Research paper

Prediction of HIV integrase resistance mutation using *in silico* approaches

Heitor Horlando Sampaio Araujo da Silva[b,c,d,*], Natalia Pereira[b,c,d], Lucas Brandão[b,d], Sergio Crovella[a,d], Ronald Moura[b,d]

[a] *Federal University of Pernambuco, Genetics Department, Recife, Brazil*
[b] *Federal University of Pernambuco, Pathology Department, Recife, Brazil*
[c] *Mauricio de Nassau University, Recife, Brazil*
[d] *Laboratory of Immunopathology Keizo Asami - UFPE, Recife, Brazil*

ARTICLE INFO

ABSTRACT

The Antiretroviral Therapy (ART) has been providing better treatment for the Human Immunodeficiency Virus 1 (HIV) infection, by reducing its viral load to undetectable levels and recovering the immune system. However, new HIV mutations could induce drug resistance to ART, increasing the viral load and disruption of immune system. One of these drugs is Dolutegravir (DTG), which inhibits HIV integrase (INT) activity.

Our objective was to predict novel HIV mutations related to DTG resistance using *in silico* approaches in order to stablish a framework of searching for new HIV drug-resistant mutations. To this end, we modelled the INT structure and produced a mutational profile to investigate hotspots that may affect INT. Being the Y226K mutation the most frequent (0.3) and with a higher $\Delta\Delta G$ ($+2.07$), we selected to test the framework. To ratify the impact of Y226K, we docked the mutant INT with the DTG and compared the results with the Wild Type (WT) with known drug-resistant mutations. Moreover, we performed molecular dynamics simulations and calculated the binding energy along the time-course. When we compared the energies of the systems, the Y226K complex showed less binding affinity ($\Delta\Delta G = 104.88$) than the other mutated complexes compared with the WT, the Y226K complex showed even less binding affinity ($\Delta\Delta G = 104.88$). This variant somehow impedes the attachment of DTG to INT, indicating this mutant as possible resistance mutation.

1. Introduction

The Antiretroviral Therapy (ART), in the last decade, has been providing a better treatment for the Human Immunodeficiency Virus 1 (HIV) infection, by reducing its viral load to undetectable levels and recovering the immune system. In fact, according to the last data from UNAIDS the mortality related to the consequences of HIV infection decreased (Gallo et al., 1983).

Nowadays, ART is based on the combination of at least 3 class of antiretroviral drugs. One of these antiretroviral drugs could be Dolutegravir (DTG), an integrase inhibitor (IN), that binds to the catalytic domain of HIV integrase enzyme (*INT*) and abolishes its activity, therefore avoiding the assembling of HIV genome into the host (Tozzi, 2010).

During the reverse transcription of HIV RNA, occasionally, new mutations occur with a rate of 5.9×10^{-4} to 5.3×10^{-5} mutations/bp/cycles (Abram et al., 2014), and somehow, increase the resistance to ART (including integrase inhibitors such as DTG) by affecting their efficacy (Krishnan et al., 2010). The appearance of novel drug-resistant HIV strains has been also related to the lack of adherence to the treatment (Michaud et al., 2012).

So, in the context of integrase inhibitors, such as DTG, the analyses of HIV integrase mutations, allows the identification of possible drug resistance, which may lead to a change of the drug that will be administered to a naïve patient, increasing the likelihood of a successful treatment (Barré-Sinoussi et al., 2013).

Rising knowledge of protein tridimensional structure allows better understanding of its activity, the structure-function relationship, the interaction with other molecules and contributes for a better and more detailed comprehension of biological processes.

All this considered, we performed as *in silico* study aimed at predicting the influence HIV Integrase mutations on the binding process of Dolutegravir, with the objective of creating a rationale for the use or not of this drug in patients carrying different HIV integrase mutation, thus ameliorating the follow-up and patients' quality of life.

2. Methods

2.1. Protein modeling and validation

The *INT* structure was modelled with the SwissModel server, using the homology method. The input INT sequence was the HXB2 reference sequence of HIV (Supplementary material 1). The server uses 5 steps to build the protein.

