Title

On the Physicochemical and Structural Modifications Associated with HIV-1 Subtype B Tropism Transition

Journal

AIDS Research and Human Retroviruses, 32(8)

ISSN

0889-2229

Authors

Lamers, Susanna L
Fogel, Gary B
Liu, Enoch S
et al.

Publication Date

2016-08-01

DOI

10.1089/aid.2015.0373

Peer reviewed

AIDS RESEARCH AND HUMAN RETROVIRUSES
Volume 32, Number 8, 2016

DOI: 10.1089/aid.2015.0373

On the Physicochemical and Structural Modifications Associated with HIV-1 Subtype B Tropism Transition

Susanna L. Lamers,[1,*] Gary B. Fogel,[2,*] Enoch S. Liu,[2] Marco Salemi,[3] and Michael S. McGrath[4]

Abstract

HIV-1 enters immune cells via binding the viral envelope to a host cell CD4 receptor, and then a secondary co-receptor, usually CCR5 (R5) or CXCR4 (X4), and some HIV can utilize both co-receptors (R5X4). Although a small set of amino-acid properties such as charge and sequence length applied to HIV-1 envelope V3 loop sequence data can be used to predict co-receptor usage, we sought to expand the fundamental understanding of the physiochemical basis of tropism by analyzing many, perhaps less obvious, amino-acid properties over a diverse array of HIV sequences. We examined 74 amino-acid physicochemical scales over 1,559 V3 loop sequences with biologically tested tropisms downloaded from the Los Alamos HIV sequence database. Linear regressions were then calculated for each feature relative to three tropism transitions (R5 → X4; R5 → R5X4; R5X4 → X4). Independent correlations were rank ordered to determine informative features. A structural analysis of the V3 loop was performed to better interpret these findings relative to HIV tropism states. Similar structural changes are required for R5 and R5X4 to transition to X4, thus suggesting that R5 and R5X4 types are more similar than either phenotype is to X4. Overall, the analysis suggests a continuum of viral tropism that is only partially related to charge; in fact, the analysis suggests that charge modification may be primarily attributed to decreased R5 usage, and further structural changes, particularly those associated with β-sheet structure, are likely required for full X4 usage.

Introduction

Human immunodeficiency virus type 1 (HIV-1) primarily infects human $CD4^+$ T-cells and macrophages. In order for HIV-1 to enter either of these cells, the viral envelope (gp120) must first bind a CD4 receptor on the cell surface and then a secondary co-receptor, usually CCR5 (R5) or CXCR4 (X4).[1] Although there are two primary cellular co-receptors, three viral phenotypes are typically described in the literature: R5, X4, and less-efficient intermediate forms of HIV-1 that retain some ability to bind either co-receptor, called dual-tropic HIV-1 (R5X4). The transition in an individual from R5 to X4 is assumed to be unidirectional.[2–4]

Both R5 and X4 co-receptors are present on T-cells and macrophages[1,5]; however, receptor density may vary among different immune cells.[6] Transmitted viruses, as well as viral populations during early infection, preferentially bind R5.[5] Viruses that bind X4 may arise, usually during late infection, and are associated with increased diversity and a high evolutionary rate.[7] Because blocking the interaction of gp120 with host co-receptors is a strategy used in anti-HIV therapy, a better understanding of tropism states may assist in the development of therapies to block viral transmission or as a treatment for those infected.

Gp120 contains five hypervariable domains (V1–V5) that possess a fluctuating glycan shield and variable epitopes near binding domains.[8,9] Of the variable domains, the V3 loop has the strongest known link with HIV-1 tropism, and, therefore, V3 is frequently the focus of studies that aim at defining tropism states. Computational approaches have been developed to classify viruses as R5 or X4 tropic based on a small number of genotypic features of the envelope V3 domain, including charge, number of glycosylation motifs and domain length,[10–14] or the ''11/25 rule'' that associates positively charged amino acids at two positions (11 and 25) in the V3 loop.[15] These methods are highly useful in classifying

[1]Bioinfoexperts, LLC, Thibodaux, Louisiana.
[2]Natural Selection, Inc., San Diego, California.
[3]Department of Pathology and Laboratory Medicine, University of Florida, Gainesville, Florida.
[4]Department of Laboratory Medicine, Pathology, and Medicine, and the AIDS and Cancer Specimen Resource, University of California, San Francisco, California.
*These authors contributed equally to this work.

HIV-1 subtypes B and C as either R5 or X4 tropic so that the appropriate therapy can be prescribed. However, it remains unclear whether R5 viruses always pass through an R5X4 stage before the development of X4 viruses or whether other less obvious structural modifications between tropism transitions exist that could benefit anti-HIV drug design.

Our goal in this study was to more completely characterize the molecular and structural basis of tropism and its evolution by exploring a diverse database of sequences (those biologically characterized at the HIV database at Los Alamos) and a larger space of possible features, including 74 amino-acid physicochemical scales.

Materials and Methods

Sequences and alignment

Publicly available V3 loop sequences (relative to HXB2 positions 7,110–7,217) for HIV-1 subtype B were downloaded from the HIV sequence database at the Los Alamos National Laboratory (www.hiv.lanl.gov/content/index) and translated into amino-acid sequences. The search criteria for each tropism were limited to ''only CCR5,'' ''only CXCR4,'' or ''only R5X4.'' The tropism fields at the Los Alamos HIV database are annotated based on only biological experiments and are not presumed using inferred genetic sequences. Identical sequences were removed from each category. The amino-acid sequences were aligned using the ClustalW algorithm that was implemented within the MEGA5 sequence analysis package[16] and then manually edited to correct for any obvious alignment errors. Diversity in each data set was calculated using the Poisson substitution model and 1,000 bootstrap replicates in MEGA5.

Feature generation and correlations

Seventy-four amino-acid scales, or ''features,'' were identified from the available literature and resources such as ProtScale and ProtParam (www.expasy.org)[17] (Table 1). These features were grouped into six major classes: (1) amino acid size, shape, or structure ($n=24$); (2) polarity ($n=6$); (3) composition ($n=5$); (4) hydrophobicity ($n=26$); (5) local features, such as charges and glycosylation motifs occurring at specific regions of the aligned data ($n=4$); and (6) other miscellaneous features such as those associated with HPLC and pKa ($n=8$).

All features were calculated for the V3 loop taken as a whole, and for specific regions of the V3 loop at the amino- (*N*) and carboxy- (*C*) ends (termed ''regional V3-loop features''). So-called ''*N*'' features corresponded to alignment positions 9 through 14 in the *N*-terminus of the V3 loop. So-called ''$C_{(1)}$'' features corresponded to alignment positions 22 through 28 in the *C*-terminus. ''$C_{(2)}$'' features corresponded to alignment positions 31 through 37 in the *C*-terminus. These two *C*-terminus labels identify regions of the V3 loop alignment as strand and helix, and they were added simply to explore which features in specific regions had more or less influence on tropism transitions. This led to the calculation of 281 total features over all regions.

Linear regressions were then calculated for each feature relative to each of three tropism transitions (R5 → X4; R5 → R5X4; R5X4 → X4). Independent correlations for each tropism transition were rank ordered, with features $R^2 \geq 0.1$ for at least one of the three tropism decisions considered potentially informative toward an understanding of the differences between these tropism types. The features with the highest correlation were used to build an understanding of the transitions among R5, R5X4, and X4 HIV V3 domains and viewed with structural models as defined next.

V3 loop structure generation

Three sequences were chosen at random from each of the three tropism phenotypes for further computational modeling and structural analysis. V3 loop structures were generated for each of these nine sequences using the I-Tasser server.[18–20] I-Tasser generates structures using an iterative approach of (1) retrieving template proteins with similar folds from the Protein Data Bank (PDB) (www.rcsb.org), (2) reassembly of matching templates into full-length models and threading of unaligned regions built by *ab initio* modeling, and (3) fragment assembly simulation guided by PDB templates and TM-align. Along with a three-dimensional structure, the server also provides information regarding secondary structure and solvent accessibility.

Energy minimization and equilibration of V3 loop structures

One of the limitations of protein threading is that it relies on existing published structures in the PDB; therefore, protein threading was followed by an energy minimization of each model. Molecular dynamics are commonly used to generate improved protein structures, which is accomplished by allowing protein atoms to adjust under defined conditions (i.e., temperature and pressure) using Newtonian equations. These simulations were carried out using the Nanoscale Molecular Dynamics program (NAMD).[21]

The models generated by I-Tasser were solvated in a water box of 10 Å in each direction. The system was neutralized with the addition of NaCl. Simulations were carried out using periodic boundary conditions at 310 K to reduce surface interactions of water molecules and to create a more accurate *in vivo* environment. Particle Mesh Ewald (PME) electrostatics were employed using a grid spacing of 1.0 Å to reproduce the charge distribution of the system. Langevin dynamics were used to control kinetic energies. A time step of two femtoseconds (fs) required the use of rigid bonds. Structures were minimized for 1,000 time steps and allowed to equilibrate for 100,000 time steps. Three replicate NAMD runs were performed for each structural model. Molecular simulations were viewed in the program visual molecular dynamics.[22]

Results

Unequal abundance and diversity in R5, R5X4, and X4 sequences

The Los Alamos HIV sequence database queries for biologically phenotyped V3 loop HIV resulted in 3452 R5 sequences, 545 R5X4 sequences, and 197 X4 sequences. After removal of all identical sequences, the final sequence population contained 1223 R5 sequences, 241 R5X4 sequences, and 95 X4 sequences for a total of 1,559 unique sequences. Some nonidentical multiple clones from the same patient were apparent and included in our analysis to capture all positional amino-acid information for each position. Although all sequences were biologically tested for tropism, different

