

RESEARCH ARTICLE

# Protein structural disorder of the envelope V3 loop contributes to the switch in human immunodeficiency virus type 1 cell tropism

Xiaowei Jiang<sup>1,2,3\*</sup>, Felix Feyertag<sup>1,4</sup>, David L. Robertson<sup>1,5\*</sup>

**1** Evolution & Genomic Sciences, School of Biological Sciences, Faculty of Life Sciences, University of Manchester, Manchester, United Kingdom, **2** Department of Genetics, University of Cambridge, Cambridge, United Kingdom, **3** Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, United Kingdom, **4** Department of Biology, University of Nevada, Reno, Nevada, United States of America, **5** MRC-University of Glasgow Centre for Virus Research, Garscube Campus, Glasgow, United Kingdom

\* [xj234@well.ox.ac.uk](mailto:xj234@well.ox.ac.uk) (XJ); [david.l.robertson@glasgow.ac.uk](mailto:david.l.robertson@glasgow.ac.uk) (DLR)



**OPEN ACCESS**

**Citation:** Jiang X, Feyertag F, Robertson DL (2017) Protein structural disorder of the envelope V3 loop contributes to the switch in human immunodeficiency virus type 1 cell tropism. *PLoS ONE* 12(10): e0185790. <https://doi.org/10.1371/journal.pone.0185790>

**Editor:** Eugene A. Permyakov, Russian Academy of Medical Sciences, RUSSIAN FEDERATION

**Received:** June 14, 2017

**Accepted:** September 19, 2017

**Published:** October 19, 2017

**Copyright:** © 2017 Jiang et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

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

**Funding:** This work was supported by Medical Research Council (G1001806/1, <http://gtr.rcuk.ac.uk/project/B49A67AE-9D16-4EB6-9024-6499FF854BAC>) and Wellcome Trust (097820/Z/11/A) to DLR and by the Issac Newton Trust and Wellcome Trust funding to John Welch, University of Cambridge. FF was supported by a BBSRC studentship and XJ by the MRC (G1001806/1) and

## Abstract

Human immunodeficiency virus type 1 (HIV-1) envelope gp120 is partly an intrinsically disordered (unstructured/disordered) protein as it contains regions that do not fold into well-defined protein structures. These disordered regions play important roles in HIV's life cycle, particularly, V3 loop-dependent cell entry, which determines how the virus uses two coreceptors on immune cells, the chemokine receptors CCR5 (R5), CXCR4 (X4) or both (R5X4 virus). Most infecting HIV-1 variants utilise CCR5, while a switch to CXCR4-use occurs in the majority of infections. Why does this 'rewiring' event occur in HIV-1 infected patients? As changes in the charge of the V3 loop are associated with this receptor switch and it has been suggested that charged residues promote structure disorder, we hypothesise that the intrinsic disorder of the V3 loop is permissive to sequence variation thus contributing to the switch in cell tropism. To test this we use three independent data sets of gp120 to analyse V3 loop disorder. We find that the V3 loop of X4 virus has significantly higher intrinsic disorder tendency than R5 and R5X4 virus, while R5X4 virus has the lowest. These results indicate that structural disorder plays an important role in HIV-1 cell tropism and CXCR4 binding. We discuss the potential evolutionary mechanisms leading to the fixation of disorder promoting mutations and the adaptive potential of protein structural disorder in viral host adaptation.

## Introduction

HIV-1 cell entry relies primarily on the binding of viral envelope protein, gp120, to two host immune cell membrane proteins, namely, the CD4 receptor and the coreceptor CCR5 or CXCR4 [1]. Virus that solely uses CCR5 as coreceptor is termed R5 tropic, whereas exclusive CXCR4 using virus is termed X4 tropic, while virus that can use both coreceptors for cell entry is termed dual tropic (R5X4). The switch from CCR5 to CXCR4 is correlated with disease progression and pathogenesis [2, 3]. Interestingly, the CCR5 receptor is not expressed in individuals that present the CCR5 delta-32 mutation [4]. This mutation confers natural resistance to

WT (097820/Z/11/A) grants to DLR. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

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

R5 tropic HIV and non-expression of this receptor appears not to have any significant effect. Based on these observations, the small molecule CCR5 antagonist maraviroc (MVC) was developed to inhibit the CCR5 receptor, thereby blocking virus cell entry [5] and helping to control virus infection.

In all documented cases of MVC virologic failure due to an observed tropism switch, a minority pre-existing X4 using viral population was present prior to therapy, which is ‘unmasked’ by the use of an entry-inhibitor [6]. This indicates that X4 populations may commonly be present, but only transiently dominant, in the life time of an infection. Indeed, in the now classic longitudinal study of Shankarrapa et al. [7], this is readily apparent (see Meehan et al. for visualisations [8]), indicating that the true rates of X4 using virus are much higher than the commonly reported figure of 50% of patients observed to progress to X4 using virus, as confirmed by recent studies [3, 9].

Investigating the HIV-1 coreceptor switch is therefore central to understanding HIV-1 infection and has clinical implications for the use of CCR5 entry inhibitors. It is generally accepted that the V3 loop of gp120 plays a primary role in determining HIV-1 cell tropism and coreceptor specificity [10]. However, our understanding of the mechanisms of HIV-1 tropism remains incomplete [2, 3]. Charged residues in the V3 loop have been shown to affect tropism [11, 12]. Moreover, N-linked glycosylation in the vicinity of the V3 loop has been shown to affect viral tropism [13], e.g., the lack of N-linked glycosylation sites is associated with X4 phenotypes [13–18]. Because N-linked glycosylation plays a role in protein folding and stability [19–22], lack of N-linked glycosylation may decrease V3 loop stability and therefore contribute to X4 using phenotypes. While these changes may collectively lead to the tropism switch, why does this occur in the majority of HIV-1 infections? How can random sequence evolution, given its blind nature, result in a predictable outcome? Could changes in the disorder of gp120 V3 loop play a role in rewiring protein-protein interactions (from CCR5 to CXCR4) or be resulting in a different structure in the context of trimeric Env that results in the change in tropism?

The classical structure-function paradigm states that protein function requires a well-defined three-dimensional structure [23, 24]. However, it has become clear that many functionally important proteins do not have well-defined structures (termed intrinsically disordered proteins), or have protein regions that lack structures (proteins with disordered regions) [23, 25–28]. In some cases the intrinsically disordered regions can form structures after binding to other molecules. HIV-1’s gp120 has five highly variable (V1–V5) and five conserved (C1–C5) protein regions [29]. The variable regions are normally missing in the solved X-ray crystal structures (V1/V2, V3, V4 and V5) of gp120 unless co-crystallized with other binding partners, presumably because they are intrinsically disordered. Disordered proteins or protein regions have been observed to be “sticky” and aggregate potentially forming promiscuous molecular interactions [25, 30, 31]. Disordered regions are frequently found to undergo post-translational modifications, such as phosphorylation or N-linked glycosylation [24, 32]. In addition, amino acid residues in proteins that give rise to structural disorder tend to be polar and charged [24, 25]. Given the disordered nature of HIV-1’s V3 loop, the documented role of charge and glycosylation in the tropism switch, we hypothesize that intrinsic disorder of the V3 loop is permissive to the variation that can lead to a change coreceptor usage and thus switches in cell tropism.

