



Università degli Studi di Cagliari

**CORSO DI DOTTORATO IN SCIENZE E TECNOLOGIE  
FARMACEUTICHE**

Cycle XXVIII

**STRUCTURE-BASED APPROACHES APPLIED TO THE STUDY OF  
PHARMACEUTICAL RELEVANT TARGETS**

CHIM08

Presentata da  
Coordinatore Dottorato  
Tutor

Giulia Bianco  
Prof. Elias Maccioni  
Dr. Simona Distinto

Esame finale anno accademico 2014-2015



## **ABSTRACT**

Computer Aided Drug Design/Discovery methods became complementary to traditional and modern drug discovery approaches. Indeed CADD is useful to improve and speed up the detection and the optimization of bioactive molecules. The present study is focused on the application of structure-based approaches to the study of pharmaceutical relevant targets. The introduction provides a quick overview on the fundamentals of computational chemistry and structure-based methods, while in the successive chapters the main targets investigated through these methods are treated. In particular we focused our attention on Reverse Transcriptase of HIV-1, Monoamine oxidase B and VP35 of Ebola virus. The last chapter is dedicated to the validation of covalent docking performed with Autodock.

# CONTENTS

<b>1</b>	<b>CHAPTER.....</b>	<b>1</b>
1.1	INTRODUCTION TO COMPUTER AIDED DRUG DESIGN.....	1
1.2	BASIS OF COMPUTATIONAL CHEMISTRY.....	4
1.2.1	<i>Basic theory of quantum mechanical methods.....</i>	<i>5</i>
1.2.2	<i>Basic theory of molecular mechanical methods.....</i>	<i>6</i>
1.3	STRUCTURE-BASED APPROACHES.....	15
1.3.1	<i>Molecular docking.....</i>	<i>16</i>
1.3.2	<i>Molecular dynamics.....</i>	<i>25</i>
1.3.3	<i>Pharmacophore modeling.....</i>	<i>28</i>
1.4	REFERENCES.....	32
<b>2</b>	<b>CHAPTER.....</b>	<b>42</b>
2.1	INTRODUCTION TO REVERSE TRANSCRIPTASE OF HIV-1.....	42
2.2	DIARYLPROPENONES AS DUAL INHIBITORS OF HIV-1 RT.....	53
2.2.1	<i>Biochemical studies.....</i>	<i>56</i>
2.2.2	<i>Molecular modeling.....</i>	<i>58</i>
2.2.3	<i>Conclusions.....</i>	<i>73</i>
2.3	INDOLINONES DERIVATIVES AS DUAL INHIBITORS OF HIV-1 RT.....	74
2.3.1	<i>Biochemical studies.....</i>	<i>77</i>
2.3.2	<i>Computational studies.....</i>	<i>84</i>
2.3.3	<i>Single point site directed mutagenesis in pocket 1.....</i>	<i>88</i>
2.3.4	<i>Single point site directed mutagenesis in pocket 2.....</i>	<i>90</i>
2.3.5	<i>Conclusions.....</i>	<i>92</i>
2.4	SUMMARY.....	92
2.5	EXPERIMENTAL PART.....	93
2.5.1	<i>Synthesis and characterization.....</i>	<i>93</i>
2.5.2	<i>Protein expression and purification.....</i>	<i>99</i>
2.5.3	<i>Biochemical assays.....</i>	<i>100</i>
2.5.4	<i>Molecular modeling.....</i>	<i>102</i>
2.5.5	<i>In vitro membrane permeation study.....</i>	<i>104</i>
2.6	REFERENCES.....	106
<b>3</b>	<b>CHAPTER.....</b>	<b>120</b>
3.1	INTRODUCTION TO MONOAMINO OXIDASE B.....	120
3.2	NEW OXADIAZOLES AS SELECTIVE INHIBITORS OF MAO-B.....	127
3.2.1	<i>Design.....</i>	<i>127</i>
3.2.2	<i>Synthesis.....</i>	<i>129</i>
3.2.3	<i>Biological activity.....</i>	<i>130</i>
3.2.4	<i>Structure Activity Relationships.....</i>	<i>132</i>
3.2.5	<i>Enantiomeric mixture resolution.....</i>	<i>133</i>

3.2.6	<i>Molecular modeling</i> .....	136
3.2.7	<i>Conclusions</i> .....	149
3.3	EXPERIMENTAL PART .....	150
3.3.1	<i>Chemistry and compounds characterization</i> .....	150
3.3.2	<i>Enantioseparation and X-ray crystal</i> .....	154
3.3.3	<i>Enzymatic assay</i> .....	156
3.3.4	<i>Molecular modelling</i> .....	156
3.4	REFERENCES .....	159
<b>4</b>	<b>CHAPTER</b> .....	<b>176</b>
4.1	INTRODUCTION TO VP35 OF EBOLA VIRUS .....	176
4.2	DEFINITION KEY RESIDUES IN dsRNA RECOGNITION OF VP35 .....	181
4.2.1	<i>dsRNA binding property of Ebola VP35 end-capping mutants</i> .....	181
4.2.2	<i>Molecular modeling</i> .....	187
4.2.3	<i>IFN-β inhibition property of EBOV VP35 mutants</i> .....	193
4.2.4	<i>Conclusions</i> .....	198
4.3	VIRTUAL SCREENING .....	199
4.3.1	<i>Conclusions</i> .....	201
4.4	EXPERIMENTAL PART .....	202
4.4.1	<i>EBOV VP35 wt and mutant plasmids</i> .....	202
4.4.2	<i>EBOV VP35 wt and mutants mammalian expression plasmid</i> .....	202
4.4.3	<i>Expression and purification of full-length wt and mutants EBOV rVP35</i> 203	
4.4.4	<i>Differential scanning fluorimetry analysis</i> .....	203
4.4.5	<i>Magnetic pull down assay</i> .....	204
4.4.6	<i>Molecular modeling</i> .....	205
4.4.7	<i>Cell line</i> .....	207
4.4.8	<i>EBOV VP35 luciferase reporter gene inhibition assay</i> .....	207
4.5	REFERENCES .....	208
<b>5</b>	<b>CHAPTER</b> .....	<b>219</b>
5.1	INTRODUCTION TO COVALENT DOCKING WITH AUTODOCK .....	219
5.2	METHODS.....	221
5.2.1	<i>Two point attractor method</i> .....	221
5.2.2	<i>Flexible sidechain method</i> .....	223
5.2.3	<i>Data set</i> .....	224
5.2.4	<i>Coordinate preparation and docking</i> .....	224
5.3	RESULTS AND DISCUSSION .....	225
5.4	CONCLUSIONS.....	234
5.5	REFERENCES .....	235
	<b>CONCLUDING REMARKS</b> .....	<b>239</b>

PUBLICATIONS .....	240
ORAL PRESENTATIONS.....	241
POSTERS .....	242
CONFERENCE PROCEEDINGS ABSTRACTS .....	243
PARTECIPATION AT SCHOOLS .....	243
GRANTS AND AWARDS.....	244
ACKNOWLEDGEMENTS .....	245

# 1 CHAPTER

## 1.1 INTRODUCTION TO COMPUTER AIDED DRUG DESIGN

During the last decades the interest on computational methodologies applied toward pharmaceutical targets became higher. These approaches became indispensable and increasingly combined with the traditional methods. In fact, Computer-aided Drug Design (CADD) methods, compared to the classical experimental procedures, allows to investigate the mechanism of action of various atomic or molecular phenomena with a remarkable saving of money and time.<sup>1</sup> Hence, from the '60 the improvements in this field have been remarkable and continuous in order to ensure reliable results related to the study of complex biological structures.<sup>2</sup> Previously computational studies were directed to the analysis of small molecules and were carried out by means of quantum mechanical approaches, based on the Schrödinger quantum mechanics concepts, then the advent of molecular mechanics made possible simulations of larger systems in reasonable time.<sup>3</sup>

In a drug discovery campaign, CADD is usually used for three major purposes:<sup>4,5</sup>

- 1) Filter large compound libraries into smaller sets of predicted active compounds that can be tested.<sup>6-9</sup>
- 2) Guide the optimization of lead compounds, to increase its affinity or to optimize drug metabolism and pharmacokinetics properties including absorption, distribution, metabolism, excretion and the potential toxicity (ADMET).<sup>4,10,11</sup>
- 3) Explain the molecular basis of therapeutic activity of compounds and rationalize it through the study of the Structure Activity Relationships (SAR).<sup>12</sup>

Of course there are also limitations and caveats to consider in the application of computational methods.<sup>1</sup> These have to be keep in mind as *in cerebro* element is essential in the application of such techniques.<sup>13</sup> It is famous the sentence of professor Hugo Kubinyi by warning students

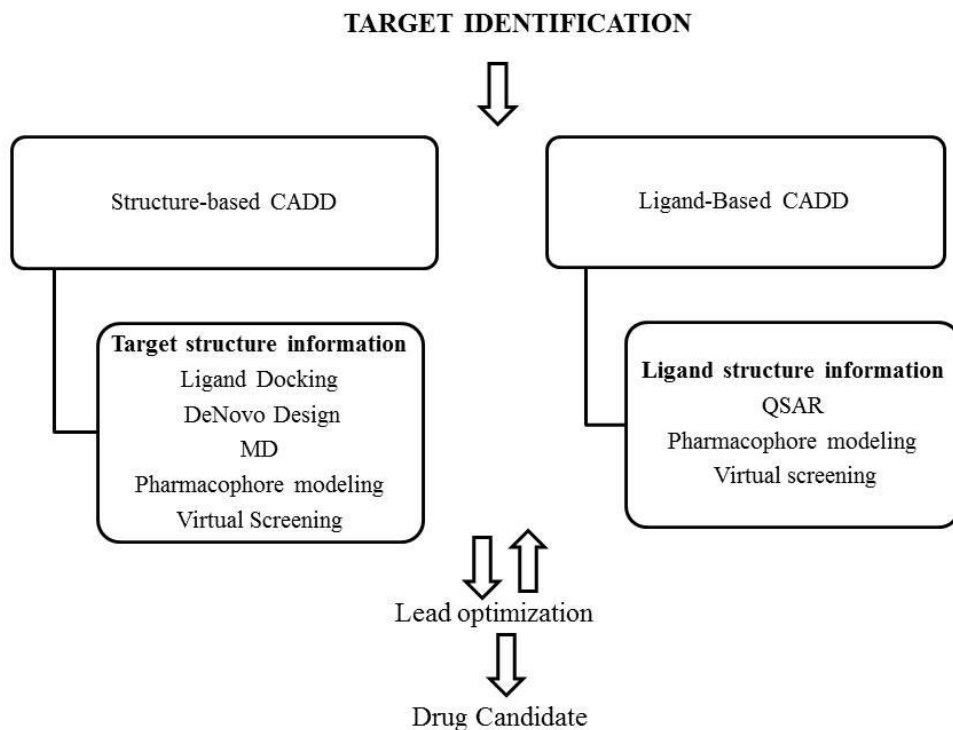
approaching this field : “Don’t switch off the brain when you switch on the computer!”.

Before gaining insight into each computational technique, it must be pointed out that CADD can be divided into two general categories: Structure-based and Ligand-based Drug Design.

Structure-based methods relies on the availability of structural data of the target macromolecule to calculate interaction energies for tested compounds,<sup>14</sup> whereas ligand-based methods exploit the knowledge of known active and inactive molecules through chemical similarity searches or construction of predictive, quantitative structure-activity relation (QSAR) models. The application of the latter ones is needed when no or little structural information of the target is available. Furthermore structure and ligand based methods can be also combined in more complex drug discovery workflows.<sup>15</sup>

The goal of these applications is to validate the activity of compounds *in vitro* and *in vivo* and to predict their putative binding site, ideally this should be then supported through a co-crystal structure.<sup>16</sup> Lead identification is often followed by several cycle of subsequent lead optimization using CADD.<sup>17</sup>





**Figure 1.** CADD in a Drug discovery pipeline and the available ligand-based and structure-based methodologies.<sup>18</sup>

In this dissertation I will focus my attention on structure-based methods which were applied for the different targets object of the studies carried out during my PhD.

## 1.2 BASIS OF COMPUTATIONAL CHEMISTRY

Computational chemistry, or molecular modeling, comprehends a series of techniques for investigating chemical problems on a computer. The investigations could interest: molecular geometry, energies of molecules, transition states, chemical reactivity and so on. In addition more complex problems can be studied, like the protein-protein interactions or the binding mode of a substrate/drug with an enzyme in order to design new potent drugs.

In order to describe a system we need four fundamental features:

- System description: type of units or “particles” and their number;
- Starting condition: information about particles position and their velocities;
- Interaction: mathematical forms necessary for describing the forces acting between the particles;
- Dynamical equation: mathematical form used to describe the behaviour of complex system evolving in time.

The choice of “particles” puts limitations on what we are ultimately able to describe: sub-atomic particle, atomic and residue level.

Computational methods can be divided into two big categories depending on whether they are based upon the concepts of quantum mechanics or molecular mechanics. Nowadays the molecular mechanicals methods are mostly used because they allow to analyse systems with tens of thousands atoms like proteins or DNA.

### 1.2.1 Basic theory of quantum mechanical methods

Quantum mechanical methods are mainly used in the computational chemistry field to collect essential parameters for the development of the molecular mechanicals methods.<sup>19</sup>

The reasons behind this poor use of these methods can be founded in their impossibility to study complex molecular systems. This branch of computational chemistry is driven by the Schrödinger equation which is shown here in the simplified form: <sup>3</sup>

$$H\psi = E\psi$$

Where  $H$  is the Hamiltonian operator,  $\psi$  a wave function and  $E$  the energy.

In this equation electrons are considered as particles that have an undulatory behaviour which mathematical description is derived from the wave function  $\psi$ . This is a probabilistic description of electrons behaviour. As such, it can describe the probability of electrons being in certain locations, but it cannot predict exactly where electrons are located.

The solution of the equation allows to establish the position of electrons and nuclei and to define their energies. Unfortunately for molecular systems, these solutions can be only approximated. Depending on the level of this approximation, quantum mechanical methods can be divided into *ab initio* and *semi-empirical* methods. These latter, unlike the *ab initio*, substitute some terms with parameters derived from experimental measurements or derived from *ab initio* calculations upon smaller models. It is easy to guess that the *ab initio* methods are characterized by a higher accuracy compared to the semi-empirical methods.

However this high accuracy requires a higher calculation cost, in fact with these methods only systems with tens of atoms can be studied, while the semi-empirical calculations allow to analyse even systems with hundreds of atoms.

In both methods nuclei and electrons are considered separately in order to investigate their space distribution and their energies.

Quantum mechanical methods are useful to calculate properties like:

- Energies of the molecular orbitals;
- Total electronic energy;
- Total repulsion energy of the nuclei;
- Heat formation;
- Atomic partial charges;
- Electrostatic potential;
- Dipole moment.

### 1.2.2 Basic theory of molecular mechanical methods

Molecular mechanical approaches are widespread in the medicinal chemistry field because of their ability to allow the molecular study of complex biological targets.

Despite what happens in the quantum mechanical methods, in the molecular mechanics, nuclei and electrons are not treated separately but they are considered as spheres with radius and electronegativity previously determined through experimental observations or quantum mechanics calculation. Since electrons are not explicitly included, electronic processes cannot be modeled, therefore only the fundamental state can be studied.<sup>19</sup>

The molecular mechanics energy expression consists of a simple algebraic equation for the energy of a compound. It does not use a wave function or total electron density. The constants in this equation are obtained either from spectroscopic data, other experimental techniques or *ab initio* calculations. The set of equations with their associated constants is called *force field (FF)*. The fundamental assumption of the molecular mechanics methods is the applicability of parameters to different molecules. This means that, the energetic penalty associated with a defined molecular motion (e. g. the stretching of C-C single bond) will be the same from one molecule to the next. This assumption allows a very simple calculation that can be applied to large molecular systems.

In molecular mechanic representations atoms are symbolized as balls and bonds as springs.

Within a molecular system there are various components of the total energies, such as the phenomena of *stretching*, *bending* and *torsion* of bonds. An important role have also the *non-bonding* interactions. The term

*non-bonding* interactions includes the hydrogen bonds, the van der Waals and electrostatic interactions.

Therefore the in a molecular mechanics simulation, the total energy of the system is the sum of all these components:

$$E_{tot} = E_s + E_b + E_t + E_{nb}$$

Where  $E_s$  is the stretching energy,  $E_b$  is the bending energy,  $E_t$  is the torsional energy and  $E_{nb}$  is the non-bonding energy.

Equations, that will be described later, and the necessary parameters for the calculation of the different energy components are included into specific database called *FF*. Therefore molecular mechanical methods are also referred as *FF* methods.

$E_s$  (stretching energy) is the sum of the forces that are exerted through the bond axis of two atoms and it is described by the Hook equation, where the energy is directly proportional to the bond length compared to the equilibrium position and it is defined by this equation:

$$E_s = \sum_{bonds} k_b (r - r_0)^2$$

Where  $K_b$  represents the elastic constant of the spring that mimics the bond,  $r_0$  the equilibrium length and  $r$  the bond strength. According to this equation, the energy trend appears to be parabolic.

$E_b$  (binding energy) describes the deformation of the bond angle compared to his optimal value. Like the  $E_s$  it is regulated by an equation based on the Hook equation:

$$E_b = \sum_{angles} k_\vartheta (\vartheta - \vartheta_0)^2$$

Where  $k_\vartheta$  indicates the elastic constant of the spring,  $\vartheta$  and  $\vartheta_0$  represent respectively, the reference position and the equilibrium one. The function that describes the bending energy has, like the stretching energy, a

parabolic trend that depends on the three atoms that define the bond angle.

Both properties have a similar behaviour and for both of them there is a direct proportionality between the elastic constants,  $k_s$  and  $k_\theta$ , and the energy variation as a function of the atoms shift.

The torsional energy ( $E_t$ ), instead, has a completely different behaviour because it depends on the positioning of four atoms and how this can affect the energy of the system. Moreover it is usually used to obtain total energy values of the system close to reference data, experimental or quantum mechanical, relative to the rotation around a bond. This is the equation that describes this component:

$$E_t = \sum_{torsion} A[1 + \cos(n\tau - \phi)]$$

Where  $A$  represents the coefficient obtained from the parametrization,  $n$  a symmetry factor,  $\tau$  the dihedral angle and  $\phi$  is the synchrony factor. Therefore, knowing that through an axes rotations of  $360^\circ$  are possible, the trend of the torsional energy is periodic.

In order to gain a detailed description of the variation of the total energetic profile of a system, it is essential to take into consideration the energy contributions due to the interaction of non-bonded atoms, such as hydrogen bonds, van Der Waals and electrostatic interactions.

Hydrogen bond takes place when particular conditions are fulfilled. First of all a donator, namely an electronegative atom to whom the hydrogen is bonded and an acceptor, represented by an electron-rich atom with, at least, one free electronic doublet must be present.

The distance between the hydrogen and the acceptor must be between 2.3 and 3.0 Å and the angle between the donator, the hydrogen and the acceptor must be between  $120^\circ$  and  $240^\circ$ . Moreover the angle between the hydrogen, the acceptor and the atom to whom this is directly bonded should not be less than  $90^\circ$ .

Another important non-bonded contribution is the van der Waals energy. This contribution is important when the distance between two

atoms is low, while loses its importance when this distance is higher. This energy is described by the Lennard-Jones equation:

$$E_{vdW} = \sum_{couple} \epsilon [(R_{ij}/r_{ij})^{12} - 2(R_{ij}/r_{ij})^6]$$

Where  $\epsilon$  represents the maximum attraction, considered as energy minimum, between the atom couple  $i$  and  $j$ ,  $R$  is the sum of their atomic ray and  $r$  their distance. In order to reduce the time calculations, the maximum value (*cut-off*) of the distance between atoms that can interact is set to 8 Å.

The electrostatic energy, instead, is able to describe long distance interactions and this plays an important role in complex systems. This type of contribution represents the charge-charge interaction and the dipole-dipole interactions.

This is described by a modified Coulomb law:

$$E_{el} = \frac{q_1 q_2}{\epsilon r_{12}}$$

Where the electrostatic term depends from the distance. As seen above for the electrostatic energy, the *cut-off* value is set to 20 Å.

As regards the charges of the atoms, they are calculated with quantum-mechanical *ab initio* or semi-empirical methods.

Molecular mechanics *FF* enable the exploration of the ligand-protein structure for rational design and other tasks and to model conformational changes and non-covalent interactions quite accurately. A successful *FF* in drug design should work well both for biological molecules and the organic molecules that interact with them.<sup>20</sup> The most used *FF* are: AMBER, CHARMM, OPLS and MMFF.

The AMBER force field, was primarily developed for protein and nucleic acid systems, and it is now widely used for the study of biological targets. In this force field, not all hydrogens are included, but only polar hydrogens, while hydrogens combined with carbon are combined into united atoms. Charges are derived from quantum chemistry calculations.

Force constants and idealized bond lengths and angles are taken from crystal structures and adapted to match normal mode frequencies for a number of peptide fragments. Finally torsion force constants are adjusted to match torsional barriers extracted from experiment or from quantum chemistry calculations. A general Amber force field (GAFF) was developed for organic molecules; it is designed to be compatible with existing Amber force field for protein and nucleic acids and has parameters for most organic and pharmaceutical molecules.<sup>21</sup>

As with AMBER, the CHARMM *FF* (Chemistry at Harvard using Molecular mechanics) was originally developed in the early 1980s and initially used as an extended atom force field with no explicit hydrogens. By 1985, this had been replaced by the CHARMM19 parameters, in which hydrogen atoms bonded to nitrogen and oxygen are explicitly represented, while hydrogens bonded to carbon or sulphur are treated as part of extended atoms.<sup>22</sup>

Another largely applied *FF* is OPLS (Optimized Potentials for Liquid Simulations). It was developed by Jorgensen and co-workers to simulate liquid state properties, initially for water and for more than 40 organic liquids. This force field places a strong emphasis on deriving non-bonded interactions by comparison to liquid state thermodynamics. The initial applications to protein used a polar-hydrogen only representation, taking the atom types and the valence (bond, angle, dihedral) parameters from the Amber force field. The parameter choices were intended to be “functional group friendly”, so that they could be easily transferred to other molecules with similar chemical groups.<sup>23</sup>

There are also other *FF* used for drug-like organic molecules like MMFF (Merck molecular force field).<sup>24</sup> This *FF* was developed in the 1990's to incorporate quantum mechanical calculations of parameters for a variety of functional groups. MMFFs and OPLS-2005 force fields have a good description of electrostatic interactions.

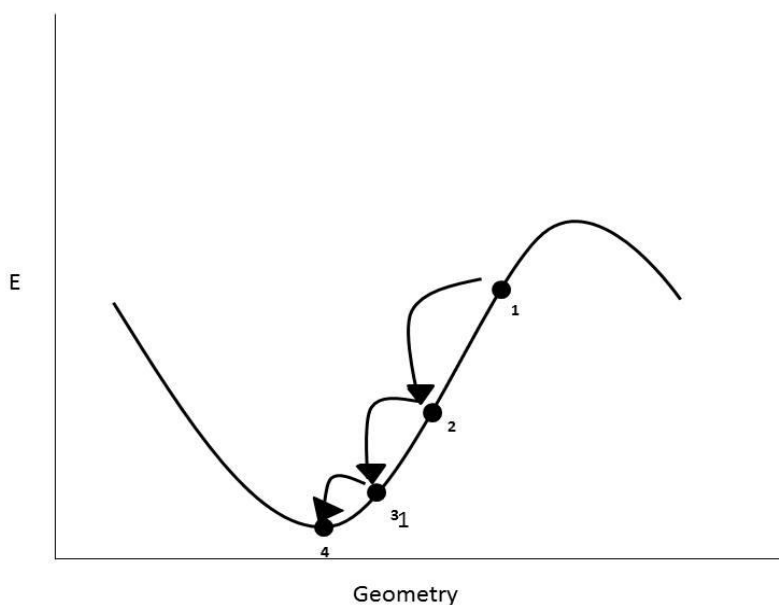
No force field has parameters for everything, therefore the reason to choose one instead another is considering the warnings about low-quality parameters issued in the log file.



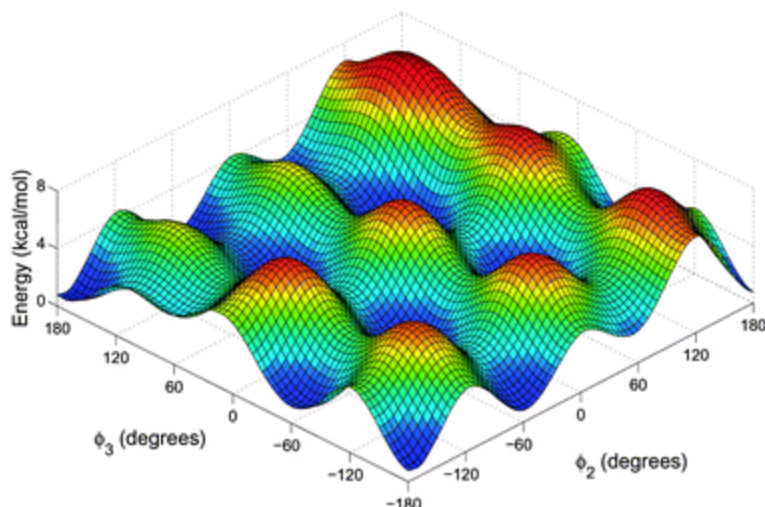
### 1.2.2.1 Energy minimization

When a molecule is built in a computational chemistry package, the initial geometry does not necessarily correspond to a stable conformer, therefore energy minimization is usually carried out to determine it.

Energy minimization is a numerical procedure for finding a minimum on the potential energy surface starting from a higher energy initial structure. During the minimization, the geometry is changed in a stepwise manner where the energy of the molecule is reduced from step 2, to 3 to 4 as shown in Figure 2. After a number of steps, a local or global minimum on the potential energy surface is reached. Most optimization methods determine the nearest stationary point, but a multidimensional function may contain many different stationary point of the same kind. The minimum with the lowest value is called global minimum, while all the other are local minima.



**Figure 2.** The process of energy minimization changes the geometry of the molecule in a step-wise fashion until a minimum is reached.



**Figure 3.** Amino acid potential energy at different conformations, minima on a potential energy surface are indicated in blue color.<sup>25</sup>

Most energy minimization methods proceed by determining the energy and the slope of a point. If the slope is positive the value of the coordinate is reduced until zero is reached, which means minimum is reached (Figure 2). There are numerous methods for varying the geometry to find the minimum. Many methods to find a minimum on the potential energy surface of a molecule use an iterative procedure to work in a step-wise mode.

- The *Newton-Raphson* method is the most computationally expensive per step of all the methods developed to perform energy minimization because it requires for each step the calculation of the matrix of the second derivative of the potential energy. This method it is able to retrieve in one step a stable conformation starting from distorted conformations but unluckily it becomes less efficient when it is applied to more complex systems.
- The *Steepest descent* method, rather than requiring the computation of numerous second derivatives, relies on an approximation. In fact, in this method, the second derivative is assumed to be a constant. This allows faster calculation than the previous. However, because of this approximation, it is not as efficient and therefore more steps are generally required to

find the minimum. It is named Steepest descendent because the direction in which the geometry is first minimized is opposite to the direction in which the gradient is largest at the initial point. Once a minimum in the first direction is reached, a second minimization is carried out starting from that point and moving in the steepest remaining direction. This process continues until a minimum has been reached in all directions within a sufficient tolerance.

- In the *Conjugate Gradient* method, the first portion of the search takes place in the opposite direction of the largest gradient, just as in the Steepest Descendent method. However, to avoid some of the oscillating back and forth that often plagues the steepest descendent method as it moves toward the minimum, the conjugated gradient method mixes in a little of the previous direction in the next search. This allows the method to move rapidly to the minimum. The equations for the conjugate gradient method are more complex than those of the other two methods.

#### **1.2.2.2 Conformational analysis**

In medicinal chemistry, it is important to predict a compound behaviour into a biological structure. Indeed only some conformations allow a good interaction with the biological target. In fact, it can happens that the interacting ligands can assume conformations that are not the minimum energy conformation, but, in the overall complex are possible because they can lead to a minor total energy, if compared with the sum of the energy of the isolated molecules: ligand and receptor.

Then, in order to comprehend the molecular mechanisms behind the activity of a drug it is of primary importance knowing the conformational properties of both the isolated molecule, and the molecule interacting with a biological target. The minimization can only lead to an optimization of the geometry and energy of a starting molecule, but does not allow the exploration of some possible conformational states. The conformational

space, is a space where are present some conditions of minimum energy (local minima) and only one condition of minimum absolute energy (global minima).

Different approaches are available for the conformational space analysis.<sup>26</sup> The most common strategies are systematic analysis, simulation methods (Monte Carlo and Molecular Dynamics) and genetic algorithm.

The systematic approach can be used only for the analysis of small molecules. This method consists in the exploration of variation of total energy of the system during the rotation of bonds. It is not possible to analyse complex systems because the calculation is not efficient because in order to find the minimum energy it is necessary the exploration of all the other conformations. Alternatively, the molecule can be fragmented and treated with systematic algorithm, this variant is call model building.

A more efficient method is the stochastic method, known as Monte Carlo method. It is a numeric method for the solution of mathematical problems by means of simulations based on variables, which value is attributed in a random way.

In the computational chemistry application there are special values attributed to the torsions of the system. The process is repeated for several times (steps) and depending on this number, conformations are generated. Some of them can be not realistic and should be excluded from the simulation. The number of conformations is a function of the applied steps, the number of torsions and the complexity of the energy surface. The Monte Carlo method, does not guarantee to find the global minimum of the energy. The probabilities are directly proportional to the number of steps and inversely proportional to the complexity of the system. This method has some limitations because if the system is really complex, a big number of steps is necessary and as a consequence copious calculation resources are required.<sup>27</sup>

Another methodology used for the exploration of the conformational space is the molecular dynamics. The main difference between MC and MD is in the way the conformational space is sampled. In MD methods, system configurations are given by integration of Newton's laws for motion over a small time-step, and new atomic positions and velocities are determined.<sup>28</sup> In some cases, a combination of these methods is used to

perform conformational search. Simulated annealing is an example. In this method, the system is initially set at a high temperature, which is gradually lowered until a configurational minimum is achieved. At each thermal step, equilibrium is reached by MC or MD implementation in the program.<sup>29</sup>

Also Genetic Algorithm (GA) can be used to successfully find low-energy conformations: a population of individuals (conformations) where individuals who are more “fit” (conformational energy) have a higher probability of surviving into subsequent generations. Therefore, in the course of a simulated evolution, the population produces conformations having increasingly lower energy.<sup>30</sup> As result of conformational search one can save the global minimum or a certain number of conformations (often considering an energetic threshold) since only the more stable conformations are of interest for 3D investigation methods.

### **1.3 STRUCTURE-BASED APPROACHES**

As previously said, computational chemistry can be divided into two big families, whether or not information about the target structure are available. In this dissertation I will discuss only about structure-based approaches, in order to match some theory with the applications in my work.

Macromolecules coordinates for structure-based approaches are derived from crystallographic experiments, or NMR spectroscopy, but there is an increasing interest also in high quality homology structures.<sup>31-33</sup> Regarding our experiments, we retrieved our targets 3D structures from the Protein Data Bank,<sup>34</sup> a free archive where 3D structures of proteins are available, with or without bounded ligands.

Herein three of the main structures-based approached will be discussed: Molecular docking, Molecular dynamics and 3D-Pharmacophore methods.

### 1.3.1 Molecular docking

Molecular docking is a key procedure in a structure-based drug design workflow. It is able to predict the ligand conformation and its orientation inside the target structure. The basis of molecular docking is the estimation of the free energy of binding of a ligand to a specific receptor site in a fixed environment. This tool is really useful but also has some scientific and mathematical issues. For example, it is currently much easier to calculate the energy/enthalpy of interaction than obtain the free energy because of the lack of efficient ways to obtain the entropic contributions. Secondly, the interactions of the ligand and the receptor with the solvent are not easy to estimate.

Docking can be applied at different levels of a Drug Discovery pipeline. It can be applied in a virtual screening workflow, to retrieve inside of a large compound library the best compounds able to interact with a single protein. The combination of docking experiments considering other targets or anti-targets (not desirable targets) could also help to find either compounds with multitarget properties, selective compounds, and compounds with less probability to have side effects.<sup>35-37</sup> Docking approach can be used to study protein-protein interactions<sup>38</sup> or protein-acid nucleic interactions.<sup>39, 40</sup> Furthermore it can give a rational explanation to biological activity and guide the lead optimization process,<sup>17</sup> or help to identify possible binding site inside a protein (blind docking).<sup>41</sup>

Another emerging application of docking method, often in combination with other approaches, is the study and prediction of compound metabolism.<sup>4, 42</sup>

Docking can be carried out by placing manually the small molecule into the binding pocket or by placing it automatically. The latter is the easiest and most modern way. In our projects automated docking was always performed.

Docking protocols consist of two steps: the first one is the search of the conformational space through a posing mechanism, where the ligand is placed inside the receptor in different orientations in order to identify the putative binding mode of the ligand. Several algorithms such as genetic algorithms, the Monte Carlo algorithm, evolutionary algorithms, simulated

annealing algorithms, empirical approaches, knowledge-based algorithms are used for the effective search of parameters space.<sup>43</sup> While the second step involves the assignment of an energy based score, a 'scoring function' is applied to evaluate the proposed binding modes referred as poses.

The purpose of the scoring function is to assess the binding affinity. However these scoring functions adopt various assumptions and simplifications. They can be listed as force-field based, empirical, knowledge-based and Consensus scoring functions.<sup>44</sup>

Classical force-field-based scoring functions use classical molecular mechanics to calculate the energy as the sum of the non-bonded (electrostatic and van der Waals) interactions through a Coulombic formulation and a Lennard-Jones potential function respectively. These functions use parameters derived from experimental data or *ab initio* calculations. The limit of these scoring functions is the slow computational speed. Extensions of force-field-based scoring functions consider the hydrogen bonds, solvation and entropy contributions.

In the empirical scoring functions, the binding energy is decomposed into several energy components, such as hydrogen bond interactions, ionic interactions, hydrophobic effect, desolvation effect and binding entropy. Empirical function terms are simple to evaluate and are based on the idea that binding energies can be approximated by a sum of individual uncorrelated terms experimentally determined.

Finally knowledge-based scoring functions use statistical analysis of ligand-protein complexes crystal structures to obtain the interatomic contact frequencies and distances between the ligand and protein. They are based on the assumption that more favourable an interaction is, the greater the frequency of occurrence will be. The score is calculated by favouring preferred contacts and penalizing repulsive interactions between each atom in the ligand and the protein within a given cut off.<sup>45</sup>

Sometimes scores can be used in parallel in order to find a consensus score between different scoring functions. Consensus scoring is a recent strategy that combines several different scores to assess the best docking conformations. A ligand could be accepted when its scores are well under a number of different scoring schemes. This allows to improve the reliability of docking scores and improving the probability of identifying accurately

docked poses.<sup>46</sup> In the combined approach each scoring function compensates for the other's weaknesses.<sup>47</sup>

It is important, before docking a molecule into the target structure, to validate the procedure. This consists usually in finding a complex of a known inhibitor with target structure from the PDB, and to dock this known inhibitor into the target (Re-Docking). The comparison of docked pose and experimental (i.e. calculating the Root Mean Square Deviation (RMSD)) can help to understand if the protocol is reliable. When this procedure is carried out considering different compounds available in different pdb entries of the same target, the validation is made through Cross-docking. The second validation is preferable whenever is possible and the structure of the target allows it (i.e. target not too flexible).<sup>9, 48</sup>

However it is always better to evaluate the goodness of the structure and check (if available) the electron density map of available complexes, since often ligands are not well defined.<sup>49-51</sup> Otherwise in case of bad pdb structure docking validation is questionable.

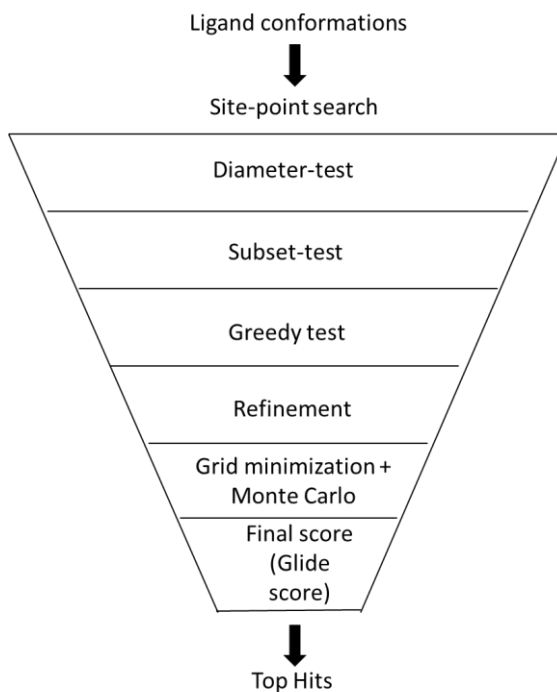
The docking programs that have been used for my PhD research work are Glide,<sup>52</sup> Autodock.<sup>53</sup>



**Table 1.** Most popular Docking programs.<sup>54</sup>

<b>Program</b>	<b>Algorithm</b>	<b>Reference</b>
AutoDock	Lamarckian Genetic Algorithm	53, 55
DOCK	Shape matching	56, 57
FlexX	Incremental construction	58
FRED	Shape matching (gaussian functions)	59
Glide	Descriptor matching/MC	52
GOLD	GA	60
FlapDock	Flexible fragment-based docking	61
Surflex Dock	Surface-based molecular similarity	62

Glide (Grid Based Ligand Docking with Energetics) uses a series of hierarchical filters to search for possible locations of the ligand inside the active site of a protein. The successive step produces a set of initial ligand conformations that are selected from an enumeration of the minima in the ligand torsion angle space (Figure 4). Given these conformations, initial screens are performed over the entire phase space available to the ligand to locate. Glide developed techniques of exhaustive systematic search, though approximations and truncations are required to achieve acceptable computational speed. Starting from the poses selected by the initial screening, the ligand is minimized in the field of the receptor using a standard molecular mechanics energy function: In this case, OPLS2005 force field. Finally, lowest poses obtained are subjected to a Monte Carlo procedure that examines nearby torsional minima.<sup>52</sup>



**Figure 4.** Glide docking “funnel”, showing the Glide docking hierarchy.

A novel protocol has been implemented in Glide: The QMPLD (Quantum Mechanics Polarized Ligand Docking), that uses *ab initio* methodology to calculate ligand charges within the protein environment.<sup>63</sup>

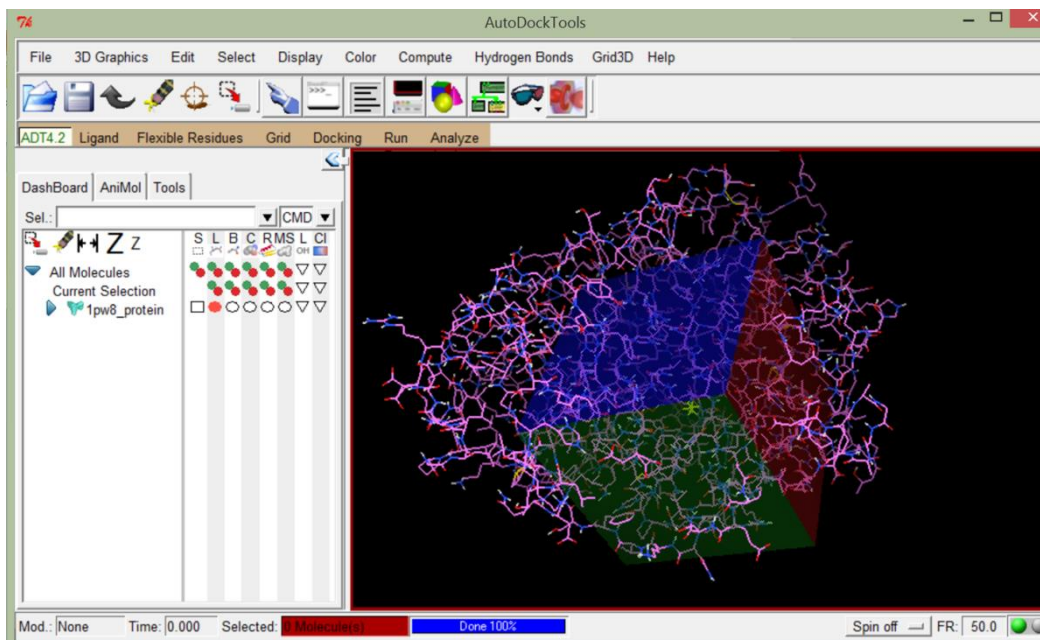
The QMPLD workflow consists of three steps: first, the protein-ligand complex is generated with Glide. Poses passed through these initial screens enter a stage, which involves the evaluation and minimization of a grid approximation to the OPLS non-bonded ligand-receptor interaction energy. Final scoring is then carried out on the energy-minimized poses. Finally, the minimized poses are rescored using Schrödinger’s proprietary GlideScore scoring function.

In the second step, a mixed quantum mechanical molecular mechanics method is used to compute the ligand charge distribution. For quantum mechanical molecular mechanics calculations, the QSITE program is used. The protein is defined as the MM region, and the ligand is defined as the QM region.<sup>64</sup> Evaluation is performed with NDDO (Neglect of Diatomic Differential Overlap) semiempirical method in MOPAC (Molecular Orbital PACKage) using Coulson charges.<sup>65</sup> In the third step, the ligand is submitted

to another Glide docking run where the ligand charges are substituted with the new charge sets calculated in the second step. This workflow was successfully applied for several targets.<sup>66</sup>

Another widely used docking program is Autodock. To allow searching of the large conformational space available to a ligand around a protein, Autodock uses a grid-based method that allows a rapid evaluation of the binding energy of trial conformations. In this method, the target protein is embedded in a grid (Figure 5). Then, a probe atom is sequentially placed at each grid point, the interaction energy between the probe and the target is computed, and the value is stored in the grid. This grid of energies may then be used as a lookup table during the docking simulation. The primary method for conformational searching is a Lamarckian genetic algorithm:<sup>67</sup> a population of trial conformations is created, and then in successive generations these individuals mutate, exchange conformational parameters, and compete in a manner analogous to biological evolution, ultimately selecting individuals with lowest binding energy. This allows individual conformations to search their local conformational space, finding local minima, and then pass this information to later generation. AutoDock uses a semi empirical free energy force field to predict free energy of small molecules to macromolecular target.<sup>53</sup>

With AutoDock Vina a higher speed of calculation is achieved.<sup>55</sup>



**Figure 5.** Graphical interface of AutoDockTools with the macromolecule embedded in a grid.

Many docking programs exist, however no one outperforms all others in all cases. Generally, programs are good on generating correct ligand poses (binding modes), but the scoring functions need improvement in order to improve the correlation between calculated and observed binding affinities. There are available different docking methodologies: rigid ligand and rigid receptor, flexible ligand and rigid receptor and flexible ligand and flexible receptor.<sup>68</sup>

When the ligand and receptor are both treated as rigid bodies, the searches space is very limited, considering only three translational and three rotational degrees of freedom. In this case, ligand flexibility could be addressed by using a set of ligand conformations.

Most common docking approaches treat the ligand as flexible while the receptor is kept rigid during docking because the computational cost is very high when the receptor is flexible. However flexibility can be taken into account modeling the conformational changes induced by ligand binding with the induced fit docking protocol. This procedure combines docking with sidechains rearrangements and minimization of the residues within the binding pocket.<sup>69</sup>

Other programs incorporate receptor flexibility in ligand docking by using 'soft' receptors, thus limiting penalties due to steric clashes, or allowing a selection of a few critical degrees of freedom in the binding site. Another strategy could be the selection (when is possible) of multiple and diverse receptor structures from the PDB to use for parallel docking experiments (Ensemble Docking).<sup>70</sup>

As previously mentioned a limit of the docking procedure is that most of the times the receptor is fixed and there is a limited treatment of the solvation effect. Therefore often the docking simulations are not able yield the best correlation between predicted affinity and biological activity of a small molecule.<sup>71</sup> To overcome this issue the docking simulations complexes are subjected to energy minimization and successive binding free energies calculations. The binding free energies are obtained by applying molecular mechanics and continuum or explicit solvation models using MM-PB/SA (Molecular mechanics/Poisson Boltzmann Surface Area) or molecular mechanics generalized Born/surface area (MM-GBSA) method.<sup>72</sup> These assume that the free energy change in a receptor-ligand binding process can be computed by only considering the difference between the unbound state and the bound state. The calculated free energy of the binding of ligands to proteins can be calculated out considering implicit (continuum) or explicit solvent. For the latter one a brief MD simulation is carried out and a set of snapshot of the protein-ligand complex structure are saved and rescored with either the PB/SA or GB/SA scoring functions, and the average interaction score of the snapshots is taken as the free energy of binding for the ligand.<sup>73</sup> When compared to docking scoring functions, the MM-GB/SA procedure is able to provide more accurate docking poses since the calculation of affinity is more precise.<sup>72</sup>

In the MM-GB(PB)SA formulation, the binding free energy of a ligand (L) to a protein (P) to form the complex (PL) is calculated as the difference:<sup>74</sup>

$$\Delta G_{\text{bind}} = G_{(\text{PL})} - G_{(\text{P})} - G_{(\text{L})}$$

The free energy of each of the three molecular systems (ligand protein and complex) can be obtained considering different contributions:

$$G_{(X)} = E_{MM(X)} + G_{solv(X)} - TS_{(X)}$$

In particular  $E_{MM}$  is the total molecular mechanics energy of a generic molecular system X in the gas phase, and it is calculated considering molecular mechanics energy function (or force field);  $G_{solv}$  is the solvation free energy, and S is the entropy of the system.

$E_{MM}$  can be considered decomposed in bonded (internal), non-bonded electrostatic and van der Waals energies:

$$E_{MM} = E_{bonded} + (E_{elec} + E_{vw})$$

The solvation free energy term  $G_{solv}$  is calculated considering both polar and non-polar contributions. The polar contributions are accounted for the generalized Born, Poisson, or Poisson-Boltzmann model, and the non-polar are assumed proportional to the solvent-accessible surface area (SASA)

$$G_{solv} = G_{PB(GB)} + G_{SASA}$$

Finally, conformational entropy S, is further decomposed into three parts, the translational, the rotational and the vibrational entropies.

Recent examples showed that MM-PBSA scoring can lead to an improvement compared to conventional scoring: re-ranking of the poses with MM-PBSA leads to a better separation between correct and incorrect poses.<sup>54</sup>

This post-docking procedure can be applied to compare docking results derived from different programs consensus docking. It consists in docking the compounds with different programs, performing post-docking procedure (minimization of complexes) and  $\Delta G$  calculation using MM-GBSA method, then best scores can be averaged and will help to select most promising compounds.

### 1.3.2 Molecular dynamics

Molecular dynamics was first developed in the late 1970<sup>75</sup> to overcome the computationally intensive quantum-mechanical calculations of the motions of big molecular systems. In fact this methodologies uses approximations based on the Newtonian physics to simulate atomic motions, thus reducing computational efforts.

Molecular dynamics is based on thermodynamic and kinetic concepts. It investigates the motion of atoms and molecules as a function of energy and time. These two quantities, constitute a conformational change in the molecule. The move of the atoms takes place in response to a force application described by the second law of Newton:

$$F = m \times a$$

Where  $F$  is the global force exercised upon the atom,  $m$  the mass of the atom and  $a$  its acceleration. The purpose of molecular dynamics is to calculate, for very short intervals of time, the force exercised upon the atoms and to use this data to derive the position of the atoms.

During a simulation, a speed in the term of kinetic energy, is assigned to each atom of the system. At the beginning, the assumption is that the system has a speed equal to zero.

To the system is applied energy in the term of heat, and this heat is transformed in kinetic energy and as a consequence the atoms get a motion. The force acting upon the atom can be obtained by analyzing the variations of the total energy during short movements. For a generic atom, the equation can be expressed in this way:

$$F_i = - \frac{dE}{dr_i}$$

Where  $F_i$  is the force acting on the atom,  $dE$  the variation of energy and  $dr_i$  the shift of the atom.

The data relative to the energy can be calculated with quantum mechanical methods, but normally are calculated with molecular mechanics

methods. Knowing the force and the mass relative to the atom, is possible to calculate their position for short intervals of time, in the femtosecond order.

The algorithm of the simulation proceeds first at the calculation of the acceleration, obtained from the product of the force and the mass of the atom  $i$ :

$$a_i = F_i \times m_i$$

Then, the velocity is obtained as the product between the acceleration and the time variation:

$$v_i = a_i \times dt$$

Finally, the atom position is obtained from the product between the velocity and the time variation:

$$d_r = v_i \times dt$$

For each iteration, where the starting point coincides with the arrival of the previous, a positioning is obtained and the set of the positioning constitutes the trajectory assumed by the atom during the simulation.