The input data, which is the FASTA protein sequence, the template search with the provided data in step 1, the server performs a BLAST and HHblits, which adds sensitivity in case of remote homology.

The template selection is based on Global Model Quality Estimate (QMQE) and Quaternary Structure Quality Estimate (QSQE).

The model building, for each selected template, a 3D protein model is automatically generated by first transferring conserved atom coordinates as defined by the target-template alignment and finally the model quality estimation, is to quantify the modeling errors and give estimates on expected model accuracy, the SwissModel relies on the QMEAN scoring function (Guex and Peitsch, 1997). As input to modelled we used the HXB2 reference protein sequence of HIV present on Los Alamos database. Based on the sequence coverage and the QMEAN score we selected CryoEM Tetramer Integrase as the best template (PDB-ID:5U1C). The structure refinement process was made with GalaxyWEB-Refinement (Ko et al., 2012) server and validated using the following algorithms: PROCHECK, Rampage, Qmean server, ProSA server, ERRAT and Verify3D. Dolutegravir (DTG) was obtained from ZINC database. The PROCHECK server, checks the stereochemical quality of a protein structure, producing plots analyzing its overall and residue-by-residue geometry; rampage server performs a geometry calculation based on Ramachandran plot, which relies on visualization of the dihedral angles Phi (Φ) and Psi (ψ) of amino acid residues in protein structure. The others algorithms is based on stereochemical methods to assign the protein structure quality (Benkert et al., 2009; Laskowski et al., 1993; Lovell et al., 2003; Wiederstein and Sippl, 2007).

2.2. Docking simulation

Docking simulations was carried out to validate the structure of the wt-*INT* and the DTG as ligand. All simulations were submitted to Vina Autodock. The Autodock Vina performs a flexible docking simulation, searching 40,000 times in the protein the best binding site for the ligand; this is a aleatory method (Trott and Olson, 2010) After the molecular docking, the residues that interact with DTG were observed with LigPlot+ software (Laskowski and Swindells, 2011).

2.3. Point mutation scanning

Once established the binding sites between wt-*INT* and DTG, we predicted the point mutations of all integrase residues using the MAESTROweb server;, this method calculates the energy variation by changing the aa in the residue (Laimer et al., 2016) to understand the impact of the mutation depending on ΔΔG. Based on this concept we have found the novel possible resistance mutations. Then we have filtered this mutation list with the two known high resistance mutation N155H and R263K, these mutations are reported in the Stanford HIV resistance database and are widely studied (Anstett et al., 2015; Liu and Shafer, 2006). The both (N155H and R263K) mutations was applied on the structure using the RosettaBackrub Point Mutation web server (Lauck et al., 2010), this was the best method to mutate the structure because the tool made an local refinement on the mutation site to eliminate wrong atoms positions.

2.4. Allele frequencies estimative

To calculate the allele frequency of these mutations 1113 aligned sequences corresponding to the INT region were downloaded from the HIV Los Alamos database. All these sequences were from treatment naïve patients worldwide retrieved until 2011. R Software (version 3.1.3) was used to create an in-house script to determine their frequency.

2.5. Molecular dynamics simulation

Molecular dynamics was performed with GROMACS software (version 2016.2) to determinate the wt-INT-DTG;N155H-INT-DTG;R263K-INT-DTG structure stability. The simulation force field used for all simulations was the GROMOS/53A6. The tridimensional structures were solvated into a cubic box with SPC/E water molecules, and neutralized by adding Na + Cl- ions. The steepest descent method for Energy minimization was used with 50,000 steps. The equilibration of the system was made in two phases: first, NVT (constant number of particles, volume and temperature) equilibration, with constant temperature (300*K*) for 100 ps; second, an NPT (constant number of particles, pressure and temperature) equilibration with constant pressure of 1 bar and constant temperature of 300 K also for 100 ps. Finally, after minimizing energy and equilibrating the system, the production phase was carried out at 310 K for 10 ns. The covalent bonds were constrained using the LINCS (Linear Constraint Solver) algorithm whereas the electrostatic interactions were assessed through the PME (Particle Mesh Ewald) method. The MD trajectories were recorded every 10 ps and the production time was 40 ns.

