

Effect of HIV-1 Subtype C integrase mutations implied using molecular modeling and docking data

Jaiprasath Sachithanandham¹, Karnati Konda Reddy², King Solomon¹, Shoba David¹, Sanjeev Kumar Singh², Veena Vadhini Ramalingam¹, Susanne Alexander Pulimood³, Ooriyapadickal Cherian Abraham⁴, Pricilla Rupali⁴, Gopalan Sridharan⁵ & Rajesh Kannangai^{1*}

Departments of Clinical Virology¹, Dermatology³, Internal Medicine⁴, Christian Medical College, Vellore, Sri Sakthi Amma Institute of Biomedical Research Institute⁵, SNHRC Vellore and Computer-Aided Drug Design and Molecular Modeling Lab², Department of Bioinformatics², Alagappa University, Karaikudi, Tamil Nadu, India. Rajesh Kannangai- Email: kannangair@cmcvellore.ac.in; Phone: 91 (416) 2282070; Fax #; 91 (416) 2232035; *Corresponding author

Received February 4, 2016; Revised February 29, 2016; Accepted March 2, 2016; Published June 15, 2016

Abstract:

The degree of sequence variation in HIV-1 integrase genes among infected patients and their impact on clinical response to Anti retroviral therapy (ART) is of interest. Therefore, we collected plasma samples from 161 HIV-1 infected individuals for subsequent integrase gene amplification (1087 bp). Thus, 102 complete integrase gene sequences identified as HIV-1 subtype-C was assembled. This sequence data was further used for sequence analysis and multiple sequence alignment (MSA) to assess position specific frequency of mutations within *pol* gene among infected individuals. We also used biophysical geometric optimization technique based molecular modeling and docking (Schrodinger suite) methods to infer differential function caused by position specific sequence mutations towards improved inhibitor selection. We thus identified accessory mutations (usually reduce susceptibility) leading to the resistance of some known integrase inhibitors in 14% of sequences in this data set. The Stanford HIV-1 drug resistance database provided complementary information on integrase resistance mutations to deduce molecular basis for such observation. Modeling and docking analysis show reduced binding by mutants for known compounds. The predicted binding values further reduced for models with combination of mutations among subtype C clinical strains. Thus, the molecular basis implied for the consequence of mutations in different variants of integrase genes of HIV-1 subtype C clinical strains from South India is reported. This data finds utility in the design, modification and development of a representative yet an improved inhibitor for HIV-1 integrase.

Key words: HIV-1, integrase, raltegravir, elvitegravir, dolutegravir, modeling, docking and binding energy

Background:

The enzyme integrase (IN) is a key retroviral enzyme that plays a significant role in the replication of the virus in its susceptible host [1]. The HIV-1 *pol* gene encodes for the IN enzyme, the reverse transcriptase enzyme (RT) and Protease (Pr). In HIV-1 replication mechanism IN catalyzes the chromosomal integration of newly synthesized double-stranded DNA by strand transfer and incorporates it into the host genomic DNA [2]. The HIV-1 IN also plays a role in the stabilization of pre-integration complex (PIC) which acts as the permanent genetic

reservoir and also in the process of initiating the production of new HIV-1 virions [3]. The HIV-1 IN enzyme is a 32 kilo Dalton protein and is composed of 288 amino acids. This gene is also an important target for molecular diagnosis in HIV-1 infected individuals because the region is highly conserved compared to reverse transcriptase and protease.

The HIV IN enzyme comprises three structurally and functionally distinct domains the N terminal domain (residues 1-50), the catalytic core domain (residues 51-212) and the C

terminal domain (residues 213-288) contains an SH3 like domain binds non specifically to DNA. The IN active site is located in the catalytic domain which is composed of 2 Asp residues and 1 Glu residue in the conserved DDE motif which are required for catalysis. The conserved DDE motif coordinates two divalent Mg^{2+} ions essential for the catalysis of integration [4, 5]. The critical step during retroviral life cycle is the integration of the viral double stranded DNA into the host chromosome [6]. The HIV-1 integrase enzyme plays a crucial role by removing a dinucleotide next to the conserved cytosine-adenine sequence from each 3' end of the viral DNA. Subsequently then integrase catalyzes by joining of the processed viral 3' end to 5' end of strand breaks in the host DNA [6].

Currently there are three regimens named Raltegravir (RAL), Elvitegravir (EVG) and Dolutegravir (DTG) which can be used as IN inhibitors and are approved by the US Food and Drug Administration. Regimen including the IN inhibitors can even be used among HIV infected individuals who are treatment naïve as per guide lines [7]. These drugs can halt the HIV-1 viral replication and significantly reduce the disease progression to AIDS. The IN inhibitors can be very useful in countries where the numbers of individuals infected with HIV are very high such as South Africa, Nigeria and India [8, 9]. The amino acid substitutions reflect resistance to RAL in three distinct genetic pathways. The two major resistance pathways are Q148HRQ/G140S and N155H/E92Q and third less frequent pathway identified as Y143CRH/T97A specifically described for raltegravir and specific resistance mutations [10]. The only described resistance pathways for EVG was T66I and S147G. The DTG has a resistance profile markedly distinct from those of RAL and EVG. The non polymorphic residues such as S153YF, R263K and G118R, have been described as being selected by DTG [10].

The molecular informatics and computer aided applications have contributed for best understanding of the mechanism of integration, drug binding and resistance development [11]. The computer based methods are becoming increasingly important and complementary to *in vitro* assays in studying the structure and function of bio molecules. Molecular docking is a frequently used tool in analyzing the structure based drug design. The most frequently used method and an effective way to predict the binding of substrate with its receptor is docking simulation, which is successfully implemented in many applications including the understanding of drug resistance [12, 13]. The study reported here was intended to assess the frequency of mutations in the IN region of HIV-1 among infected treatment naïve (IN inhibitor) population and also to predict the response to treatment of strains with mutations by minimal molecular modeling and docking studies.

Methodology:

The study was carried out at the department of Clinical Virology of a tertiary care hospital in South India with the approval from its Institutional Review Board. The analysis was carried out in a university attached Computer-Aided Drug

Design and Molecular Modeling Lab in South India. Blood samples were collected from HIV-1 infected individuals who had come to the department Clinical Virology for CD4+ T- cell estimation and/or HIV-1 drug resistance testing. The study was conducted during the years 2009-2012. A total of 128 samples collected from treatment naïve HIV infected individuals and 33 samples collected from HIV-1 infected individuals on anti retroviral therapy (ART with NRTI/NNRTI/PI) showing clinical failure. None of these 33 individuals were on IN inhibitors. All study participants were enrolled for the study only after getting the informed consent.

