Computational Biology and Bioinformatics
2015; 3(2): 21-30
Published online March 27, 2015 (http://www.sciencepublishinggroup.com/j/cbb)
doi: 10.11648/j.cbb.20150302.11
ISSN: 2330-8265 (Print); ISSN: 2330-8281 (Online)

HIV Tropism Prediction: Digital Signal Processing-Based Bioinformatics ApproachisNon-Sequence Alignment Dependent

Norbert Nwankwo[1, *], Michael Adikwu[2], Ignatus Okafor[3]

[1]Department of Clinical Pharmacy, Faculty of Pharmacy, Madonna University, Elele, Rivers State, Nigeria
[2]Office of the Vice Chancellor, University of Abuja, Abuja, Federal Capital Territory, Nigeria
[3]Departments of Pharmaceutics and Pharmaceutical Technology, Faculty of Pharmacy, University of Jos, Jos, Plateau State, Nigeria

Email address:
nobatn@gmail.com(N. Nwankwo), adikwum@yahoo.com (M. Adikwu), okaforis@yahoo.com (I. Okafor)

To cite this article:
Norbert Nwankwo, Michael Adikwu, Ignatus Okafor. HIV Tropism Prediction: Digital Signal Processing-Based Bioinformatics Approach is Non-Sequence Alignment Dependent. *Computational Biology and Bioinformatics*. Vol. 3, No. 2, 2015, pp. 21-30.
doi: 10.11648/j.cbb.20150302.11

Abstract: Phenotypic and genotypic predictors forHIV/SIV tropism are available. The genotypic predictors are more rational. However, they are sequence alignment dependent only. Regrettably, non-homologous proteins are found to display common biological functionality. This indicates that sequence-dependent predictors cannot be trusted with appropriate classification of the HIV and SIV isolates, especially if the isolates belong to same tropic group but share divergent sequence alignment. There is therefore need for genotypic predictors that will incorporate embedded intrinsic biological characteristics of the HIV and SIV isolates in the determination of HIV tropism. Secondly, more than 30 positions with at least single mutation outside the V3 domain have been found to influence HIV Tropism. Disappointingly, the available sequence alignment-based HIV genotypic predictors engage only the hyper-variable region (V3) of the HIV gp120. This has resulted in inaccurate classification of HIV and SIV strains. Finally, available HIV genotypic predictors are found to lack the ability to identify and accurately evaluate the sequences of most HIV-1 non B clades, HIV-2 and SIV. Against this background, the ability of the Digital Signal Processing (DSP) Technique called Informational Spectrum Method (ISM), which does not engage sequence similarity but the embedded bio-functionalities of the entire gp120 sequence length to predict HIV and SIV tropism is therefore investigated. 83 isolates of HIV and SIV are subjected to ISM and three other procedures. Results are generated and findings correlated. For isolates, which are analyzable by the four procedures, the results from ISM and three other procedures are found to correlate. Using 50% affinity for the host CD4 as the cut-off, the tropism of the uncategorized isolates are predicted. ISM-based technique is adjudged a better procedure. It analyzes the sequences of all HIV (HIV-1 together with non B, and HIV-2) as well as SIV isolates including those that could not be investigated by other genotypic predictors. It engages the embedded biological characteristics rather than sequence similarity and utilizes the entire HIV gp120 sequence-length instead of V3 domain. This makes ISM-based procedure a better tool for over 180,000 isolates of HIV-1, HIV-2 and SIV in the UNIPROT database. Clinical approaches are unfeasible. This study recommends ISM technique principally for viral tropism prediction as it does not discriminate against HIV/SIVcategories. It suggests that further work be done to determine a suitable cut-off (as in geno2pheno[CORECEPTOR]), and the procedure in combination with other genotypic predictors be engaged in developing an algorithm for determining viral tropism.

Keywords: CD4, Charge Rule, Digital Signal Processing, Geno2pheno, Genotypic, HIV/AIDS, HIV Surface Protein, Informational Spectrum Method, Phenotypic, Position-Specific Scoring Matrix

1. Introduction

Multi-step procedures are required for the HIV to gain entry into the host cell and replicate [1]. The first step is the binding interaction between the HIV Surface glycoprotein (HIV gp120) and the host target, the CD4 [1, 2]. This is followed by another interaction between the HIV Transmembrane (HIV gp41) and chemokine co-receptors such as CXCR4 and CCR5 [1-3].

Some HIV isolates prefer the CCR5 co-receptor. Others engage the CXCR4 [2, 4, 5] or a combination of CCR5 and CXCR4 co-receptors (dual) [6]. There exist biological characteristics associated with each category. CXCR4-tropics,

which are mostly T-Cell Lympho-tropics [7], are also known to offer high degree of affinity to the CD4 [8, 9] and pre-dominate the late of HIV/AIDS progression [6, 8, 9]. The CCR5-tropics on the other hand are mostly Macrophage lovers. They dominate the viral population of the newly infected, asymptomatic individuals [6].

The level of affinity between the HIV isolates and the CD4 is a biological characteristic, which has helped identify the mechanism by which HIV infection translate to AIDS infection [8-10]. Additionally, the place of the CCR5 antagonists in the management of HIV/AIDS has necessitated the determination of patients' pre-treatment HIV Co-receptor status [11]. This has helped highlight the relevance of HIV co-receptor tropism.

Another biological property of the HIV isolates, which has helped determine HIV level of infectivity is the Syncytium Inducing (SI) capacity [12]. SI signifies the ability of the isolates to infect the MT-cells and elicit multi-nucleated transfer cell formation [13]. HIV isolates, which acquire this ability, belong to the Syncytium Inducing (SI) category. On the other hand, HIV strains, which lack these capabilities, are classified as Non- Syncytium Inducing (NSI) [12, 14]. HIV isolates that possess SI capacity are mostly CXCR4 whereas those with NSI status are mainly CCR5 [7, 15]. Knowledge of these HIV bio-functionalities is vital in the designing and development of HIV therapeutic interventions [8-10]. Though Digital Signal Processing-based approaches such as the ISM-based have been engaged in assessing various bio-functionalities of HIV [8, 9] and Influenza viruses [16, 17], they have not been engaged in the prediction of HIV/SIV tropism.

