservation (Farzan et al). Consequently, the absence of sulfation on *CCR5*'s Tyr15 might have influenced the polymorphic behavior observed in this residue. Within our analysis of *gp120*/*CCR5* docking, our focus centered on the residues situated in the N-terminal domain.

However, it's crucial to acknowledge that the *V3*/*CCR5* docking encompasses a broader array of amino acid residues, as evidenced by previous studies (Tamamis & Floudas, 2014; Farzan et al.). These additional residues actively contribute to the overall docking score, suggesting a more intricate interplay at play in this interaction. While our study demonstrated some *gp120*/*CCR5* interactions, the results were inconsistent preventing us from drawing a solid conclusion. We believe, that further *CCR5* focused *in vitro* analysis can provide with more specific details about the viral/host gene interaction.

The ecological distribution of primates likely plays a significant role in shaping HIV-host interactions, potentially influencing both viral transmission and evolution. The diverse habitats of primates across Africa, Asia, and South America create varied environmental contexts for host-pathogen interactions. These ecological differences may contribute to the genetic diversity observed in both host receptors and viral strains. In Africa, where HIV is believed to have originated, the coexistence of various primate species in close proximity could theoretically provide opportunities for cross-species transmission events (Sharp and Hahn; Gao et al., 1999). The complex ecological networks in these regions, involving both non-human primates and human populations, may have facilitated the initial zoonotic transmissions and subsequent spread of the virus. Geographical isolation, such as that experienced by lemurs in Madagascar, could potentially lead to unique evolutionary trajectories for both hosts and pathogens (Setash et al., 2017). Conversely, widely distributed species like macaques across Asia might experience different selective pressures, possibly resulting in varied immune responses and receptor characteristics. The genetic diversity of HIV within human populations may also be influenced by ecological factors. Research conducted in controlled laboratory environments has shown evidence of parallel evolution in HIV. This suggests that the evolutionary trajectory of HIV-1 remains relatively consistent even when subjected to minor environmental variations (Bons et al., 2020). Different environmental conditions across geographical regions could potentially exert varying selection pressures on the virus, contributing to the emergence of distinct HIV subtypes and recombinant forms. However, it's important to note that human factors such as migration patterns and societal practices also play crucial roles in viral distribution and evolution. At the molecular level, environmental factors could theoretically affect the expression and functionality of HIV co-receptors. Factors such as UV radiation exposure, which varies with geographical location, might influence cellular processes, potentially impacting virus-host interactions. However, more research is needed to fully understand these complex relationships.

While direct causal relationships between specific ecological factors and HIV dynamics are challenging to establish, the ecological perspective provides a valuable framework for understanding the complex interplay between virus, host, and environment. Future research integrating ecological data with molecular and epidemiological studies could potentially offer deeper insights into HIV transmission and evolution across diverse populations.

In conclusion, our findings suggest the new candidate site R58 through *in silico* comparative approaches for HIV-host protein interactions. However, uncertainties persist regarding these interactions in other primates, which can be validated in future *in vitro* studies. This research aims to clarify the intricate dynamics of *gp120*/*CD4* affinities in diverse primate species, contributing insights into species-specific vulnerabilities to HIV-1 infection. This endeavor builds upon our current knowledge and underscores our commitment to understanding the multifaceted interplay between *gp120* and *CD4*. Through this effort, we anticipate enriching our comprehension of this complex subject and contributing to the broader understanding of HIV-1 interactions and their evolutionary implications.

CRediT authorship contribution statement

Vladimir Li: Writing – review & editing, Writing – original draft, Visualization, Validation, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Chul Lee:** Writing – review & editing, Writing – original draft, Visualization, Validation, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **TaeHyun Park:** Writing – review & editing, Writing – original draft, Methodology, Investigation, Conceptualization. **Erich D. Jarvis:** Writing – review & editing, Supervision, Project administration, Methodology. **Heebal Kim:** Writing – review & editing, Supervision, Project administration, Methodology, Funding acquisition, Conceptualization.

Data availability statement

All data essential to reproduce the findings of this study are presented within the manuscript. *CD4* and *CCR5* sequences were retrieved from ENSEMBL (https://www.ensembl.org) and NCBI databases (https://www.ncbi.nlm.nih.gov). HIV sequences were retrieved from Los Alamos National Laboratory HIV Database (https://www.hiv.lanl.gov).

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

This work was supported by the BK21 FOUR Program of the Department of Agricultural Biotechnology, Seoul National University, Seoul, South Korea and was supported by HHMI, USA.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.idm.2024.10.003.

