Structural and dynamical properties comparison between the binding of Efavirenz and Nevirapine to K103 Mutation of HIV-1 Reverse Transcriptase

Zambelli Francesco (2029014)

Academic Year 2021-2022

Abstract

HIV reverse transcriptase is an important protein for the HIV virus, and for this reason is very studied in the pharmaceutical and medical world. Many drugs where developed to inhibit the function of this protein, and in this work some relations between them and the HIV-RT protein will be inspected. In particular the process of molecular dynamics will be used, which allows to simulate the movements of the molecules atoms and in this way to understand some of their characteristics. In the first part, some theoretical background regarding molecular dynamics and the description of the methodology used to produce the complexes trajectories are reported. Once the trajectories are produced an analysis is performed to search for quantities which could have a biological relevance, following the considerations reported in the article that describe the complexes structure on the PDB web page[2]. In particular are analyzed:

- the RMSD and the RMSF of the drugs to understand how much and how they change conformation during the time evolution
- the trend of some specific dihedrals angles that help to understand which parts of the drug are involved in the major structural changes
- some distance between atoms that could present and hydrogen bond, which gives stability to the binding
- the position of the drug with respect to the protein active site
- the comparison of the residues composing the active site positions in the two different complexes

The main idea behind the analysis is to inspect the binding affinity between the drug and the active site of the protein.

1 Introduction

HIV-1 Reverse transcriptase (HIV-RT) is a protein which function is to transcribe the HIV-1 RNA genome into double stranded DNA, which is subsequently integrated into the host cell genome by an integrase enzyme[1]. To cure such a disease, many drugs and therapies have been proposed, but we are far from being able to say that the virus is beaten. This is mainly due to the fact that the virus, and in particular the HIV-RT protein, is able to adapt to the presence of the drug and mutate accordingly. There are different types of drug used, and good results are achieved with the class of non-nucleoside reverse-transcriptase inhibitors (NNRTIs). NNRTIs operate allosterically, binding approximately 10 Å from the polymerase active site in a pocket[2]. These drugs can be further divided in two sub-classes, the first generation drugs, as nevirapine (NVP), and the second generation ones, such as efavirenz (EFZ). Experimentally it was seen that second generation drugs are more resistant to point-mutation of the protein structure than the first generation ones[2].

In this work we are going to study the structural property of one of the most frequent mutations, the K103 in which the Leu103 is substituted with the Asn103, in presence of both nevirapine and efavirenz. The PDB codes are respectively 1fko[3] and 1fkp[4]. The goal is to explore both the structural and dynamical properties in the two protein complexes and understand what could be the main differences between first and second generation drugs.

2 Methods

A MD simulation was performed to equilibrate and study the structural and dynamical properties of the systems. The process is the same for HIV-RT bounded with EFZ and NVP. Given the fact that the point of interest of the work is essentially to study the interactions between the drug and the active site of the protein, only chain A of the HIV-RT was considered, in order to speed up the simulation.

The program used for the simulations is Amber[5], for the parametrization of the acting forces between atoms the force fields `ff99SB` and `gaff` were used and the solvent water molecules were generated by using the `tip3p` water model.

Before starting with the actual description of the simulation setup, a brief theoretical explanation about Molecular Dynamics will be presented.

2.1 Brief Theoretical Introduction

In this work different tools were used to reach the final goal. The main production of the configurations to be studied comes from long Molecular Dynamics runs, but this can be done only after some preliminary steps.

The first thing to do is to parametrize the molecule, which originally is a PDB file with coordinates, type and belonging residue for each atom in it, in order to create a structure in which also bond length, bond angles, dihedral angles and charge are defined. After this the potential energy structure that characterize the interactions between the molecule's atoms can be built by using the so called Force Fields (FF), a computational method that is used to estimate the forces between atoms within molecules and also between molecules. More precisely, the force field refers to the functional form and parameter sets used to calculate the potential energy of a system of atoms in molecular mechanics, molecular dynamics, or Monte Carlo simulations. If at the end of such procedure the molecule is characterized by a net charge different from zero, it's recommended to add ions of the opposite sign to balance the charge and make the final system neutral.

When dealing with biological molecules, an important aspect to point out is the role of the solvent, essential to produce a realistic simulation. The two main options are: implicit and explicit solvent. With the explicit solvent, each solvent particle is generates and added to the coordinates file of our system, while with the implicit solvents a mean-field approach is used. Using an explicit solvent is computationally expensive, requiring inclusion of roughly ten times more particles in the simulation, but the granularity and viscosity of explicit solvent is essential to reproduce certain properties of the solute molecules. On the other hand the implicit solvent acts on the solute atoms with actions that simulate the presence of the solvent, such as slowing of dynamics due to viscosity/friction and acting as a "heat bath" which can exchange energy with the solute.

In the following setup the explicit solvent option is chosen. This requires the introduction of Periodic Boundary Conditions (PBC), which goal is to minimize the boundary effects that could compromise the goodness of the simulation. PBC are a set of boundary conditions which make it possible to approximate an infinite system by using only a small part of it (unit cell). A unit cell in MD is usually referred to as periodic box, i.e. a fixed region of space in which both the solute, in our case HIV-RT complex, and the solvent, water molecules, are contained. To implement PBC the unit cell is surrounded by translated copies in all directions to approximate an infinitely large system. When one molecule diffuses across the boundary of the simulation box it reappears on the opposite side. So each molecule always interacts with its neighbours even though they may be on opposite sides of the simulation box.