The heat and the duration of the simulation are really important in determining the efficacy of the simulation. In fact, the energy barriers that separate the different states of minimum energy can be exceeded with the applied heat. It is then, necessary to apply a sufficient temperature but not so high because that can cause a damage in the integrity of bonds. As the heat, also the duration of the simulation is important, because the exploration of the system, in particular the number of conformational and configurational conversions, depends on the duration of the simulation. When there are not variations of the energetic values or there is a periodic trend, it can be affirmed that the time of conformational search is sufficient.

The Energy (E), as a function of the positioning of all the atoms of the system, is obtained from two components (Figure6):



$$E_{total} = \underbrace{\sum_{bonds} K_r (r - r_{eq})^2 + \sum_{angles} K_\theta (\theta - \theta_{eq})^2 + \sum_{dihedrals} \frac{V_n}{2} [1 + \cos(n\phi - \gamma)]}_{\text{Bonded}} + \underbrace{\sum_{i < j} \left[ \frac{A_{ij}}{R_{ij}^{12}} - \frac{B_{ij}}{R_{ij}^6} + \frac{q_i q_j}{\epsilon R_{ij}} \right]}_{\text{Non-bonded}}$$

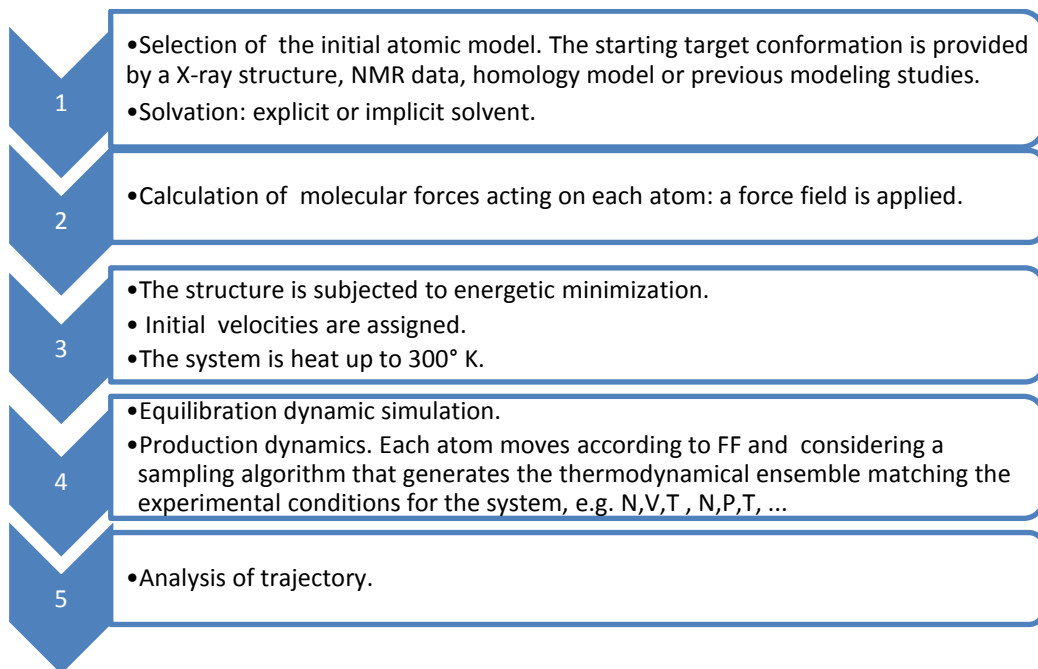
**Figure 6.** Calculation of the total Energy as sum of bonded and non-bonded term.<sup>76</sup>

$E_{bonded}$  takes into account all the interactions between two covalent bonded atom that depend on the length, angles and rotations of the bonds.  $E_{non-bonded}$  instead, takes into account all the interactions between atoms non-covalently bonded such as Van der Waals and Coulomb interactions.<sup>77</sup>

In order to reproduce the actual motion of molecules within a biological system, the energy terms described above are parametrized to fit quantum-mechanical calculations and experimental data. This parameters are again the Force fields that are able to describe the contributions of the various atomic forces that govern molecular dynamics. Several force fields are commonly used including AMBER, CHARMM and GROMOS. In the micro-canonical, or NVE ensemble, the system is isolated from changes in moles (N), volume (V) and energy (E). It corresponds to an adiabatic process with no heat exchange.

In the canonical ensemble, amount of substance (N), volume (V) and temperature (T) are conserved. A variety of thermostat algorithms are available to add and remove energy from the boundaries of a MD simulation.

In the isothermal–isobaric ensemble, amount of substance (N), pressure (P) and temperature (T) are conserved (NPT). In addition to a thermostat, a barostat is needed. It corresponds most closely to laboratory conditions with a flask open to ambient temperature and pressure.



**Figure 7.** Process of molecular dynamic simulation.

Molecular dynamics approach can be applied to every type of system but it has some limitations. In fact this method is affected by the complexity of the model and by the long duration of the simulations. Another limitation to consider is that the force fields used are also approximations of quantum-mechanical calculations.

Nevertheless molecular dynamics is unique in capturing dynamic events in biological systems. Many pharmaceutical phenomena of scientific interest occur on time scale that are computationally demanding but MD it is able to elucidate these important biological processes. Often MD has been used coupled with other tools, such as docking.<sup>76</sup>

### 1.3.3 Pharmacophore modeling

The concept of pharmacophore is based on the assumption that the interaction between a ligand and a biological structure is possible when some features are complementary. These feature are hydrogen-bond

donors and acceptors, positively and negatively charged or polarizable, hydrophobic regions or metal-ion interactions.

In particular:

1. The pharmacophore describes the steric and electronic characteristics determining the interaction within the biological target.
2. The pharmacophore does not represent a real molecule but an abstract concept that takes into account common features between the compound and its target structure.<sup>78</sup>

This can be summarized by Wermuth definition of pharmacophore: “A pharmacophore is the ensemble of steric and electronic features that is necessary to ensure the optimal supra-molecular interactions with a specific biological target and to trigger (or block) its biological response”.<sup>79</sup>

The most frequent use of pharmacophore approach is in virtual screening, where it is possible to filter large libraries of compounds based on the features selected. Several programs are used, such as Catalyst, Phase, LigandScout, Galahad, Flap and the pharmacophore module on Moe.<sup>80, 81</sup>

The use of pharmacophore models presents some advantageous characteristics: the models are universal because they represent chemical functions valid for all the molecules. Furthermore, they are computationally efficient: in fact because of their simplicity they can be used for large scale virtual screening.

A pharmacophore model can be built starting from ligand information or from target structure information. In the first case this can be achieved through the exploration of conformational space.

When information about the target structure are available, it is possible to build a pharmacophore model through GRID interaction fields that convert regions of high interaction energy into pharmacophore point locations and constrains.<sup>82</sup>

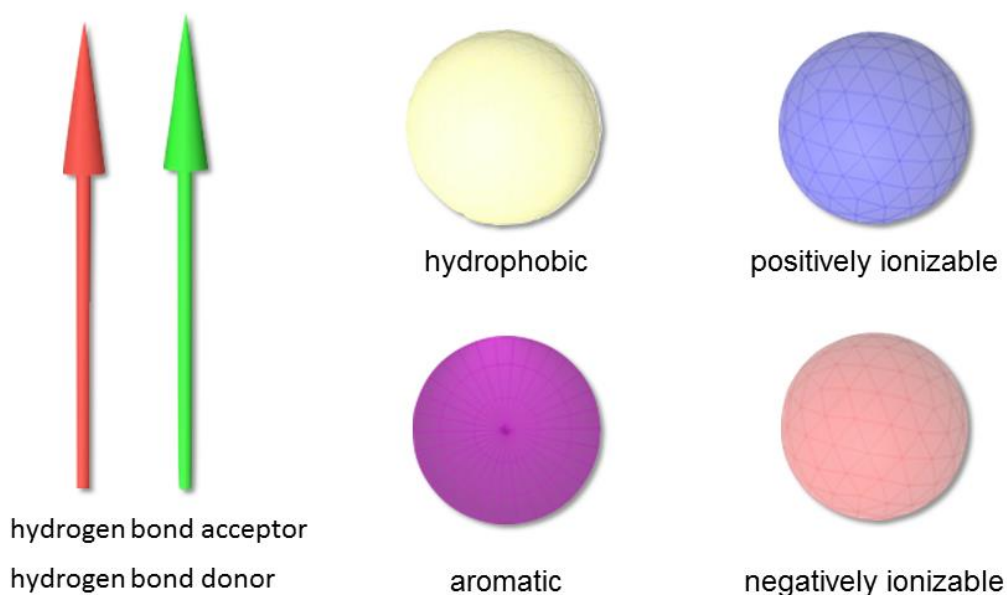
In alternative, starting from ligand-structure complex it is possible to convert interaction patterns into pharmacophore point locations and constraints.<sup>83</sup>

Structure-based methods in pharmacophore-modeling aim to be complementary to docking procedure, but are less demanding in terms of computational effort and much more efficient.

During this project we applied this method by means of LigandScout software.<sup>83</sup>

The first step, during the building of a pharmacophore model is the ligand perception and interpretation. This is performed through the perception and correction of plausible molecular topology including ring perception and through the interpretation and subsequent assignment of hybridization states and bonds types from geometrical information.

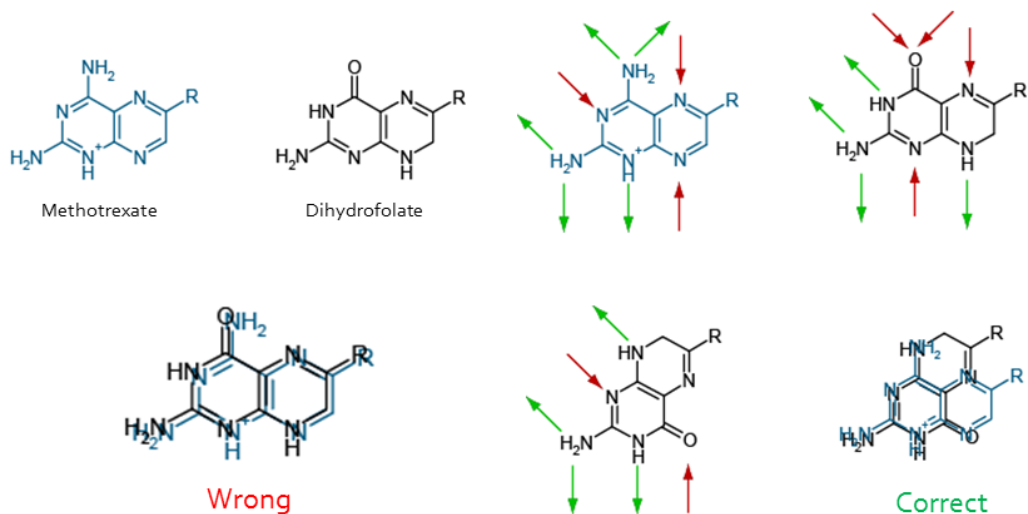
After these preliminary steps, the pharmacophore model can be built with the introduction of general chemical features definitions such as charge transfers, lipophilic groups and H-bond interactions that are able to describe the binding mode in a general way. The resulting model is a universal model that can lack of selectivity. Therefore, in the common pharmacophore creation and validation process, general feature descriptions are changed in order to improve the selectivity. The pharmacophore tool created in LigandScout is useful to create a pharmacophore model that is still universal but yet selective enough to reflect the specific ligand-receptor interactions. The features that describe a pharmacophore are: Hydrogen-bond interactions, hydrophobic areas, aromatic  $\pi$ - $\pi$  interactions and charge transfer interactions (Figure 8).



**Figure 8.** Pharmacophoric features representation by LigandScout.

After the recognition of the feature in the ligand, the successive step, is the search for corresponding features in the protein, the addition of the interaction features to the model only if a corresponding feature pair is found in the complex and the addition of excluded volume spheres for opposite hydrophobic features.<sup>83</sup>

Ligand-based pharmacophores are built considering the information relative to active and inactive compounds. Programs able to do it align the molecules in an automated way completely unbiased (Figure 9). This method allows medicinal chemists to see the key features in their small molecules data-set. After the building of the pharmacophore (ligand or structure based) the user can easily adapt his model, by omitting features for example considering only common feature (shared-pharmacophore) or by merging selective pharmacophore models. This interesting method can be used alone or in combination with other computational methods in complex workflows in successive or parallel way.



**Figure 9.** Example of wrong and correct alignment of pharmacophore points.<sup>84</sup>

## 1.4 REFERENCES

1. Kapetanovic, I. M., Computer-Aided Drug Discovery and Development (CADD): in silico-chemico-biological approach. *Chem-Biol. Interact.* **2008**, 171, 165-176.
2. Lewars, E. G., *Computational chemistry: introduction to the theory and applications of molecular and quantum mechanics*. Springer Science & Business Media: 2010.
3. Schrödinger, E., Quantisierung als Eigenwertproblem. *Annalen der Physik* **1926**, 385, 437-490.
4. Kirchmair, J.; Goeller, A. H.; Lang, D.; Kunze, J.; Testa, B.; Wilson, I. D.; Glen, R. C.; Schneider, G., Predicting drug metabolism: experiment and/or computation? *Nat. Rev. Drug Discovery* **2015**, 14, 387-404.
5. Macalino, S. J. Y.; Gosu, V.; Hong, S.; Choi, S., Role of computer-aided drug design in modern drug discovery. *Arch. Pharmacol Res.* **2015**, 38, 1686-1701.
6. Braga, R. C.; Alves, V. M.; Silva, A. C.; Nascimento, M. N.; Silva, F. C.; Liao, L. M.; Andrade, C. H., Virtual Screening Strategies in Medicinal Chemistry: The State of the Art and Current Challenges. *Curr. Top. Med. Chem. (Sharjah, United Arab Emirates)* **2014**, 14, 1899-1912.
7. Shin, W.-H.; Zhu, X.; Bures, M. G.; Kihara, D., Three-dimensional compound comparison methods and their application in drug discovery. *Molecules* **2015**, 20, 12841-12862.
8. Xia, J.; Tilahun, E. L.; Reid, T.-E.; Zhang, L.; Wang, X. S., Benchmarking methods and data sets for ligand enrichment assessment in virtual screening. *Methods (Amsterdam, Neth.)* **2015**, 71, 146-157.

9. Kirchmair, J.; Distinto, S.; Schuster, D.; Spitzer, G.; Langer, T.; Wolber, G., Enhancing drug discovery through in silico screening: Strategies to increase true positives retrieval rates. *Curr. Med. Chem.* **2008**, *15*, 2040-2053.
10. Devereux, M.; Popelier, P. L. A., In silico techniques for the identification of bioisosteric replacements for drug design. *Curr. Top. Med. Chem.* **2010**, *10*, 657-68.
11. Vangrevelinghe, E.; Rudisser, S., Computational approaches for fragment optimization. *Curr. Comput.-Aided Drug Des.* **2007**, *3*, 69-83.
12. Wassermann, A. M.; Wawer, M.; Bajorath, J., Activity Landscape Representations for Structure-Activity Relationship Analysis. *J. Med. Chem.* **2010**, *53*, 8209-8223.
13. Kubinyi, H., Drug research: myths, hype and reality. *Nat. Rev. Drug Discov.* **2003**, *2*, 665-668.
14. Kalyanamoorthy, S.; Chen, Y. P., Structure-based drug design to augment hit discovery. *Drug Discov. Today* **2011**, *16*, 831-9.
15. Gregory, L. W.; Markus, A. L. Integrating structure-and ligand-based approaches for computer-aided drug design. In *In Silico Drug Discovery and Design*; Future Science Ltd: 2013, pp 190-202.
16. Klebe, G., Recent developments in structure-based drug design. *J. Mol. Med.* **2000**, *78*, 269-281.
17. Chen, L.; Morrow, J. K.; Tran, H. T.; Phatak, S. S.; Du-Cuny, L.; Zhang, S., From laptop to benchtop to bedside: structure-based drug design on protein targets. *Curr. Pharm. Des.* **2012**, *18*, 1217-39.
18. Sliwoski, G.; Kothiwale, S.; Meiler, J.; Lowe, E. W., Jr., Computational methods in drug discovery. *Pharmacol. Rev.* **2014**, *66*, 334-95.

19. Young, D., *Computational chemistry: a practical guide for applying techniques to real world problems*. John Wiley & Sons: 2004.
20. Ponder, J. W.; Case, D. A., Force fields for protein simulations. *Adv. Protein Chem.* **2003**, 66, 27-85.
21. Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A., Development and testing of a general amber force field. *J. Comput. Chem.* **2004**, 25, 1157-74.
22. Brooks, B. R.; Brucoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M., CHARMM: A program for macromolecular energy, minimization, and dynamics calculations. *J. Comput. Chem.* **1983**, 4, 187-217.
23. Jorgensen, W. L. OPLS Force Fields. In *Encyclopedia of Computational Chemistry*; John Wiley & Sons, Ltd: 2002.
24. Halgren, T. A., Merck molecular force field. I. Basis, form, scope, parameterization, and performance of MMFF94. *J. Comput. Chem.* **1996**, 17, 490-519.
25. Sturdy, Y. K.; Clary, D. C., Torsional anharmonicity in the conformational analysis of tryptamine. *Phys. Chem. Chem. Phys.* **2007**, 9, 2065-2074.
26. Morley, S. D. The Global Minimum Problem in Molecular Mechanics: Simulated Annealing and Related Techniques. In *Molecular Modelling and Drug Design*, Vinter, J. G.; Gardner, M., Eds.; Macmillan Education UK: 1994; Chapter 3, pp 89-136.
27. Chang, G.; Guida, W. C.; Still, W. C., An internal-coordinate Monte Carlo method for searching conformational space. *J. Am. Chem. Soc.* **1989**, 111, 4379-4386.



28. Jorgensen, W. L.; Tirado-Rives, J., Monte Carlo vs Molecular Dynamics for Conformational Sampling. *The Journal of Physical Chemistry* **1996**, 100, 14508-14513.
29. Naidoo, K. J.; Brady, J. W., The application of simulated annealing to the conformational analysis of disaccharides. *Chem. Phys.* **1997**, 224, 263-273.
30. Judson, R. S.; Jaeger, E. P.; Treasurywala, A. M.; Peterson, M. L., Conformational searching methods for small molecules. II. Genetic algorithm approach. *J. Comput. Chem.* **1993**, 14, 1407-1414.
31. Ilari, A.; Savino, C. Protein Structure Determination by X-Ray Crystallography. In *Bioinformatics*, Keith, J., Ed.; Humana Press: 2008; Vol. 452, Chapter 3, pp 63-87.
32. Göbl, C.; Tjandra, N., Application of Solution NMR Spectroscopy to Study Protein Dynamics. *Entropy* **2012**, 14, 581.
33. Meier, A.; Söding, J., Automatic Prediction of Protein 3D Structures by Probabilistic Multi-template Homology Modeling. *PLoS Comput. Biol.* **2015**, 11, e1004343.
34. Berman, Helen M.; Kleywegt, Gerard J.; Nakamura, H.; Markley, John L., The Protein Data Bank at 40: Reflecting on the Past to Prepare for the Future. *Structure* **2012**, 20, 391-396.
35. Ellingson, S. R.; Smith, J. C.; Baudry, J., Polypharmacology and supercomputer-based docking: opportunities and challenges. *Mol. Simul.* **2014**, 40, 848-854.
36. Garcia-Sosa, A. T.; Maran, U., Improving the Use of Ranking in Virtual Screening against HIV-1 Integrase with Triangular Numbers and Including Ligand Profiling with Antitargets. *J. Chem. Inf. Model.* **2014**, 54, 3172-3185.

37. Liu, X.; Zhu, F.; Ma, X. H.; Shi, Z.; Yang, S. Y.; Wei, Y. Q.; Chen, Y. Z., Predicting targeted polypharmacology for drug repositioning and multi-target drug discovery. *Curr. Med. Chem.* **2013**, *20*, 1646-1661.
38. Sheng, C.; Dong, G.; Miao, Z.; Zhang, W.; Wang, W., State-of-the-art strategies for targeting protein-protein interactions by small-molecule inhibitors. *Chem. Soc. Rev.* **2015**, *44*, 8238-8259.
39. Burge, R. G.; Martinez-Yamout, M. A.; Dyson, H. J.; Wright, P. E., Structural Characterization of Interactions between the Double-Stranded RNA-Binding Zinc Finger Protein JAZ and Nucleic Acids. *Biochemistry* **2014**, *53*, 1495-1510.
40. Robertson, T.; Varani, G. Prediction of protein-nucleic acid interactions. 2009; John Wiley & Sons, Inc.: 2009; pp 593-613.
41. Hetényi, C.; van der Spoel, D., Toward prediction of functional protein pockets using blind docking and pocket search algorithms. *Protein Science : A Publication of the Protein Society* **2011**, *20*, 880-893.
42. Kirchmair, J.; Williamson, M. J.; Tyzack, J. D.; Tan, L.; Bond, P. J.; Bender, A.; Glen, R. C., Computational Prediction of Metabolism: Sites, Products, SAR, P450 Enzyme Dynamics, and Mechanisms. *J. Chem. Inf. Model.* **2012**, *52*, 617-648.
43. Brooijmans, N.; Kuntz, I. D., Molecular Recognition and Docking Algorithms. *Annu. Rev. Biophys. Biomol. Struct.* **2003**, *32*, 335-373.
44. Wang, R.; Lu, Y.; Wang, S., Comparative Evaluation of 11 Scoring Functions for Molecular Docking. *J. Med. Chem.* **2003**, *46*, 2287-2303.
45. Kitchen, D. B.; Decornez, H.; Furr, J. R.; Bajorath, J., Docking and scoring in virtual screening for drug discovery: methods and applications. *Nat. Rev. Drug Discovery* **2004**, *3*, 935-949.

46. Houston, D. R.; Walkinshaw, M. D., Consensus Docking: Improving the Reliability of Docking in a Virtual Screening Context. *J. Chem. Inf. Model.* **2013**, 53, 384-390.
47. Cheng, T.; Li, X.; Li, Y.; Liu, Z.; Wang, R., Comparative Assessment of Scoring Functions on a Diverse Test Set. *J. Chem. Inf. Model.* **2009**, 49, 1079-1093.
48. Moitessier, N.; Englebienne, P.; Lee, D.; Lawandi, J.; Corbeil, C. R., Towards the development of universal, fast and highly accurate docking/scoring methods: a long way to go. *Br. J. Pharmacol.* **2008**, 153, S7-S26.
49. Kleywegt, G. J., On vital aid: the why, what and how of validation. *Acta Crystallographica Section D: Biological Crystallography* **2009**, 65, 134-139.
50. Kirchmair, J.; Markt, P.; Distinto, S.; Schuster, D.; Spitzer, G. M.; Liedl, K. R.; Langer, T.; Wolber, G., The Protein Data Bank (PDB), its related services and software tools as key components for in silico guided drug discovery. *J. Med. Chem.* **2008**, 51, 7021-7040.
51. Pozharski, E.; Weichenberger, C. X.; Rupp, B., Techniques, tools and best practices for ligand electron-density analysis and results from their application to deposited crystal structures. *Acta Crystallographica Section D* **2013**, 69, 150-167.
52. Friesner, R. A.; Banks, J. L.; Murphy, R. B.; Halgren, T. A.; Klicic, J. J.; Mainz, D. T.; Repasky, M. P.; Knoll, E. H.; Shelley, M.; Perry, J. K.; Shaw, D. E.; Francis, P.; Shenkin, P. S., Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. *J. Med. Chem.* **2004**, 47, 1739-49.

53. Morris, G. M.; Huey, R.; Lindstrom, W.; Sanner, M. F.; Belew, R. K.; Goodsell, D. S.; Olson, A. J., AutoDock4 and AutoDockTools4: Automated Docking with Selective Receptor Flexibility. *J. Comput. Chem.* **2009**, 30, 2785-2791.
54. Kroemer, R. T., Structure-based drug design: Docking and scoring. *Curr. Protein Pept. Sci.* **2007**, 8, 312-328.
55. Trott, O.; Olson, A. J., AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J. Comput. Chem.* **2010**, 31, 455-461.
56. Kuntz, I. D.; Blaney, J. M.; Oatley, S. J.; Langridge, R.; Ferrin, T. E., A geometric approach to macromolecule-ligand interactions. *J. Mol. Biol.* **1982**, 161, 269-88.
57. Gschwend, D.; Kuntz, I., Orientational sampling and rigid-body minimization in molecular docking revisited: On-the-fly optimization and degeneracy removal. *J. Comput-Aided. Mol. Des.* **1996**, 10, 123-132.
58. Rarey, M.; Kramer, B.; Lengauer, T.; Klebe, G., A fast flexible docking method using an incremental construction algorithm. *J. Mol. Biol.* **1996**, 261, 470-89.
59. McGann, M. R.; Almond, H. R.; Nicholls, A.; Grant, J. A.; Brown, F. K., Gaussian docking functions. *Biopolymers* **2003**, 68, 76-90.
60. Jones, G.; Willett, P.; Glen, R. C.; Leach, A. R.; Taylor, R., Development and validation of a genetic algorithm for flexible docking. *J. Mol. Biol.* **1997**, 267, 727-48.
61. Artese, A.; Cross, S.; Costa, G.; Distinto, S.; Parrotta, L.; Alcaro, S.; Ortuso, F.; Cruciani, G., Molecular interaction fields in drug discovery: recent advances and future perspectives. *Wiley Interdisciplinary Reviews: Computational Molecular Science* **2013**, 3, 594-613.

62. Jain, A. N., Surflex: fully automatic flexible molecular docking using a molecular similarity-based search engine. *J. Med. Chem.* **2003**, 46, 499-511.
63. Chung, J. Y.; Hah, J.-M.; Cho, A. E., Correlation between Performance of QM/MM Docking and Simple Classification of Binding Sites. *J. Chem. Inf. Model.* **2009**, 49, 2382-2387.
64. Murphy, R. B.; Philipp, D. M.; Friesner, R. A., A mixed quantum mechanics/molecular mechanics (QM/MM) method for large-scale modeling of chemistry in protein environments. *J. Comput. Chem.* **2000**, 21, 1442-1457.
65. Stewart, J., Optimization of parameters for semiempirical methods V: Modification of NDDO approximations and application to 70 elements. *Journal of Molecular Modeling* **2007**, 13, 1173-1213.
66. Cho, A. E.; Guallar, V.; Berne, B. J.; Friesner, R., Importance of accurate charges in molecular docking: Quantum mechanical/molecular mechanical (QM/MM) approach. *J. Comput. Chem.* **2005**, 26, 915-931.
67. Morris, G. M.; Goodsell, D. S.; Halliday, R. S.; Huey, R.; Hart, W. E.; Belew, R. K.; Olson, A. J., Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function. *J. Comput. Chem.* **1998**, 19, 1639-1662.
68. Totrov, M.; Abagyan, R., Flexible ligand docking to multiple receptor conformations: a practical alternative. *Curr. Opin. Struct. Biol.* **2008**, 18, 178-184.
69. Sherman, W.; Day, T.; Jacobson, M. P.; Friesner, R. A.; Farid, R., Novel Procedure for Modeling Ligand/Receptor Induced Fit Effects. *J. Med. Chem.* **2006**, 49, 534-553.

70. Huang, S.-Y.; Zou, X., Ensemble docking of multiple protein structures: Considering protein structural variations in molecular docking. *Proteins-Structure Function and Bioinformatics* **2007**, 66, 399-421.
71. Carlson, H. A.; McCammon, J. A., Accommodating protein flexibility in computational drug design. *Mol. Pharmacol.* **2000**, 57, 213-8.
72. Massova, I.; Kollman, P., Combined molecular mechanical and continuum solvent approach (MM-PBSA/GBSA) to predict ligand binding. *Perspectives in Drug Discovery and Design* **2000**, 18, 113-135.
73. Godschalk, F.; Genheden, S.; Soderhjelm, P.; Ryde, U., Comparison of MM/GBSA calculations based on explicit and implicit solvent simulations. *Phys. Chem. Chem. Phys.* **2013**, 15, 7731-7739.
74. Genheden, S.; Ryde, U., The MM/PBSA and MM/GBSA methods to estimate ligand-binding affinities. *Expert Opinion on Drug Discovery* **2015**, 10, 449-461.
75. McCammon, J. A.; Gelin, B. R.; Karplus, M., Dynamics of folded proteins. *Nature* **1977**, 267, 585-590.
76. Durrant, J. D.; McCammon, J. A., Molecular dynamics simulations and drug discovery. *BMC Biol.* **2011**, 9, 1-9.
77. Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A., A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *J. Am. Chem. Soc.* **1995**, 117, 5179-5197.
78. Mannhold, R.; Kubinyi, H.; Folkers, G.; Langer, T.; Hoffmann, R. D., *Pharmacophores and pharmacophore searches*. John Wiley & Sons: 2006; Vol. 32.

79. Wermuth, C.; Ganellin, C.; Lindberg, P.; Mitscher, L., Glossary of terms used in medicinal chemistry (IUPAC Recommendations 1998). *Pure Appl. Chem.* **1998**, 70, 1129-1143.
80. Markt, P.; Schuster, D.; Langer, T. Pharmacophore Models for Virtual Screening. In *Virtual Screening*; Wiley-VCH Verlag GmbH & Co. KGaA: 2011, pp 115-152.
81. Cross, S.; Baroni, M.; Carosati, E.; Benedetti, P.; Clementi, S., FLAP: GRID Molecular Interaction Fields in Virtual Screening. Validation using the DUD Data Set. *J. Chem. Inf. Model.* **2010**, 50, 1442-1450.
82. Ortuso, F.; Langer, T.; Alcaro, S., GBPM: GRID-based pharmacophore model: concept and application studies to protein-protein recognition. *Bioinformatics* **2006**, 22, 1449-55.
83. Wolber, G.; Langer, T., LigandScout: 3-D Pharmacophores Derived from Protein-Bound Ligands and Their Use as Virtual Screening Filters. *J. Chem. Inf. Model.* **2005**, 45, 160-169.
84. Sippl, W. Chapter 28 - Pharmacophore Identification and Pseudo-Receptor Modeling. In *The Practice of Medicinal Chemistry (Third Edition)*, Wermuth, C. G., Ed.; Academic Press: New York, 2008, pp 572-586.

## **2 CHAPTER**

### **2.1 INTRODUCTION TO REVERSE TRANSCRIPTASE OF HIV-1**

According to UNAIDS, at the end of 2015, an estimated 37 million people were living with HIV worldwide. However the scenario is more optimistic than at the beginning of 2000s. In fact the number of people dying of AIDS-related causes fell to 1.2 million in 2014. Furthermore new HIV infections have fallen by 35% since 2000 (by 58% among children) and AIDS-related deaths have decreased by 42% since the peak in 2004. As of June 2015, 15.8 million people were accessing treatment. These results were considered impossible 15 years ago. The main priority for the next 15 years is ending the AIDS epidemic by 2030.<sup>1</sup>

After the identification of HIV-1 as the causative agent of AIDS, more than 30 antiretroviral drugs have been approved for the clinical treatment of HIV infected patients targeting different steps of the HIV replication cycle.<sup>2</sup>

However, although a number of drugs reached the market, there are still problems associated with drug toxicity and resistance, therefore the development of new HIV drugs is an ongoing process.

In fact HIV virus is able to evolve sufficiently rapidly that, if the therapy is not well designed, resistance will develop in treated patients. A shortcut to overcome this problem is to use a combination of drugs to completely block the viral replication. Thus different drugs can be associated in the so called Highly Active Antiretroviral Therapy (HAART). This type of therapy changed the prognosis of HIV infected patients from high probability of mortality to a chronic infection.

Usually these drugs are able to inhibit the virus replication cycle at different levels: by inhibiting the Reverse transcriptase, the Protease and the Integrase.<sup>3</sup>

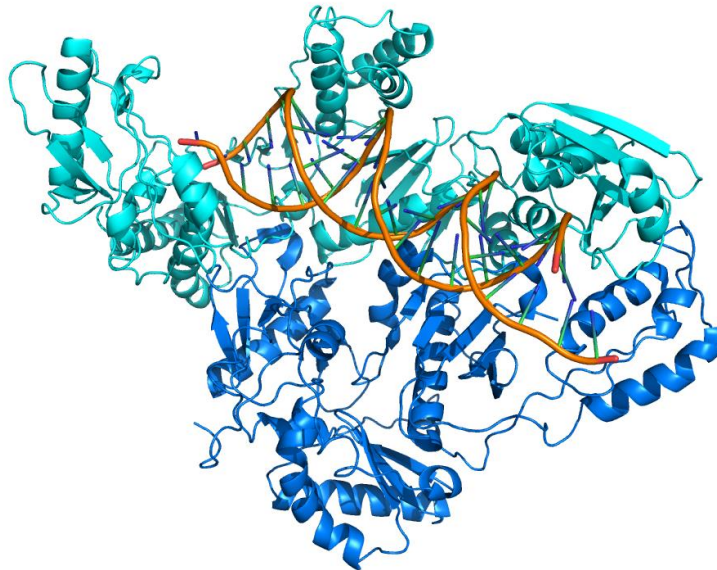
To overcome toxicity and resistance of existing drugs, new drugs characterized by a new mechanism of action are strongly needed.

The replication cycle of HIV consists of different steps that are targeted by these drugs, leading to an inhibition of virus replication.<sup>2</sup> Among these, one of the most attractive and explored target is the Reverse Transcriptase (RT) which is responsible for the retrotranscription. This crucial step of the HIV converts the viral single-stranded RNA genome into



integration-competent double stranded DNA through the formation of a RNA/DNA hybrid intermediate. This process requires both viral and cellular elements, among which the most important is the virus-coded RT protein.

RT consists of two subunits of different lengths, p66 and p51 which are combined in a stable asymmetric heterodimer.<sup>2</sup> The p66 domain possess the catalytic activity, while the p51 it is important in keeping the p66 in the correct folding.



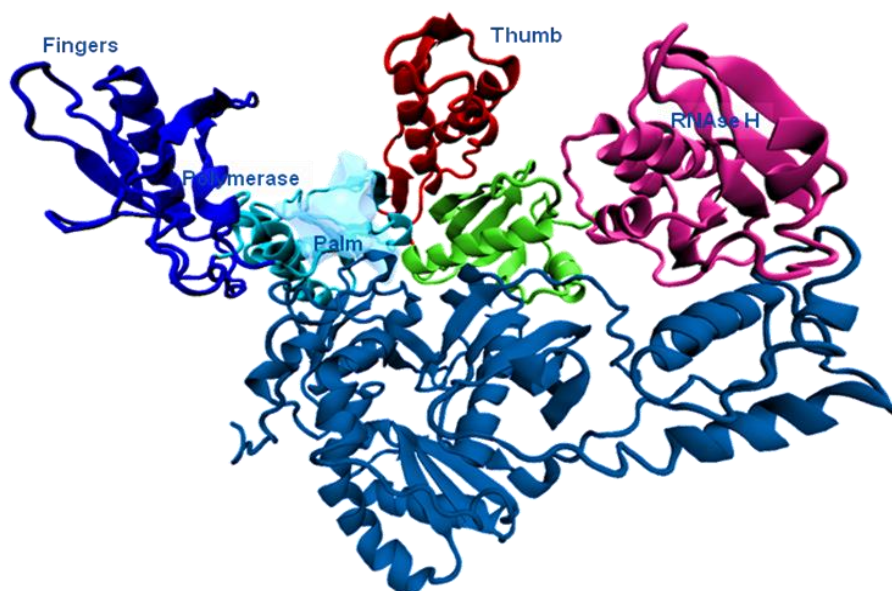
**Figure 1.** Structure of Reverse Transcriptase complexed with polypurine tract RNA:DNA (1hys pdb)<sup>4</sup>. In light blue the p66 subunit and in blue the p51 subunit.

The subunit p66 comprises different domains: the polymerase domain, the connection domain and the RNase H domain. The polymerase domain has the RNA-polymerase DNA-dependent (RDDP) function, which is able to synthesize double stranded DNA starting from viral RNA. While the RNase H domain has a RNase-Hydrolase function, which task is to cleave the RNA from the double strand RNA-DNA. This conversion takes place in the cytoplasm of the infected cell and after DNA synthesis has been completed, the resulting linear double-stranded viral DNA is translocated to the nucleus where the viral DNA is inserted into the host genome by the Integrase.

To exploit its function, the polymerase domain holds the nucleic acid in a cleft formed by its subdomains as a hand. Commonly the four

subdomains are identified as fingers, palm, connection and thumb (Figure 2). P51 folds into the same four subdomains like the polymerase domain, however the positions of the subdomains relative to each other are different and the absence of the RNase H domain make this subunit different from the p66 subunit.

The connection and thumb subdomains of p51 form the floor of the binding cleft and throughout this the nucleic acid is directed towards RNase H domain.<sup>3</sup>



**Figure 2.** Structure of Reverse Transcriptase with indication of domains and subdomains.

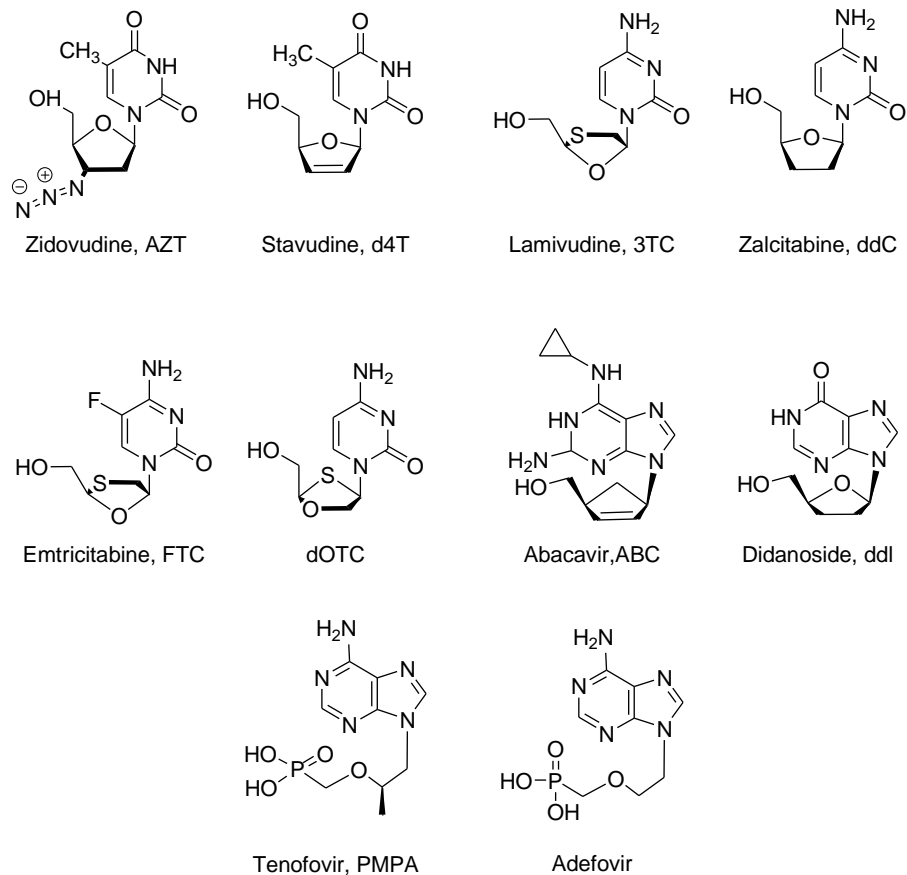
Currently two different classes of Reverse Transcriptase, targeting the RDDP function, have been approved for the treatment of HIV-1: nucleoside/nucleotide (NRTIs/NtRTIs) and non-nucleoside RT inhibitors (NNRTIs). NRTIs are analogs of the natural substrate (dNTP) and inhibit RDDP function by a competitive mechanism at the active site.

To act on such a mechanism, these inhibitors must necessarily lack a free OH group in 3' position. Zidovudine (AZT, 3'-azidothymidine) was identified as the first NRTI acting as pro-drug. It requires successive phosphorylation steps and operates through its triphosphate metabolite.<sup>5</sup>

The active form of the drug is used as a false substrate during reverse transcription of viral RNA.

Currently eight NRTIs are clinically available, structurally resembling either pyrimidine or purine analogues.

In the pyrimidine nucleoside analogues both thymidine and cytosine analogues are included. The most representative thymidine analogues are 3'-azido-2',3'-dideoxythymidine (zidovudine, AZT) and 2',3'-didehydro-2',3'-dideoxythymidine (stavudine, d4T), while cytosine analogues are (-)-2',3'-dideoxy-3'-thiacytidine (lamivudine, 3TC), (-)-2',3'-dideoxy-5-fluoro-3'-thiacytidine (emtricitabine, FTC) and [(-)-2'-deoxy-3'-oxa-4'-thiacytidine] (dOTC), 2',3'-dideoxycytidine (zalcitabine, ddC) which, however, is no longer recommended due to its peripheral neuropathy side effect,. Purine nucleoside analogues include (1S-4R)-4-[2-amino-6(cyclopropylamino)-9H-purin-9yl]-2-cyclopentane-1-methanol (abacavir, ABC) and 2',3'-dideoxyinosine (didanosine, ddl) as guanosine and adenine analogues, respectively (Figure 3).



**Figure 3.** NRTIs clinically available

Unfortunately drug resistant viral mutants can gain a competitive advantage over wild type virus under selective drug pressure, almost becoming the dominant species.

Generally two different mechanisms lead to HIV-1 resistance to NRTIs. The first consists of NRTI discrimination leading to a reduction of incorporation rate, the second consists of NRTI excision that unblocks NRTI-terminated primers. Typically discrimination occurs due to steric hindrance leading to a selective alteration of the NRTI binding and/or incorporation rate.<sup>6</sup>

Regarding NRTI excision, it is mostly increased through mutations, located around the dNTP binding pocket and also in terms of thymidine analogues mutations (TAMs).

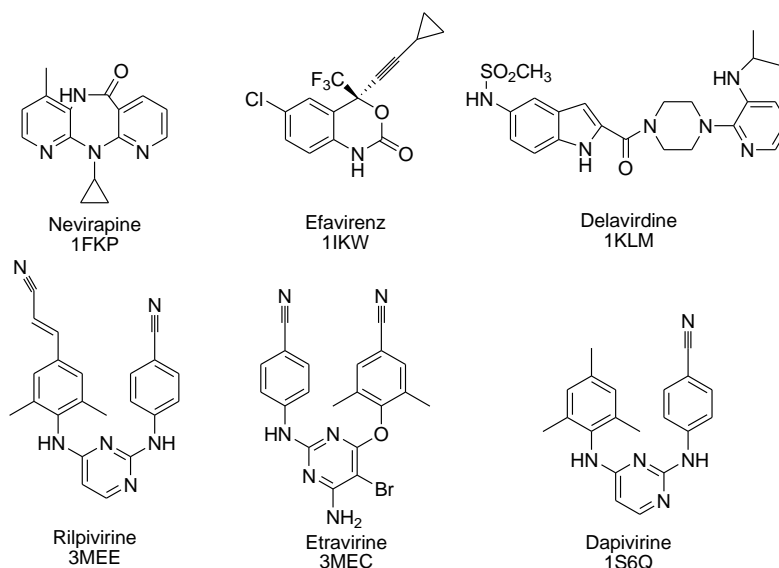
NRTI resistance could also be conferred by mutations in the connection and RNase H domains.<sup>7-11</sup>

The NtRTIs, such as adefovir [9-(2-phosphonyl-methoxyethyl) adenine (PMEA)] and tenofovir [(R)-9-(2-phosphonyl-methoxypropyl) adenine (PMPA)] are acyclic phosphonate analogues of adenine, therefore, only need two phosphorylation steps to be converted into the active drug.

As the NRTIs, they act as obligatory chain terminators.<sup>12</sup>

Along with exploration of the NRTI binding pocket to obtain analogous molecules with improved drug-like properties and effective against many NRTI drug-resistant RT variants,<sup>13</sup> recently, two families of compounds have been reported as new class of “nucleotide-competing RT inhibitors” (NcRTIs). The first is represented by indolopyridones (INDOPY) derivatives. Although structurally different from classical NRTI it seems that this series can occupy the active site of the enzyme (or a site in close proximity) and competes with natural dNTP substrates.<sup>14, 15</sup> The second class includes 4-dimethylamino-6-vinylpyrimidine derivatives (DAVP) whose binding site is close to the polymerase active site.<sup>16</sup>

In contrast with the NRTI class, NNRTIs are a family of compounds characterized by a high variety of structures (Figure 4).



**Figure 4.** NNRTIs in the market and under clinical evaluation and relative examples of crystal structures pdb codes

Unlike the other class of compounds, they do not need intracellular activation. Many different classes of NNRTIs could be distinguished and five drugs, acting as NNRTIs, have been approved for HIV-1 treatment so far.

It should be observed that, while in the case of first generation NNRTIs, like delavirdine and nevirapine, single mutations (Y181C, K103N and Y188C) could lead to drug resistance, in the case of the more bulky second generation NNRTIs, like efavirenz and dapivirine, more than one mutation is generally required to induce drug resistance.

More than 30 NNRTIs have been reported even though resistance and toxicity are some of the most important disadvantages of this class of compounds.<sup>17, 18</sup> More recently, molecules with a higher flexibility, although less favoured by a thermodynamic point of view, have been proposed as NNRTIs.<sup>19, 20</sup> Molecules like etravirine and rilpivirine are successful examples of this new approach as well as dapivirine actually under clinical evaluation.<sup>21</sup>

NNRTIs cause a distortion of the protein structure that inhibits the polymerase activity. They bind in a hydrophobic pocket (non-nucleoside inhibitors binding pocket, ) located in the palm domain of the p66 subunit of the heterodimeric RT, approximately 10 Å from the catalytic site of the enzyme. This pocket contains the side chains of aromatic and hydrophobic amino acid residues Y181, Y188, F227, W229, Y318, P95, L100, V106, V108, V179, L234, and P236 from the p66 subunit. It is flexible and its conformation depends on the size, shape, and binding mode of the different NNRTIs. It can accommodate a space of about 620-720 Å<sup>3</sup>, which is approximately more than twice the volume occupied by most of the present NNRTIs.<sup>2</sup> This explains the large variety of chemical scaffolds of this class of inhibitors whose shapes inspired authors to create imaginative names to describe them (e. g “butterfly”,<sup>22</sup> “horseshoe”,<sup>19</sup> and “dragon”<sup>23</sup>).

All these classes of compounds act by inhibiting the DNA-polymerase of the Reverse Transcriptase, while in therapy no drugs targeting the RT associated RNase H activity are available, even though some inhibitors have recently designed and studied.<sup>24</sup>

This target is crucial for the virus infectivity and it was reported that the block of this function through a single-point mutation could lead to a total loss of the virus infectivity.<sup>25</sup>

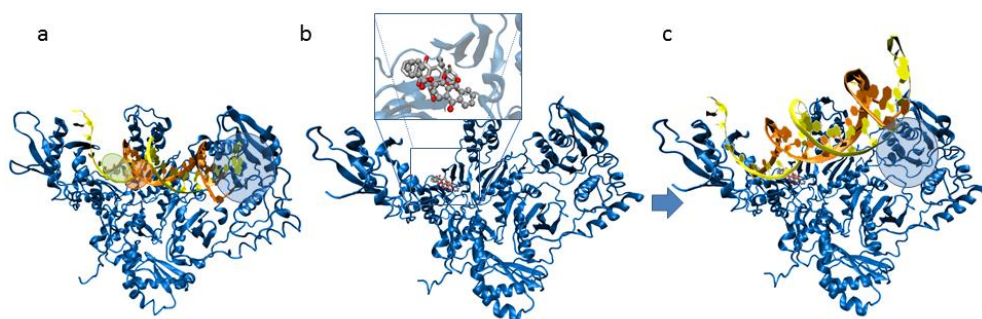
Most of the RNase inhibitors identified so far act by chelating the divalent metal ion  $Mg^{2+}$  that is coordinated in the active site by the catalytic residues D443, E478, D498 and D549. The issue with these compounds is that they showed toxicity due to lack of specificity.

Recently new compounds with novel allosteric mechanisms of action have been discovered.

Vinylogous urea derivatives, such as NSC727447, were proposed to bind to an allosteric pocket outside the RNase H catalytic site. NSC727447 does not inhibit the RT-associated RDDP activity and does not show metal-chelating properties.<sup>26</sup>

Furthermore it was reported that some hydrazones,<sup>27</sup> naphthyridinone<sup>28</sup> and anthraquinone<sup>29</sup> derivatives are able to inhibit the HIV-1 RNase H function through the binding to an allosteric pocket located between the polymerase catalytic region and the NNRTIs binding pocket, which is 50 Å away from the RNase H catalytic site.