Preliminarily, phenotypic approaches have been employed in investigating HIV Phenotypic properties. These techniques have however been found to be resource-wasting, expensive and slow [18]. For example, CCR5 assay techniques engage ELISA-based methods, which have been acknowledged to be resource-consuming. They employ several antagonists of CCR5 such as TAK-779, Regulated on Activation Normal T-cell Expressed and Secreted (RANTES) [19, 20]. CXCR4 assay also uses expensive techniques like Flow Cytometry analysis. This procedure uses costly ligand such as Stromal cell Derivative Factor 1 (SDF-1) [20, 21]. Phenoscript is a recombinant phenotypic assay technique has been described [22]. These phenotypic approaches are considered wasteful as a result of the amount of reagents, equipments, animal tissues and time involved.

Unlike the Phenotypic procedures, the genotypic predictors utilize computational procedures. They do not engage clinical experimentations. They are known to be resource and time saving [23, 24]. The genotypic predictors include Charge Rule [25], Position-Specific Scoring Matrix (PSSM) [26], Geno2pheno [CORECEPTOR] [27] methods. Unfortunately, all available HIV genotypic predictors are known to engage sequence alignment procedures [28, 29]. As a result, there are no genotypic predictors for the determination of HIV tropism that incorporates the viral embedded intrinsic biological characteristics though their need currently, is well-understood.

Secondly, non-sequence alignment techniques have become necessary. This is because; it has been acknowledged that most non-homologous proteins display common biological functionalities [29, 30]. Most HIV isolates therefore, which belong to same tropic group but share divergent sequence alignment will inevitably be incorrectly classified.

Thirdly, available sequence alignment-based HIV genotypic predictors are acknowledged to engage only the hyper-variable region (V3) of the HIV gp120 [28]. Regrettably, more than 30 positions with at least single mutation outside the V3 domain have been identified to impact on the HIV Tropism [8]. Employing HIV genotypic predictors that engage only the V3 motifs only may therefore result in disengaging mutations outside this domain that would help correctly help predict appropriately, the viral tropism leading to incorrect categorization.

Finally, it has been reported that the available genotypic predictors (Charge Rule, PSSM and geno2pheno [CORECEPTOR]) lack the ability to predict the HIV Tropism of most HIV-1 non B Clades, all HIV-2 and some SIV isolates [8]. As a result, genotypic predictors have been criticized, and christened the HIV B clades predictor [28]. This is because these techniques recognize and analyze only the sequences in the V3 region of these isolates [25-27].

All proteins have been found to be analyzable using Digital Signal Processing (DSP) techniques [29-31]. DSP techniques include ISM [16, 17], and Resonant Recognition Model (RRM) [29, 30]. DSP-based, non-sequences alignment-oriented approaches are therefore found more appropriate in categorizing all HIV and SIV isolates. This is because they engage intrinsic biological characteristics that are uncovered from the sequences rather the sequence alignments. These procedures will therefore be of immense help in categorizing HIV/SIV isolates whose gp120 sequences are non-homologous though they share common biological characteristics.

In this study, 83 HIV and SIV isolates, are subjected to ISM procedure, as well as Charge Rule, PSSM and geno2pheno techniques.

The results derived from the four procedures are correlated and presented in section 3.

2. Methods

Four techniques namely Charge Rule [25], Position-Specific Scoring Matrix (PSSM) [26], geno2pheno [27], and a Digital Signal Processing-based technique called Informational Spectrum Method (ISM) [16, 17] are employed.

2.1. Materials

The protein residues of the Surface Glycoprotein (gp120) belonging to 83 HIV and SIV isolates, listed in Tables 2-7 are retrieved from the UNIPROT [32] and engaged.

2.2. Experimental Procedure

The techniques, Informational Spectrum Method (ISM) [16,

17], Charge Rule [25], Position-Specific Scoring Matrix for the CXCR4 and CCR5 co-receptors ($WebPSSM_{X4/R5}$) [26] and geno2pheno[CORECEPTOR] [27] are engaged in this study. Bioinformatics algorithms for Charge Rule, $WebPSSM_{X4/R5}$ and geno2pheno [CORECEPTOR] have been created and websites provided while ISM procedures have been explained [16, 17, 33]. Protein sequences of the V3 domain are employed in the analysis involving Charge Rule, $WebPSSM_{X4/R5}$ and geno2pheno[CORECEPTOR] predictive techniques using these algorithms. The ISM analysis engaged full amino acid length of the gp120 belonging to the HIV and SIV strains.

2.2.1. Genotypic Predictors-Procedures

Protein residues of the V3 domainsfrom the 83 HIV and SIV isolates are retrieved from UNIPROT databaseand analyzed using the already provided programs. For Charge Rule [25] and Position-Specific Scoring Matrix for the CXCR4 and CCR5 co-receptors ($WebPSSM_{X4/R5}$) [26] procedures, HIV/SIV protein residues of the V3 domain were analyzed at [http://indra.mullins.microbiol.washington.edu/webpssm/]. In the case of geno2pheno[CORECEPTOR][27], geno2pheno predictions were carried out at [http://coreceptor.bioinf.mpi-inf.mpg.de/] using 5% False Positive Rate (FPR),and protein residues of the V3 domains only. This process is represented as $geno2pheno_{FPR=5\%}$ [22].

2.2.2. Informational Spectrum Method (ISM)-Procedures

Informational Spectrum Method (ISM) entails bio-molecular interactions encompassing two fundamental procedures namely, inter-molecular and intra-molecular interactions. ISM procedures engaged are summarily provided below, though detailed descriptions exist [16, 17, 33].

Inter-molecular interactions such as bio-recognition and bio-attachment describe the long-range interaction at distances greater than 10 Armstrong. On the other hand, intra-molecular interactions define the physiological and structural interactions at distances less 5 Armstrong [16, 17, 33].

The long-range interactions engage the Electron-Ion Interaction Potential (EIIP), which expresses the energy term of valence electrons [16, 17, 33]. For bio-molecules, their EIIP values have been calculated using equation obtained from the General Model Pseudo-potential [33]:

$$W=0.25Z^{*}\sin(1.04\pi Z^{*})2\pi \quad (1)$$

Z^{*} is the average quasivalence number (AQVN) obtained by an expression:

$$Z^{*}=\sum^{m} niZi/N) \quad (2)$$

Zi expresses the valence number of the i-th atomic constituent; ni represents the number of atoms of the i-th constituent while m is the number of atomic constituents in the molecule. N is the total number of atoms. EIIP values derived using these equations are expressed in terms Rydbergs (Ry).