TABLE 1. PHYSICOCHEMICAL AND STRUCTURAL FEATURES USED FOR THE STUDY

Class	*Features*		
A. Size, shape, structure	Molecular weight[46]	Bulkiness[47]	Antiparallel beta strand[48]
	Beta turn[49]	Beta-sheet Levitt[50]	Alpha Chou and Fasman[51]
	Coil[49]	Beta strand[48]	Parallel beta[48]
	Surface area[52]	Volume[52]	Surface exposure[53]
	Average flexibility[54]	Mol. fraction of buried Res.[55]	Beta Chou and Fasman[51]
	% Accessible residues[55]	2D Propensity[53]	Transmembrane[56]
	Avg. area buried[57]	Membership class[53]	Mass membership class[53]
	Recognition factors[58]	Alpha helix Levitt[50]	Alpha helix[49]
B. Polarity	Polarity[54]	Charge[a]	Charge polarity[47]
	Charge scale[54]	Grantham[54]	
C. Composition	Amino acid composition[59]	Length of V3[15]	Amino acid composition swiss prot[46]
	Relative mutability[60]	Total sequence glycosylation[61]	
D. Hydrophobicity	Sweet et al.[62]	Kyte and Doolittle[63]	Hydrophobicity[53]
	Abraham and Leo[64]	Bull and Breese[65]	Guy[66]
	Roseman[67]	Wolfen et al.[68]	Wilson et al.[69]
	Eisenberg et al.[70]	Hopp and Woods[71]	Manvalan et al.[72]
	Fauchere et al.[73]	Janin[55]	Rao and Argos[74]
	Tanford[75]	Welling et al.[76]	Chothia[77]
	Cowan and Whittaker[78]	Parker[79]	Browne[80]
	Meek[81]	Aboderin[82]	Rose et al.[57]
	Black and Mould[83]	Miyazawa et al.[84]	
E. Local features	Glycosylation at positions 6–8[61]	Glycosylation at positions 5,7,9[61]	Charge at position 12[15]
	Charge at position 30[15]		
F. HPLC and other	Retention at pH 2.1[81]	HPLC/TFA[80]	Refractivity[85]
	HP scale[53]	pKa alpha carboxylate[86]	Exchange[53]
	pKa amine[86]	pI at 25°C[86]	

[a]"Charge" feature = [+1] for K and R; [−1] for D and E.

experimental methods in different laboratories could have marginally impacted tropism determination. The significant overabundance of R5 sequences in the available public databases suggests that R5 sequences greatly predominate over X4 variants in nature. The sequence alignment spanned 40 amino acids, including gapped positions, and allowed for the inclusion of all sequences, even those that were unusually long or those with varied sequence and that were derived from the same individual. Because of this, charged positions 11 and 25, previously associated with tropism,[15] corresponded to positions 12 and 30 in our alignment, respectively. A representative alignment of three sequences chosen at random from each known tropism is provided in Figure 1A. Figure 1B provides additional structural models for the V3 loop showing positional charge variation and slight structural differences commonly observed when threading sequences using major PDB models for different tropisms.

In Figure 2, we show an almost stepwise progression in the diversity of sequences for each of these tropisms as described by their internal pairwise distances. The mean pairwise amino-acid distance for R5 sequences was 18.2%, which was nearly doubled for both R5X4 at 32.9% and X4 at 38.8%. These diversity measurements are not meant to guide tropism prediction; instead, they are merely meant to elucidate the stepwise increase in the diversity trend of R5 to R5X4 to X4 sequence populations. Although the inclusion of some nonidentical sequences derived from the same individual in the study could bias the R5 sequence population toward reduced diversity, this effect would have been minimal when considering that the overall R5 sequence population was almost 17 times larger than the X4 sequence population. Furthermore, it is generally known that X4 variants maintain more diverse amino-acid substitutions.

A different set of features was associated with each tropism transition and suggested a progression or continuum of viral tropism that is only partially related to charge. In Table 2, we provide a list of all features examined and their correlation to tropism relationships. A correlation of $R^2 \geq 0.1$ was chosen arbitrarily as a lower cutoff for useful tropism understanding. Only 84 features had an $R^2 \geq 0.1$ for at least one of the three tropism comparisons. The remaining 197 features were considered to have little utility for tropism insight and were eliminated from further consideration.

Although amino-acid charges are commonly used features to discriminate R5 from X4 sequences in public computational algorithms that predict viral tropism, in our transitional analysis, a set of 12 features identified R5 → R5X4 at $R^2 \geq 0.1$, with the top four features associated with charge. However, 12 other amino-acid characteristics surpassed charge as a feature in distinguishing the R5X4 → X4 transition and 25 noncharge-associated features (shape, size, or structure) with an $R^2 > 0.2$ were determined to be important in the R5 → X4 transition. With this in mind, it may be that computational algorithms used to determine tropism are biased toward

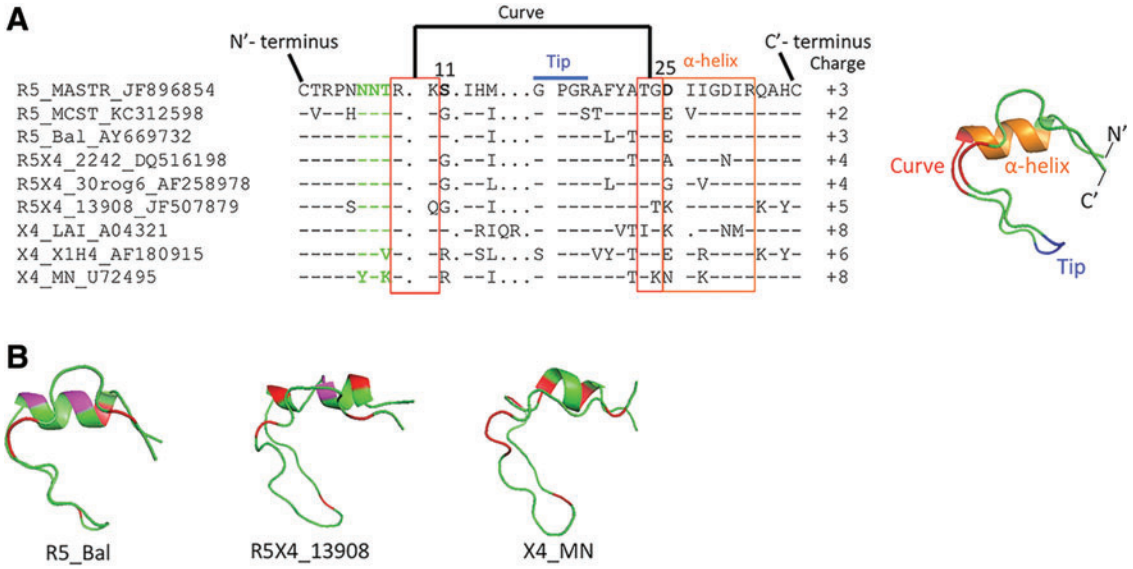


FIG. 1. Representative HIV-1 V3 loop phenotypes. **(A)** A sequence alignment containing three V3 loop sequences for each phenotype is shown. The sequence name consists of the biologically tested viral tropism, the sequence name, and the GenBank accession number. A *period* in the aligned sequences represents a gap inserted to maximize the alignment of all 1,559 unique sequences in the study. A *dash* represents identity to the top sequence in the displayed alignment (JF896874). An amino acid is listed when the amino acid varies from JF896874. A gap is inserted every 10 amino acids. A three-amino-acid region where a glycosylation motif (NX[T or S], where X = any amino acid except P) commonly occurs is colored *green*. Positions associated with the ''11/25 charge rule'' are indicated. Other distinguishing structural features are also indicated in the alignment and are mapped in the structure on the *right*. The overall charge of the V3 loop sequence for each isolate is shown at the end of the sequence. **(B)** To highlight charge variation between structures of different tropisms, three structural models of the V3 loop are shown that are representative of an R5 (Bal), R5X4 (13908), and X4 (MN) structure, respectively, with positively charged residues in *red* and negatively charged residues in *purple*. Color images available online at www.liebertpub.com/aid

assigning actual intermediate R5X4 V3 loop sequences as X4 variants, or that additional features added to existing algorithms could improve their accuracy. Also, the analysis suggests that key features associated with true X4 viruses have not been adequately teased out of the structures.

In the N-terminus of the V3 loop, three features were identified for the R5 → R5X4 transition with an $R^2 \geq 0.1$: 2D Membership Class *N* (structural), Grantham (polarity), and Beta Chou and Fasman (structural). However, in the transition from R5X4 to X4 and from R5 to X4, 24 structural features associated within the $C_{(1)}$ region were calculated to have an $R^2 \geq 0.1$. Many of these features were associated with beta-sheet structure. Also noteworthy in the R5X4 → X4 and R5 → X4 transition were the 2 pKa-associated features (one associated with the amine and one with the carboxylate) with higher R^2, which could mean that the transition to X4 from either R5 or R5X4 is zwitterion like, meaning that the overall effect can be a neutral molecule with an increase in both positive and negative electrical charges.