After the parametrization and the definition of the solvent are done, a first quick phase of minimization is performed in order to rapidly get away from the non-stable initial configuration. This is

	Not optimized	Optimized
Rectangular	31985	23411
Octagonal	39367	39386

Table 1: Number of solvent molecules generated for all the possible combinations in the case of EFZ used as drug

done by using Molecular Mechanics (MM), in which the gradient of all Potential Energy surface (PES) is calculated with respect to each d.o.f. of the system (in this phase only positions are considered) and proceeding in the opposite direction, from the properties of the gradient, the system goes in the direction of a minimum. This procedure doesn't reproduce realistic movements of the molecules, because it doesn't follow the physics law, but it's a simple mathematical optimization procedure.

After this, it's necessary to generate the set of velocities related to the atoms, and reach a distribution of position and momenta which is compatible with a desired system's temperature. Both this step and the generation of the actual trajectory that will be studied are done by using Molecular Dynamics. Molecular dynamics (MD) is a computer simulation method for analyzing the physical movements of atoms and molecules. The trajectories of atoms and molecules are determined by numerically solving Newton's equations of motion for a system of interacting particles, where forces between the particles and their potential energies are calculated using the molecular mechanics force fields cited before.

2.2 Preparation

To create a structure that can be used to run a MD simulation of the HIV-RT complex, the first necessary step is to refine the raw PDB file downloaded from the PDB site. The function `pdb4amber` from Amber is used to convert the syntax of the original file into the one used in `tlep`. After this the hydrogen atoms are added to the structure with the `reduce` command, then with the `antechamber` program, the parametrization of the system is done and bonds, angles, dihedrals and charges are defined. The chosen force field is applied to the complex by using the `tleap` program, firstly only to the drug molecule to generate a *.lib* file that describes it, and then to the whole complex (protein+solvent+drug). In this way the inter-atomic forces are defined, and the system is ready to be used for MD simulations.

2.3 Solvation

Explicit solvent is used for the simulation, in order to increase the precision as much as possible given the fact that the used FF is not so accurate, and so, also PBC has to be introduced. The unit cell can be generated with `tleap`, by giving in input the solute parameter file and some characteristics of both the shape of the cell we desire (which can be cubic, rectangular, octagonal ...) and the libraries necessary to describe the solvent molecules, in our case water. This chosen water model is `tip3p`.

The chosen cell shape is the rectangular one, which appear as the shape more fitted for the spatial conformation of the complex, but before proceeding with the other steps the protein orientation is optimized in order to minimize as much as possible the number of solvent molecules required to fill the cell. This is done by exploiting a function from the `Simulaid` program[6], in particular following the *Geometry Optimization > Optimize orientation in bounding cube/rectangle* commands path. Tab.1 clearly shows the good impact of this process when dealing with the complex HIV-RT plus EFZ (similar quantities can be found also in the case of the combination with NVP).

2.4 Minimization

After generating the necessary structure it's possible to proceed with the minimization phase. In order to do so the sander program from Amber is used, by setting the input's file parameter `imin` equal to 1. A cutoff of 10Å is used, which means that for every atom interactions are only computed up to a fixed atomic distance. The chosen value is high enough to obtain good results but low enough to avoid

interactions between contiguous cells in the PBC framework. The cutoff distance is the same also for the next passages of the simulation. The minimization is performed in two steps.

In the first one all the residues of the protein and also the drug are kept fixed in order to minimize only the positions of water molecules and ions. The restraints are introduced by setting the input file's parameter `ntr=1`. This has the effect to impose harmonic potentials between atoms of the selected residues. The rigidity of such interactions can be explicitly set, and in this way, by imposing a large enough value (in our case a value of 500.0 kcal/mol $Å^2$ is chosen), it's possible to keep almost fixed the selected units. In this step the chosen number of steps is equal to 1000, of which in the first 500 the steepest descent algorithm is used, while for the remaining ones it is changed with the conjugate gradient one.

In the second step the restraints are relaxed and the whole system is allowed to move freely. The objective is reaching a more favorable energetic state from where to start the equilibration process. The results for both EFZ and NVP are reported in fig.1. As can be seen in both cases the energy quickly decrease in the first steps while at the end it reaches an almost constant behaviour. The y axis of the gradient graphs (fig.1 on the left) have a logarithmic scale, and indicates that in the last iterations the value of the maximum gradient decreased of almost two orders of magnitude and also the fluctuations are much smaller.

2.5 Heating

After the minimization step, in which only the positions of the atoms and the interactions forces are considered, it's necessary to generate also the velocities associated to each atom. In order to do so, the sander program from amber is used, with parameter `imin` set equal to 0, which corresponds to the MD setup. Imposing the parameter `irest=0`, random initial velocities are generated from a Boltzmann distribution. Then the Langevin thermostat, which regulated the random interaction by an ideal thermal bath and our system (`ntt=3`), is declared and the associated collision frequency is set to 1.0 ps^{-1}. Then initial and target temperature for the heating procedure are set, respectively equal to 0 and 300 K. The last one corresponds more or less to the temperature inside the human body, the environment in which we are interested to study the drug+protein complex. Another important aspect is the use of the SHAKE algorithm (`ntc=2`), which allows to keep some specific bounds fixed, which in our case correspond to the one involving hydrogen atoms. This allows to speed up the computation and the ignored movements don't produce a big loss in accuracy as regards the simulation.

As in the previous point, two steps are performed. The first one with harmonic restraints of the protein plus drug complex, but this time the force constant is chosen equal to 10.0 kcal/mol $Å^2$[9]. This weak constraints ensure that this first stage of equilibration happens without wild fluctuations in our system. The chosen number of iterations is 10000 with time-step of 1 fs. In the second step we release the constraints as before allowing the whole system to move freely. Here the iteration number is still 10000 but the time-step is increased to 2 fs, because we suppose that up to this stage, the system will be less susceptible to wild fluctuations.