Sample collection

Eight ml of whole blood samples were collected in a sterile K2 EDTA vacutainer (Beckton Dickinson, New Jersey, USA) from study subjects who were not on any IN inhibitors. Fresh whole blood was used for CD4 estimation and in parallel plasma samples are separated as required and made in to multiple aliquots and stored at $-70^{\circ}C$ until testing. The CD4 + T cell count estimation for all these patients was performed using the standard procedure by Guava Technologies with Easy CD4™ and cytosoft™ software version 2.2 (Guava Technologies, California, USA) or BD FACS Count system with FACS Count CD4/CD4 SW, Version 1.0 as described earlier [14].

HIV-1 integrase gene amplification

The plasma samples were used to extract HIV-1 RNA using QIAamp® viral RNA extraction (Qiagen GmbH, Hilden, Germany) reagents as per the manufacturer's instruction. The amplification of HIV-1 IN gene was carried out by an in-house nested PCR assay using 20µl of extracted RNA with reagents from Qiagen one step RT PCR assay (Hilden, Germany) which combines both conversion of RNA to cDNA and amplification of IN using specific primers. The 1st round of PCR amplification was done using 20 picomoles of forward GCAGGATTCGGGATTAGAAG and reverse primers CTTTCTCCTGTATTCAGACC. The thermal cycling conditions used were as follows; 50°C for 30 minutes, 95°C for 15 min for activation of the Hot start Taq enzyme followed by 40 cycles of 94°C for 30 sec, 52°C for 30 sec and 68°C for 1 minute followed by a final extension at 68°C for 7 minutes. The second round PCR was carried out using 2 µL of first round DNA input to a PCR mix containing 2.5 units of Hot star Taq master mix (Qiagen, Hilden, Germany) and 20 pico moles of forward AAGGTCTATCTGGCATGGGTA and reverse primers TCCCCTAGTGGGATGTGTACTTC in a total reaction volume of 50 µL. The PCR was carried out in thermal cycler PTC-100 (MJ research, California, USA) with following thermal cycling conditions; 95°C for 15 min for activation of the Hot start Taq polymerase enzyme followed by 40 cycles of 94°C for 30 sec, 52°C for 30 sec and 72°C for 1 min followed by a final extension at 68°C for 7 min. Following amplification the PCR products were first analyzed on an ethidium bromide stained agarose (2%) gel to check for the 1087 bp specific amplicon using the gel documentation system Gel doc 2000 (Bio-Rad, California, USA) using the software Quantity one version 4.6.9.

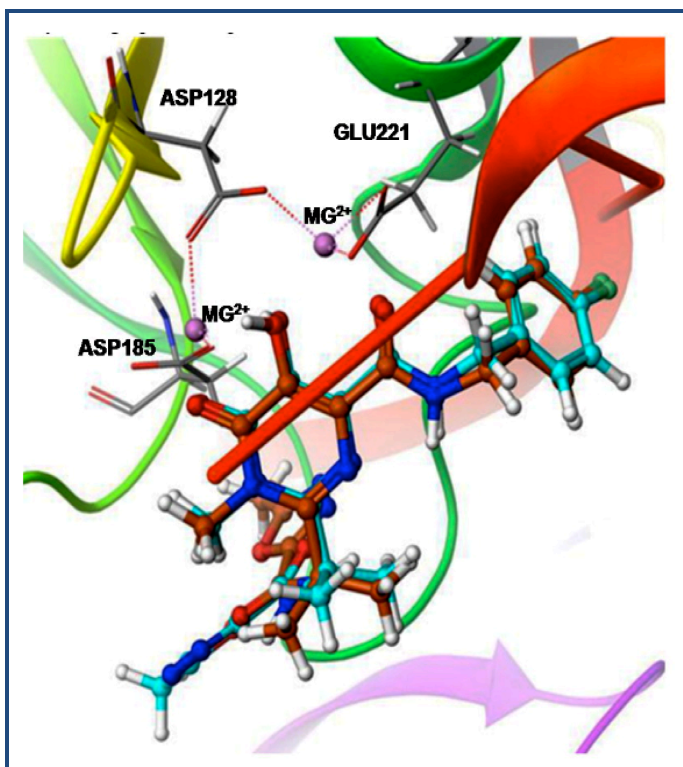


Figure 3: Showing the validation of docking protocol by superposition of Glide docked RAL and its original structure in the crystallographic complex of Prototype Foamy Virus intasome (PFV). **Legend:** Docked conformational view of RAL (C₂₀H₂₁FN₆O₅) RAL IUPAC (N-[(4-fluorophenyl) methyl]-5-hydroxy-1-methyl-2-[2-[(5-methyl-1, 3, 4-oxadiazol-2-yl) formamido] propan-2-yl]-6-oxo-1, 6-dihydropyrimidine-4-carboxamide) with PFV intasome (PDB ID: 3OYA). This was generated using software Carbon atoms in PFV are colored brown and RAL are colored in cyan. This docking protocol using PFV was only used for validation purpose for showing the accuracy of our docking method.

Sequencing of the PCR products

The PCR amplified products were subjected to pre-sequencing purification using sequencing clean up reagents (Millipore Corporation, Montage, MA, USA). The sequencing PCR was carried out with respective forward and reverse primers used for 2nd round PCR amplification. The sequencing reaction was carried out using Big Dye Terminator Cycle Sequencing assay kit (Applied Biosystems, CA, USA), subsequently run on ABI PRISM 310 genetic analyzer (Applied Bio-System, California, USA) followed by post-sequencing PCR clean up by the Millipore clean up system (Montage, MA, USA). The sequences were aligned using the Bio Edit software [www.mbio.ncsu.edu/bioedit/bioedit.html] and Finch TV software (version 1.4.0). These sequences were submitted to the Stanford HIV drug resistance database (<http://hivdb.stanford.edu>) for analysis and interpretation on a regular basis.

HIV-1 IN gene functional domain analysis

The IN study strains with amino acid sequences were aligned using the Bio Edit software. The study strains amino acid sequences were also constructed and compared to the global subtype C and the global subtype B consensus sequences [2]. The amino acid sequence alignment shown in our study strains had complete IN region (288 residues). The functional domains (F1-A49, M50-E212 and L213-D288) on the test consensus were compared with HIV-1 global C and global B subtypes, in order to determine conserved and substituted amino acid residues. The mean genetic distances among the amino acid sequences were determined using Kimura 2-parameter model [2].