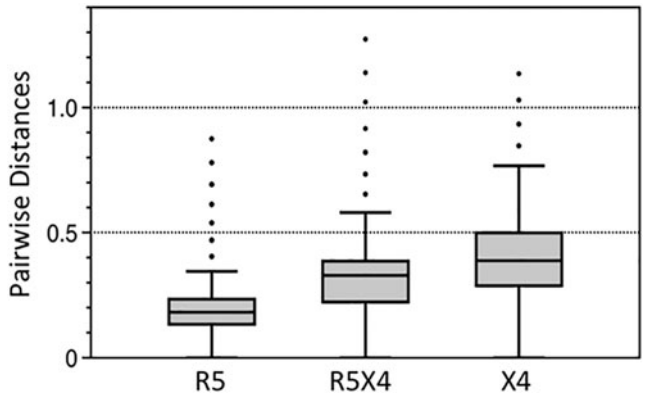


FIG. 2. Pairwise distances calculated for all sequences in each tropism category. Category is shown on the *x-axis*, and pairwise distance is indicated on the *y-axis*. Distances are plotted in a box and whisker format with the median (*horizontal line*), upper and lower quartiles, and range. *Dots* represent outliers.

No structural features from the $C_{(2)}$ region were considered important for any tropism transition. These results suggest that features that distinguish R5 from X4 viruses are found either throughout the V3 loop or within the *N*-terminus, whereas the transition from R5X4 to X4 virus results from changes in the $C_{(1)}$ domain of the V3 loop. Four polarity features (Charge Scale $C_{(1)}$, Charge Polarity $C_{(1)}$, Grantham $C_{(1)}$, and Polarity $C_{(1)}$) were also central in the transition from R5X4 to X4.

Interestingly, glycosylation was not identified for either the R5 → R5X4 or the R5X4 → X4 transition, but it had a higher R^2 when transitioning from R5 to X4, suggesting that this feature may be a final adjustment in the transition from R5 to X4 co-receptor usage. Alternatively, it may be that an X4 glycosylation motif arises unsystematically due to the required shape/size adjustment for the final X4 usage; this could explain why many X4 variants do not, in fact, possess an additional glycosylation motif. Also interesting is that sequence length was identified as only a minimally useful indicator during the

Table 2. Linear Regression Analysis for Tropism Transitions

R5 vs. R5X4 features	R^2	*R5X4 vs. X4 features*	R^2	*R5 vs. X4 features*	R^2
Charge position 12[e]	0.227	Beta sheet $C_{(1)}$[a]	0.68	Beta sheet $C_{(1)}$[a]	0.585
Charge[b]	0.208	pKa—amine $C_{(1)}$[f]	0.653	pKa alpha carboxylate $C_{(1)}$[f]	0.453
Charge position 30[e]	0.189	pKa alpha carboxylate $C_{(1)}$[f]	0.649	Beta Sheet Levitt $C_{(1)}$[a]	0.449
Charge scale[b]	0.172	Rose $C_{(1)}$[d]	0.592	pKa—amine $C_{(1)}$[f]	0.425
Janin[d]	0.136	Surface Area $C_{(1)}$[a]	0.592	Charge position 12[e]	0.41
2D propensity membership class N[a]	0.135	Beta Levitt $C_{(1)}$[a]	0.589	Beta Chou and Fasman $C_{(1)}$[a]	0.401
Volume[a]	0.126	Manavalan $C_{(1)}$[d]	0.581	Rose $C_{(1)}$[d]	0.373
Eisenberg[d]	0.125	Polarity $C_{(1)}$[b]	0.576	Surface area $C_{(1)}$[a]	0.373
Grantham N[b]	0.114	Beta Chou and Fasman $C_{(1)}$[a]	0.574	Manavalan $C_{(1)}$[d]	0.371
Refractivity[f]	0.113	Average Flex $C_{(1)}$[a]	0.569	Polarity $C_{(1)}$[b]	0.358
Beta sheet $C_{(1)}$[a]	0.113	Coil $C_{(1)}$[a]	0.541	Coil $C_{(1)}$[a]	0.331
Beta Chou and FasmanN[a]	0.104	Recognition factors $C_{(1)}$[a]	0.516	Average Flex $C_{(1)}$[a]	0.331
Beta strand $C_{(1)}$[a]	0.098	Charge scale $C_{(1)}$[b]	0.457	Charge[b]	0.286
Eisenberg $C_{(2)}$[d]	0.089	2D propensity membership class $C_{(1)}$[a]	0.449	Exchange $C_{(1)}$[f]	0.277
Alpha helix N[a]	0.084	pI at 25C $C_{(1)}$[f]	0.445	AA Comp $C_{(1)}$[c]	0.264
Average area buried[a]	0.082	Hydrophobicity membership class $C_{(1)}$[d]	0.426	Janin[d]	0.254
Charge polarity[b]	0.082	Charge polarity $C_{(1)}$[b]	0.415	Charge polarity $C_{(1)}$[b]	0.252
Beta sheet $C_{(2)}$[a]	0.081	Surface exposure membership class $C_{(1)}$[a]	0.383	Recognition factors $C_{(1)}$[a]	0.245
Tanford[d]	0.08	Miyazawa $C_{(1)}$[d]	0.377	Charge scale[b]	0.236
Wolfen[d]	0.079	Exchange $C_{(1)}$[f]	0.367	Rao and Argos $C_{(1)}$[d]	0.233
Relative mutability[c]	0.078	Rao and Argos $C_{(1)}$[d]	0.362	Hydrophobicity Membership Class $C_{(1)}$[d]	0.231
Abraham.Leo[d]	0.076	AA Comp $C_{(1)}$[c]	0.358	Eisenberg[d]	0.228
Bulkiness[a]	0.076	Mol fraction of buried res $C_{(1)}$[a]	0.351	AA Comp Swiss Prot $C_{(1)}$[c]	0.227
Fauchere $C_{(2)}$[d]	0.072	% Accessible residues $C_{(1)}$[a]	0.344	Miyazawa $C_{(1)}$[d]	0.222
Chou and Fasman $C_{(1)}$[a]	0.072	Beta-sheet Levett $C_{(1)}$[a]	0.329	% Accessible residues $C_{(1)}$[a]	0.219
Rao Argos[d]	0.071	AA Comp Swiss Prot $C_{(1)}$[c]	0.328	2D propensity membership class $C_{(1)}$[a]	0.214
Beta strand[a]	0.067	Black and Mould $C_{(1)}$[d]	0.303	Surface exposure membership class $C_{(1)}$[a]	0.213
Molecular weight[a]	0.065	Chothia $C_{(1)}$[d]	0.303	Volume[a]	0.208
Polarity[b]	0.065	Aboderin $C_{(1)}$[d]	0.232	2D propensity membership class N[a]	0.206
Chothia[d]	0.061	MW $C_{(1)}$[a]	0.232	Black and Mould $C_{(1)}$	0.204
Antiparallel beta[a]	0.061	Alpha Levitt $C_{(1)}$[a]	0.202	Refractivity[f]	0.179
Abraham and Leo $C_{(2)}$[d]	0.061	Chou and Fasman $C_{(1)}$[a]	0.181	Tanford[d]	0.173
2D propensity membership class[a]	0.06	Avg. area buried $C_{(1)}$[a]	0.173	Aboderin $C_{(1)}$[d]	0.165
Guy[d]	0.058	Alpha Chou and Fasman $C_{(1)}$[a]	0.172	Mol fraction of buried res $C^{(1)a}$	0.164
Relative mutability $C_{(2)}$[c]	0.058	Length $C_{(1)}$[c]	0.167	Glycosylation NXT[e]	0.156
Antiparallel beta $C_{(1)}$[a]	0.056	Janin[d]	0.155	Chothia $C_{(1)}$[d]	0.153
Beta Sheet $C_{(1)}$[a]	0.053	Grantham $C_{(1)}$[b]	0.145	Alpha Levitt $C_{(1)}$[a]	0.15
Roseman[d]	0.052	Eisenberg[d]	0.139	Abraham.Leo[d]	0.149
Hopp and Woods[d]	0.052	Volume $C_{(1)}$[a]	0.135	pI at 25C $C_{(1)}$[f]	0.149
Charge scale NN[e]	0.045	Beta strand $C_{(1)}$[a]	0.134	Wolfen[d]	0.147
Browne N[d]	0.043	Charge[b]	0.132	Chothia[d]	0.145
Chothia $C_{(2)}$[d]	0.042	Tanford[a]	0.126	Avg. area buried N[a]	0.145
Alpha helix $C_{(1)}$[a]	0.041	Antiparallel beta $C_{(1)}$[a]	0.126	Beta Chou and FasmanN[a]	0.143
Beta-sheet Levett $C_{(1)}$[a]	0.041	Mass membership class $C_{(1)}$[a]	0.124	Charge scale $C_{(1)}$[b]	0.14
Avg. area buried N[a]	0.039	Refractivity[f]	0.12	Average area buried[a]	0.133
pI at 25C N[f]	0.039	Volume[a]	0.115	Molecular weight[a]	0.131
Mass membership class N[a]	0.038	Eisenberg $C^{(1)d}$	0.112	Refractivity N[f]	0.131
AA Comp $C_{(1)}$[c]	0.035	Welling $C_{(1)}$[d]	0.11	Alpha Chou and Fasman $C_{(1)}$[a]	0.128
Glycosylation NXT[e]	0.035	Relative mutability $C_{(1)}$[c]	0.108	Mass membership class N[a]	0.125
Alpha Levitt $C_{(1)}$[a]	0.034	Wolfen[d]	0.106	2D propensity membership class[a]	0.124
Polarity N[b]	0.033	Beta sheet $C_{(1)}$[a]	0.099	Guy[d]	0.122