2.6. MM/PBSA simulation

MMPBSA was used to understand the variation Gibbs binding free energies, with this simulation we could understand if during the simulation the ligation with INT-DTG was stable or not. This analysis was performed with the *g mmpbsa* tool (Kumari et al., 2014). For each of these (N155H, R263K and Y226K) mutations and the WT protein we used a scale to qualify ΔΔG as: a) destabilizing mutation, when ΔΔG > 1 Kcal/mol; b) stabilizing mutation when ΔΔG < −1 Kcal/mol; c) neutral mutations when −1 Kcal/mol < ΔΔG < 1 Kcal/mol. All the steps from the framework can be visualized on this diagram (Fig. 1).

3. Results

3.1. Modeling and validation of HIV integrase

The best and the selected model from the homology prediction of the integrase structure showed a root-mean-square deviation (RMSD) and TM-score (Template Model score) of wildtype (wt)-*INT* were 1.640 Å and 0.823, respectively. The Ramachandran plots show the phi (φ)-psi (ψ) torsion angles for every residue of a protein. The final model had 95.6% residues in the most favored regions and 4.4% in additional allowed regions on the plot generated by PROCHECK (Supplementary Material 2A). Analyzing the results obtained from RAMPAGE, the final model had 98.6% residues in favored region, 1.4% in allowed region (Supplementary Material 2B). The expected value for PROCHECK in favored regions was over than 90% and for RAMPAGE the values expected is about 98.0%.

The ERRAT plot for the final model was 83.571 overall quality factor, the good quality scored higher than 50 are considered acceptable (Supplementary Material 2C). The other analysis (Qmean server, ProSA server and Verify3D) also indicates that the structure is consistent to continue the study (Supplementary Material 3).

3.2. Mutational hotspots

The sensitivity profile of the initial structure was important to understand the weight of each mutation in promoting stabilization/destabilization of the INT. A heatmap plot was produced to rank hotspot
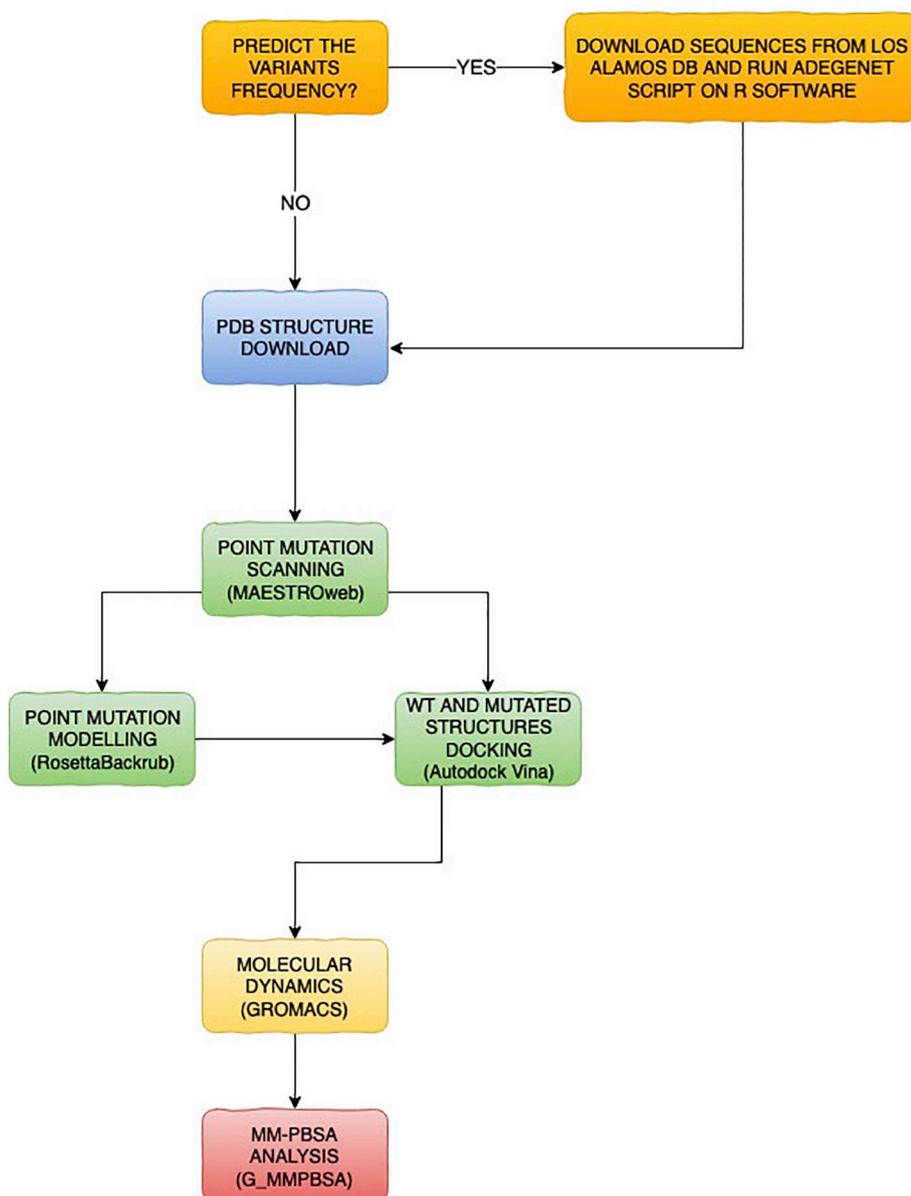
HIV-DRPRED