Table 1. *Electron-Ion Interaction Potential (EIIP) values, showing the degree of the individual participation of the 20 essential amino acids that constitute protein in the interaction.*

Amino Acid	EIIP	Amino Acid	EIIP	Amino Acid	EIIP	Amino Acid	EIIP
A	0.0373	Q	0.0761	L	0.0000	S	0.0829
R	0.0959	E	0.0058	K	0.0371	T	0.0941
N	0.1263	G	0.0050	M	0.0823	W	0.0548
D	0.0036	H	0.0242	F	0.0946	Y	0.0516
C	0.0829	1	0.0000	P	0.0198	V	0.0057

The inter-molecular procedure considers the bio-recognition and bio-attachment between bio-molecules (such as proteins, peptides, etc) and engages EIIP, which is applied in this study. It involves three processes.

They are:

(i) Translation of the Alphabetic Codes of the Protein Residues into Numerical Sequences (Signal)

This entails the exchange of the alphabetic codes representing the protein residues of the gp120 belonging to the HIV and SIV isolates with the corresponding values of the EIIP (Table 1). By this, the protein residues are translated into numerical sequences (also called signals). The signals represent the proteins residues of the gp120 of the HIV and SIV in terms of their affinity with other bio-molecules.

Prior to decomposition using the Discrete Fourier Transform, the numerical sequences are zero-padded. This involves adding zero to the shorter sequences in order to bring them to same window length.

(ii) Decomposition of the Signals Using Discrete Fourier Transform (DFT)

This is the processing of the signals (numerical sequences) obtained from the 83 HIV/SIV isolates by means of Discrete Fourier Transform in order to reveal embedded biological information in terms of binding interaction.

Discrete Fourier Transform [34], expressed as:

$$X(n)= \sum x(m)e^{-j(2/N)nm} \quad (3)$$

Here, n=0, 1, 2, … , N/2, (m) is the m-th member of a given numerical sequence; N, the total number of points in this sequence, while the X(n) are the Discrete Fourier Transform coefficients, describing the level of interaction as amplitude and frequency as the point of interaction in the spectrum.

Bio-functionalities embedded in the proteins are therefore uncovered as Spectral Characteristics or Informational Spectrum, which are defines as below:

$$S(n)=X(n)X^{*}(n)=|X(n)|^{2} \quad (4)$$

where n=0, 1, 2, …, N/2

The distance between the protein residues are assumed to be of equal (equidistant) with d=1. As such, the maximum frequency (F) is expressed as:

$$F=1/2d= 0.5 \quad (5)$$

(iii) Common Informational Spectrum

Proteins with common biological functionalities are known to display maximum amplitude at a Consensus Frequency or position [16, 17, 31] when their Spectral Characteristics or Informational Spectra are point-wise multiplied.

Point-wise multiplication of the spectral characteristics is expressed as:

$$C(j)=\prod S(j) \quad (6)$$

S(j) represents the j-th component of the power spectrum. C(j) is the j-th component of the CIS. Proteins with common bio-functionalities are found to demonstrate common peak called Consensus Frequency (CF) [16, 17, 30, 31].

3. Results

The results presented in this section (Tables 2-7) comprise of the outcomes of the analysis of the 83 isolates using the three genotypic predictors and ISM technique. The outcomes consist of Percentile (describing the genuineness of the sequence), False Positive Rate (FPR), Prediction (using the PSSM and Charge Rule program), as well as the number of Positively Charged (Poschg) amino acid residues. Percentile, Prediction and Poschg are obtained using PSSM and Charge Rule program. FPR is derived by means of geno2pheno software. The ISM-based results (designated ISM) are the level of affinity of the surface proteins (gp120) of the HIV and SIV to the CD4 of the host. NI stands for Not Investigated. ISM of some of the isolates has preliminarily been studied [8].

CTRPNNNTRKSINIGPRALYTTGEIIGDIRQAHC

Amino acid sequences of the V3 region of the Yu-2 isolate with 6 occurrences of the Positively Charged (Poschg) amino acid residues namely R, K, and H (bold). Yu-2 isolate is known to belong to the CCR5 and NSI group.

CTRPNKRKRIHIGPRAFYTTKNIIGTIRQAHC

Amino acid sequences of the V3 region of the MN isolate with 10 occurrences of the Positively Charged (Poschg) amino acid residues (bold). MN isolate belongs to the CXCR4 and SI group.

Figure 1. *Showing the amino acid sequences of the V3 region of Yu-2 and MN isolates. The Figure indicates number of occurrences (10 in MN, a CXCR4/SI category) and (6 in Yu-2, a CCR5/NSI category) of the Positively Charged (Poschg) amino acid residues (bold). As indicated, the CXCR4/SI category has more Poschg values than the CCR5/NSI.*

Table 2. *This is the results of the clinically categorized HIV-1 T-tropics (CXCR4 and SI) viruses. ISM findings show consistency with the preliminary clinical outcomes except for CDC-451 and Z6 (30%and 37% affinity for the CD4, respectively), a characteristic of CCR5/NSI.Thegenotypic predictors also demonstrated reliability.*

Protein ID	Isolate	Percentile %	FPR %	Prediction	Poschg	ISM %	Consistency%
P04578	HXB2	81.68	0.00	CXCR4	9	92.21	100
P03377	BRU/LA1	78.27	0.50	CXCR4	8	84.77	100
P03375	BH10	78.27	0.50	CXCR4	8	98.14	100
P05877	MN	78.71	0.20	CXCR4	10	58.61	100
P18799	NDK	98.8	0.20	CXCR4	8	61.41	100
P04582	BH8	81.39	0.20	CXCR4	8	84.28	100
P05879	CDC-451	70.98	0.20	CXCR4	9	30	75
P12488	BRVA	57.46	0.20	CXCR4	8	68.30	100
P05878	SC	23.18	80.10	CCR5	6	67.13	50
P31872	WMJ1	63.81	1.70	CXCR4	9	48.90	75
P05880	WMJ2	94.64	0.70	CXCR4	8	49.88	75
P04580	Z6	98.77	19.10	CXCR4	8	37.30	75
P03378	ARV2/SF2	56.07	1.50	CXCR4	9	65.01	100
P04624	HXB3	81.39	0.10	CXCR4	8	96.55	100
P19551	MFA	78.27	0.50	CXCR4	8	100	100
P04581	ELI	97.95	1.30	CXCR4	6	45.21	75
P05881	Z321	34.36	0.50	CCR5	6	62.53	75
O89292	93BR020	46.32	8.60	CXCR4	8	57.9	100
P19549	SF33	81.61	0.70	CXCR4	8	67.13	100
Q3ZLG7	90TH_BK	83.72	0.2	CXCR4	7	68.09	100
12E694	90TH_BK	68.42	0.2	CXCR4	8	65.19	100