(*continued*)

TABLE 2. (CONTINUED)

R5 vs. R5X4 features	R^2	*R5X4 vs. X4 features*	R^2	*R5 vs. X4 features*	R^2
Alpha Chou and Fasman $C_{(1)}$[a]	0.031	2D propensity membership class[a]	0.098	Volume N[a]	0.122
Volume $C_{(1)}$[a]	0.031	Chothia[d]	0.094	Browne N[d]	0.121
Rao and Argos $C_{(1)}$[a]	0.03	Guy[d]	0.094	Alpha helix N[a]	0.116
Refractivity N[f]	0.029	Abraham.Leo[d]	0.093	Beta-sheet Levett $C_{(1)}$[a]	0.116
AA Comp Swiss Prot $C_{(1)}$[c]	0.026	Alpha helix $C_{(1)}$	0.093	Bull and Breese $C_{(1)}$[d]	0.11
Volume N[a]	0.026	Charge[b]	0.09	pl at 25C N[f]	0.109
Avg. area buried $C_{(1)}$[a]	0.026	Molecular weight[a]	0.084	Charge scale N[e]	0.108
Charge polarity $C_{(1)}$[b]	0.023	Roseman[d]	0.08	Chothia $C_{(2)}$[d]	0.108
Bull and Breese $C_{(1)}$[d]	0.022	Average area buried[a]	0.076	Bulkiness[a]	0.106
Molecular weight N[a]	0.021	Hopp and Woods[d]	0.074	Charge polarity[b]	0.105
% accessible residues $C_{(1)}$[a]	0.018	Refractivity N[f]	0.065	Rao Argos[d]	0.104
Recognition factors $C_{(1)}$[a]	0.017	RaoArgos[d]	0.062	Polarity N[b]	0.102
pKa alpha carboxylate $C_{(1)}$[f]	0.016	Glycosylation NXT[e]	0.062	Molecular weight N[a]	0.102

Features over the entire V3 sequence with $R^2 \geq 0.1$ for at least one of the three tropism transitions are provided and associated with the feature class as follows: [a]amino acid size, shape, or structure; [b]polarity; [c]composition; [d]hydrophobicity; [e]regional features; [f]miscellaneous other features. R^2 values are provided for each feature for each tropism decision.

transition of R5X4 to X4 with an $R^2 \geq 0.167$. This observation is likely due to the fact that although X4 tropism may allow for increased V3 loop length, the majority of X4 viruses we analyzed were no longer than R5 or R5X4 HIV in the V3 loop region.

Overall, most molecular features that are important for the R5X4 → X4 transition were not found to be important for the R5 → R5X4 transition. Further, transitional features between R5 → R5X4 tended to also perform well for the transition R5 → X4. These interesting results suggest that there are fundamental biological differences during the transitions between tropisms. For example, while examining the top 10 features for each tropism decision, a pattern emerges (Fig. 3). During the R5 → R5X4 transition, changes occur that affect charge at specific positions in the V3 loop. These changes also affect overall structural characteristics such as volume, hydrophobicity, and refractivity. During the conversion of R5X4 → X4, changes occur that affect β-sheet development, surface area, flexibility, and pKa of the V3 loop. These features are quite important to the development of X4 viruses. Note that with the exception of only two features (charge at position 12 and average Flex $C_{(1)}$) the top 10 features for the R5X4 → X4 are the same as the top 10 features for distinguishing R5 → X4. This also indicates that what we currently refer to as R5X4 is probably functionally closer to R5 than X4 in its sequence and structure.

To summarize, these statistical results suggest a series of amino-acid alterations during the transition of viral tropism from R5 to X4, including: (1) charge modifications and structural *N* changes reduce the ability of the virus to utilize R5 receptors; (2) structural changes, especially in the $C_{(1)}$ region, allow for the adaptation of V3 loop to X4 receptor binding; and (3) increased glycosylation may be a final feature in structural adjustments, allowing the final transition

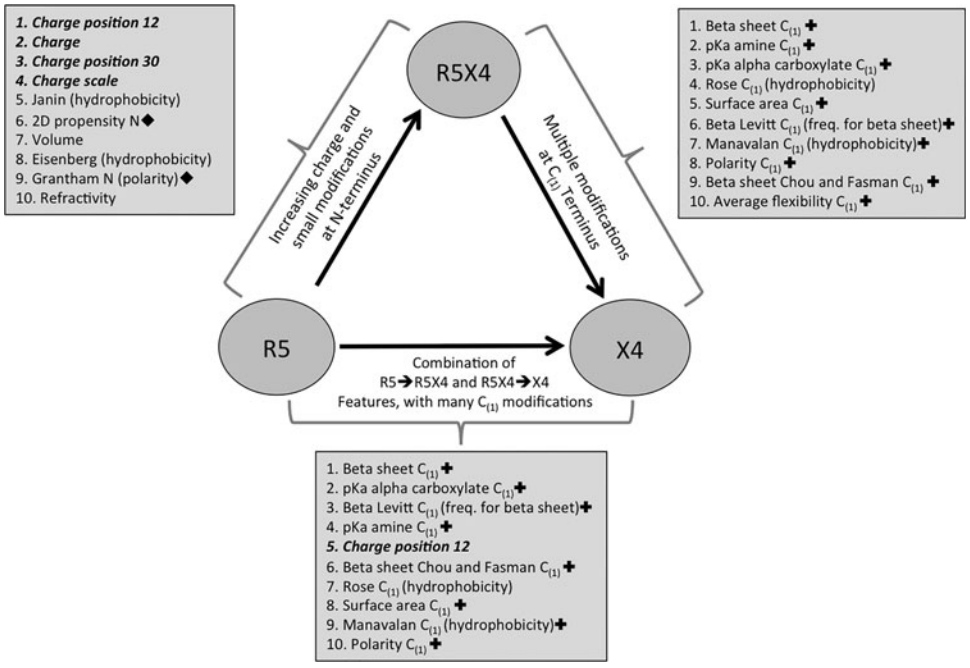


FIG. 3. Top 10 features for each tropism transition. The top 10 scoring features with an $R^2 \geq 0.1$ associated with transitions among tropism types are shown in *boxes*. A *diamond* indicates features associated with N-terminus transitions, and a *plus sign* indicates features associated with the $C_{(1)}$ terminus transitions. Charge-associated features are emphasized using *bold text*.

Table 3. HIV V3 Loop Threading Scores

Tropism	*Isolate*	*C-score*	*TM-score*	*RMSD(Å)*
R5	MASTR	–0.21	0.69 ± 0.12	2.4 ± 1.8
	MCST	–0.18	0.69 ± 0.12	2.3 ± 1.8
	HIVBal	–0.13	0.70 ± 0.12	2.2 ± 1.7
R5X4	2242	–0.18	0.69 ± 0.12	2.3 ± 1.8
	30 Rog	–0.23	0.68 ± 0.12	2.4 ± 1.8
	13908	–0.42	0.66 ± 0.13	2.8 ± 2.0
X4	HIVLai	–0.69	0.63 ± 0.14	3.3 ± 2.3
	X1H4	–0.74	0.62 ± 0.14	3.4 ± 2.3
	MN	–0.52	0.65 ± 0.13	2.9 ± 2.1

Confidence score (C-score) based on significance of the threading; between –5 and +2 is considered accurate. A measure of structural similarity in modeling (TM-score) of >0.5 indicates a model of correct topology. The root mean square deviation (RMSD) from native models used for threading in the PDB.
PDB, Protein Data Bank.

from R5X4 to X4. These findings could be tested on controlled data sets, such as those used to develop co-receptor algorithms, to confirm their utility.

Energy minimization and equilibration of V3 loop structures revealed a more organized secondary structure in R5 than in R5X4 or X4 structures. The most commonly used threading templates used by the I-Tasser server were PDB models 1ce4A,[23] 4ncoA,[24] and 3tygA,[25] all of which have an R5-like tropism, but still produced reasonable structures for further examination and resolution using molecular dynamics simulations. The generation of structures using this server produces a confidence score (C-score) based on the significance of the threading. The resulting top scoring models based on three confidence measurements (C-score, TM-score, and RMSD) are described in Table 3. All models produced were within the range of models of reasonable topology, with the R5 models exhibiting the lowest C-scores, which is not surprising considering that the native models in the PDB had an R5-like tropism.

Another result from protein modeling is secondary structure prediction at each position in the model and predicted solvent accessibility at each position (Figs. 4 and 5). The N-terminus of the structure and the ''tip'' of the loop are highly conserved in terms of secondary structure for all tropisms. It is in the α-helix domain, at the C-terminus, that a less ordered structure is particularly apparent (Fig. 4). Minimal differences are noted for solvent accessibility after modeling (Fig. 5) between R5 and R5X4 virus; however, X4 has some potential interesting changes from two other tropism classes.