Fig. 1. HIV-DRpred framework diagram.

mutations destabilizing the INT. 241 mutations were retrieved assuming the ΔΔG values bellow −1.00 or above +1.00 that represent significant changes in the structure (being stabilizing/destabilizing) (Capriotti et al., 2008). To test our framework we ranked the mutations according to its frequency based on INT sequences available on Los Alamos database, through Adegenet package; the most frequent mutation (0.30) was Y226K. This variation had a dramatic impact on protein stability (ΔΔG = 2.07). Interestingly, variations known to promote drug resistance did not exceed the threshold, such as R263K (frequency = 0.26; ΔΔG = 0.32) and N155H (frequency = 0.15; ΔΔG = −0.22).

3.3. Interaction between Integrase and Dolutegravir

The results from Autodock showed that the WT-*INT*-DTG complex has a ΔG = −7.40. This value was used as reference for other docking simulations. The LigPlot showed that the residues Arg263, Arg228, Lys160, Lys46, Leu45, Gly47, Pro233 and Trp235 are interacting directly with DTG. The interaction is composed by three hydrogen bonds (Arg228, Arg263 and Gly47) and five hydrophobic interactions (Lys160, Lys46, Leu45, Pro233 and Trp235) (Fig. 2). We have docked the R263K, Y226K and N155H, but the ΔΔG did not show significance results, being the values of −0.4, 0.2 and −0.3, respectively. Also, these three mutated structures were not able to change the binding site of INT.
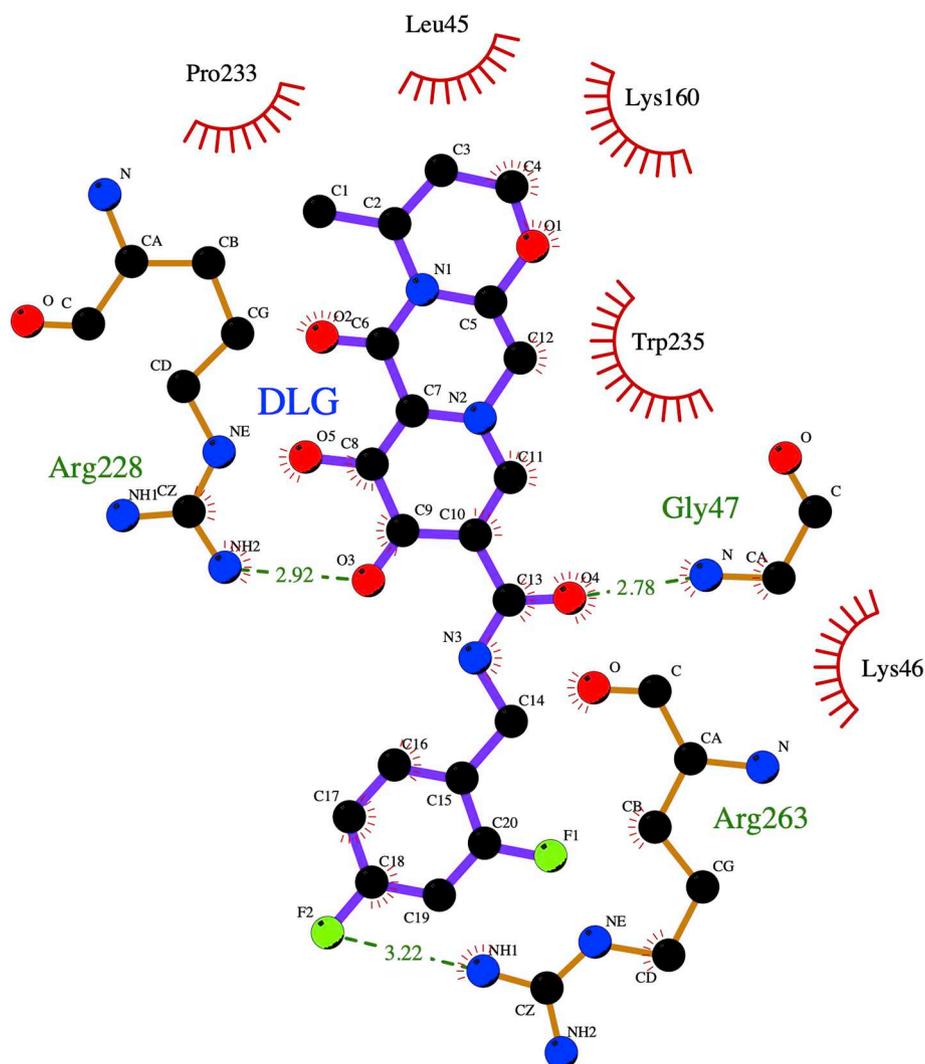