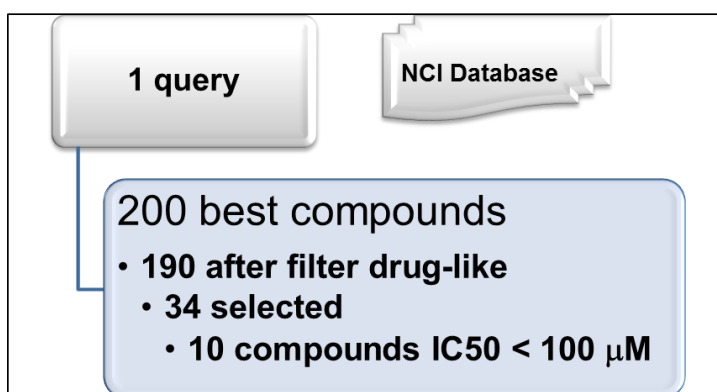
It seems that these compounds act through an innovative mechanism of action: by deviating the trajectory of the nucleic acid and preventing the bound with the RNase H site (Figure 5).<sup>27, 28</sup>



**Figure 5.** Putative mechanism of action of allosteric RNase H inhibitors: a) the usual trajectory of nucleic acid allow the substrate to reach RNase H catalytic pocket, b) compounds able to bind close to the primer grip, between polymerase catalytic site and NNRTI binding pocket, could be able to redirect the trajectory of the substrate (c) so that the RNA strand is not close enough to the RNase H active site for cleavage to occur.

Taking into account this innovative mechanism, a combined shape, 2D fingerprint and pharmacophore-based Virtual screening (VS) has been conducted by our group using as a query the first crystallized compound in this new pocket, which has a hydrazone moiety (dihydroxy benzoyl naphthyl hydrazone (DHBNH)).<sup>30</sup>

The overall VS protocol consisted of two consecutive screening processes. In the first, Rapid Overlay of Chemical Structures (ROCS)<sup>31, 32</sup> was used to perform *in silico* shape-based similarity screening on the NCI compounds database (Figure 6).<sup>33, 34</sup>



**Figure 6.** Workflow followed for the first selection after shape based VS

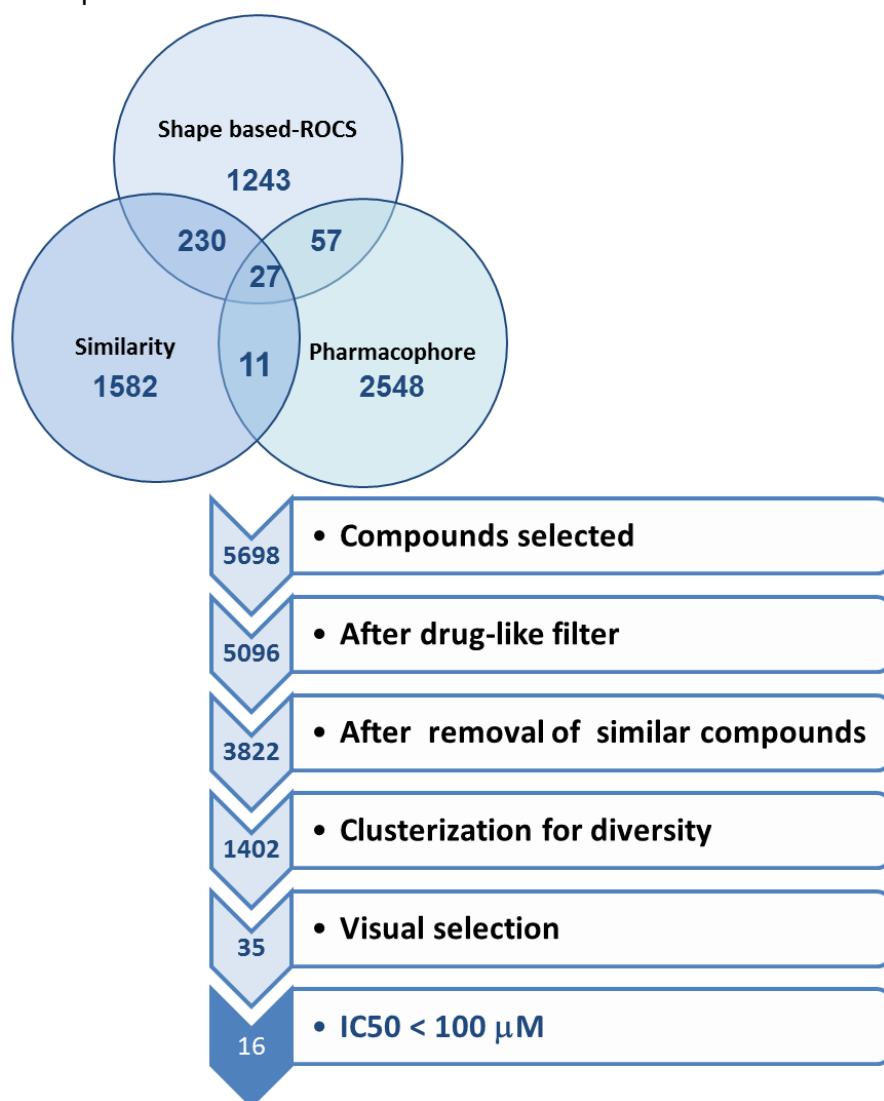
As a result, 34 hit molecules were selected and assayed on both RT-associated functions. Out of the 34 tested compounds, 10 inhibited the HIV-1 RNase H function with different potencies in the micromolar range, and were tested also for their effect on the RT-associated RDDP function. Subsequently, the four most active compounds toward RNase H ( $IC_{50}$  values  $\leq 21 \mu M$ ) were selected as queries for a second VS which combined three different ligand based methods: shape-based, 2D-fingerprint and 3D-pharmacophore.

The best-scored compounds with each method were selected and the results were analyzed and compared (Figure 7). Compounds were prioritized according to a consensus model: compounds predicted to be active by all three methods were selected, than the ones predicted by two methods, finally, we have also chosen some compounds retrieved by only one method based on proved strong points and complementarities of each approach in



finding active compounds.<sup>35</sup> In particular, 27 compounds were selected by all three methods and, out of these, 11 were chosen based on their chemical structure and availability from NCI. A total of 287 compounds were selected by two methods (Figure 7), and based on their structures, 10 were obtained by NCI. Finally, 14 compounds were selected even though they were picked up by only one method based on their molecular diversity, according to visual inspection of clusters.

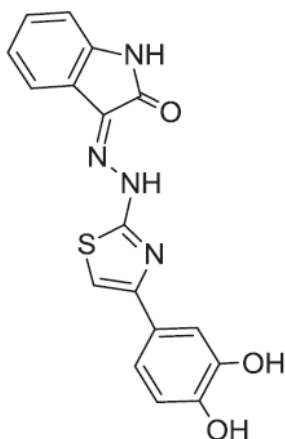
Therefore, other 35 compounds were selected at the end of the second VS procedure and tested on the HIV-1 RT-associated activities.



**Figure 7.** Overall screening process applied for the second VS selection

Results showed that 8 out of the 11 compounds selected by all three methods inhibited both RT-associated functions.

Among the 10 compounds selected by two VS methods, only one was active on both RT functions, compound **46** (numbered as in the paper)<sup>30</sup> (Figure 8) IC<sub>50</sub> values were 2  $\mu$ M for RNase H activity and 1.4 for RDDP activity, which, noteworthy, was the most potent inhibitor among all the tested compounds.



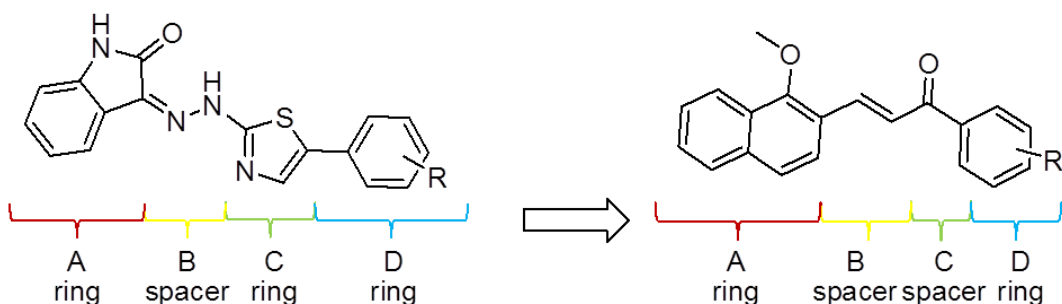
**Figure 8.** Compound **46**: Most active compound found by second VS round.

Considering the good results of this VS, new series of compounds were synthesized.

The purpose was to confirm the hydrazoneindolin-2-one scaffold as promising for the development of dual inhibitors of RT both associated functions (scaffold validation), and to modify the scaffold in order to maintain the pharmacophoric features through bioisosteric substitutions. Furthermore we wanted understand the dual inhibition mechanism because this information could help further scaffold optimization.

## 2.2 DIARYLPROPENONES AS DUAL INHIBITORS OF HIV-1 RT

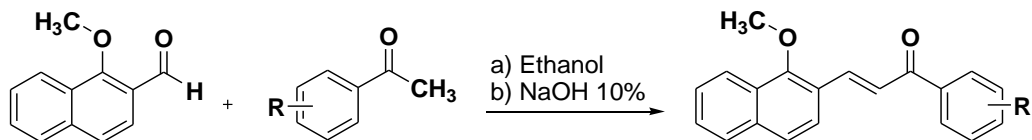
On the basis of compound **46** (Figure 8) as a hit compound, bioisosteric substitutions have been applied for the design of novel compounds as potential dual inhibitors (Figure 9). Hence, a new series of 1,3-diarylpropenones have been designed and synthesized.<sup>36</sup>



**Figure 9.** Bioisosteric substitutions starting from the compound **46**

The main structural features are an aromatic portion (A ring) a hydrazine spacer (B) and a thiazole ring (C) bearing a second aromatic ring (D) at the 4 position.

The indolinone ring was replaced by the 1-methoxynaphthalene moiety, the hydrazine spacer was substituted by a vinyl group, and the thiazole was replaced by the bioisosteric carbonyl feature.<sup>37</sup> The entire series had the E configuration, according to the coupling constants for the proton on the C=C bond (see Material and Methods).



R = 4-Br, 4-F, 4-OCH<sub>3</sub>, 4-Cl, 3NO<sub>2</sub>, 4-Ph

Scheme 1. Synthetic pathway to compounds EMAC 2000-2005.

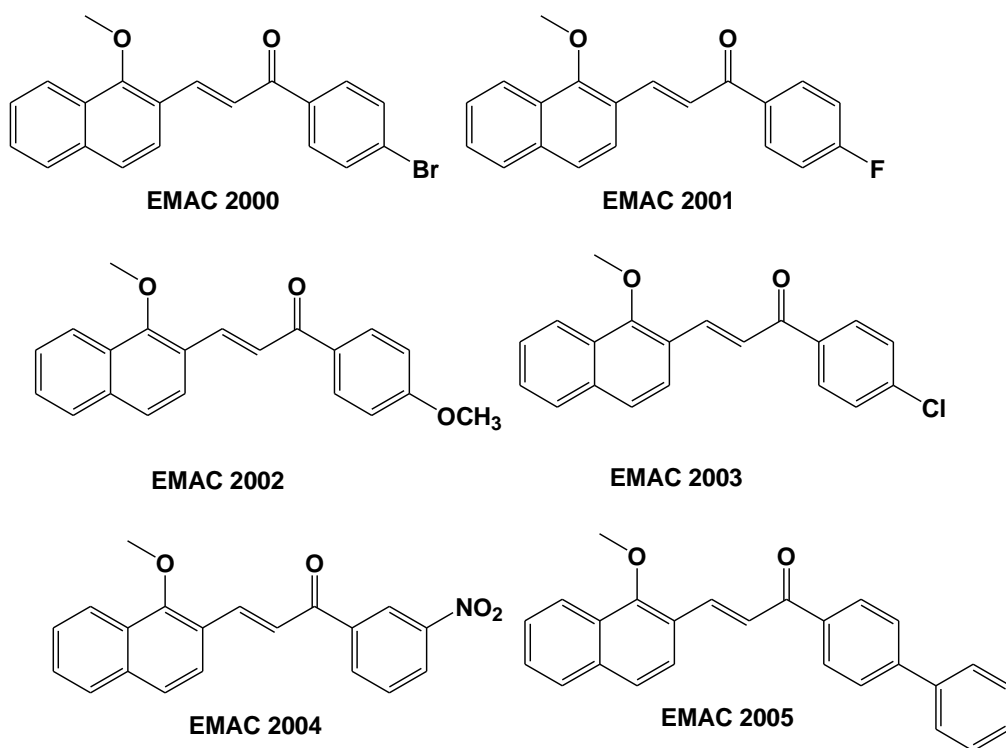
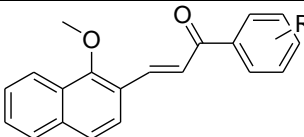


Figure 10. Diarylpropenones derivatives EMAC2000-2005.

The ability of the synthesized 1,3-diarylpropenone derivatives of inhibiting both RT-associated functions was measured by biochemical assays using **RDS1643**<sup>38</sup> and efavirenz as positive controls (Table 1). Most potent inhibitors were **EMAC2005** and **EMAC2002**. Preliminary structure–activity relationships analysis showed that although the RDDP activity was not affected by variation of the substituent at the 4-position of the D ring, the RNase activity was strongly influenced. In particular, bulky and strongly/weakly activating groups (e.g., methoxy and phenyl) were

preferred with respect to deactivating substituents (e.g., halogens). The introduction of a nitro group at the 3-position of the D ring was slightly more tolerated, probably because of its minor conjugative electron-withdrawing effect.

**Table 1.** HIV-1 RT-associated activity inhibition by 1,3-diarilpropenones derivatives.

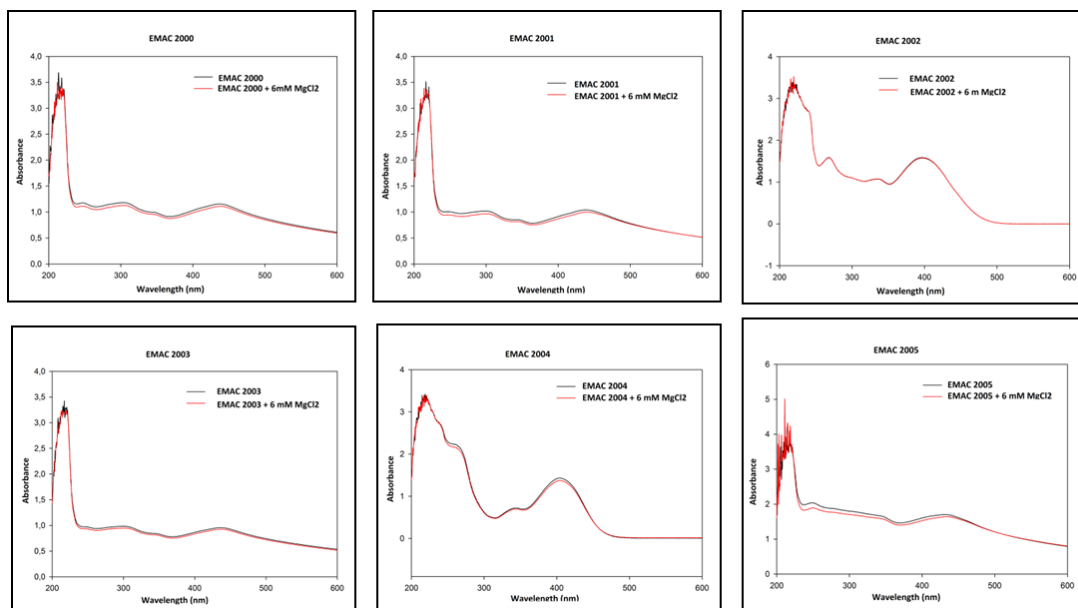
			
Compound	R	<sup>a</sup> IC <sub>50</sub> (μM) RNase H	<sup>b</sup> IC <sub>50</sub> (μM) RDDP
EMAC2000	-4 Br	47 ± 1	6 ± 1
EMAC2001	-4 F	23 ± 3	5 ± 1
EMAC2002	-4 OCH <sub>3</sub>	9 ± 2	6 ± 2
EMAC2003	-4 Cl	76 ± 11	5 ± 1
EMAC2004	-3 NO <sub>2</sub>	31 ± 4	5 ± 1
EMAC2005	-4 C <sub>6</sub> H <sub>5</sub>	6 ± 2	4 ± 1
<b>RDS1643</b>		13 ± 4	> 100
<b>Efavirenz</b>		> 10	0.003 ± 0.002
<sup>a</sup> Compound concentration required to reduce the HIV-1 RT associated RNase H activity by 50%. <sup>b</sup> Compound concentration required to reduce the HIV-1 RT associated RDDP activity by 50%.			

The activity of the compounds was tested on the replication ability of HIV-1 in a single round of infection in Jurkat cells. All compounds were not able to inhibit HIV-1 replication within these experimental conditions. Thus, we performed in vitro permeability assays to assess if a reduced or absent transmembrane permeation could explain these results. As results we obtained that the trans-membrane permeation profile of **EMAC2005** was characterized by a prolonged lag-time followed by a gradual permeation.

However since our main goal was to gain more insight into the mechanism of dual inhibition we have carried out biochemical and computational experiments.

## 2.2.1 Biochemical studies

Since several classes of HIV-1 RNase H generally act by chelating the magnesium(II) (MgII) ions within the active site, the ability of **EMAC2000-EMAC2005** to chelate Mg ions was tested measuring the absorbance spectrum in absence and presence of MgCl<sub>2</sub> observing no differences (Figure 11). Thus we can exclude they will act with this mechanism.

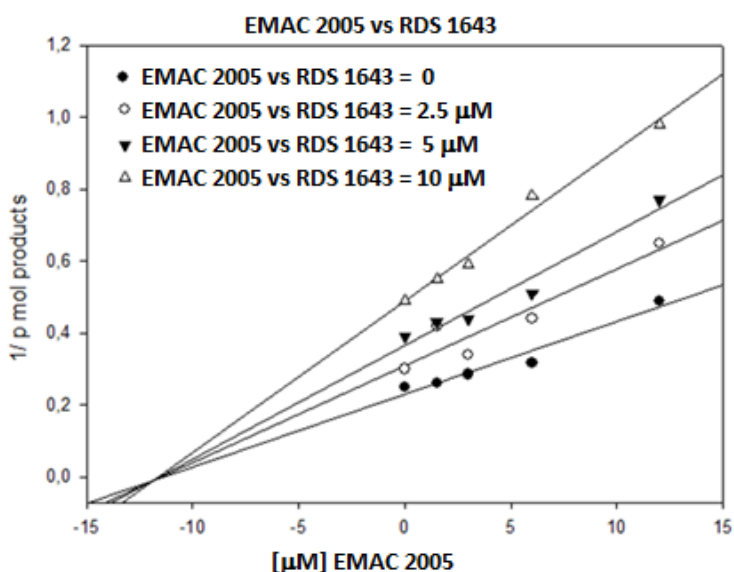


**Figure 11.** Spectrum of absorbance of the EMAC series in absence (blue) and presence of MgCl<sub>2</sub> (red) **EMAC2000-EMAC2005**.

Subsequently, we further investigated if **EMAC2005** and the diketo acid derivative **RDS1643**, an RNase H catalytic site inhibitor,<sup>39</sup> were able to simultaneously bind to RT. Such an evaluation was carried out by means of the Yonetani revised Yonetani-Theorell model that allows to determine when two inhibitors of a certain enzyme compete for the same binding site or act on two non-overlapping binding sites. In this revised model, the plot of the reverse reaction velocity ( $1/v$ ) observed in the presence of different concentrations of the first inhibitor, in the absence or in the contemporaneous presence of the second inhibitor, leads to a series of lines

which are parallel if the two inhibitors compete for the same binding site, whereas they intersect if the inhibitors bind to different enzyme sites.<sup>40</sup>

Therefore, the HIV-1 RT RNase H activity was measured in the presence of increasing concentrations of both **EMAC2005** and **RDS1643** and analysed with the Yonetani-Theorell plot (Figure 12). Results showed that the slope of the plots of  $1/v$  versus **EMAC2005** concentration decreased at increasing **RDS1643** concentrations, confirming that the two compounds are not kinetically mutually exclusive. Overall, these data support the hypothesis that **EMAC2005** does not bind to the HIV-1 RNase H catalytic site.



**Figure 12.** Yonetani-Theorell plots of the interaction between **EMAC2005** and **RDS1643** on the HIV-1 RNase H activity. HIV-1 RT was incubated in the presence of different concentrations of **EMAC2005** and in the absence (●) or in the presence of 2.5 μM (▼), 5 μM (△) or 10 μM (■) **RDS1643**.

Next, we evaluated the effects of **EMAC2005** on Y181C and K103N mutated RTs, involved in NNRTIs resistance and located close to RNase allosteric pocket and NNIBP (Table 2). Results showed that when tested on the K103N RT, **EMAC 2005** was 10 fold less potent on the RNase H. On the contrary, no influence of K103N mutation on the RDDP activity was observed. In the case of Y181C mutation a more dramatic effect, with respect to K103N, was observed: the activity towards RNase H activity was

almost suppressed, while the activity towards the RT-associated RDDP was only slightly affected. This behaviour might be explained either by the binding of **EMAC2005** to a single site close to Y181, the hydrazones pocket or the NNIBP, or by the interaction of the compound with an interdependent pocket, whose conformation is affected by the Y181C mutation.

**Table 2.** Inhibition of drug resistant HIV-1 mutated RT associated functions by **EMAC2005**.

	IC <sub>50</sub> (μM)			
	Y181C RT		K103N RT	
	<sup>a</sup> RNase H	<sup>b</sup> RDDP	RNase H	RDDP
<b>EMAC2005</b>	> 100	8 ± 3	59 ± 8	3 ± 1
Efavirenz	--	0.40 ± 0.03	--	0.68 ± 0.05

<sup>a</sup>Compound concentration required to reduce the HIV-1 RT associated RNase H activity by 50%.  
<sup>b</sup>Compound concentration required to reduce the HIV-1 RT associated RDDP activity by 50%.

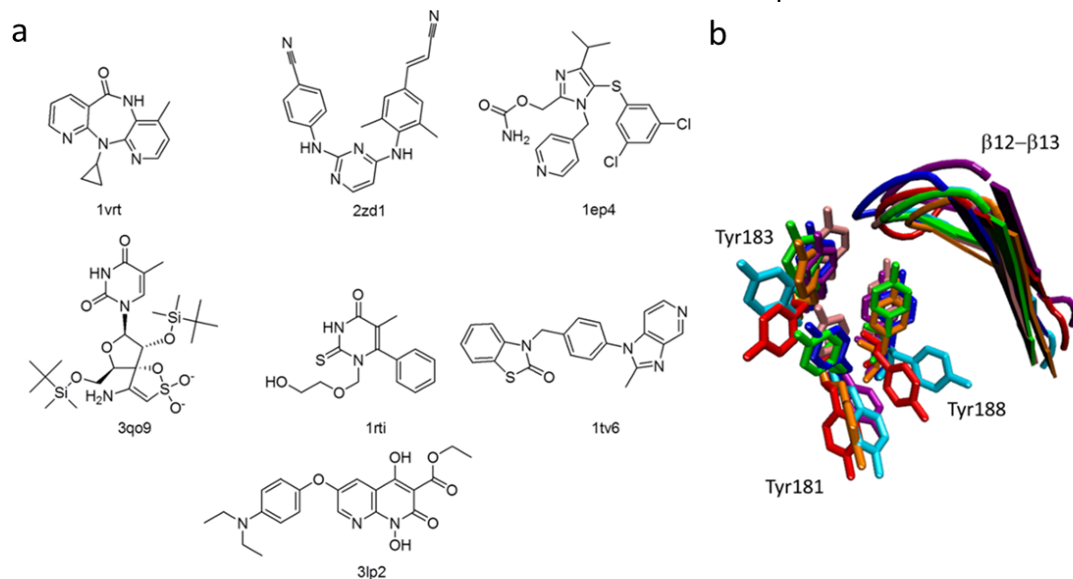
### 2.2.2 Molecular modeling

We applied a computational strategy based on molecular docking and molecular dynamics to gain insight into the mechanism of action of this new class of 1,3-diarylpropenones. The studies were focused on the most active compound **EMAC2005**. According to the available literature information, dual inhibitory activity could be achieved either by inhibitor binding into two different sites<sup>27</sup> or by its binding into a single site.<sup>27, 28, 30, 41</sup> Therefore, we investigated both possibilities.

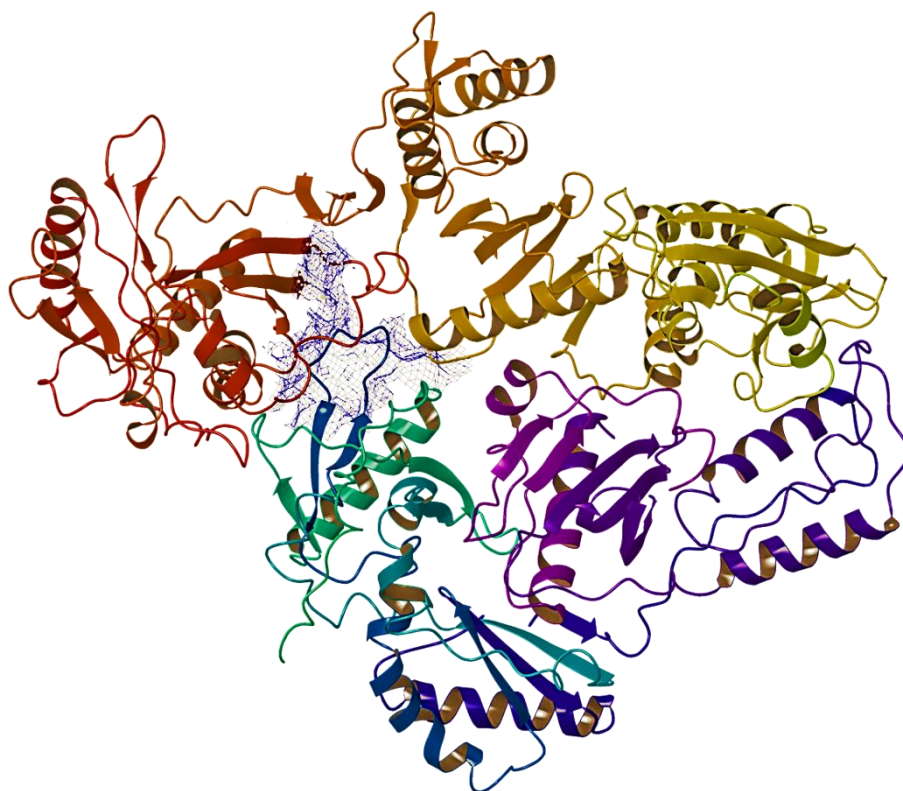
NNRTIs are characterized by a high diversity of chemical structures. They are able to bind allosterically in the hydrophobic NNIBP and lock the enzyme into an inactive form. Due to the flexibility of the target and to the different shapes of known inhibitors (Figure 13a), the major conformational changes in the NNIBP were taken into account to perform a clusterization of the available RT-NNRTIs complexes. In particular, the orientation of Y181,



Y188, Y183 and primer grip  $\beta$ 12- $\beta$ 13 hairpin were considered. A representative of each different cluster was picked<sup>42</sup> and the 3D structure of HIV RT was retrieved from the Protein Data Bank<sup>43</sup> (Figure 13b). Hence we carried out ensemble docking experiments using seven different crystal structures.<sup>44</sup> It was observed that NNIBP, when Y181 and Y188 adopt a parallel position, communicates with RNase H allosteric pocket.<sup>27, 28</sup> The overall shape of this pocket is characterized by a L shape (Figure 14) and offers many possibility of binding to small molecules. Two crystals, 1tv6 and 3lp2, are characterized by this conformation and we have chosen to use both since they have co-crystallized compounds in different position: the first in the NNIBP and the second in the allosteric RNase H pocket.



**Figure 13.** a) Chemical and crystal structures pdb codes of co-crystallized NNRTIs selected for the ensemble docking procedure b) Stereoview of the primer grip region and residues Y181, Y183 and Y188 of the selected pdb complexes.

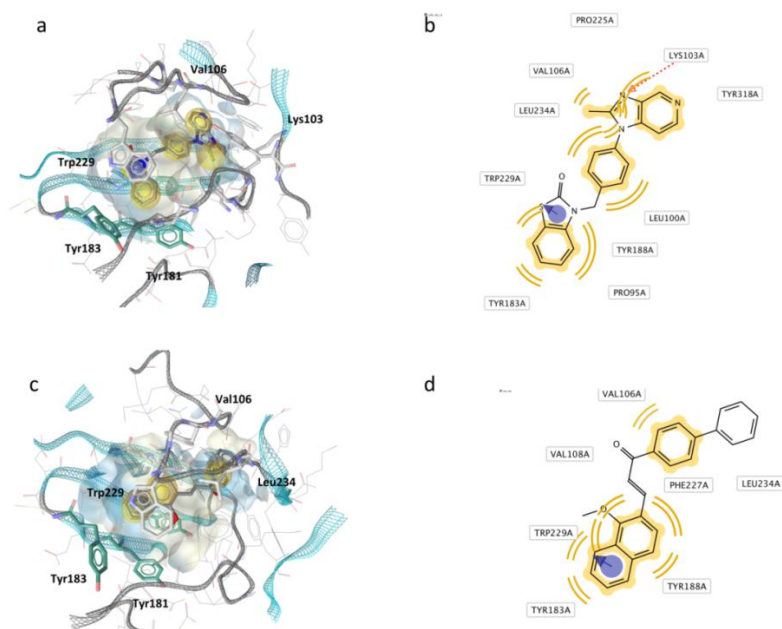


**Figure 14.** L shape binding pocket in the polymerase domain of HIV 1.

The docking protocol employed for this project has involved the QM (quantum mechanical) polarized Docking<sup>45</sup> (see 1.3.1 for more details). The protocol has been validated through the redocking of the co-crystallized ligands (data not shown). In order to consider the whole enzyme, two grids box were considered: one centered in W229 (NNIBP) and one in Q500 (RNase domain). The obtained **EMAC2005-RT** complexes were subjected to a post-docking procedure based on energy minimization and successive binding free energy calculation. The binding free energies ( $\Delta G(\text{Bind})$ ) were obtained applying molecular mechanics and continuum solvation models using the molecular mechanics generalized Born/surface area (MM-GBSA) method.<sup>46</sup> As reported in Table 2, by comparing the  $\Delta G\text{-MMGBSA}$  values, we can assert that the most probable binding mode in the NNIBP is obtained by docking the compound into the RT conformation model reported in 1tv6 pdb<sup>47</sup> (Table 3 and Figure 15).

**Table 3.** Ensemble docking scores and binding free energies of [EMAC2005-RT] complexes

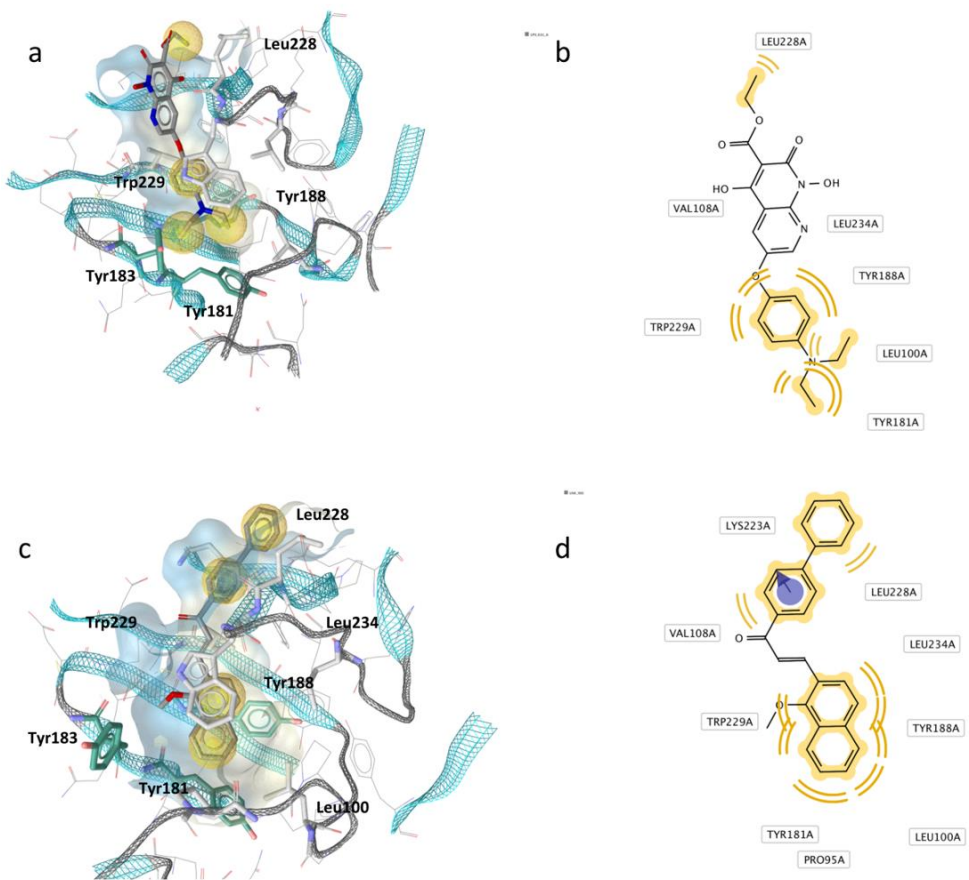
Docking site and box centre	pdb ID	Co-crystallised ligand	G-score	$\Delta G(\text{MMGBSA})$	Best Poses
NNIBP (centre on W229)	1vrt <sup>48</sup>	Nevirapine	-4.18	-30.6	
	2zd1 <sup>49</sup>	Rilpivirine	-8.43	-30.7	
	1ep4 <sup>50</sup>	Capravirine	- 8.26	-38.8	
	3qo9 <sup>23</sup>	TSAO-T	-9.21	-35.0	
	1rti <sup>48</sup>	HEPT	-11.04	-34.1	
	1tv6 <sup>47</sup>	CP-94,707	-9.53	-48.0	A
RNase H allosteric pocket (centre on W229)	3lp2 <sup>28</sup>	MK3	-6.19	-38.4	B
RNase H domain (centre on Q500)	1tv6		-7.50	-38.6	C
	-	EN37 <sup>51</sup>	-5.58	-24.6	



**Figure 15** Comparison between a), b) HIV-1 RT co-crystallized ligand CP-94,707 binding mode reported in 1tv6 crystal structure and c) d) optimized **EMAC2005** best docking pose in the same site (Pose A). Yellow spheres show hydrophobic contacts, red arrows: HB acceptor, violet circle: aromatic interaction. Binding pocket surface is drawn as solid and coloured according to lipophilicity: pale yellow indicates lipophilic residues and light blue hydrophilic residues.

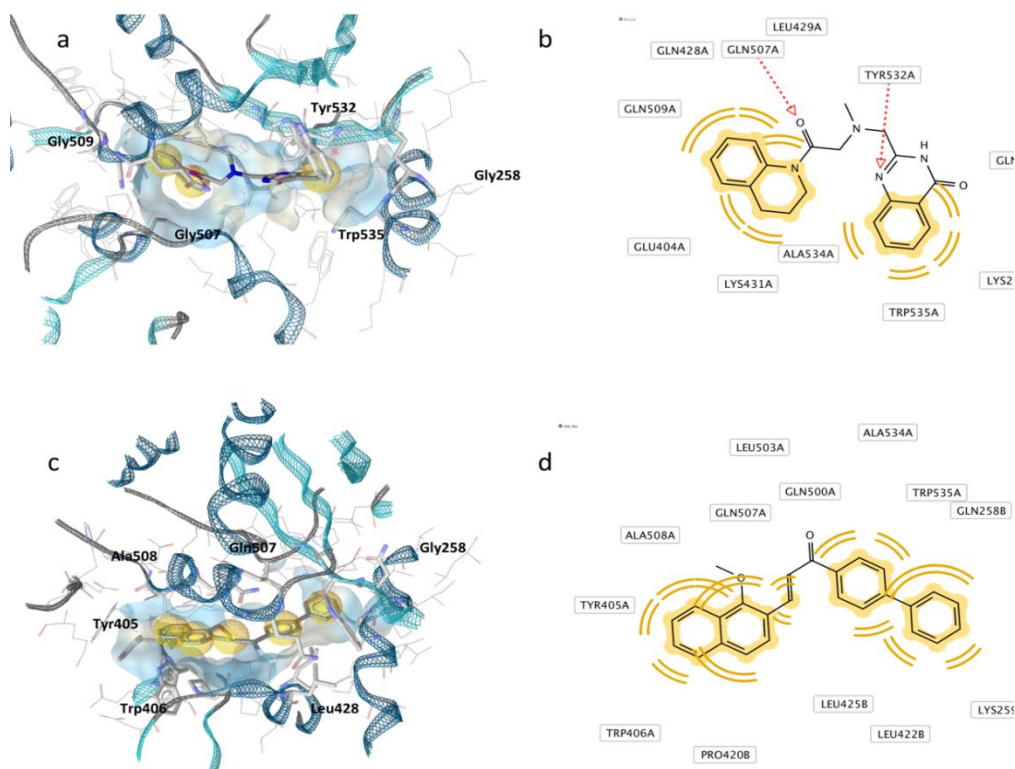
To assess which binding pocket was responsible for the RNase H inhibitory activity, we considered the reported binding pockets: one is located in the catalytic domain and the other is an allosteric site described by Himmel as hydrazone site<sup>28</sup> and already considered in our previous studies.<sup>30, 41</sup> Furthermore, recently, new RNase H inhibitors have been described that bind in a cleft below the RNase H catalytic site, close to Q500.<sup>51</sup> Probably, these compounds induce an RNase H domain conformation which prevent his function. Hence we have to consider also this second allosteric pocket for RNase H inhibitory activity.

Biochemical experiment directed to verify the ability of this series of compounds to coordinate the metal ions indicated that the chelation mechanism can be excluded (Figure 11) so we excluded that our compounds bind to the catalytic site. However, we could not ignore the possibility of a binding site close to the RNase H catalytic residues as reported by Felts.<sup>51</sup> Hence, in order to include the whole RNase H domain for investigation, in our docking experiments the binding site was defined by a regular box of 97336 Å<sup>3</sup>, centred on residue Q500. The RT conformations adopted for docking experiments were: 1tv6 X-ray and the crystallographic model reported by Felts *et al.* but not available in the PDB yet.<sup>51</sup> In Table 3 are reported the  $\Delta G$  of the best poses in the first and second allosteric pocket of RNase H. While, in the following figures the best poses are depicted (Figure 16 and 17).



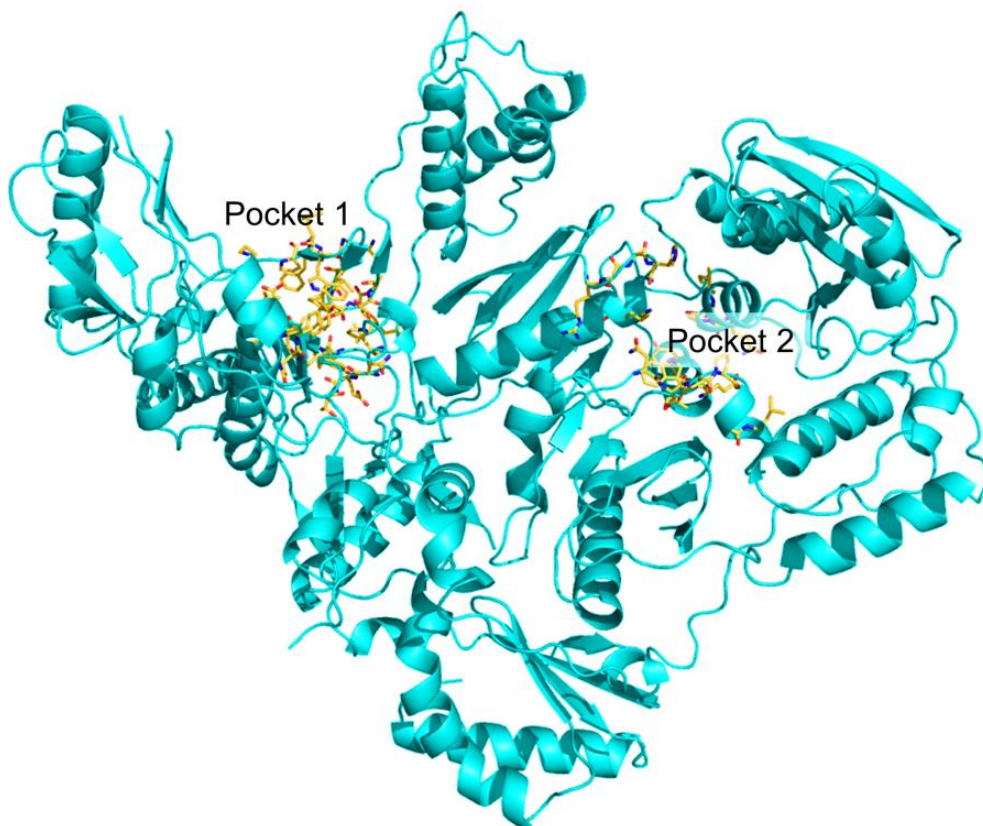
**Figure 16.** a) b) MK3 binding mode reported into 3lp2 crystal structure and c) d) optimized EMAC2005 best docking pose (Pose B). Interactions and rendering according to previous Figure

15



**Figure 17.** a), b) EN37 binding mode reported by Felts *et al.* and c), d) optimized **EMAC2005** best docking pose into 1tv6 crystal structure (Pose C). Interactions and rendering according to previous Figure 15

It can be noticed that the predicted  $\Delta G$  of the two complexes which consider the two allosteric RNase H sites are comparable (Pose B and C). Hence a better evaluation of the energy of the complexes may help to discriminate between the two binding modes. Therefore we decided to perform molecular dynamic simulations considering the best scored poses after post-docking protocol: **EMAC2005** bound in the NNIBP (A); **EMAC2005** bound in the two RNase H allosteric pockets described by Himmel (B) and by Felts (C). Spatially we have that the most favourable poses occupy the L pocket of the palm subdomain (Figure 14 and Figure 18): pose A and B (pocket 1); while the pose C is located below the RNase H catalytic site (pocket 2) close to the interface between p66 and p51.



**Figure 18.** Binding sites individuated after docking experiments on the wt HIV-1RT structure.

We run the molecular dynamic (MD) simulations up to 6 ns using Desmond Molecular Dynamics System ver. 2.4,<sup>52</sup> keeping the whole enzyme free to move into explicit solvent water environment.

Docking and MD simulation were carried out also on the mutated enzyme complexes containing single point mutation Y181C and K103N. (see Table 2 for IC<sub>50</sub>). We applied the same computational protocol described for wt RT: docking, energy minimization,  $\Delta G$ -MMGBSA calculation and MD simulations.

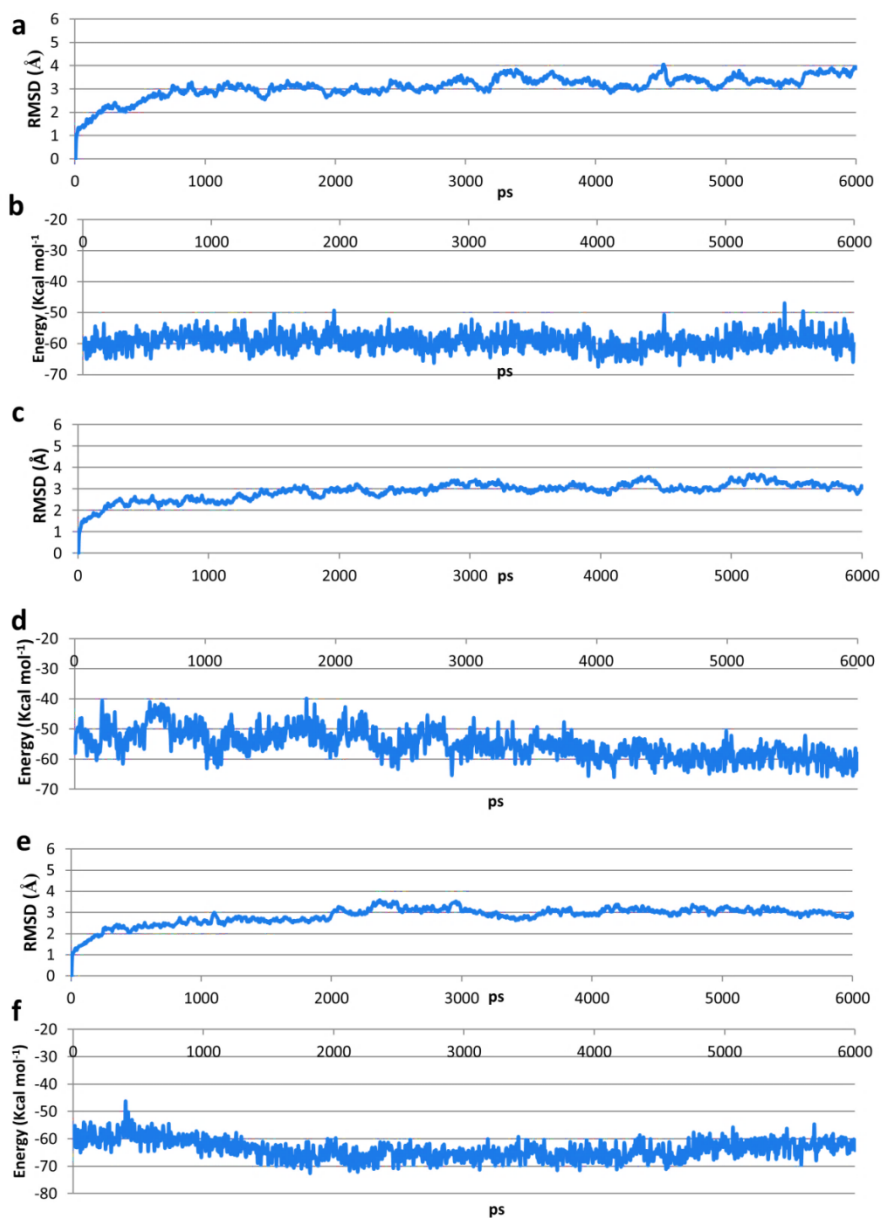
The interaction energy values of [EMAC2005-RT] complexes are reported in Table 4, while their variations, sampled at regular intervals during the simulations over the entire MD trajectory, are illustrated in Figures 19-21. In particular the analysis of the Root-Mean-Square Deviation (RMSD) for wt and mutated enzymes during the MD, shows that the system

is structurally stable during the simulation. The effect of mutation was analysed for both poses A and B which are close to mutated residues and may be affected strongly by the mutations.

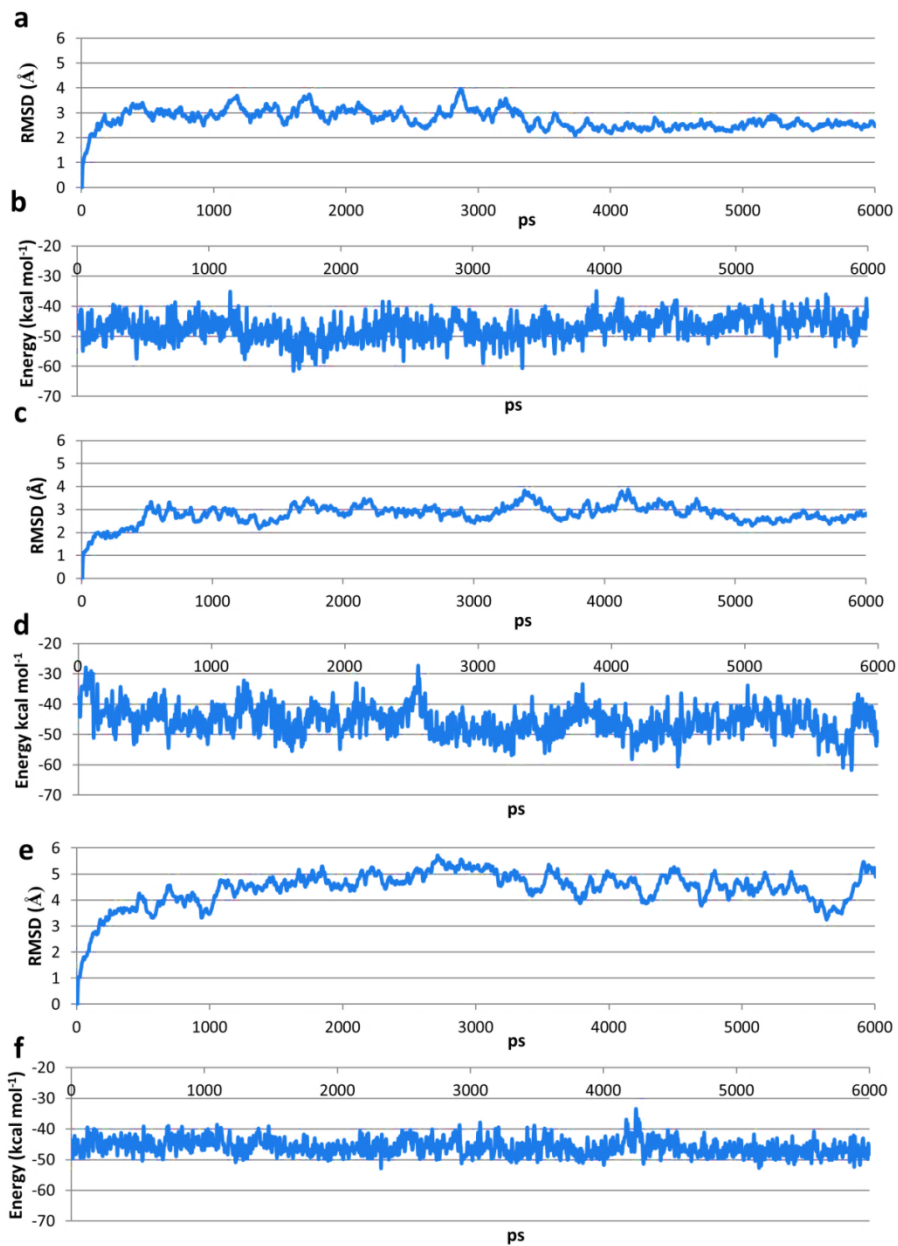
**Table 4 EMAC2005** docking GScoreXP,  $\Delta G_{\text{MMGBSA}}$  and Total Interaction Energy (IE) variation after MD simulation. Terms expressed in kcal/mol per complex.