Table 3. *This shows the results of the clinically categorized HIV-1 M-tropics (CCR5 and NSI) viruses. The results indicate consistency with the preliminarily determined clinical outcomes except for the LW123 (82.96%) for CD4.*

Protein ID	Isolate	Percentile %	FPR %	Prediction	Poschg	ISM %	Consistency %
Q70626	LW123	83.85	0.00	CXCR4	9	82.96	0
P35961	YU-2	39.29	75.60	CCR5	6	40.48	100
P04579	RF/HAT3	72.70	2.60	CCR5	8	27.29	50
O41803	92NG083	24.23	98.80	CCR5	5	28.68	100
Q75008	ETH2220	30.69	98	CCR5	5	26.27	100
Q9WC60	SE9280	70.59	92.60	CCR5	6	41.03	100
Q9WC69	SE9173	53.54	71.50	CCR5	6	57.26	75
O91086	YBF30	95.37	17.80	CCR5	3	23.03	75
P05882	Z84	95.77	0.70	CCR5	6	20.78	75
P20871	JRCSF	14.64	31.70	CCR5	7	66.08	50
O12164	92BR025	5.97	38.40	CCR5	6	61.84	75
O70902	90CF056	12.82	49.70	CCR5	6	51.74	75

Table 2 presents the results of the clinically categorized HIV-1 T-tropics (CXCR4 and SI) viruses. From Table 2, the following are observed:

The Percentile of protein residues of V3 region belonging to all the HIV-1 T-tropics (CXCR4 and SI) viruses are below 95%.

All the isolates except SC (80.10%) have False Positive Rate (FPR) below 20%.

The PSSM and Charge Rule-based CXCR4 and CCR5 predictions are consistent with the results obtained clinically, except for the isolates, SC and Z321 (CCR5).

None of the clinically categorized HIV-1 T-tropics (CXCR4 and SI) viruses in Table 2 has Positively Charged (Poschg) amino acid residues less than 6.

In the case of the ISM-based analysis, all the isolates except CDC-451 (30%) and Z6 (37.30%) have more than 40% binding interaction with the host CD4.

Table 3 presents the results of the clinically categorized HIV-1 M-tropics (CCR5 and NSI) viruses. The results present the following:

The Percentile of sequences obtained from the HIV gp120 Variable 3 (V3) motifs belonging to the HIV-1 M-tropics (CCR5 and NSI) viruses, which are engaged in this study do not exceed 95%.

All the isolates except LW123 (0%), RF/HAT3 (2.60%), and Z84 (0.70%) have False Positive Rate (FPR) greater than 20%.

Based on the PSSM-Charge Rule algorithm, the CXCR4 and CCR5 predictions are found to be consistent with the clinically experimented findings, except for an isolate called LW123.

In terms of Positively Charged (Poschg) amino acid residues, only LW123 (9), RF/HAT3 (8) and JRCSF isolates exceeded 6, characteristics found more in the CXCR4/T-Tropics.

The ISM-based prediction shows that only three isolates LW123 (82.96%), JRCSF (66.08%) and 92BR025 (61.84%) have binding interaction with the CD4 greater than 60%.

Table 4. *This is the results of the clinically categorized HIV-1 Dual-tropics viruses.ISM outcomes are in accord with the clinical characteristics of the viruses. Other outcomes show consistency with clinically derived results except results from Prediction algorithm.*

Protein ID	Isolate	Percentile %	FPR %	Prediction	Poschg	ISM %	Consistency%
P19550	SF162M	9.51	42.7	CCR5	6	37.50	100
P19550	SF162T	50.07	6.90	CCR5	7	41.08	25
Q73372	89.6	66.00	0.20	CXCR4	9	66.28	100

Table 5. *This is the results of the clinically categorized HIV-2 Viruses. Percentile result, which determines the genuineness of the sequence alignment, could not be generated. This signifies unsuitability of other predictions. ISM is non-sequence alignment dependent.*

Protein ID	Isolate	Percentile %	FPR %	Prediction	Poschg	ISM %
P32536	ST/24.1#2	No value	No value	CXCR4	7	34.13
P15831	D205	No value	No value	CXCR4	8	40.85
P24105	CAM2	No value	No value	CXCR4	9	41.14
P18040	Ghana-1	No value	No value	CXCR4	6	22.49
Q76638	UC1	No value	No value	CXCR4	6	42.28
P17755	D194	No value	No value	CXCR4	9	28.87
P04577	ROD (A)	No value	No value	CCR5	10	38.07
P05883	NIH-Z	No value	No value	CCR5	8	10.82

Table 6. *This illustrates the results of the clinically categorized SIV Viruses. All sequences in this category show percentile greater than 95%, signifying lack of authenticity. Only ISM technique, which is non-sequence dependent could therefore assess the sequences and provided acceptable results.*

Protein ID	Isolate	Percentile %	FPR %	Prediction	Poschg	ISM %
Q17281	cpz GAB1	95.12	83.30	CXCR4	4	37.82
Q1A261	MB66	96.54	19.80	CXCR4	3	31.94
Q8A1H5	TAN1	95.66	17.10	CXCR4	4	34.49
Q02837	AGM gr-1	No value	No value	CCR5	7	51.42
P08810	Mm251	No value	No value	CCR5	6	37.73

Table 4 presents the outcome of the clinically categorized HIV-1 Dual-tropics. These are strains that use both CXCR4 and CCR5 co-receptors and also show both SI and NSI characteristics [6]. The findings are presented as follows:

The sequences of the Variable 3 (V3) motif of the HIV gp120 belonging to the dual strains have Percentile lower than 95%.

By means of the geno2pheno program, which assessed the FPR, the SF162M and its dual counterpart, the SF162T presented a wider difference SF162M (6.9%) and SF162T (42.7%).

The predicted results are in agreementwith the clinical results except for the SF162T, which is predicted as CCR5 while it belongs to the CXCR4.