Fig. 2. Ligplot from WT-DTG. Green indicates the Hydrogen Bond and Red the hydrophobic interactions. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

3.4. Molecular dynamics

RMSD of backbone Cα atoms from the initial structures of production phase of the dynamics have been calculated and plotted in Fig. 3A. Two points along the simulation are noteworthy: at 10 ns, the Y226K complex start to increase its instability from around 0.5 nm to almost 1 nm; at 20 ns, the WT and R263K complexes increased its instability from 0.5 nm to around 1 nm, while Y226K reached back 0.5 nm of RMSD. Despite those changes, the binding of INT with DTG remained throughout the end of the simulation. The other parameters (Radius of Gyration, SASA and RMSF) demonstrated that the drug-resistance mutations, including Y226K, became more compact and less accessible solvent (Fig. 3B, C and D).

3.5. Binding stability

To get a deeper understanding of the effects of point mutations on the interactions between INT-DTG, the binding free energies and the

individual energy components were calculated employing the MM-PBSA method. The R263K and N155H variants showed a decrease in the binding affinity between the INT and DTG (ΔΔG = 66.61 and 50.74, respectively), with major decrease in the presence of R263K with respect to N155H, which is in agreement with previous experimental results.(Anstett et al., 2015) The mutation that we found, Y226K, showed an even more accentuated decrease in the binding affinity (ΔΔG = 104.88) (Table 1).