The goal of this phase is to progressively tune the atoms velocities in order to reach a distribution compatible with our target system temperature. The results for both EFZ and NVP are presented in fig.2. As can be seen in both cases the target temperature is reached after almost 5 ps, and also the different energy quantities reaches a plateau at this instant. Nevertheless the total simulation lasts 25 more seconds, this because even if the target temperature is reached and the fluctuations are much smaller than in the beginning, this doesn't mean the system is in equilibrium, also because our protein plus drug complex is only a small part of the total system, in which the majority of the atoms belongs to the solvent.

2.6 Production Run

Finally, once the equilibrium is reached, a long MD run was performed in order to explore the configuration space of the complex and study its dynamic properties. The sander program from Amber was used with the parameter `imin=0`, which means MD setting. As in the previous point, the SHAKE applied to the bonds evolving hydrogens was used to speed up the simulation, supposing that tracking

such movements is not so fundamental for the goal of this work, while no harmonic constraints were applied.

Different runs were performed in order to span a large time interval, always keeping a time-step of 2fs. In the first one the last frame generated by the heating phase was taken as starting point, while in the the others, as starting point of a run it was used the final frame of the previous one.

At the end of this phase, the program simulated 900ps for the case of the complex with EFZ, while in the case of NVP the total simulated time interval is 920ps.

3 Results and Discussion

In the following paragraph will be presented the analysis of the simulations results done by following the biological considerations expressed in the article associated with the PBD page of 1fk0 and 1fk9, with some more attempt to recognize quantities that could be of biological interest.

The visualization of the molecule was mainly done using the `VMD` program [7], while the analysis of the features of interests is performed using both `VMD` and the `mdtraj` [8] library of python.

In all the graphs and in the discussion below, the temporal evolution of the trajectory is defined in progression of frames. Given the fact that the step size in the MD production simulation is set to 2fs and that each trajectory frame is printed every 500 steps, every frame is separated from the previous one with a temporal distance of 1ps.

3.1 Fluctuations and stability

The first thing to inspect is the conformational stability of the two drugs bounded to the protein.

First generation drugs as NVP, are generally bigger and less movable than second generation drugs [2]. It's possible to verify so by computing the RMSD for the two drugs in the complex and compare them. The results are shown in fig.3, 4, 5. By comparing fig.3 and 4 it's possible to see that in the first stage of the simulation, EFZ seems to be more movable and changes often conformation, while for NVP it seems that there are two preferred conformation, with periodic jumps from one to the other. Some more information about such a behaviour can be extracted by looking at the fig.5. From image *a* we can see that the magnitudes of the fluctuations for the two drugs are comparable, but, by looking at image *b*, we can see that the fluctuation in the case of EFZ are bigger than for NVP. By looking at the line corresponding to NVP in image *b*, it's clear that the two main configuration for the drug are also characterized by different magnitude of fluctuations. In particular we can see that the standard deviation of the RMSD alternates interval of low values with period of large values, which could correspond to different configurations of the molecule. By looking at the visual representations of the molecules during the simulation (fig.7, *b* and *c*) the movement responsible for the grater variability of the drug's structure seems to be the rotation of the NVP's cyclopropyl group (fig.6 *b c*). Another important hint can be obtained by observing the RMSF graph for the NVP molecule (fig.6), in which the highest values are reached for the hydrogen atoms of the cyclopropyl group. To quantify such an observation, the dihedral angle between atoms CC, CA, N1, C2 of the NVP molecule (image a of fig.7) is computed for the whole trajectory, and two distinct regimes can be clearly observed. For simplicity the configuration with such dihedral angle that fluctuates between 0 and -50º will be denoted config. A, while the one in which the dihedral is around 180º will be denoted config. B. From the dynamical representation of the complex, it appears that config. B, which is the one characterized by the larger fluctuations, is the one in which the cyclopropyl group is rotated within the center of the protein, and, being placed in a more free space, it's allowed to move more freely, while in config. A the cyclopropyl group seems to be in tight contact with some water molecules close to the active site which could reduce the movement capabilities of the atomic group.

As regards EFZ, it looks that in the first part of the simulation the drug conformation changes frequently, leading to large values of the RMSD. On the other hand, also by looking at fig.5 *b*, around frame 400 the system starts to become more stable and the fluctuations decreases accordingly. As in the case of before, the majority of the movement of EFZ are consequence of the rotation of the cyclopropyl group. Analyzing the variation of the dihedral angle between C6, C7, C10, C12 of EFZ (fig.8) it's possible to appreciate such dumping as the simulation progresses.

3.2 Distance from the active site

To inspect the relations between the drugs and their closest residues, reciprocal positions and movements between the active site and the drugs are studied. First of all, the active site is chosen as the set of residues which posses some atoms within a distance of 4Å from the drug. This was done by using the selection language of `VMD` and the result is the following list: Leu100, Lys101, Asn103, Val106, Val179, Tyr181, Tyr188, Val189, Gly190, Pro225, Phe227, Trp229, Leu234, Pro236, Tyr318 (fig.10). Two quantities were then computed:

- the distance between the center of mass of such residues group
- the average of the distances between the center of mass of each atoms in the active site

The first one could represent how much the drug is centred within the region inside the active site, while the second one quantifies how much the active side is tightened around the drug.

The results are presented in fig.11. It shows that the first distance is almost constant for the case of NVP, while for EFZ is generally smaller and also seems to decrease in the first part of the simulation, signal that the residues gets progressively closer to the drug. Also the distance between the center of mass of the active site group and the drug decreases in the case of EFZ while for NVP fluctuates around a mean value. All this signal could suggest that in the case of EFZ the active site is able to adapt better to the shape of the drug, and this could possibly mean that the binding affinity is grater.

