

1

# 2 **Protein structural disorder of the envelope V3 loop** 3 **contributes to the switch in human immunodeficiency** 4 **virus type 1 cell tropism**

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

6

7 <sup>1</sup>Computational and Evolutionary Biology, Faculty of Life Sciences, University of  
8 Manchester, Manchester, United Kingdom; <sup>2</sup>Department of Genetics, University  
9 of Cambridge, Cambridge, United Kingdom.

10 \*Corresponding authors: [jiangx@tcd.ie](mailto:jiangx@tcd.ie); [david.robertson@manchester.ac.uk](mailto:david.robertson@manchester.ac.uk)

11

12 **Running title:**

13 Envelope V3 loop disorder contributes to HIV1 tropism

14

## 15 **Abstract**

16 Human immunodeficiency virus type 1 (HIV-1) envelope gp120 is partly an  
17 intrinsically disordered (unstructured/disordered) protein as it contains regions  
18 that do not fold into well-defined protein structures. These disordered regions  
19 play important roles in HIV's life cycle, particularly, V3 loop-dependent cell entry,  
20 which determines how the virus uses two coreceptors on immune cells, the  
21 chemokine receptors CCR5 (R5), CXCR4 (X4) or both (R5X4 virus). Most  
22 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 plays a role in determining 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 determining HIV-1 cell tropism and CXCR4 binding. We speculate that changes in N-linked glycosylation associated with tropism change (from R5 to X4) are required to stabilise the V3 loop with increased disorder tendency during HIV-1 evolution. 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.

## IMPORTANCE

HIV-1 cell entry relies on the V3 loop of its heavily glycosylated envelope protein gp120 to bind to a host coreceptor CCR5 or CXCR4. Unraveling the mechanism whereby HIV-1 switches host coreceptor is critical to understanding HIV-1 pathogenesis and development of novel intervention strategies. However, a mechanistic understanding of the switch is limited as no gp120-CCR5/CXCR4 complex is available, due to the intrinsically disordered nature of the V3 loop responsible for coreceptor switch. We hypothesise that shifts of V3 disorder may contribute to HIV-1 coreceptor switch and cell tropism. In this study we compared the disorder tendency of the V3 loop before and after the coreceptor

switch. We find that the coreceptor switch is associated with a significant increase of V3 loop disorder from CCR5 to CXCR4 using. This result provides a mechanistic explanation of coreceptor switch that increasingly disordered V3 loop results in use of a different host coreceptor.

**Keywords:** HIV1 coreceptor switch, HIV1 cell tropism, protein intrinsic disorder, protein-protein interaction

## 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 hypothesised to play an important role in 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 R5 tropic HIV and non-expression of this receptor appears not to have any significant affect. 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)). This indicates that the true rates of X4 using virus are much higher than the commonly reported figure that 50% of patients are observed to progress to X4 using virus, as confirmed by recent studies (3, 9).

Investigating 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 it's blind nature, result in a predictable outcome? Could changes of flexibility in the gp120 V3 loop

play a role in rewiring protein-protein interactions (from CCR5 to CXCR4) and therefore be a determinant of 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). 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 plays a critical role in determining coreceptor usage and switches in cell tropism: rewiring protein-protein interactions by changes in intrinsic disorder.

## 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 Shankarrapa(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.

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

Bioinformatics methods for charactering 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 (2) physiochemical properties and pairwise amino acid interaction energies can be calculated to

determine likely disordered regions of sequences. The former approach is prone to errors in the solved protein structures deposited in Protein Data Bank (<http://www.rcsb.org/>) (35). We therefore opted to use a method, IUPred (36) based on the latter approach to predict differences in intrinsic disorder regions between CCR5 and CXCR4 using virus in our datasets. A recent study comparing this method to other intrinsic disorder predictive methods concluded that IUPred makes the most accurate predictions (35, 37).

For each amino acid residue in a protein sequence, IUPred 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 ( $\geq 0.4$ ) (37). Because IUPred uses a sliding window approach to calculate disorder over a range, 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 (38) (two sample comparisons: nonparametric Behrens-Fisher problem in paired data, <http://cran.r-project.org/web/packages/nparcomp/index.html>). 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 to be smaller or larger than a). Second, the null hypothesis is  $H_0 : P(a, b) \geq 1/2$  (b tends to be similar to/ larger than a); the alternative hypothesis is  $H_a : P(a, b) < 1/2$  (b tends to be smaller than a). Finally, the null

hypothesis is  $H_0 : P(a, b) \leq 1/2$  (b tends to be similar to/smaller than a); the