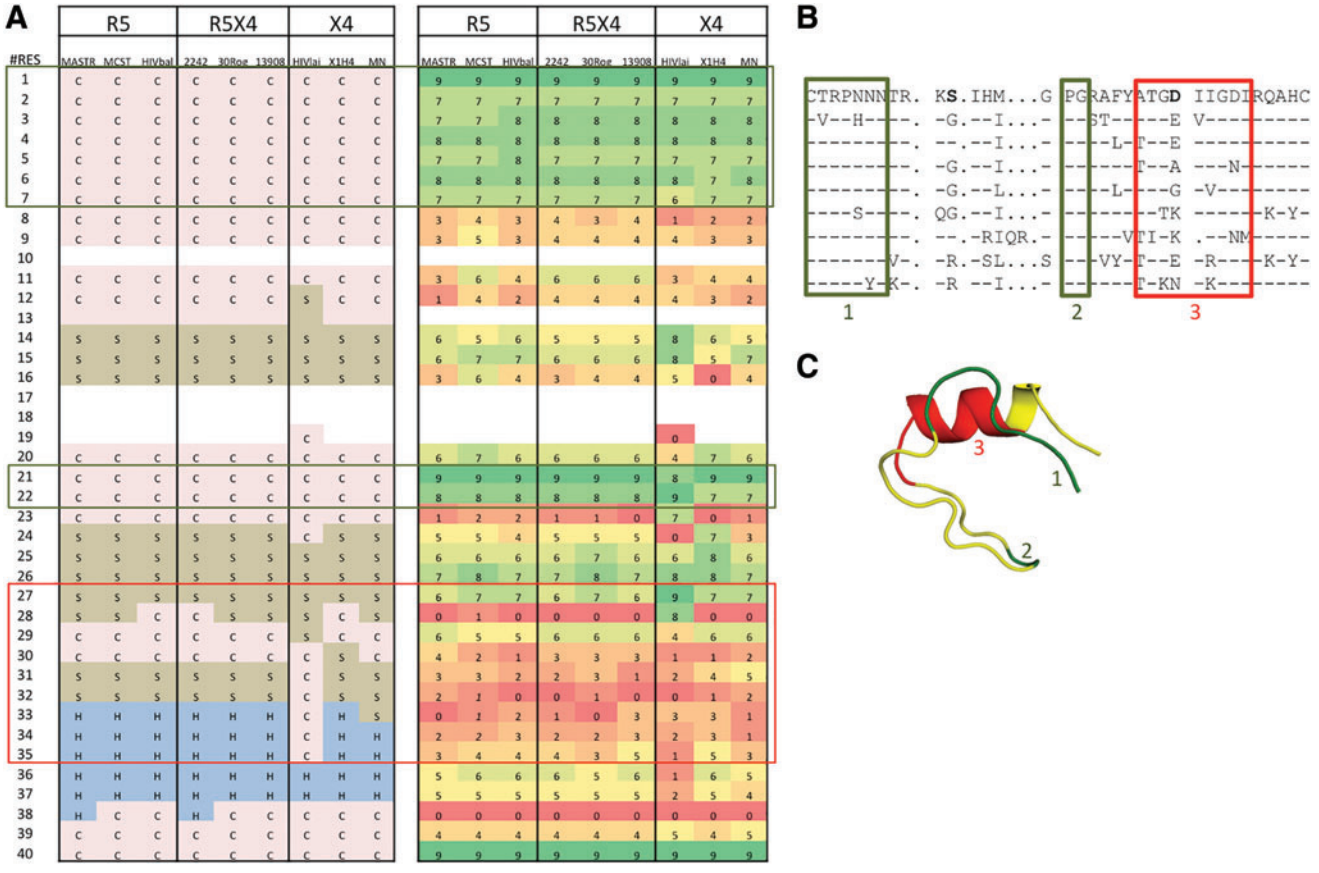
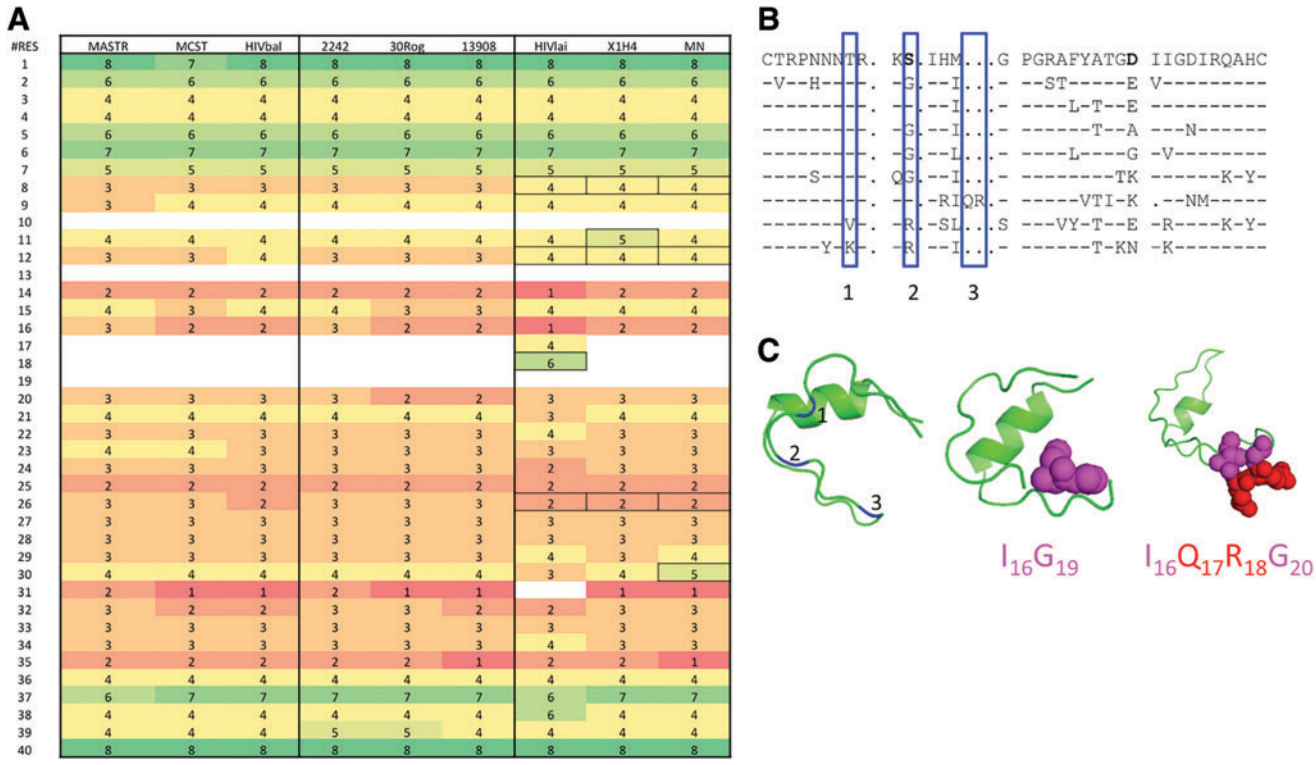


FIG. 4. Predicted secondary structure and confidence after initial modeling using I-Tasser. Numbered residues, corresponding to the aligned sequences, are shown in the *first column*. **(A)** The table on the *left* shows the predicted secondary structure with C = random coil (*pink*), S = beta strand (*gold*), and H = alpha helix (*blue*). The table on the *right* shows the confidence of the secondary structure prediction with 0 = least confident (*red*) to 9 = most confident (*green*). **(B)** The sequences in the alignment are the same as in Figure 1A. *Thick green* and *red boxes* indicate the regions in the tables in Panel A that are highly structured (*dark green*) or less structured (*red*). **(C)** The V3 loop structure depicting these results with *green* = most confident prediction and *red* = least confident. Color images available online at www.liebertpub.com/aid



#RES	MASTR	MCST	HIVbal	2242	30Rog	13908	HIVlai	X1H4	MN
1	8	7	8	8	8	8	8	8	8
2	6	6	6	6	6	6	6	6	6
3	4	4	4	4	4	4	4	4	4
4	4	4	4	4	4	4	4	4	4
5	6	6	6	6	6	6	6	6	6
6	7	7	7	7	7	7	7	7	7
7	5	5	5	5	5	5	5	5	5
8	3	3	3	3	3	3	4	4	4
9	3	4	4	4	4	4	4	4	4
10									
11	4	4	4	4	4	4	4	5	4
12	3	3	4	3	3	3	4	4	4
13									
14	2	2	2	2	2	2	1	2	2
15	4	3	4	4	3	3	4	4	4
16	3	2	2	3	2	2	1	2	2
17							4		
18							6		
19									
20	3	3	3	3	2	2	3	3	3
21	4	4	4	4	4	4	3	4	4
22	3	3	3	3	3	3	4	3	3
23	4	4	3	3	3	3	3	3	3
24	3	3	3	3	3	3	2	3	3
25	2	2	2	2	2	2	2	2	2
26	3	3	2	3	3	3	2	2	2
27	3	3	3	3	3	3	3	3	3
28	3	3	3	3	3	3	3	3	3
29	3	3	3	3	3	3	4	3	4
30	4	4	4	4	4	4	3	4	5
31	2	1	1	2	1	1		1	1
32	3	2	2	3	3	2	2	3	3
33	3	3	3	3	3	3	3	3	3
34	3	3	3	3	3	3	4	3	3
35	2	2	2	2	2	1	2	2	1
36	4	4	4	4	4	4	4	4	4
37	6	7	7	7	7	7	6	7	7
38	4	4	4	4	4	4	6	4	4
39	4	4	4	5	5	4	4	4	4
40	8	8	8	8	8	8	8	8	8

FIG. 5. Predicted solvent accessibility after modeling. Numbered residues, corresponding to the aligned sequences, are shown in the *first column*. **(A)** Values in the table range from 0 (buried residue-*red*) to 9 (highly exposed-*green*). Small differences in exposed regions in X4 models are *boxed*. **(B)** The sequences in the alignment are the same as in Figure 1A. *Blue boxes* indicate potentially exposed regions in X4-tropic HIV. **(C)** Three V3 loop structures are shown. On the far *left*, *blue* designates the location of more exposed residues in X4 tropic viruses. The *middle* structure contains *purple* spheres that correspond to positions 16 and 19 in the alignment of a V3 loop without length variation at positions 17–18. The structure on the *right* shows a V3 loop structure with additional residues at positions 17 and 18. Color images available online at www.liebertpub.com/aid

In particular, the charged amino-acid position at position 12 and amino-acid positions that are observed as insertions appear exposed on the structure's surface.