3.3 Configuration of the active site

From the analysis of the conformation of the residues composing the active site during the trajectory, some more considerations can be done. If the trajectories of the complexes containing EFZ and NVP are aligned by minimizing the distances of the active site's residues, it's possible to see that almost all of them are characterize by a similar position inside the region, with the exception of Tyr181 (fig.12). Its aromatic ring has opposite orientation in the case of EFZ and NVP, such a difference characterize the whole trajectory. The consequence is that is the case of EFZ the aromatic ring "close" the region in which the contact between the drug and the water molecule are more frequent, while in the case of NVP such connection is "open". In this way it's more likely that NVP is in direct contact with water molecule than EFZ.

3.4 Hydrogen bond

In the article [2] the authors proposed that between the bonzoxanin-2-one ring nitrogen of EFZ and the O1 atom of the the Asn 103 residue (the mutated one) there could be an hydrogen bond, but the measured distance equal to 3.7Å is too large to allow a strong interaction. In order to verify this, such distance was computed for the simulated trajectory, and the results reported in fig.9 *b*, were obtained. In particular the average value of the separation between these two atoms, equal to 3.21 ± 0.32Å is definitely smaller than 3.7Å, and partially inside the interval $2.7 - 3.3$Å which characterize the optimal inter-atomic distance for hydrogen bonding[10].

In the article it's also highlighted the presence of another hydrogen bond between the O of the Lys 101 residue and the same nitrogen of the EFZ, which is fundamental for the increased binding intensity of this second generation drug with respect to first generation one. The distance between these two atoms during the time evolution is reported in fig.9 *a*. On average the distance is equal to 2.88 ± 0.13Å, and it's completely inside the $2.7 - 3.3$Å interval, which suggest that between these two atoms there is actually a strong hydrogen bond.

On the other hand NVP doesn't seem to have any potential hydrogen bond, and this has as consequence a reduced interaction strength with the active site.

4 Conclusion

From the analysis results it's possible to conclude that the complex with EFZ possesses different traits that suggests a grater binding affinity between the drug and the protein, such as the presence of two

hydrogen bonds, the tighten disposition of the active site residues and the progressive stabilization of the drug shape. On the other hand the NVP complex, which doesn't show any of this traits, can be considered to be less stable.

Nevertheless, the simulations time is definitely insufficient to have insightful results, in fact 1ns second of simulation won't be able to capture a representative population of the two complexes conformations. Better results could be achieved by running a simulations for much large time, maybe of the order of milliseconds, by taking in consideration even the chain B of HIV-RT and finally by using a more accurate force field.

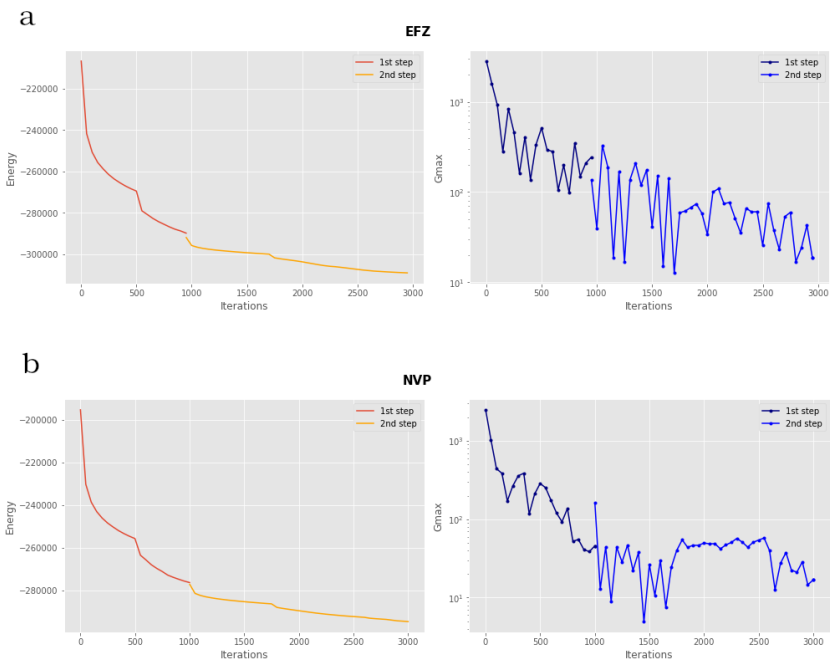

Figure 1: Key quantities for the minimization procedure for both presence of EFZ (a) and NVP (b). In the graphs on the left is reported the behaviour of the energy, and it's divided in a region with dark red line, corresponding to the phase in which only the solvent and the ions are allowed to move, and the one with orange line, in which all the constraints are released. On the right there are reported the trends of the gradient, and as before the dark blue line corresponds to the phase with constraints, the light blue one line to the phase without constraints. For both energies and gradient, the trend of the curves indicate that the minimization converged.