HIV Sub typing and phylogenetic analysis

All the samples were subjected to HIV-1 sub typing using the REGA HIV-1 Sub typing Tool - version 2.0 in HIV drug resistance Stanford data base (<http://hivdb.stanford.edu>). The HIV-1 sub typing was also determined by phylogenetic tree analysis using the nucleotide sequences of the study strains aligned using clustal W with reference strains of HIV-1 subtypes (A-D, F-H, J and K) taken from the previously published literature [2]. The tree was generated by the neighbor joining method using 1000 bootstrap replicates in MEGA4 software. All the IN sequences generated from this study was submitted to Gen Bank and accession numbers obtained for the study samples [KF870715-KF870774, KF870776-KF870797, KF870801, KF870805-KF870813, KF870816, KF870822-KF870826, KF870829 -KF870832].

Criteria used for construction of IN 3D modeling

The consensus amino acid sequence generated from the total study samples (n=102) was utilized for the 3D modeling. The generated consensus amino acid sequence had not shown any major or minor mutations and hence was considered as wild type for 3D model. The same amino acid sequence was further incorporated with appropriate mutations in single or in combinations at appropriate positions while generating the mutation models. The model was built by introducing 3 individual accessory mutations (L74M, T97A and G163R), two major mutations (Q148H, N155H) also combination of accessory (L74M, T97A and G163R) mutations and accessory with major mutations (L74M, T97A, G163R and Q148H), (L74M, T97A, G163R and N155H).

Model selection

The full length structure of HIV-1 IN is not available till date. The prototype foamy virus (PFV) (PDB ID: 3OYA) is the surrogate model used in many studies to analyze the binding affinity of several strand transfer inhibitors including RAL [15-17]. Though the PFV IN has structural and functional homology with HIV IN with virtually identical active sites but the overall HIV-1 IN specific sequence similarity is very limited [4]. Our aim of the study was to generate HIV-1 IN sequence specific clinical mutation models for this reason we had not used PFV (PDB ID: 3OYA). The 1WKN was chosen because it contains sequence similarity of full length HIV-1 IN and also the presence in its active site of both two important Mg²⁺ ions

opportunistically chelated by the D64, D116, and E152 (Conserved DDE motif) required for catalysis [18]. This 1WKN model lack the correct fitting of DNA. Hence, we chose IN from 1WKN and we superimposed to fit the DNA from PFV, which is surrogate model of HIV-1 IN. Though 1WKN model was built in 2004 but it was refined by the same group very recently and also this model was used in many recent studies [18, 19-21, 10]. Though there are surrogate reference model like PFV intasome for Integrase we used 1WKN for our docking studies. The 1WKN is not considered as superior as PFV intasome but we attempted to use because it contains sequence similarity of full length HIV-1 IN. Our main aim was to build models based on the HIV-1 integrase sequence specific data obtained from clinical samples. So we used this model (1WKN) as reference since it had full length HIV-1 IN sequence similarity.

Building HIV-1 IN-DNA models

Homology models of full length HIV-1 IN-DNA complex were created based on the two PDB structures (PDB ID: 1WKN and 3OYA). The full length IN model was generated using the template from 1WKN using Modeller 9v7 [22]. DNA and two Mg²⁺ were taken from PDB ID: 3OYA and aligned over homology model of IN using PyMol [23].

Protein model preparation

The wild and mutant HIV-1 IN and viral DNA complexes were prepared by a multi-step process through Schrödinger Protein Preparation Wizard [24]. The right bond orders as well as charges and atom types were assigned, and hydrogen atoms were added to the structures. Each model was subjected to energy minimization using the Optimized Potentials for Liquid Simulations (OPLS)-2005 force field with implicit solvation. Each model was minimized and the minimization was terminated when root mean square deviation (RMSD) of heavy atoms in the minimized structure relative to the model structure exceeded 0.3 Å. We prepared one wild type, five single mutants (L74M, T97A, G163, Q148H, N155H), combination of accessory mutations (L74M, T97A and G163R), and combination of major with accessory mutations (L74M, T97A, G163R, Q148H), (L74M, T97A, G163R and N155H).

Ligand preparation

The HIV-1 IN inhibitors RAL, EVG and DTG were prepared using the LigPrep [25]. First all the hydrogen atoms were added to the ligand molecules as they had implicit hydrogen atoms. The bond orders of these ligands were fixed. The ionization states of the ligands were generated in the pH range of 5.0-9.0 using Epik [26]. Most probable tautomers and all possible stereo isomers were generated to study the activity of individual stereotypes of each ligand. In the final stage of Lig prep, compounds were minimized with OPLS-2005 force field.

Molecular Docking

Molecular docking was analyzed to study the relative estimates of interaction between IN molecule of the clinical isolate having different mutations and with the IN therapeutic drugs. The 3D modeling of HIV-1 IN mutation that can lead to potential low level drug resistance with accessory mutations (L74M, T97A

and G163R) and high level drug resistance with major mutation (Q148H, N155H) were chosen for docking analysis. We performed standard precision (SP) docking using the program Glide and semi-flexible docking protocols [27, 28]. The ligands being docked were kept flexible, in order to explore an arbitrary number of torsional degrees of freedom spanned by the translational and rotational parameters. The final energy evaluation is performed with Glide Score. The Glide Score was used as a function of fitness and best-fit pose for a given ligand was determined by the E_{model} score. The choice of the best docked structure for each ligand was made using model energy score E_{model} that combines Glide score, the non bonded interaction energy and the excess internal energy of the generated ligand conformation. The IN inhibitors like RAL, EVG and DTG were chosen to investigate their binding mode within the active site of modeled structure. Glide SP was carried out in each modeled structure.

Structural validation

The structural validation was done for all the generated HIV-1 IN models using the web server based Ramachandran plot (<http://mordred.bioc.cam.ac.uk/~rapper/rampage.php>). The Ramachandran plot shows the phi (φ) psi (ψ) torsion angles for all residues except glycine and proline in the structure.