The Positively Charged (Poschg) amino acid residues belonging to the CXCR4/SI category (SF162T) demonstrated an additional amino acid residue charge (7).

The ISM-based procedure revealed a little difference in their gp120-CD4 interaction. They are SF162M (37.5%) and SF162M (41.08%), respectively.

Table 5 presents the findings of the analysis of the clinically categorized HIV-2 viruses. The findings present the following:

According to the algorithm [27], Percentile and False Positive Result (FPR) of the sequences of all the HIV-2 strains stated are unreliable.

Only PSSM-based Prediction and the number of Positively Charged (Poschg) amino acid residues are obtainable. Based on the PSSM-based prediction, ROD (A) and NIH-Z are assigned CCR5 while others are CXCR4. In the case of Poschg, none is less than 6.

Of all the 8 HIV-2 isolates investigated, only one isolate, UC1 has binding interaction with the CD4 as high as 42.28%.

Table 6 shows the findings of the analysis of the clinically categorized SIV. The results present the following:

All the sequences from the SIV strains displayed False Positive Rate (FPR) greater than 95%.

All the strains except AGM gr-1 and Mm251 are predicted as CCR5 by the PSSM algorithm.

The number of Positively Charged (Poschg) amino acid residues of all the SIV strains fall below 6 except for the AGM gr-1.

Only AGM gr-1 has the binding interaction with the CD4 greater than 50% as obtained using ISM-based technique.

4. Discussions

This section interprets the results presented in section 3. It further correlates the outcomes of the programs engaged and the clinically derived findings. Prior to these elucidations, the correlation guideline and rules for the explanation of the results are set as below.

4.1. The Correlation Guideline and Rules

The intention of this research is to predict the viral tropism of the HIV and SIV isolates using ISM technique and further correlate the ISM-based results with those derived by means of the three predictors. In an attempt to appropriately predict HIV/SIV isolates and also correlate with other findings, the following rules are observed in interpreting the findings:

According to the European guidelines on the clinical management of HIV-1 tropism testing [35], HIV and maybe SIV isolates, which demonstrate False Positive Rate (FPR) 20% and below are categorized as CXCR4. Similarly, those with FPR 20% and above belong to the CCR5.

Secondly, isolates that recorded 6 and above, the number of Positively Charged (Poschg) amino acid residues are classified as CXCR4, while those with 6 and below are likely CCR5. This is in regards to the Charge Rule [24-26].

The PSSM algorithm has viral prediction certified by it. This is engaged and the results incorporated in the study.

The last set of results employed is obtained using Informational Spectrum Method. This represents the degree of affinity between the gp120 of the isolates and host CD4. Some of these findings have preliminarily been published [8]. Isolates with 50% and above are considered to belong to the family of the CXCR4 while those with 50% and below are believed to belong to the CCR5. CXCR4/SI has been found to possess high binding interaction unlike the CCR5/NSI category [8, 9], though some CXCR4/SI isolates such as SC are found to provide affinity less than 50% and few CCR5/NSI isolate including LW123 provided more than 50% affinity for the host CD4.

Each criterion for the correlation is awarded 25%. There are four criteria.

The results obtained in this study are interpreted in the following order:

4.2. Interpretations of Results Presented in Table 2

This belongs to the analysis of the clinically categorized HIV-1 T-tropics (CXCR4 and SI) viruses assessed using three Genotype Predictors and ISM techniques. They demonstrated that:

The sequences of the HIV-1 T-tropics (CXCR4 and SI) viruses engaged possess genuine HIV gp120 Variable 3 (V3) motif except for the Zairian isolates NDK (98.8%), Z6

(98.77%), and ELI (97.95%). This is because; none of the isolates in this category has the Percentile of its protein residues for their V3 region greater than 95%. This is in accordance to the geno2pheno program.

All the isolates except SC (80.10%) have acceptable level of False Positive Rate (FPR) for the CXCR5/SI category signifying that the genotypic predictor appropriately predicted the isolates in this group. The European guidelines on the clinical management of HIV-1 tropism testing take 20% FPR as maximum value for the CXCR4/SI isolates [35].

Additionally, the PSSM-based CXCR4 and CCR5 predictions obtained here appear to be consistent with the results obtained clinically reported outcomes, except for the isolates, SC and Z321, which are classified as CCR5.

None of the isolates in Table 2 have the number of Positively Charged (Poschg) amino acid residues less than 6. This is in accord with the characteristics of the group.

On the part of the ISM-based analysis, all the isolates except CDC-451 (30%) and Z6 (37.30%) have reasonable level of binding interaction with the host CD4.

The two varieties of HIV-1 isolate of 90TH_BK with accession numbers Q3ZLG7 and 12E694, which are also, examined demonstrated characteristics attributable to the group, CXCR4 by means of the four procedures.

4.3. Interpretations of Results Presented in Table 3

From the analysis of the results obtained using clinically categorized HIV-1 M-tropics (CCR5 and NSI) viruses and assessed using three Genotype Predictors and ISM techniques as presented in Table 3, it can be deduced that:

The sequences obtained from the HIV gp120 Variable 3 (V3) motifs belonging to the HIV-1 M-tropics (CCR5 and NSI) viruses, which are engaged in this study are found to be genuine. This is because, the Percentile of protein residues their V3 region do not exceed 95% according to the geno2pheno program.

All the isolates except LW123 (0%), RF/HAT3 (2.60%), and Z84 (0.70%) have acceptable level of FPR for the CCR5 and NSI category. This is an indication that the isolates (except LW123, RF/HAT3 and Z84) clinically classified here as M-tropic (CCR5 and NSI) displayed a value above 20%. This is in accord with the European guidelines on the clinical management of HIV-1 tropism testing [35]. The acceptable lower limit for the CCR5 according to the guideline is 20%.

As shown in Table 3, by means of the PSSM algorithm, the CXCR4 and CCR5 predictions derived are also found to be consistent with the clinically experimented findings, except for an isolate called LW123. LW123 demonstrated a characteristic CXCR4 that is associated with the CXCR4/T –tropics.

In terms of Poschg, the LW123 (9) and RF/HAT3 (8) isolates have the number of positive amino acids residues greater than 6. This characteristic is mostly found in the CXCR4/T-Tropics.

The ISM-based prediction shows that four isolates LW123, JRCSF 92BR025, and 90CF056 have binding interaction with the CD4 above 50%.