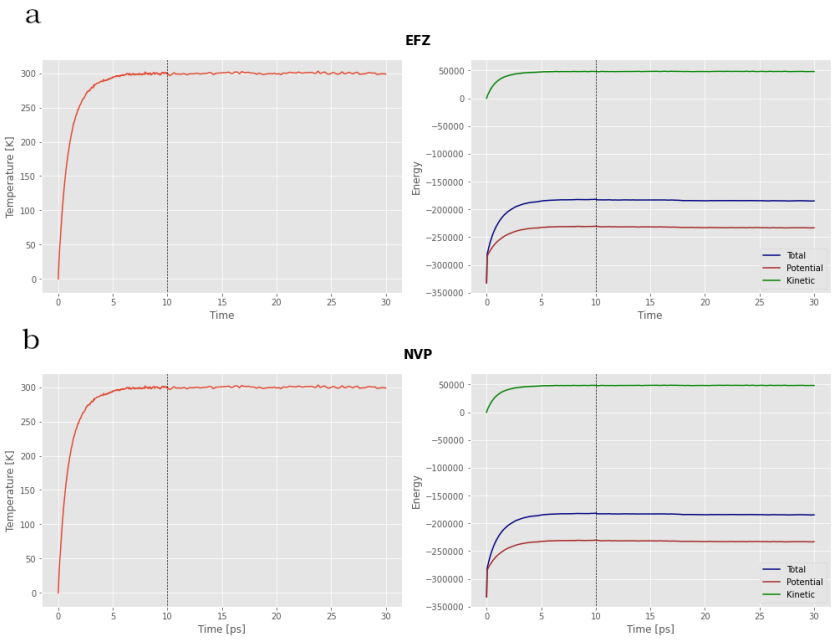

Figure 2: Key quantities related to the heating phase for both presence of EFZ (a) and NVP (b). In the graphs on the left is reported the trend of the temperature, and the dashed vertical line separates the regions in which the constraints are present from the one in which they are not. On the right there are reported the trends of the energies, with the green line corresponding to the total energy, the blue line the potential energy and the red line the kinetic energy. The black vertical line indicates the same separation as above. For both energies and temperature, the trend of the curves indicate that the equilibration procedure converged.

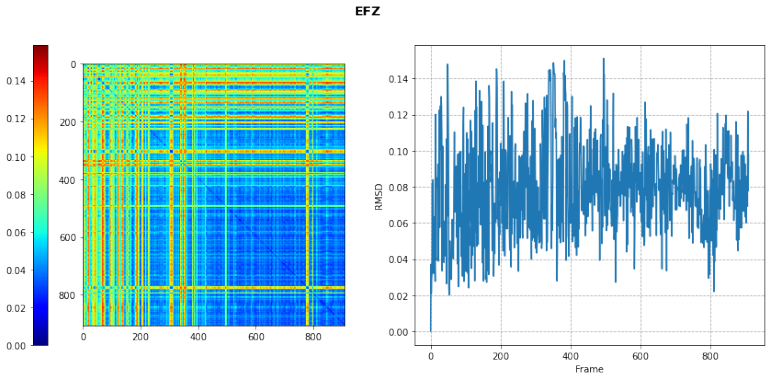

Figure 3: RMSD of the EFZ drug, on the left the 2D representation, on the right the 1D one.

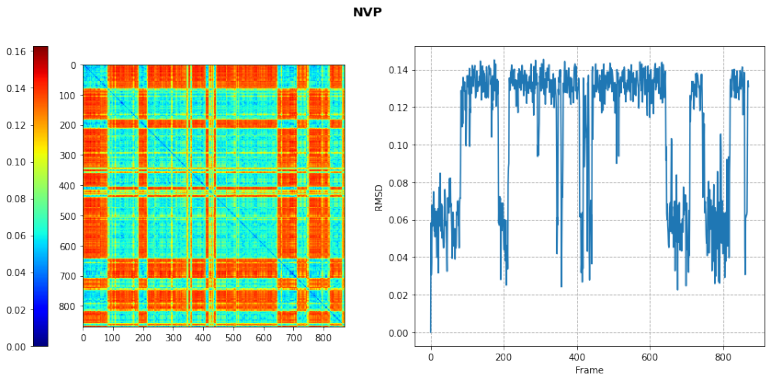

Figure 4: RMSD of the NVP drug, on the left the 2D representation, on the right the 1D one.

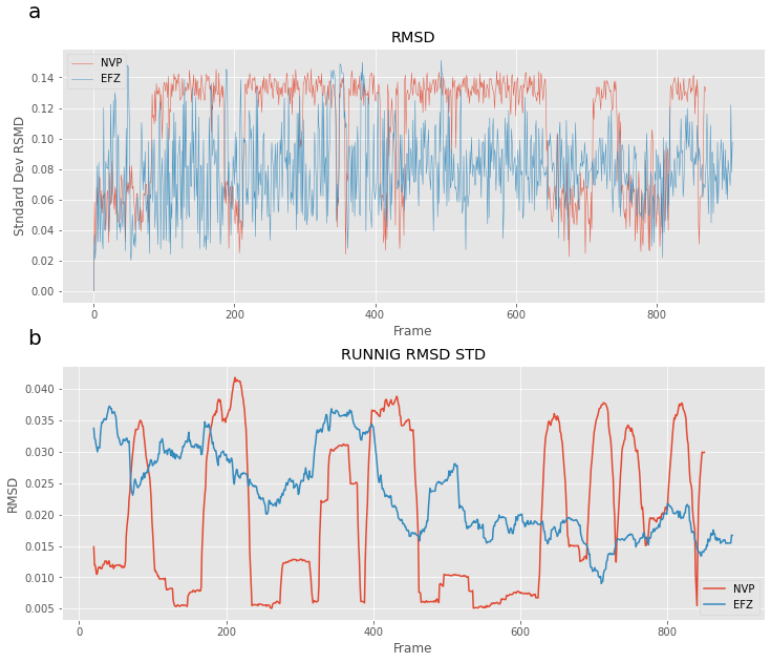

Figure 5: In figure a it's reported the comparison of the 1D RMSD trends between EFZ (blue) and NVP(red). In figure b the lines correspond to the values of the standard deviation of intervals of 40 points of the 1d RMSD. This last graph helps to understand the width of the fluctuations in the different intervals of the trajectory.