Pocket	Docking site and box centre	Enzyme	G-score	$\Delta G_{\text{MMGBSA}}$	Total-IE
Pocket 1	NNIBP (centre on W229)	wt	-9.53	-48.00	-52.01
		Y181C	-9.90	-44.30	-48.03
		K103N	-9.59	-45.50	-50.78
	RNase H allosteric pocket (centre on W229)	wt	-6.19	-38.40	-40.37
		Y181C	-5.40	-37.10	-38.79
		K103N	-6.06	-36.40	-39.06
Pocket 2	RNase H domain (centre on Q500)	wt	-7.50	-38.60	-42.40

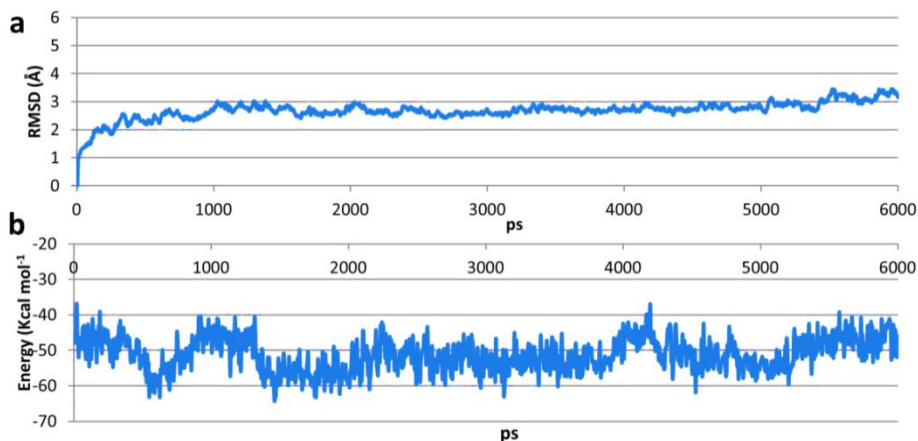




**Figure 19.** MD Analysis of EMAC2005 bound in NNIBP of RT (pocket 1). Plots depict RMSD, calculated considering the RT heavy atoms, and total interaction energy variation a) b) [EMAC2005-wt-RT], c) d) [EMAC2005-Y181C-RT], e) f) [EMAC2005-K103N RT] complexes, values were sampled at regular intervals during the simulation over the entire MD trajectory.



**Figure 20.** MD Analysis of **EMAC2005** bound in RNase H allosteric site of RT (pocket1). Plots depict RMSD, calculated considering the RT heavy atoms, and total interaction energy variation a) b) [EMAC2005-wt-RT], c) d) [EMAC2005-Y181C-RT], e) f) [EMAC2005-K103N RT] complexes, values were sampled at regular intervals during the simulation over the entire MD trajectory.



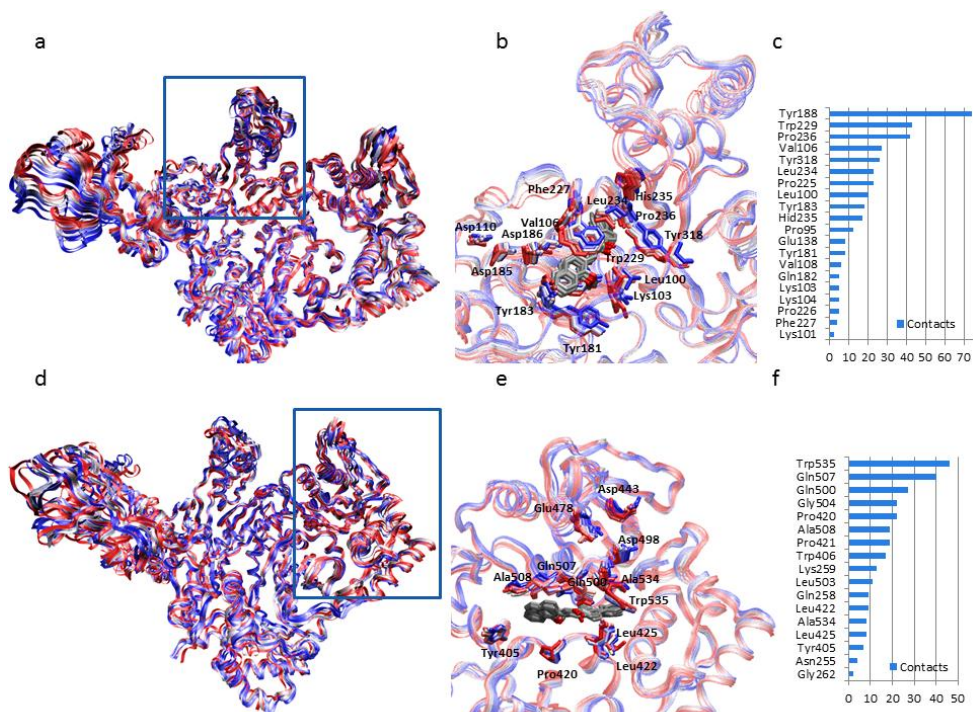
**Figure 21.** MD Analysis of **EMAC2005** bound in RNase H allosteric site of RT (pocket2). Plots depict RMSD, calculated considering the RT heavy atoms, and total interaction energy variation a) b) [**EMAC2005**-RT] values were sampled at regular intervals during the simulation over the entire MD trajectory.

Considering the MD results reported in Table 4 the binding in the allosteric pocket of RNase H close to NNIBP cannot explain the loss of **EMAC2005** inhibitory activity toward RNase H. In fact when Y181 is mutated in cysteine and K103 is mutated in asparagine, we could not observe any significant variation of the energy of the complex. While a significant loss of activity was observed experimentally. Consequently we can conclude that the binding of **EMAC2005** in this pocket is not the most favourite.

Hence we can suppose that the polymerase inhibition is due to the binding of **EMAC2005** into the NNIBP (Figure 19). This hypothesis is also supported by the similar behaviour of **EMAC2005** and CP-94,707 when tested in the mutated enzymes. In fact, also CP-94,707 acts in the same manner and both compounds interestingly retain RDDP activity.<sup>47</sup> Differently, both biochemical and modelling studies seem to confirm that the binding into an allosteric site close to RNase H catalytic residues is responsible for RNase H inhibitory activity (Pocket2, Figure 17). Both binding modes are retained during MD (Figure 22).

To further corroborate this hypothesis we evaluated the activity of **EMAC2005** on A502F RT, a residue close to the RNase H allosteric pocket.<sup>53</sup> As we expected, this mutation does not affect the activity towards the RDDP activity. On the contrary, the activity of compound **EMAC2005** on the RNase

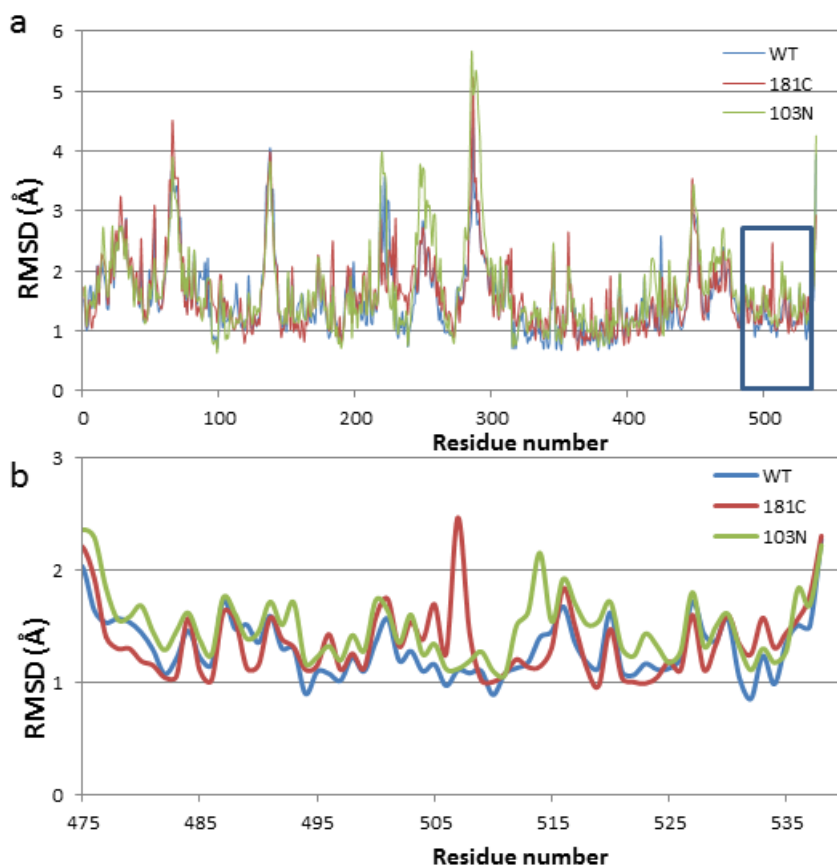
H function is fivefold lower, with respect to the wt RT, indicating that residue A502 is close to the **EMAC 2005** allosteric binding pocket.



**Figure 22.** Superimposed structures of 6 ns MD simulation frames of [EMAC2005-RT] complex: initial, final along with intermediate structures snapshots. a) Overall structure of the HIV-1 RT heterodimer with the NNIBP occupied; b) Close-up of the binding cavity; d) Overall structure of the HIV-1 RT heterodimer with the Site Q500 occupied; e) close-up of the binding cavity. c) f) Residues involved in the complex stabilization sorted by number of contacts between ligand and receptor. Interacting and catalytic residues are represented in sticks.

Still it remained to be understood why the **EMAC2005** inhibitory activity towards RNase H is modified in the mutated enzymes. In this regard there is a considerable evidence that the binding of NNRTIs as well as the mutations in the allosteric pocket in the RT DNA polymerase domain affect the activity of the spatially remote RNase H domain. The mechanisms involved in this long-range alteration of RNase H activity are not entirely understood, but likely involve changes in the positioning of the RNA/DNA duplex nucleic acid.<sup>54, 55</sup> Therefore we examined long-range effects of both mutations by checking the fluctuations of the residues (RMSF) during MDs of the wt and mutated enzymes (Figure 23a). We noticed that while wt-RT does not show relevant fluctuations in the site Q500 involved in the binding

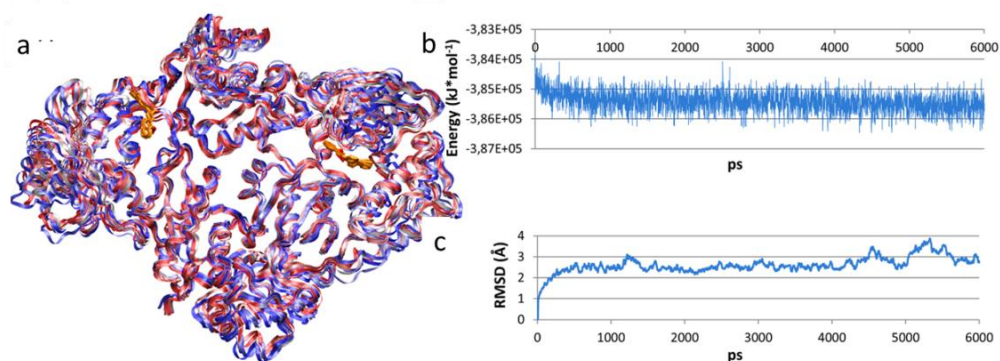
of **EMAC2005**, some residues in the mutated RTs showed a huge fluctuation which may disturb the binding of **EMAC2005** (Figure 23b).



**Figure 23.** RMSF of subunit p66 during MD a) Entire wt and mutated RT. b) Close up of site Q500 residues.

Finally we have evaluated the stability of the ternary complex enzyme-**EMAC2005** bound in the two allosteric sites (Figure 24).

Plots for potential energy and RMSD fluctuations, related to the complex, are depicted in Figure 24b,c. The analysis shows that the structure reached equilibrium and the low fluctuations support the stability of the intermolecular interactions.



**Figure 24.** Superimposed structures of 6 ns MD simulations frames of with **EMAC2005**-RT ternary complex: initial, final along with intermediate structures snapshots. a) Overall structure of the HIV-1 RT heterodimer; b) potential energy of the complex during the MD simulation; c) RMSD fluctuations during trajectory.

### 2.2.3 Conclusions

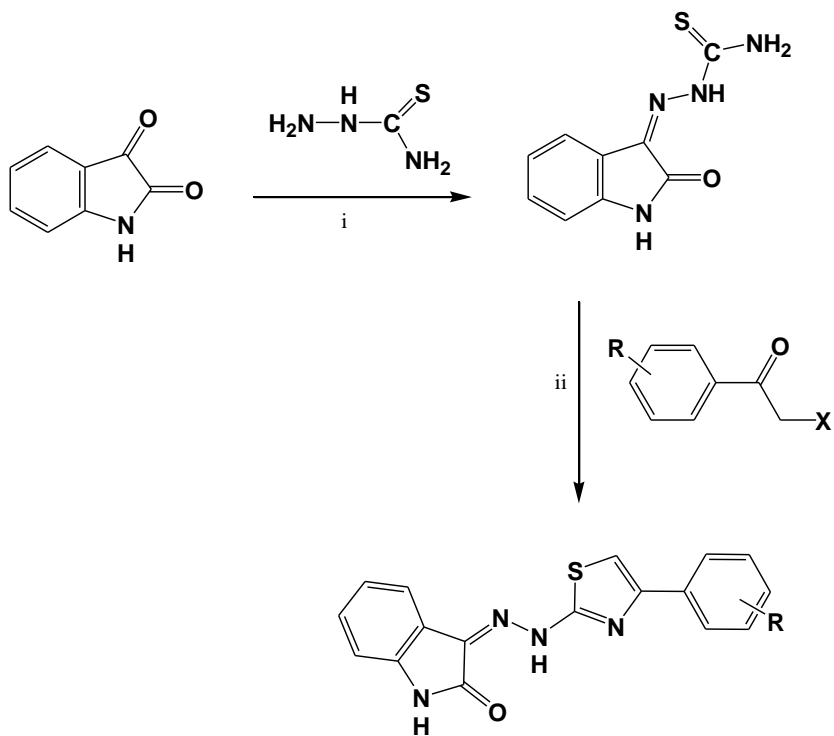
In conclusion, with the aim to obtain dual inhibitors of the RT associated functions, a small series of 1,3-diarylpropenones was designed, synthesised and tested. The activity of some compounds and the profile towards mutated enzymes were remarkable and suggestive for further modifications and studies. Moreover, investigating the possible mechanism of action of the most promising compound **EMAC2005**, we found that its inhibitory activity could be addressed to the binding at two different enzyme clefts: the NNIBP and an allosteric site close to the RNase H catalytic DEDD motif (close to Q500). We highlighted that the compound is accommodated in a pocket with Y181 and Y188 in close conformation (PDB code 1tv6) better than in open conformation as most of NNRTIs. This facilitates the enzyme recognition when common mutations, such as Y181C and K103N, occur and therefore RDDP activity is not impaired. The **EMAC2005** binding mode confirms the known key role of W229 and Y188 in the stabilization of the complex. Other interacting residues, L100, P225, L234, Y318, V106, P236, highlighted the importance of hydrophobic contacts. Instead, most likely, the loss and decrease of RNase H inhibitory potency is due to the improbable entrance of **EMAC2005** into the second pocket (close to residue Q500) when Y181C and K103N mutations occur. This hypothesis is further

confirmed by single point mutation experiment on A502 residue. It was found that while the inhibition potency of compound **EMAC2005** towards the RDDP function of HIV-1 RT A502F is almost not modified ( $IC_{50}$  increased by 1.5 folds), the inhibition of the RNase function is remarkably affected with a 5 fold reduction of the potency. Thus compound **EMAC2005** behaves, most likely, as a dual site dual function inhibitor.

### **2.3 INDOLINONES DERIVATIVES AS DUAL INHIBITORS OF HIV-1 RT**

The second class of compounds studied was derived directly from the most active compound **46** identified with the VS (Figure 8).<sup>30</sup> Indeed it is important after VS to validate the scaffold in order to see if also derivatives retain the same activity. Therefore a small library of indolinones compounds with different substituents on the aromatic portion has been synthesised (Scheme1, Figure 25).<sup>56</sup>





EMAC 2072-2083

R: 4-Cl, 4-F, 4-Br, 4-NO<sub>2</sub>, 4-C<sub>6</sub>H<sub>5</sub>, 4-CN, 2,4-F, 3-NO<sub>2</sub>, 4-CH<sub>3</sub>, 4-OCH<sub>3</sub>,  
H, 2,4-Cl X: Cl or Br

**Scheme 2.** Synthesis of (Z)-3-(2-(4-arylthiazol-2-yl)hydrazono)indolin-2-one derivatives  
**EMAC 2072-2083.** Reagents and solvents: (i) 2-propanol, AcOH; (ii) 2-propanol, room  
temperature (r.t).

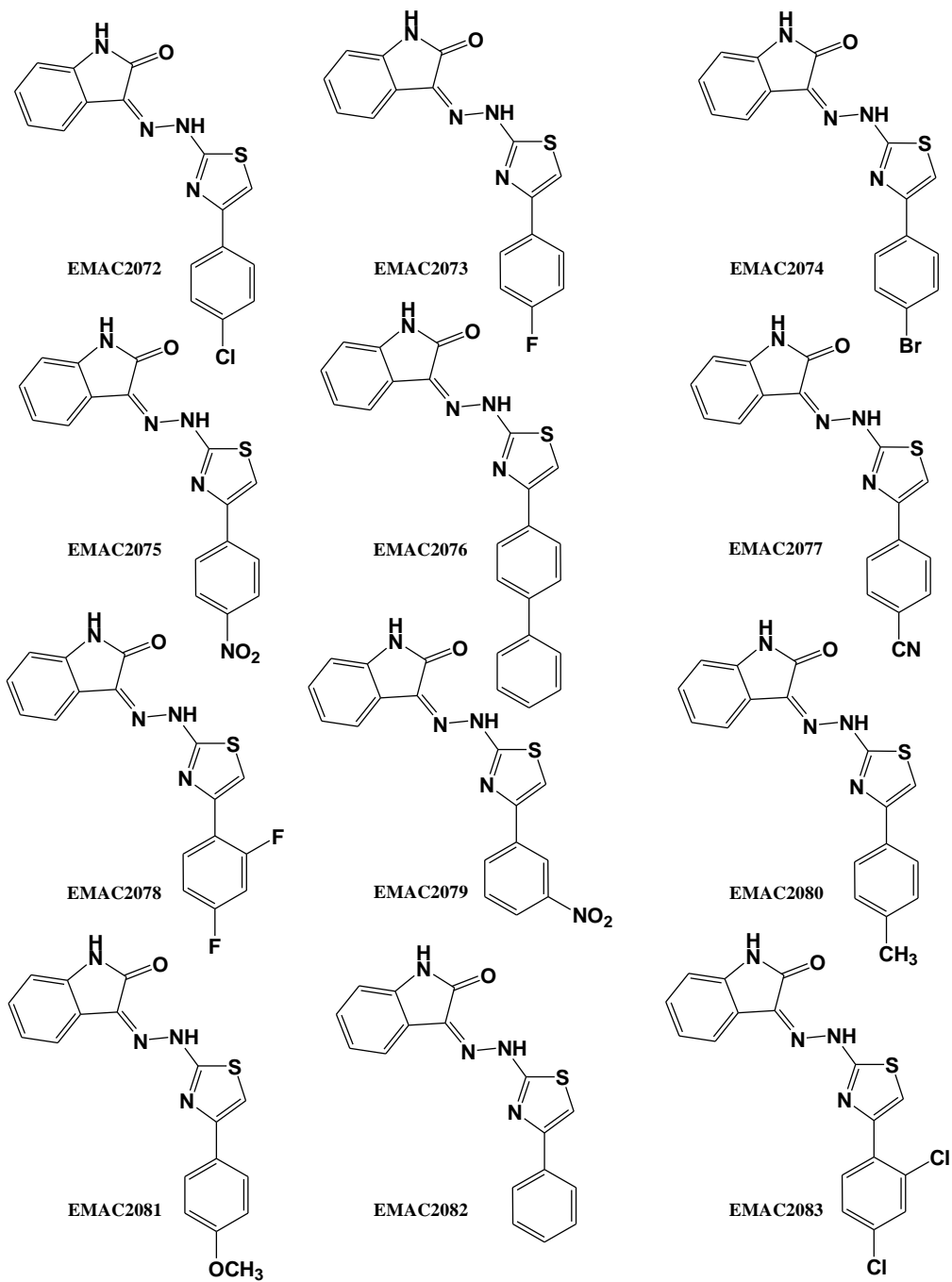
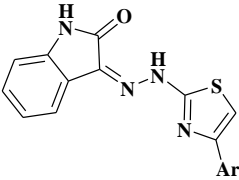
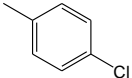
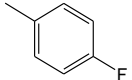
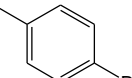
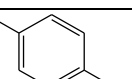
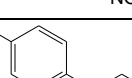
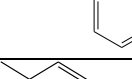
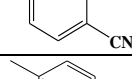
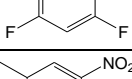
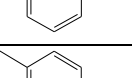


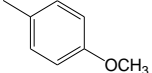
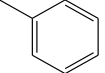
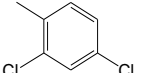
Figure 25. Structures of compounds with the indolinone scaffold.

### 2.3.1 Biochemical studies

The activity of these compounds was tested toward both RT activities (Table 5) and the most active compound was the derivative bearing the Cyano group on the para position aromatic portion ((Z)-4-(2-(2-(2-oxoindolin-3-ylidene)hydrazinyl)thiazol-4-yl)benzonitrile (**EMAC2077**) with an IC<sub>50</sub> against RNase H of 1.3 ± 0.3 μM and against RDDP of 9.8 ± 1.4 μM.

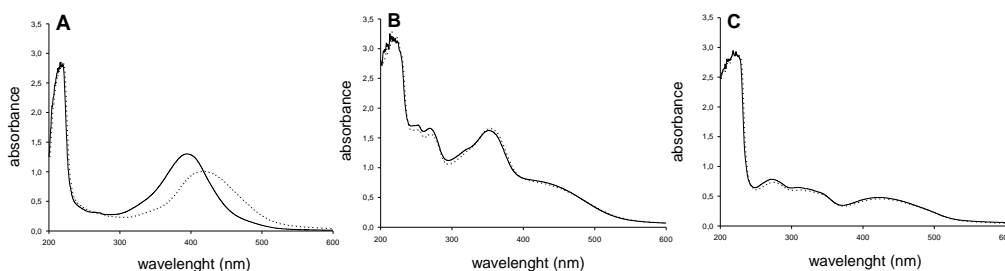
**Table 5. EMAC2072-2083 derivatives effects on the HIV-1 RT-associated functions.**

				
Compound	Ar	RNase H <sup>a</sup> IC <sub>50</sub> (μM)	DP <sup>b</sup> IC <sub>50</sub> (μM)	<sup>c</sup> Spl
EMAC2072		10.6 ± 0.8	20.0 ± 6.0	0.53
EMAC2073		6.4 ± 1.5	23.9 ± 5.7	0.26
EMAC2074		14.9 ± 2.8	81.0 ± 15	0.18
EMAC2075		2.9 ± 0.8	34.0 ± 5	0.08
EMAC2076		2.5 ± 0.4	22.0 ± 1	0.11
EMAC2077		1.3 ± 0.3	9.8 ± 1.4	0.13
EMAC2078		2.6 ± 0.5	18.5 ± 2.5	0.14
EMAC2079		2.8 ± 0.8	13.0 ± 5	0.21
EMAC2080		9.0 ± 0.9	45.0 ± 2.5	0.20

<b>EMAC2081</b>		4.7 ± 0.5	26.0 ± 5	0.18
<b>EMAC2082</b>		100	71.6 ± 16	1.3
<b>EMAC2083</b>		3.9 ± 1.5	19.5 ± 2.5	0.20
<b>Efavirenz</b>		ND <sup>d</sup>	0.023 ± 0.006	
<b>RDS1643</b>		10.1 ± 2.2	ND	
<sup>a</sup> Compound concentration required to reduce the HIV-1 RT-associated RNase H activity by 50%; <sup>b</sup> Compound concentration required to reduce the HIV-1 RT-associated RNA-dependent DNA polymerase activity by 50%; <sup>c</sup> Specificity Index: ratio between compound concentration required to reduce the HIV-1 RT-associated RNase H activity by 50% and compound concentration required to reduce the HIV-1 RT-associated DP activity by 50% (IC <sub>50</sub> RNase H/IC <sub>50</sub> DP); <sup>d</sup> ND, not done.				

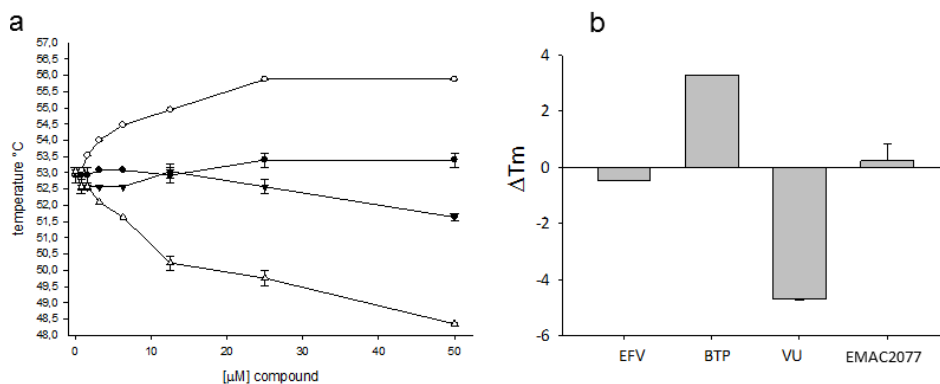
As we did for diarylpropenones, we wanted to verify if this dual activity was due to the binding to a single site or to two different sites.

Knowing that most of the RNase H inhibitors act by chelating the Mg<sup>2+</sup> in the active site<sup>38</sup> we firstly analyzed chelating potential of **EMAC2077** by measuring its UV spectra in the absence and in the presence of magnesium. Results showed that **EMAC2077** maximum of absorbance did not shift in the presence of 6 mM MgCl<sub>2</sub>, excluding the involvement of chelation in the mechanism of action (Figure 26). Therefore, the RNase H activity of **EMAC2077** is not due to its binding in the catalytic site, as we previously have seen with **EMAC2005**.



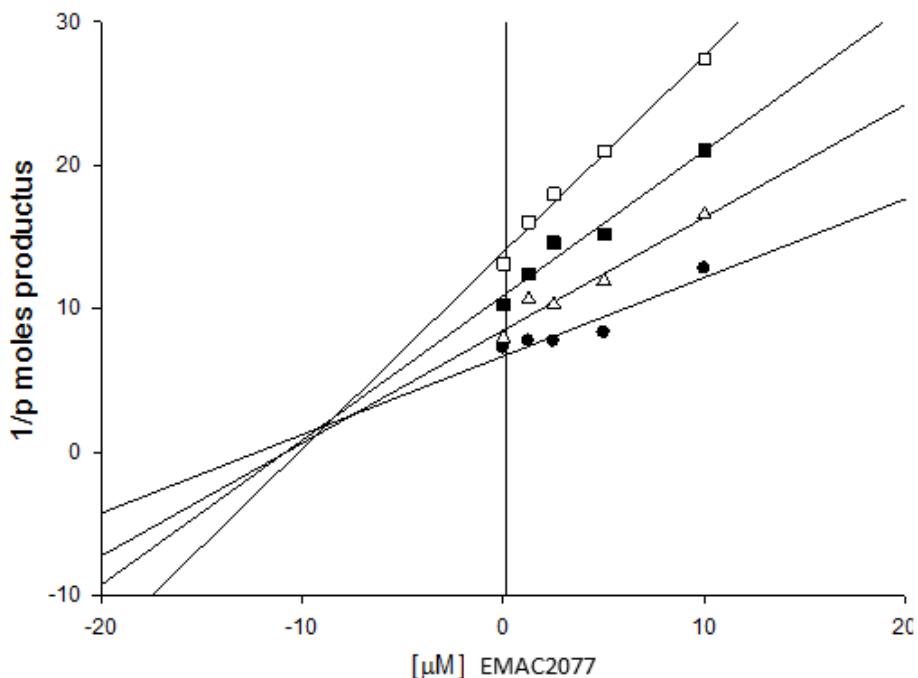
**Figure 26.** Effect of  $\text{MgCl}_2$  on the spectrum of absorbance of RHIs. Chelation of  $\text{Mg}^{2+}$  by **RHS1643** (panel A), **Compound 46** (panel B) and **EMAC2077** (panel C). UV-vis spectrum was measured with compound alone (unbroken line) or in the presence of 6 mM  $\text{MgCl}_2$  (dotted line).

RNase H inhibitors such as vynologous ureas (VUs) are known to destabilize the RT heterodimer by binding to an allosteric pocket in the RNase H domain at the interface between p66 and p51 subunits. Docking studies suggested that the residue C280 in helix  $\alpha$ I of p51 subunit was particularly important for ligands binding.<sup>57</sup> Therefore, we examined alterations in HIV-1 RT thermal stability by differential scanning fluorimetry<sup>58</sup> in the presence of increasing concentrations of **EMAC2077** as well as known inhibitors such as the NNRTI efavirenz (EFV), the RNase H active-site inhibitor hydroxytropolone  $\beta$ -thujaplicinol (BTP) and the allosteric RNase H inhibitor 2-(3,4-dihydroxyphenyl)-5,6-dimethylthieno[2,3-d]pyrimidin-4(3H)-one (VU).<sup>59</sup> In agreement with previous studies,<sup>57,60</sup> the RNase H active-site inhibitor BTP, that has been shown to stabilize the RT against thermal denaturation,<sup>60</sup> causing a  $T_m$  increase of  $> 2.0$  °C in the presence of  $\text{Mg}^{2+}$ , while the interface inhibitor VU, that has been shown to destabilize the HIV-1 RT<sup>57</sup> decreasing the  $T_m$  by 5.5 °C (Figure 27). In contrast, **EMAC2077** did not affect significantly RT thermal stability, showing behavior similar to EFV and suggesting that **EMAC2077** may have an allosteric binding mode different from VU but possibly similar to the one shown by EFV (Figure 27).



**Figure 27.** Effect of RT inhibitors on the thermal stability of p66/p51 HIV-1 RT. a) The melting temperature of HIV-1 RT was measured by differential scanning fluorimetry in presence of increasing concentrations of different inhibitors: (▼) EFV, (○) BTP, (Δ) VU and (●) **EMAC2077**. b) Maximum HIV-1 RT thermal shift ( $\Delta T_m$ ) observed in the presence of 50  $\mu\text{M}$  concentration of compounds.  $\Delta T_m$  values are the average of triplicate analysis, standard deviations are indicated as bars.

To investigate the **EMAC2077** mode of action with respect to EFV, we performed a Yonetani-Theorell analysis<sup>40</sup> on the combined effects of **EMAC2077** and EFV on RDDP function. Such an analysis reveals whether simultaneous binding (or inhibition) of two compounds is possible or not. Results (Figure 28) showed that **EMAC2077** and EFV inhibition are not mutually exclusive. However, it is worth to note that the calculated interaction constant  $\alpha$  had the value of 1.2, suggesting a negative interference between the two compounds (i.e. EFV binding has a negative influence on **EMAC2077** binding, and vice versa) suggesting a close binding site.



**Figure 28.** Yonetani–Theorell plot of the combination of **EMAC2077** and EFV on the HIV-1 RT RNA-dependent DNA polymerase activity. HIV-1 RT was incubated in the presence of **EMAC2077** alone (●) or in presence different concentrations of EFV: 4nM (Δ), 8 nM(■), 16nM (□).

Furthermore, we tested **EMAC2077** against several mutants conferring resistance to NNRTIs such as K103N, Y181C and Y188L, and against the HIV-1 group O RT which shows natural resistance to NNRTIs<sup>61</sup> due to the presence of the amino acid substitutions A98G, V179E and Y181C (Figure 29).

The DNA polymerase activity of HIV-1 group O RT as well as the one of wild-type and mutant of HIV-1 group M subtype B RTs were all susceptible to **EMAC2077** in RDDP assays ( $IC_{50}$  values in the range 5.4-20.1  $\mu$ M, for a maximum 2 fold increases in the  $IC_{50}$  values relative to the wild type HIV-1 group M subtype B RT). In contrast, HIV-1 group O RT and mutants K103N and Y188L showed decreased susceptibility to EFV (of 7.7- to 9.2-fold increases in the  $IC_{50}$  values) (Table 6). These observations suggest different modes of action for **EMAC2077** and EFV.

Interestingly, the HIV-1 group O RT was also inhibited by BTP and **EMAC2077** in RNase H activity assays, although  $IC_{50}$  values were 5 to 6.5 times higher than those obtained with the prototypic wild-type HIV-1 group

M subtype B RT. These effects could be addressed to differences that affect the overall structure of the RNase H domain in both enzymes. Previous studies showed that HIV-2 RT was about 3.7 times less susceptible to BTP than the HIV-1 enzyme,<sup>62</sup> while the crystal structure of BTP bound to HIV-1 group M subtype B RT revealed that major interactions involved in inhibitor binding affected RNase H active site residues.<sup>28</sup> However, RNase H active site residues are conserved in all HIV-1 and HIV-2 clades. Outside the active site, the most significant differences between group M subtype B and group O HIV-1 RTs are found around positions 460-471, 482-492, and 502-511 (Figure 29).



```

GroupM-subB      PISPIETVPVKLPGMDGPKVKQWPLTEEKIKALVEICTEMEKEGKISKIGPENPYNTPV 60
GroupO           PISPIAPVPVKLPGMDGPKVKQWPLSKEKIEALTAICQEMEQEKGKISRIGPENPYNTPI 60
***** .*****::**:* * **:******:*****:

GroupM-subB      FAIKKDDSTKWRKLVDFRELNKRTQDFWEVQLGIPHPAGLKKKSVTLVDGDAYFSVPL 120
GroupO           FAIKKDDSTKWRKLVDFRELNKRTQDFWEVQLGIPHPGGLKQKRSVTLVDGDAYFSCPL 120
*****:*****:*****:*****:*****

GroupM-subB      DEDFRKYTAFTIPSVNNETPGIRYQYNVLPQGWKGSPAIFQSSMTKILEPFRKQNPDIVI 180
GroupO           DPDFRKYTAFTIPSVNNETPGVRYQYNVLPQGWKGSPAIFQSSMTKILDPRKDNPELEI 180
* *****:*****:*****:*****:*****:*****:*****

181
GroupM-subB      YQYMDL YVGS DLEIGQHRTKIEELRQHLLRWGLTTPDKKHQK EPPFLWMGYELHPDKWT 240
GroupO           CQYMDL YVGS D LPLAEHRKRVESLREHLYQWGF TTPDKKHQK EPPFMWGYELHPDKWT 240
***** .:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*

GroupM-subB      VQPIVLEKDSWTVNDIQKLVGLMNASQIYPGKIVRQLCKLLRGTKALTEVIPLTEAE 300
GroupO           VQPIKLPNKDWTVNDIQKLVGLMNASQIYQGIRVRELCCLIRGTSKSLTEVWPLSKEAE 300
**** *:*:* *****:***** *:*:*:*:*:*:*:*:*:*:*:*:*

GroupM-subB      LELAENREILKEPVHGVVYDPSKDLIAEIQQQGQGWTYQIYQEPFKNLKTGKYARMRGA 360
GroupO           MELEENREKLKEPVHGVVYQPKDLWNVIQKQEGEGWTYQIYQDEHKNLKTGKYTRQKAS 360
:* * * * * *****:* * * * * :*****:*****: *****:* * * *

GroupM-subB      HTNDVKQLTEAVQKITTESIWIWGKTPKFKLPIQKETWETWTEYVQATWIPEWVNTNP 420
GroupO           HTNDIRQLAEVIQKVSQESIWIWGKLPKFKLPVTRETRWADYVQATWIPEWDYVSTP 420
*****:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*

GroupM-subB      PLVKLWYQLEKEPIVGAETFYVDGAANRETKLGKAGYVTRGRQKVVTLTDTTNQKTELQ 480
GroupO           PLIKLWYRLESEPIRGAETFYVDGAANRDTKLGKAGYVTEQGKQKIIKLDETTNQKAEML 480
**:*:*:*:*:* * * * * * *****:*****:*****:*****:*****

GroupM-subB      AIYLALQDSGLEVNIVTDSQYALGIIQAQPDSSELVNQIIEQLIKKVKVYLAWPAHK 540
GroupO           AVLLALQDSKEKVNIVTDSQYVGLGIISQPSSESPIVQQIIEELTKKEQVYLAWPAHK 540
*:* *****:*****:*****:*****:*****:*****:*****

GroupM-subB      GIGGNEQVDKLVSAGIRKVL 560
GroupO           GIGGNEKIDKLVSKDIRRVL 560
*****:***** .:*:*

```

**Figure 29.** Clustal O alignment<sup>63</sup> of the Group M and Group O sequences. In yellow are marked mutated residues located in **EMAC2077** putative binding sites. An \* (asterisk) indicates positions which have a single, fully conserved residue. A : (colon) indicates conservation between groups of strongly similar properties; A . (period) indicates conservation between groups of weakly similar properties

**Table 6.** Susceptibility to **EMAC2077** of wt and NNRTI-resistant HIV-1 RTs measured in RNase H and RDDP activity assays.

RT	EMAC2077		BTP	EFV
	RNase H	RDDP	RNase H	RDDP
	IC <sub>50</sub> ( $\mu$ M) <sup>a</sup>	IC <sub>50</sub> ( $\mu$ M) <sup>b</sup>	IC <sub>50</sub> ( $\mu$ M) <sup>a</sup>	IC <sub>50</sub> (nM) <sup>b</sup>
<b>wt</b>	1.3 $\pm$ 0.3	9.8 $\pm$ 1.4	0.19 $\pm$ 0.03	23 $\pm$ 2.7
<b>K103N</b>	2.3 $\pm$ 0.1	13.6 $\pm$ 1.0	0.22 $\pm$ 0.08	176 $\pm$ 25
<b>Y181C</b>	2.1 $\pm$ 0.6	5.4 $\pm$ 0.3	0.23 $\pm$ 0.05	49.7 $\pm$ 9.1
<b>Y188L</b>	1.6 $\pm$ 0.5	16.5 $\pm$ 3.7	0.08 $\pm$ 0.05	198 $\pm$ 60
<b>Group O</b>	8.5 $\pm$ 2.6	20.1 $\pm$ 7.4	0.91 $\pm$ 0.01	212 $\pm$ 46
<sup>a</sup> Compound concentration required to reduce the HIV-1 RT-associated RNase H activity by 50%; <sup>b</sup> Compound concentration required to reduce the HIV-1 RT-associated RNA-dependent DNA polymerase activity by 50.				

### 2.3.2 Computational studies

To achieve further insights into the **EMAC2077** binding mode, we performed blind docking studies on the wt HIV-1 group M subtype B RT heterodimer and **EMAC2077**, by using the QM-polarized ligand docking protocol (QMPLD).<sup>64</sup>

As we did for the diarylpropenones, we carried out ensemble docking experiments, due to the flexibility of the target and different shapes of known inhibitors (Figure 13).<sup>36</sup> The major conformational changes in the NNRTIs binding pocket were taken into account to cluster the available RT complexes. In particular, the orientation of amino acid residues Y181, Y188, Y183 and primer grip  $\beta$ 12- $\beta$ 13 hairpin were considered.<sup>42</sup> A representative of each different cluster was picked and the three-dimensional structure of HIV RT was retrieved from the Protein Data Bank. The obtained [**EMAC2077**•RT] complexes were then subjected to a post-docking procedure, based on energy minimization and successive binding free

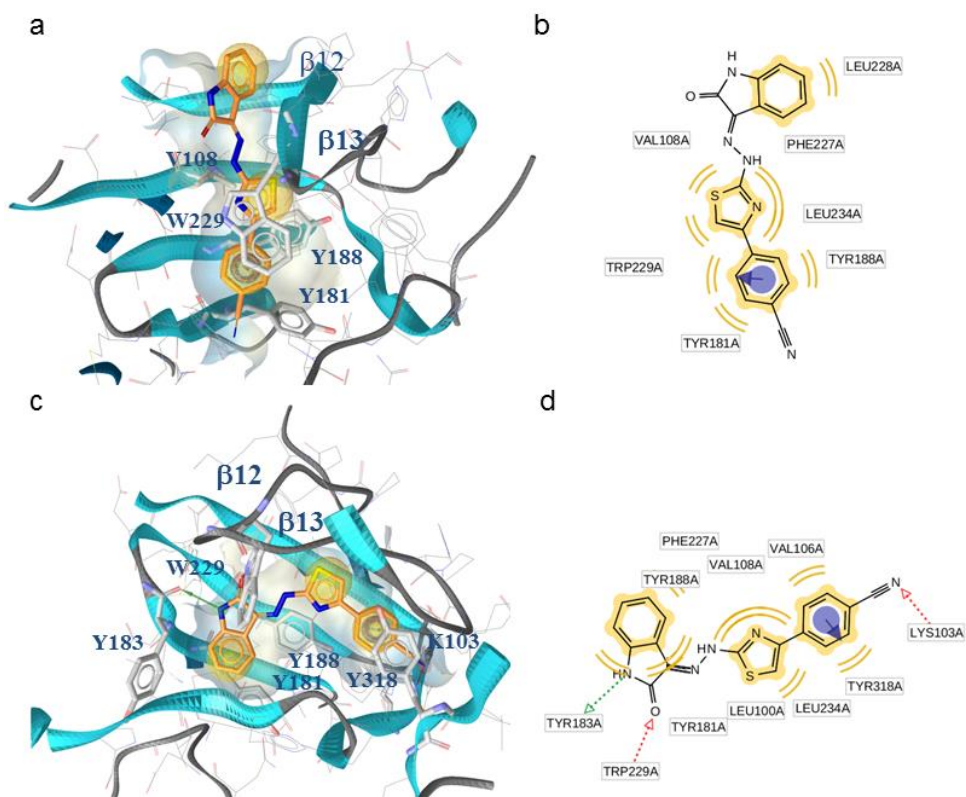
energies calculation. The binding free energies ( $\Delta G(\text{Bind})$ ) were obtained by applying molecular mechanics and continuum solvation models using the molecular mechanics generalized Born/surface area (MM-GBSA) method.<sup>46</sup> Blind docking calculations indicated the presence of two energetically favored binding pockets sites (Figure 18) for indolinone derivative **EMAC2077** (Table 7).

**Table 7.** Ensemble docking results: binding free energies of [RMNC•RT] complexes. The most likely binding poses are indicated in bold.

RT Xray (pdb code)	Pocket	$\Delta G_{\text{MMGBSA}}$ (Kcal*mol)	Best Poses
<b>1ep4</b>	1	-41.23	
<b>1ep4</b>	2	-36.22	
<b>3lp2</b>	1	<b>-44.85</b>	A
<b>3lp2</b>	2	-42.58	
<b>1rti</b>	1	-33.26	
<b>1rti</b>	2	-40.49	
<b>1tv6</b>	1	<b>-51</b>	B
<b>1tv6</b>	2	<b>-48.74</b>	C
<b>1vrt</b>	1	-35.9	
<b>1vrt</b>	2	-32.44	
<b>2zd1</b>	1	-41.96	
<b>2zd1</b>	2	-32.06	
<b>3q09</b>	1	-39.25	
<b>3q09</b>	2	-38.48	

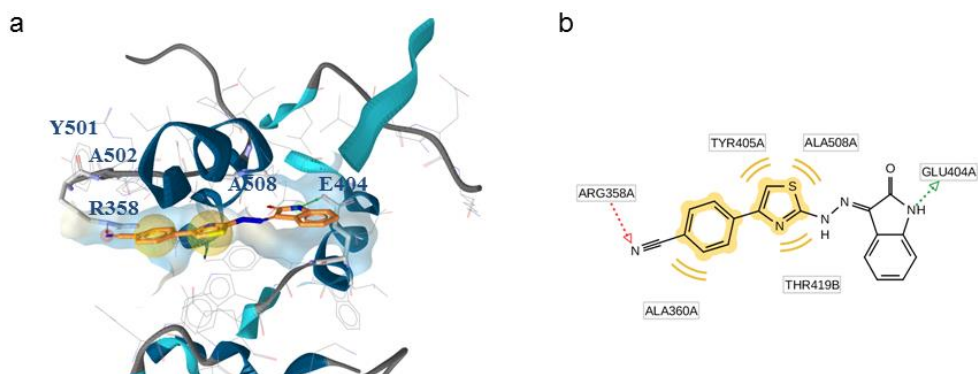
Pocket 1 is located close to the DNA polymerase catalytic center and is contiguous to the NNRTIs binding pocket, having an “L shape” (Figure 14). According to the first orientation (pose A) (Figure 30a, b), the compound is accommodated in a pocket located between polymerase catalytic domain and NNRTIs binding pocket. This pocket was recently described as allosteric pocket of HIV-1 RNase H function.<sup>27-29</sup> However the most stable pose (Figure 30 c, d) involves amino acids L100, K103, V106, V108, Y183, Y188, L234, W229 and Y318 (Pose B). According to both putative binding modes (A and B), different residues of the hairpin constituted by the  $\beta$ 12 and  $\beta$ 13 sheet<sup>47</sup>

are involved with **EMAC2077** binding and for this reason were subjected to alanine scanning mutagenesis (see below paragraph 2.3.3).



**Figure 30.** Putative binding modes of **EMAC2077** and critical residues individuated for **EMAC2077** binding in the pocket 1: a) binding mode A; b) 2D depiction of **EMAC2077** and its respective interactions with RT residues. c) binding mode B pale yellow sphere indicates hydrophobic interactions with lipophilic residues. Red arrow indicates an hydrogen bond (HB) acceptor interaction, green HB donor, while the violet sphere represents the aromatic  $\pi$ - $\pi$  stacking interaction.

The second putative binding pocket (pocket 2) is located in the RNase H domain, between the RNase H active site and the primer grip region, close to the p66/p51 interface. This pocket is the same pocket found during docking calculations of compound **EMAC2005** and discussed before.<sup>36</sup> Docking experiments suggest that, in this site, **EMAC2077** could be partially sandwiched between different secondary structural units, such as namely  $\beta$ 21 strand and the  $\alpha$ H helix in p51, and the  $\alpha$ B helix in p66 (residues 500-508) (Figure 31).



**Figure 31.** a) Putative binding mode of **EMAC2077** and interacting residues in pocket 2.b) 2D depiction of **EMAC2077** and its respective interactions with RT residues.

Overall, such *in silico* analysis support alternative modes of action for **EMAC2077**, since the compound could bind (i) preferentially only to one of the two pockets and have both short- and long-range effects; or (ii) to both sites and act by short-range effects on the two functions. According to the first mode of action, **EMAC2077** would inhibit RDDP and RNase H activities through binding into the sole pocket 1, by having a short-range inhibitory effect on RDDP activity and a long-range inhibitory effect on RNase H activity binding in the RNase H allosteric site (pose A, Figure 30 a,b). Hence, it would act differently from classical NNRTIs such as nevirapine and EFV that were shown to destabilize the 3'-end of the DNA primer in the DNA polymerase active site and promote RT-mediated polymerase-independent RNase H cleavages.<sup>65</sup>

Alternatively, according to the first mode of action, **EMAC2077** could bind to the sole pocket 2 and it would have a short-range inhibitory effect on RNase H activity and a long-range effect on the RDDP activity.