## Materials and methods

### Data and tropism prediction

We retrieved three data sets: (i) Full-length envelope protein (*Env*) sequences of HIV-1 subtype B with tropism information (R5, R5X4 or X4) from the Los Alamos National Laboratory

HIV sequence database (<http://www.hiv.lanl.gov/content/index>). These sequences have annotated coreceptor usage based on biological data only, which is not from sequence-based bioinformatics prediction (<http://www.hiv.lanl.gov/components/sequence/HIV/search/help.html#coreceptor>). And, two inpatient data sets: (ii) A dataset containing C2 to V5 regions of *Env*, as used by Shankarrappa[7]. (iii) A dataset consisting of V1 to V3 regions of *Env*, as used by Mild [16, 33]. We predicted tropism using Geno2pheno[454] with default parameters, and a false positive rate (FPR) cutoff of 5% was used to class sequences as CXCR4-using ('true') or R5 tropic ('false') [34]. Note that Geno2pheno does not distinguish between X4 and R5X4 viruses, this means that sequences classed as true could be either X4- or dual-tropic, which has important implications when interpreting the statistical test results. All sequences used in this study with tropism annotation are in [S1 File](#).

## Comparison of protein disorder in the V3 loop of CCR5, CXCR4 and CCR5-CXCR4 tropic viruses

Bioinformatics methods for characterizing probable disordered regions in amino acid sequences generally fall into two categories: (1) machine-learning methods trained on missing (presumed to be disordered) regions of experimentally solved X-ray crystallographic structures and these trained models used to predict disordered regions in protein sequences or on mobile regions characterised by nuclear magnetic resonance (NMR) [35, 36], or physiochemical properties and pairwise amino acid interaction energies can be calculated to determine likely disordered regions of sequences (2). The former approach is prone to errors in the solved protein structures deposited in Protein Data Bank (<http://www.rcsb.org/>) [37]. We therefore opted to use several methods, IUPred [38], DISOPRED3 [36], PONDR VL-XT [39] and PONDR VSL2 [40], to predict differences in the intrinsic disorder regions between CCR5 and CXCR4 using virus in our datasets, namely R5, R5X4 and X4 envelope sequences.

For each amino acid residue in a protein sequence, IUPred and other methods will report a disorder score from 0 to 1 ranging from complete order (0) to complete disorder (1). We use a cut-off 0.4 to indicate structural disorder ( $\geq 0.4$ ) [41]. We first performed disorder prediction on the full available envelope sequences in our three datasets, and then extracted the V3 region to analyze disorder within the V3 loop. We then group the extracted scores into two groups (CCR5-using and CXCR4-using) according to tropism prediction by Geno2pheno, and compare these statistically using a nonparametric method in the R statistical package [42] (two sample comparisons: nonparametric Behrens-Fisher problem in paired data, <http://cran.r-project.org/web/packages/nparcomp/index.html>). The R scripts for the nonparametric test and V3 loop disorder scores of all patients are in [S2 File](#). We performed three statistical tests. First, the null hypothesis is  $H_0: P(a,b) = 1/2$  (b tends to be similar to a); the alternative hypothesis is  $H_a: P(a,b) < 1/2$  or  $P(a,b) > 1/2$  (b tends not to be similar to a). Second, the null hypothesis is  $H_0: P(a,b) \geq 1/2$  (b tends to be greater than or similar to a); the alternative hypothesis is  $H_a: P(a,b) < 1/2$  (b tends to be less than a). Finally, the null hypothesis is  $H_0: P(a,b) \leq 1/2$  (b tends to be less than or similar to a); the alternative hypothesis is  $H_a: P(a,b) > 1/2$  (b tends to be greater than a).

## Results

First, we investigated whether there were statistically significant differences in structural disorder associated with coreceptor usage. We calculated the disorder tendency of amino acids in the full-length consensus envelope protein sequences (gp120 and gp41) of R5, R5X4 and X4 tropic viruses by four different methods, respectively. This allowed us to obtain an overview of the extent of protein disorder in HIV-1 envelope proteins by tropism with different methods.

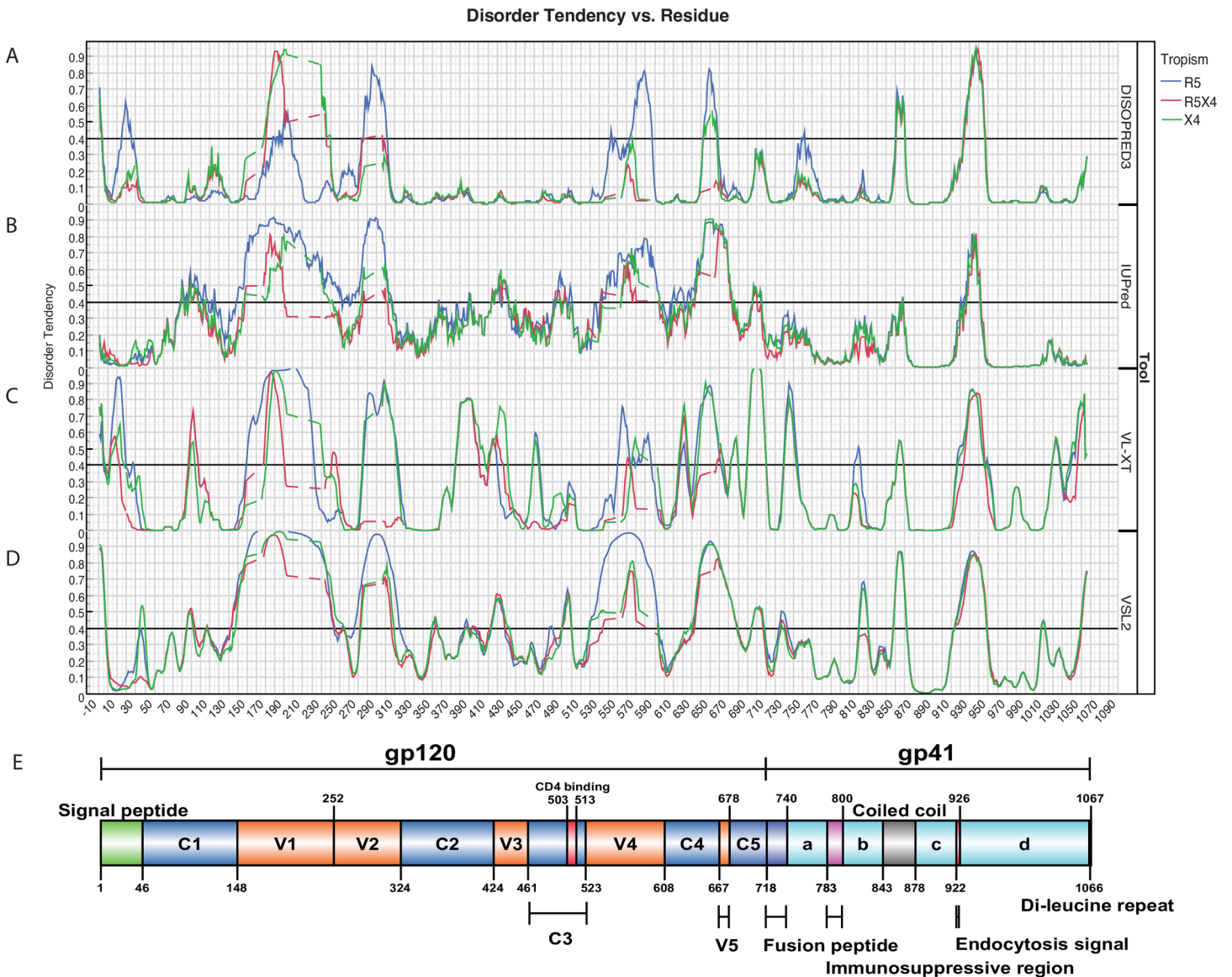
The predicted disorder tendency from the four methods generally coincides with regions missing from experimentally solved crystallized protein structures of gp120 [43] and gp41 [44] (Fig 1). The V3 loop may be folded in the trimeric Env, however, structural studies of the Env trimers demonstrated that the variable regions are indeed structurally flexible (e.g., V1-V3) [45–49]. For example, the electron densities of the longer intrinsically disordered regions such as V1/V2, V4 and V3 are difficult to obtain monomerically without co-crystallization with a stabilizing binding partner. There are several noticeable differences between the predictions of the four methods. First, there are over-predictions of disorder in either N- and/or C-terminus by DISOPRED3, PONDR VL-XT and PONDR VSL2 except in IUPred as noticed before [41]. Second, strikingly, PONDR VL-XT gives direct support of our hypothesis that X4 has the highest disorder propensity of the V3 loop (Fig 1C and 1E) than R5/R5X4 V3 loop. Finally, as IUPred specifically offers an option to predict long disordered regions (>30 amino acids) and there is no over-prediction of the termini (Fig 1) [41], we therefore decided to use disorder tendency scores predicted by IUPred for subsequent statistical analysis.