Molecular dynamics simulations allow atoms and molecules to interact under a controlled set of circumstances (i.e., temperature and pressure) that mimic natural molecular forces and lead to an improved three-dimensional protein structure. These simulations allowed for structural relaxation and adjustments for each V3 loop model after the protein threading stage. Certain trends were observed after this process of modeling (Fig. 6). For example, R5 models almost always retained a more closed ''C-shaped'' loop structure with a well-defined α-helix at the C-terminus of the loop. In the case of the MASTR HIV isolate, secondary structure was not visible after threading, but it appeared in all structures after allowing the model to adjust over time and under controlled temperature and pressure. Loss of the α-helix was observed in two structures from the 30Rog isolate after energy minimization. In the X1H4 isolate, no recovery of the α-helix domain was observed after energy minimization.

More frequently, a less-organized structure was observed in R5X4 and X4 variants in comparison to the R5 isolates. These experiments also highlighted amino acids that exhibited more freedom of movement during the simulation, where red amino acids exhibited more freedom, blue amino acids demonstrated less movement, and green amino acids were intermediate (Fig. 6). The alpha helix in most R5 isolates appears blue to green.

Discussion

The V3 loop of the HIV-1 envelope was identified as a key determinant of tropism decades ago,[26] and three discrete tropism states (R5, R5X4, and X4) are commonly discussed in the literature. In this study, our goal was not to discriminate between tropism phenotypes; rather, it was to identify amino-acid features beyond those more obvious features used in computational algorithms that could potentially be harnessed for therapeutic design or improved co-receptor evaluation.

We identified that a variety of amino-acid features that have not previously been taken into account reside in the HIV-1 V3 loop and appear necessary for HIV-1 to transition between tropism states. Our analysis supports the hypothesis that a continuum of tropisms exists from R5 to X4 through a spectrum of dual-tropic intermediates.[14] Of course, it is also possible that in many cases a virus population can slip between tropism states, only to return to an R5 population after rounds of selection. Our analysis also reinforces the notion that R5X4 and X4 HIV contain a less-ordered structure than R5 structures, particularly in the capacity to maintain an organized secondary structure and a compact configuration.

Early studies have described the change from R5 to X4 HIV-1 as a ''switch,''[27,28] with each tropism having a

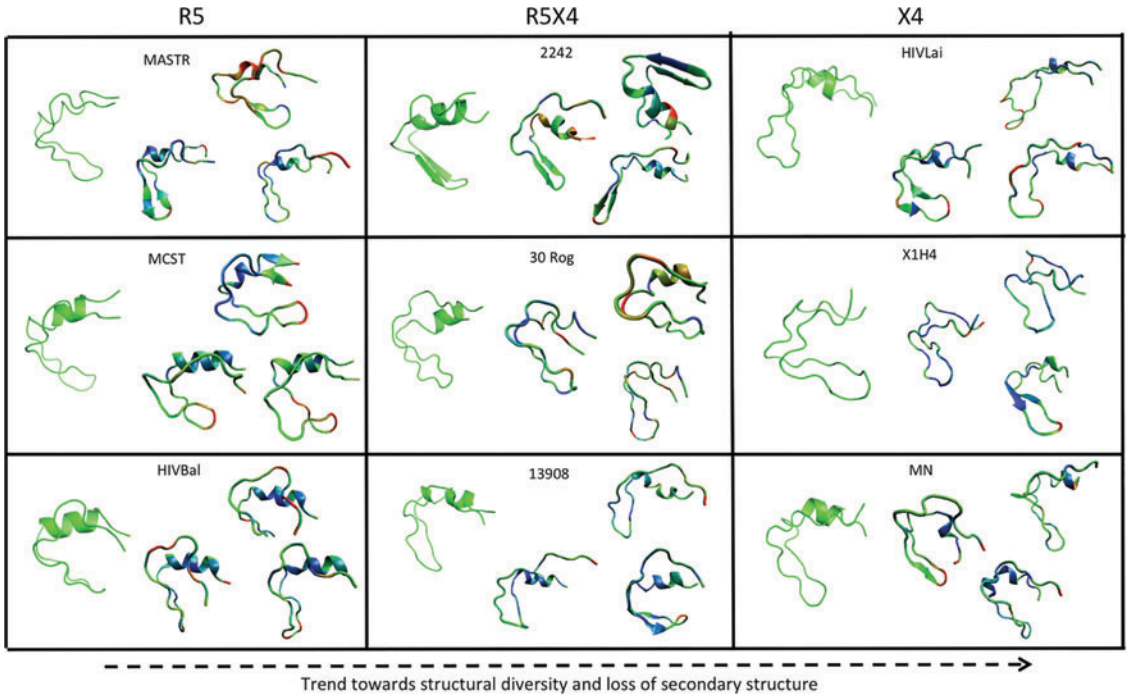


FIG. 6. V3 loop structures for viruses with known tropisms both before and after molecular dynamics simulations. The structures are displayed in a cartoon format. For each isolate, a *green* image is shown on the *left* that represents the structure before minimization; the structures on the *right* show the results of three equilibration experiments with amino acids colored by the average RMSD per residue. *Red* denotes more mobile residues, and *blue* denotes residues that move less during equilibration. Color images available online at www.liebertpub.com/aid

preference for tissue macrophages or T-cells (the primary cellular targets of HIV-1). R5 HIV are more commonly identified within infected patients and also the usual transmitted phenotype.[29,30] X4 HIV were first suggested as contributing to rapid AIDS progression, because they appeared later in disease[31]; however, it is now well understood that AIDS progression is not always associated with X4 viruses, as most late-stage HIV-infected patients contain predominantly R5 HIV-1 or progress only to R5X4.[30]

These transitions among tropisms can be attributed partly to gp120, which is the target for neutralizing antibodies; it is this interplay of gp120 with the immune system that drives a substantial amount of selection against the least-fit variants during successive rounds of viral replication and results in the genetic variability of gp120.[7] However, during complete immune failure, this balance breaks down and lesser-fit isolates (X4) can emerge. In fact, a study recently showed that X4 viruses were closely linked to low nadir CD4 T-cells <100 cells/mm.[32] Therefore, the X4 phenotype is not the driver of disease progression; rather, it is the interaction between gp120 and a weakened immune system occurring during late-stage AIDS progression that allows for reduced viral selection and the evolution of the less structured and more openly configured X4 phenotype.

Dual-tropic HIV-1 can utilize either co-receptor; whereas later studies showed a preference for these to use the X4 receptor in lymphocytes but they use both co-receptors equally in macrophages.[33] This is relevant considering that circulating plasma T-cell levels can vary greatly over the course of infection, whereas HIV-1 is readily amplified from a variety of anatomical tissues when plasma T-cell levels are nonexistent,[34–36] thus implying that a persistently infected tissue-based reservoir still exists during complete immunodeficiency and could contribute to the tropism states observed. Furthermore, under cART, HIV tissue reservoirs are of increasing interest, as viral suppression from plasma fails to fully eradicate infection from an individual[37] and, in fact, lymphoid tissues have recently been implicated as a sanctuary for productive virus under Cart.[38]

The unidirectional transition of R5 →R5X4→X4 observed is a reflection of similar ordinal selective pressures generated in different individuals who generate similar viral phenotypes until such time as the cellular environment (and resulting selective pressure) changes in perhaps less predictable ways. Thus, the specific timing of each of these events is relative to the condition of each patient, and the speed at which HIV progresses from R5 to R5X4 to X4 is a reflection of these differences. To summarize, selection against X4, a dynamic immune system, and the persistence of HIV in anatomical reservoirs where varied HIV tropisms could replicate efficiently in different immune microenvironments (i.e., diseased tissue) could all contribute to a continuum of tropism states during the span of HIV infection. Further studies of tropism modulation in anatomical reservoirs are underway.

Maraviroc is a CCR5 antagonist currently approved by the U.S. Food and Drug Administration for the treatment of patients infected with R5-tropic HIV.[39] Although this therapy in combination with other antiretrovirals has had some success, it still does not greatly alter disease course in HIV-infected patients who routinely follow clinical antiretroviral HIV treatment protocols.[40,41] The partial success of Maraviroc may be due to the ability of HIV to readily infect tissue-based HIV reservoirs (i.e., tissue macrophages),[37] where immune

cells with a more varied receptor concentration may thrive, thus supplying the blood with a low-level consistent source of new virus. For example, in one study, Maraviroc did not affect biomarkers of monocyte/macrophage activation.[42] It may be that future therapeutic approaches could harness the increasing knowledge of V3 loop variability to better block viral entry for cells that are both circulating and sequestered in anatomical sites.

Importantly, viral tropism is likely related to changes outside of the V3 loop domain, either somewhere else within the HIV envelope or even in co-evolution of CD4 with HIV gp120, which has been suggested in several reports.[43–45] The present study has treated V3 loop features associated with tropism as being independent; however, it is likely the case that their nonlinear interaction plays a key role in tropism. For the purpose of this investigation, we chose to assume feature independence to make the correlation between sequence and structure easier to understand. Future work will examine their true interdependencies. Future work will also assess whether the amino-acid features identified as related to tropism shifts in this work also correlate to amino-acid changes found in more controlled assays where HIV is harvested from indicator cells expressing exogenous CD4 and either R5 or R4.