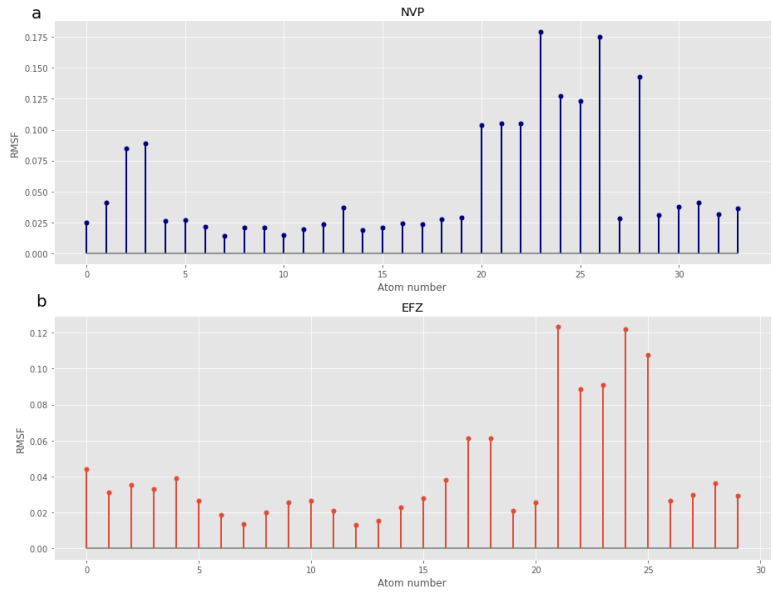

Figure 6: RMSF for the drug's atoms in the case of NVP(A) and EFZ(b).

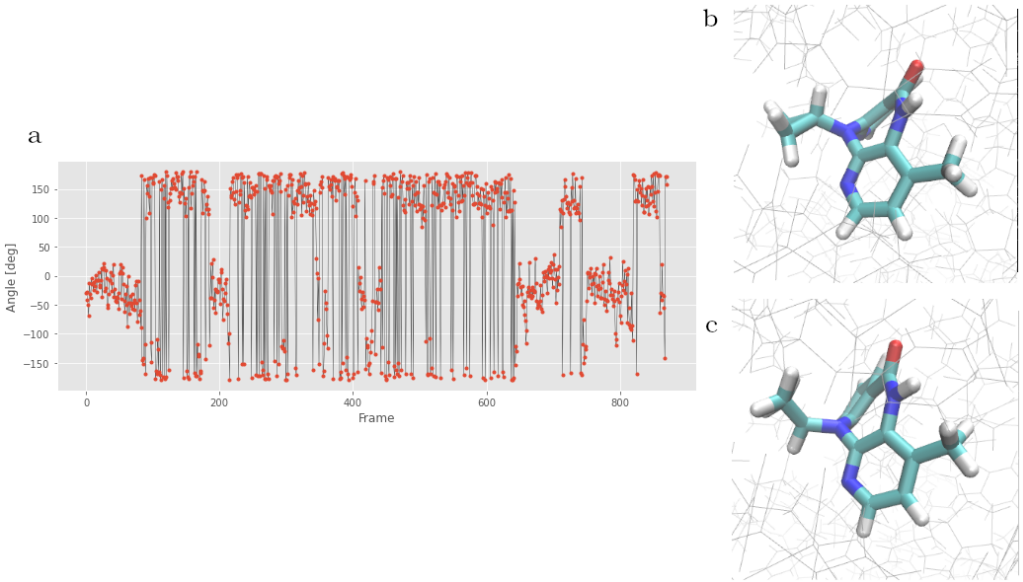

Figure 7: Figure a represents the trend of the dihedral angle between the atoms CC, CA, N1, C2 of NVP. In figure b and c are reported the two main configuration in which NVP appears, characterized by the change of the orientation of the cyclopropyl group.

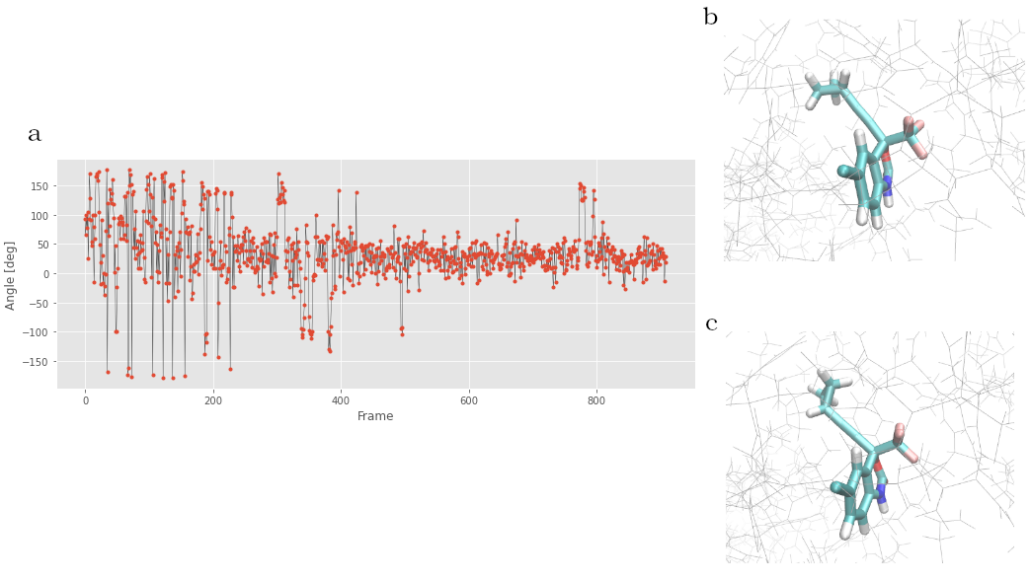

Figure 8: Figure a represents the trend of the dihedral angle between the atoms C6, C7, C10, C12 of EFZ. In figure b and c are reported the two main configuration in which EFZ appears, characterized by the change of the orientation of the cyclopropyl group.