References

Affinity – Professional Creative Software. Affinity https://affinity.serif.com/en-us/.
Agy, M. B., et al. (1992). Infection of *Macaca nemestrina* by human immunodeficiency virus type-1. *Science, 257*, 103–106.
Barclay, A. N. (1999). Ig-like domains: Evolution from simple interaction molecules to sophisticated antigen recognition. *Proc Natl Acad Sci U S A, 96*, 14672–14674.
Bbosa, N., Kaleebu, P., & Ssemwanga, D. (2019). HIV subtype diversity worldwide. *Current Opinion in HIV and AIDS, 14*, 153–160.
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.
Biswas, S., & Akey, J. M. (2006). Genomic insights into positive selection. *Trends in Genetics, 22*, 437–446.
Bons, E., Leemann, C., Metzner, K. J., & Regoes, R. R. (2020). Long-term experimental evolution of HIV-1 reveals effects of environment and mutational history. *PLoS Biology, 18*, Article e3001010.
Bour, S., Geleziunas, R., & Wainberg, M. A. (1995). The human immunodeficiency virus type 1 (HIV-1) CD4 receptor and its central role in promotion of HIV-1 infection. *Microbiological Reviews, 59*, 31.
Bradley, P., Misura, K. M. S., & Baker, D. (2005). Toward High-Resolution de Novo Structure Prediction for Small Proteins. *Science, 309*, 1868–1871.
Casella, G. (1985). An introduction to empirical Bayes data analysis. *The American Statistician, 39*, 83–87.
Cayley, A. (1857). *On the theory of the analytical forms called trees, 13*(85), 172–176 (The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science.
Chaudhury, S., et al. (2011). Benchmarking and analysis of protein docking performance in Rosetta v3.2. *PLoS One, 6*, Article e22477.
Chaudhury, S., & Gray, J. J. (2008). Conformer selection and induced fit in flexible backbone protein–protein docking using computational and NMR ensembles. *Journal of Molecular Biology, 381*, 1068–1087.
DiMaio, F., Tyka, M. D., Baker, M. L., Chiu, W., & Baker, D. (2009). Refinement of protein structures into low-resolution density maps using Rosetta. *Journal of Molecular Biology, 392*, 181–190.
Dutrillaux, B., Couturier, J., Muleris, M., Lombard, M., & Chauvier, G. (1982). Chromosomal phylogeny of forty-two species or subspecies of cercopithecoids (Primates Catarrhini). *Annales de Genetique, 25*, 96–109.
Explore — Opendatasoft. https://public.opendatasoft.com/explore/?sort=modified.
Farzan, M. *et al.* Tyrosine sulfation of the amino terminus of CCR5 facilitates HIV-1 entry. 10.
Federhen, S. (2012). The NCBI Taxonomy database. *Nucleic Acids Research, 40*, D136–D143.
Fontenot, D., et al. (2007). Critical role of Arg59 in the high-affinity gp120-binding region of CD4 for human immunodeficiency virus type 1 infection. *Virology, 363*, 69–78.
Gao, F., et al. (1999). *Origin of HIV-1 in the chimpanzee, 397* (p. 6).
Gorantla, S., Poluektova, L., & Gendelman, H. E. (2012). Rodent models for HIV-associated neurocognitive disorders. *Trends in Neurosciences, 35*, 197–208.
Hirsch, V. M., Olmsted, R. A., Murphey-Corbt, M., Purcell, R. H., & Johnson, P. R. (1989). *An African primate lentivirus (SIVsm> closely related to HIV-2, 339* (p. 4).
HIV data and statistics. (2023). World Health Organization. https://www.who.int/teams/global-hiv-hepatitis-and-stis-programmes/hiv/strategic-information/hiv-data-and-statistics.
Howe, K. L., et al. (2021). Ensembl 2021. *Nucleic Acids Research, 49*, D884–D891.
IUCN. 2024. The IUCN Red List of Threatened Species. Version 2024-2. https://www.iucnredlist.org. Accessed on June 17, 2022.
Karolchik, D., et al. (2003). The UCSC genome browser database. *Nucleic Acids Research, 31*, 51–54.
Krykbaev, R., McKeating, J., & Jones, I. (1997). Mutant CD4 molecules with improved binding to HIV envelope protein gp120 selected by phage display. *Virology, 234*, 196–202.
Los Alamos National Laboratory. (2021). HIV databases. https://www.hiv.lanl.gov/content/index.
Löytynoja, A., & Goldman, N. (2010). webPRANK: a phylogeny-aware multiple sequence aligner with interactive alignment browser. *BMC Bioinformatics, 11*, 579.