Differently, according to the second mode of action **EMAC2077** would bind to both individual pockets and its activity would be due to short-range inhibition effects, so that its binding to each site would be responsible for the inhibition of one function.

To dissect this diverse scenario of possibilities we performed site-directed mutagenesis studies.

### 2.3.3 Single point site directed mutagenesis in pocket 1

Considering the size of pocket 1, its surface was explored in detail by site-directed mutagenesis. Since the hairpin comprising strands  $\beta$ 12 and  $\beta$ 13 was identified as potentially involved in **EMAC2077** binding alanine-scanning mutagenesis was performed for amino acids 224-231. Mutants V108A and V106A were also obtained since Val108 and Val106 were identified as potentially important for inhibitors binding (pose A and pose B, Figure 30). The susceptibility of mutant RTs to **EMAC2077** was also tested in RNase H and RDDP activity assays, using BTP and EFV as positive controls (Table 8). Interestingly, the E224A RT showed a 6-fold reduced susceptibility to **EMAC2077** in RNase H assays, while mutant enzymes such as V108A and V106A showed  $IC_{50}$  values 4 and 10 times higher, respectively, in comparison with the wt RT. In contrast, we did not observed major differences in the  $IC_{50}$  values obtained with other mutant RTs (Table 8). When the RDDP function of mutant RTs was assayed, a similar pattern emerged, even though the extent of the reduction in the  $IC_{50}$  values was lower (1.4, 3.3 and 2.5  $\mu$ M for E224A, V106A and V108A RTs, respectively).

**Table 8.** Effects of selected amino acid substitutions in pocket 1 of HIV-1 RT in the susceptibility to **EMAC2077** in RNase H and RDDP activity assays

RT	EMAC2077		BTP	EFV
	RNase H	RDDP	RNase H	RDDP
	IC <sub>50</sub> (μM) <sup>a</sup>	IC <sub>50</sub> (μM) <sup>b</sup>	IC <sub>50</sub> (μM) <sup>a</sup>	IC <sub>50</sub> (nM) <sup>b</sup>
<b>V106A</b>	5.3 ± 0.6 (4.1) <sup>c</sup>	25.1 ± 2.3 (2.5)	0.16 ± 0.03 (0.8)	35.4 ± 2.2 (1.5)
<b>V108A</b>	13.1 ± 2.5 (10)	33.0 ± 5.2 (3.4)	0.18 ± 0.04 (0.9)	21.3 ± 3.6 (0.9)
<b>Y188A</b>	3.3 ± 1.5 (2.5)	20.9 ± 4.0 (2.1)	0.19 ± 0.08 (1.0)	28.3 ± 7.4 (1.2)
<b>E224A</b>	7.9 ± 1.1 (6.1)	13.8 ± 4.2 (1.4)	0.15 ± 0.03 (0.8)	24.9 ± 0.7 (1.1)
<b>P225A</b>	1.7 ± 0.2 (1.3)	3.1 ± 0.3 (0.3)	0.13 ± 0.06 (0.7)	19.5 ± 2.9 (0.8)
<b>P226A</b>	1.9 ± 0.1 (1.5)	18.9 ± 0.7 (1.9)	0.29 ± 0.08 (1.5)	18.9 ± 2.4 (0.9)
<b>F227A</b>	0.7 ± 0.1 (0.5)	15.2 ± 0.9 (1.5)	0.22 ± 0.04 (1.2)	61.5 ± 1.4 (2.7)
<b>L228A</b>	1.5 ± 0.1 (1.1)	13.1 ± 1.9 (1.3)	0.10 ± 0.06 (0.5)	21.8 ± 3.8 (1.0)
<b>W229A</b>	1.9 ± 0.5 (1.5)	15.4 ± 2.8 (1.6)	0.15 ± 0.08 (0.8)	22.9 ± 5.2 (1.1)
<b>M230A</b>	0.5 ± 0.1 (0.4)	15.0 ± 3.5 (1.5)	0.14 ± 0.04 (0.7)	29.1 ± 2.2 (1.3)
<b>G231A</b>	1.7 ± 0.4 (1.3)	10.8 ± 2.4 (1.1)	0.25 ± 0.03 (1.3)	67.7 ± 3.5 (2.9)

<sup>a</sup>Concentration required to inhibit HIV-1 RT-associated RNase H activity by 50% obtained by three independent experiments (reported as average ± standard deviation).  
<sup>b</sup>Concentration required to inhibit HIV-1 RT-associated RDDP activity by 50% obtained by three independent experiments (reported as average ± standard deviation).  
<sup>c</sup>Fold of increase with respect to wt RT.

Hence, the results of the assays with the mutant RTs did not provide incontrovertible evidence of **EMAC2077** binding in one position. Instead the results could not exclude the **EMAC2077** binding to pocket 1 in two different ways. In fact, given the amplitude of the pocket and the high RT flexibility, the very limited IC<sub>50</sub> reduction towards RDDP function observed could be determined by a different orientation of **EMAC2077** while accommodating in pocket 1, as a response to the introduced amino acid changes. It is worth noting that the kinetic analysis of the interaction between **EMAC2077** and EFV, showed a negative interference between the two compounds, even if they are not kinetically mutually exclusive, supporting the possibility that

**EMAC2077** could accommodate in the large pocket 1 even in the presence of EFV. Hence these mutagenesis experiments seem to confirm the binding modes predicted by docking studies.

#### 2.3.4 Single point site directed mutagenesis in pocket 2

Pocket 2, located in the RNase H domain, was already predicted<sup>51</sup> and investigated in the context of hydrazone derivatives binding studies<sup>53,66</sup> and diarylpropenones described above (paragraph 2.2).<sup>36</sup> According to our modeling results, when bound to this site, **EMAC2077** might nudge the RNase H domain to a position in which the active site might no longer be able to catalyze hydrolysis cleavage of the RNA strand in the of RNA:DNA duplex. To investigate possible **EMAC2077** binding into this pocket, we mutated residues Ala502 to Phe and Ala508 to Val in an attempt to reduce the space available for **EMAC2077** accommodation. Hence, we tested **EMAC2077** effects on both RNase H and RDDP functions of mutant RTs, using BTP and EFV as positive controls (Table 9). The RNase H function of mutant RTs A502F and A508V showed 4- to 10-fold reduced susceptibility to **EMAC2077** in RNase H assays.



**Table 9.** Effects of selected amino acid substitutions in pocket 2 of HIV-1 RT in the susceptibility to **EMAC2077** in RNase H and RDDP activity assays

RT	EMAC2077		BTP	EFV
	RNase H	RDDP	RNase H	RDDP
	IC <sub>50</sub> (μM) <sup>a</sup>	IC <sub>50</sub> (μM) <sup>b</sup>	IC <sub>50</sub> (μM) <sup>a</sup>	IC <sub>50</sub> (nM) <sup>b</sup>
<b>A502F</b>	13.0 ± 0.7 (10.0)	17.1 ± 2.1 (1.8)	0.17 ± 0.03 (0.9)	22.3 ± 1.5 (1.6)
<b>A508V</b>	6.5 ± 0.7 (5.0)	19.3 ± 2.8 (2.0)	0.16 ± 0.05 (0.8)	24.7 ± 2.4 (1.8)

<sup>a</sup>Concentration required to inhibit HIV-1 RT-associated RNase H activity by 50% obtained by three independent experiments (reported as average ± standard deviation).  
<sup>b</sup>Concentration required to inhibit HIV-1 RT-associated RDDP activity by 50% obtained by three independent experiments (reported as average ± standard deviation).  
<sup>c</sup>Fold of increase with respect to wt RT.

These results strongly support the hypothesis that **EMAC2077** interacts with pocket 2 and that this binding is responsible for short-range inhibition of the RNase H function.

With respect to pocket 2, the sequence alignment of HIV-1 RTs of group O and group M subtype<sup>67, 68</sup> revealed many differences (Figure 29). Thus, comparing the pocket 2 residues of prototypic group M subtype B enzyme versus HIV-1 group O RT we have the following differences: E404D, A502V, A508S and D511T. The 6.5 folds reduction in **EMAC2077** susceptibility obtained with HIV-1 group O RT in RNase H activity assays (Table 6) is consistent with the effects observed with single-mutants A502F or A508V (Table 9), since amino acid substitutions found in the HIV-1 group O RT would also contribute to reducing the size of the putative **EMAC2077** binding pocket in the RNase H domain. In addition, the NNRTI binding pocket of the HIV-1 group M subtype B RT differs from that of the group O enzyme mainly at the 181 position (Cys in HIV-1 group O RT and Tyr in HIV-1 group M subtype B RT), but this does not affect the inhibitory efficiency of **EMAC2077**. Taken together, our data reveal a relevant role for pocket 2 in **EMAC2077** binding.

### 2.3.5 Conclusions

In conclusion, our study demonstrates that **EMAC2077** is a novel dual function RT inhibitor that has an allosteric binding mode, different from that shown by known NNRTIs. Our results support the hypothesis that **EMAC2077** binds two different sites in HIV-1 RT. This double-site binding mode seems to confirm that it could be possible to target both RT-associated functions by a single molecule, retaining full potency of inhibition on drug-resistant mutant RTs. RNase H inhibition by **EMAC2077** seems to be mainly, but not exclusively, due to its interaction with pocket 2, close to the RNase H active site. The long-range effects on RNase H inhibition produced by **EMAC2077**, and observed when amino acids in the polymerase domain were mutated, are probably due to an alteration of the substrate binding pocket in the RNase H domain that, in turn, alters **EMAC2077** binding or efficacy. On the contrary, HIV-1 RT-associated RDDP inhibition by **EMAC2077** seems to be related to its binding in the polymerase domain in the NNRTI binding pocket. However, the size of the pocket and the plasticity of the enzyme may be responsible for the conservation of the RDDP inhibition of the compound on mutant RTs.

## 2.4 SUMMARY

Although this target has been studied since more than 30 year for drug development, it is still actual and offers new opportunity to design inhibitors with new mechanism of action. Its flexibility, the presence of two catalytic functions make this enzyme hard to examine with both biochemical and computational methods. Nevertheless, considering both series synthesised and studied in this project we can assert that the overall, the dual site-dual RT-functions inhibition by these compounds is an attractive possibility to strongly reduce drug resistance occurrence. Furthermore, since pocket 2 is specific for HIV RT, compounds that target this site will not interfere with human RNase H1. This pocket was deeply investigated in the past, but seems to be promising. These results prompt us to undergo further studies to better define the compounds interactions within these two pockets in

order to perform a rational drug design of dual inhibitors acting on both binding sites.

## 2.5 EXPERIMENTAL PART

### 2.5.1 Synthesis and characterization

**Material and General methods.** Starting materials and reagents were obtained from commercial suppliers and were used without purification. All melting points were determined by the capillary method on a Büchi-540 capillary melting points apparatus and are uncorrected. Electron ionization mass spectra were obtained by a Fisons QMD 1000 mass spectrometer (70 eV, 200 mA, ion source temperature 200°C). Samples were directly introduced into the ion source. Found mass values are in agreement with theoretical ones.

For the first series all samples were measured in CDCl<sub>3</sub> at 278.1 K temperature on a Varian Unity 300 spectrometer. In the signal assignments the proton chemical shifts are referred to the solvent (1H:  $\delta = 7.24$  ppm,).

For the second all samples were measured in DMF-*d*<sub>7</sub> or DMSO-*d*<sub>6</sub> solvent at 278.1 K temperature on a Bruker AVANCE III 500 MHz spectrometer. Chemical shifts are reported relative to TMS ( $\delta = 0$ ) and/or referenced to the solvent in which they were measured. (In the <sup>15</sup>N chemical shift assignments we applied the spectrometer's digital reference which is calibrated to liq. NH<sub>3</sub>  $\delta = 0$  ppm.) Coupling constants *J* are expressed in hertz (Hz).

Both series <sup>13</sup>C were recorded on a Varian Unity 500 spectrometer using CDCl<sub>3</sub> DMF-*d*<sub>7</sub> or DMSO-*d*<sub>6</sub> as a solvent at 278.1 K.

Elemental analyses were obtained on a Perkin–Elmer 240 B microanalyser. Analytical data of the synthesised compounds are in agreement within  $\pm 0.4$  % of the theoretical values. TLC chromatography was performed using silica gel plates (Merck F 254), spots were visualised by UV light.

**Synthetic procedures:** 1,3-Diarylpropenones were synthesised according to a slightly modified Claisen-Schmidt reaction (Scheme 1). NMR

analysis supports the "E" configuration, according to the double bond protons coupling constants that ranges from 15 to 16 Hz.

#### General procedure.

*(E)*-1-(4-bromophenyl)-3-(1-methoxynaphthalen-2-yl)prop-2-en-1-one (**EMAC2000**). 0.9 g of 1-(4-bromophenyl)ethanone (4.5 mmol) were dissolved in ethanol and a solution of NaOH 10% was added dropwise at r.t. The mixture was stirred for 10 min then 1 g of 1-methoxy-2-naphthaldehyde (5.4 mmol) in ethanol solution was added. 24 hours later the reaction was completed (monitored by TLC, eluent: n-hexane: ethyl acetate 2:1), and a pale yellow crystalline solid was filtered, washed with water, crystallised by a mixture of water-ethanol, and characterised.

Yellow crystal, Yield 67%, mp: 110-112°C. <sup>1</sup>H-NMR (300 MHz, CDCl<sub>3</sub>): δ= 4.06 (s, 3H, OCH<sub>3</sub>), 7.3 (d, 1H, J: 9.1, Ar-CH), 7.41 (t, 1H, J: 7.5, Ar-CH), 7.55 (t, 1H, J: 7.5, Ar-CH), 7.65 (d, 2H, J: 8.3, Ar-CH), 7.82 (d, 1H, J: 7.5, Ar-CH), 7.85 (d, 1H, J: 15.6, -CH=), 7.9 (d, 1H, J: 9.0, Ar-CH), 7.93 (d, 2H, J: 8.3, Ar-CH), 8.25 (d, 1H, J: 8.6, Ar-CH), 8.51 (d, 1H, J: 15.6, -CH=). <sup>13</sup>C-NMR (100 MHz, CDCl<sub>3</sub>): δ= 56.1, 112.7, 117.1, 123.3, 124.1, 126.6, 127.6, 128.7, 129.0, 130.0, 130.1 (2C), 131.8 (2C), 132.1, 133.1, 137.3, 138.3, 157.2, 190.2.

According to this general method, the following compounds were synthesised:

*(E)*-1-(4-fluorophenyl)-3-(1-methoxynaphthalen-2-yl)prop-2-en-1-one (**EMAC 2001**). Yellow crystal, Yield 81%, mp: 93-95°C. <sup>1</sup>H-NMR (300 MHz, CDCl<sub>3</sub>): δ= 4.07 (s, 3H, OCH<sub>3</sub>), 7.11 (t, 1H, J: 8.5, Ar-CH), 7.26 (d, 2H, J: 9.1, Ar-CH), 7.42 (d, 1H, J: 7.1, Ar-CH), 7.50 (t, 1H, J: 8.5, Ar-CH), 7.57 (d, 1H, J: 7.1, Ar-CH), 7.78 (d, 1H, J: 15.9, -CH=), 7.93 (d, 1H, J: 8.6, Ar-CH), 8.09 (t, 2H, J<sub>H-H</sub>: 9.1, J<sub>H-F</sub>: 9.3, Ar-CH), 8.18 (d, 1H, J: 8.5, Ar-CH), 8.43 (d, 1H, J: 15.9, -CH=). <sup>13</sup>C-NMR (100 MHz, CDCl<sub>3</sub>): δ= 56.5, 112.7, 115.7, 123.3, 124.0, 126.8, 127.6, 128.6 (2C), 129.0, 130.0 (2C), 131.1, 131.3, 132.0, 133.0, 136.2, 138.3, 157.2, 190.0.

*(E)*-1-(4-methoxyphenyl)-3-(1-methoxynaphthalen-2-yl)prop-2-en-1-one (**EMAC 2002**). Yellow crystal, Yield 83%, mp: 137-139°C. <sup>1</sup>H-NMR (300 MHz, CDCl<sub>3</sub>): δ= 3.9 (s, 3H, OCH<sub>3</sub>), 4.05 (s, 3H, OCH<sub>3</sub>), 7.00 (d, 1H, J: 8.9, Ar-CH), 7.33 (d, 2H, J: 9.0, Ar-CH), 7.70 (t, 1H, J: 8, Ar-CH), 7.47 (d, 1H, J: 16.0, -CH=), 7.55 (t, 1H, J: 8, Ar-CH), 7.82 (d, 1H, J: 8.0, Ar-CH), 7.89 (d, 1H, J: 9.0, Ar-CH), 8.08 (d, 1H, J: 9.0, Ar-CH), 8.28 (d, 2H, J: 9.0, Ar-CH), 8.45 (d, 1H, -CH=). <sup>13</sup>C-

NMR (100 MHz, CDCl<sub>3</sub>):  $\delta$ = 55.4, 56.5, 112.7, 113.5, 117.7, 123.5, 124.1, 127.2, 127.6, 128.6, 129.1, 130.9 (2C), 131.5 (2C), 131.7, 133.1, 137.0, 156.9, 163.3, 189.5.

*(E)*-1-(4-chlorophenyl)-3-(1-methoxynaphthalen-2-yl)prop-2-en-1-one (**EMAC 2003**). Yellow crystal, Yield 53%, mp: 108-109°C. <sup>1</sup>H-NMR (300 MHz, CDCl<sub>3</sub>):  $\delta$ = 4.06 (s, 3H, OCH<sub>3</sub>), 7.33 (d, 1H, J: 9.1, Ar-CH), 7.41 (t, 1H, J: 7.8, Ar-CH), 7.49 (d, 2H, J: 8.4, Ar-CH), 7.53 (d, 1H, J: 15.6, -CH=), 7.55 (t, 1H, J: 7.8, Ar-CH), 7.74 (d, 1H, J: 8.1, Ar-CH), 7.90 (d, 1H, J: 9.6, Ar-CH), 8.01 (d, 2H, J: 8.3, Ar-CH), 8.25 (d, 1H, J: 8.6, Ar-CH), 8.51 (d, 1H, J: 15.6, -CH=). <sup>13</sup>C-NMR (500 MHz, CDCl<sub>3</sub>):  $\delta$ = 56.3, 112.7, 117.1, 123.3, 124.0, 126.7, 127.6, 128.6, 128.8 (2C), 129.0, 130.0 (2C), 132.1, 133.1, 136.9, 138.3, 139.0, 157.2, 190.0.

*(E)*-1-(3-nitrophenyl)-3-(1-methoxynaphthalen-2-yl)prop-2-en-1-one (**EMAC 2004**). Pale orange crystal, Yield 64%, mp: 143-145°C. <sup>1</sup>H-NMR (300 MHz, CDCl<sub>3</sub>):  $\delta$ = 4.10 (s, 3H, OCH<sub>3</sub>), 7.35 (d, 1H, J: 8.9, Ar-CH), 7.42 (t, 1H, J: 7.8, Ar-CH), 7.58 (t, 1H, J: 8.5, Ar-CH), 7.72 (t, 1H, J: 7.8, Ar-CH), 7.83 (d, 1H, J: 8.0, Ar-CH), 7.93 (d, 1H, J: 8.5, Ar-CH), 7.94 (d, 1H, J: 15.6, -CH=), 8.25 (d, 1H, J: 8.5, Ar-CH), 8.39 (d, 1H, J: 8.5, Ar-CH), 8.44 (d, 1H, J: 8.0, Ar-CH), 8.60 (d, 1H, J: 15.6, -CH=), 8.89 (s, 1H, Ar-CH). <sup>13</sup>C-NMR (500 MHz, CDCl<sub>3</sub>):  $\delta$ = 56.3, 112.6, 116.6, 123.1, 123.4, 124.1, 125.7, 126.8, 127.8, 128.8, 129.0, 129.8, 132.7, 133.1, 134.1, 139.5, 140.0, 148.3, 157.6, 189.0.

*(E)*-1-(4-biphenyl)-3-(1-methoxynaphthalen-2-yl)prop-2-en-1-one (**EMAC 2005**). Yellow crystal, Yield 87%, mp: 104-105°C. <sup>1</sup>H-NMR (300 MHz, CDCl<sub>3</sub>):  $\delta$ = 4.07 (s, 3H, OCH<sub>3</sub>), 7.34- 7.50 (m, 2H, Ar-CH), 7.56 (t, 1H, J: 8.1, Ar-CH), 7.69 (d, 2H, J: 7.8, Ar-CH), 7.75 (d, 2H, J: 7.8, Ar-CH), 7.83(d, 1H, J: 8.8, Ar-CH), 7.88 (d, 1H, J: 8.0, Ar-CH), 7.91 (d, 2H, J: 8.2, Ar-CH), 7.95 (d, 1H, J: 15.9, -CH=), 8.04 (d, 2H, J: 8.2, Ar-CH), 8.19 (d, 1H, J: 8.8, Ar-CH), 8.29 (d, 1H, J: 8.1, Ar-CH), 8.54 (d, 1H, J: 15.9, -CH=). <sup>13</sup>C-NMR (100 MHz, CDCl<sub>3</sub>):  $\delta$ = 56.4, 112.8, 117.4, 123.5, 124.0, 127.2 (3C), 127.3 (3C), 127.5, 128.1, 128.6, 128.9 (2C), 129.1, 129.2 (2C), 131.8, 137.3, 137.7, 140.1, 145.3, 157.1, 190.7.

**Synthetic procedures:** Indolinones were synthesised according to the following procedure. NMR analysis supports the “Z” configuration: This was based on the selective NOE experiment, where we observed NOE interaction between the indole NH and CH protons, while no correlation was seen between the indole CH and =N-NH- hydrogens.

**Synthesis of compound EMAC 2072 (Z)-3-(2-(4-(4-chlorophenyl)thiazol-2-yl)hydrazono)indolin-2-one (general procedure)**

The synthesis is accomplished by a two steps procedure (Scheme 2)

**a) Synthesis of 2-(2-oxoindolin-3-ylidene)-hydrazinecarbothioamide**

0,5g (3.8 mmol) of indolinone-2,3-dione and 0,35g (3.8 mmol) of thiosemicarbazide were introduced in a three necked flask and dissolved with 25 mL of 2-propanol at 50°C. 5 drops of CH<sub>3</sub>COOH were added to the mixture reaction as catalyst. After a few minutes the formation of an abundant yellow precipitate is observe. The reaction was monitored with TLC (eluent ethylacetate: exane 7:2). After 24 hours the reaction is completed and the solid filtered. The desired compound is a yellow solid. R.f.: 0.33 (eluent ethylacetate: exane 7:2); M.P.: >250°C; Yield: 97%

<sup>1</sup>H NMR (DMF-d<sub>7</sub>) δ(ppm): 12.72 (s; 1H); 11.34 (s; 1H); 9.21 (s; 1H); 9.05 (s; 1H); 7.66-7.64 (m; 1H); 7.42-7.34 (m; 1H); 7.17-7.10 (m; 1H); 7.06-7.02 (m; 1H). <sup>13</sup>C NMR (DMF) δ(ppm): 180.1; 163.3; 143.1; 132.4; 131.6; 122.8; 121.1; 120.7; 111.5. <sup>15</sup>N NMR (DMF) δ(ppm): 170.0; 135;6; 110.4

**b) Synthesis of (Z)-3-(2-(4-(4-chlorophenyl)thiazol-2-yl)hydrazono)indolin-2-one EMAC 2072**

0.5g (2.3 mmol) of (Z)-1-(2-oxoindolin-3-ylidene)thiosemicarbazide and 0.54g (2.3 mol) of 2-bromo-4'-chloroacetophenone were stirred at r.t. in 30 ml of 2-propanol. After few minutes the formation of an abundant light yellow-orange precipitate was observed. The reaction was monitored by TLC (eluent exane: ethylacetate 7:2) and after 24 hours it was completed. The desired compound is a yellow-orange solid. R.f.:0.85 (eluent exane:ethylacetate 7:2); M

**NMR Final Compounds characterisation indolinones series:**

**(Z)-3-(2-(4-(4-chlorophenyl)thiazol-2-yl)hydrazono)indolin-2-one (EMAC2072).** Yellow-orange solid, Yield 73%, mp: >250°. <sup>1</sup>H NMR (DMF-d<sub>7</sub>) δ(ppm): 13.55 (s, 1H); 11.43 (s; 1H); 8.07-8.02 (m; 2H); 7.82 (s; 1H); 7.63 (dm; J=7.6 Hz; 1H); 7.58-7.53 (m; 2H); 7.41 (td; J=7.6 Hz; 1.1 Hz; 1H); 7.16 (td; J=7.6 Hz; 1.0 Hz; 1H); 7.10 (dm; J=7.6 Hz; 1H). <sup>13</sup>C NMR (DMF-d<sub>7</sub>) δ(ppm): 166.9; 164.0; 150.6; 142.1; 133.6; 133.1; 132.7; 130.9; 129.2; 127.9; 122.8; 120.5; 120.3; 111.5; 107.7. <sup>15</sup>N NMR (DMF-d<sub>7</sub>) δ(ppm): 156.2; 135.7

*(Z)-3-(2-(4-(4-fluorophenyl)thiazol-2-yl)hydrazono)indolin-2-one*  
**(EMAC2073)**. Yellow-orange solid, Yield 57%, mp: >250°. <sup>1</sup>H NMR (DMSO-d6) δ(ppm): 13.33 (s; 1H); 11.25 (s; 1H); 7.97-7.90 (m; 2H); 7.60 (s; 1H); 7.53 (dm; *J* = 7.5 Hz; 1H); 7.34 (tm; *J* = 7.5 Hz; 1H); 7.29-7.17 (m; 2H); 7.09 (tm; *J* = 7.5 Hz; 1H); 6.99-6.92 (m; 1H). <sup>13</sup>C NMR (DMSO-d6) δ(ppm): 166.4; 163.2; 161.9 (d; *J* = 244.6 Hz); 150.0; 141.3; 132.2; 130.6; 130.5; 127.8 (d; *J* = 8.3 Hz); 122.4; 119.9; 119.8; 115.6 (d; *J* = 21.6 Hz); 111.1; 106.6. <sup>15</sup>N NMR (DMSO-d6) δ(ppm): 137.1

*(Z)-3-(2-(4-(4-bromophenyl)thiazol-2-yl)hydrazono)indolin-2-one*  
**(EMAC2074)**. Yellow-orange solid, Yield 95%, mp: >250°. <sup>1</sup>H NMR (DMF-d7) δ(ppm): 13.53 (s; 1H); 11.42 (s; 1H); 7.99-7.93 (m; 2H); 7.82 (s; 1H); 7.71-7.65 (m; 2H); 7.61 (dm; *J* = 7.6 Hz; 1H); 7.39 (tm; *J* = 7.6 Hz; 1H); 7.14 (tm; *J* = 7.6 Hz; 1H); 7.08 (dm; *J* = 7.6 Hz; 1H). <sup>13</sup>C NMR (DMF-d7) δ(ppm): 167.0; 164.0; 150.7; 142.1; 133.9; 132.7; 132.1; 130.9; 128.1; 122.8; 121.6; 120.5; 120.3; 111.5; 107.8. <sup>15</sup>N NMR (DMF-d7) δ(ppm): 156.1; 135.7.

*(Z)-3-(2-(4-(4-nitrophenyl)thiazol-2-yl)hydrazono)indolin-2-one*  
**(EMAC2075)**. Yellow-orange solid, Yield 84%, mp: >250°. <sup>1</sup>H NMR (DMF-d7) δ(ppm): 13.55 (s; 1H); 11.42 (s; 1H); 8.38-8.32 (m; 2H); 8.30-8.24 (m; 2H); 8.12 (s; 1H); 7.62 (dm; *J* = 7.7 Hz; 1H); 7.40 (td; *J* = 7.7 Hz; 1.2 Hz; 1H); 7.15 (td; *J* = 7.7 Hz; 0.7 Hz; 1H); 7.08 (dm; *J* = 7.7 Hz; 1H). <sup>13</sup>C NMR (DMF-d7) δ(ppm): 167.3; 163.9; 149.7; 147.2; 142.2; 140.7; 133.1; 131.1; 127.0; 124.5; 122.9; 120.4; 120.3; 111.6; 111.6. <sup>15</sup>N NMR (DMF-d7) δ(ppm): 155.8; 135.7.

*(Z)-3-(2-(4-(4-biphenyl)thiazol-2-yl)hydrazono)indolin-2-one*  
**(EMAC2076)**. Yellow-orange solid, Yield 96%, mp: >250°. <sup>1</sup>H NMR (DMF-d7) δ(ppm): 13.56 (s; 1H); 11.41 (s; 1H); 8.13-8.08 (m; 2H); 7.85-7.75 (m; 5H); 7.63 (d; *J* = 7.6 Hz; 1H); 7.55-7.49 (m; 2H); 7.44-7.36 (m; 2H); 7.15 (tm; *J* = 7.6 Hz; 1H); 7.08 (dm; *J* = 7.6 Hz; 1H). <sup>13</sup>C NMR (DMF-d7) δ(ppm): 166.8; 164.0; 151.6; 142.1; 140.4; 140.3; 133.8; 132.6; 130.9; 129.4; 127.9; 127.4; 127.0; 126.7; 122.8; 120.5; 120.2; 111.5; 107.1. <sup>15</sup>N NMR (DMF-d7) δ(ppm): 156.5; 135.7

*(Z)-3-(2-(4-(4-bromophenyl)thiazol-2-yl)hydrazono)indolin-2-one*  
**(EMAC2077)**. Yellow-orange solid, Yield 99%, mp: >250°.

<sup>1</sup>H NMR (DMF-d7) δ(ppm): 13.54 (s; 1H); 11.42 (s; 1H); 8.23-8.17 (m; 2H); 8.04 (s; 1H); 7.97-7.93 (m; 2H); 7.62 (d; *J* = 7.6 Hz; 1H); 7.40 (tm; *J* = 7.6

Hz; 1H); 7.15 (tm; J: 7.6 Hz; 1H); 7.08 (d; J: 7.6 Hz; 1H). <sup>13</sup>C NMR (DMF) δ(ppm): 167.2; 163.9; 150.1; 142.2; 138.8; 133.2; 133.0; 131.0; 126.8; 122.9; 120.4; 120.3; 119.3; 111.6; 110.8; 110.6. <sup>15</sup>N NMR (DMF) δ(ppm): 155.9; 135.6.

*(Z)-3-(2-(4-(2,4-difluorophenyl)thiazol-2-yl)hydrazono)indolin-2-one* (**EMAC2078**). Yellow-orange solid, Yield 66%, mp: >250°. <sup>1</sup>H NMR (DMF-d7) δ(ppm): 13.52 (s; 1H); 11.39 (s; 1H); 8.23-8.14 (m; 1H); 7.62 (d; J= 7.6 Hz; 1H); 7.58(d; J= 2.4 Hz; 1H); 7.45-7.36 (m; 2H); 7.25 (td; J= 8.5 Hz; 2.5 Hz; 1H); 7.15 (td; J= 7.6 Hz; 0.7 Hz; 1H); 7.07 (dm; J= 7.6 Hz; 1H). <sup>13</sup>C NMR (DMF-d7) δ(ppm): 166.4; 164.0; 162.3 (dd; J= 248.0 Hz; 12.4 Hz); 160.4 (dd; J= 251.9 Hz; 12.3 Hz); 144.7 (dd; J= 2.6 Hz; 0.9 Hz); 142.1; 132.8; 131.2 (dd; J= 9.6Hz; 4.8 Hz); 131.0; 122.9; 120.5; 120.3; 119.1 (dd; J= 11.3Hz; 3.7 Hz); 112.3 (dd; J= 21.4Hz; 3.4 Hz); 111.5; 111.2 (d; J= 15.0 Hz); 104.9 (t; J= 26.5 Hz). <sup>15</sup>N NMR (DMF-d7) δ(ppm): 155.9; 135.5

*(Z)-3-(2-(4-(3-nitrophenyl)thiazol-2-yl)hydrazono)indolin-2-one* (**EMAC2079**). Yellow-orange solid, Yield 86%, mp: >250°. <sup>1</sup>H NMR (DMF-d7) δ(ppm): 13.57 (s; 1H); 11.41 (s; 1H); 8.84-8.76 (m; 1H); 8.46 (dm; J= 8.1 Hz; 1H); 8.26 (dm; J= 8.1 Hz; 1H); 8.09 (s; 1H); 7.80 (t; J= 8.1 Hz; 1H); 7.63 (d; J= 7.6 Hz; 1H); 7.41 (tm; J= 7.6 Hz; 1H); 7.15 (tm; J= 7.6 Hz; 1H); 7.08 (dm; J= 7.6 Hz; 1H). <sup>13</sup>C NMR (DMF-d7) δ(ppm): 167.3; 164.0; 149.5; 149.0; 142.2; 136.3; 133.0; 132.2; 131.0; 130.7; 122.9; 122.9; 120.6; 120.4; 120.3; 111.5; 109.6. <sup>15</sup>N NMR (DMF-d7) δ(ppm): 155.9; 135.5.

*(Z)-3-(2-(4-p-tolylthiazol-2-yl)hydrazono)indolin-2-one* (**EMAC2080**). Yellow-orange solid, Yield 72%, mp: >250°.

<sup>1</sup>H NMR (DMF-d7) δ(ppm): 13.53 (s; 1H); 11.38 (s; 1H); 7.91-7.86 (m; 2H); 7.66 (s; 1H); 7.62 (d; J= 7.6 Hz; 1H); 7.39 (td; J= 7.6 Hz; 1.2 Hz; 1H); 7.30-7.25 (m; 2H); 7.15 (td; J= 7.6 Hz; 0.8 Hz; 1H); 7.07 (dm; J= 7.6 Hz; 1H); 2.35 (s; 3H). <sup>13</sup>C NMR (DMF-d7) δ(ppm): 166.6; 164.0; 152.0; 142.0; 138.0; 132.4; 132.1; 130.8; 129.7; 126.1; 122.8; 120.5; 120.2; 111.5; 106.0; 20.8. <sup>15</sup>N NMR (DMF-d7) δ(ppm): 156.4; 135.5.

*(Z)-3-(2-(4-(4-methoxyphenyl)thiazol-2-yl)hydrazono)indolin-2-one* (**EMAC2081**). Yellow-orange solid, Yield 84%, mp: >250°. <sup>1</sup>H NMR (DMF-d7) δ(ppm): 13.53 (s; 1H); 11.40 (s; 1H); 7.95-7.91 (m; 2H); 7.63-7.59 (dm; J= 7.7 Hz; 1H); 7.56 (s; 1H); 7.39 (td; J= 7.7 Hz; 1.1 Hz; 1H); 7.15 (td; J= 7.7 Hz; 1.0 Hz; 1H); 7.08 (dm; J= 7.7 Hz; 1H); 7.06-7.02 (m, 2H); 3.85 (s; 3H). <sup>13</sup>C NMR



(DMF-d7)  $\delta$ (ppm): 166.6; 164.0; 159.9; 151.8; 142.0; 132.4; 130.8; 127.5; 127.5; 122.8; 120.5; 120.2; 114.3; 111.5; 104.7; 55.3.  $^{15}\text{N}$  NMR (DMF-d7)  $\delta$ (ppm): 156.6; 135.6.

*(Z)*-3-(2-(4-phenylthiazol-2-yl)hydrazono)indolin-2-one (**EMAC2082**).

Yellow-orange solid, Yield 92%, mp: >250°.

$^1\text{H}$  NMR (DMF-d7)  $\delta$ (ppm): 13.54 (s; 1H); 11.41 (s; 1H); 8.01-7.97 (m; 2H); 7.74 (s; 1H); 7.62 (dm;  $J$  = 7.7 Hz; 1H); 7.49-7.44 (m; 2H); 7.42-7.34 (m; 2H); 7.17-7.10 (m; 1H); 7.08 (dm;  $J$  = 7.7 Hz; 1H).  $^{13}\text{C}$  NMR (DMF-d7)  $\delta$ (ppm): 166.8; 164.0; 151.9; 142.1; 134.7; 132.6; 130.9; 129.1; 128.3; 126.1; 122.8; 120.5; 120.2; 111.5; 106.9.  $^{15}\text{N}$  NMR (DMF-d7)  $\delta$ (ppm): 156.2; 135.6.

*(Z)*-3-(2-(4-(2,4-dichlorophenyl)thiazol-2-yl)hydrazono)indolin-2-one

**(EMAC2083)**. Yellow-orange solid, Yield 90%, mp: >250°.  $^1\text{H}$  NMR (DMF-d7)  $\delta$ (ppm): 13.51 (s; 1H); 11.39 (s; 1H); 8.05 (d;  $J$  = 8.5 Hz; 1H); 7.85 (s; 1H); 7.75 (d;  $J$  = 2.2 Hz; 1H); 7.63 (dm;  $J$  = 7.6 Hz; 1H); 7.58 (dd;  $J$  = 8.5 Hz; 2.2 Hz; 1H); 7.40 (td;  $J$  = 7.6 Hz; 1.2 Hz; 1H); 7.15 (td;  $J$  = 7.6 Hz; 1.0 Hz; 1H); 7.07 (dm;  $J$  = 7.6 Hz; 1H).  $^{13}\text{C}$  NMR (DMF-d7)  $\delta$ (ppm): 166.1; 164.0; 147.1; 142.1; 133.6; 132.6; 132.9; 132.2; 132.1; 131.0; 130.4; 128.1; 122.9; 120.5; 120.3; 112.9; 111.5.  $^{15}\text{N}$  NMR (DMF-d7)  $\delta$ (ppm): 155.6; 135.5

## 2.5.2 Protein expression and purification.

**HIV-1 RT group M subtype B.** The recombinant HIV-1 RT protein, whose coding gene was subcloned in the p6HRT\_prot plasmid, was expressed in *E. coli* strain M15.<sup>69, 70</sup> The bacteria cells were grown up to an OD<sub>600</sub> of 0.8 and induced with 1.7 mM IPTG for 5 hrs. HIV-1 RT purification was carried out as described.<sup>71</sup> Briefly, cell pellets were resuspended in Lyses Buffer (20 mM Hepes pH 7.5, 0.5 M NaCl, 5 mM  $\beta$ -mercaptoethanol, 5 mM imidazole, 0.4 mg/mL lysozyme), incubated on ice for 20 min, sonicated and centrifuged at 30,000 x  $g$  for 1 hr. The supernatant was applied to a His-binding resin column and washed thoroughly with wash buffer (20 mM Hepes pH 7.5, 0.3 M NaCl, 5 mM  $\beta$ -mercaptoethanol, 60 mM imidazole, 10% glycerol). RT was eluted by imidazole gradient and the enzyme-containing fractions were pooled, dialyzed and aliquots were stored at -80 °C.

**Recombinant HIV-1 group O RT** was expressed and purified as previously described<sup>72, 73</sup>.

It was obtained as a heterodimer composed of subunits of 66 and 51 kDa, with a polyhistidine tag at the C-terminus of p66. For this purpose, the RT p66 subunit encoded within a p66RTB expression vector was co-expressed with the HIV-1 protease in *E. coli* XL-1 Blue, and the obtained heterodimers were then purified by ionic exchange followed by chromatography on Ni<sup>2+</sup>-nitriloacetic acid agarose. The enzyme was quantified by active site titration before biochemical studies.

**Site directed mutagenesis.** Aminoacid substitutions were introduced into the p66 HIV-1 RT subunit coded in a p6HRT-prot plasmid using the QuikChange mutagenesis kit by following the manufacturer's instructions (Agilent Technologies Inc., Santa Clara, CA).

### 2.5.3 Biochemical assays.

**RNase H polymerase-independent cleavage assay.** The HIV-1 RT associated RNase H activity was measured as described<sup>74</sup> in 100  $\mu$ L reaction volume containing 50 mM Tris HCl pH 7.8, 6 mM MgCl<sub>2</sub>, 1 mM dithiothreitol (DTT), 80 mM KCl, hybrid RNA/DNA (5'-GTTTTCTTTTCCCCCTGA C-3'-Fluorescein, 5'-CAAAGAAAAGG GGGGACUG-3'-Dabcyl) and increasing concentrations of inhibitors, whose dilution were made in water, and different amounts of enzymes according to a linear range of dose-response curve: 6 ng wt RT; 30 ng K103N RT; 12 ng V106ART; 19 ng V108A RT; 1.51.5 ng Y181C RT; 45 ng Y188A RT; 15 ng Y188L RT; 30 ng E224A RT; 15 ng P225A RT; 18 ng P226A RT; 23 ng F227A RT; 15 ng L228A RT; 30 ng W229A RT; 30 ng M230A RT; 15 ng G231A RT; 15 ng A502F RT; 19 ng A508V RT; 38 ng HIV-1 group O RT. After enzyme addition, the reaction mixture was incubated for 30 min at 37 °C and the stopped by addition of EDTA. Reaction products were detected by picogreen addition and measured with a multilabel counter plate reader Victor 3 (Perkin Elmer model 1420-051) equipped with filters for 502/523 nm (excitation/emission wavelength).

**DNA polymerase assay.** the HIV-1 RT associated (RDDP) activity was measured using Invitrogen EnzCheck Reverse Transcriptase Assay Kit, in 50

μL volume containing 60 mM Tris-HCl pH 8.1, 8 mM MgCl<sub>2</sub>, 60 mM KCl, 13 mM Dithiothreitol, 100 μM dTTP, and poly(A)-oligo(dT) and increasing concentrations of inhibitors, whose dilution were made in water, and different amount of enzyme according to a linear range of dose-response curve: 20 ng of WT RT; 60 ng K103N RT; 37.5 ng V106A RT; 75 ng V108A RT; 5 ng Y181C RT; 50 ng Y188A RT; 30 ng Y188L RT; 100 ng E224A RT; 37.5 ng P225A RT; 20 ng P226A RT; 18 ng F227A RT; 30 ng L228A RT; 30 ng W229A RT; 10 ng M230A RT; 30 ng G231A RT; 100 ng A502F RT; 37.5 ng A508V RT; 38 ng HIV-1 group O RT. The reaction mixture was incubated for 1 h at 37 °C, stopped by addition of EDTA and products were measured with a multilabel counter plate reader Victor 3 (Perkin Elmer model 1420-051) equipped with filters for 490/528 nm (excitation/emission wavelength).

**The Yonetani-Theorell analysis.** The Yonetani-Theorell analysis was performed according to the official protocol<sup>75</sup>

The calculation of the interaction constant  $\alpha$ , allows to estimate the degree of interference of the two inhibitors for the binding, was performed according to the following equation, assuming that  $K_{EFV} = IC_{50}$  for non-competitive inhibitors in accordance with Prusoff–Cheng equation<sup>76</sup> and equal to 12 nM in our system.

$$\frac{\text{slope with EFV}}{\text{slope without EFV}} = 1 + \frac{[\text{nM EFV}]}{\alpha K_{EFV}}$$

### **Detection of protein inhibitor interactions by Differential Scanning Fluorimetry.**

Thermal stability assays were performed according to Nettleship et al.<sup>77</sup>. In a LightCycler 480 96-well plate (Roche) we incubated 10 μM inhibitor, in 50 μl of reaction buffer containing 20 mM HEPES, pH 7.5, 10 mM MgCl<sub>2</sub>, 100 mM NaCl, 300 nM of HIV-1 RT and a 1:1000 dilution of Sypro Orange dye (Invitrogen).

The mixture was heated from 30 to 90 °C in increments of 0.2 °C. Fluorescence intensity was measured using excitation and emission wavelengths of 483 and 568 nm, respectively. Changes in protein thermal stability ( $\Delta T_m$ ) upon inhibitor binding were analyzed by using the LightCycler 480 software. All assays were performed in triplicate.

**Cell lines and virus.** The human embryonic c kidney cells 293T and the human T-lymphoid Jurkat cell line (clone E6-1) were from the American Type Culture Collection and maintained in DMEM or RPMI medium (Invitrogen), respectively, containing 10% fetal bovine serum (Invitrogen), at 37 °C in a humidified 5% CO<sub>2</sub> atmosphere. Recombinant viral stock was produced by transient transfection of 293T cells as previously described<sup>78</sup> and used to transduce Jurkat cells. In this context, an *env*-defective provirus encoding the bacterial chloramphenicol acetyltransferase (CAT) gene was complemented in *trans* by the envelope glycoprotein derived from the laboratory-adapted T-cell-tropic strain HXBc2. The level of CAT expression in the infected cells reflects the efficiency of a single round of the retroviral infection cycle.

**Cytotoxicity assay.** For cytotoxicity assays, cell lines were seeded in 96-well plates (Falcon) at an initial density of 10<sup>5</sup> cells/100 μL in medium containing 10% FBS, in the absence or presence of serial dilutions of test compounds. Plates were incubated for 72 hrs at 37 °C in a humidified 5% CO<sub>2</sub> atmosphere. Cell viability was determined by using Cell Proliferation Kit I (MTT) (Roche).

#### 2.5.4 Molecular modeling

**Ligand preparation.** Theoretical 3D models of the compounds were built by means of Maestro.<sup>79</sup> Starting conformations were optimized by means of an energy minimization carried out using the MMFFs force field,<sup>80</sup> the GB/SA<sup>81</sup> water implicit solvation model and the Polak-Ribier Coniugate Gradient (PRCG) method, converging on gradient with a threshold of 0.05 kJ(mol\*Å)<sup>-1</sup>.

**Protein preparation.** The coordinates for reverse transcriptase enzymes were taken from the RCSB Protein Data Bank<sup>43</sup> (PDB codes 1vrt<sup>48</sup>, 2zd1<sup>49</sup>, 1ep4<sup>50</sup>, 3qo9<sup>23</sup>, 1rti<sup>48</sup>, 1tv6<sup>47</sup>, 3lp2<sup>28</sup>). The proteins were prepared by using the Maestro Protein Preparation Wizard protocol. Original water molecules were removed and termini were capped. The bond orders and formal charges were added for hetero groups, and all the hydrogen atoms were added in the structure. Missing atoms and residues were included.

After preparation, the structures were refined in order to optimize the hydrogen bond network using OPLS\_2005<sup>82</sup> force field. The minimization was terminated when the energy converged or the RMSD reached a maximum cut-off of 0.30 Å.

**Docking protocol.** Molecular docking studies were performed using QMPL workflow protocol.<sup>45</sup> Grids were defined around the refined structure by centering on the residue indicated in the text (i.e W229 and Q500) and fixing the box volume at 97336 Å<sup>3</sup>. The extra precision (XP) docking algorithm was applied for scoring theoretical poses. The other settings were left as default. The same protocol was applied for all docking simulations.

**Post-docking protocol.** 10.000 steps of the Polak-Ribier conjugate gradient (PRCG) minimization method were conducted on the top ranked theoretical complexes using OPLS\_2005 force field. The optimization process was performed up to the derivative convergence criterion equal to 0.01 kJ/(mol\*Å)<sup>-1</sup>. The binding free energies ( $\Delta G(\text{Bind})$ ) were computed applying molecular mechanics and continuum solvation models with the molecular mechanics generalized Born/surface area (MM-GBSA) method.<sup>46</sup>

Furthermore for **EMAC2005** mechanism investigation, best docking complexes were submitted to 6 ns of MD by using Desmond ver. 2.4.<sup>52</sup> The complexes were solvated with a TIP3P (Transferable Intermolecular Potential 3-Point)<sup>83</sup> box of water and counter ions were added to neutralize the system net charge. The solvated models were optimised, and subsequently the MTK\_NPT (Martyna-Tobias-Klein with constant Number of particles, Pressure and Temperature) ensemble was employed.<sup>84</sup> The default stages in the relaxation process for the NPT ensemble include two energy minimizations and four simulation steps. During the energy minimizations, two runs of 2000 iteration were processed using the steepest descent method: during the first run, the protein structure was fixed by a force restraint constant of 50 kcal/(molÅ) and in the second all restraints were removed. With the first simulation, at NVT (constant Number of particles, Volume, and Temperature) ensemble, the system reached a temperature of 10 K. In the following three simulations in the NPT ensemble, the system was heated up to 300 K and the pressure was kept constant at 1 bar using the Berendsen thermostat-barostat. During the production phase, temperature and pressure were kept constant using the Nosè-Hoover

thermostat–barostat. The energy and trajectory were recorded every 1.2 ps and 4.8 ps, respectively. For multiple time step integration, RESPA (REversible reference System Propagator Algorithm)<sup>85</sup> was applied to integrate the equation of motion with Fourier-space electrostatics computed every 6 fs, and all remaining interactions computed every 2 fs. All chemical bond lengths involving hydrogen atoms were fixed with SHAKE.<sup>86</sup> Short range cut-off was set to 9 Å and the smooth particle mesh Ewald method (PME)<sup>87</sup> was used for long range electrostatic interaction. The resulting seven trajectories were analyzed in terms of interaction energies and geometries.

The same protocol was applied for the **EMAC2005-RT** ternary complex.