Acknowledgments

The authors would like to thank the Los Alamos National Laboratory HIV Sequence Database for the HIV sequence information used in this study. This work was supported by grants from the U.S. National Institutes of Health under grant R01 MH100984 to M.S.M. and grants P50 GM103297-01 and R01 NS063897-01A2 to M.S. The funders had no role in the study design, data collection and interpretation, or the decision to submit the work for publication.

Author Disclosure Statement

No competing financial interests exist.

References

1. Goodenow MM, Collman RG: HIV-1 coreceptor preference is distinct from target cell tropism: A dual-parameter nomenclature to define viral phenotypes. J Leukoc Biol 2006;80:965–972.
2. Tersmette M, Gruters RA, de Wolf F, de Goede RE, Lange JM, Schellekens PT, Goudsmit J, Huisman HG, Miedema F: Evidence for a role of virulent human immunodeficiency virus (HIV) variants in the pathogenesis of acquired immunodeficiency syndrome: Studies on sequential HIV isolates. J Virol 1989;63:2118–2125.
3. Koot M, Keet IP, Vos AH, de Goede RE, Roos MT, Coutinho RA, Miedema F, Schellekens PT, Tersmette M: Prognostic value of HIV-1 syncytium-inducing phenotype for rate of CD4+ cell depletion and progression to AIDS. Ann Intern Med 1993;118:681–688.
4. Karlsson A, Parsmyr K, Sandstrom E, Fenyo EM, Albert J: MT-2 cell tropism as prognostic marker for disease progression in human immunodeficiency virus type 1 infection. J Clin Microbiol 1994;32:364–370.
5. Scarlatti G, Tresoldi E, Bjorndal A, Fredriksson R, Colognesi C, Deng HK, Malnati MS, Plebani A, Siccardi AG, Littman DR, *et al.*: In vivo evolution of HIV-1 co-receptor usage and sensitivity to chemokine-mediated suppression. Nat Med 1997;3:1259–1265.
6. Platt EJ, Wehrly K, Kuhmann SE, Chesebro B, Kabat D: Effects of CCR5 and CD4 cell surface concentrations on infections by macrophagetropic isolates of human immunodeficiency virus type 1. J Virol 1998; 72:2855–2864.
7. Mild M, Gray RR, Kvist A, Lemey P, Goodenow MM, Fenyo EM, Albert J, Salemi M, Esbjornsson J, Medstrand P: High intrapatient HIV-1 evolutionary rate is associated with CCR5-to-CXCR4 coreceptor switch. Infect Genet Evol 2013;19:369–377.
8. Wyatt R, Kwong PD, Desjardins E, Sweet RW, Robinson J, Hendrickson WA, Sodroski JG: The antigenic structure of the HIV gp120 envelope glycoprotein. Nature 1998;393: 705–711.
9. Kwong PD, Doyle ML, Casper DJ, Cicala C, Leavitt SA, Majeed S, Steenbeke TD, Venturi M, Chaiken I, Fung M, *et al.*: HIV-1 evades antibody-mediated neutralization through conformational masking of receptor-binding sites. Nature 2002;420:678–682.
10. Resch W, Hoffman N, Swanstrom R: Improved success of phenotype prediction of the human immunodeficiency virus type 1 from envelope variable loop 3 sequence using neural networks. Virology 2001;288:51–62.
11. Brumme ZL, Dong WW, Yip B, Wynhoven B, Hoffman NG, Swanstrom R, Jensen MA, Mullins JI, Hogg RS, Montaner JS, *et al.*: Clinical and immunological impact of HIV envelope V3 sequence variation after starting initial triple antiretroviral therapy. AIDS 2004;18:F1–F9.
12. Poveda E, Briz V, Roulet V, Del Mar Gonzalez M, Faudon JL, Skrabal K, Soriano V: Correlation between a phenotypic assay and three bioinformatic tools for determining HIV co-receptor use. AIDS 2007;21:1487–1490.
13. Delgado E, Fernandez-Garcia A, Vega Y, Cuevas T, Pinilla M, Garcia V, Sanchez M, Gonzalez M, Sanchez AM, Thomson MM, *et al.*: Evaluation of genotypic tropism prediction tests compared with in vitro co-receptor usage in HIV-1 primary isolates of diverse subtypes. J Antimicrob Chemother 2012;67:25–31.
14. Lin NH, Kuritzkes DR: Tropism testing in the clinical management of HIV-1 infection. Curr Opin HIV AIDS 2009;4:481–487.
15. Milich L, Margolin B, Swanstrom R: V3 loop of the human immunodeficiency virus type 1 Env protein: Interpreting sequence variability. J Virol 1993;67:5623–5634.
16. Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S: MEGA5: Molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol Biol Evol 2011;28:2731–2739.
17. Wilkins MR, Gasteiger E, Bairoch A, Sanchez JC, Williams KL, Appel RD, Hochstrasser DF: Protein identification and analysis tools in the ExPASy server. Methods Mol Biol 1999;112:531–552.
18. Roy A, Kucukural A, Zhang Y: I-TASSER: A unified platform for automated protein structure and function prediction. Nat Protoc 2010;5:725–738.
19. Zhang Y: I-TASSER server for protein 3D structure prediction. BMC Bioinformatics 2008;9:40.
20. Roy A, Yang J, Zhang Y: COFACTOR: An accurate comparative algorithm for structure-based protein function annotation. Nucleic Acids Res 2012;40:W471–W477.

21. Phillips JC, Braun R, Wang W, Gumbart J, Tajkhorshid E, Villa E, Chipot C, Skeel RD, Kale L, Schulten K: Scalable molecular dynamics with NAMD. J Comput Chem 2005; 26:1781–1802.
22. Humphrey W, Dalke A, Schulten K: VMD: Visual molecular dynamics. J Mol Graph 1996;14:33–38, 27–38.
23. Vranken WF, Budesinsky M, Fant F, Boulez K, Borremans FA: The complete Consensus V3 loop peptide of the envelope protein gp120 of HIV-1 shows pronounced helical character in solution. FEBS Lett 1995;374:117–121.
24. Julien JP, Cupo A, Sok D, Stanfield RL, Lyumkis D, Deller MC, Klasse PJ, Burton DR, Sanders RW, Moore JP, *et al.*: Crystal structure of a soluble cleaved HIV-1 envelope trimer. Science 2013;342:1477–1483.
25. Pejchal R, Doores KJ, Walker LM, Khayat R, Huang PS, Wang SK, Stanfield RL, Julien JP, Ramos A, Crispin M, *et al.*: A potent and broad neutralizing antibody recognizes and penetrates the HIV glycan shield. Science 2011;334: 1097–1103.
26. 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;253:71–74.
27. Kuiken CL, de Jong JJ, Baan E, Keulen W, Tersmette M, Goudsmit J: Evolution of the V3 envelope domain in proviral sequences and isolates of human immunodeficiency virus type 1 during transition of the viral biological phenotype. J Virol 1992;66:4622–4627.
28. Jensen MA, Li FS, van 't Wout AB, Nickle DC, Shriner D, He HX, McLaughlin S, Shankarappa R, Margolick JB, Mullins JI: 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. J Virol 2003;77:13376–13388.
29. Grivel JC, Shattock RJ, Margolis LB: Selective transmission of R5 HIV-1 variants: Where is the gatekeeper? J Transl Med 2011;9 Suppl 1:S6.
30. Gorry PR, Churchill M, Crowe SM, Cunningham AL, Gabuzda D: Pathogenesis of macrophage tropic HIV-1. Curr HIV Res 2005;3:53–60.
31. Penn ML, Grivel JC, Schramm B, Goldsmith MA, Margolis L: CXCR4 utilization is sufficient to trigger CD4+ T cell depletion in HIV-1-infected human lymphoid tissue. Proc Natl Acad Sci U S A 1999;96:663–668.
32. Sechet M, Roussel C, Schmit JL, Saroufim C, Ghomari K, Merrien D, Cordier F, Pik JJ, Landgraf N, Douadi Y, *et al.*: X4 tropic virus prediction is associated with a Nadir CD4 T-cell count below 100 cells/mm. Intervirology 2015;58: 155–159.
33. Yi Y, Shaheen F, Collman RG: Preferential use of CXCR4 by R5X4 human immunodeficiency virus type 1 isolates for infection of primary lymphocytes. J Virol 2005;79:1480–1486.
34. Lamers SL, Thomas J, Fogel GB, McGrath MS: HIV-1 nef protein structures associated with brain tissues. In: *Zing Structural Biology and DRug Discovery Conference*, December 4–7; Puerto Morelos, Mexico, 2010.
35. Lamers SL, Gray RR, Salemi M, Huysentruyt LC, McGrath MS: HIV-1 phylogenetic analysis shows HIV-1 transits through the meninges to brain and peripheral tissues. Infect Genet Evol 2011;11:31–37.
36. Galiwango RM, Lamers SL, Redd AD, Manucci J, Tobian AA, Sewankambo N, Kigozi G, Nakigozi G, Serwadda D, Boaz I, *et al.*: HIV type 1 genetic variation in foreskin and blood from subjects in Rakai, Uganda. AIDS Res Hum Retroviruses 2012;28:729–733.
37. Rappaport J: Editorial: The monocyte/macrophage in the pathogenesis of AIDS: The next frontier for therapeutic intervention in the CNS and beyond: Part I. Curr HIV Res 2014;12:75–76.
38. Rothenberger MK, Keele BF, Wietgrefe SW, Fletcher CV, Beilman GJ, Chipman JG, Khoruts A, Estes JD, Anderson J, Callisto SP, *et al.*: Large number of rebounding/founder HIV variants emerge from multifocal infection in lymphatic tissues after treatment interruption. Proc Natl Acad Sci U S A 2015;112:E1126–E1134.
39. Woollard SM, Kanmogne GD: Maraviroc: A review of its use in HIV infection and beyond. Drug Des Devel Ther 2015;9:5447–5468.
40. Cillo AR, Hilldorfer BB, Lalama CM, McKinnon JE, Coombs RW, Tenorio AR, Fox L, Gandhi RT, Ribaudo H, Currier JS, *et al.*: Virologic and immunologic effects of adding maraviroc to suppressive antiretroviral therapy in individuals with suboptimal CD4+ T-cell recovery. AIDS 2015;29:2121–2129.
41. Rusconi S, Vitiello P, Adorni F, Colella E, Foca E, Capetti A, Meraviglia P, Abeli C, Bonora S, D'Annunzio M, *et al.*: Maraviroc as intensification strategy in HIV-1 positive patients with deficient immunological response: An Italian randomized clinical trial. PLoS One 2013;8:e80157.
42. Patro SC, Azzoni L, Joseph J, Fair MG, Sierra-Madero JG, Rassool MS, Sanne I, Montaner LJ: Antiretroviral therapy in HIV-1-infected individuals with CD4 count below 100 cells/mm3 results in differential recovery of monocyte activation. J Leukoc Biol 2015;pii:jlb.5AB0915-406R.
43. Ghaffari G, Tuttle DL, Briggs D, Burkhardt BR, Bhatt D, Andiman WA, Sleasman JW, Goodenow MM: Complex determinants in human immunodeficiency virus type 1 envelope gp120 mediate CXCR4-dependent infection of macrophages. J Virol 2005;79:13250–13261.
44. Dimonte S, Mercurio F, Svicher V, D'Arrigo R, Perno CF, Ceccherini-Silberstein F: Selected amino acid mutations in HIV-1 B subtype gp41 are associated with specific gp120v(3) signatures in the regulation of co-receptor usage. Retrovirology 2011;8:33.
45. Cicala C, Arthos J, Fauci AS: HIV-1 envelope, integrins and co-receptor use in mucosal transmission of HIV. J Transl Med 2011;9 Suppl 1:S2.
46. Gasteiger E, Hoogland C, Gattiker A, Duvaud S, Wilkins MR, Appel RD, Bairoch A: *Protein Identification and Analysis Tools on the ExPASy Server*. Humana Press, 2005.
47. Zimmerman JM, Eliezer N, Simha R: The characterization of amino acid sequences in proteins by statistical methods. J Theor Biol 1968;21:170–201.
48. Lifson S, Sander C: Antiparallel and parallel beta-strands differ in amino acid residue preferences. Nature 1979;282: 109–111.
49. Deleage G, Roux B: An algorithm for protein secondary structure prediction based on class prediction. Protein Eng 1987;1:289–294.
50. Levitt M: Conformational preferences of amino acids in globular proteins. Biochemistry 1978;17:4277–4285.
51. Chou PY, Fasman GD: Prediction of the secondary structure of proteins from their amino acid sequence. Adv Enzymol Relat Areas Mol Biol 1978;47:45–148.
52. Darby NJ, Creighton TE: *Protein Structure: In Focus* Oxford University Press, 1993.
53. Lamers SL, Salemi M, McGrath MS, Fogel GB: Prediction of R5, X4, and R5X4 HIV-1 coreceptor usage with evolved neural networks. IEEE/ACM Trans Comput Biol Bioinform 2008;5:291–300.