Next, we focus on the V3 loop to examine how structural disorder changes between coreceptor switches. We found that the disorder tendency of amino acids in the V3 loop of X4 virus was significantly higher than in R5 virus ( $p = 2.35 \times 10^{-3}$ , Fig 2) and R5X4 virus ( $p = 1.87 \times 10^{-16}$ , Fig 2), respectively. Interestingly, there was significantly higher structural disorder in R5 virus compared to R5X4 virus ( $p = 7.49 \times 10^{-6}$ , Fig 2). Thus, dual tropic R5X4 virus has the lowest V3 loop structural disorder tendency, compared to R5 and X4 virus. Collectively, these results suggest that the V3 domain of X4 virus has significantly greater structural disorder tendency than that of R5 and R5X4 virus during coreceptor switch. Dual tropic R5X4 virus may therefore have a less flexible V3 domain than R5 and X4 virus, which may be more prone to neutralizing antibody binding. This is consistent with a previous study that showed dual tropic virus had lower fitness, and was therefore more sensitive to antibody inhibitors and neutralization [50].

To further test our hypothesis, we analyzed two data sets (Shankarappa et al. 1999 and Mild et al. 2013) with longitudinal time points and multiple patients [7, 33]. The Mild et al. dataset also had experimentally verified coreceptor calls. In the Shankarappa dataset, we first predicted coreceptor usage for all sequences at all visits across all nine patients (patient 1, 2, 3, 5, 6, 7, 8, 9 and 11). We compared the structural disorder tendency between X4-capable and R5 virus at each visit that has coreceptor switch events in eight patients (patient 1, 2, 3, 5, 7, 8, 9 and 11). We showed the results for two patients (patient 2 and 9) that are representative of all eight patients, as suggested in Shankarappa et al. study (results for patient 1, 3, 5, 6, 7, 8 and 11 are reported in S1 Fig). In patient 2 we found that X4-capable V3 loop had significantly higher disorder tendency than R5 using V3 loop at visits 7 ( $p < 0.001$ ), 10 ( $p < 0.05$ ), 12 ( $p < 0.001$ ), 13 ( $p < 0.001$ ), 14 ( $p < 0.05$ ), 15 ( $p < 0.001$ ), 17 ( $p < 0.001$ ), 19 ( $p < 0.001$ ) and 23 ( $p < 0.001$ ) except at visit 16 (Fig 3 patient 2). At visit 16 the R5 using V3 loop has significantly higher disorder tendency than the X4-capable V3 loop (the mean disorder tendency of the R5 using V3 loop is also higher than the X4-capable V3 loop), which may indicate the predicted X4-capable virus is actually dual tropic R5X4. For patient 9, it was only at visit 20 that the disorder tendency of the X4-capable V3 loop is significantly higher than that of the R5 using V3 loop ( $p < 0.001$  Fig 3 patient 9). At both visits 3 and 21 the R5 using V3 loop disorder tendency is significantly higher than the X4-capable V3 loop indicating the predicted tropism may be dual tropic R5X4. However, at both visits 17 and 22 there is no significant difference of disorder tendency during coreceptor switch ( $p > 0.05$  Fig 3 patient 9), which may suggest the prediction of tropism is false positive and they have the same tropism.

In the Mild et al. data we first predicted the coreceptor usage for all sequences at each time point within patients. Patients who are predicted to have coreceptor switch events are

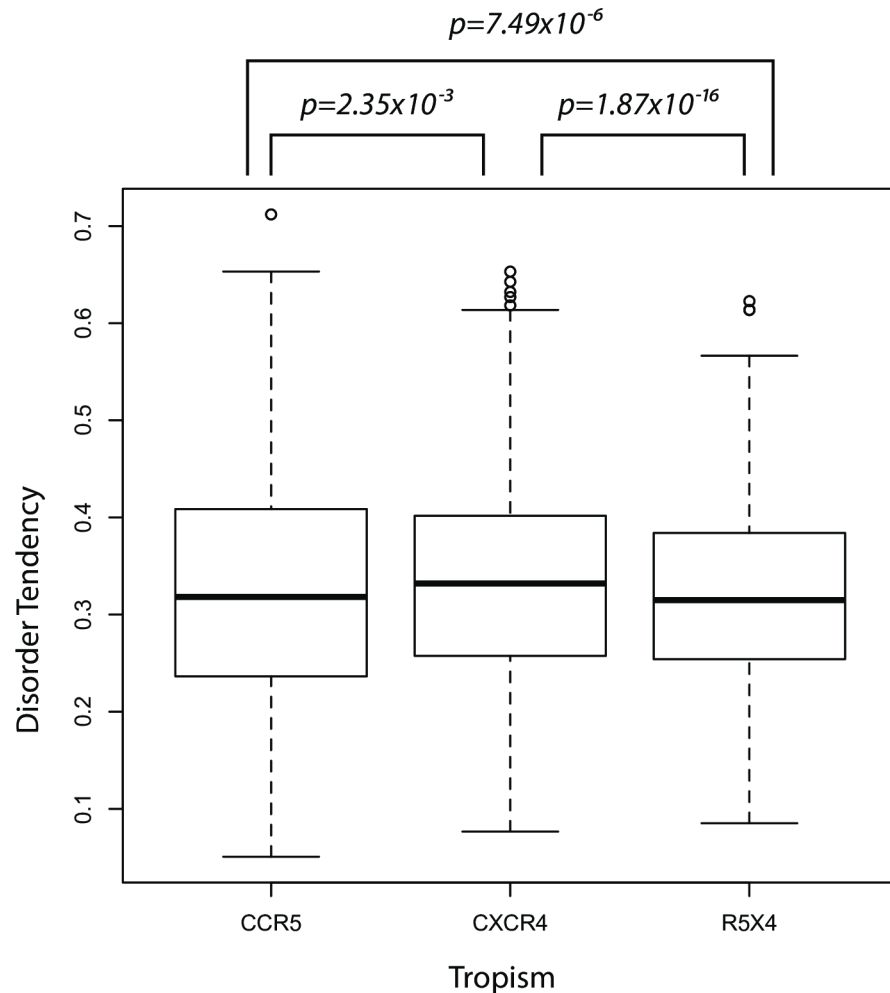




**Fig 1. Plot of predicted structural disorder tendency of the consensus envelope protein.** Disorder tendency for each amino acid is predicted by four prediction methods (0 represents complete order; 1 represents complete disorder; scores over 0.4 represent intrinsically disordered). Disorder tendency scores are plotted for consensus R5, R5X4 and X4 envelope sequence for DISOPRED3 (A), IUPred (B), PONDR VL-XT (C) and PONDR VSL2 (D), respectively. (E) The consensus protein sequences of R5, R5X4 and X4 are aligned to HXB2 envelope sequence (gp120 and gp41) for indicating the structure locations of the disordered residues (adapted from Jiang et al. 2015 [61]). Gp120 and gp41 protein domains are numbered and color-coded for visualisation. Note that here the length of the Env sequence is longer than the normal one because gaps are introduced in the sequence alignment and dashed lines are used for linking sites with missing prediction scores in the gaps.