alternative hypothesis is  $H_a : P(a, b) > 1/2$  (b tends to be larger 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, respectively. This allowed us to obtain an overview of the extent of protein disorder in HIV-1 envelope proteins by tropism. The predicted disorder tendency within the three proteins groups coincides with regions missing from experimentally solved crystallized protein structures of gp120 (39) and gp41 (40)(fig. 1). For example, the electron densities of the longer intrinsically disordered regions such as V1/V2, V4 and V3 are difficult to obtain independently without co-crystallization with a stabilizing binding partner (protein regions longer than 30 amino acids, if solved, they are normally presented as flexible loops in the structure, see Fig 1A, B, C and D).

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. In



sum, 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 (41).

To further test our hypothesis, we analyzed two data sets (Shankarappa et al. data and Mild et al. data) 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 supplementary fig s1). 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 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 finally 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 ( $p < 0.001$ ). 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, which may suggest that the predicted tropism is due to false prediction and therefore we may compare two groups of virus with the same tropism. Moreover, the mean disorder tendency in each comparison was also consistent with our hypothesis, although some comparisons did not reach statistical significance (except time point 45 months in patient 2242).

## 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. Strikingly, there seems to be an increase of 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.

264  
 265 Our statistical comparisons of the predicted disorder tendency between  
 266 disordered V3 loops capable of using different chemokine receptors (CCR5  
 267 and/or CXCR4) are consistent with our hypothesis that 1) the V3 loop of X4  
 268 using virus has significantly higher disorder tendency than that of R5 virus; and  
 269 2) the V3 loop of X4 and R5 virus has significantly higher disorder than the dual  
 270 tropic (R5X4) virus. As protein disorder is a source of novel protein-protein  
 271 interactions (23, 24, 30, 35, 42), such a significant shift of protein disorder  
 272 tendency is presumably contributing to rewiring of protein-protein interactions  
 273 (42).

274  
 275 In terms of a mechanism for the switch, intrinsically disordered regions are  
 276 relatively unconstrained resulting in the accumulation of neutral or nearly  
 277 neutral residue changes in line with the neutral theory of molecular evolution  
 278 (43, 44). Random changes may lead to some affinity for the CXCR4 coreceptor,  
 279 i.e., dual tropic intermediate viruses (41) which selection may then act upon  
 280 resulting in X4 virus.

281  
 282 The higher disorder tendency of the X4 virus V3 loop suggests it may be 'stickier'  
 283 and able to use CXCR4 coreceptor more efficiently and therefore cause further  
 284 infection of immune cells that express CXCR4. Alternatively, increased disorder  
 285 tendency of the V3 loop may make the virus more promiscuous in binding to  
 286 immune cell coreceptors for cell entry (e.g., not just CXCR4) (30, 45).

287

We speculate that that structural disorder of the V3 loop contributes to function by being stabilized by secondary interactions leading to coreceptor switch, which could be another protein (e.g., CXCR4 or other chemokine receptors (45)) 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 (46). Third, many previous studies demonstrate that increased charge in the V3 loop play a significant role in coreceptor switch (13, 16, 17, 47, 48). 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. Here 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

312 promoting mutations in the virus-host protein interaction networks, as these  
313 mutations may be key to host adaptations and cross-species transmission.

314

315

316

317

318

319

## Acknowledgements

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 to DLR. XJ was supported by the MRC (G1001806/1) and WT (097820/Z/11/A) to DLR and by the Issac Newton Trust and WT (PCJW/GAAB) funding to John Welch, University of Cambridge.

## Author contributions

XJ and DLR conceived and designed the study. XJ performed the research. XJ, DLR and FF analysed the data. XJ, DLR and FF wrote the paper. All authors commented on and approved a final version of paper.

## Reference:

1. **Wu L, Gerard NP, Wyatt R, Choe H, Parolin C, Ruffing N, Borsetti A, Cardoso AA, Desjardin E, Newman W, Gerard C, Sodroski J.** 1996. CD4-induced interaction of primary HIV-1 gp120 glycoproteins with the chemokine receptor CCR-5. *Nature* **384**:179-183.
2. **Gorry PR, Ancuta P.** 2011. Coreceptors and HIV-1 pathogenesis. *Current HIV/AIDS reports* **8**:45-53.
3. **Regoes RR, Bonhoeffer S.** 2005. The HIV coreceptor switch: a population dynamical perspective. *Trends in microbiology* **13**:269-277.
4. **Dean M, Carrington M, Winkler C, Huttley GA, Smith MW, Allikmets R, Goedert JJ, Buchbinder SP, Vittinghoff E, Gomperts E, Donfield S, Vlahov D, Kaslow R, Saah A, Rinaldo C, Detels R, O'Brien SJ.** 1996. 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* **273**:1856-1862.
5. **Dorr P, Westby M, Dobbs S, Griffin P, Irvine B, Macartney M, Mori J, Rickett G, Smith-Burchnell C, Napier C, Webster R, Armour D, Price D, Stammen B, Wood A, Perros M.** 2005. 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* **49**:4721-4732.