4. Discussion

The HIV literature is overflowed by several *in vitro* studies focusing on the discovery and evaluation of the presence of drug-resistant strains, mainly caused by changes in key proteins of viral life cycle. Nowadays, the main tool to categorize the drug-resistant mutation levels is the HIV Stanford database. However, this database should be fed by experimental findings, with delay occurring between the release of novel medications and the information concerning the resistance
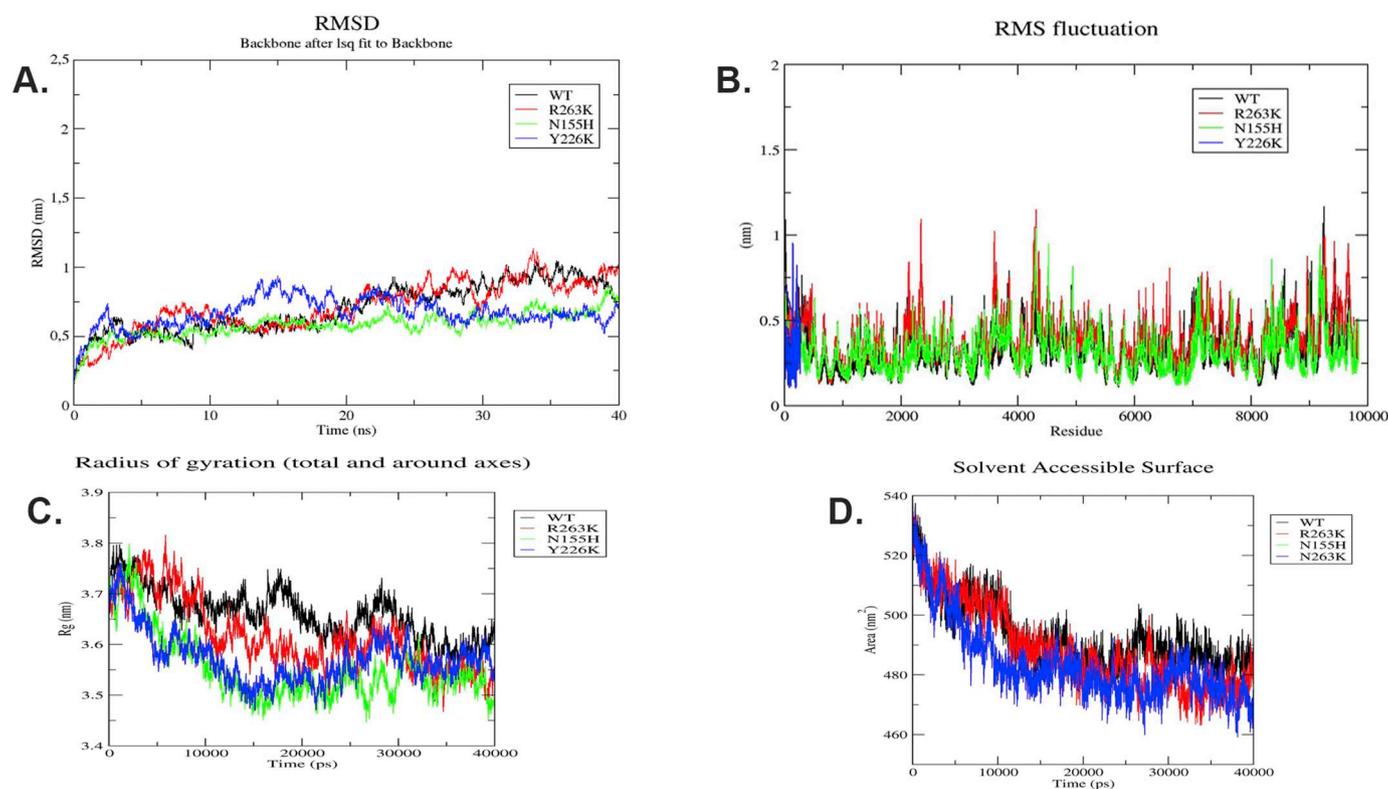
Fig. 3. RMSD of backbone Cα atoms of INT-Complex(A) and other molecular dynamics results, RMSF(B), Radius of Gyration(C) and SASA(D).

Table 1
Binding free energies and individual energy between INT and DLG calculated with MM-PBSA (Unit: kcal/mol).