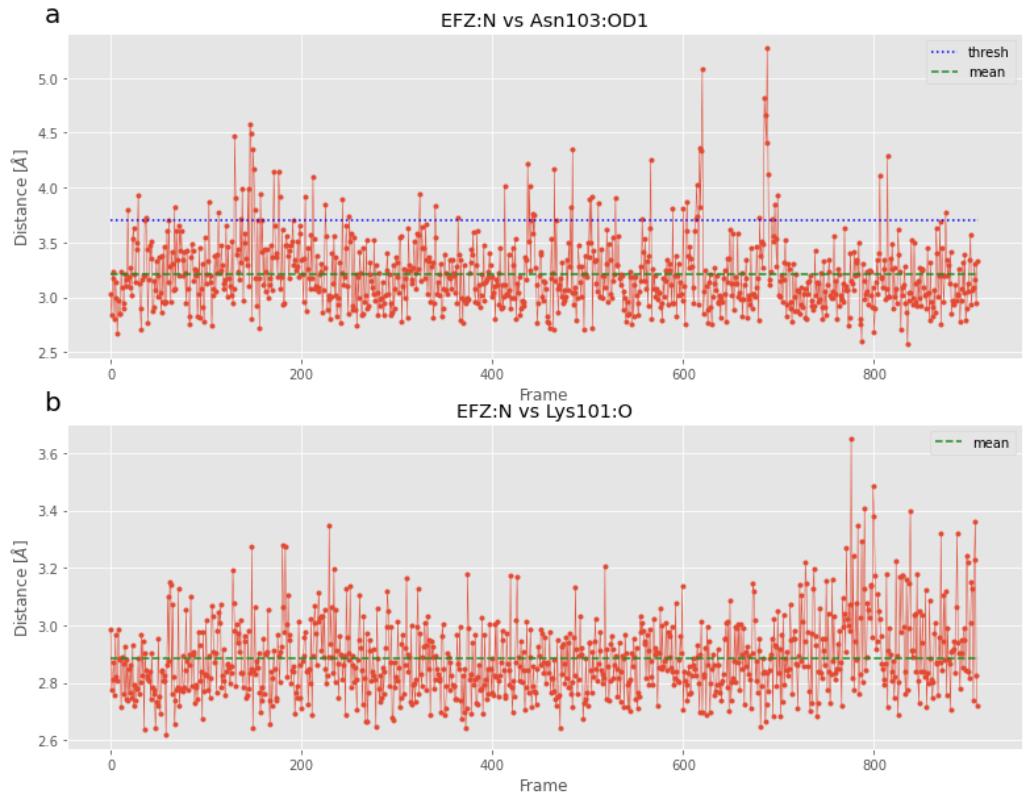

Figure 9: The distance between the OD1 atom from the residue Asn103 and N from EFZ (a). The dashed blue line corresponds to the value of 3.7Å reported in [2], while the green dashed line is the mean value of the distance over time, corresponfing to 3.21 ± 0.32Å. In figure b it's reported the distance between the O atom from the residue Lys101 and N from EFZ. The green dashed line indicates the mean value of such distance, corresponding to 2.88 ± 0.13Å.

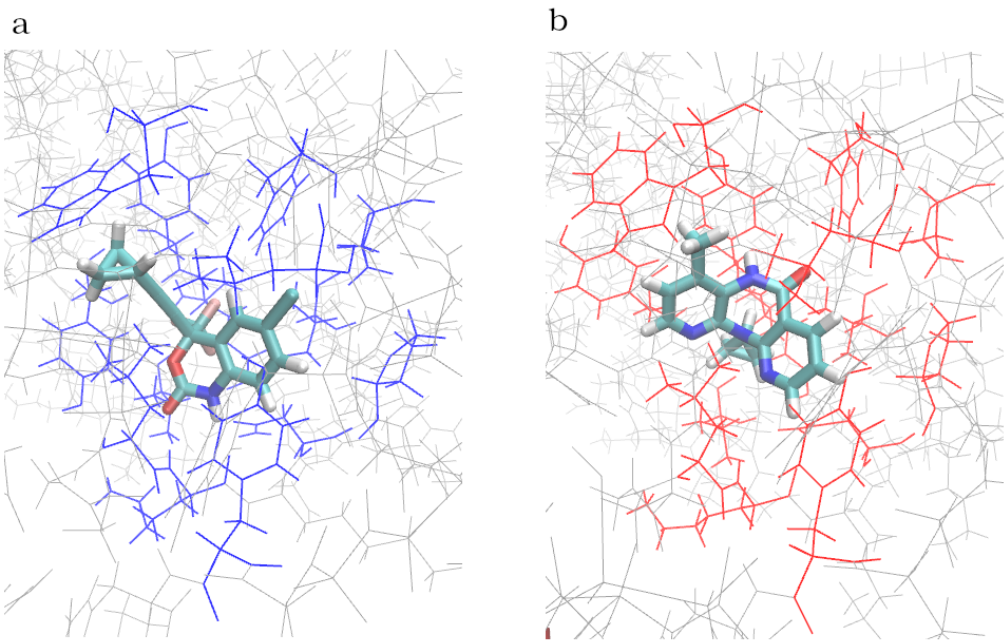

Figure 10: The drug, drawn as licorice, the active site residues, thick colored lines and the other residues of the protein, thin grey line, for both the EFZ complex (a) and the NVP one (b).