Results:

A total of 143 samples of 161 tested were amplified and of these only 102 samples were successfully sequenced for the complete length of the IN region from *pol* protein (288 amino acid). Of the 102 sequences with complete IN amino acid residues, 81 were from ART naïve while 21 samples were from individuals who were on NRTI/NNRTI/PI drug regimen and showed clinical failure. None of them were on any IN inhibitors. Fourteen (14%) among the 102 strains showed accessory mutations that can contribute to potential low level IN inhibitor drug resistance. None of the individuals had any major mutations that can contribute drug resistance to the IN inhibitors. The mutation data profile on the HIV-1 IN drug resistance accessory mutations among the study subjects and the significance of the mutation on the drug activity is shown in **Table 1**. Among the 102 individuals the CD4 counts were available for 101 individuals, these ranged from 2- 1295 cells/μl. The CD4 counts did not show any significant difference ($p = 0.39$) between the individuals who presented with or without mutations in the IN region.

Six strains (6%) showed mutation at position 74 while 3 strains (3%) had mutation at 95th amino acid position. The remaining 1 each strain showed accessory mutations at positions 97, 138, 157, 163 and 230. All the 102 sequences from the clinical samples were identified as HIV-1 subtype C according to REGA HIV-1 Sub typing Tool. Based on the phylogenetic tree analysis all the clinical strains were also identified as subtype C and this data was shown in **Figure 1**. The amino acid residues of these study samples were aligned with global subtype B and C consensus of HIV-1 (residues 1-288) using Bio Edit software and this comparison is shown in **Figure 2**. The genetic variability of the IN gene among the study strains was very low

as shown by the kimura 2-parameter analysis. The mean genetic distance observed was in the range of 0.0012 to 0.0413 (0.1 – 4%). The complete predicted study amino acid sequence alignment (up to 288 amino acid residues) showed that the consensus of the study strains was identical to the global subtype C consensus, except at 8 positions (D25E, K42Q, M50I, K110V, V111K, D182I, T218I and R269K). It differed from the global subtype B consensus at 12 positions (R14K, K42Q, V72I, F100Y, L101I, K110V, V111K, V113I, N134G, E167D, D182I and D278A).

The structural validation based on the Ramachandran plot showed a good validation score for all generated HIV-1 IN mutation models. The distribution of φ , ψ angles which describes the rotations of the polypeptide backbone around the bonds between N-C α (Phi, φ) and C α -C (Psi, ψ) showed a range

from 88.4- 90.5% residues in the most favorable region, 6.7- 7.7% range of residues in allowed region and 2.8- 3.9% residues in the generous region. Overall 96.1-97.1% of the residues were within the allowed region. Molecular docking was analyzed to study the relative estimates of interaction between the clinical isolate protein sequences with mutations and the therapeutic drug. The molecular docking scores obtained for various mutation models like 3 individual accessory mutations (L74M, T97A and G163R), major mutation (Q148H, N155H) and also in combination of accessory (L74M, T97A and G163R), accessory and major mutations (L74M, T97A, G163R and Q148H), (L74M, T97A, G163R and N155H) were compared with wild type strain. The molecular docking scores using the study strains showed significant effect on the molecule binding affinity compared with wild type.

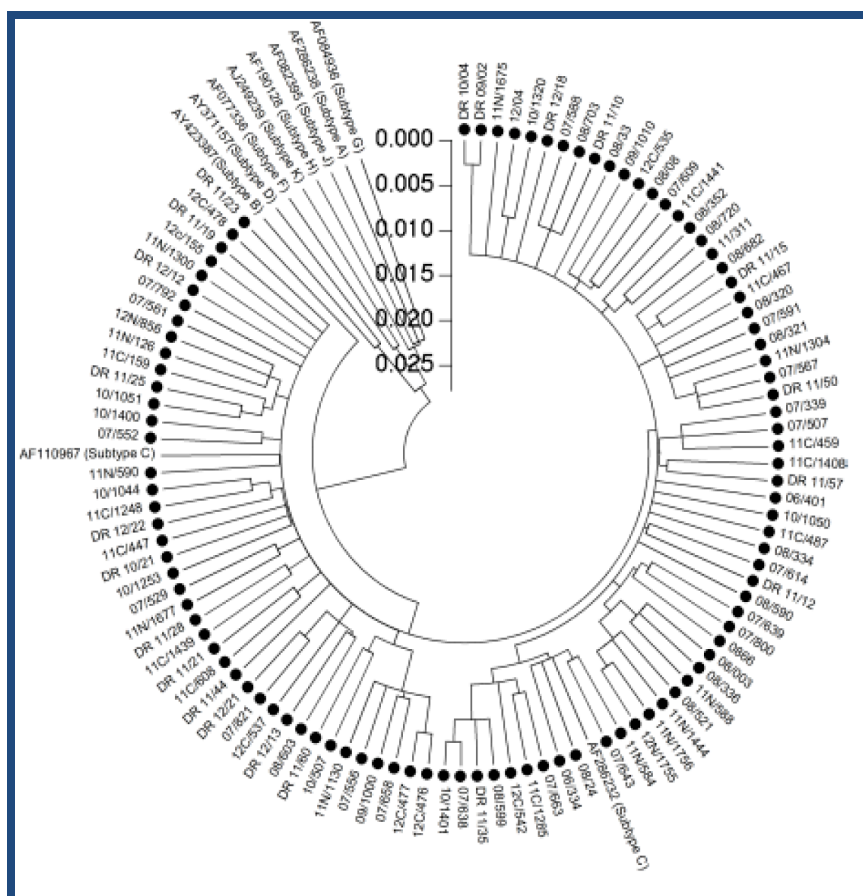


Figure 1: Phylogenetic tree (circular view) of HIV-1 integrase genes to mark the subtype in the collected clinical samples GenBank accession numbers are used to show HIV-1 subtypes (A-D, F-H, J and K). Sequences representing clinical samples used in this study are shown using dark large dots (•). The GenBank accession number AF286238 (Subtype A), AY423387 (Subtype B), AF110967 (Subtype C), AF286232 (Subtype C), AY371157 (Subtype D), AF077336 (Subtype F), AF084936 (Subtype G), AF190128 (Subtype H), AF082395 (Subtype J) and AJ249239 (Subtype K) for respective HIV-1 subtypes were taken from literature as described elsewhere [2]. The tree was generated by the neighbor joining (N-J) method with 1000 bootstrap replicates using MEGA4 software.