356 6. **Archer J, Rambaut A, Taillon BE, Harrigan PR, Lewis M, Robertson DL.**  
357 2010. The evolutionary analysis of emerging low frequency HIV-1 CXCR4  
358 using variants through time--an ultra-deep approach. PLoS computational  
359 biology **6**:e1001022.

360 7. **Shankarappa R, Margolick JB, Gange SJ, Rodrigo AG, Upchurch D,**  
361 **Farzadegan H, Gupta P, Rinaldo CR, Learn GH, He X, Huang XL,**  
362 **Mullins JI.** 1999. Consistent viral evolutionary changes associated with  
363 the progression of human immunodeficiency virus type 1 infection.  
364 Journal of virology **73**:10489-10502.

365 8. **Meehan CJ, Hedge JA, Robertson DL, McCormack GP, Travers SA.** 2010.  
366 Emergence, dominance, and possible decline of CXCR4 chemokine  
367 receptor usage during the course of HIV infection. Journal of medical  
368 virology **82**:2004-2012.

369 9. **Esbjornsson J, Mansson F, Martinez-Arias W, Vincic E, Biague AJ, da**  
370 **Silva ZJ, Fenyo EM, Norrgren H, Medstrand P.** 2010. Frequent CXCR4  
371 tropism of HIV-1 subtype A and CRF02\_AG during late-stage disease--  
372 indication of an evolving epidemic in West Africa. Retrovirology **7**:23.

373 10. **Hwang SS, Boyle TJ, Lierly HK, Cullen BR.** 1991. Identification of the  
374 envelope V3 loop as the primary determinant of cell tropism in HIV-1.  
375 Science **253**:71-74.

376 11. **Moore JP, Kitchen SG, Pugach P, Zack JA.** 2004. The CCR5 and CXCR4  
377 coreceptors--central to understanding the transmission and pathogenesis  
378 of human immunodeficiency virus type 1 infection. AIDS research and  
379 human retroviruses **20**:111-126.

- 380 12. **Bartlett AD, Macartney MJ, Conibear TC, Feyertag F, Smith CJ,**  
381 **Johnson MA, Hyams C, Garcia-Diaz A, McCormick AL, Booth C,**  
382 **Robertson DL, Webster DP.** 2014. The utility of different bioinformatics  
383 algorithms for genotypic HIV-1 tropism testing in a large clinical cohort  
384 with multiple subtypes. *Aids*.  
385 13. **Pollakis G, Kang S, Kliphuis A, Chalaby MI, Goudsmit J, Paxton WA.**  
386 2001. N-linked glycosylation of the HIV type-1 gp120 envelope  
387 glycoprotein as a major determinant of CCR5 and CXCR4 coreceptor  
388 utilization. *The Journal of biological chemistry* **276**:13433-13441.  
389 14. **Li Y, Rey-Cuille MA, Hu SL.** 2001. N-linked glycosylation in the V3 region  
390 of HIV type 1 surface antigen modulates coreceptor usage in viral  
391 infection. *AIDS research and human retroviruses* **17**:1473-1479.  
392 15. **Polzer S, Dittmar MT, Schmitz H, Schreiber M.** 2002. The N-linked  
393 glycan g15 within the V3 loop of the HIV-1 external glycoprotein gp120  
394 affects coreceptor usage, cellular tropism, and neutralization. *Virology*  
395 **304**:70-80.  
396 16. **Mild M, Kvist A, Esbjornsson J, Karlsson I, Fenyo EM, Medstrand P.**  
397 2010. Differences in molecular evolution between switch (R5 to  
398 R5X4/X4-tropic) and non-switch (R5-tropic only) HIV-1 populations  
399 during infection. *Infection, genetics and evolution : journal of molecular*  
400 *epidemiology and evolutionary genetics in infectious diseases* **10**:356-  
401 364.  
402 17. **Edo-Matas D, Rachinger A, Setiawan LC, Boeser-Nunnink BD, van 't**  
403 **Wout AB, Lemey P, Schuitemaker H.** 2012. The evolution of human  
404 immunodeficiency virus type-1 (HIV-1) envelope molecular properties

405 and coreceptor use at all stages of infection in an HIV-1 donor-recipient  
406 pair. *Virology* **422**:70-80.