<https://doi.org/10.1371/journal.pone.0185790.g001>

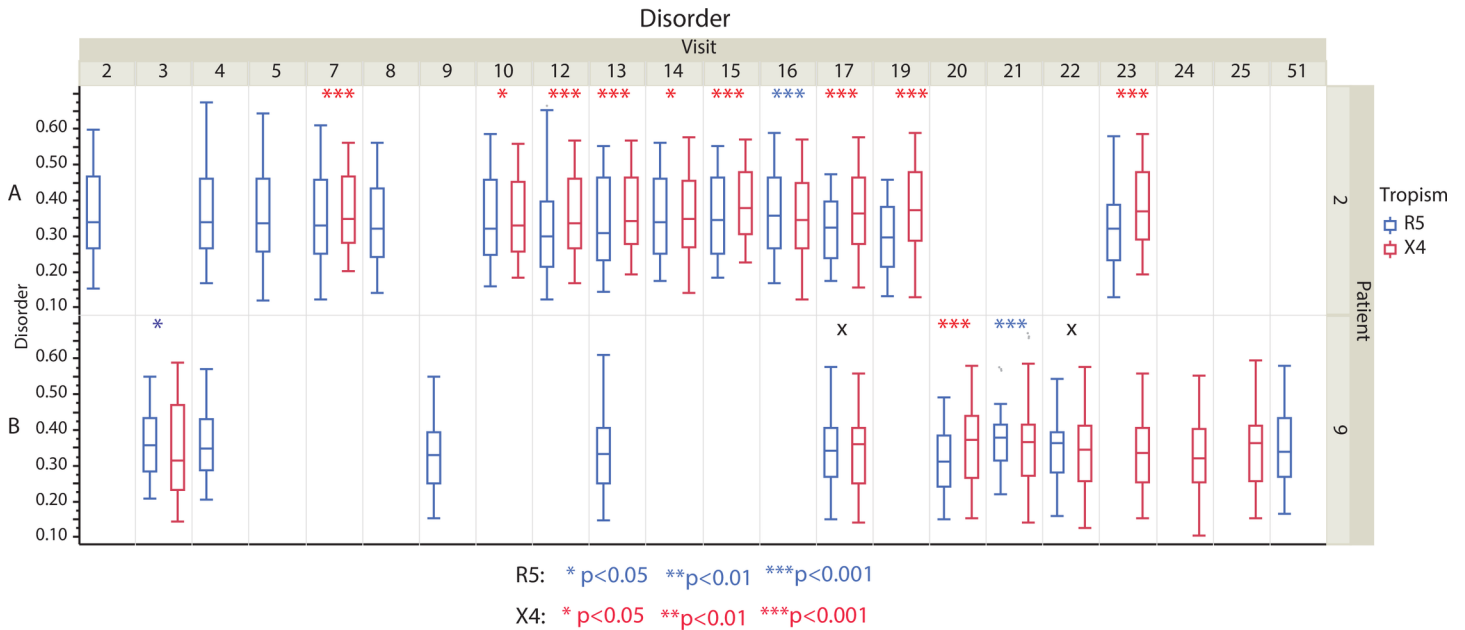
consistent with the experiment verification except 25 months, 45 months, 10 months in patient 2239, 2242 and 2282, respectively (patient 2239, 2242 and 2282 are included in our study while patient 1865 was omitted due to dual tropic R3X4 virus). We then calculated the disorder tendency in R5 and X4 virus (Fig 4), respectively. We then compared the disorder tendency between R5 and X4 virus at each time point of all three patients that have coreceptor switch events. In patient 2239 we find that there are three time points that have coreceptor switch events (time points 25 months, 68 months and 88 months). At two of the three time points the X4-capable V3 loop has significantly higher disorder tendency than the R5 using V3 loop



**Fig 2. Boxplot and nonparametric comparisons of the V3 loop disorder tendency.** Disorder scores are predicted by IUPred for all R5, R5X4 and X4 envelope sequences, respectively. Disorder tendency scores are compared by a nonparametric method described in Materials and Methods. The X4 V3 loop has significantly higher disorder scores than R5 ( $p = 2.35 \times 10^{-3}$ ) and R5X4 ( $p = 1.87 \times 10^{-16}$ ) virus, respectively. The V3 loop of R5 virus also has significantly higher disorder scores than R5X4 virus ( $p = 7.49 \times 10^{-6}$ ).

<https://doi.org/10.1371/journal.pone.0185790.g002>

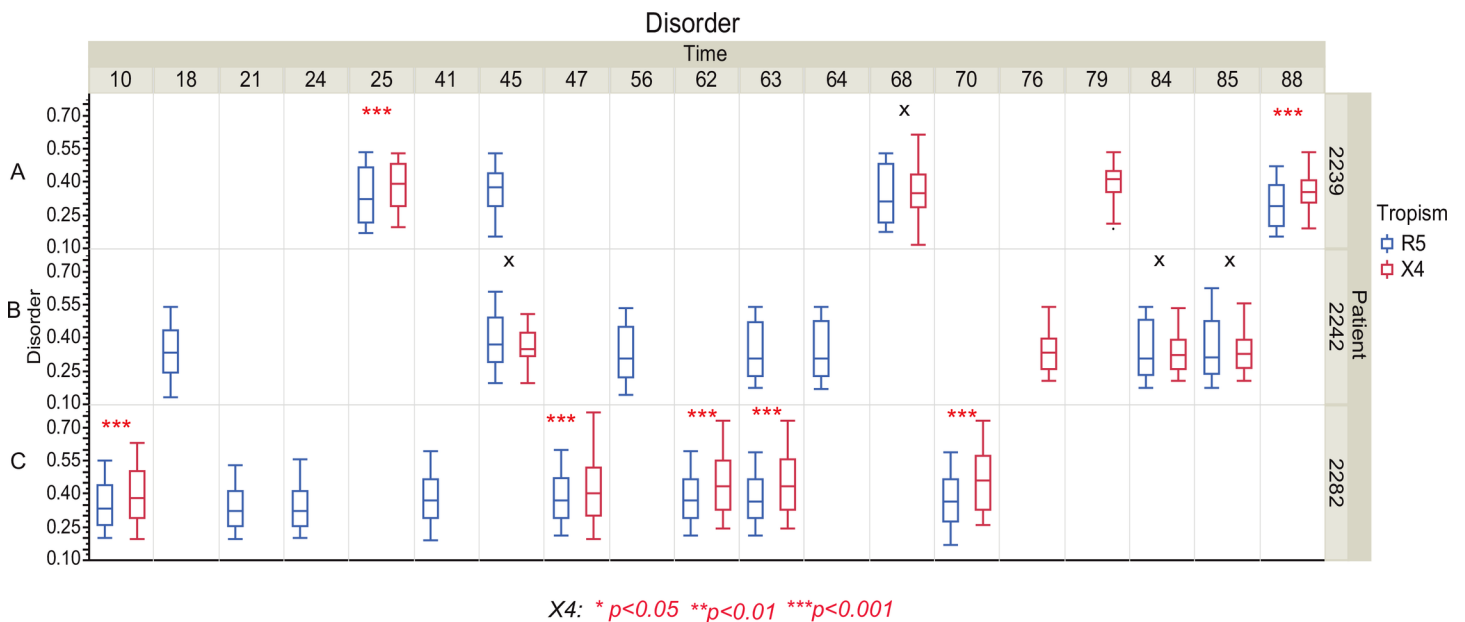
( $p < 0.001$ ). In patient 2242 we find three time points that have coreceptor switch events (time points 45 months, 84 months and 85 months). However, none of the three switch events involves any significant change of V3 loop disorder tendency ( $p > 0.05$ , Fig 4 patient 2242). In patient 2282 we find 5 time points that have coreceptor switch events (time points 10 months, 47 months, 62 months, 63 months and 70 months, Fig 4 patient 2282). Interestingly, all X4 V3 loops have significantly higher structural disorder tendency than R5 using V3 loops ( $p < 0.001$ , Fig 4 patient 2282). These results are compatible with our hypothesis: Firstly the X4-capable virus had significantly higher structural disorder than the R5 tropic virus (time points 25 months and 88 months in patient 2239; time points 10 months, 47 months, 62 months 63 months and 70 months in patient 2282, Fig 4). Secondly, the structural disorder is not significantly different between R5 and X4 tropic virus (time points 45 months, 84 months and 85 months in patient 2242, Fig 4). This is not consistent with our hypothesis indicating the predicted tropism is due to a false prediction and therefore we may compare viruses with the same tropism. Moreover, the mean disorder tendency in each comparison was also consistent