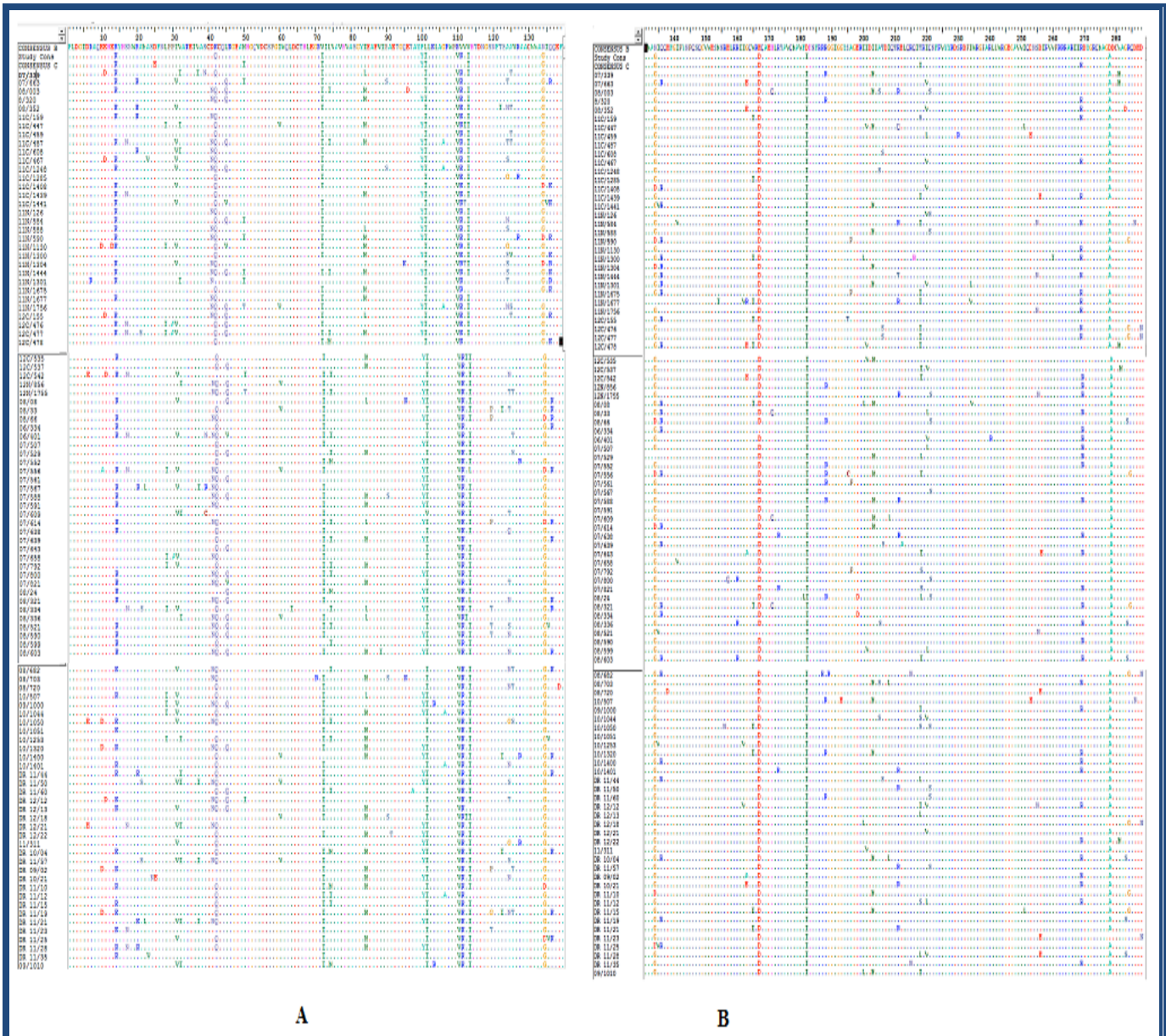


Figure 2: Multiple sequence alignment (MSA) of HIV-1 subtype C integrase protein sequences used in this study. The global subtype B and C consensus sequences were obtained from HIVsequence database. MSA was drawn using the BioEdit software. Sequence positions matching the global subtype B consensus sequence are denoted by a dot (.).

The docking algorithm used in the study was validated to reproduce the co-crystallized pose of RAL in the PFV IN-DNA binding pocket. The docking study yielded a good agreement between the crystal and docked structure (RMSD 0.6 Å) shown in **Figure- 3**. This docking protocol using PFV was only used for validation purpose for showing the accuracy of our docking method. The most significant method of evaluating the accuracy of a docking procedure is to determine how closely the lowest energy poses (binding conformation) predicted by the object scoring function, Glide score. The model with single accessory mutation at positions like 74, 97, 163 had a mean

difference in binding score of -0.26 (range -0.14 to -0.39) with RAL, -0.68 (range -0.38 to -1.13) with EVG and -0.55 (range -0.22 to -0.71) with DTG docking analysis. The mean difference in the binding score for the two major mutations Q148H and N155H were -0.23, -0.62 and -0.80 respectively for RAL, EVG and DTG. However, when there were three accessory mutation co-exist the reduction in the binding scores for all 3 inhibitors showed -0.25, -0.35 and -0.22 respectively. This difference in the binding energy score was higher when there is a combination of accessory and major mutations (-2.07 to -2.53 for RAL, -1.82 to -2.41 for EVG and -1.92 to -2.98 for DTG). The predicted

binding affinity, energy of best docked compound and their interaction in active site for the individual mutation models and wild type with RAL, EVG, and DTG analogs is shown in

Table 2. The representation of 3D and 2D docking pose of HIV-1 wild type strain structure with RAL, EVG, and DTG analogs is shown in **Figure 4**.