Molecular modeling figures were depicted by LigandScout<sup>88, 89</sup>, VMD ver. 1.8.7.<sup>90</sup> and Maestro.<sup>79</sup>

### 2.5.5 *In vitro* membrane permeation study

**HPLC determination of EMAC2005.** Sensitive HPLC method with UV detection was developed for the quantitative determination of EMAC2005. The chromatographic system was a HPLC Jasco mod. PU-1580 (Tokyo, Japan) with a 20 µl loop injection valve. The chromatographic system was equipped with a Jasco MD 1510 diode array detector, which was set at  $\lambda_{\text{max}}$  296 for EMAC2005. The separation was performed using a C18 reverse-phase Phenomenex column (Jupiter 250 × 4.60 mm, 5 µm particle size), which was maintained at room temperature. The mobile phase was pH 9 water (eluent A) and acetonitrile (eluent B) and it was delivered at a flow rate of 1 ml/min. Solvents were degassed by sonication for 15 min. A gradient elution method was applied for the determination of EMAC2005. The gradient was set as follow, eluent A/eluent B 70:30 0-7 min, linear increase of eluent B to 60% 7-9 min, linear increase of eluent B to 70 % 9-12 min, eluent A/eluent B 30:70 was maintained for 7 min (12-19 min), linear decrease of eluent B to 50 % 19-21 min and then the system was linearly returned to original conditions 21-25 min (see supplementary materials for further details).

HPLC data were processed using the Borwin chromatography

software (Version 1.5) from Jasco. A pure ethanol solutions of the compound EMAC2005 was prepared (1 mg/ml) and used as a stock solution for the calibration curve. EMAC2005 quantification was carried using a calibration curve in a linear concentration interval ranging from 0.1 to 100 µg/ml, according to the following equation:

$$\text{Eq. 1} \quad \text{AUC} = 0.5598x - 1.94010$$

where AUC is the area under the curve (mAu × min) and x is the drug concentration (µg/ml).

***In vitro* trans-membrane permeation.** A model of biological membrane was prepared as elsewhere reported.<sup>91</sup> Briefly, a polycarbonate membrane (50 nm pore size) was pre-soaked in pH 7.4 isotonic phosphate buffer for 3 h and layered on a synthetic cellulose nitrate membrane (molecular weight cut-off 10,000 Da), which was previously impregnated with a liquid paraffin-lauryl alcohol (2.1:10 w/w) mixture up to the doubling of the weight. Flow Franz diffusion cells were used for the trans-membrane permeation of EMAC2005 and they were characterized by a surface area of 0.75 cm<sup>2</sup> and a nominal receiving volume of 4.75 ml. The model of biological membrane was placed horizontally between the donor and receptor compartments. An ethanol/water mixture (50/50 v/v) was used to fill the receptor compartment. The same mixture (200 µl) was used to suspend the drug. This suspension was placed into the donor compartments. The receptor fluid was constantly stirred at 600 rpm during experiments by means of a magnetic anchor and warmed (GR 150 thermostat, Grant Instruments Ltd, Cambridge, UK) to 37 °C. These conditions were maintained throughout the experiments. At predetermined times, 400 µl of the receptor compartment were withdrawn using a Minipuls 3 peristaltic pump (Gilson Italia S.r.l., Cinisello Balsamo (MI), Italy) connected to a FC 204 fraction collector (Gilson Italia S.r.l., Cinisello Balsamo (MI), Italy) and immediately replaced with the same volume of fresh solution. The amount of EMAC2005, which permeated through the membranes, was immediately analyzed by HPLC. Experiments were carried out in triplicate and results were the mean of five different experiments ± standard deviations.

## 2.6 REFERENCES

1. [www.unaids.org/sites/default/files/media\\_asset/AIDS\\_by\\_the\\_numbers\\_2015\\_en.pdf](http://www.unaids.org/sites/default/files/media_asset/AIDS_by_the_numbers_2015_en.pdf).
2. Mehellou, Y.; De Clercq, E., Twenty-six years of anti-HIV drug discovery: where do we stand and where do we go? *J. Med. Chem.* **2010**, 53, 521-38.
3. Sarafianos, S. G.; Marchand, B.; Das, K.; Himmel, D.; Parniak, M. A.; Hughes, S. H.; Arnold, E., Structure and function of HIV-1 reverse transcriptase: molecular mechanisms of polymerization and inhibition. *J. Mol. Biol.* **2009**, 385, 693-713.
4. Sarafianos, S. G.; Das, K.; Tantillo, C.; Clark, A. D.; Ding, J.; Whitcomb, J. M.; Boyer, P. L.; Hughes, S. H.; Arnold, E., Crystal structure of HIV-1 reverse transcriptase in complex with a polypurine tract RNA:DNA. *The EMBO Journal* **2001**, 20, 1449-1461.
5. Cihlar, T.; Chen, M. S., Incorporation of selected nucleoside phosphonates and anti-human immunodeficiency virus nucleotide analogues into DNA by human DNA polymerases  $\alpha$ ,  $\beta$  and  $\gamma$ . *Antivir. Chem. Chemother.* **1997**, 8, 187-195.
6. Esposito, F.; Corona, A.; Tramontano, E., HIV-1 Reverse Transcriptase Still Remains a New Drug Target: Structure, Function, Classical Inhibitors, and New Inhibitors with Innovative Mechanisms of Actions. *Molecular Biology International* **2012**, 2012, 23.
7. Hachiya, A.; Kodama, E. N.; Sarafianos, S. G.; Schuckmann, M. M.; Sakagami, Y.; Matsuoka, M.; Takiguchi, M.; Gatanaga, H.; Oka, S., Amino acid mutation N348I in the connection subdomain of human immunodeficiency



virus type 1 reverse transcriptase confers multiclass resistance to nucleoside and nonnucleoside reverse transcriptase inhibitors. *J. Virol.* **2008**, 82, 3261-3270.

8. Brehm, J. H.; Koontz, D.; Meteer, J. D.; Pathak, V.; Sluis-Cremer, N.; Mellors, J. W., Selection of mutations in the connection and RNase H domains of human immunodeficiency virus type 1 reverse transcriptase that increase resistance to 3'-azido-2',3'-dideoxythymidine. *J. Virol.* **2007**, 81, 7852-7859.

9. Delviks-Frankenberry, K. A.; Nikolenko, G. N.; Barr, R.; Pathak, V. K., Mutations in human immunodeficiency virus type 1 RNase H primer grip enhance 3-Azido-3'-deoxythymidine resistance. *J. Virol.* **2007**, 81, 6837-6845.

10. Yap, S.-H.; Sheen, C.-W.; Fahey, J.; Zanin, M.; Tyssen, D.; Lima, V. D.; Wynhoven, B.; Kuiper, M.; Sluis-Cremer, N.; Harrigan, P. R.; Tachedjian, G., N348I in the connection domain of HIV-1 reverse transcriptase confers Zidovudine and Nevirapine resistance. *PLoS Med.* **2007**, 4, e335.

11. Nikolenko, G. N.; Delviks-Frankenberry, K. A.; Palmer, S.; Maldarelli, F.; Fivash, M. J., Jr.; Coffin, J. M.; Pathak, V. K., Mutations in the connection domain of HIV-1 reverse transcriptase increase 3'-azido-2',3'-deoxythymidine resistance. *Proc. Natl. Acad. Sci. U. S. A.* **2007**, 104, 317-322.

12. Balzarini, J.; Naesens, L.; Aquaro, S.; Knispel, T.; Perno, C. F.; De Clercq, E.; Meier, C., Intracellular metabolism of CycloSaligenyl 3'-azido-2',3'-dideoxythymidine monophosphate, a prodrug of 3'-azido-2',3'-dideoxythymidine (zidovudine). *Mol. Pharmacol.* **1999**, 56, 1354-1361.

13. Cihlar, T.; Ray, A. S., Nucleoside and nucleotide HIV reverse transcriptase inhibitors: 25 years after zidovudine. *Antivir. Res.* **2010**, 85, 39-58.

14. Jochmans, D.; Deval, J.; Kesteleyn, B.; Van Marck, H.; Bettens, E.; De Baere, I.; Dehertogh, P.; Ivens, T.; Van Ginderen, M.; Van Schoubroeck, B.; Ehteshami, M.; Wigerinck, P.; Götte, M.; Hertogs, K., Indolopyridones Inhibit Human Immunodeficiency Virus Reverse Transcriptase with a Novel Mechanism of Action. *J. Virol.* **2006**, *80*, 12283-12292.
15. Ehteshami, M.; Scarth, B. J.; Tchesnokov, E. P.; Dash, C.; Le Grice, S. F. J.; Hallenberger, S.; Jochmans, D.; Goette, M., Mutations M184V and Y115F in HIV-1 Reverse Transcriptase Discriminate against "Nucleotide-competing Reverse Transcriptase Inhibitors". *J. Biol. Chem.* **2008**, *283*, 29904-29911.
16. Freisz, S.; Bec, G.; Radi, M.; Wolff, P.; Crespan, E.; Angeli, L.; Dumas, P.; Maga, G.; Botta, M.; Ennifar, E., Crystal Structure of HIV-1 Reverse Transcriptase Bound to a Non-Nucleoside Inhibitor with a Novel Mechanism of Action. *Angew. Chem., Int. Ed.* **2010**, *49*, 1805-1808.
17. Sluis-Cremer, N.; Tachedjian, G., Mechanisms of inhibition of HIV replication by non-nucleoside reverse transcriptase inhibitors. *Virus Res.* **2008**, *134*, 147-156.
18. Zhan, P.; Chen, X.; Li, D.; Fang, Z.; De Clercq, E.; Liu, X., HIV-1 NNRTIs: Structural diversity, pharmacophore similarity, and implications for drug design. *Med. Res. Rev.* **2011**, DOI: 10.1002/med.20241.
19. Das, K.; Lewi, P. J.; Hughes, S. H.; Arnold, E., Crystallography and the design of anti-AIDS drugs: conformational flexibility and positional adaptability are important in the design of non-nucleoside HIV-1 reverse transcriptase inhibitors. *Prog. Biophys. Mol. Biol.* **2005**, *88*, 209-231.
20. Zhan, P.; Liu, X.; Li, Z.; Pannecouque, C.; De Clercq, E., Design Strategies of Novel NNRTIs to Overcome Drug Resistance. *Curr. Med. Chem.* **2009**, *16*, 3903-3917.

21. Bunge, K. E.; Dezzutti, C. S.; Rohan, L. C.; Hendrix, C. W.; Marzinke, M. A.; Richardson-Harman, N.; Moncla, B. J.; Devlin, B.; Meyn, L. A.; Spiegel, H. M.; Hillier, S. L., A Phase 1 trial to assess the safety, acceptability, pharmacokinetics and pharmacodynamics of a novel dapivirine vaginal film. *J. Acquir. Immune Defic. Syndr.* **2015**.
22. Ding, J. P.; Das, K.; Moereels, H.; Koymans, L.; Andries, K.; Janssen, P. A. J.; Hughes, S. H.; Arnold, E., Structure of HIV-1 RT/TIBO R 86183 complex reveals similarity in the binding of diverse nonnucleoside inhibitors. *Nat. Struct. Biol.* **1995**, *2*, 407-415.
23. Das, K.; Bauman, J. D.; Rim, A. S.; Dharia, C.; Clark, A. D.; Camarasa, M. a.-J.; Balzarini, J.; Arnold, E., Crystal Structure of tert-Butyldimethylsilyl-spiroaminoxathioledioxide-thymine (TSAO-T) in Complex with HIV-1 Reverse Transcriptase (RT) Redefines the Elastic Limits of the Non-nucleoside Inhibitor-Binding Pocket. *J. Med. Chem.* **2011**, *54*, 2727-2737.
24. Klumpp, K.; Mirzadegan, T., Recent progress in the design of small molecule inhibitors of HIV RNase H. *Curr. Pharm. Des.* **2006**, *12*, 1909-22.
25. Schatz, O.; Cromme, F. V.; Gruninger-Leitch, F.; Le Grice, S. F., Point mutations in conserved amino acid residues within the C-terminal domain of HIV-1 reverse transcriptase specifically repress RNase H function. *FEBS Lett.* **1989**, *257*, 311-4.
26. Wendeler, M.; Lee, H. F.; Bermingham, A.; Miller, J. T.; Chertov, O.; Bona, M. K.; Baichoo, N. S.; Ehteshami, M.; Beutler, J.; O'Keefe, B. R.; Gotte, M.; Kvaratskhelia, M.; Le Grice, S., Vinylogous Ureas as a Novel Class of Inhibitors of Reverse Transcriptase-Associated Ribonuclease H Activity. *ACS Chem. Biol.* **2008**, *3*, 635-644.
27. Himmel, D. M.; Sarafianos, S. G.; Dharmasena, S.; Hossain, M. M.; McCoy-Simandle, K.; Ilina, T.; Clark, A. D., Jr.; Knight, J. L.; Julias, J. G.; Clark,

- P. K.; Krogh-Jespersen, K.; Levy, R. M.; Hughes, S. H.; Parniak, M. A.; Arnold, E., HIV-1 reverse transcriptase structure with RNase H inhibitor dihydroxy benzoyl naphthyl hydrazone bound at a novel site. *ACS Chem. Biol.* **2006**, *1*, 702-12.
28. Himmel, D. M.; Maegley, K. A.; Pauly, T. A.; Bauman, J. D.; Das, K.; Dharia, C.; Clark, A. D., Jr.; Ryan, K.; Hickey, M. J.; Love, R. A.; Hughes, S. H.; Bergqvist, S.; Arnold, E., Structure of HIV-1 reverse transcriptase with the inhibitor beta-Thujaplicinol bound at the RNase H active site. *Structure* **2009**, *17*, 1625-35.
29. Kharlamova, T.; Esposito, F.; Zinzula, L.; Floris, G.; Cheng, Y.-C.; Ginger, E. D.; Tramontano, E., Inhibition of HIV-1 Ribonuclease H Activity by Novel Frangula-Emodine Derivatives. *Med. Chem.* **2009**, *5*, 398-410.
30. Distinto, S.; Esposito, F.; Kirchmair, J.; Cardia, M. C.; Gaspari, M.; Maccioni, E.; Alcaro, S.; Markt, P.; Wolber, G.; Zinzula, L.; Tramontano, E., Identification of HIV-1 reverse transcriptase dual inhibitors by a combined shape-, 2D-fingerprint- and pharmacophore-based virtual screening approach. *Eur. J. Med. Chem.* **2012**, *50*, 216-29.
31. OpenEye *ROCS*, 3.0; Santa Fe, NM, 2009.
32. Nicholls, A.; McGaughey, G. B.; Sheridan, R. P.; Good, A. C.; Warren, G.; Mathieu, M.; Muchmore, S. W.; Brown, S. P.; Grant, J. A.; Haigh, J. A.; Nevins, N.; Jain, A. N.; Kelley, B., Molecular Shape and Medicinal Chemistry: A Perspective. *J. Med. Chem.* **2010**, *53*, 3862-3886.
33. NCIDatabase, In <http://cactus.nci.nih.gov/ncidb2/download.html>.
34. Irwin, J. J.; Shoichet, B. K., ZINC – A Free Database of Commercially Available Compounds for Virtual Screening. *J. Chem. Inf. Model.* **2004**, *45*, 177-182.

35. Krüger, M. D.; Evers, A., Comparison of Structure- and Ligand-Based Virtual Screening Protocols Considering Hit List Complementarity and Enrichment Factors. *ChemMedChem* **2009**, 5, 148-158.
36. Meleddu, R.; Cannas, V.; Distinto, S.; Sarais, G.; Del Vecchio, C.; Esposito, F.; Bianco, G.; Corona, A.; Cottiglia, F.; Alcaro, S.; Parolin, C.; Artese, A.; Scalise, D.; Fresta, M.; Arridu, A.; Ortuso, F.; Maccioni, E.; Tramontano, E., Design, synthesis, and biological evaluation of 1,3-diarylpropenones as dual inhibitors of HIV-1 reverse transcriptase. *ChemMedChem* **2014**, 9, 1869-79.
37. Rosen, T.; Nagel, A. A.; Rizzi, J. P.; Ives, J. L.; Daffeh, J. B.; Ganong, A. H.; Guarino, K.; Heym, J.; McLean, S.; Nowakowski, J. T.; et al., Thiazole as a carbonyl bioisostere. A novel class of highly potent and selective 5-HT<sub>3</sub> receptor antagonists. *J. Med. Chem.* **1990**, 33, 2715-20.
38. Tramontano, E.; Esposito, F.; Badas, R.; Di Santo, R.; Costi, R.; La Colla, P., 6-[1-(4-Fluorophenyl)methyl-1H-pyrrol-2-yl]-2,4-dioxo-5-hexenoic acid ethyl ester a novel diketo acid derivative which selectively inhibits the HIV-1 viral replication in cell culture and the ribonuclease H activity in vitro. *Antiviral Res.* **2005**, 65, 117-24.
39. Tramontano, E.; Esposito, F.; Badas, R.; Di Santo, R.; Costi, R.; La Colla, P., 6-[1-(4-Fluorophenyl)methyl-1H-pyrrol-2-yl]-2,4-dioxo-5-hexenoic acid ethyl ester a novel diketo acid derivative which selectively inhibits the HIV-1 viral replication in cell culture and the ribonuclease H activity in vitro. *Antivir. Res.* **2005**, 65, 117-124.
40. Yonetani, T., The Yonetani-Theorell graphical method for examining overlapping subsites of enzyme active centers. *Methods Enzymol.* **1982**, 87, 500-9.

41. Esposito, F.; Kharlamova, T.; Distinto, S.; Zinzula, L.; Cheng, Y.-C.; Dutschman, G.; Floris, G.; Markt, P.; Corona, A.; Tramontano, E., Alizarine derivatives as new dual inhibitors of the HIV-1 reverse transcriptase-associated DNA polymerase and RNase H activities effective also on the RNase H activity of non-nucleoside resistant reverse transcriptases. *FEBS J.* **2011**, *278*, 1444-1457.
42. Paris, K. A.; Haq, O.; Felts, A. K.; Das, K.; Arnold, E.; Levy, R. M., Conformational Landscape of the Human Immunodeficiency Virus Type 1 Reverse Transcriptase Non-Nucleoside Inhibitor Binding Pocket: Lessons for Inhibitor Design from a Cluster Analysis of Many Crystal Structures. *J. Med. Chem.* **2009**, *52*, 6413-6420.
43. Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E., The Protein Data Bank. *Nucleic Acids Res.* **2000**, *28*, 235-242.
44. Huang, S.-Y.; Zou, X., Ensemble docking of multiple protein structures: Considering protein structural variations in molecular docking. *Proteins: Struct., Funct., Bioinf.* **2007**, *66*, 399-421.
45. Schrödinger LLC. *QMPolarized protocol*, 2012; New York, NY, USA.
46. Kollman, P. A.; Massova, I.; Reyes, C.; Kuhn, B.; Huo, S.; Chong, L.; Lee, M.; Lee, T.; Duan, Y.; Wang, W.; Donini, O.; Cieplak, P.; Srinivasan, J.; Case, D. A.; Cheatham, T. E., Calculating Structures and Free Energies of Complex Molecules: Combining Molecular Mechanics and Continuum Models. *Acc. Chem. Res.* **2000**, *33*, 889-897.
47. Pata, J. D.; Stirtan, W. G.; Goldstein, S. W.; Steitz, T. A., Structure of HIV-1 reverse transcriptase bound to an inhibitor active against mutant reverse transcriptases resistant to other nonnucleoside inhibitors. *Proc. Natl. Acad. Sci. U. S. A.* **2004**, *101*, 10548-53.

48. Ren, J. S.; Esnouf, R.; Garman, E.; Somers, D.; Ross, C.; Kirby, I.; Keeling, J.; Darby, G.; Jones, Y.; Stuart, D.; Stammers, D., High-resolution structures of HIV-1 rt from 4 RT-inhibitor complexes. *Nat. Struct. Biol.* **1995**, *2*, 293-302.
49. Das, K.; Bauman, J. D.; Clark, A. D.; Frenkel, Y. V.; Lewi, P. J.; Shatkin, A. J.; Hughes, S. H.; Arnold, E., High-resolution structures of HIV-1 reverse transcriptase/TMC278 complexes: Strategic flexibility explains potency against resistance mutations. *Proc. Natl. Acad. Sci. U. S. A.* **2008**, *105*, 1466-1471.
50. Ren, J. S.; Nichols, C.; Bird, L. E.; Fujiwara, T.; Sugimoto, H.; Stuart, D. I.; Stammers, D. K., Binding of the second generation non-nucleoside inhibitor S-1153 to HIV-1 reverse transcriptase involves extensive main chain hydrogen bonding. *J. Biol. Chem.* **2000**, *275*, 14316-14320.
51. Felts, A. K.; Labarge, K.; Bauman, J. D.; Patel, D. V.; Himmel, D. M.; Arnold, E.; Parniak, M. A.; Levy, R. M., Identification of alternative binding sites for inhibitors of HIV-1 ribonuclease H through comparative analysis of virtual enrichment studies. *J. Chem. Inf. Model.* **2011**, *51*, 1986-98.
52. Bowers, K. J.; Dror, R. O.; Shaw, D. E., The midpoint method for parallelization of particle simulations. *The Journal of Chemical Physics* **2006**, *124*, 184109-11.
53. Gong, Q.; Menon, L.; Ilina, T.; Miller, L. G.; Ahn, J.; Parniak, M. A.; Ishima, R., Interaction of HIV-1 reverse transcriptase ribonuclease H with an acylhydrazone inhibitor. *Chem. Biol. Drug Des.* **2011**, *77*, 39-47.
54. Seckler, J. M.; Barkley, M. D.; Wintrode, P. L., Allosteric Suppression of HIV-1 Reverse Transcriptase Structural Dynamics upon Inhibitor Binding. *Biophys. J.* **2011**, *100*, 144-153.

55. Iliina, T.; LaBarge, K.; Sarafianos, S. G.; Ishima, R.; Parniak, M. A., Inhibitors of HIV-1 Reverse Transcriptase—Associated Ribonuclease H Activity. *Biology* **2012**, *1*, 521-541.
56. Meleddu, R.; Distinto, S.; Corona, A.; Bianco, G.; Cannas, V.; Esposito, F.; Artese, A.; Alcaro, S.; Matyus, P.; Bogdan, D.; Cottiglia, F.; Tramontano, E.; Maccioni, E., (3Z)-3-(2-[4-(aryl)-1,3-thiazol-2-yl]hydrazin-1-ylidene)-2,3-dihydro-1H-indol-2-one derivatives as dual inhibitors of HIV-1 reverse transcriptase. *Eur. J. Med. Chem.* **2015**, *93*, 452-60.
57. Masaoka, T.; Chung, S.; Caboni, P.; Rausch, J. W.; Wilson, J. A.; Taskent-Sezgin, H.; Beutler, J. A.; Tocco, G.; Le Grice, S. F. J., Exploiting Drug-Resistant Enzymes as Tools to Identify Thienopyrimidinone Inhibitors of Human Immunodeficiency Virus Reverse Transcriptase-Associated Ribonuclease H. *J. Med. Chem.* **2013**, *56*.
58. Cummings, M. D.; Farnum, M. A.; Nelen, M. I., Universal screening methods and applications of ThermoFluor. *J. Biomol. Screen.* **2006**, *11*, 854-63.
59. Chung, S.; Miller, J. T.; Johnson, B. C.; Hughes, S. H.; Le Grice, S. F., Mutagenesis of human immunodeficiency virus reverse transcriptase p51 subunit defines residues contributing to vinyllogous urea inhibition of ribonuclease H activity. *J. Biol. Chem.* **2012**, *287*, 4066-75.
60. Su, H. P.; Yan, Y.; Prasad, G. S.; Smith, R. F.; Daniels, C. L.; Abeywickrema, P. D.; Reid, J. C.; Loughran, H. M.; Kornienko, M.; Sharma, S.; Grobler, J. A.; Xu, B.; Sardana, V.; Allison, T. J.; Williams, P. D.; Darke, P. L.; Hazuda, D. J.; Munshi, S., Structural basis for the inhibition of RNase H activity of HIV-1 reverse transcriptase by RNase H active site-directed inhibitors. *J. Virol.* **2010**, *84*, 7625-33.



61. Quinones-Mateu, M. E.; Soriano, V.; Domingo, E.; Menendez-Arias, L., Characterization of the reverse transcriptase of a human immunodeficiency virus type 1 group O isolate. *Virology* **1997**, *236*, 364-73.
62. Budihas, S. R.; Gorshkova, I.; Gaidamakov, S.; Wamiru, A.; Bona, M. K.; Parniak, M. A.; Crouch, R. J.; McMahon, J. B.; Beutler, J. A.; Le Grice, S. F., Selective inhibition of HIV-1 reverse transcriptase-associated ribonuclease H activity by hydroxylated tropolones. *Nucleic Acids Res.* **2005**, *33*, 1249-56.
63. Sievers, F.; Wilm, A.; Dineen, D.; Gibson, T. J.; Karplus, K.; Li, W.; Lopez, R.; McWilliam, H.; Remmert, M.; Söding, J.; Thompson, J. D.; Higgins, D. G., Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.* **2011**, *7*, 539-539.
64. Cho, A. E.; Guallar, V.; Berne, B. J.; Friesner, R., Importance of accurate charges in molecular docking: quantum mechanical/molecular mechanical (QM/MM) approach. *J. Comput. Chem.* **2005**, *26*, 915-31.
65. Herman, B. D.; Sluis-Cremer, N., Transient kinetic analyses of the ribonuclease H cleavage activity of HIV-1 reverse transcriptase in complex with efavirenz and/or a beta-thujaplicinol analogue. *Biochem. J.* **2013**, *455*, 179-84.
66. Christen, M. T.; Menon, L.; Myshakina, N. S.; Ahn, J.; Parniak, M. A.; Ishima, R., Structural basis of the allosteric inhibitor interaction on the HIV-1 reverse transcriptase RNase H domain. *Chem. Biol. Drug Des.* **2012**, *80*, 706-16.
67. Julias, J. G.; McWilliams, M. J.; Sarafianos, S. G.; Arnold, E.; Hughes, S. H., Mutations in the RNase H domain of HIV-1 reverse transcriptase affect the initiation of DNA synthesis and the specificity of RNase H cleavage in vivo. *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99*, 9515-20.

68. Rausch, J. W.; Lener, D.; Miller, J. T.; Julias, J. G.; Hughes, S. H.; Le Grice, S. F., Altering the RNase H primer grip of human immunodeficiency virus reverse transcriptase modifies cleavage specificity. *Biochemistry* **2002**, *41*, 4856-65.
69. Tramontano, E.; Cheng, Y. C., HIV-1 reverse-transcriptase inhibition by a dipyrindodiazepinone derivative: BI-RG-587. *Biochem. Pharmacol.* **1992**, *43*, 1371-1376.
70. Mellors, J. W.; Im, G. J.; Tramontano, E.; Winkler, S. R.; Medina, D. J.; Dutschman, G. E.; Bazmi, H. Z.; Piras, G.; Gonzalez, C. J.; Cheng, Y. C., A single conservative amino acid substitution in the reverse transcriptase of human immunodeficiency virus-1 confers resistance to (+)-(5S)-4,5,6,7-tetrahydro-5-methyl-6-(3-methyl-2-butenyl)imidazo[4,5, 1-jk][1,4]benzodiazepin-2(1H)-thione (TIBO R82150). *Mol. Pharmacol.* **1993**, *43*, 11-16.
71. Suchaud, V.; Bailly, F.; Lion, C.; Tramontano, E.; Esposito, F.; Corona, A.; Christ, F.; Debyser, Z.; Cotelle, P., Development of a series of 3-hydroxyquinolin-2(1H)-ones as selective inhibitors of HIV-1 reverse transcriptase associated RNase H activity. *Bioorg. Med. Chem. Lett.* **2012**, *22*, 3988-3992.
72. Álvarez, M.; Barrioluengo, V.; Afonso-Lehmann, R. N.; Menéndez-Arias, L., Altered error specificity of RNase H-deficient HIV-1 reverse transcriptases during DNA-dependent DNA synthesis. *Nucleic Acids Res.* **2013**, *41*, 4601-12.
73. Alvarez, M.; Matamoros, T.; Menéndez-Arias, L., Increased thermostability and fidelity of DNA synthesis of wild-type and mutant HIV-1 group O reverse transcriptases. *J. Mol. Biol.* **2009**, *392*, 872-84.
74. Distinto, S.; Esposito, F.; Kirchmair, J.; Cardia, M. C.; Gaspari, M.; Maccioni, E.; Alcaro, S.; Markt, P.; Wolber, G.; Zinzula, L.; Tramontano, E.,

Identification of HIV-1 reverse transcriptase dual inhibitors by a combined shape-, 2D-fingerprint- and pharmacophore-based virtual screening approach. *European Journal of Medicinal Chemistry* **2012**, 50, 216-229.

75. Yonetani, T., The Yonetani-Theorell graphical method for examining overlapping subsites of enzyme active centers. *Methods in enzymology* **1982**, 87, 500-9.

76. Cheng, Y.; Prusoff, W. H., Relationship between the inhibition constant ( $K_1$ ) and the concentration of inhibitor which causes 50 per cent inhibition ( $I_{50}$ ) of an enzymatic reaction. *Biochem. Pharmacol.* **1973**, 22, 3099-108.

77. Nettleship, J. E.; Brown, J.; Groves, M. R.; Geerlof, A., Methods for protein characterization by mass spectrometry, thermal shift (ThermoFluor) assay, and multiangle or static light scattering. *Methods Mol. Biol.* **2008**, 426, 299-318.

78. Parolin, C.; Gatto, B.; Del Vecchio, C.; Pecere, T.; Tramontano, E.; Cecchetti, V.; Fravolini, A.; Masiero, S.; Palumbo, M.; Palù, G., New Anti-Human Immunodeficiency Virus Type 1 6-Aminoquinolones: Mechanism of Action. *Antimicrob. Agents Chemother.* **2003**, 47, 889-896.

79. Schrödinger LLC. *Maestro GUI*, New York, NY, USA, 2013.

80. Halgren, T., Merck molecular force field. II. MMFF94 van der Waals and electrostatic parameters for intermolecular interactions. *J. Comput. Chem.* **1996**, 17, 520-552.

81. Hasel, W.; Hendrickson, T. F.; Still, W. C., A rapid approximation to the solvent accessible surface areas of atoms. *Tetrahedron Computer Methodology* **1988**, 1, 103-116.

82. Kaminski, G. A.; Friesner, R. A.; Tirado-Rives, J.; Jorgensen, W. L., Evaluation and Reparametrization of the OPLS-AA Force Field for Proteins

- via Comparison with Accurate Quantum Chemical Calculations on Peptidest. *The Journal of Physical Chemistry B* **2001**, 105, 6474-6487.
83. Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L., Comparison of simple potential functions for simulating liquid water. *The Journal of Chemical Physics* **1983**, 79, 926-935.
84. Martyna, G. J.; Tobias, D. J.; Klein, M. L., Constant pressure molecular dynamics algorithms. *The Journal of Chemical Physics* **1994**, 101, 4177-4189.
85. Gibson, D. A.; Carter, E. A., Time-reversible multiple time-scale ab-initio molecular-dynamics. *J. Phys. Chem.* **1993**, 97, 13429-13434.
86. Ryckaert, J.-P.; Ciccotti, G.; Berendsen, H., Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *J. Comput. Phys.* **1977**, 23, 327-341.
87. Darden, T.; York, D.; Pedersen, L., Particle mesh Ewald: An N [center-dot] log(N) method for Ewald sums in large systems. *The Journal of Chemical Physics* **1993**, 98, 10089-10092.
88. InteLigand Software GmbH *LigandScout 3.0*, Maria Enzersdorf, Austria.
89. Wolber, G.; Langer, T., LigandScout: 3-D pharmacophores derived from protein-bound ligands and their use as virtual screening filters. *J. Chem. Inf. Model.* **2004**, 45, 160-169.
90. Humphrey, W.; Dalke, A.; Schulten, K., VMD: Visual molecular dynamics. *J. Mol. Graph.* **1996**, 14, 33-38.
91. Cavallaro, G.; Fresta, M.; Giammona, G.; Puglisi, G.; Villari, A., Entrapment of  $\beta$ -lactams antibiotics in polyethylcyanoacrylate nanoparticles: Studies on the possible in vivo application of this colloidal delivery system. *Int. J. Pharm.* **1994**, 111, 31-41.



### 3

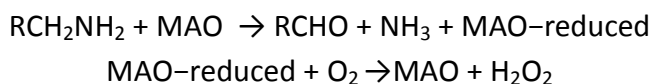
## CHAPTER

### 3.1 INTRODUCTION TO MONOAMINO OXIDASE B

In this third chapter I will discuss about the study conducted on a class of new oxadiazoles synthesised and tested on Monoamine Oxidase enzymes (MAO). But firstly, a brief introduction of these target is necessary, in order to understand their role in the human organism.

MAO are ubiquitous enzymes that contain a flavin adenine dinucleotide cofactor (FAD) and play a key role in the degradation of exogenous and endogenous amines.

MAO catalyzes the oxidative deamination with this simple reaction:



The mammalian family of MAOs consists of two isozymes, namely MAO-A and MAO-B, differing in their selectivity versus substrates and inhibitors, serotonin (5HT) and norepinephrine (NE) are preferentially deaminated by the A isoform, whereas 2-phenylethylamine and benzylamine are MAO-B substrates. Both isoforms act on dopamine (DA) and tryptamine.<sup>1</sup> These enzymes are tightly bound to the mitochondrial outer membrane. Although MAOs are widely distributed in various organs, most of the studies concerning their functional properties and involvement in pathological processes have been mainly focused on the central nervous system. In the periphery, MAO-A and MAO-B are differently expressed in a variety of tissues: MAO-A is predominant in heart, adipose tissue, and skin fibroblasts, MAO-B is the major form found in platelets and lymphocytes, whereas both isoenzymes are expressed in kidney and liver.<sup>2,3</sup>

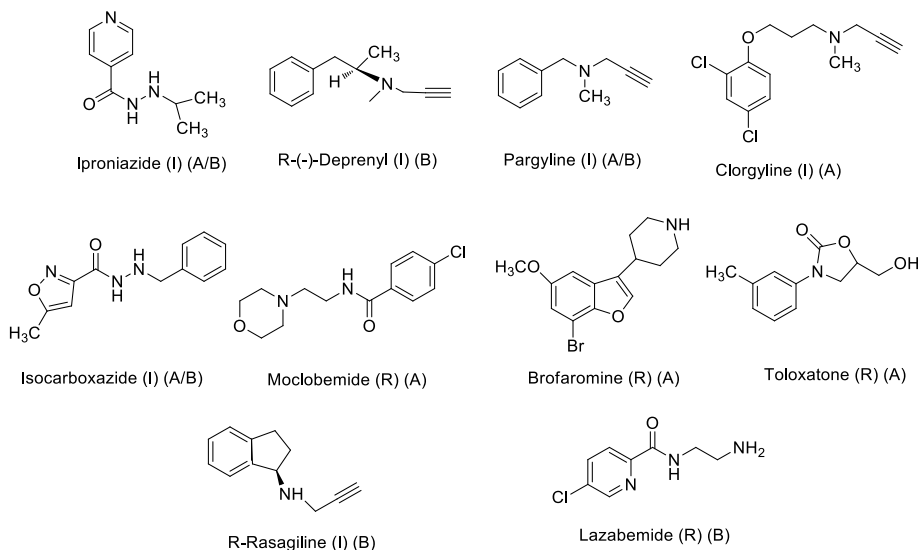
From an historical point of view, the application of MAO inhibitors in therapy showed a fluctuating trend: After the first enthusiastic interest on their use as antidepressants, the incidence of side effects and hypertensive crisis, due to the interference with metabolism of dietary amines (known as *cheese effect*), slowly let these drugs disappear from clinical practice. These severe side effects are due to the irreversible and unselective inhibition of MAO first generation drugs. Then the research on this target has led to

focusing efforts on finding first, selective inhibitors (second generation), and successively selective and reversible compounds for the two enzyme isoforms (third generation).<sup>4-6</sup>

Reversible inhibitors of monoamine oxidase A (RIMA) led to reexamination of MAOIs in psychiatry and they represent the first choice drugs in the treatment of atypical depression, anxiety and panic disorders. In fact, such inhibitors can be displaced by high concentration of xenobiotic amines allowing their deamination by the enzyme.

Instead selective irreversible human MAO-B inhibitors (hMAO-BIs) are used alone or in combination with other drugs, in the treatment of both Parkinson's (PD)<sup>7, 8</sup> and Alzheimer's (AD) diseases.<sup>9-12</sup>

In fact, increased levels of MAO-B were observed in both PD patients and elderly population<sup>13, 14</sup> which lead to an increased production of hydrogen peroxide and other reactive oxygen species. These substances are responsible for neuron degeneration in the central nervous system (CNS) and in particular in the "*substantia nigra*".<sup>15-18</sup> MAO-A expression, differently from MAO-B, does not increase with age, suggesting that a totally independent mechanism regulates the expression of the two enzymatic isoforms.<sup>19, 20</sup>



**Figure 1.** Irreversible (I), reversible (R), and/or selective hMAO-A and hMAO-B inhibitors.

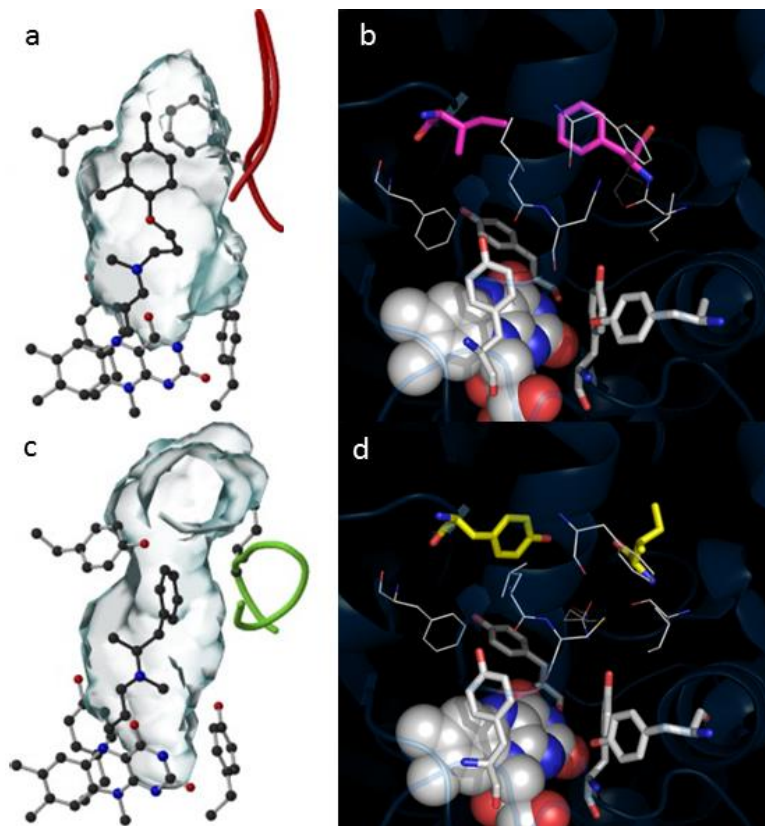
An increase of the DA levels as well as a neuroprotective effect can be observed following the inhibition of MAO-B.<sup>2, 3, 6, 21, 22</sup>

Although the two isoforms share a high sequence similarity, they differ in shape and volume of the catalytic site allowing to discover selective binders. In fact, structural studies reported the presence of a single cavity of about 550 Å<sup>3</sup> for MAO-A (Figure 2a,b), whereas MAO-B has a tight and longer dipartite cleft, named entrance and substrate cavities. These two pockets can merge into a single cavity of ~700 Å<sup>3</sup> (Figure 2c,d).<sup>23</sup>

The availability of crystal structures of MAO played a central role in galvanizing researchers to synthesise new inhibitors. In the last decade several crystallographic models have been published in the Protein Data Bank<sup>24</sup> (Table 1). Apart from revealing the differences in volume and shape of the binding pocket of the two isoforms, crystallographic and mutagenesis studies showed which residues are pivotal in catalysis and inhibition. All structural studies converge on considering the residues Ile199 and Tyr326, which separate the “access” from “substrate” cavity of MAO-B, as determinants of enzyme specificity and selectivity.<sup>25,26,27</sup> In some complexes with reversible MAO-BIs it has been highlighted that the residue Ile199 is in “open” conformation behaving as “gate” between the two cavities.<sup>28,29</sup>

The access of substrates or inhibitors is modulated also by the loop 99-112.<sup>30</sup> Recent mutagenesis studies followed by crystallographic studies corroborated the fundamental role of the “aromatic cage” on the deamination mechanism.<sup>31</sup> This explains why the aromatic pair of tyrosine involved in the “cage” is conserved in homologues and related enzymes, such as polyamine oxidase (PAO). The achievement of high-resolution crystal structures of several protein-ligand complexes made structure-based computational studies possible.





**Figure 2.** Main differences between the a) MAO-A and b) MAO-B active site.

The structural information helped to understand the reason of selectivity of potent cavity-spanning MAO-BIs such as safinamide (Figure 3),<sup>32, 33</sup> coumarin derivatives,<sup>29</sup> farnesol,<sup>34</sup> caffeine derivatives<sup>35, 36</sup> and the anti-diabetes drug pioglitazone.<sup>37</sup>

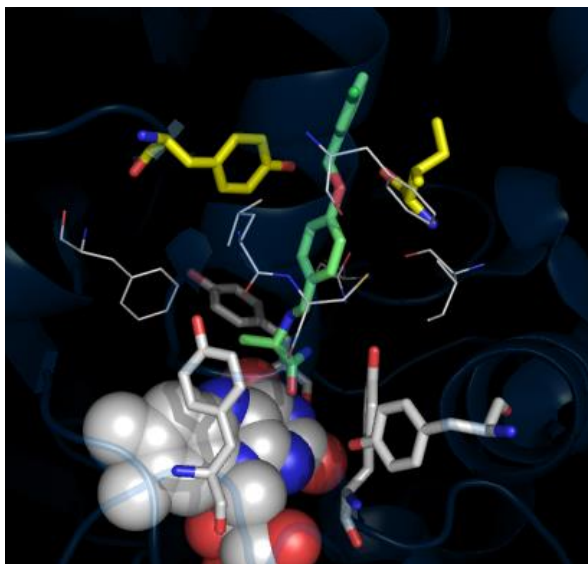


Figure 3. Saffinamide-MAO-B complex.

Table 1. PDB models MAO B and MAO A.

Isoform	PDB	Res.	Mutation	Ligand	Binding	Year	Ref.
hMAO-B	1GOS	3		pargyline	Cov	2002	38
	1OJ9	2.3		1,4-diphenyl-2-butene	Non Cov	2003	39
	1OJA	1.7		isatin	Non Cov		
	2XFU	2.2		tranylcypromine	Cov		
	1OJC	2.4		N-(2-aminoethyl)-p-chlorobenzamide	Cov		
	1OJD	3.1		lauryl-N,N-dimethylamine-N-oxide ( <i>detergent</i> )	Non Cov		
	1S2Q	2.1		rasagiline	Cov	2004	40
	1S2Y	2.1		N-propargyl-1(S)-aminoindan	Cov		
	1S3B	1.7		N-methyl-N-propargyl-1(R)-aminoindan	Cov		
	1S3E	1.6		6-hydroxy-N-propargyl-1(R)-aminoindan	Cov		
	2BK3	1.8		farnesol	Non Cov	2005	34

2BK 4	1.9	I199F	rasagiline	Cov		
2BK 5	1.8	I199F	isatin	Non Cov		
2BY B	2.2		selegiline	Cov	2005	41
2C6 4	2.2		6-hydroxy-N-methyl-N-propargyl-1(R)-aminoinda	Cov	2005	34
2C6 5	1.7		4-(N-methyl-N-ethyl-carbamoyloxy)-N-methyl-N-propargyl-1(R)-aminoindan	Cov		
2C6 6	2.5		4-hydroxy-N-propargyl-1(R)-aminoindan	Cov		
2C6 7	1.7		N-methyl-1(R)-aminoindan	Non Cov		
2C7 0	2.1		p-nitrophenylmethylamine ( <i>substrate</i> )	Cov	2006	42
2C7 2	2	Y435H	rasagiline	Cov		
2C7 3	2.2	Y435F	rasagiline	Cov		
2C7 5	1.7	Y435L	rasagiline	Cov		
2C7 6	1.7	Y435W	rasagiline	Cov		
2V5 Z	1.6		safinamide	Non Cov	2007	29
2V6 0	2		7-[(3-chlorobenzyl)oxy]-2-oxo-2H-chromene-4-carbaldehyde	Non Cov		
2V6 1	1.7		7-[(3-chlorobenzyl)oxy]-4-[(methylamino)methyl]-2H-chromen-2-one	Non Cov		
2VR L	2.4		benzylhydrazine	Cov	2008	43
2VR M	2.3		phenyletylhydrazine	Cov		
2VZ 2	2.3		mofegiline	Cov	2008	44
2XF N	1.6		2-(1-benzofran)imidazole	Non Cov	2010	45
2XC G	1.9		tranylcypromine & 2-(1-benzofran)imidazole	Cov		
2XF P	1.7		isatin & 2-(1-benzofran)imidazole	Non Cov		
2XF	2.2		rasagiline & 2-(1-	Cov		

	Q			benzofran)imidazole			
	2XFO	2.1	I199A	tranylcypromine & 2-(1-benzofran)imidazole	Cov		
	3PO7	1.8		zonsinamide	Non Cov	2011	46
	2ZYX	2.2	I199A I326A	methylen blue	Non Cov	2011	47
	4A79	1.89		pioglitazone	Non-Cov	2011	37
	4CRT	1.8		Ass234	Cov	2014	48
<b>Isoform</b>	<b>PDB</b>	<b>Res.</b>	<b>Mutation</b>	<b>Ligand</b>	<b>Binding</b>	<b>Year</b>	<b>Ref.</b>
<b>hMAO-A</b>	2BXR	3		clorgyline	Cov	2005	41
	2BXS	3.2		clorgyline			
	2Z5X	2.2		harmine	Non Cov	2008	27
	2Z5Y	2.2	G110A	harmine	Non Cov		
<b>rMAO-A</b>	1O5W	3.2		clorgyiline	Cov	2004	49

Many scaffolds have been proposed for MAO-BIs both from nature<sup>50, 51</sup> or obtained by synthesis: chalcones,<sup>52-54</sup> chromanones,<sup>55-58</sup> pyridines,<sup>59</sup> benzoxathiolones,<sup>60</sup> pyrroles,<sup>61, 62</sup> indanones,<sup>63</sup> pyridazinones,<sup>63</sup> and many others.

An interesting prospective in the field of neurodegenerative diseases therapy is the application of MAO-BIs with multi-target activity.<sup>6, 35, 64</sup>

The introduction of a carbamate group or the presence of substituted or unsubstituted thiocarbamoyl, acetyl group often leads also to the acetylcholinesterase inhibition.<sup>65-68</sup> This could be advantageous for possible application on AD therapy.<sup>69, 70</sup> Also the presence of moieties able to chelate iron ions seems to be promising: It was found that iron dramatically increases in patients with PD and produces radical oxygen species involved in neurodegeneration.<sup>12</sup> Recently, antagonists of the adenosine receptor A2A with methylxanthine group or benzimidazole group have been studied as MAO-BIs and proposed as new neuroprotective agents dual-target-directed.<sup>36, 71, 72</sup> Finally, several compounds from selegiline and

propargilamine derivatives<sup>73</sup> to lazabemide,<sup>74</sup> quinoline<sup>12</sup> and coumarin derivatives,<sup>75-82</sup> show, besides the MAO-BI activity, antioxidant potential.