**Fig 3. Boxplot and nonparametric comparisons of the V3 loop disorder tendency between R5 and X4 viruses from Shankarappa et al. data [7].** Boxplot and statistical comparisons are shown for patient 2 (A) and 9 (B), respectively (\* $p < 0.05$ , \*\* $p < 0.01$ , \*\*\* $p < 0.001$ , asterisks colored with blue and red represent results for R5 and X4 virus, respectively; X means  $p > 0.05$ ).

<https://doi.org/10.1371/journal.pone.0185790.g003>

with our hypothesis, although some comparisons did not reach statistical significance (except time point 45 months in patient 2242).



**Fig 4. Boxplot and nonparametric comparisons of the V3 loop disorder tendency between R5 and X4 viruses from Mild et al. data [33].** Boxplot and statistical comparisons are shown for patient 2239 (A), 2242 (B) and 2282 (C), respectively (\* $p < 0.05$ , \*\* $p < 0.01$ , \*\*\* $p < 0.001$ , asterisks colored with red represent results for X4 virus; X means  $p > 0.05$ ).

<https://doi.org/10.1371/journal.pone.0185790.g004>

## Discussion

In this study we have used a nonparametric statistical approach to investigate if there are distinct patterns of predicted structural disorder in the V3 domain of HIV-1 envelope protein between R5, R5X4 and X4 tropic viruses. Interestingly, there seems to be an increase in the predicted structural disorder tendency of V3 domain from R5/R5X4 tropic virus to X4 tropic virus, which indicates that structural disorder (flexibility) of the V3 loop plays a role in HIV-1 cell tropism. Our disorder prediction of gp120 and gp41 generally agrees with previous studies [51]. The difference in disorder propensity of the V3 loop for preferential coreceptor usage (CCR5 and/or CXCR4) with the binding of gp120 to CD4 offers a new model to understand HIV-1 tropism and coreceptor switch. The flexibility of V3 loop and other variable regions (e.g., V1 and V2) are well known. The “snapshots” of the V3 loop presented in crystal or cryoEM structures are not at their functional state as there is no experimental gp120-CCR5/CXCR4 complex available with the V3 loop present. Moreover, in order to get either the crystal or the cryo-electron microscopy (cryoEM) structures particular chemicals and/or antibodies are added before and/or after the purification of the protein for structural determination (see the “Methods” or “Experimental procedures” section in previous studies such as [44]). Thus, the appearance of a predicted disordered region in a structure does not necessarily mean it is not flexible or disordered as it can be stabilised by added chemicals and/or antibodies, and it also does not necessarily mean the “snapshot” of the disordered region is in its native functional state. The flexibility of the V3 loop in trimeric Env is also demonstrated in several structural and molecular dynamic simulation studies [45–49]. Even if the V3 loop is structured in the context of the trimer it is this context that is giving the structure.

Our statistical comparisons of the predicted disorder tendency between disordered V3 loops capable of using different chemokine receptors (CCR5 and/or CXCR4) are consistent with our hypothesis that 1) the V3 loop of X4 using virus has a significantly higher disorder tendency than that of R5 virus; and 2) the V3 loop of X4 and R5 virus has significantly higher disorder than the dual tropic (R5X4) virus. As protein disorder is a source of novel protein-protein interactions [23, 24, 30, 37, 52], such a significant shift of protein disorder tendency is presumably contributing to rewiring of protein-protein interactions [52].

In terms of a mechanism for the switch, intrinsically disordered regions are relatively unconstrained resulting in the accumulation of neutral or nearly neutral residue changes in line with the neutral theory of molecular evolution [53, 54]. Random changes may lead by chance to some affinity for the CXCR4 coreceptor, i.e., dual tropic intermediate viruses [50] which selection may then act upon resulting in X4 virus. Given the main receptor used in recognition by HIV-1 is CD4, switching between the spatially proximal (and paralogous) coreceptor molecules CCR5 and CXCR4 on the different cell types could be relatively straightforward due to the unconstrained nature of disordered regions.

The higher disorder tendency of the X4 virus V3 loop suggests it may be ‘stickier’ and able to use CXCR4 coreceptor more efficiently and therefore cause further infection of immune cells that express CXCR4. Or alternatively it forms a different structure in the context of the trimeric Env and this lead to the switch in tropism. On the other hand, the increased disorder tendency of the V3 loop may make the virus more promiscuous in binding to immune cell coreceptors for cell entry (e.g., not just CXCR4) [30, 55].

We speculate that that structural disorder of the V3 loop contributes to function by being stabilized by either the context of the Env trimer and/or secondary interactions leading to coreceptor switch, which could be another protein (e.g., CXCR4 or other chemokine receptors [55]) or posttranslational modifications such as N-linked glycosylations as suggested in other studies [13–20, 22]. Evidence from several previous studies supports our hypothesis. First,



non-switching viral populations (remain R5 using) have increased N-linked glycosylation sites, which has a stabilization role [17, 19–22]. Second, molecular-dynamics simulation study also suggests that X4 binding stabilizes the V3 loop [56]. Third, many previous studies demonstrate that increased charge in the V3 loop play a significant role in coreceptor switch [13, 16, 17, 57, 58]. It is well known that charged residues promote structural disorder [24], which leads to increased disorder in X4 using V3 loop. However, if the V3 loop with increased disorder can be stabilized by increased N-glycans or a combination of N-glycans and other secondary interactions the virus could remain R5 tropic [13–18, 22]. In conclusion, understanding the evolutionary mechanisms that lead to the fixation of virus protein structural disorder may hold the key to understand viral host adaptation and develop novel intervention strategies. The prediction of intrinsic disorder can be informative in determining that a region may be more likely to form protein-protein interactions with other proteins or under post-translational modifications (PTM) [59], such as glycosylation, phosphorylation, disulfide-bond formation and many others [60]. So disorder analysis can be a useful approach to identify candidate regions for PTM and protein-protein interaction [59]. This also helps to explain why a particular N-linked glycosylation may be important to determine CCR5 tropism [13]. Nevertheless, we caution that further experimental studies must be used to confirm our hypothesis. The HIV-1 structural disorder mediated coreceptor switch provides a unique model to study change in specificity of a protein-protein interaction, which provides a mechanistic understanding of HIV-1 cell tropism. Future research aiming at understanding viral host adaptation could benefit from elucidating the molecular evolutionary mechanisms that lead to the fixation of structural disorder promoting mutations in the virus-host protein-protein interaction networks, as these mutations may be key to host adaptations and cross-species transmission.