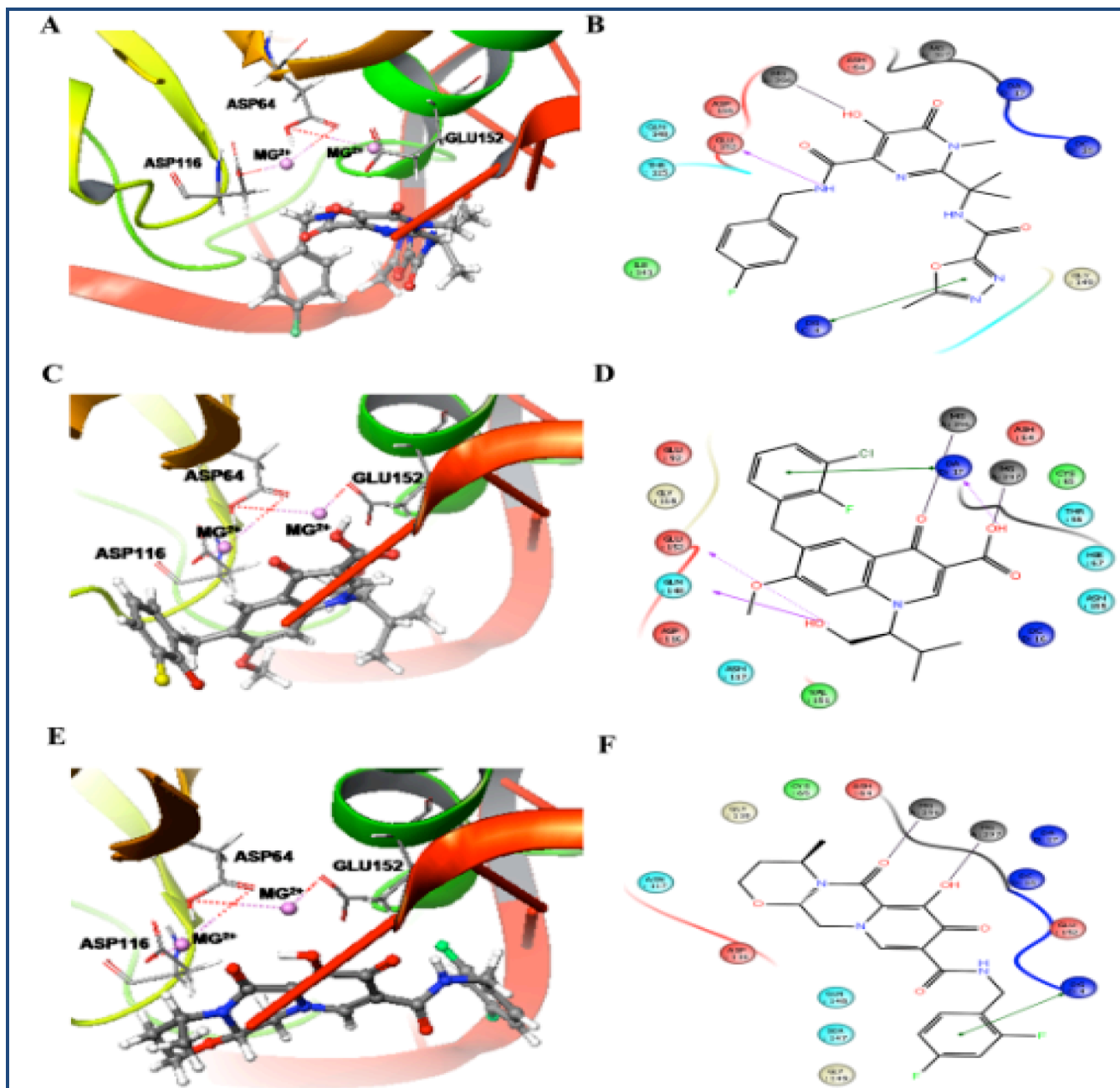


Figure 4: A representative binding mode of integrase inhibitors (Raltegravir, Elvitegravir and Dolutegravir) in the active site of modeled HIV-1 integrase structure. Legend: Raltegravir (A &B), Elvitegravir (C &D) and Dolutegravir (E &F) in the active site of modeled HIV-1 wild type IN structure. The divalent Mg^{2+} are shown as magenta spheres, three catalytic residues D64, D116 and E152 are highlighted as sticks and ligands are presented as ball and sticks. The coordinate bonds between the ligands and divalent Mg^{2+} ions are shown as dashed bonds and the 3' viral DNA are shown in orange color. Raltegravir shown interaction with catalytic amino acid residue Glu152 and also π -stacking interaction with DNA. Elvitegravir shown interaction with catalytic amino acid residue Glu152, Gln148, and π -stacking and H-bond interaction with DNA. Dolutegravir shown π -stacking interaction with DNA. All three compounds shown coordination bonds with Mg^{2+} .

Table 1: Frequency of accessory mutations in the integrase sequences among HIV-1 subtype C infected individuals (n=102) with known drug resistance data.

| No | Accessory mutation | Frequency | Significance (based on the HIV drug resistance Stanford data base [44]) |
|----|--------------------|------------|---|
| 1 | L74M | 6/102 (6%) | It reduces RAL susceptibility in combination with other INI-resistance mutations. |
| 2 | Q95K | 3/102 (3%) | Non polymorphic INI resistance mutation. |
| 3 | T97A | 1/102 (1%) | Minimally polymorphic INI resistance mutation. |
| 4 | E138D | 1/102 (1%) | Unusual mutation in non-polymorphic sites. |
| 5 | E157Q | 1/102 (1%) | Minimally polymorphic mutation reduces RAL and EVG susceptibility. |
| 6 | G163R | 1/102 (1%) | Non polymorphic INI resistance mutation it reduces RAL susceptibility in combination with other INI-resistance mutations. |
| 7 | S230R | 1/102 (1%) | Non polymorphic INI resistance mutation with minimal effect on INI susceptibility |

RAL; raltegravir, INI; integrase, EVG; elvitegravir

Table 2: Molecular docking analysis of known HIV-1 integrase mutants with Raltegravir, Elvitegravir and Dolutegravir analogs

| Model | Raltegravir | | | Elvitegravir | | | Dolutegravir | | |
|-------------------|-------------|----------|--------------|--------------|----------|--------------|--------------|----------|--------------|
| | glide score | *E model | glide energy | glide score | *E model | glide energy | glide score | *E model | glide energy |
| WT | -4.6 | -64.1 | -52.0 | -5.6 | -66.1 | -47.8 | -5.6 | -61.0 | -45.8 |
| L74M | -4.4 | -62.4 | -50.3 | -5.2 | -60.4 | -43.4 | -4.9 | -59.6 | -48.6 |
| T97A | -4.5 | -61.5 | -50.4 | -5.0 | -46.1 | -37.6 | -4.9 | -59.8 | -48.6 |
| G163R | -4.2 | -62.6 | -51.3 | -4.4 | -57.8 | -45.2 | -5.3 | -58.9 | -47.6 |
| **3 minor | -4.4 | -62.3 | -50.3 | -5.2 | -46.0 | -37.3 | -5.3 | -58.0 | -46.6 |
| Q148H | -4.3 | -59.3 | -48.7 | -4.7 | -56.4 | -46.3 | -4.6 | -56.1 | -47.0 |
| N155H | -4.6 | -63.6 | -51.1 | -5.2 | -47.3 | -38.2 | -4.9 | -59.4 | -48.2 |
| 3 minor and Q148H | -2.5 | -39.4 | -40.6 | -3.7 | -40.6 | -39.1 | -3.6 | -43.8 | -38.3 |
| 3minor and N155H | -2.1 | -34.2 | -37.5 | -3.1 | -36.0 | -34.7 | -2.6 | -46.7 | -41.7 |