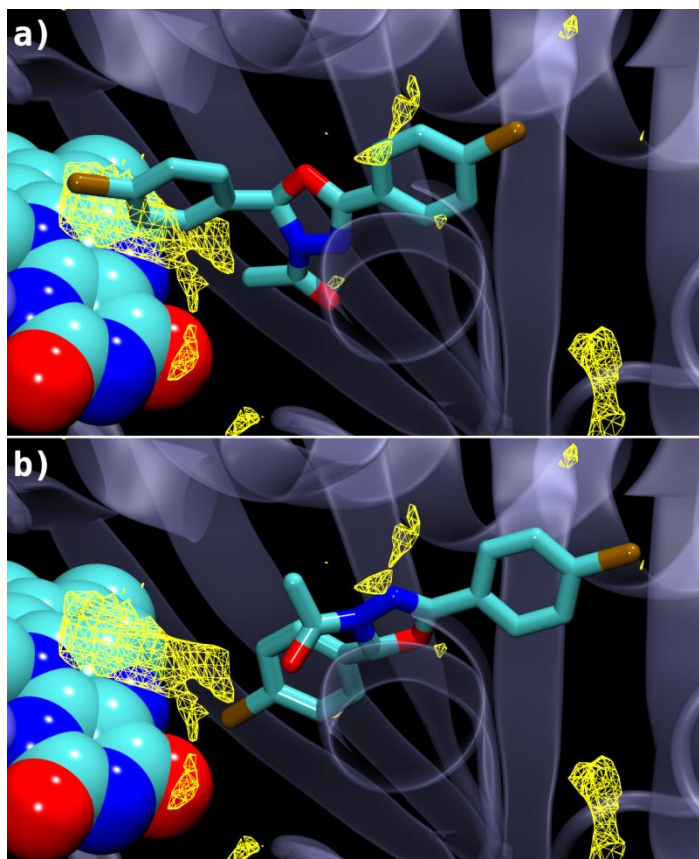
Our group is working in this field since many years and several series of compounds have been synthesised.<sup>83-87</sup> Recently a first series of oxadiazoles was synthesised and tested toward MAO showing a good activity and selectivity toward MAO-B.<sup>88, 89</sup>

## **3.2 NEW OXADIAZOLES AS SELECTIVE INHIBITORS OF MAO-B**

### **3.2.1 Design**

Starting from previously reported theoretical complexes between MAO-B and the 2-(4-chlorophenyl)-3-acetyl-5-(4-chlorophenyl)-2,3-dihydro-1,3,4-oxadiazole<sup>88</sup> we carried out a molecular interaction fields (MIFs) analysis: after removing the inhibitor, the interaction capabilities of the MAO-B active site were investigated by means of CL probe as implemented in the program GRID.<sup>90</sup> We subsequently monitored the resulting isocontour maps taking into account the two most energy stabilized previously reported poses ( $P_a$  and  $P_b$ ) of the inhibitor into the hMAO-B binding cleft.<sup>88</sup>

MIFs indicated a strong interaction between the target and organic chlorine suggesting the possibility to include a second chlorine atom on the *p*-chlorophenyl ring substituent located at the position 2 of the dihydrooxadiazole ring (Figure 4).

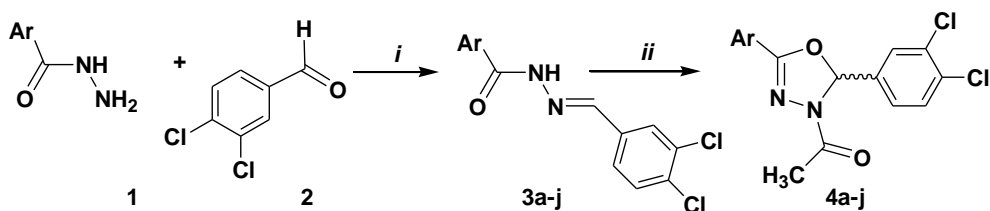


**Figure 4.** Chlorine molecular interaction field (yellow wireframe grids) displayed on previous reported a) P<sub>a</sub> and b) P<sub>b</sub> theoretical complexes of compound **4p**-MAO-B.<sup>88</sup> The FAD cofactor is reported in spacefill notation and the ligand in sticks cyan carbon and brown chlorine atoms colored. The rest of the enzyme is depicted in transparent violet cartoon.

An important requirement for selectivity is the ability of the inhibitors to fit within the elongated bipartite cavity of MAO-B, which is bigger (700 Å<sup>3</sup>) and narrower than the single substrate cavity of MAO-A (550 Å<sup>3</sup>).<sup>23</sup> Therefore the skeleton of the series was maintained because of its ability to occupy both cavities. In fact previous series showed good selectivity toward MAO-B.<sup>88</sup>

### 3.2.2 Synthesis

The **4a-j** series was synthesised according to a previously described procedure<sup>88</sup> as illustrated in Scheme 1.



*i*: CH<sub>3</sub>COOH (CAT)/ EtOH, reflux; *ii*: (CH<sub>3</sub>CO)<sub>2</sub>O, reflux

**Scheme 1.** Synthetic pathway to compounds **4a-j**

Briefly, the suspension in acetic anhydride of the appropriate benzoylhydrazide was heated under reflux until the formation of a yellow-orange solution was observed. The reaction mixture was then poured onto ice water and vigorously stirred. The obtained precipitate was then washed with 10% aqueous NaHCO<sub>3</sub> solution, water, and purified by crystallization. All of the products were characterised by means of both analytical and spectroscopic methods (Experimental section) and tested towards MAO-A and MAO-B.

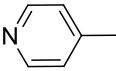
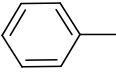
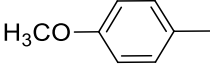
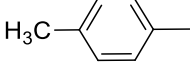
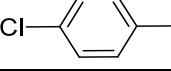
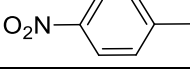
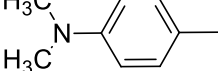
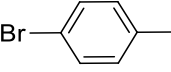
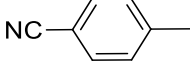
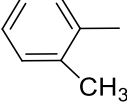
### 3.2.3 Biological activity

The inhibitory activities of compounds **4a-j** on human recombinant MAO-A and -B isoforms, expressed in baculovirus infected BTI infected cells, as  $IC_{50}$ , are reported in Table 2.

Tested compounds demonstrated no interference with the measurements, since they were unable to react directly with the Amplex Red reagent. The kinetic parameters of *h*MAO-A and *h*MAO-B were evaluated in the presence of different tyramine concentrations. In our experiments, *h*MAO-A displayed a Michaelis constant ( $K_m$ ) of  $514 \pm 46.8 \mu\text{M}$  and a maximum reaction velocity ( $V_{max}$ ) of  $301.4 \pm 27.9 \text{ nmol/min/mg protein}$ , whereas *h*MAO-B showed a  $K_m$  of  $104.7 \pm 16.3 \mu\text{M}$  and a  $V_{max}$  of  $28.9 \pm 6.3 \text{ nmol/min/mg protein}$  ( $n = 5$ ). Active compounds showed reversible behaviour according to the method proposed by Cer et al.<sup>91</sup> Therefore, reported  $IC_{50}$  is a useful tool to determine the relative activity of the compounds within the series as well as to determine key substituents in the scaffold.



**Table 2** Inhibitory activities towards hMAO A and hMAO-B of 3-acetyl-2-(3,4-dichlorophenyl)-5-aryl-2,3-dihydro-1,3,4-oxadiazoles derivatives **4a-j**.

Compound	Ar	hMAO-A ( $IC_{50}$ )	hMAO-B ( $IC_{50}$ )
<b>4a</b>		**	69.36 ± 3.51 μM
<b>4b</b>		**	***
<b>4c</b>		**	217.99 ± 10.62 nM
<b>4d</b>		**	19.35 ± 0.68 nM
<b>4e</b>		**	626.82 ± 36.52 nM
<b>4f</b>		**	9.46 ± 0.57 nM
<b>4g</b>		**	**
<b>4h</b>		**	67.69 ± 4.27 nM
<b>4i</b>		**	**
<b>4j</b>		**	2.46 ± 0.17 μM
<b>Clorgyline</b>		4.6 ± 0.3 nM	61.35 ± 1.13 μM
<b>Deprenyl</b>		67.25 ± 1.02 μM	0.019 ± 8,6*10 <sup>-4</sup> μM
<b>Iproniazide</b>		6.56 ± 0.76 μM	7.54 ± 0.36 μM
<b>Moclobemide</b>		3.61 ± 1.94 μM	69.36 ± 3.51 μM

All  $IC_{50}$  values shown in this table are the mean ± S.E.M. from five experiments.  
 \* Inactive at 1 mM (highest concentration tested).  
 \*\* Inactive at 100 μM (highest concentration tested). At higher concentration the compounds precipitate.  
 \*\*\* 100 μM inhibits the corresponding MAO activity by approximately 40-45%. At higher concentration the compounds precipitate.

Half of the tested compounds exhibited inhibitory activity towards hMAO-B at nM concentration, with the exception of compound **4b** which inhibited the corresponding MAO activity by approximately 40-50% at the concentration of 100  $\mu$ M and compounds **4g** and **4i** that were inactive at 100  $\mu$ M. Compounds **4a** and **4j** were effective in the  $\mu$ M range. Interestingly, none of the tested compounds exhibited activity towards the A isoform of the enzyme at the tested concentration of 100  $\mu$ M.

### 3.2.4 Structure Activity Relationships

Biological tests confirmed our hypothesis, in other words, that the introduction of a 3,4-dichlorophenyl moiety in the position 5 of the dihydrooxadiazole ring could lead to more effective and selective compounds with respect to the previously reported mono-chloro derivatives (i.e. the mono-chloro analog of **4f** was inactive on MAO-A whereas demonstrated a MAO-B  $IC_{50}$  equal to  $121.62 \pm 9.63$  nM).<sup>88</sup> Although an excellent remote Hammett correlation ( $\sigma_p$  or  $\sigma_p^+$ ) has been found for *para* substitution in the two aryl rings at the positions 2 and 5 of the dihydrooxadiazole,<sup>89</sup> the electronic effect of the substituent in the *para* position of the phenyl moiety at the position 5 of the dihydrooxadiazole did not seem to play a relevant role in determining the potency of the inhibitors. On the contrary, just the presence of a substituent in this position seemed to be essential for the activity, probably due to the need of bulky substituent in this position. This hypothesis was corroborated by the low activity exhibited by compound **4a** and **4b** bearing a pyridyl and a phenyl substituent respectively. However either the introduction of a dimethylamino moiety or of a cyano as in compound **4g** and **4i**, led to a decrease of the inhibitory activity. It is worth to note that the introduction of a methyl substituent in the ortho position, as in compound **4j**, led to decrease of the inhibitory potency more than 100 fold higher than the corresponding *para*-substituted compound **4d**. Hence a complex blend of electronic and steric effects should be considered to determine the best substitutions for the biological activity.

### 3.2.5 Enantiomeric mixture resolution

All the synthesised compounds have a stereogenic centre at the position 2 of the heterocyclic ring. **4d**, **4e** and **4f** were chosen to perform the separation of the single enantiomers and to evaluate the influence of stereochemistry on their biological properties. Repetitive semipreparative HPLC resolutions carried out on the Chiralpak IA chiral stationary phase (CSP) using the mixture dichloromethane/DEA 100/0.1 (v/v) as eluent enabled us to easily collect tens of mg of enantiopure samples (Table 3). The enantiomeric nature of the samples obtained on mg-scale was demonstrated by polarimetric (Table 3) and circular dichroism (CD) analysis (Figure 5).

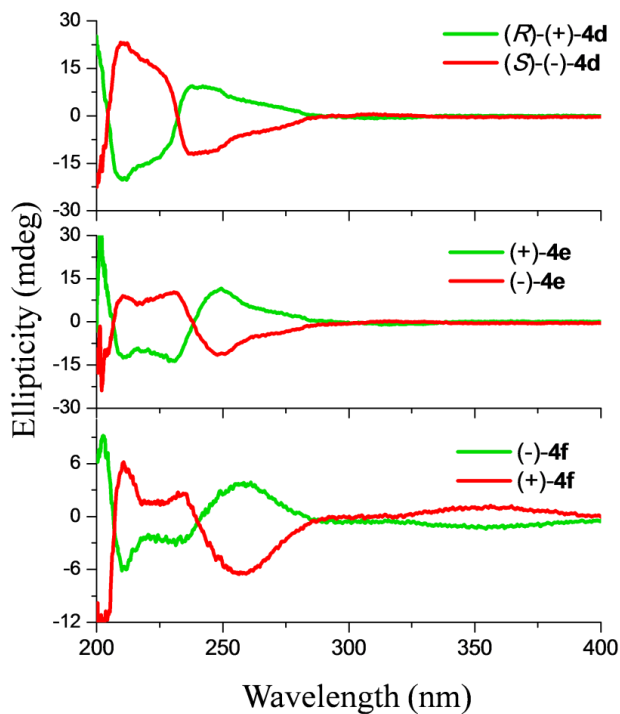
**Table 3.** Chromatographic and polarimetric analysis of the pooled fractions containing the first (F1) and second (F2) eluted enantiomers of **4d-f**.

Compound	$k_1^a$	$k_2^b$	$A^c/V^d$	F1 <sup>e</sup>		F2 <sup>e</sup>	
				e.e.(%)	$[\alpha]_D^{20}$	e.e.(%)	$[\alpha]_D^{20}$
<b>4d</b>	0.14	0.46	10/0.1	>99.0	+15 (c = 0.1, CH <sub>2</sub> Cl <sub>2</sub> )	>99.0	-15 (c = 0.1, CH <sub>2</sub> Cl <sub>2</sub> )
<b>4e</b>	0.10	0.40	6/0.05	>99.0	+7 (c = 0.1, CH <sub>2</sub> Cl <sub>2</sub> )	>99.0	-7 (c = 0.1, CH <sub>2</sub> Cl <sub>2</sub> )
<b>4f</b>	0.10	0.35	5/0.3	>99.0	-35 (c = 0.1, CH <sub>2</sub> Cl <sub>2</sub> )	99.0	+34 (c = 0.1, CH <sub>2</sub> Cl <sub>2</sub> )

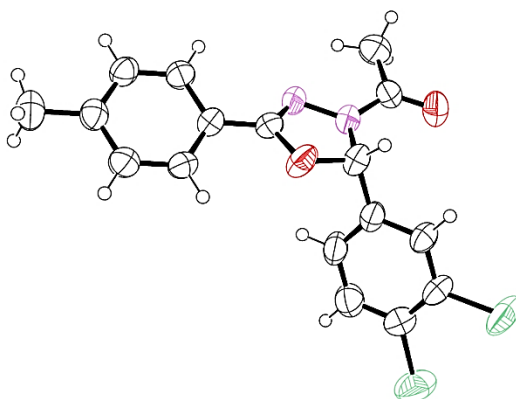
<sup>a</sup> retention factor for the first eluted enantiomer. <sup>b</sup> retention factor for the second eluted enantiomer. <sup>c</sup> amount of sample (in mg) resolved in a single semipreparative run. <sup>d</sup>: volume of sample (in mL). <sup>e</sup> enantiomeric purity and polarimetric data for the pooled fractions containing the first (F1) and second (F2) eluted enantiomers.

The chiroptical properties of the enantiopure antipodes were perfectly specular (Table 3 and Figure 5). The absolute configuration of the enantiomers of **4e** and **4f** was empirically established by CD correlation method using the enantiomers of **4d** as reference samples. Crystallization of the second eluted enantiomer of (-)-**4d** (in the chromatographic conditions reported in Table 3) from ethanol/water afforded single crystals which were suitable for X-ray diffraction analysis. An Oak Ridge Thermal Ellipsoid Plot Program (ORTEP) view of (S)-(-)- **4d** is showed in Figure 6. The replacement of the methyl group on the phenyl ring (compound **4d**) by a chlorine atom

(compound **4e**) or a nitro group (compound **4f**) did not produce significantly alteration in the spectral location of the maximum and minimum of the representative CD bands recorded in ethanol solution (Figure 5).



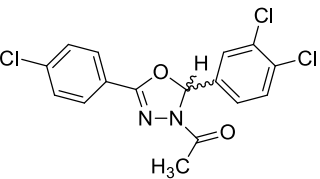
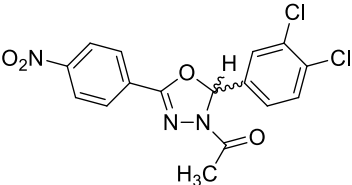
**Figure 5.** CD spectra of the enantiomers of **4d-f** recorded in ethanol.



**Figure 6.** An ORTEP view of the molecular structure of (*S*)-(-)-**4d**

The pure enantiomers of **4e** and **4f**, respectively the less active and the most active compounds in the nM range, were then evaluated for their ability to inhibit the two different MAO isoforms. Also in this case no activity towards MAO-A at the concentration of 100 nM was observed. The results for MAO-B isoform are reported in Table 4: The activity is exclusively (**4e**) or mostly (**4f**) associated with the (*R*)- enantiomer.

**Table 4** Inhibitory activities towards hMAO-B of 3-acetyl-2-(3,4-dichlorophenyl)-5-aryl-2,3-dihydro-1,3,4-oxadiazoles **4e** and **4f** as pure enantiomers and racemates. Data reported in nM.

Compound	Structure	hMAO-B ( $IC_{50}$ )
(±) <b>4e</b>		626.82 ± 36.52
( <i>R</i> )-(+) <b>4e</b>		203.59 ± 18.61
( <i>S</i> )-(-) <b>4e</b>		**
(±) <b>4f</b>		9.46 ± 0.57
( <i>R</i> )-(+) <b>4f</b>		7.61 ± 0.64
( <i>S</i> )-(-) <b>4f</b>		523.46 ± 41.30
Each $IC_{50}$ values is the mean ± S.E.M. from five experiments. ** Inactive at 100000 nM (highest concentration tested).		

### 3.2.6 Molecular modeling

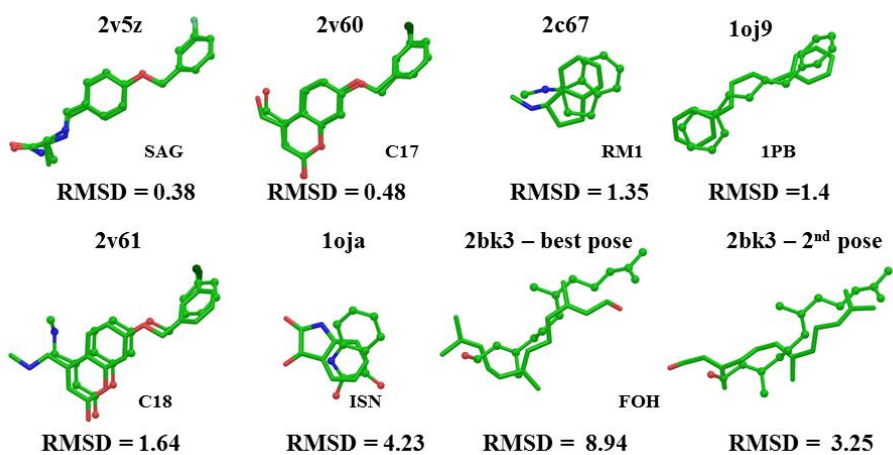
In order to rationalize at molecular level the enzyme inhibition of most promising compounds, suggesting their putative binding mode, we carried out docking and molecular dynamic (MD) simulations, but before that we validated both docking and MD protocol.

Among the crystal structure available (Table 1), we selected crystal structure of non-covalent inhibitors complexes, with a resolution < 2.5 Å,  $K_i$  < 40 μM. Then, docking protocol was validated performing self and cross-docking of seven crystallized reversible inhibitors (Table 5), into the hMAO-B complex crystal structure with the best resolution, having the code 2V5Z.<sup>29</sup>

**Table 5.** PDB crystal structures, of MAO-B in complex with no-covalent inhibitors, employed for docking validation.

PDB entry	Lig. abbrev	Ligand	X-ray res. (Å)	Ki (μM)	ΔG <sub>diss</sub> (kcal/mol)
1OJ9 <sup>92</sup>	1PB	1,4-diphenyl-2-butene[(1E)-4-phenylbut-1-enyl]benzene	2.3	35	6.34
1OJA <sup>92</sup>	RM1	N-methyl-1(R)-aminoindan	1.7	3	7.87
2C67 <sup>93</sup>	ISN	Isatin	1.7	17	6.79
2BK3 <sup>34</sup>	FOH	Farnesol	1.8	2.3	8.04
2V5Z <sup>29</sup>	SAG	Safinamide	1.6	0.45 ± 0.13	9.05
2V60 <sup>29</sup>	C17	7-[(3-chlorobenzyl)oxy]-2-oxo-2H-chromene-4-carbaldehyde	2.0	0.40 ± 0.02	9.12
2V61 <sup>29</sup>	C18	7-[(3-chlorobenzyl)oxy]-4-[(methylamino)methyl]-2H-chromen-2-one	1.7	0.10 ± 0.02	9.96

The purpose of this step was to estimate the efficiency of our docking protocol in predicting the correct orientation of co-crystallized inhibitors into the substrate binding cavity. The biological activity of reference compounds varies with a Ki ranging from 35 to 0.10 μM. Among the docking settings implemented in Glide software (Glide SP,<sup>94</sup> Glide XP,<sup>95</sup> Quantum Mechanics-polarized ligand docking (QMPLD)<sup>96, 97</sup> and Induced Fit workflow protocol<sup>98</sup>), QMPLD workflow<sup>96, 97</sup> successfully reproduced the crystallographic poses of 5 out of 7 compounds within 1.6 Å root-mean-square deviation (RMSD). However, also the docking pose of farnesol, the co-crystallized ligand reported in 2BK3<sup>34</sup> PDB model, in absence of electron density maps could be considered good. In fact, farnesol has a plausible pose rotated of almost of 180° and the second best pose, is slightly shifted compared to experimental data even if with an RMSD value > 2 Å. Finally isatin is quite small and could adopt several binding modes inside the catalytic pocket (Figure 7) (Table 6)



**Figure 7.** Results of seven self and cross-docking experiments performed using 2V5Z as receptor model. Below each docking superimposition is reported the RMSD between co-crystallized (in ball and sticks) and first pose (in sticks). For 2BK3 ligand is reported also the second docking pose.



**Table 6.** Results of 7 (self and cross)-docking experiments performed using 2V5Z as receptor model without water co-crystallized (2V5Z) and with crystallized water (2V5Z-w). Below each setting of docking experiment is reported the RMSD between experimental and best 5 docked poses.

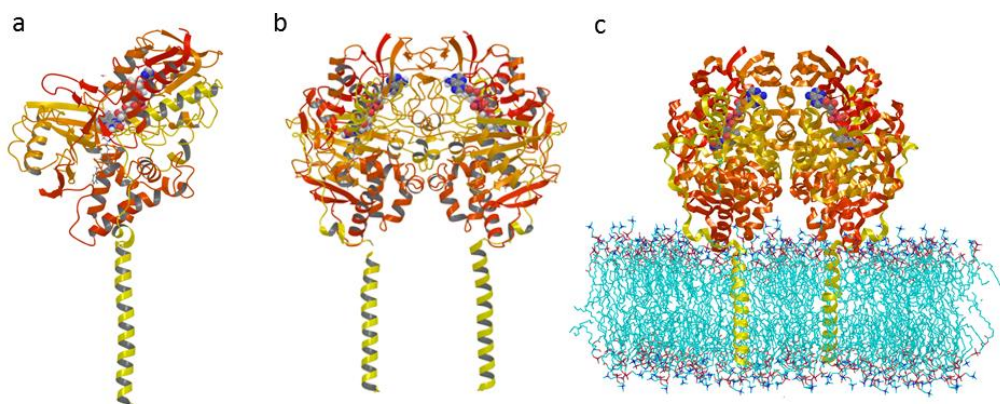
PDB	Lig	2V5Z					2V5Z-w				
		SP	XP	Qmp I	QmplX P	lfd	SP	XP	Qmp I	QmplX P	lfd
<b>1OJ 9</b>	<b>1PB</b>	1.57	1.6 0	<b>1.40</b>	1.60	8.65	8.40	8.3 5	8.42	8.35	1.60
		2.11	1.6 1	1.26	1.66	2.96	8.17	8.2 0	0.88	8.20	
		8.17		1.33	1.67		1.36	1.3 8	1.36	1.38	
		8.40		1.57	0.89		7.89	0.9 0	1.36	0.90	
				8.12	1.37		1.58	1.6 2	1.34	1.62	
<b>1OJ A</b>	<b>ISN</b>	4.28	4.2 7	<b>4.24</b>	4.24	3.61	2.26	9.4 1	2.27	9.41	0.68
		2.07		4.24		4.53	4.69	4.1 0	2.25	4.10	
		3.19		4.24			3.71	4.0 3	2.27	4.03	
		3.56		4.24			4.82	3.7 2	2.26	3.72	
		3.93		4.24			4.99	3.7 2	2.21	3.72	
<b>2BK 3</b>	<b>FO H</b>	3.39	2.2 5	8.94	2.04	3.36	1.98	2.7 2	2.00	2.72	
		3.81		<b>3.25</b>	2.00	8.06	2.50	2.0 4	2.01	2.04	
		2.93		3.29	2.52		1.99	2.3 0	2.30	2.30	
		2.92		3.24	2.66		1.35	8.9 0	2.27	8.90	
		3.41		3.29	9.12		2.23		2.14		
<b>2C6 7</b>	<b>RM 1</b>	5.08	5.0 2	<b>1.35</b>	1.61	1.93	4.83	3.7 6	4.82	3.76	11.2 2
		1.36		1.35	1.74	5.07	10.8 8	1.6 1	4.83	1.61	1.67

		10.20		1.11	10.26	1.94	10.82	3.28	4.83	3.28	
		4.78		0.89	3.26	10.34	10.58		4.83		
		4.75		1.65	9.95		3.27		4.82		
2V5Z	SAG	0.40	2.03	<b>0.38</b>	2.09	1.95	0.36	0.44	0.35	0.44	2.53
		1.54		0.38	0.41	2.28	1.66	0.44	0.35	0.44	2.53
		1.10		0.37	1.76	2.52	1.08	1.68	0.34	1.68	2.54
		1.61		0.37	2.37	2.42	1.91	1.68	0.36	1.68	
				1.10		1.87	1.24		0.36		
2V60	C17	0.47	3.23	<b>0.43</b>	8.39	2.03	0.49	1.40	0.44	1.40	3.07
		0.86	8.39	0.43	3.23	0.84	1.93	8.33	0.44	8.33	1.41
		2.69		0.42	2.27		0.72	8.31	0.44	8.31	2.40
		3.25		0.43	8.45		2.05	2.39	0.50	2.39	
		1.89		0.40	3.36				1.93		
2V61	C18	1.70	3.75	<b>1.64</b>	3.33	2.04	1.95	1.86	1.82	1.86	1.98
		2.18		1.64	2.55	1.70	2.58	1.83	1.84	1.83	1.86
		3.33		1.63	3.77	9.01	8.70	2.48	1.77	2.48	2.50
		2.74		1.63	3.72		3.35	2.45	1.86	2.45	
		2.81		1.64	3.83				2.65		

After this validation step, we applied then the best protocol to dock the separated enantiomers (*R*)-**4e** and (*R*)-**4f**. Due to lack of accuracy of docking scores, the simple docking experiment alone is not often sufficient to rank binding scores with a linear correlation with experimentally measured binding affinities of known complexes.<sup>99</sup>

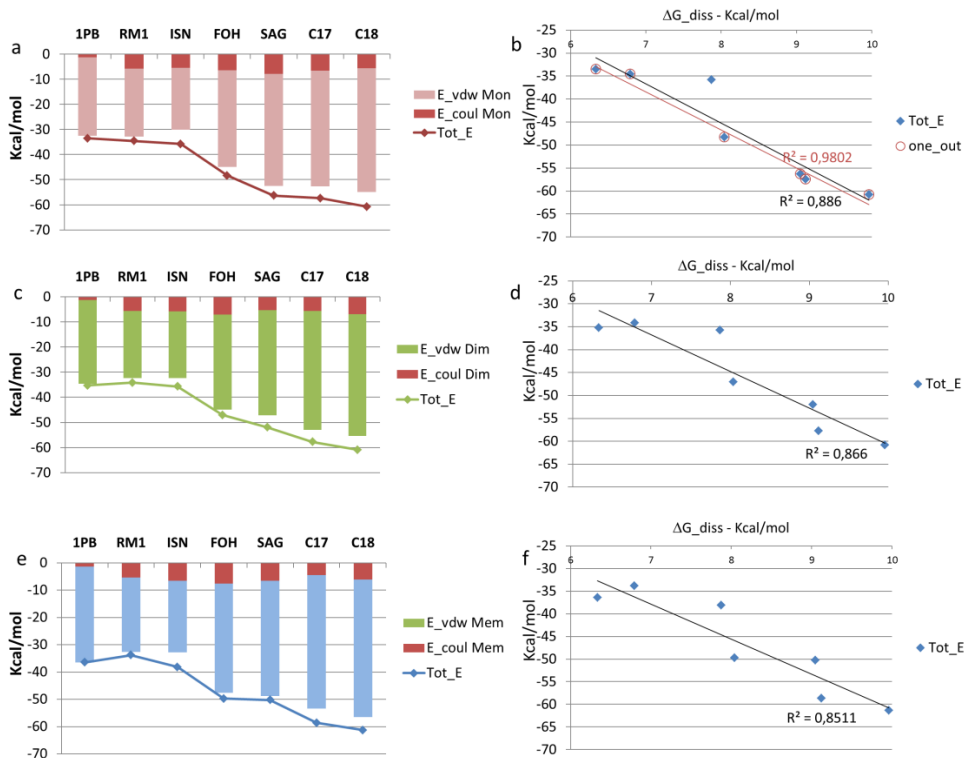
Therefore we decided to use a combination of docking and MD. Actually, explicit solvent MD, coupled with efficient free-energy sampling algorithms, can potentially offer accurate prediction of ligands to proteins binding free energies.

Firstly, as we did for the docking procedure, we validated the molecular dynamic simulation considering the 7 crystal taken into account for docking validation. We wanted to evaluate the importance of the mitochondrial membrane therefore we prepared three different systems settings: monomer, dimer and dimer in membrane (Figure 8).



**Figure 8.** Overview of MAO B a) monomer; b) homodimer; c) homodimer embedded in a lipid bilayer.

We decided to consider the enzyme embedded into the mitochondrial membrane since in literature has been reported its key role in compound recognition. In particular, it was reported that the loop 85-112 is involved in the substrate catalytic site entrance modulation.<sup>100</sup> We observed that the estimated interaction energies well correlate with  $\Delta G$  of dissociation data with all three settings (Figure 9). However, the monomer showed the best correlation with a  $r^2 = 0.89$  (0.98 if we leave the crystal 1OJA (with co-crystallized isatin (ISN)) out of the evaluation).



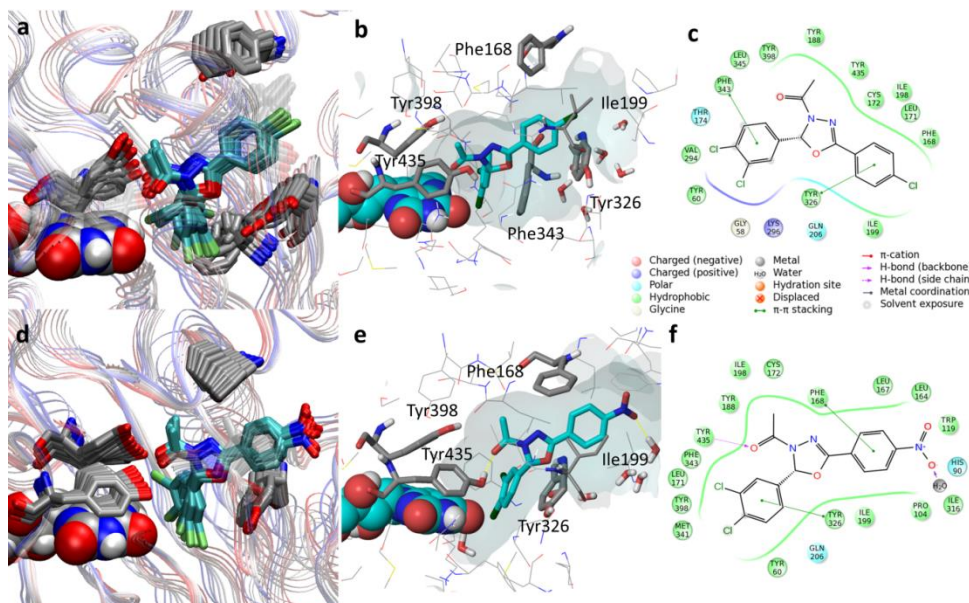
**Figure 9.** MD energy analysis. Inhibitors are arranged from left to right by increasing magnitude of  $\Delta G$  of dissociation. Sum of binding energy components for each inhibitor and setting (a) monomer, (c) dimer, (e) membrane. The Total energy ( $Tot\_E$ ) of interaction is shown with a line. Besides each histogram: figures (b), (d) and (f), the related linear regression plot and  $r^2$  values for the sum of all interaction energies between each inhibitor and MAO B.

According to these data, the presence of the membrane into the theoretical models, although relevant in the stage of compound entrance and recognition, was found to be not essential for the correct prediction of the interaction energy.

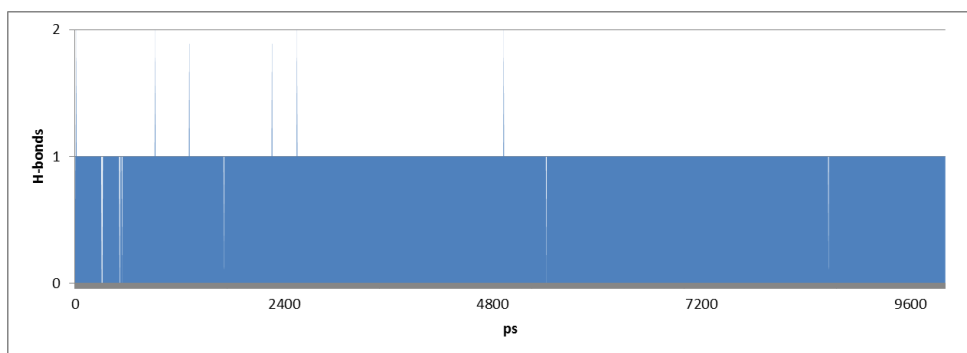
Hence, to better estimate the new ligands-receptor interaction energies, the poses, reporting an RMSD  $> 2 \text{ \AA}$  within the top five ranked docking poses, were submitted to 10 ns of MD simulations (pose <sup>1</sup> and <sup>2</sup>, Table 7).

Thanks to the previously performed studies, the monomer MD setting was used to run further simulation taking into account two starting poses per compound. The best poses results are illustrated in Figure 10.

Compound **4e** was able to enter deeply in the binding cavity and was stabilized by hydrophobic and aromatic interactions (Figure 10a-c). Although the dichlorophenyl moiety was able to rotate, still the  $\pi$ - $\pi$  interaction with Phe343 was conserved. On the contrary, compound **4f** was kept between the entrance and the catalytic cavity separated by the Ile199 due to the interaction of the bulky nitro moiety with residues and waters in this region. More in details, compound **4f** resulted stabilized within the binding site through an array of aromatic (Tyr326, Phe168), hydrophobic (Leu171, Ile199, Leu167, Ile198), and H-bond interactions with binding site residues (Figure 10d-f and Figure 11).

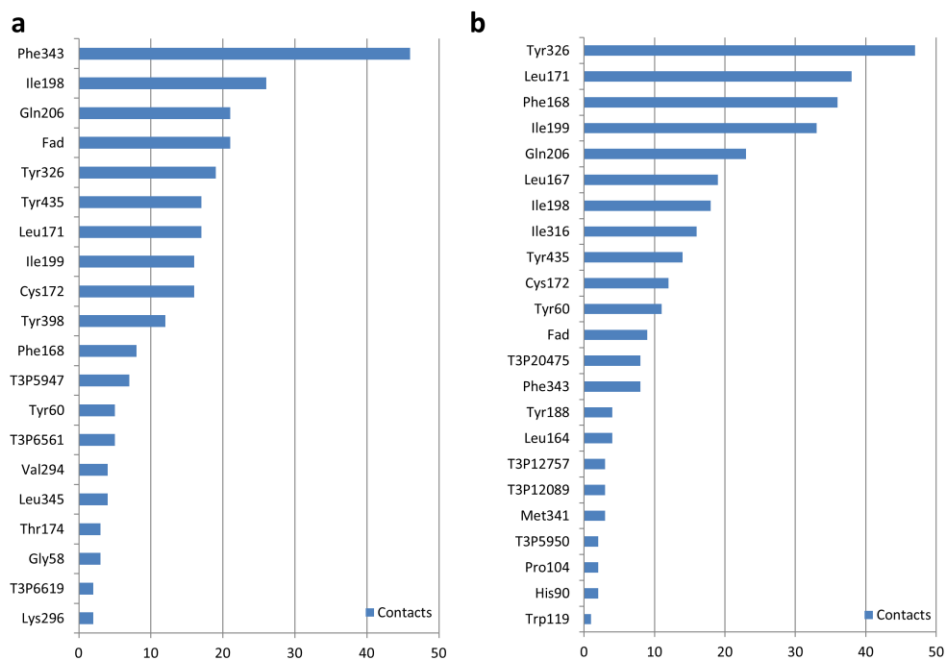


**Figure 10.** a) d) Superimposed structures of 10 ns MD simulations frames of **[4e 4f- MAO-B]** complex coloured by time-step: initial (red), final (blue) along with intermediate structures snapshots. b) e) close-up of the binding cavity; c) f) 2D representing compound and interacting residues.



**Figure 11.** MD trajectory analysis. Number of hydrogen bonds between **4f** and MAO-B

The H-bond between the **4f** carbonyl and the hydroxyl group of the important residue Tyr435, which, together with FAD and Tyr398, delimit the aromatic cage, pivotal for MAO activity, was monitored and found to be stable along the whole MD simulation (Figure 11). In addition, the number of good contacts was higher for **4f** compared to **4e** (Figure 12).



**Figure 12.** Residues involved in the complex stabilization sorted by number of contacts between ligand and receptor. a) **4e**; b) **4f**.

Furthermore, the averaged energy of interaction (Table 7) was in accordance with the difference of activity and underlines the higher weight of van der Waals interactions contribution.

**Table 7** Averaged energies of interaction between new compounds and hMAO-B. Tot-E, Coul-E and vdW-E are respectively the total interaction energy, its electrostatic and van der Waals terms expressed in kcal/mol. The different starting poses are indicated by <sup>1</sup> and <sup>2</sup>.

	Tot-E (kcal/mol)	Coul-E (kcal/mol)	vdW-E (kcal/mol)
<b>(R)-4e<sup>1</sup></b>	-56.76	-5.38	-45.78
<b>(R)-4e<sup>2</sup></b>	-56.19	-2.98	-51.79
<b>(R)-4f<sup>1</sup></b>	-64.15	-9.26	-50.39
<b>(R)-4f<sup>2</sup></b>	-63.75	-6.89	-48.12





With the aim to estimate the drug-likeness of the compounds, we carried out *in silico* ADME-tox prediction.<sup>101</sup> Since **4f** was identified as the best inhibitor of our series, we investigated the prediction of its absorption, distribution, metabolism, excretion and toxicity (ADMEtox) properties using the Qikprop<sup>102</sup> software.

Such an approach provides molecular properties by means of calculation and comparison of the predicted value for the novel compound with respect to the 95% of known drugs. The computed molecular descriptors are shown in Table 8.

All properties were within the ranges specified. The predicted IC<sub>50</sub> for human Ether-à-go-go-Related Gene (HERG) K<sup>+</sup> Channel Blockage was borderline in its value.

However, considering the possibility of a limited error of prediction in discriminating a safe or not-safe compound, biological experiments are needed. Tissue distribution is an important element of the pharmacokinetic (PK) profile of a drug united with knowledge of the *in vitro* activity. The activity of this compound should occur in the brain, therefore we paid great attention on analysing *in silico* molecular descriptors that are required for a CNS active agent like the polar surface area (PSA) and the logarithm of the blood-brain barrier partition (LogBB). The values indicate that newly synthesized molecule would cross blood-brain barrier (BBB) with a low rate. Hence, we checked if this was an issue of other interesting compounds of the series and in particular of **4d**, which also has good inhibitory activity (19.35 ± 0.68 nM). In this case the prediction of CNS activity was positive. Therefore, this compound appears more interesting in this respect. Finally, qualitative human oral absorption was predicted with favourable score. In conclusion, **4d** revealed to have slightly more favourable physicochemical properties and therefore seems to be a good candidate for future development.

**Table 8.** Summary results ADME/Tox prediction for **4f** and **4d** solute, in brackets are reported the ranges observed in 95% of known drugs. (SASA: Solvent accessible surface area; SA: Surface area).

Descriptors	4f	4d	Range 95% of Drug
Solute Molecular Weight	380.19	349.22	( 130.0 / 725.0)
Solute Dipole Moment (D)	9.29	6.01	( 1.0 / 12.5)
Solute Total SASA	612.26	604.33	( 300.0 /1000.0)
Solute Hydrophobic SASA	78.70	166.91	( 0.0 / 750.0)
Solute Hydrophilic SASA	161.56	64.02	( 7.0 / 330.0)
Solute Carbon Pi SASA	241.58	243.00	( 0.0 / 450.0)
Solute Weakly Polar SASA	130.41	130.41	( 0.0 / 175.0)
Solute Molecular Volume (A <sup>3</sup> )	1053.95	1039.37	( 500.0 /2000.0)
Solute vdW Polar SA (PSA)	96.57	51.56	( 7.0 / 200.0)
Solute No. of Rotatable Bonds	0.00	0.00	( 0.0 / 15.0)
Solute as Donor - Hydrogen Bonds	0.00	0.00	( 0.0 / 6.0)
Solute as Acceptor - Hydrogen Bonds	5.50	4.50	( 2.0 / 20.0)
Solute Globularity (Sphere = 1)	0.82	0.82	( 0.75 / 0.95)
Solute Ionization Potential (eV)	9.55	8.96	( 7.9 / 10.5)
Solute Electron Affinity (eV)	1.78	0.75	( -0.9 / 1.7)
QP Polarizability (Angstroms <sup>3</sup> )	37.690M	37.762M	( 13.0 / 70.0)
QP log P for hexadecane/gas	11.725M	10.907M	( 4.0 / 18.0)
QP log P for octanol/gas	17.318M	15.717M	( 8.0 / 35.0)
QP log P for water/gas	9.177M	7.740M	( 4.0 / 45.0)
QP log P for octanol/water	3.16	4.22	( -2.0 / 6.5)
QP log S for aqueous solubility	-5.35	-5.07	( -6.5 / 0.5)
QP log S - conformation independent	-5.55	-5.36	( -6.5 / 0.5)
QP log K <sub>hsa</sub> Serum Protein Binding	0.16	0.43	( -1.5 / 1.5)
QP log BB for brain/blood	-0.79	0.25	( -3.0 / 1.2)
No. of Primary Metabolites	2.00	2.00	( 1.0 / 8.0)
<b>Predicted CNS Activity (-- to ++)</b>	-	+	
HERG K <sup>+</sup> Channel Blockage: log IC <sub>50</sub>	-5.63	-5.59	(concern below -5)
Apparent Caco-2 Permeability (nm/sec)	290.00	2448.00	(<25 poor, >500 great)
Apparent MDCK Permeability (nm/sec)	674.00	6745.00	(<25 poor, >500 great)
QP log K <sub>p</sub> for skin permeability	-3.55	-1.84	(K <sub>p</sub> in cm/hr)
J <sub>m</sub> , max transdermal transport rate	0.00	0.01	(micrograms/cm <sup>2</sup> -hr)
Lipinski Rule of 5 Violations	0.00	0.00	(maximum is 4)
Jorgensen Rule of 3 Violations	0.00	0.00	(maximum is 3)
% Human Oral Absorption in GI (+-20%)	90.00	100.00	(<25% is poor)
Qual. Model for Human Oral Absorption	HIGH	HIGH	(>80% is high)

### 3.2.7 Conclusions

Given our interest in the development of MAO-BIs as a potential therapeutic strategy for neurodegenerative diseases, we developed molecules with optimized activity toward B isoform. Starting from the 2-(4-chlorophenyl)-3-acetyl-5-(4-chlorophenyl)-2,3-dihydro-1,3,4-oxadiazole we found out, through Molecular Interaction Field analysis that the introduction of a second chlorine atom in the 5-phenyl substituent improved the potency.

Therefore, we synthesised a new series of 2-(3,4-dichlorophenyl)-3-acetyl-5-aryl-2,3-dihydro-1,3,4-oxadiazoles. We found out that several compounds showed inhibition properties in the nM range and a high selectivity toward MAO-B. As previously observed for similar compounds, the (*R*)-enantiomer exhibits a higher inhibitory activity compared to both the racemic mixture and the (*S*)-enantiomer.

In addition we have validated a computational protocol of molecular docking and molecular dynamic, obtaining a good qualitative and quantitative accord between experimental and theoretical activities. Therefore, the devised workflow will be applied for future studies on MAO-B.

### 3.3 EXPERIMENTAL PART

#### 3.3.1 Chemistry and compounds characterization

**General methods.** Melting points were uncorrected and were determined on a Reichert Kofler thermopan apparatus. <sup>1</sup>H-NMR spectra were recorded on a Varian Unity 500 and on a Bruker 400, using tetramethylsilane (TMS) as internal standard (chemical shifts in  $\delta$  values). . Electron ionisation (EI) mass spectra were obtained by a Fisons QMD 1000 mass spectrometer (70 eV, 200  $\mu$ A, ion source temperature 200 °C). The samples were introduced directly into the ion source. TLC analyses were performed on silica gel 60 F254 plates. All synthesized compounds were purified by crystallization from an appropriate solvent. Elemental analyses were obtained on a Perkin Elmer 240 B microanalyser. Analytical data of the synthesised compounds are in agreement within  $\pm 0.4$  % of the theoretical values.

**Synthesis of compounds 4a-j.**<sup>88</sup> Equal amounts of aromatic aldehyde (0.018 mol) and the appropriate arylhydrazide (0.018 mol) were refluxed in ethanol (60 mL) under vigorous stirring from 1 to 3 h in presence of acetic acid as catalyst (0.5mL). With respect to the different aromatic substituents in positions 2 and 5 of the oxadiazole ring, a differently coloured solution is obtained. The reaction solution is monitored by TLC (eluent dichloromethane: methanol 20:1). After the mixture is cooled at room temperature, the precipitated product is filtered off and then washed with isopropyl ether. Obtained arylidenearylhydrazides (0.003 mol) are refluxed in 6 mL of acetic anhydride under vigorous stirring from 15 min to 2 h. The suspension is monitored by TLC (eluent chloroform: methanol 20:1). The solution is then poured onto ice-water (100 g) and vigorously stirred. A precipitate is formed which is washed with NaHCO<sub>3</sub> (10% aqueous solution) to remove the acetic acid. The obtained solid is further purified by crystallization.

*2-(3,4-dichlorophenyl)-3-acetyl-5-pyridin-2,3-dihydro-1,3,4-oxadiazole*  
**(4a)**

C<sub>15</sub>H<sub>11</sub>Cl<sub>2</sub>N<sub>3</sub>O<sub>2</sub>. Pink powder, M.p. 117-118°C (isopropyl ether), MS (*m/z*) 294, yield 92%. CHN %: Calc. C, 53.59; H, 3.30; N, 12.50, Found C, 53.55; H, 3.29; N, 12.46

<sup>1</sup>H-NMR (500 MHz, DMSO): δ 2.39 (s, 3H), 7.24 (s, 1H), 7.50 (d, 1H, *J* = 8 Hz), 7.73 (m, 3H, *J* = 8, and 4 Hz), 7.82 (s, 1H), 8.75 (s, 2H). <sup>13</sup>C-NMR (100 MHz, DMSO): δ 21.38, 91.36, 120.39, 127.23, 129.26, 131.39, 131.71, 132.94, 137.15, 150.80 (2C), 153.15, 157.01, 166.44, 166.55

**4b:** *2-(3,4-dichlorophenyl)-3-acetyl-5-phenyl-2,3-dihydro-1,3,4-oxadiazole* **(4b)**

C<sub>16</sub>H<sub>12</sub>Cl<sub>2</sub>N<sub>2</sub>O<sub>2</sub>. White crystals, M.p. 99-100°C (ethanol), MS (*m/z*) 335, yield 43%. CHN %: Calc. C, 57.33; H, 3.61; N, 8.36, Found C, 57.36; H, 3.63; N, 8.37

<sup>1</sup>H-NMR (400 MHz, DMSO): δ 2.27 (s, 3H), 7.22 (s, 1H), 7.48 (dd, 1H, *J* = 8.3 and 2.0 Hz), 7.54 (t, 2H, *J* = 7.6 Hz), 7.61 (t, 1H, *J* = 7.2 Hz), 7.73 (d, 1H, *J* = 8.4 Hz), 7.95 (d, 1H, *J* = 2 Hz), 7.84 (m, 2H). <sup>13</sup>C-NMR (100 MHz, DMSO): δ 21.39, 90.56, 123.89, 126.83 (2C), 127.03, 129.12, 129.32 (2C), 131.46, 131.68, 132.24, 132.79, 137.57, 154.80, 167.27.

*2-(3,4-dichlorophenyl)-3-acetyl-5-(4-methoxyphenyl)-2,3-dihydro-1,3,4-oxadiazole* **(4c)**

C<sub>17</sub>H<sub>14</sub>Cl<sub>2</sub>N<sub>2</sub>O<sub>3</sub>. Pale yellow crystals, M.p. 118-120°C (ethanol), MS (*m/z*) 364, yield 45%. CHN %: Calc. C, 55.91; H, 3.86; N, 7.67, Found C, 55.89; H, 3.86; N, 7.65

<sup>1</sup>H-NMR (500 MHz, DMSO): δ 2.25 (s, 3H), 3.82 (s, 3H), 7.07 (d, 2H, *J* = 9 Hz), 7.17 (1H, s), 7.45 (dd, 1H, *J* = 8.5 and 2.0 Hz), 7.71 (d, 1H, *J* = 8.5 Hz), 7.75 (d, 1H, *J* = 2 Hz), 7.77 (d, 2H, *J* = 8.5 Hz). <sup>13</sup>C-NMR (100 MHz, DMSO): δ 21.32, 55.66, 90.16, 114.75 (2C), 115.93, 126.89, 128.69 (2C), 128.99, 131.40, 131.61, 132.67, 137.67, 154.81, 162.29, 167.01.