## Supporting information

**S1 File. All sequences with tropism annotation used in this study.**

(ZIP)

**S2 File. The R script and V3 loop disorder scores of all patients for the nonparametric tests.**

(ZIP)

**S1 Fig. Boxplot and nonparametric comparisons of the V3 loop disorder tendency between R5 and X4 viruses from all patients (1, 2, 3, 5, 6, 7, 8, 9 and 11) in Shankarappa et al. data (1).** (\* $p < 0.05$ , \*\* $p < 0.01$ , \*\*\* $p < 0.001$ , asterisks colored with blue and red represent results for R5 and X4 virus, respectively; X means  $p > 0.05$ ).

(PDF)

## Acknowledgments

We would like to thank Simon Lovell for helpful discussion, and Frank Konietschke and Marius Placzek for providing an updated version of the R Nparcomp package. FF was supported by a BBSRC studentship and XJ by the MRC (G1001806/1) and WT (097820/Z/11/A) grants to DLR, and by the Issac Newton Trust and WT funding to John Welch, University of Cambridge.

## Author Contributions

**Conceptualization:** Xiaowei Jiang, David L. Robertson.

**Data curation:** Xiaowei Jiang.

**Formal analysis:** Xiaowei Jiang, Felix Feyertag.

**Funding acquisition:** David L. Robertson.

**Investigation:** Xiaowei Jiang, Felix Feyertag, David L. Robertson.

**Methodology:** Xiaowei Jiang, Felix Feyertag.

**Software:** Xiaowei Jiang.

**Supervision:** Xiaowei Jiang, David L. Robertson.

**Writing – original draft:** Xiaowei Jiang, Felix Feyertag, David L. Robertson.

**Writing – review & editing:** Xiaowei Jiang, David L. Robertson.

## References

1. Wu L, Gerard NP, Wyatt R, Choe H, Parolin C, Ruffing N, et al. CD4-induced interaction of primary HIV-1 gp120 glycoproteins with the chemokine receptor CCR-5. *Nature*. 1996; 384(6605):179–83. <https://doi.org/10.1038/384179a0> PMID: 8906795.
2. Gorry PR, Ancuta P. Coreceptors and HIV-1 pathogenesis. *Current HIV/AIDS reports*. 2011; 8(1):45–53. Epub 2010/12/29. <https://doi.org/10.1007/s11904-010-0069-x> PMID: 21188555.
3. Regoes RR, Bonhoeffer S. The HIV coreceptor switch: a population dynamical perspective. *Trends in microbiology*. 2005; 13(6):269–77. <https://doi.org/10.1016/j.tim.2005.04.005> PMID: 15936659.
4. Dean M, Carrington M, Winkler C, Huttley GA, Smith MW, Allikmets R, et al. Genetic restriction of HIV-1 infection and progression to AIDS by a deletion allele of the *CCR5* structural gene. Hemophilia Growth and Development Study, Multicenter AIDS Cohort Study, Multicenter Hemophilia Cohort Study, San Francisco City Cohort, ALIVE Study. *Science*. 1996; 273(5283):1856–62. PMID: 8791590.
5. Dorr P, Westby M, Dobbs S, Griffin P, Irvine B, Macartney M, et al. Maraviroc (UK-427,857), a potent, orally bioavailable, and selective small-molecule inhibitor of chemokine receptor CCR5 with broad-spectrum anti-human immunodeficiency virus type 1 activity. *Antimicrobial agents and chemotherapy*. 2005; 49(11):4721–32. <https://doi.org/10.1128/AAC.49.11.4721-4732.2005> PMID: 16251317; PubMed Central PMCID: PMC1280117.
6. Archer J, Rambaut A, Taillon BE, Harrigan PR, Lewis M, Robertson DL. The evolutionary analysis of emerging low frequency HIV-1 CXCR4 using variants through time—an ultra-deep approach. *PLoS computational biology*. 2010; 6(12):e1001022. <https://doi.org/10.1371/journal.pcbi.1001022> PMID: 21187908; PubMed Central PMCID: PMC3002995.
7. Shankarappa R, Margolick JB, Gange SJ, Rodrigo AG, Upchurch D, Farzadegan H, et al. Consistent viral evolutionary changes associated with the progression of human immunodeficiency virus type 1 infection. *Journal of virology*. 1999; 73(12):10489–502. PMID: 10559367; PubMed Central PMCID: PMC113104.
8. Meehan CJ, Hedge JA, Robertson DL, McCormack GP, Travers SA. Emergence, dominance, and possible decline of CXCR4 chemokine receptor usage during the course of HIV infection. *Journal of medical virology*. 2010; 82(12):2004–12. <https://doi.org/10.1002/jmv.21922> PMID: 20981786.
9. Esbjornsson J, Mansson F, Martinez-Arias W, Vincic E, Biague AJ, da Silva ZJ, et al. Frequent CXCR4 tropism of HIV-1 subtype A and CRF02\_AG during late-stage disease—indication of an evolving epidemic in West Africa. *Retrovirology*. 2010; 7:23. <https://doi.org/10.1186/1742-4690-7-23> PMID: 20307309; PubMed Central PMCID: PMC2855529.
10. Hwang SS, Boyle TJ, Lyerly HK, Cullen BR. Identification of the envelope V3 loop as the primary determinant of cell tropism in HIV-1. *Science*. 1991; 253(5015):71–4. PMID: 1905842.
11. Moore JP, Kitchen SG, Pugach P, Zack JA. The CCR5 and CXCR4 coreceptors—central to understanding the transmission and pathogenesis of human immunodeficiency virus type 1 infection. *AIDS research and human retroviruses*. 2004; 20(1):111–26. <https://doi.org/10.1089/088922204322749567> PMID: 15000703.
12. Bartlett AD, Macartney MJ, Conibear TC, Feyertag F, Smith CJ, Johnson MA, et al. The utility of different bioinformatics algorithms for genotypic HIV-1 tropism testing in a large clinical cohort with multiple subtypes. *Aids*. 2014. <https://doi.org/10.1097/QAD.000000000000288> PMID: 24785955.
13. Pollakis G, Kang S, Kliphuis A, Chalaby MI, Goudsmit J, Paxton WA. N-linked glycosylation of the HIV type-1 gp120 envelope glycoprotein as a major determinant of CCR5 and CXCR4 coreceptor utilization.