*Emodel is a specific combination of Glide Score, the non-bonded interaction energy between the ligand and the receptor and the internal torsion energy of the ligand conformer, expressed in kcal/mol. Minor mutations - L74M, T97A, G163R, Major mutations - Q148H, N155H, **3 minor - combination of L74M, T97A, G163R; The difference in the binding energy score was higher when there is a combination of accessory and major mutations.

Discussion:

In HIV-1 infected individuals the antiretroviral therapy with integrase inhibitors (INIs) has become a major breakthrough in clinical management especially for those in whom resistance to first line and second line of treatment develop. The IN inhibitors, in particular RAL with its potential of high antiviral activity, better adherence and good tolerability has significantly reduced the virological failure among the HIV-1 infected individuals [29]. Clinical trials with other two FDI approved INIs also shown to be very promising and the single dose combination with NRTI has made INI based regimen successful with less adverse effects [30]. The major mutations on IN inhibitors were not reported in many studies on treatment naïve populations. However, the studies have reported very low IN accessory mutations irrespective of HIV-1 subtypes [2, 31-34]. Many of the studies showed the rate of integrase polymorphism was variable at various positions from various populations from different parts of the world [32-34]. However these studies showed the significance of subtype-specific polymorphisms in both ARV-naïve and experienced individuals. [33-34]

Our study where in HIV-1 subtype C strains were analyzed and did not demonstrate any major mutations while 14 (14%) individuals had accessory IN mutations at 7 positions of which majority were L74M. This was similar to the findings from a Brazilian study where 8% of the study population had accessory mutations like V72I and V201I among treatment naïve individuals among the HIV-1 B, C, and F subtypes [35]. Our study also demonstrated these two polymorphisms, V72I and V201I reported in the Brazilian study. Studies from Europe and USA also reported absence of major IN mutations and presence of accessory mutations among treatment naïve subjects infected with HIV-1 B and non- B subtypes [36, 37]. Compared to the European study our frequency of accessory mutations were less but were higher than the one reported from USA [37]. Among the RAL treated individuals the major IN resistance mutations reported were G140S, Q148H and N155H, V151I, E92EQ, V151I, G163R [1]. These major mutations were not present in any of our study strains.

The reported genetic variability of IN gene based on the mean genetic distance among HIV-1 subtype C from south Africa varied from 0.0124 to 0.1004 [2]. Compare to this our subtype C

study strains had a lower genetic variability of the IN gene with a mean genetic distance range of 0.0012 to 0.0413 using the kimura 2-parameter method. Since the mean genetic distance was low the cross contamination between sequences in the data set were also verified and ruled out.

A study from reported Canada had analysed the structural modeling of R263K mutation to show a change in 3' processing and strand transfer activities with viral LTR DNA and target DNA compared to the wild type. Based on their modeling studies they reported that IN containing R263K mutation against DTG showed a slight decrease in 3' processing and strand transfer activities compared to the wild type. They also reported based on their structural modeling analysis by *in vitro* IN-DNA binding assays that R263K mutation can affect the IN DNA interactions [37]. Another study analyzed the molecular dynamics approach to estimate the binding energy of HIV-1 IN inhibitors (including RAL and EVG) and correlated with *in vitro* activity [38]. They reported that RAL and EVG interaction energy values closely matched with the experimental binding energy [35]. In one of the structural analysis it was found that two mutations like G118R and E138K lead to the introduction of a positively charged arginine at position 118, near the catalytic amino acid 116, this might decrease Mg (2+) binding, compromising the enzyme function and which can lead to significant reduction in viral integration and replication capacity [39].

In yet another study difference in the susceptibility of IN strand transfer inhibitors (RAL and EVG) among HIV-1 Subtypes B and CRF02_AG were compared by looking at the binding affinity. The study showed that the sequence variations between the INs of CRF02_AG and B strains did not lead to any notable difference in the structural features of the enzyme and did not impact the susceptibility to the IN inhibitors [40]. Studies had also evaluated various IN inhibitors including RAL and EVG impact on the specific mutations associated with IN resistance [41, 11]. Those investigators had stated that distinct binding energy difference was observed between different classes of IN inhibitors which also correlated with resistance patterns [11].

The phenotypic analysis of clinically derived IN for G118R and F121Y was analysed in a study which showed high resistance to all 3 inhibitors with a fold change >100. This finding was also supported by their results which showed G118R and F121Y was associated with reduced binding affinities to each of the inhibitors and with a decreased number of hydrogen bonds compared with the wild-type complexes [10]. Marsden *et al* 2011 in their experimental analysis found the presence of a single primary raltegravir resistance mutation (Q148H, Q148R, N155H, or N155S) is sufficient to provide resistance to RAL during macrophage infection. They also had reported that N155 pathway mutations typically produce lower-level resistance than Q148 pathway mutations [42]. Another study from USA had looked at the genotypic and phenotypic susceptibility to IN inhibitors using 39 clinical isolates. In their study the susceptibility to DTG and RAL was determined for RAL

resistant clinical isolates. Of 39 clinical isolates 30 isolates had genotypic and phenotypic resistance to RAL, in which the median level of phenotypic resistance to RAL was high, while the DTG containing the Y143 and N155 mutations level of resistance to was close to that of wild-type variants. But for isolates with Q148 plus additional IN mutations showed more reduced susceptibility to DTG [43]. Based on these phenotypic studies it was observed that there was reduced drug susceptibility when clinical isolates contain combination of mutations. When compared to these studies our docking analysis also supports these findings.

In our study we look at the drug binding affinity with mutation singly or in combination. This difference in the binding energy was observed for RAL, EVG and DTG analogs. The binding energy was decreased with either a single accessory or major mutation. But in combinations of accessory mutations as well as in combination of major mutations with accessory mutations the binding energy has been significantly reduced. Our findings with the docking experiments is well correlating with HIV stanford drug resistance data base findings as it suggests these accessory mutations cause low level resistance when presented individually (<http://hivdb.stanford.edu>). But when presented together they cause high level resistance against IN inhibitors among infected individuals.