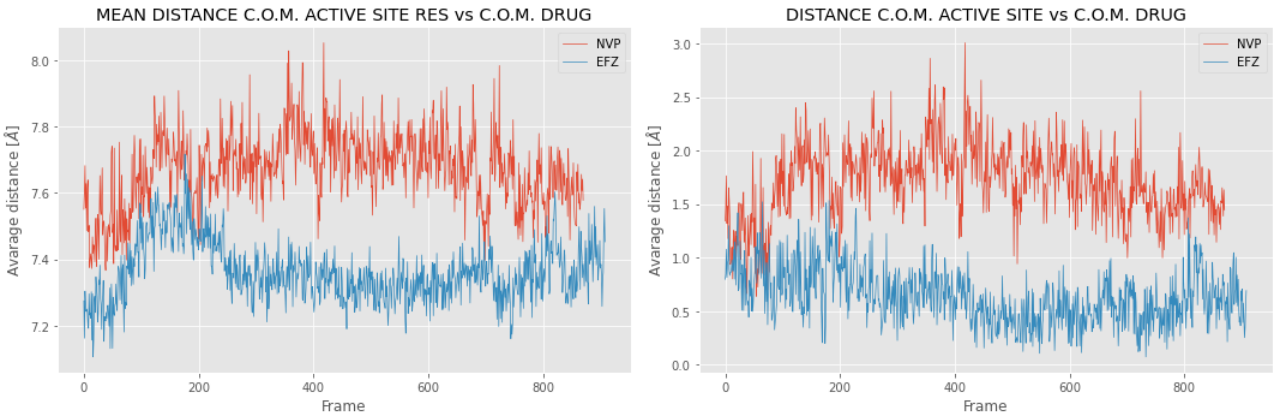

Figure 11: On the left: mean distance between the center of mass of the drugs and the centers of mass of the active site's residues. On the right: distance between the center of mass of the drugs and the one of the group of the active site residues.

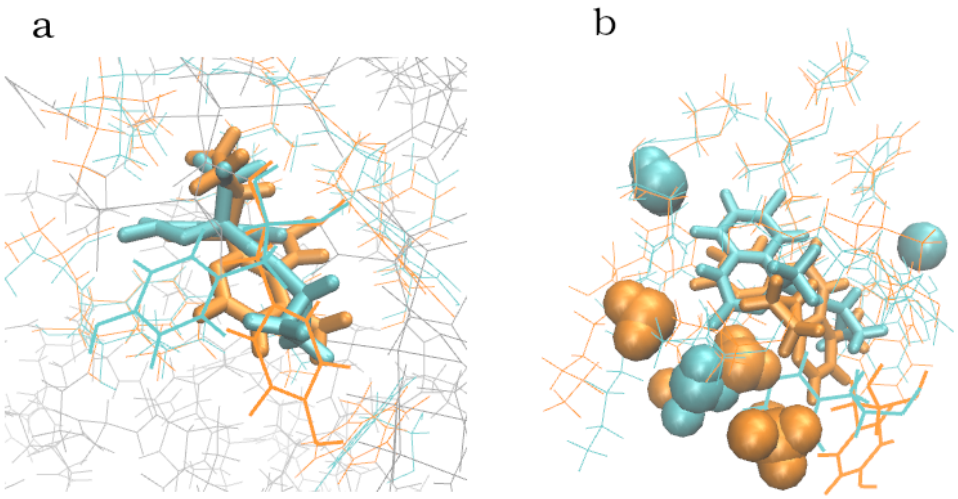

Figure 12: In both the images the licorice representation indicates the drug, the thick lines correspond to the Tyr181 residue, the thin colored lines are the active site residues and the grey line the other protein's residues. The cyan color is for the EFZ complex, the orange one for the NVP one. In image a it's highlighted the difference in the position of the residue Tyr181 in the two cases, while in b are also reported the water molecules within a distance of 5Å from the drug, and are represented as spheres.

References

[1] Use of MM-PBSA in Reproducing the Binding Free Energies to HIV-1 RT of TIBO Derivatives and Predicting the Binding Mode to HIV-1 RT of Efavirenz by Docking and MM-PBSA, Junmei Wang, Paul Morin, Wei Wang, and Peter A. Kollman

[2] Structural basis for the resilience of efavirenz (DMP-266) to drug resistance mutations in HIV-1 reverse transcriptase, Ren, J., Milton, J., Weaver, K.L., Short, S.A., Stuart, D.I., Stammers, D.K.

[3] https://www.rcsb.org/structure/1FKO

[4] https://www.rcsb.org/structure/1FKP

[5] D.A. Case, H.M. Aktulga, K. Belfon, I.Y. Ben-Shalom, J.T. Berryman, S.R. Brozell, D.S. Cerutti, T.E. Cheatham, III, G.A. Cisneros, V.W.D. Cruzeiro, T.A. Darden, R.E. Duke, G. Giambasu, M.K. Gilson, H. Gohlke, A.W. Goetz, R. Harris, S. Izadi, S.A. Izmailov, K. Kasavajhala, M.C. Kaymak, E. King, A. Kovalenko, T. Kurtzman, T.S. Lee, S. LeGrand, P. Li, C. Lin, J. Liu, T. Luchko, R. Luo, M. Machado, V. Man, M. Manathunga, K.M. Merz, Y. Miao, O. Mikhailovskii, G. Monard, H. Nguyen, K.A. O'Hearn, A. Onufriev, F. Pan, S. Pantano, R. Qi, A. Rahnamoun, D.R. Roe, A. Roitberg, C. Sagui, S. Schott-Verdugo, A. Shajan, J. Shen, C.L. Simmerling, N.R. Skrynnikov, J. Smith, J. Swails, R.C. Walker, J Wang, J. Wang, H. Wei, R.M. Wolf, X. Wu, Y. Xiong, Y. Xue, D.M. York, S. Zhao, and P.A. Kollman (2022), Amber 2022, University of California, San Francisco

[6] https://mezeim01.u.hpc.mssm.edu/simulaid/simulaid.html

[7] William Humphrey and Andrew Dalke and Klaus Schulten,VMD – Visual Molecular Dynamics, Journal of Molecular Graphics, 1996, volume 14, pages 33-38

[8] https://www.mdtraj.org/1.9.8.dev0/index.html

[9] https://ambermd.org/tutorials/basic/tutorial1/section5.htm

[10] COMPUTATIONAL TECHNIQUES, Duncan E. McRee, in Practical Protein Crystallography (Second Edition), 1999