- The Journal of biological chemistry. 2001; 276(16):13433–41. <https://doi.org/10.1074/jbc.M009779200> PMID: 11278567.
14. Li Y, Rey-Cuille MA, Hu SL. N-linked glycosylation in the V3 region of HIV type 1 surface antigen modulates coreceptor usage in viral infection. *AIDS research and human retroviruses*. 2001; 17(16):1473–9. <https://doi.org/10.1089/08892220152644179> PMID: 11709091.
  15. Polzer S, Dittmar MT, Schmitz H, Schreiber M. The N-linked glycan g15 within the V3 loop of the HIV-1 external glycoprotein gp120 affects coreceptor usage, cellular tropism, and neutralization. *Virology*. 2002; 304(1):70–80. PMID: 12490404.
  16. Mild M, Kvist A, Esbjornsson J, Karlsson I, Fenyo EM, Medstrand P. Differences in molecular evolution between switch (R5 to R5X4/X4-tropic) and non-switch (R5-tropic only) HIV-1 populations during infection. *Infection, genetics and evolution: journal of molecular epidemiology and evolutionary genetics in infectious diseases*. 2010; 10(3):356–64. <https://doi.org/10.1016/j.meegid.2009.05.003> PMID: 19446658.
  17. Edo-Matas D, Rachinger A, Setiawan LC, Boeser-Nunnink BD, van 't Wout AB, Lemey P, et al. The evolution of human immunodeficiency virus type-1 (HIV-1) envelope molecular properties and coreceptor use at all stages of infection in an HIV-1 donor-recipient pair. *Virology*. 2012; 422(1):70–80. <https://doi.org/10.1016/j.virol.2011.10.005> PMID: 22047989.
  18. Tsuchiya K, Ode H, Hayashida T, Kakizawa J, Sato H, Oka S, et al. Arginine insertion and loss of N-linked glycosylation site in HIV-1 envelope V3 region confer CXCR4-tropism. *Scientific reports*. 2013; 3:2389. <https://doi.org/10.1038/srep02389> PMID: 23925152; PubMed Central PMCID: PMC3737504.
  19. Scanlan CN, Offer J, Zitzmann N, Dwek RA. Exploiting the defensive sugars of HIV-1 for drug and vaccine design. *Nature*. 2007; 446(7139):1038–45. <https://doi.org/10.1038/nature05818> PMID: 17460665.
  20. Culyba EK, Price JL, Hanson SR, Dhar A, Wong CH, Gruebele M, et al. Protein native-state stabilization by placing aromatic side chains in N-glycosylated reverse turns. *Science*. 2011; 331(6017):571–5. <https://doi.org/10.1126/science.1198461> PMID: 21292975; PubMed Central PMCID: PMC3099596.
  21. Land A, Braakman I. Folding of the human immunodeficiency virus type 1 envelope glycoprotein in the endoplasmic reticulum. *Biochimie*. 2001; 83(8):783–90. PMID: 11530211.
  22. Wood NT, Fadda E, Davis R, Grant OC, Martin JC, Woods RJ, et al. The influence of N-linked glycans on the molecular dynamics of the HIV-1 gp120 V3 loop. *PloS one*. 2013; 8(11):e80301. <https://doi.org/10.1371/journal.pone.0080301> PMID: 24303005; PubMed Central PMCID: PMC3841175.
  23. Dyson HJ, Wright PE. Intrinsically unstructured proteins and their functions. *Nature reviews Molecular cell biology*. 2005; 6(3):197–208. <https://doi.org/10.1038/nrm1589> PMID: 15738986.
  24. Habchi J, Tompa P, Longhi S, Uversky VN. Introducing Protein Intrinsic Disorder. *Chemical reviews*. 2014. <https://doi.org/10.1021/cr400514h> PMID: 24739139.
  25. Uversky VN, Oldfield CJ, Dunker AK. Intrinsically disordered proteins in human diseases: introducing the D2 concept. *Annual review of biophysics*. 2008; 37:215–46. <https://doi.org/10.1146/annurev.biophys.37.032807.125924> PMID: 18573080.
  26. Ferreon AC, Ferreon JC, Wright PE, Deniz AA. Modulation of allostery by protein intrinsic disorder. *Nature*. 2013; 498(7454):390–4. <https://doi.org/10.1038/nature12294> PMID: 23783631; PubMed Central PMCID: PMC3718496.
  27. Uversky VN. A decade and a half of protein intrinsic disorder: Biology still waits for physics. *Protein science: a publication of the Protein Society*. 2013. <https://doi.org/10.1002/pro.2261> PMID: 23553817.
  28. Tompa P. Intrinsically disordered proteins: a 10-year recap. *Trends in biochemical sciences*. 2012; 37(12):509–16. <https://doi.org/10.1016/j.tibs.2012.08.004> PMID: 22989858.
  29. Starcich BR, Hahn BH, Shaw GM, McNeely PD, Modrow S, Wolf H, et al. Identification and characterization of conserved and variable regions in the envelope gene of HTLV-III/LAV, the retrovirus of AIDS. *Cell*. 1986; 45(5):637–48. PMID: 2423250.
  30. Vavouri T, Semple JI, Garcia-Verdugo R, Lehner B. Intrinsic protein disorder and interaction promiscuity are widely associated with dosage sensitivity. *Cell*. 2009; 138(1):198–208. <https://doi.org/10.1016/j.cell.2009.04.029> PMID: 19596244.
  31. Levy ED, De S, Teichmann SA. Cellular crowding imposes global constraints on the chemistry and evolution of proteomes. *Proceedings of the National Academy of Sciences of the United States of America*. 2012; 109(50):20461–6. <https://doi.org/10.1073/pnas.1209312109> PMID: 23184996; PubMed Central PMCID: PMC3528536.
  32. Bailey JR, Lassen KG, Yang HC, Quinn TC, Ray SC, Blankson JN, et al. Neutralizing antibodies do not mediate suppression of human immunodeficiency virus type 1 in elite suppressors or selection of plasma virus variants in patients on highly active antiretroviral therapy. *Journal of virology*. 2006; 80(10):4758–70. <https://doi.org/10.1128/JVI.80.10.4758-4770.2006> PMID: 16641269; PubMed Central PMCID: PMC1472047.