Component	WT	R263K	N155H	Y226K
ΔEele[a]	−132.57	−85.18	−88.51	−66.47
ΔEvdw[b]	−279.62	−162.04	−194.57	−286.70
ΔGpol[c]	323.74	−220.36	241.07	365.00
ΔGnonpol[d]	−15.79	−10.77	−11.49	−11.19
ΔGbind[e]	−104.25	−37.63	−53.50	0.63

[a] Eletrostatic interaction energies between INT and DLG.
[b] van der Waals interaction energies between INT and DLG.
[c] Polar contributions to the solvation free energy.
[d] Nonpolar contributions to the solvation free energy.
[e] Binding free energy contribution: ΔGbind = ΔEele + ΔEvdw + ΔGpol + ΔGnonpol.

against the novel drugs.

Here we propose a relatively faster, inexpensive and robust method to find and describe possible novel HIV drug-resistance mutations based on computational methods. Our *in silico* approach is not aimed at substituting the experimental findings contained in databases such the HIV Stanford, but it is useful to complement the database information.

We tested our framework using the INT. As the original structure of INT (PDB id: 5U1C) presents several issues on the atoms positions, possibly interfering with the docking results, this was caused because the protein structure is out of their cryoEM map, we noted this error downloading the original map, in the EMDataBank.

We have tried several known methods to fitting the structure in the original map but without successful results. To circumvent this problem, we decided to infer the INT structure of the tetramer by comparative modeling (RMSD < 2.0 Å, and TM-scores tending to 1.0).

Along the DTG mutations reported on the HIV Stanford database, the R263K and N155H variants have been classified as mutations with low-level of drug-resistance (Anstett et al., 2015) (25 and 10 on Stanford mutation score, respectively), whereas the Y226K mutation has never been studied before. Based on our results, we hypothesize that this mutation appears to be a drug-resistance mutation as well, since the parameters that we analyzed were more alike the R263K and N155H variants than the WT. We believe that there could be more mutations associated with resistance to DTG, however since the use of this drug is relatively recent, only a couple of variants were found until now.

Interestingly, Y226K is not localized at the canonical binding site of the DTG, but nearby this region. Thus, drug-resistance mutations could not just interact directly with the drug but cause an allosteric change in the INT, making the binding site less accessible to the ligand. To confirm this, we superposed the mutated structure with the wild-type. Only some parts of the protein have changed their conformation and only two aa (D167 and F185) have changed their secondary structure.

Alone, the docking simulation was not capable to predict if the mutations are related to resistance or not, since the $\Delta\Delta$G between the mutated structures and the WT did not show a significant difference. Therefore, we made the Molecular Dynamics simulation to evaluate if the mutations affect not the attachment but the stability of the system.

The molecular dynamics results have showed that the INT mutated structures, including the Y226K variant, were more compact than the WT. We believe that this compactness made the DTG binding unstable. We carried out a MM-PBSA analysis to see the changes in the free energy along time of the simulations.

In conclusion, this framework has been useful to identify possible novel mutations conferring resistance to DTG, and it could be extended to other drugs and proteins. Our *in silico* framework could be used together with the experimental HIV Stanford database to improve the rate of drugs treatment's success in HIV patients.

Acknowledgments

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.meegid.2018.11.014.

References

Abram, M.E., Ferris, A.L., Das, K., Quinones, O., Shao, W., Tuske, S., Alvord, W.G., Arnold, E., Hughes, S.H., 2014. Mutations in HIV-1 reverse transcriptase affect the errors made in a single cycle of viral replication. J. Virol. https://doi.org/10.1128/JVI.00302-14.

Anstett, K., Fusco, R., Cutillas, V., Mesplède, T., Wainberg, M.A., 2015. Dolutegravir-selected HIV-1 containing the N155H and R263K resistance substitutions does not acquire additional compensatory mutations under drug pressure that lead to higher-level Resistance and increased replicative capacity. J. Virol. 89, 10482–10488. https://doi.org/10.1128/JVI.01725-15.

Barré-Sinoussi, F., Ross, A.L., Delfraissy, J.-F., 2013. Past, present and future: 30 years of HIV research. Nat. Rev. Microbiol. 11, 877–883. https://doi.org/10.1038/nrmicro3132.

Benkert, P., Künzli, M., Schwede, T., 2009. QMEAN server for protein model quality estimation. Nucleic Acids Res. 37, W510–W514. https://doi.org/10.1093/nar/gkp322.