4.4. Interpretations of Results Presented in Table 4

In the case of Table 4, which presents the findings made using HIV-1 Dual-tropics isolates that are known to use both CXCR4 and CCR5 co-receptors and also show both SI and NSI characteristics, it is disclosed that:

The sequences of the Variable 3 (V3) motif of the HIV gp120 belonging to the dual strains are genuine. This is because their Percentile falls below 95% as specified by the geno2pheno program.

The geno2pheno program produced better results than the ISM procedure. This is because the SF162M (6.9%) and its dual counterpart, the SF162T (42.7%) presented a wider difference in their False Positive Rate (FPR) compared to the SF162M (37.5%) and SF162M (41.08%).

The PSSM and Charge Rule-based genotypic predictions (CXCR4 and CCR5 features) as well as the clinically obtained results are in agreement except for the SF162T, which is predicted as CCR5 though it belongs to the CXCR4.

The Positively Charged (Poschg) amino acid residues belonging to the CXCR4/SI category (SF162T) demonstrated an additional amino acid residue charge (7).

Though the ISM results agreed with the clinically established finding, the outcome revealed a little margin. They are SF162M (37.5%) and SF162M (41.08%), respectively.

4.5. Interpretations of Results Presented in Table 5

Table 5 displays the deductions from the results of the clinically categorized HIV-2 viruses studied using the three Genotype Predictors and ISM techniques.

The sequences of all the HIV-2 strains are found not to possess the genuine HIV gp120 Variable 3 (V3) motif. This is because they have Percentile of 100%. However, geno2pheno program recommends Percentile less than 95%.

Since the sequences of the HIV gp120 Variable 3 (V3) motif are declared invalid, all genotypic predictions (FPR, Poschg, and CXCR4/CCR5) are therefore incorrect

In this case, ISM has advantage. Its results are not based on sequence alignment rather intrinsic biological functionalities. Of all the gp120 belonging to the HIV-2 isolates investigated, none is found to possess a binding interaction as much as 50%. This is in line with the clinical results and is associated with the physiological characteristics of the HIV-2 category. They are known to circumvent CD4 interaction during infection [36].

Clinically, ROD (A) and NIH-Z are already categorized as CCR5/NSI [38]. This is in agreement with their prediction here (CCR5) and level of binding with the host CD4 (38.07% and 10.82% respectively).

4.6. Interpretations of Results Presented in Table 6

Table 6 shows the findings of the analysis of the clinically categorized SIV viruses studied using the three Genotype Predictors and ISM techniques. It is demonstrated here that:

The sequences of all the SIV strains are found not to possess the genuine HIV gp120 Variable 3 (V3) motif. This is because they have Percentile of 100% while the geno2pheno program

proposes values not more than 95%.Therefore, genotypic predictions (FPR, Poschg, and CXCR4/CCR5) will be inaccurate.

This is unlike the ISM-based results that engage the biological features rather than sequence information. The results of the five (5) SIV isolates obtained by ISM showed that only one isolate (AGM gr-1) has its gp120 achieving over 50% binding interaction with the CD4. AGM gr-1 has clinically been identified to belong to the CXCR4/SI.

ISM findings are in agreement with the clinical results. SIV isolates are known to engage other chemokines instead of CD4 [36] and are therefore known to be less attracted to the CD4.

Summarily, only one isolate (SC) did not agree with the clinical findings made in CXCR4/SI group. In the case of CCR5/NSI category, isolates LW123 and SE9173 show reasonable variation while JRCSF, 92BR025 and 90CF056 demonstrated weak deviation.

As shown in Table 3, LW123 though an M-tropic isolate, has all the characteristics of the T-tropic virus namely FPR (0%), Prediction (CXCR4), Poschg (9), ISM (82.96%). This can be explained by the fact that it is a clone of a T-tropic isolate LAI/111B [37]. For the dual-tropic viruses, HIV-2 and SIV the results presented are in agreement with the clinically derived outcomes.

ISM procedure appears to show more consistency with the outcomes clinically derived as it assesses all HIV and SIV isolates. It is therefore engaged in further prediction of the viral tropism of the clinically uncategorized isolates.

Table 7. *This presents the results of the clinically uncategorized HIV-1 Isolates. HIV tropism is predicted using ISM technique. 50% affinity for the host CD4 was chosen to be the cut-off.Authentication of the results is carried out using results from the three Genotype Predictors (FPR, Prediction, Poschg).*

Protein ID	Isolate	Percentile %	FPR %	Prediction	Poschg	ISM %	ISM-based Predicted Group
P04583	MAL	69.48	11.00	CCR5	6	29.21	CCR5/NSI
P12489	JH32	83.3	0.20	CXCR4	10	35.46	CXCR4/SI
P20888	OYI	76.37	1.10	CXCR4	8	87.51	CXCR4/SI
P31819	KB-1/ETR	66.38	0.20	CXCR4	8	67.19	CXCR4/SI
Q9QBY2	96CM-MP535	9.13	38.00	CCR5	7	12.36	CCR5/NSI
Q9QBZ8	97ZR-EQTB11	7.87	61.10	CCR5	7	46.19	CCR5/NSI
Q9QBZ0	96CM-MP257	30.94	20.80	CCR5	8	44.86	CCR5/NSI
Q9QBZ4	96CM-MP255	51.04	47	CCR5	7	27.90	CCR5/NSI
Q9QSQ7	V1850	36.45	87.40	CCR5	7	35.26	CCR5/NSI
P12487	Z2/CDC-Z34	98.8	0.20	CXCR4	6	37.56	CCR5/NSI
P12490	NY5	78.71	2.40	CXCR4	8	NI	None
P12491	Z3	93.97	1.50	CXCR4	8	27.02	CCR5/NSI
Q9Q714	VI99I	56.58	9.6	CXCR4	6	29.48	CCR5/NSI
Q77377	ANT70	98.72	8.5	CCR5	3	30.25	CCR5/NSI
Q3ZLG6	98US_MSC5016	47.82	22	CXCR4	5	24.34	CCR5/NSI
Q9IDV2	YBF106	95.44	9.6	CXCR4	3	22.58	CCR5/NSI
Q4QWT0	98US_MSC5007	28.65	84.9	CXCR4	6	40.23	CCR5/NSI

Table 8. *Results of the clinically categorized HIV-2 and SIV Viruses, studied using three Genotype Predictors only. The results of the Percentile and False Positive Result (FPR) are presented as "Results are questionable, please check your sequence".*