407 18. **Tsuchiya K, Ode H, Hayashida T, Kakizawa J, Sato H, Oka S, Gatanaga**  
408 **H.** 2013. Arginine insertion and loss of N-linked glycosylation site in HIV-  
409 1 envelope V3 region confer CXCR4-tropism. *Scientific reports* **3**:2389.

410 19. **Scanlan CN, Offer J, Zitzmann N, Dwek RA.** 2007. Exploiting the  
411 defensive sugars of HIV-1 for drug and vaccine design. *Nature* **446**:1038-  
412 1045.

413 20. **Culyba EK, Price JL, Hanson SR, Dhar A, Wong CH, Gruebele M,**  
414 **Powers ET, Kelly JW.** 2011. Protein native-state stabilization by placing  
415 aromatic side chains in N-glycosylated reverse turns. *Science* **331**:571-  
416 575.

417 21. **Land A, Braakman I.** 2001. Folding of the human immunodeficiency  
418 virus type 1 envelope glycoprotein in the endoplasmic reticulum.  
419 *Biochimie* **83**:783-790.

420 22. **Wood NT, Fadda E, Davis R, Grant OC, Martin JC, Woods RJ, Travers**  
421 **SA.** 2013. The influence of N-linked glycans on the molecular dynamics of  
422 the HIV-1 gp120 V3 loop. *PloS one* **8**:e80301.

423 23. **Dyson HJ, Wright PE.** 2005. Intrinsically unstructured proteins and their  
424 functions. *Nature reviews. Molecular cell biology* **6**:197-208.

425 24. **Habchi J, Tompa P, Longhi S, Uversky VN.** 2014. Introducing Protein  
426 Intrinsic Disorder. *Chemical reviews*.

427 25. **Uversky VN, Oldfield CJ, Dunker AK.** 2008. Intrinsically disordered  
428 proteins in human diseases: introducing the D2 concept. *Annual review of*  
429 *biophysics* **37**:215-246.

- 430 26. **Ferreon AC, Ferreon JC, Wright PE, Deniz AA.** 2013. Modulation of  
431 allostery by protein intrinsic disorder. *Nature* **498**:390-394.
- 432 27. **Uversky VN.** 2013. A decade and a half of protein intrinsic disorder:  
433 Biology still waits for physics. *Protein science : a publication of the*  
434 *Protein Society.*
- 435 28. **Tompa P.** 2012. Intrinsically disordered proteins: a 10-year recap.  
436 *Trends in biochemical sciences* **37**:509-516.
- 437 29. **Starcich BR, Hahn BH, Shaw GM, McNeely PD, Modrow S, Wolf H,**  
438 **Parks ES, Parks WP, Josephs SF, Gallo RC, et al.** 1986. Identification and  
439 characterization of conserved and variable regions in the envelope gene  
440 of HTLV-III/LAV, the retrovirus of AIDS. *Cell* **45**:637-648.
- 441 30. **Vavouri T, Semple JI, Garcia-Verdugo R, Lehner B.** 2009. Intrinsic  
442 protein disorder and interaction promiscuity are widely associated with  
443 dosage sensitivity. *Cell* **138**:198-208.
- 444 31. **Levy ED, De S, Teichmann SA.** 2012. Cellular crowding imposes global  
445 constraints on the chemistry and evolution of proteomes. *Proceedings of*  
446 *the National Academy of Sciences of the United States of America*  
447 **109**:20461-20466.
- 448 32. **Bailey JR, Lassen KG, Yang HC, Quinn TC, Ray SC, Blankson JN,**  
449 **Siliciano RF.** 2006. Neutralizing antibodies do not mediate suppression  
450 of human immunodeficiency virus type 1 in elite suppressors or selection  
451 of plasma virus variants in patients on highly active antiretroviral therapy.  
452 *Journal of virology* **80**:4758-4770.
- 453 33. **Mild M, Gray RR, Kvist A, Lemey P, Goodenow MM, Fenyo EM, Albert J,**  
454 **Salemi M, Esbjornsson J, Medstrand P.** 2013. High inpatient HIV-1

455 evolutionary rate is associated with CCR5-to-CXCR4 coreceptor switch.