Lusso, P., et al. (1988). Cell-mediated immune response toward viral envelope and core antigens in gibbon apes (Hylobates lar) chronically infected with human immunodeficiency virus-1. *J Immunol, 141*, 2467–2473.
Maanen, M., & Sutton, R. (2003). Rodent models for HIV-1 infection and disease. *CHR, 1*, 121–130.
Misura, K. M. S., & Baker, D. (2005). Progress and challenges in high-resolution refinement of protein structure models. *Proteins: Structure, Function, and Bioinformatics, 59*, 15–29.
Moebius, U., Clayton, L., Abraham, S., Harrison, S., & Reinherz, E. (1992). The human immunodeficiency virus gp120 binding site on CD4: Delineation by quantitative equilibrium and kinetic binding studies of mutants in conjunction with a high-resolution CD4 atomic structure - PubMed. *Journal of Experimental Medicine, 176*, 507–517.
Murphy, W. J., et al. (2001). Resolution of the early placental mammal radiation using Bayesian phylogenetics. *Science, 294*, 2348–2351.
Mwaengo, D. M., & Novembre, F. J. (1998). Molecular cloning and characterization of viruses isolated from chimpanzees with pathogenic human immunodeficiency virus type 1 infections. *Journal of Virology, 72*, 8976–8987.
Niehrs, C., Beilßwanger, R., & Huttner, W. B. (1993). *Protein tyrosine sulfation, 1993 - an update*. https://doi.org/10.1016/0009-2797(94)90068-X
Novembre, F. J., et al. (1997). Development of AIDS in a chimpanzee infected with human immunodeficiency virus type 1. *Journal of Virology, 71*, 4086–4091.
Number of people (all ages) living with HIV. https://www.who.int/data/gho/data/indicators/indicator-details/GHO/estimated-number-of-people-(all-ages)-living-with-hiv, (2021).
Prévost, J., et al. (2020). The HIV-1 env gp120 inner domain shapes the Phe43 cavity and the CD4 binding site. *mBio, 11*, Article e00280, 20.
The PyMOL molecular Graphics System, version 2.0. Schrödinger, LLC.
QGIS. QGIS: A Free and Open Source Geographic Information System https://www.qgis.org/en/site/.
Setash, C. M., Zohdy, S., Gerber, B. D., & Karanewsky, C. J. (2017). A biogeographical perspective on the variation in mouse lemur density throughout Madagascar. *Mammal Review, 47*, 212–229.
Shaik, M., et al. (2019). *Structural basis of coreceptor recognition by HIV-1 envelope spike, 565* (p. 24).
Sharp, P. M. & Hahn, B. H. Origins of HIV and the AIDS pandemic. 22.
Sircar, A., Chaudhury, S., Kilambi, K. P., Berrondo, M., & Gray, J. J. (2010). A generalized approach to sampling backbone conformations with RosettaDock for CAPRI rounds 13-19: Sampling Backbone Conformations with RosettaDock. *Proteins, 78*, 3115–3123.
Stalker, J., et al. (2004). The Ensembl web site: Mechanics of a genome browser. *Genome Research, 14*, 951–955.
Studer, G., et al. (2021). ProMod3—a versatile homology modelling toolbox. *PLoS Computational Biology, 17*, Article e1008667.
Su, C. T.-T., Ling, W.-L., Lua, W.-H., Poh, J.-J., & Gan, S. K.-E. (2017). The role of Antibody Vκ Framework 3 region towards Antigen binding: Effects on recombinant production and Protein L binding. *Scientific Reports, 7*, 3766.
Tamamis, P., & Floudas, C. A. (2014). Molecular recognition of CCR5 by an HIV-1 gp120 V3 loop. *PLoS One, 9*, Article e95767.
Warren, C. J., et al. (2019). Selective use of primate CD4 receptors by HIV-1. *PLoS Biology, 17*, Article e3000304.

Waterhouse, A., Bertoni, M., Bienert, S., Studer, G., Tauriello, G., Gumienny, R., … Schwede, T. (2018). SWISS-MODEL: Homology modelling of protein structures and complexes. *Nucleic Acids Research, 46*, W296–W303.
Waterhouse, A. M., Procter, J. B., Martin, D. M. A., Clamp, M., & Barton, G. J. (2009). Jalview Version 2—a multiple sequence alignment editor and analysis workbench. *Bioinformatics, 25*, 1189–1191.
Welch, B. L. (1947). The generalization of 'Student's' problem when several different population variances are involved. *Biometrika, 34*, 28–35.
Wiederstein, M., & Sippl, M. J. (2007). ProSA-web: Interactive web service for the recognition of errors in three-dimensional structures of proteins. *Nucleic Acids Research, 35*, W407–W410.
Yang, Z. (2007). Paml 4: Phylogenetic analysis by maximum likelihood. *Molecular Biology and Evolution, 24*, 1586–1591.