Capriotti, E., Fariselli, P., Rossi, I., Casadio, R., 2008. A three-state prediction of single point mutations on protein stability changes. BMC Bioinforma. 9, S6. https://doi.org/10.1186/1471-2105-9-S2-S6.

Gallo, R.C., Sarin, P.S., Gelmann, E.P., Robert-Guroff, M., Richardson, E., Kalyanaraman, V.S., Mann, D., Sidhu, G.D., Stahl, R.E., Zolla-Pazner, S., Leibowitch, J., Popovic, M., 1983. Isolation of human T-cell leukemia virus in acquired immune deficiency syndrome (AIDS). Science 220 (865 LP-867).

Guex, N., Peitsch, M.C., 1997. SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling. Electrophoresis 18, 2714–2723. https://doi.org/10.1002/elps.1150181505.

Ko, J., Park, H., Heo, L., Seok, C., 2012. GalaxyWEB server for protein structure prediction and refinement. Nucleic Acids Res. 40, W294–W297. https://doi.org/10.1093/nar/gks493.

Krishnan, L., Li, X., Naraharisetty, H.L., Hare, S., Cherepanov, P., Engelman, A., 2010. Structure-based modeling of the functional HIV-1 intasome and its inhibition. Proc. Natl. Acad. Sci. U. S. A. 107, 15910–15915.

Kumari, R., Kumar, R., Lynn, A., Lynn, A., 2014. *g_mmpbsa* —a GROMACS tool for high-throughput MM-PBSA calculations. J. Chem. Inf. Model. 54, 1951–1962. https://doi.org/10.1021/ci500020m.

Laimer, J., Hiebl-Flach, J., Lengauer, D., Lackner, P., 2016. MAESTROweb: a web server for structure-based protein stability prediction. Bioinformatics 32, 1414–1416. https://doi.org/10.1093/bioinformatics/btv769.

Laskowski, R.A., Swindells, M.B., 2011. LigPlot + : multiple ligand–protein interaction diagrams for drug discovery. J. Chem. Inf. Model. 51, 2778–2786. https://doi.org/10.1021/ci200227u.

Laskowski, R.A., MacArthur, M.W., Moss, D.S., Thornton, J.M., 1993. PROCHECK: a program to check the stereochemical quality of protein structures. J. Appl. Crystallogr. https://doi.org/10.1107/S0021889892009944.

Lauck, F., Smith, C.A., Friedland, G.F., Humphris, E.L., Kortemme, T., 2010. RosettaBackrub—a web server for flexible backbone protein structure modeling and design. Nucleic Acids Res. 38, W569–W575. https://doi.org/10.1093/nar/gkq369.

Liu, T.F., Shafer, R.W., 2006. Web resources for HIV Type 1 genotypic-resistance test interpretation. Clin. Infect. Dis. 42, 1608–1618. https://doi.org/10.1086/503914.

Lovell, S.C., Davis, I.W., Arendall, W.B., de Bakker, P.I.W., Word, J.M., Prisant, M.G., Richardson, J.S., Richardson, D.C., 2003. Structure validation by Cα geometry: ϕ,ψ and Cβ deviation. Proteins 50, 437–450. https://doi.org/10.1002/prot.10286.

Michaud, V., Bar-Magen, T., Turgeon, J., Flockhart, D., Desta, Z., Wainberg, M. a, 2012. The dual role of pharmacogenetics in HIV treatment: mutations and polymorphisms regulating antiretroviral drug resistance and disposition. Pharmacol. Rev. 64, 803–833. https://doi.org/10.1124/pr.111.005553.

Tozzi, V., 2010. Pharmacogenetics of antiretrovirals. Antivir. Res. 85, 190–200. https://doi.org/10.1016/j.antiviral.2009.09.001.

Trott, O., Olson, A.J., 2010. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. J. Comput. Chem. 31, 455–461. https://doi.org/10.1002/jcc.21334.

Wiederstein, M., Sippl, M.J., 2007. ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. Nucleic Acids Res. https://doi.org/10.1093/nar/gkm290.