54. Grantham R: Amino acid difference formula to help explain protein evolution. *Science* 1974;185:862–864.
55. Janin J: Surface and inside volumes in globular proteins. *Nature* 1979;277:491–492.
56. Zhao G, London E: An amino acid ''transmembrane tendency'' scale that approaches the theoretical limit to accuracy for prediction of transmembrane helices: Relationship to biological hydrophobicity. Protein Sci 2006;15:1987–2001.
57. Rose GD, Geselowitz AR, Lesser GJ, Lee RH, Zehfus MH: Hydrophobicity of amino acid residues in globular proteins. *Science* 1985;229:834–838.
58. Hofmann HJ, Hadge D: On the theoretical prediction of protein antigenic determinants from amino acid sequences. Biomed Biochim Acta 1987;46:855–866.
59. McCaldon P, Argos P: Oligopeptide biases in protein sequences and their use in predicting protein coding regions in nucleotide sequences. *Proteins* 1988;4:99–122.
60. Dayhoff MO, Schwartz RM, Orcutt BC: *Atlas of Protein Sequence and Structure*, Vol. 5, 1978.
61. Van Baelen K, Vandenbroucke I, Rondelez E, Van Eygen V, Vermeiren H, Stuyver LJ: HIV-1 coreceptor usage determination in clinical isolates using clonal and population-based genotypic and phenotypic assays. J Virol Methods 2007;146:61–73.
62. Sweet RM, Eisenberg D: Correlation of sequence hydrophobicities measures similarity in three-dimensional protein structure. J Mol Biol 1983;171:479–488.
63. Kyte J, Doolittle RF: A simple method for displaying the hydropathic character of a protein. J Mol Biol 1982;157:105–132.
64. Abraham DJ, Leo AJ: Extension of the fragment method to calculate amino acid zwitterion and side chain partition coefficients. Proteins 1987;2:130–152.
65. Bull HB, Breese K: Surface tension of amino acid solutions: A hydrophobicity scale of the amino acid residues. Arch Biochem Biophys 1974;161:665–670.
66. Guy HR: Amino acid side-chain partition energies and distribution of residues in soluble proteins. Biophys J 1985;47:61–70.
67. Roseman MA: Hydrophilicity of polar amino acid side-chains is markedly reduced by flanking peptide bonds. J Mol Biol 1988;200:513–522.
68. Wolfenden R, Andersson L, Cullis PM, Southgate CC: Affinities of amino acid side chains for solvent water. Biochemistry 1981;20:849–855.
69. Wilson KJ, Honegger A, Stotzel RP, Hughes GJ: The behaviour of peptides on reverse-phase supports during high-pressure liquid chromatography. *Biochem J* 1981;199:31–41.
70. Eisenberg D, Schwarz E, Komaromy M, Wall R: Analysis of membrane and surface protein sequences with the hydrophobic moment plot. J Mol Biol 1984;179:125–142.
71. Hopp TP, Woods KR: Prediction of protein antigenic determinants from amino acid sequences. Proc Natl Acad Sci U S A 1981;78:3824–3828.
72. Manavalan P, Ponnuswamy PK: Hydrophobic character of amino acid residues in globular proteins. *Nature* 1978;275:673–674.
73. Fauchere JL, Pliska VE: Hydrophobic parameters-pi of amino-acid side-chains from the partitioning of N-acetyl-amino-acid amides. Eur J Med Chem 1983;18:369–375.
74. Mohana Rao JK, Argos P: A conformational preference parameter to predict helices in integral membrane proteins. Biochim Biophys Acta 1986;869:197–214.
75. Tanford C: Contribution of hydrophobic interactions to the stability of the globular conformation of proteins. J Am Chem Soc 1962;84:4240–4274.
76. Welling GW, Weijer WJ, van der Zee R, Welling-Wester S: Prediction of sequential antigenic regions in proteins. FEBS Lett 1985;188:215–218.
77. Chothia C: The nature of the accessible and buried surfaces in proteins. J Mol Biol 1976;105:1–12.
78. Cowan R, Whittaker RG: Hydrophobicity indices for amino acid residues as determined by high-performance liquid chromatography. Pept Res 1990;3:75–80.
79. Parker JM, Guo D, Hodges RS: New hydrophilicity scale derived from high-performance liquid chromatography peptide retention data: Correlation of predicted surface residues with antigenicity and X-ray-derived accessible sites. *Biochemistry* 1986;25:5425–5432.
80. Browne CA, Bennett HP, Solomon S: The isolation of peptides by high-performance liquid chromatography using predicted elution positions. Anal Biochem 1982;124:201–208.
81. Meek JL: Prediction of peptide retention times in high-pressure liquid chromatography on the basis of amino acid composition. Proc Natl Acad Sci U S A 1980;77:1632–1636.
82. Aboderin AA: An empirical hydrophobicity scale for a-aminoacids and some of its applications. Int J Biochem 1971;2:537–544.
83. Black SD, Mould DR: Development of hydrophobicity parameters to analyze proteins which bear post- or co-translational modifications. Anal Biochem 1991;193:72–82.
84. Miyazawa S, Jernigan RL: Residue-residue potentials with a favorable contact pair term and an unfavorable high packing density term, for simulation and threading. J Mol Biol 1996;256:623–644.
85. Jones DD: Amino acid properties and side-chain orientation in proteins: A cross correlation appraoch. J Theor Biol 197550:167–183.
86. pK and pI values of amino acids. Available at www.anaspec.com/html/pK_n_pl_Values_of_AminoAcids.html

Address correspondence to:
Susanna L. Lamers
Bioinfoexperts, LLC
PO BOX 693
Thibodaux, LA 70301

E-mail: susanna@bioinfox.com