Protein ID	Isolates	Poschg	Protein ID	Isolates	Poschg
	Isolate 20.77	6		Isolate 18.21	8
Q89607	EHO	8		Isolate 14.26	10
	Isolate 21.27	6		Isolate 18.26	8
P18094	Ben	6		AGM155 20.82	7
P12449	SBLISY	10		AGM3 18.92	7
	Isolate 23.57	7		AGM 20.82	7
	Mn14283.91	6		AGM 20.24	7
	Isolate 22.10	6		PBj14/BCL3 20.08	6

Predictions 1: HIV-1 Isolates

Using 50% cut-off, MAL, JH32, OYI, KB-1/ETR are predicted as CXCR4/SI while 96CM-MP535, 97ZR-EQTB11, 96CM-MP257, 96CM-MP255, V1850, Z2/CDC-Z34, Z3, VI99I, ANT70, 98US_MSC5016, YBF106 and 98US_MSC5007 as CCR5/NSI (Table 7). However, it is recommended that further work be done on the choice of cut-off as in the case of the Geno2pheno [CORECEPTOR] [27] so as to determine a more suitable cut-off for the ISM-based viral tropism prediction. This will help fine-tune the prediction.

Predictions 2: HIV-2 and SIV Isolates

In the case of the HIV-2 and SIV, their sequences are found to be unauthentic. Results obtained using these unacceptable outcomes are reported as invalid. The ISM technique, unlike others does not depend on the sequence similarity of the V3 domain. As a result, it analyzes the sequences of the HIV-2 and SIV strains and could therefore predict their viral tropism. ISM technique is consequently recommended.

5. Conclusions

HIV Phenotypic predictive approaches as well as the genotypic predictors abound. The genotypic predictors consume less resources and time but are sequence alignment dependent only. Incorporation of embedded intrinsic biological characteristics of the HIV and SIV isolates in the determination of HIV tropism has become essential since non-homologous proteins are found to display common biological functionalities. Engaging such techniques will help appropriately classify most HIV isolates, which belong to same tropic group but share divergent sequence alignment.

Though there are more than 30 positions with at least single mutation outside the V3 domain have been found to influence HIV Tropism, available sequence alignment-based HIV genotypic predictors could only engage the hyper-variable region (V3) of the HIV gp120. This has introduced inaccuracy in the classification. Available HIV genotypic predictors have also been denounced for being fashioned only for the B subtype as they lack the ability to identify and accurately evaluate the sequences of most HIV-1 non B clades, HIV-2 and SIV. As evident in the analysis, the sequences of the HIV-1 non B, HIV-2 and SIV could not be analyzed by means of the genotypic predictive programs unlike the ISM technique. This makes ISM a better procedure than the three existing genotypic predictors as it catered for the inadequacies of the genotypic predictors mentioned above.

In addition, the results obtained using the three genotypic predictors and the ISM techniques are found to correlate with the preliminarily derived clinical findings for the already clinically categorized isolates.ISM procedure was therefore further engaged in the prediction of 17 clinically uncategorized HIV isolates using 50% cut-off.This procedure can therefore suitably and appropriately be employed in the tropism of over 180,000 of the reviewed (curated) and un-reviewed HIV and SIV isolates deposited in the UNIPROT database.Clinical approaches to the prediction of their tropism are challenging. DSP-based procedures have become essential tools in the determination of viral tropism. Knowledge of the HIV tropism has become necessary as designing of therapeutic interventions including vaccines [30, 39] has remained unachievable partly due to diversity in the viral biological characteristics including tropism.

References

[1] Adamson C and Freed E, “Novel approaches to inhibiting HIV-1 replication” Antiviral Research vol. 85(2010) pp. 119-141, 2010.

[2] Doms R W, “Beyond receptor expression: the influence of receptor conformation, density, and affinity in HIV-1 infection”,Virology, vol. 276(2) pp. 229-237, 2000.

[3] Berger E A, Murphy P M, Farber J M, “Chemokine receptors as HIV-1 co-receptors: roles in entry, tropism, and disease”, Annu. Rev. Immunol., vol. 17, pp. 657-700, 1999.

[4] Lusso P, “HIV and the chemokine system: 10 years later”, The EMBO Journal, vol. 25, pp. 447-456, 2006.

[5] Berger E R, Doms E, Fenyo B, Korber D, *et al,* “A new classification for HIV-1”, Nature, vol. 391, p. 240, 1998.

[6] Doranz BJ, Yi Y, Smyth RJ, “A Dual-Tropic primary HIV-1 isolate that uses Fusin and the B-chemokine receptors CKR-5, CKR-3, CKR-2 as fusion cofactors”,Cell,vol. 85 pp. 1149-1158, 1996.

[7] Schols D, Struyf S, Van Damme J, *et al.* 1997. “Inhibition of T-tropic HIV strains by selective antagonization of the chemokine receptor cxcr4”. J. Exp. Med. 186(8):1383-1388.

[8] Nwankwo N, “Signal processing-based Bioinformatics methods for characterization and identification of Bio-functionalities of proteins”, PhD Thesis (submitted), De Montfort University, Leicester, United Kingdom, 2012. (Available at www.openthesis.org).

[9] Nwankwo N, Seker H, “HIV progression to AIDS: Bioinformatics approach to determining the mechanism of Action”,Curr. HIV Res.,vol. 11(1), pp. 30-42, 2013.

[10] Collman R, Balliet JW, Gregory SA, “An infectious molecular clone of an unusual macrophage-tropic and highly cytopathic strains of Human Immunodeficiency Virus type 1”, Journal of Virology, vol. 66(12), pp. 7517-7521 1992.

[11] Svicher V, D’Arrigo R, Alteri C, *et al*, “Performance of genotypic tropism testing in clinical practice using the Enhanced Sensitivity version of Trofile as reference assay: results from the Oscar study group”, New Microbiol., vol. 33(3) pp. 195-206, 2010.

[12] Blaak H, Vant-Wout AB, Brouwer M, *et al*, “In vivo hiv-1 infection of cd45ra+ cd4+ t cells is established primarily by syncytium-inducing variants and correlates with the rate of cd41 t cell decline”,PNAS,vol. 97(3), pp. 1269-1274, 1999.

[13] Nagy K, Clapham P, “Human T-cell leukemia virus type 1: Induction of Syncytia and inhibition by patients’ sera”, International Journal of Cancer, vol. 32(3) pp. 321-328, 1983.