Conclusion:

This study reports the analysis and interpretation of HIV-1 subtype C integrase sequences with identified accessory mutations among infected individuals from South India for the first time. This data aided with molecular models and docked inhibitor conformers finds utility in understanding the functional consequences of mutations among integrase variants for predicting drug susceptibility and or resistance. An accessory mutation in addition to other major mutations in HIV-1 subtype C integrase variants is a challenge for inferring inhibitor resistance in HIV-1 infected individuals in South India when combinational regimens are scheduled.

Ethical clearance:

It should be noted that ethical clearance board of the institute has reviewed the use of samples from human subjects for this study and appropriate approval was given in this regard.

Conflict of Interest:

The authors declare no conflicts of interests.

Acknowledgment:

We gratefully acknowledge the funding received from Indian Council of Medical Research (ICMR) [Ref no Immuno/18/11/15/2008-ECD-I] towards this study.

References:

- [1] Parczewski M *et al. BMC Infect Dis.* 2012 12: 368 [PMID: 23259737]
- [2] Bessong PO & Nwobegahay J. *Int J Mol Sci.* 2013 14: 5013 [PMID: 23455469]

- [3] Ceccherini-Silberstein F *et al.* *J Antimicrob Chemother.* 2010 **65**: 2305 [PMID: 20817922]
- [4] Reddy KK & Singh SK. *Chem Biol Interact.* 2014 **218**: 71 [PMID: 24792351]
- [5] Hu G *et al.* *J Mol Model.* 2012 **18**: 4995 [PMID: 22733274]
- [6] Gupta P *et al.* *Chem Biol Drug Des.* 2012 **79**: 835 [PMID: 22233531]
- [7] Hicks CB, *J Watch AIDS Clin Care.* 2013 **25**: 89 [PMID: 26204591]
- [8] Dubey S *et al.* *Eur J Med Chem.* 2007 **2**: 1159 [PMID: 17367896]
- [9] Squires KE *et al.* *AIDS Res Hum Retroviruses.* 2013 **29**: 859 [PMID: 23351187]
- [10] Malet I *et al.* *J Antimicrob Chemother.* 2014 **69**: 2118 [PMID: 24710029]
- [11] Loizidou EZ *et al.* *Bio org Med Chem.* 2009 **17**: 4806 [PMID: 19450984]
- [12] Li L *et al.* *J Chem Inf Model.* 2012 **52**: 2715 [PMID: 22998323]
- [13] Sun H *et al.* *J Chem Inf Model.* 2013. **53**: 2376 [PMID: 23952683]
- [14] Kandathil AJ *et al.* *Clin Diagn Lab Immunol.* 2005 **12**: 1006 [PMID: 16085920]
- [15] Hare S *et al.* *Proc Natl Acad Sci U S A.* 2010. **107**: 20057 [PMID: 21030679]
- [16] Hu JP *et al.* *J Biomol Struct Dyn.* 2013 **31**: 734 [PMID: 22913375]
- [17] De Luca L *et al.* *Eur J Med Chem.* 2011 **46**: 756 [PMID: 21227550]
- [18] De Luca L *et al.* *Biochem Biophys Res Commun.* 2005 **336**: 1010 [PMID: 16165087]
- [19] Ferro S *et al.* *J Med Chem.* 2009 **52**: 569 [PMID: 19090766]
- [20] Barreca ML *et al.* *Chem Med Chem.* 2009 **4**: 1446 [PMID: 19544345]
- [21] Majerz-Maniecka K *et al.* *Bio org Med Chem.* 2011 **19**: 1606 [PMID: 21316973]
- [22] Eswar N *et al.* *Curr Protoc Bioinformatic.* doi: 10.1002/0471250953.bi0506s15.
- [23] <https://www.pymol.org/citing>
- [24] <http://www.schrodinger.com>
- [25] <http://www.schrodinger.com/citations/41/10/1/>
- [26] <http://www.schrodinger.com/citations/41/4/1/>
- [27] <http://www.schrodinger.com/citations/41/5/1/>
- [28] Friesner RA *et al.* *J Med Chem.* 2004 **47**: 1739 [PMID: 15027865]
- [29] Marianne H, *Therapeutics.* 2009 **1**: 183
- [30] Raffi F & Wainberg MA, *Retrovirology.* 2012 **9**: 110 [PMID: 23253887]
- [31] Passaes CB *et al.* *J Acquir Immune Defic Syndr.* 2009 **51**: 7 [PMID: 19262402]
- [32] Mulu A *et al.* *J Transl Med.* 2015 **1**: 13 [PMID: 26626277]
- [33] Meixenberger K *et al.* *J Int AIDS Soc.* 2014 **2**: 17 [PMID: 25397491]
- [34] van Hal SJ *et al.* *Retrovirology.* 2009 **6**: 12 [PMID: 19203393]
- [35] Garrido C *et al.* *J Anti microb Chemother.* 2010 **65**: 320 [PMID: 20007331]
- [36] Varghese V *et al.* *AIDS Res Hum Retroviruses.* 2010 **26**: 1323 [PMID: 20961278]
- [37] Quashie PK *et al.* *J Virol.* 2012 **86**: 2696 [PMID: 22205735]
- [38] Johnson BC *et al.* *Antimicrob Agents Chemother.* 2012 **56**: 411 [PMID: 22037850]
- [39] Bar-Magen T *et al.* *J Virol.* 2010 **4**: 9210 [PMID: 20610719]
- [40] Ni X *et al.* *Adv Virol.* 2012 **2012**: 548657 [PMID :22829822]
- [41] Serrao E *et al.* *Retrovirology.* 2009 **6**: 25 [PMID: 19265512]
- [42] Marsden MD *et al.* *Antimicrob Agents Chemother.* 2011 **55**: 3696 [PMID: 21628534]
- [43] Underwood MR *et al.* *J Acquir Immune Defic Syndr.* 2012 **61**: 297 [PMID: 22878423]

Edited by P Kanguane

Citation: Sachithanandham *et al.* *Bioinformation* 12(3): 221-230 (2016)

License statement: This is an Open Access article which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly credited. This is distributed under the terms of the Creative Commons Attribution License