456 Infection, genetics and evolution : journal of molecular epidemiology and

457 evolutionary genetics in infectious diseases.

458 34. **Lengauer T, Sander O, Sierra S, Thielen A, Kaiser R.** 2007.

459 Bioinformatics prediction of HIV coreceptor usage. Nature biotechnology

460 **25**:1407-1410.

461 35. **Hegy H, Tompa P.** 2008. Intrinsically disordered proteins display no

462 preference for chaperone binding in vivo. PLoS computational biology

463 **4**:e1000017.

464 36. **Dosztanyi Z, Csizmok V, Tompa P, Simon I.** 2005. The pairwise energy

465 content estimated from amino acid composition discriminates between

466 folded and intrinsically unstructured proteins. Journal of molecular

467 biology **347**:827-839.

468 37. **Ortiz JF, MacDonald ML, Masterson P, Uversky VN, Siltberg-Liberles J.**

469 2013. Rapid evolutionary dynamics of structural disorder as a potential

470 driving force for biological divergence in flaviviruses. Genome biology and

471 evolution **5**:504-513.

472 38. **Konietschke F, Pauly M.** 2012. A studentized permutation test for the

473 nonparametric Behrens-Fisher problem in paired data. Electron J Stat

474 **6**:1358-1372.

475 39. **Kwong PD, Wyatt R, Robinson J, Sweet RW, Sodroski J, Hendrickson**

476 **WA.** 1998. Structure of an HIV gp120 envelope glycoprotein in complex

477 with the CD4 receptor and a neutralizing human antibody. Nature

478 **393**:648-659.

479 40. **Weissenhorn W, Dessen A, Harrison SC, Skehel JJ, Wiley DC.** 1997.  
 480 Atomic structure of the ectodomain from HIV-1 gp41. *Nature* **387**:426-  
 481 430.

482 41. **Tasca S, Ho SH, Cheng-Mayer C.** 2008. R5X4 viruses are evolutionary,  
 483 functional, and antigenic intermediates in the pathway of a simian-human  
 484 immunodeficiency virus coreceptor switch. *Journal of virology* **82**:7089-  
 485 7099.

486 42. **Mosca R, Pache RA, Aloy P.** 2012. The role of structural disorder in the  
 487 rewiring of protein interactions through evolution. *Molecular & cellular*  
 488 *proteomics : MCP* **11**:M111 014969.

489 43. **Kimura M.** 1983. The neutral theory of molecular evolution. Cambridge  
 490 University Press, Cambridge Cambridgeshire ; New York.

491 44. **Ohta T.** 1973. Slightly deleterious mutant substitutions in evolution.  
 492 *Nature* **246**:96-98.

493 45. **Berger EA, Murphy PM, Farber JM.** 1999. Chemokine receptors as HIV-1  
 494 coreceptors: roles in viral entry, tropism, and disease. *Annual review of*  
 495 *immunology* **17**:657-700.

496 46. **Tamamis P, Floudas CA.** 2013. Molecular recognition of CXCR4 by a dual  
 497 tropic HIV-1 gp120 V3 loop. *Biophysical journal* **105**:1502-1514.

498 47. **Jensen MA, Li FS, van 't Wout AB, Nickle DC, Shriner D, He HX,**  
 499 **McLaughlin S, Shankarappa R, Margolick JB, Mullins JL.** 2003.  
 500 Improved coreceptor usage prediction and genotypic monitoring of R5-  
 501 to-X4 transition by motif analysis of human immunodeficiency virus type  
 502 1 env V3 loop sequences. *Journal of virology* **77**:13376-13388.

503 48. **Kalinina OV, Pfeifer N, Lengauer T.** 2013. Modelling binding between  
 504 CCR5 and CXCR4 receptors and their ligands suggests the surface  
 505 electrostatic potential of the co-receptor to be a key player in the HIV-1  
 506 tropism. *Retrovirology* **10**:130.

507

508

## Figure legends

Fig 1. Plot of predicted structural disorder tendency of the consensus envelope protein. Disorder tendency for each amino acid is predicted by IUPred (0 represents complete order; 1 represents complete disorder; scores over 0.4 represent intrinsically disordered). Disorder tendency scores are plotted for consensus (A) R5, (B) R5X4 and (C) X4 envelope sequence, respectively. (D) 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. Gp120 and gp41 protein domains are numbered and color-coded for visualisation.

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 ( $p= 7.49 \times 10^{-6}$ ).

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  $p > 0.05$ ).

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  $p > 0.05$ ).

Disorder Tendency vs. Residue Grouped by Tropism