[14] Nkengasong JN, Peeters M, Nys P, *et al*. 1995, “Infectious virus titer, replicative and syncytium-inducing capacity of human immunodeficiency virus type 1”, J Med Virol., vol. 45(1) pp. 78-81, 1995.

[15] Blaak H, Vant-Wout AB, Brouwer M, “In vivo hiv-1 infection of cd45ra+ cd4+ t cells is established primarily by syncytium-inducing variants and correlates with the rate of cd41 t cell decline”, PNAS, vol. 97(3) pp. 1269-1274, 1999.

[16] Veljkovic V, Niman HL, Glisic S, *et al*, “Identification of hemagglutinin structural domain and polymorphisms which may modulate swine h1n1 interactions with human receptor”, BMC Structural Biology, vol. 9(62) pp. 1-11, 2009.

[17] Veljkovic V. and Veljkovic N, “Characterization of conserved properties of hemagglutinin of h5n1 and human inuenza viruses: possible consequences for therapy and infection control”, BMC Structural Biology, vol. 9(21) pp. 1-11, 2009.

[18] W. O. F. A. H. (OIE), “Laboratory Methodologies for Bacterial Antimicrobial Susceptibility Test”, 2008.

[19] Zhao Q, *et al*, “A novel assay to identify entry inhibitors that block binding of HIV-1 gp120 to CCR5”, Virology, vol. 326(2) pp. 299-309, 2004.

[20] Princen K, *et al*, "Evaluation of SDF-1/CXCR4-induced Ca2+ signaling by fluorometeric imaging plate reader (FLIPR) and flow cytometry",Cytometry A, vol. 51(1) pp. 35-45, 2003.

[21] JapourI,*et al*, "Standardized microtiter assay for determination of Syncytium Inducing Phenotype of Clinical Human Immunodeficiency Virus type 1 isolates", J ClinMicrobiol., vol. 32(9), pp. 2291-2294, 1994.

[22] Secle´n E, Garrido C, Gonzalez MM, "High sensitivity of specific genotypic tools for detection of X4 variants in antiretroviral-experienced patients suitable to be treated with CCR5 antagonists", AntimicrobChemother, vol. 65, pp. 1486–1492, 2010.

[23] W. O. F. A. H. (OIE), "International Standards on Antimicrobial Resistance", 2008.

[24] Schweiker KL, Makhatadze GI, "A computational approach for the rational design of stable proteins and enzymes: optimization of surface charge-charge interactions", Methods Enzymol, vol. 454, pp. 175-211, 2009.

[25] Jensen, MA, Li FS, van'tWout AB, *et al*, "Improved coreceptor usage prediction and genotypic monitoring of R5-to-X4 transition by motif analysis of HIV-1 env V3 loop sequences", Journal of Virology, vol. 77, pp. 13376-13388, 2003.

[26] Jensen MA, Coetzer M, van 't Wout AB, *et al*, "A reliable phenotype predictor for human immunodeficiency virus type 1 subtype C based on Envelope V3 sequences", Journal of Virology, vol. 80, pp. 4698-4704, 2006.

[27] Obermeier M, Ehret R, Berg T, *et al*, "Genotypic HIV-coreceptor tropism prediction with geno2pheno [CORECEPTOR]: differences depending on HIV-1 subtype", Journal of the International AIDS Society, vol. 15(Suppl 4), pp. 1821-1824, 2012

[28] Garrido C, Roulet V, Chueca N, *et al, "*Evaluation of eight different bioinformatics tools to predict viral tropism in different human immunodeficiency virus type 1 subtypes", J ClinMicrobiol, vol. 46(3), pp. 887-91,2008.

[29] Cosic I, "Macromolecular Bioactivity: Is It Resonant Interaction between Macromolecules?-Theory and Application", IEEE Transactions on Biomedical Engineering, vol. 41(I), pp. 1101-1114, 1994.

[30] Veljkovic V, Cosic I, Dimitrijevic B, *et al*, "Is it possible to analyze DNA and protein sequence by the method of digital signal processing", IEEE Trans Biomed Eng., vol. 32(5), pp. 337-341, 1985.

[31] Veljkovic V, Glisic S, Veljkovic N, et al, Assessment of Hepatitis C Virus protein sequences with regard to interferon/ribavirin combination therapy response in patients with HCV genotype 1b, Vaccine, vol 32(48), pp. 6569-75, 2014.

[32] Jain E, Bairoch A, Duvaud S, *et al*, "Infrastructure for the life sciences: design and implementation of the uniprot website", BMC Bioinformatics, vol. 10, pp. 136-154, 2009.

[33] Perovic VR, Muller CP, Niman HL, *et al*, "Novel Phylogenetic Algorithm to Monitor Human Tropism in Egyptian H5N1-HPAIV Reveals Evolution toward Efficient Human-to-Human Transmission", PLoS ONE, vol. 8(4) pp. e61572, 2013. doi:10.1371/journal.pone.0061572

[34] Smith SW, "The Scientist and Engineer's Guide to Digital Signal Processing", California Technical Publishing, 2002.

[35] Vandekerckhove LPR, Wensing AMJ, Kaiser R, *et al*,"European guidelines on the clinical management of HIV-1 tropism testing", Lancet Infect Dis., vol. 11, pp. 394–407, 2011.

[36] Jeffs SA, Shotton C, Balfe P, *et al*, "Truncated gp120 envelope glycoprotein of human immunodeficiency virus 1 elicits a broadly reactive neutralizing immune response", J Gen Virol., vol. 83(Pt 11), pp. 2723-2732, 2002.

[37] Miller ED, Duus KM, Townsend M, *et al,* "Human immunodeficiency virus type 1 IIIB selected for replication in vivo exhibits increased envelope glycoproteins in virions without alteration in co-receptor usage: Separation of in vivo replication from macrophage tropism", Journal of Virology, vol. 75(18), pp. 8498-8506, 2001.

[38] Collman R, Balliet JW, Gregory SA, *et al,* "An infectious molecular clone of an unusual macrophage-tropic and highly cytopathic strain of human immunodeficiency virus type 1", Journal of Virology, vol. 66(12), pp. 7517-7521, 1992.

[39] Doliana R, Veljkovic V, "Emilins interact with anthrax protective antigen and inhibit toxin action in vitro",Matrix Biology, vol. 27, pp. 96-106, 2008.