33. Mild M, Gray RR, Kvist A, Lemey P, Goodenow MM, Fenyo EM, et al. High inpatient HIV-1 evolutionary rate is associated with CCR5-to-CXCR4 coreceptor switch. *Infection, genetics and evolution: journal of molecular epidemiology and evolutionary genetics in infectious diseases*. 2013. <https://doi.org/10.1016/j.meegid.2013.05.004> PMID: 23672855.
34. Lengauer T, Sander O, Sierra S, Thielen A, Kaiser R. Bioinformatics prediction of HIV coreceptor usage. *Nature biotechnology*. 2007; 25(12):1407–10. <https://doi.org/10.1038/nbt1371> PMID: 18066037.
35. Oates ME, Romero P, Ishida T, Ghalwash M, Mizianty MJ, Xue B, et al. D(2)P(2): database of disordered protein predictions. *Nucleic Acids Res*. 2013; 41(Database issue):D508–16. <https://doi.org/10.1093/nar/gks1226> PMID: 23203878; PubMed Central PMCID: PMC3531159.
36. Jones DT, Cozzetto D. DISOPRED3: precise disordered region predictions with annotated protein-binding activity. *Bioinformatics*. 2015; 31(6):857–63. <https://doi.org/10.1093/bioinformatics/btu744> PMID: 25391399; PubMed Central PMCID: PMC34380029.
37. Hegyi H, Tompa P. Intrinsically disordered proteins display no preference for chaperone binding in vivo. *PLoS computational biology*. 2008; 4(3):e1000017. <https://doi.org/10.1371/journal.pcbi.1000017> PMID: 18369417; PubMed Central PMCID: PMC2265518.
38. Dosztanyi Z, Csizmek V, Tompa P, Simon I. The pairwise energy content estimated from amino acid composition discriminates between folded and intrinsically unstructured proteins. *Journal of molecular biology*. 2005; 347(4):827–39. <https://doi.org/10.1016/j.jmb.2005.01.071> PMID: 15769473.
39. Romero P, Obradovic Z, Li X, Garner EC, Brown CJ, Dunker AK. Sequence complexity of disordered protein. *Proteins*. 2001; 42(1):38–48. PMID: 11093259.
40. Peng K, Radivojac P, Vucetic S, Dunker AK, Obradovic Z. Length-dependent prediction of protein intrinsic disorder. *BMC Bioinformatics*. 2006; 7:208. <https://doi.org/10.1186/1471-2105-7-208> PMID: 16618368; PubMed Central PMCID: PMC1479845.
41. Ortiz JF, MacDonald ML, Masterson P, Uversky VN, Siltberg-Liberles J. Rapid evolutionary dynamics of structural disorder as a potential driving force for biological divergence in flaviviruses. *Genome biology and evolution*. 2013; 5(3):504–13. <https://doi.org/10.1093/gbe/evt026> PMID: 23418179; PubMed Central PMCID: PMC3622304.
42. Konietzschke F, Pauly M. A studentized permutation test for the nonparametric Behrens-Fisher problem in paired data. *Electron J Stat*. 2012; 6:1358–72. doi: [10.1214/12-Ejs714](https://doi.org/10.1214/12-Ejs714). WOS:000306921000001.
43. Kwong PD, Wyatt R, Robinson J, Sweet RW, Sodroski J, Hendrickson WA. Structure of an HIV gp120 envelope glycoprotein in complex with the CD4 receptor and a neutralizing human antibody. *Nature*. 1998; 393(6686):648–59. Epub 1998/06/26. <https://doi.org/10.1038/31405> PMID: 9641677.
44. Weissenhorn W, Dessen A, Harrison SC, Skehel JJ, Wiley DC. Atomic structure of the ectodomain from HIV-1 gp41. *Nature*. 1997; 387(6631):426–30. Epub 1997/05/22. <https://doi.org/10.1038/387426a0> PMID: 9163431.
45. Ozorowski G, Pallesen J, de Val N, Lyumkis D, Cottrell CA, Torres JL, et al. Open and closed structures reveal allostery and pliability in the HIV-1 envelope spike. *Nature*. 2017; 547(7663):360–3. <https://doi.org/10.1038/nature23010> PMID: 28700571; PubMed Central PMCID: PMC5538736.
46. Moscico CG, Xing L, Hui J, Hu J, Kalkhoran MB, Yenigun OM, et al. Trimeric HIV Env provides epitope occlusion mediated by hypervariable loops. *Scientific reports*. 2014; 4:7025. <https://doi.org/10.1038/srep07025> PMID: 25395053; PubMed Central PMCID: PMC4231788.
47. Bartesaghi A, Merk A, Borgnia MJ, Milne JL, Subramaniam S. Prefusion structure of trimeric HIV-1 envelope glycoprotein determined by cryo-electron microscopy. *Nat Struct Mol Biol*. 2013; 20(12):1352–7. <https://doi.org/10.1038/nsmb.2711> PMID: 24154805; PubMed Central PMCID: PMC3917492.
48. Pancera M, Zhou T, Druz A, Georgiev IS, Soto C, Gorman J, et al. Structure and immune recognition of trimeric pre-fusion HIV-1 Env. *Nature*. 2014; 514(7523):455–61. <https://doi.org/10.1038/nature13808> PMID: 25296255; PubMed Central PMCID: PMC4348022.
49. Chandramouli B, Chillemi G, Desideri A. Structural dynamics of V3 loop in a trimeric ambience, a molecular dynamics study on gp120-CD4 trimeric mimic. *J Struct Biol*. 2014; 186(1):132–40. <https://doi.org/10.1016/j.jsb.2014.02.014> PMID: 24583235.
50. Tasca S, Ho SH, Cheng-Mayer C. R5X4 viruses are evolutionary, functional, and antigenic intermediates in the pathway of a simian-human immunodeficiency virus coreceptor switch. *Journal of virology*. 2008; 82(14):7089–99. <https://doi.org/10.1128/JVI.00570-08> PMID: 18480460; PubMed Central PMCID: PMC2446975.
51. Xue B, Mizianty MJ, Kurgan L, Uversky VN. Protein intrinsic disorder as a flexible armor and a weapon of HIV-1. *Cell Mol Life Sci*. 2012; 69(8):1211–59. <https://doi.org/10.1007/s00018-011-0859-3> PMID: 22033837.

52. Mosca R, Pache RA, Aloy P. The role of structural disorder in the rewiring of protein interactions through evolution. *Molecular & cellular proteomics: MCP*. 2012; 11(7):M111 014969. <https://doi.org/10.1074/mcp.M111.014969> PMID: 22389433; PubMed Central PMCID: PMC3394948.
53. Kimura M. *The neutral theory of molecular evolution*. Cambridge Cambridgehire; New York: Cambridge University Press; 1983. xv, 367 p. p.
54. Ohta T. Slightly deleterious mutant substitutions in evolution. *Nature*. 1973; 246(5428):96–8. PMID: 4585855.
55. Berger EA, Murphy PM, Farber JM. Chemokine receptors as HIV-1 coreceptors: roles in viral entry, tropism, and disease. *Annual review of immunology*. 1999; 17:657–700. <https://doi.org/10.1146/annurev.immunol.17.1.657> PMID: 10358771.
56. Tamamis P, Floudas CA. Molecular recognition of CXCR4 by a dual tropic HIV-1 gp120 V3 loop. *Biophysical journal*. 2013; 105(6):1502–14. <https://doi.org/10.1016/j.bpj.2013.07.049> PMID: 24048002; PubMed Central PMCID: PMC3785887.
57. Jensen MA, Li FS, van 't Wout AB, Nickle DC, Shriner D, He HX, et al. Improved coreceptor usage prediction and genotypic monitoring of R5-to-X4 transition by motif analysis of human immunodeficiency virus type 1 env V3 loop sequences. *Journal of virology*. 2003; 77(24):13376–88. <https://doi.org/10.1128/JVI.77.24.13376-13388.2003> PMID: 14645592; PubMed Central PMCID: PMC296044.
58. Kalinina OV, Pfeifer N, Lengauer T. Modelling binding between CCR5 and CXCR4 receptors and their ligands suggests the surface electrostatic potential of the co-receptor to be a key player in the HIV-1 tropism. *Retrovirology*. 2013; 10:130. <https://doi.org/10.1186/1742-4690-10-130> PMID: 24215935; PubMed Central PMCID: PMC3833284.
59. Tompa P, Schad E, Tantos A, Kalmar L. Intrinsically disordered proteins: emerging interaction specialists. *Curr Opin Struct Biol*. 2015; 35:49–59. <https://doi.org/10.1016/j.sbi.2015.08.009> PMID: 26402567.
60. Walsh CT, Garneau-Tsodikova S, Gatto GJ Jr. Protein posttranslational modifications: the chemistry of proteome diversifications. *Angew Chem Int Ed Engl*. 2005; 44(45):7342–72. <https://doi.org/10.1002/anie.200501023> PMID: 16267872.
61. Jiang X, Feyertag F, Meehan CJ, McCormack GP, Travers SA, Craig C, et al. Characterizing the Diverse Mutational Pathways Associated with R5-Tropic Maraviroc Resistance: HIV-1 That Uses the Drug-Bound CCR5 Coreceptor. *Journal of virology*. 2015; 89(22):11457–72. <https://doi.org/10.1128/JVI.01384-15> PMID: 26339063; PubMed Central PMCID: PMC4645647.