*2-(3,4-dichlorophenyl)-3-acetyl-5-(4-methylphenyl)-2,3-dihydro-1,3,4-oxadiazole* **(4d)**

C<sub>17</sub>H<sub>14</sub>Cl<sub>2</sub>N<sub>2</sub>O<sub>2</sub>. Pale yellow crystals, M.p. 120-122°C (isopropyl ether), MS (*m/z*) 348, yield 75%. CHN %: Calc. C, 58.47; H, 4.04; N, 8.02, Found C, 58.50; H, 4.03; N, 8.05

$^1\text{H-NMR}$  (400 MHz, DMSO):  $\delta$  2.26 (s, 3H), 2.38 (s, 3H), 7.2 (s, 1H), 7.34 (d, 2H,  $J = 8$  Hz), 7.46 (dd, 1H,  $J = 8$  and 2 Hz), 7.72 (m, 3H,  $J = 8$ , and 2 Hz), 7.77 (d, 1H,  $J = 2$  Hz).  $^{13}\text{C-NMR}$  (100 MHz, DMSO):  $\delta$  21.32, 21.38, 90.37, 121.06, 126.83 (2C), 126.98, 129.07, 129.87 (2C), 130.17, 131.47, 132.76, 137.63, 142.44, 154.94, 167.17.

*2-(3,4-dichlorophenyl)-3-acetyl-5-(4-chlorophenyl)-2,3-dihydro-1,3,4-oxadiazole(4e)*

$\text{C}_{16}\text{H}_{11}\text{Cl}_3\text{N}_2\text{O}_2$ . Pale yellow crystals, M.p. 119-120°C (ethanol); MS ( $m/z$ ) 367, yield 92%. CHN %: Calc. C, 51.99; H, 3.00; N, 7.58 Found C, 52.02; H, 3.02; N, 7.60

$^1\text{H-NMR}$  (400 MHz, DMSO):  $\delta$  2.27 (3H, s), 7.22 (1H, s), 7.48 (1H, dd,  $J = 8.4$  and 2.0 Hz), 7.61 (2H, d,  $J = 8.8$  Hz), 7.72 (1H, d,  $J = 8.4$  Hz), 7.80 (1H, d,  $J = 2.0$  Hz), 7.84 (2H, d,  $J = 8.8$  Hz).  $^{13}\text{C-NMR}$  (100 MHz, DMSO):  $\delta$  21.39, 90.88, 122.83, 127.10, 128.63 (2C), 129.17, 129.49 (2C), 131.44, 131.70, 132.85, 136.88, 137.41, 154.02, 167.34.

*2-(3,4-dichlorophenyl)-3-acetyl-5-(4-nitrophenyl)-2,3-dihydro-1,3,4-oxadiazole(4f)*

$\text{C}_{16}\text{H}_{11}\text{Cl}_2\text{N}_3\text{O}_4$ . Yellow crystals, M.p. 165-166°C (ethanol); MS ( $m/z$ ) 348, yield 47%. CHN %: Calc. C, 50.55; H, 2.92; N, 11.05 Found C, 50.53; H, 2.91; N, 11.03

$^1\text{H-NMR}$  (500 MHz, DMSO):  $\delta$  2.29 (3H, s), 7.26 (1H, s), 7.51 (1H, dd,  $J = 8$  and 1.5 Hz), 7.72 (1H, d,  $J = 8.5$  Hz), 7.83 (1H, d,  $J = 2.0$  Hz), 8.07 (2H, d,  $J = 8$  Hz), 8.35 (2H, d,  $J = 9$  Hz).

$^{13}\text{C-NMR}$  (100 MHz, DMSO):  $\delta$  21.38, 91.46, 122.51, 124.41 (2C), 127.20, 128.12 (2C), 129.23, 129.83, 131.38, 132.94, 137.14, 149.25, 153.26, 167.55.

*2-(3,4-dichlorophenyl)-3-acetyl-5-(4-N,N-dimethylaminophenyl)-2,3-dihydro-1,3,4-oxadiazole(4g)*

$\text{C}_{18}\text{H}_{17}\text{Cl}_2\text{N}_3\text{O}_2$ . Yellow crystals, M.p. 156-158°C (isopropyl ether); MS ( $m/z$ ) 377, yield 76%. CHN %: Calc. C, 57.16; H, 4.53; N, 11.11 Found C, 57.18; H, 4.55; N, 11.13

$^1\text{H-NMR}$  (400 MHz, DMSO):  $\delta$  2.23 (3H, s), 2.99 (6H, s), 6.77 (2H, d,  $J = 8.8$  Hz), 7.13 (1H, s), 7.43 (1H, dd,  $J = 8.4$ , and 2 Hz), 7.63 (2H, d,  $J = 9.2$  Hz), 7.72 (2H, m,  $J = 8.4$  Hz).  $^{13}\text{C-NMR}$  (100 MHz, DMSO):  $\delta$  21.36, 39.80 (2C),

89.55, 109.72, 111.68 (2C), 126.85, 128.20 (2C), 128.94, 131.45, 131.61, 132.58, 137.99, 152.58, 155.69, 166.68.

*2-(3,4-dichlorophenyl)-3-acetyl-5-(4-bromophenyl)-2,3-dihydro-1,3,4-oxadiazole(4h)*

C<sub>16</sub>H<sub>11</sub>BrCl<sub>2</sub>N<sub>2</sub>O<sub>2</sub>. White cottony solid, M.p. 140-142°C (ethanol); MS (*m/z*) 411, yield 47%. CHN %: Calc. C, 46.41; H, 2.68; N, 6.77 Found C, 46.40; H, 2.69; N, 6.78.

<sup>1</sup>H-NMR (400 MHz, DMSO): δ 2.27 (3H, s), 7.22 (1H, s), 7.48 (1H, dd, *J* = 8.4 and 2.0 Hz), 7.72 (1H, d, *J* = 8.4 Hz), 7.76 (4H, m, *J* = 2.0 Hz), 7.79 (1H, d, *J* = 2 Hz). <sup>13</sup>C-NMR (100 MHz, DMSO): δ 21.39, 90.88, 123.17, 125.77, 127.10, 128.74 (2C), 129.17, 131.44, 131.70, 132.41 (2C), 132.85, 137.40, 154.13, 167.34.

*2-(3,4-dichlorophenyl)-3-acetyl-5-(4-cyanophenyl)-2,3-dihydro-1,3,4-oxadiazole (4i)*

C<sub>17</sub>H<sub>11</sub>Cl<sub>2</sub>N<sub>3</sub>O<sub>2</sub>. Pale yellow crystals, M.p. 148-150°C (ethanol/isopropanol); MS (*m/z*) 359, yield 86%. CHN %: Calc. C, 56.69; H, 3.08; N, 11.67; N, 6.77 Found C, 56.68; H, 3.09, N, 11.68

<sup>1</sup>H-NMR (400 MHz, DMSO): δ 2.28 (3H, s), 7.25 (1H, s); 7.50 (1H, dd, *J* = 8.4, and 2.0 Hz); 7.72 (1H, d, *J* = 8 Hz), 7.83 (1H, d, *J* = 2.0 Hz), 7.99 (4H, m, *J* = 8.6 Hz). <sup>13</sup>C-NMR (100 MHz, DMSO): δ 21.39, 91.34, 114.13, 118.30, 127.22, 127.48 (2C), 128.23, 128.72, 129.25, 131.41, 131.73, 133.21 (2C), 137.22, 153.51, 167.54.

*2-(3,4-dichlorophenyl)-3-acetyl-5-(2-methylphenyl)-2,3-dihydro-1,3,4-oxadiazole (4j)*

C<sub>17</sub>H<sub>14</sub>Cl<sub>2</sub>N<sub>2</sub>O<sub>2</sub>. White solid, M.p. 147-150°C (hexane); MS (*m/z*) 348, yield 32%. CHN %: Calc. C, 58.47; H, 4.04; N, 8.02 Found C, 58.49; H, 4.06; N, 8.04

<sup>1</sup>H-NMR (400 MHz, DMSO) : δ 2.43 (3H, s), 2.70 (3H, s), 7.45 (1H, s), 7.54 (1H, t, CH, *J* = 7.33 Hz), 7.78 (1H, d, *J* = 7.6 Hz), 7.88 (1H, dd, *J* = 8.2 and 2.2), 8.09 (1H, td, *J* = 7.6 and 1.5 Hz), 8.14 (3H, m). <sup>13</sup>C-NMR (100 MHz, DMSO): δ 20.18, 23.11, 91.37, 126.39, 128.28, 128.39, 128.56, 128.95, 129.71, 130.58, 131.34, 132.12, 132.98, 133.35, 136.50, 154.70, 157.99

### 3.3.2 Enantioseparation and X-ray crystal

**HPLC resolution of 4d-f.** The enantioseparations of **4d-f** were performed by using the stainless-steel Chiralpak IA (250 mm x 4.6 mm i.d. and 250 x 10 mm i.d.) (Chiral Technologies Europe, Illkirch, France) columns. All chemicals and solvents for HPLC were purchased from Aldrich (Italy) and used without further purification. The analytical HPLC apparatus consisted of a Perkin-Elmer (Norwalk, CT, USA) 200 LC pump equipped with a Rheodyne (Cotati, CA, USA) injector, a 20- $\mu$ L sample loop, a HPLC Dionex CC-100 oven (Sunnyvale, CA, USA) and a Jasco (Jasco, Tokyo, Japan) Model CD 2095 Plus UV/CD detector. For semipreparative separations a Perkin-Elmer 200 LC pump equipped with a Rheodyne injector, a 1 mL sample loop, a Perkin-Elmer LC 101 oven and Waters 484 detector (Waters Corporation, Milford, MA, USA) were used. The signal was acquired and processed by Clarity software (DataApex, Prague, The Czech Republic).

**Circular dichroism.** The CD spectra were measured in ethanol solution by using a Jasco Model J-700 spectropolarimeter. The concentration was 0.2 mg/mL. The optical path and temperature were set at 0.1 mm and 20°C, respectively. The spectra were average computed over three instrumental scans and the intensities were presented in terms of ellipticity values (mdeg).

**Polarimetry.** Specific rotations were measured at 589 nm by a PerkinElmer polarimeter model 241 equipped with a Na/Hg lamp. The volume of the cell was 1 cm<sup>3</sup> and the optical path was 10 cm. The system was set at 20 °C.

**Crystal structure determination for compound 4d.** C<sub>17</sub>H<sub>14</sub>Cl<sub>2</sub>N<sub>2</sub>O<sub>2</sub> M=349.21, Monoclinic, space group P 2<sub>1</sub>, a=9.095(1), b=6.552(1), c=13.870(1)Å,  $\beta$ =95.982(4), V=822.0(2)Å<sup>3</sup>, Z=2 D<sub>c</sub>=1.415,  $\mu$ =3.643mm<sup>-1</sup>, F(000)= 362. 3040 reflections were collected with a 4.89< $\theta$ <69.78 range with a completeness to theta 95.7%; 2307 were unique, the parameters were 213 and the final R index was 0.0446 for reflections having I>2 $\sigma$ I and 0.0689 for all data. A colourless prismatic shaped crystal (0.08x0.06x0.02) was used for data collection. Hydrogen atoms were assigned in calculated positions, except H2 (on asymmetric C2) which was found in the F.D map. RX-analysis was carried out with a Goniometer Oxford Diffraction KM4



Xcalibur2 at room temperature. Cu/K $\alpha$  radiation (40mA/-40KV), monochromated by an Oxford Diffraction Enhance ULTRA assembly, and an Oxford Diffraction Excalibur PX Ultra CCD were used for cells parameters determination and data collection. The integrated intensities, measured using the  $\omega$  scan mode, were corrected for Lorentz and polarization effects.<sup>102</sup> Direct methods of SIR2004<sup>103</sup> were used in solving the structures and they were refined using the full-matrix least squares on F2 provided by SHELXL97.<sup>104</sup> Multi-scan symmetry-related measurement was used as experimental absorption correction type. The non-hydrogen atoms were refined anisotropically whereas hydrogen atoms were refined as isotropic. The X-ray CIF file for this structure has been deposited at the Cambridge Crystallographic Data Center and allocated with the deposition number CCDC 1030508. Copies of the data can be obtained, free of charge, from CCDC, 12 Union Road, Cambridge, CB2 1EZ UK (e-mail: deposit@ccdc.cam.ac.uk; <http://www.ccdc.cam.ac.uk>).

### 3.3.3 Enzymatic assay

MAO inhibition measurements were evaluated following the general procedure previously described by us<sup>86</sup>. Briefly, test drugs and adequate amounts of recombinant hMAO-A or hMAO-B (Sigma-Aldrich Química S.A., Alcobendas, Spain) required and adjusted to oxidize 165 pmol of p-tyramine/min in the control group, were incubated for 15 min at 37 °C in a flat-black-bottom 96-well microtest plate (BD Biosciences, Franklin Lakes, NJ) placed in the dark fluorimeter chamber. The reaction was started by adding 200  $\mu$  M Amplex Red reagent (Molecular Probes, Inc., Eugene, OR), 1 U/mL horseradish peroxidase, and 1 mM p-tyramine and the production of resorufin, was quantified at 37 °C in a multidetection microplate fluorescence reader (FLX800, Bio-Tek Instruments, Inc., Winooski, VT) based on the fluorescence generated (excitation, 545 nm; emission, 590 nm). The specific fluorescence emission was calculated after subtraction of the background activity, which was determined from vials containing all components except the hMAO isoforms, which were replaced by a sodium phosphate buffer solution.

### 3.3.4 Molecular modelling

**MAO-B active site characterization.** The previous reported<sup>1</sup> theoretical complex between the PDB MAO-B and the most active compound 2-(4-chlorophenyl)-3-acetyl-5-(4-chlorophenyl)-2,3-dihydro-1,3,4-oxadiazole was considered for the active site characterization. The binding cleft was defined by means of a regular box of 1000 Å<sup>3</sup> centred onto the ligand position. After removing the inhibitor, GRID<sup>90</sup> software was applied for mapping the target previously defined using CL probe. Default software directives were taken into account except for the number of planes per Angstrom (NPLA) that was increased till 3 corresponding to a grid spacing equal to 0.333 Å. Computed isocontour maps were displayed using VMD ver. 1.9.1<sup>105</sup> and graphical inspected considering an energy threshold equal to -5.5 kcal/mol.

**Docking protocol.** Three-dimensional coordinates of the receptor<sup>29</sup> were obtained from the Protein Data Bank (PDB). The protein was processed and the internal hydrogen bonding network of the receptor was optimized using the algorithm implemented in Protein Preparation wizard.<sup>106</sup> In order to compare energetic analysis it has been necessary to complete the transmembrane portion considering the aminoacid sequence and  $\alpha$ -helix packing reported in literature.<sup>30,49</sup> The mitochondrial bilayer membrane was simulated with an Dipalmitoylphosphatidylcholine (DPPC) bilayer.<sup>107</sup> Three-dimensional ligand structures for all docking experiments were built using Maestro GUI. Structures were optimized by means of energy minimization carried out using the Merck Molecular Force Field (MMFFs)<sup>108</sup>, the Generalized Born/Surface Area (GB/SA)<sup>109</sup> water implicit solvation model and the Polak-Ribier Coniugate Gradient (PRCG) method, converging on gradient with a threshold of  $0.05 \text{ kJ (mol}\cdot\text{\AA)}^{-1}$  as implemented in Macromodel. Molecular docking studies were performed using Glide SP,<sup>94</sup> Glide XP,<sup>95</sup> QMPLD<sup>96, 97</sup> and Induced Fit workflow protocol.<sup>98</sup> Grids were defined around the refined structure by centring on co-crystallized ligand reported into the PDB entry 2V5Z.<sup>29</sup>

**Molecular dynamics.** The docking resulting complexes were solvated with a box of TIP3P (Transferable Intermolecular Potential 3-Point) water<sup>110</sup> and counter ions were added creating an overall neutral system simulating approximately 0.15 M NaCl. The ions were equally distributed in a water box. The final system was subjected to a MD simulation up to 10 ns using Desmond.<sup>111</sup> Restraint force of  $50 \text{ kcal}/(\text{mol}\cdot\text{\AA})$  was added to monomer and dimer transmembrane portion whereas the dimer in membrane was left without any restraint. The solvated models were optimized, and subsequently the MTK\_NPT (Martyna-Tobias-Klein with constant Number of particles, Pressure and Temperature) ensemble was employed<sup>112</sup>. The default stages in the relaxation process for the NPT ensemble included two energy minimizations and four simulation steps. During the energy minimizations, two runs of 2000 iteration were processed using the steepest descent method: during the first run, the protein structure was fixed by a force restraint constant of  $50 \text{ kcal}/(\text{mol}\cdot\text{\AA})$  and in the second all restraints were removed. With the first simulation, at NVT (constant Number of particles, Volume, and Temperature) ensemble, the system reached a

temperature of 10 K. In the following three simulations in the NPT ensemble, the system was heated up to 325 K and the pressure was kept constant at 1 bar using the Berendsen thermostat–barostat. During the production phase, temperature and pressure were kept constant using the Nosè-Hoover thermostat–barostat. The energy and trajectory were recorded every 1.2 ps and 4.8 ps, respectively. For multiple time step integration, RESPA (REversible reference System Propagator Algorithm)<sup>113</sup> was applied to integrate the equation of motion with Fourier-space electrostatics computed every 6 fs, and all remaining interactions computed every 2 fs. All chemical bond lengths involving hydrogen atoms were fixed with SHAKE.<sup>114</sup> Short range cut-off was set to 9 Å and the smooth particles mesh Ewald method (PME)<sup>115</sup> were used for long range electrostatic interaction. The MD trajectories were analysed in terms of interaction energies and geometries.

**ADME-Tox prediction.** The energy minimized ligands structures were evaluated for their drug-like properties using Qikprop<sup>102</sup> which predicts physically significant descriptors and pharmaceutically relevant properties of organic molecules.<sup>116, 117</sup>

### 3.4 REFERENCES

1. Amit, S. K.; Neal, C.; Bernard, T., Selective inhibitors of monoamine oxidase (MAO-A and MAO-B) as probes of its catalytic site and mechanism. *Med. Res. Rev.* **1995**, 15, 325-388.
2. Collins, G. G. S.; Sandler, M.; Williams, E. D.; Youdim, M. B. H., Multiple forms of human brain mitochondrial monoamine oxidase. *Nature* **1970**, 225, 817-820.
3. O'Carroll, A.-M.; Fowler, C.; Phillips, J.; Tobbia, I.; Tipton, K., The deamination of dopamine by human brain monoamine oxidase. *Naunyn-Schmiedeberg's Arch. Pharmacol.* **1983**, 322, 198-202.
4. Youdim, M. B. H.; Collins, G. G. S.; Sandler, M.; Jones, A. B. B.; Pare, C. M. B.; Nicholson, W. J., Human brain monoamine oxidase. Multiple forms and selective inhibitors. *Nature (London, U. K.)* **1972**, 236, 225-8.
5. Riederer, P.; Danielczyk, W.; Gruenblatt, E., Monoamine oxidase-B inhibition in Alzheimer's disease. *Neurotoxicology* **2004**, 25, 271-277.
6. Youdim, M. B. H.; Edmondson, D.; Tipton, K. F., The therapeutic potential of monoamine oxidase inhibitors. *Nat. Rev. Neurosci.* **2006**, 7, 295-309.
7. Mínguez-Mínguez, S.; Solís-García del Pozo, J.; Jordán, J., Rasagiline in Parkinson's disease: A review based on meta-analysis of clinical data. *Pharmacol. Res.* **2013**, 74, 78-86.
8. Unzeta, M.; Sanz, E. Novel MAO-B inhibitors: Potential therapeutic use of the selective MAO-B inhibitor PF9601N in Parkinson's Disease. In *Int. Rev. Neurobiol.*, Moussa, B. H. Y.; Peter, D., Eds.; Academic Press: 2011; Vol. 100, pp 217-236.

9. Saura, J.; Luque, J. M.; Cesura, A. M.; Da Prada, M.; Chan-Palay, V.; Huber, G.; Loffler, J.; Richards, J. G., Increased monoamine oxidase B activity in plaque-associated astrocytes of Alzheimer brains revealed by quantitative enzyme radioautography. *Neuroscience* **1994**, 62, 15-30.
10. Sano, M.; Ernesto, C.; Thomas, R. G.; Klauber, M. R.; Schafer, K.; Grundman, M.; Woodbury, P.; Growdon, J.; Cotman, C. W.; Pfeiffer, E.; Schneider, L. S.; Thal, L. J., A controlled trial of selegiline, alpha-tocopherol, or both as treatment for Alzheimer's disease. *N. Engl. J. Med.* **1997**, 336, 1216-1222.
11. Weinreb, O.; Amit, T.; Bar-Am, O.; Youdim, M. B. H. A novel anti-Alzheimer's disease drug, Ladostigil: neuroprotective, multimodal brain-selective monoamine oxidase and cholinesterase inhibitor. In *Int. Rev. Neurobiol.*, Moussa, B. H. Y.; Peter, D., Eds.; Academic Press: 2011; Vol. 100, pp 191-215.
12. Zheng, H.; Gal, S.; Weiner, L. M.; Bar-Am, O.; Warshawsky, A.; Fridkin, M.; Youdim, M. B. H., Novel multifunctional neuroprotective iron chelator-monoamine oxidase inhibitor drugs for neurodegenerative diseases: In vitro studies on antioxidant activity, prevention of lipid peroxide formation and monoamine oxidase inhibition. *J. Neurochem.* **2005**, 95, 68-78.
13. Nicotra, A.; Pierucci, F.; Parvez, H.; Senatori, O., Monoamine oxidase expression during development and aging. *Neurotoxicology* **2004**, 25, 155-165.
14. Kumar, M. J.; Andersen, J., Perspectives on MAO-B in aging and neurological disease. *Mol. Neurobiol.* **2004**, 30, 77-89.
15. Palmer, A. M.; DeKosky, S. T., Monoamine neurons in aging and Alzheimer's disease. *Journal of Neural Transmission / General Section JNT* **1993**, 91, 135-159.

16. Alper, G.; Girgin, F. K.; Ozgönül, M.; Menteş, G.; Ersöz, B., MAO inhibitors and oxidant stress in aging brain tissue. *Eur. Neuropsychopharmacol.* **1999**, 9, 247-252.
17. Good, P. F.; Werner, P.; Hsu, A.; Olanow, C. W.; Perl, D. P., Evidence for neuronal oxidative damage in Alzheimer's disease. *Am. J. Pathol.* **1996**, 149, 21-28.
18. Mallajosyula, J. K.; Kaur, D.; Chinta, S. J.; Rajagopalan, S.; Rane, A.; Nicholls, D. G.; Di Monte, D. A.; Macarthur, H.; Andersen, J. K., MAO-B elevation in mouse brain astrocytes results in Parkinson's pathology. *PLoS One* **2008**, 3, e1616.
19. Cash, A. D.; Perry, G.; Smith, M. A., Therapeutic Potential in Alzheimer Disease. *Curr. Med. Chem.* **2002**, 9, 1605-1610.
20. Yu, P. H., Pharmacological and clinical implications of MAO-B inhibitors. *General Pharmacology: The Vascular System* **1994**, 25, 1527-1539.
21. Volz, H. P.; Gleiter, C. H., Monoamine oxidase inhibitors. A perspective on their use in the elderly. *Drugs Aging* **1998**, 13, 341-355.
22. Tatton, W. G.; Ju, W. Y. L.; Holland, D. P.; Tai, C.; Kwan, M., (-)-Deprenyl reduces PC12 cell apoptosis by inducing new protein synthesis. *J. Neurochem.* **1994**, 63, 1572-1575.
23. Edmondson, D. E.; Binda, C.; Mattevi, A., Structural insights into the mechanism of amine oxidation by monoamine oxidases A and B. *Arch. Biochem. Biophys.* **2007**, 464, 269-276.
24. Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E., The Protein Data Bank. *Nucleic Acids Res.* **2000**, 28, 235-242.

25. Geha, R. M.; Chen, K.; Wouters, J.; Ooms, F.; Shih, J. C., Analysis of conserved active site residues in monoamine oxidase A and B and their three-dimensional molecular modeling. *J. Biol. Chem.* **2002**, *277*, 17209-16.
26. Hubalek, F.; Binda, C.; Khalil, A.; Li, M.; Mattevi, A.; Castagnoli, N.; Edmondson, D. E., Demonstration of isoleucine 199 as a structural determinant for the selective inhibition of human monoamine oxidase B by specific reversible inhibitors. *J. Biol. Chem.* **2005**, *280*, 15761-6.
27. Son, S. Y.; Ma, J.; Kondou, Y.; Yoshimura, M.; Yamashita, E.; Tsukihara, T., Structure of human monoamine oxidase A at 2.2-Å resolution: the control of opening the entry for substrates/inhibitors. *Proc. Natl. Acad. Sci. U. S. A.* **2008**, *105*, 5739-44.
28. Binda, C.; Li, M.; Hubalek, F.; Restelli, N.; Edmondson, D. E.; Mattevi, A., Insights into the mode of inhibition of human mitochondrial monoamine oxidase B from high-resolution crystal structures. *Proc. Natl. Acad. Sci. U. S. A.* **2003**, *100*, 9750-5.
29. Binda, C.; Wang, J.; Pisani, L.; Caccia, C.; Carotti, A.; Salvati, P.; Edmondson, D. E.; Mattevi, A., Structures of human monoamine oxidase B complexes with selective noncovalent inhibitors: Safinamide and coumarin analogs. *J. Med. Chem.* **2007**, *50*, 5848-5852.
30. Binda, C.; Hubalek, F.; Li, M.; Edmondson, D. E.; Mattevi, A., Crystal structure of human monoamine oxidase B, a drug target enzyme monotonically inserted into the mitochondrial outer membrane. *FEBS Lett.* **2004**, *564*, 225-8.
31. Li, M.; Binda, C.; Mattevi, A.; Edmondson, D. E., Functional role of the "aromatic cage" in human monoamine oxidase B: structures and catalytic properties of Tyr435 mutant proteins. *Biochemistry* **2006**, *45*, 4775-84.
32. Fariello, R. G., Safinamide. *Neurotherapeutics* **2007**, *4*, 110-116.



33. Müller, T., Current status of Safinamide for the drug portfolio of Parkinson's disease therapy. *Expert Rev. Neurother.* **2013**, 13, 969-977.
34. Hubálek, F.; Binda, C.; Khalil, A.; Li, M.; Mattevi, A.; Castagnoli, N.; Edmondson, D. E., Demonstration of isoleucine 199 as a structural determinant for the selective inhibition of human monoamine oxidase B by specific reversible inhibitors. *J. Biol. Chem.* **2005**, 280, 15761-15766.
35. Chen, J.-F.; Steyn, S.; Staal, R.; Petzer, J. P.; Xu, K.; Van der Schyf, C. J.; Castagnoli, K.; Sonsalla, P. K.; Castagnoli, N.; Schwarzschild, M. A., 8-(3-Chlorostyryl)caffeine may attenuate MPTP neurotoxicity through dual actions of monoamine oxidase inhibition and A2A receptor antagonism. *J. Biol. Chem.* **2002**, 277, 36040-36044.
36. Petzer, J.; Castagnoli, N.; Schwarzschild, M.; Chen, J.-F.; Schyf, C., Dual-target-directed drugs that block monoamine oxidase B and adenosine A2A receptors for Parkinson's disease. *Neurotherapeutics* **2009**, 6, 141-151.
37. Binda, C.; Aldeco, M.; Geldenhuys, W. J.; Tortorici, M.; Mattevi, A.; Edmondson, D. E., Molecular insights into human monoamine oxidase B inhibition by the glitazone antidiabetes drugs. *ACS Medicinal Chemistry Letters* **2011**, 3, 39-42.
38. Binda, C.; Newton-Vinson, P.; Hubalek, F.; Edmondson, D. E.; Mattevi, A., Structure of human monoamine oxidase B, a drug target for the treatment of neurological disorders. *Nat. Struct. Mol. Biol.* **2002**, 9, 22-26.
39. Binda, C.; Li, M.; Hubálek, F.; Restelli, N.; Edmondson, D. E.; Mattevi, A., Insights into the mode of inhibition of human mitochondrial monoamine oxidase B from high-resolution crystal structures. *Proceedings of the National Academy of Sciences* **2003**, 100, 9750-9755.
40. Binda, C.; Hubálek, F.; Li, M.; Herzig, Y.; Sterling, J.; Edmondson, D. E.; Mattevi, A., Crystal Structures of Monoamine Oxidase B in Complex with

Four Inhibitors of the N-Propargylaminoindan Class. *J. Med. Chem.* **2004**, *47*, 1767-1774.

41. De Colibus, L.; Li, M.; Binda, C.; Lustig, A.; Edmondson, D. E.; Mattevi, A., Three-dimensional structure of human monoamine oxidase A (MAO A): Relation to the structures of rat MAO A and human MAO B. *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *102*, 12684-12689.

42. Li, M.; Binda, C.; Mattevi, A.; Edmondson, D. E., Functional Role of the "Aromatic Cage" in Human Monoamine Oxidase B: Structures and Catalytic Properties of Tyr435 Mutant Proteins†,‡. *Biochemistry* **2006**, *45*, 4775-4784.

43. Binda, C.; Wang, J.; Li, M.; Hubalek, F.; Mattevi, A.; Edmondson, D. E., Structural and Mechanistic Studies of Arylalkylhydrazine Inhibition of Human Monoamine Oxidases A and B†‡. *Biochemistry* **2008**, *47*, 5616-5625.

44. Milczek, E. M.; Bonivento, D.; Binda, C.; Mattevi, A.; McDonald, I. A.; Edmondson, D. E., Structural and Mechanistic Studies of Mofegiline Inhibition of Recombinant Human Monoamine Oxidase B||. *J. Med. Chem.* **2008**, *51*, 8019-8026.

45. Bonivento, D.; Milczek, E. M.; McDonald, G. R.; Binda, C.; Holt, A.; Edmondson, D. E.; Mattevi, A., Potentiation of Ligand Binding through Cooperative Effects in Monoamine Oxidase B. *J. Biol. Chem.* **2010**, *285*, 36849-36856.

46. Binda, C.; Aldeco, M.; Mattevi, A.; Edmondson, D. E., Interactions of Monoamine Oxidases with the Antiepileptic Drug Zonisamide: Specificity of Inhibition and Structure of the Human Monoamine Oxidase B Complex. *J. Med. Chem.* **2010**, *54*, 909-912.

47. Milczek, E. M.; Binda, C.; Rovida, S.; Mattevi, A.; Edmondson, D. E., The 'gating' residues Ile199 and Tyr326 in human monoamine oxidase B

function in substrate and inhibitor recognition. *FEBS J.* **2011**, 278, 4860-4869.

48. Esteban, G.; Allan, J.; Samadi, A.; Mattevi, A.; Unzeta, M.; Marco-Contelles, J.; Binda, C.; Ramsay, R. R., Kinetic and structural analysis of the irreversible inhibition of human monoamine oxidases by ASS234, a multi-target compound designed for use in Alzheimer's disease. *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics* **2014**, 1844, 1104-1110.

49. Ma, J.; Yoshimura, M.; Yamashita, E.; Nakagawa, A.; Ito, A.; Tsukihara, T., Structure of rat monoamine oxidase A and its specific recognitions for substrates and inhibitors. *J. Mol. Biol.* **2004**, 338, 103-114.

50. Carradori, S.; D'Ascenzio, M.; Chimenti, P.; Secci, D.; Bolasco, A., Selective MAO-B inhibitors: a lesson from natural products. *Mol. Diversity* **2014**, 18, 219-243.

51. Mazzi, E.; Deiab, S.; Park, K.; Soliman, K., High throughput Screening to Identify Natural Human Monoamine Oxidase B Inhibitors. *Phytother. Res.* **2013**, 27, 818-828.

52. Choi, J. W.; Jang, B. K.; Cho, N.-c.; Park, J.-H.; Yeon, S. K.; Ju, E. J.; Lee, Y. S.; Han, G.; Pae, A. N.; Kim, D. J.; Park, K. D., Synthesis of a series of unsaturated ketone derivatives as selective and reversible monoamine oxidase inhibitors. *Bioorg. Med. Chem.* **2015**, 23, 6486-6496.

53. Morales-Camilo, N.; Salas, C. O.; Sanhueza, C.; Espinosa-Bustos, C.; Sepulveda-Boza, S.; Reyes-Parada, M.; Gonzalez-Nilo, F.; Caroli-Rezende, M.; Fierro, A., Synthesis, Biological Evaluation, and Molecular Simulation of Chalcones and Aurones as Selective MAO-B Inhibitors. *Chem. Biol. Drug Des.* **2015**, 85, 685-695.

54. Mathew, B.; Mathew, G. E.; Ucar, G.; Baysal, I.; Suresh, J.; Vilapurathu, J. K.; Prakasan, A.; Suresh, J. K.; Thomas, A., Development of

fluorinated methoxylated chalcones as selective monoamine oxidase-B inhibitors: Synthesis, biochemistry and molecular docking studies. *Bioorg. Chem.* **2015**, 62, 22-29.

55. Gomes, L. R.; Low, J. N.; Cagide, F.; Chavarria, D.; Borges, F., New insights in the discovery of novel h-MAO-B inhibitors: structural characterization of a series of N-phenyl-4-oxo-4H-chromene-3-carboxamide derivatives. *Acta Crystallogr., Sect. E: Crystallogr. Commun.* **2015**, 71, 547-554.

56. Lan, J.-S.; Xie, S.-S.; Huang, M.; Hu, Y.-J.; Kong, L.-Y.; Wang, X.-B., Chromanones: selective and reversible monoamine oxidase B inhibitors with nanomolar potency. *MedChemComm* **2015**, 6, 1293-1302.

57. Pisani, L.; Farina, R.; Nicolotti, O.; Gadaleta, D.; Soto-Otero, R.; Catto, M.; Di Braccio, M.; Mendez-Alvarez, E.; Carotti, A., In silico design of novel 2H-chromen-2-one derivatives as potent and selective MAO-B inhibitors. *Eur. J. Med. Chem.* **2015**, 89, 98-105.

58. Matos, M. J.; Vina, D.; Vazquez-Rodriguez, S.; Uriarte, E.; Santana, L., Focusing on new monoamine oxidase inhibitors: differently substituted coumarins as an interesting scaffold. *Curr. Top. Med. Chem. (Sharjah, United Arab Emirates)* **2012**, 12, 2210-2239.

59. Park, H. R.; Kim, J.; Kim, T.; Jo, S.; Yeom, M.; Moon, B.; Choo, I. H.; Lee, J.; Lim, E. J.; Park, K. D.; Min, S.-J.; Nam, G.; Keum, G.; Lee, C. J.; Choo, H., Oxazolopyridines and thiazolopyridines as monoamine oxidase B inhibitors for the treatment of Parkinson's disease. *Bioorg. Med. Chem.* **2013**, 21, 5480-5487.

60. Mostert, S.; Petzer, A.; Petzer, J. P., Inhibition of monoamine oxidase by benzoxathiolone analogues. *Bioorg. Med. Chem. Lett.* **2016**, Ahead of Print.

61. La Regina, G.; Silvestri, R.; Artico, M.; Lavecchia, A.; Novellino, E.; Befani, O.; Turini, P.; Agostinelli, E., New Pyrrole Inhibitors of Monoamine Oxidase: Synthesis, Biological Evaluation, and Structural Determinants of MAO-A and MAO-B Selectivity. *J. Med. Chem.* **2007**, 50, 922-931.
62. Ogunrombi, M. O.; Malan, S. F.; Terre'Blanche, G.; Castagnoli, N.; Bergh, J. J.; Petzer, J. P., Structure-activity relationships in the inhibition of monoamine oxidase B by 1-methyl-3-phenylpyrroles. *Bioorg. Med. Chem.* **2008**, 16, 2463-2472.
63. Teran Moldes, M. d. C.; Besada Pereira, P.; Costas Caamano, T.; Costas Lago, M. d. C.; Vila Molares, N.; Vina Castelao, D. Preparation of pyridazin-3(2H)-one derivatives as selective monoamine oxidase B inhibitors. ES2544869A1, 2015.
64. Geldenhuys, W. J.; Schyf, C. J. V. d., Rationally Designed Multi-Targeted Agents Against Neurodegenerative Diseases. *Curr. Med. Chem.* **2013**, 20, 1662-1672.
65. Bruehlmann, C.; Ooms, F.; Carrupt, P.-A.; Testa, B.; Catto, M.; Leonetti, F.; Altomare, C.; Carotti, A., Coumarins derivatives as dual inhibitors of acetylcholinesterase and monoamine oxidase. *J. Med. Chem.* **2001**, 44, 3195-3198.
66. Sterling, J.; Herzig, Y.; Goren, T.; Finkelstein, N.; Lerner, D.; Goldenberg, W.; Miskolczi, I.; Molnar, S.; Rantal, F.; Tamas, T.; Toth, G.; Zagyva, A.; Zekany, A.; Lavian, G.; Gross, A.; Friedman, R.; Razin, M.; Huang, W.; Kraus, B.; Chorev, M.; Youdim, M. B.; Weinstock, M., Novel Dual Inhibitors of AChE and MAO Derived from Hydroxy Aminoindan and Phenethylamine as Potential Treatment for Alzheimer's Disease. *J. Med. Chem.* **2002**, 45, 5260-5279.

67. Ucar, G.; Gokhan, N.; Yesilada, A.; Bilgin, A. A., 1-N-Substituted thiocarbamoyl-3-phenyl-5-thienyl-2-pyrazolines: novel cholinesterase and selective monoamine oxidase B inhibitors for the treatment of Parkinson's and Alzheimer's diseases. *Neurosci. Lett.* **2005**, 382, 327-331.
68. Kumar, V.; Chadha, N.; Tiwari, A. K.; Sehgal, N.; Mishra, A. K., Prospective atom-based 3D-QSAR model prediction, pharmacophore generation, and molecular docking study of carbamate derivatives as dual inhibitors of AChE and MAO-B for Alzheimer's disease. *Med. Chem. Res.* **2014**, 23, 1114-1122.
69. Matilde, Y.; Dolores, V., Dual Inhibitors of Monoamine Oxidase and Cholinesterase for the Treatment of Alzheimer Disease. *Curr. Top. Med. Chem.* **2013**, 13, 1692-1706.
70. Weinstock, M.; Goren, T.; Youdim, M. B. H., Development of a novel neuroprotective drug (TV3326) for the treatment of Alzheimer's disease, with cholinesterase and monoamine oxidase inhibitory activities. *Drug Dev. Res.* **2000**, 50, 216-222.
71. Preti, D.; Baraldi, P. G.; Moorman, A. R.; Borea, P. A.; Varani, K., History and Perspectives of A2A Adenosine Receptor Antagonists as Potential Therapeutic Agents. *Med. Res. Rev.* **2015**, 35, 790-848.
72. Petzer, J. P.; Petzer, A., Caffeine as a Lead Compound for the Design of Therapeutic Agents for the Treatment of Parkinson's Disease. *Curr. Med. Chem.* **2015**, 22, 975-988.
73. Xie, S.; Chen, J.; Li, X.; Su, T.; Wang, Y.; Wang, Z.; Huang, L.; Li, X., Synthesis and evaluation of selegiline derivatives as monoamine oxidase inhibitor, antioxidant and metal chelator against Alzheimer's disease. *Bioorg. Med. Chem.* **2015**, 23, 3722-9.

74. Toprakci, M.; Yelekci, K., Docking studies on monoamine oxidase-B inhibitors: Estimation of inhibition constants ( $K_i$ ) of a series of experimentally tested compounds. *Bioorg. Med. Chem. Lett.* **2005**, 15, 4438-4446.
75. Carotti, A.; Altomare, C.; Catto, M.; Gnerre, C.; Summo, L.; De Marco, A.; Rose, S.; Jenner, P.; Testa, B., Lipophilicity plays a major role in modulating the inhibition of monoamine oxidase B by 7-substituted coumarins. *Chem Biodivers* **2006**, 3, 134-149.
76. Carotti, A.; Melloni, P.; Thaler, F.; Caccia, C.; Maestroni, S.; Salvati, P. Preparation of substituted aminoalkyl- and amidoalkyl-benzopyran derivatives as selective and reversible MAO-B inhibitors. 2006-EP1572 2006102958, 20060222., 2006.
77. Catto, M.; Nicolotti, O.; Leonetti, F.; Carotti, A.; Favia, A. D.; Soto-Otero, R.; Mendez-Alvarez, E.; Carotti, A., Structural Insights into Monoamine Oxidase Inhibitory Potency and Selectivity of 7-Substituted Coumarins from Ligand- and Target-Based Approaches. *J. Med. Chem.* **2006**, 49, 4912-4925.
78. Chimenti, F.; Secci, D.; Bolasco, A.; Chimenti, P.; Granese, A.; Befani, O.; Turini, P.; Alcaro, S.; Ortuso, F., Inhibition of monoamine oxidases by coumarin-3-acyl derivatives: biological activity and computational study. *Bioorg. Med. Chem. Lett.* **2004**, 14, 3697-3703.
79. Gnerre, C.; Catto, M.; Leonetti, F.; Weber, P.; Carrupt, P.-A.; Altomare, C.; Carotti, A.; Testa, B., Inhibition of Monoamine Oxidases by Functionalized Coumarin Derivatives: Biological Activities, QSARs, and 3D-QSARs. *J. Med. Chem.* **2000**, 43, 4747-4758.
80. Novaroli, L.; Daina, A.; Favre, E.; Bravo, J.; Carotti, A.; Leonetti, F.; Catto, M.; Carrupt, P.-A.; Reist, M., Impact of Species-Dependent Differences

on Screening, Design, and Development of MAO B Inhibitors. *J. Med. Chem.* **2006**, *49*, 6264-6272.

81. Matos, M. J.; Rodríguez-Enríquez, F.; Borges, F.; Santana, L.; Uriarte, E.; Estrada, M.; Rodríguez-Franco, M. I.; Laguna, R.; Viña, D., 3-Amidocoumarins as Potential Multifunctional Agents against Neurodegenerative Diseases. *ChemMedChem* **2015**, *10*, 2071-2079.

82. Gaspar, A.; Matos, M. J.; Garrido, J.; Uriarte, E.; Borges, F., Chromone: A Valid Scaffold in Medicinal Chemistry. *Chem. Rev.* **2014**, *114*, 4960-4992.

83. Chimenti, F.; Maccioni, E.; Secci, D.; Bolasco, A.; Chimenti, P.; Granese, A.; Befani, O.; Turini, P.; Alcaro, S.; Ortuso, F.; Cardia, M. C.; Distinto, S., Selective inhibitory activity against MAO and molecular modeling studies of 2-thiazolylhydrazone derivatives. *J. Med. Chem.* **2007**, *50*, 707-712.

84. Chimenti, F.; Maccioni, E.; Secci, D.; Bolasco, A.; Chimenti, P.; Granese, A.; Befani, O.; Turini, P.; Alcaro, S.; Ortuso, F.; Cirilli, R.; La Torre, F.; Cardia, M. C.; Distinto, S., Synthesis, molecular modeling studies, and selective inhibitory activity against monoamine oxidase of 1-thiocarbamoyl-3,5-diaryl-4,5-dihydro-(1H)-pyrazole derivatives. *J. Med. Chem.* **2005**, *48*, 7113-7122.

85. Chimenti, F.; Secci, D.; Bolasco, A.; Chimenti, P.; Granese, A.; Carradori, S.; MacCioni, E.; Cardia, M. C.; Yáñez, M.; Orallo, F.; Alcaro, S.; Ortuso, F.; Cirilli, R.; Ferretti, R.; Distinto, S.; Kirchmair, J.; Langer, T., Synthesis, semipreparative HPLC separation, biological evaluation, and 3D-QSAR of hydrazothiazole derivatives as human monoamine oxidase B inhibitors. *Bioorg. Med. Chem.* **2010**, *18*, 5063-5070.



86. Distinto, S.; Yáñez, M.; Alcaro, S.; Cardia, M. C.; Gaspari, M.; Sanna, M. L.; Meleddu, R.; Ortuso, F.; Kirchmair, J.; Markt, P.; Bolasco, A.; Wolber, G.; Secci, D.; MacCioni, E., Synthesis and biological assessment of novel 2-thiazolyhydrazones and computational analysis of their recognition by monoamine oxidase B. *Eur. J. Med. Chem.* **2012**, *48*, 284-295.
87. MacCioni, E.; Alcaro, S.; Orallo, F.; Cardia, M. C.; Distinto, S.; Costa, G.; Yanez, M.; Sanna, M. L.; Vigo, S.; Meleddu, R.; Secci, D., Synthesis of new 3-aryl-4,5-dihydropyrazole-1-carbothioamide derivatives. An investigation on their ability to inhibit monoamine oxidase. *Eur. J. Med. Chem.* **2010**, *45*, 4490-4498.
88. Maccioni, E.; Alcaro, S.; Cirilli, R.; Vigo, S.; Cardia, M. C.; Sanna, M. L.; Meleddu, R.; Yanez, M.; Costa, G.; Casu, L.; Matyus, P.; Distinto, S., 3-Acetyl-2,5-diaryl-2,3-dihydro-1,3,4-oxadiazoles: A New Scaffold for the Selective Inhibition of Monoamine Oxidase B. *J. Med. Chem.* **2011**, *54*, 6394-6398.
89. Cerioni, G.; Maccioni, E.; Cardia, M. C.; Vigo, S.; Mocci, F., Characterization of 2,5-diaryl-1,3,4-oxadiazolines by multinuclear magnetic resonance and density functional theory calculations. investigation on a case of very remote hammett correlation. *Magn. Reson. Chem.* **2009**, *47*, 727-733.
90. Goodford, P. J., A computational procedure for determining energetically favorable binding sites on biologically important macromolecules. *J. Med. Chem.* **1985**, *28*, 849-857.
91. Cer, R. Z.; Mudunuri, U.; Stephens, R.; Lebeda, F. J., IC50-to-K-i: A web-based tool for converting IC50 to K-i values for inhibitors of enzyme activity and ligand binding. *Nucleic Acids Res.* **2009**, *37*, W441-W445.
92. Binda, C.; Li, M.; Hubálek, F.; Restelli, N.; Edmondson, D. E.; Mattevi, A., Insights into the mode of inhibition of human mitochondrial monoamine

oxidase B from high-resolution crystal structures. *Proc. Natl. Acad. Sci. U. S. A.* **2003**, 100, 9750-9755.

93. Binda, C.; Hubálek, F.; Li, M.; Herzig, Y.; Sterling, J.; Edmondson, D. E.; Mattevi, A., Binding of Rasagiline-Related Inhibitors to Human Monoamine Oxidases: A Kinetic and Crystallographic Analysis. *J. Med. Chem.* **2005**, 48, 8148-8154.

94. Friesner, R. A.; Banks, J. L.; Murphy, R. B.; Halgren, T. A.; Klicic, J. J.; Mainz, D. T.; Repasky, M. P.; Knoll, E. H.; Shelley, M.; Perry, J. K.; Shaw, D. E.; Francis, P.; Shenkin, P. S., Glide: A New Approach for Rapid, Accurate Docking and Scoring. 1. Method and Assessment of Docking Accuracy. *J. Med. Chem.* **2004**, 47, 1739-1749.

95. Friesner, R. A.; Murphy, R. B.; Repasky, M. P.; Frye, L. L.; Greenwood, J. R.; Halgren, T. A.; Sanschagrín, P. C.; Mainz, D. T., Extra precision Glide: Docking and scoring incorporating a model of hydrophobic enclosure for protein–ligand complexes. *J. Med. Chem.* **2006**, 49, 6177-6196.

96. Mohamadi, F.; Richards, N. G. J.; Guida, W. C.; Liskamp, R.; Lipton, M.; Caufield, C.; Chang, G.; Hendrickson, T.; Still, W. C., MacroModel—an integrated software system for modeling organic and bioorganic molecules using molecular mechanics. *J. Comput. Chem.* **1990**, 11, 440-467.

97. Chung, J. Y.; Hah, J.-M.; Cho, A. E., Correlation between performance of QM/MM docking and simple classification of binding sites. *J. Chem. Inf. Model.* **2009**, 49, 2382-2387.

98. Sherman, W.; Day, T.; Jacobson, M. P.; Friesner, R. A.; Farid, R., Novel procedure for modeling ligand/receptor induced fit effects. *J. Med. Chem.* **2005**, 49, 534-553.

99. Warren, G. L.; Andrews, C. W.; Capelli, A.-M.; Clarke, B.; LaLonde, J.; Lambert, M. H.; Lindvall, M.; Nevins, N.; Semus, S. F.; Senger, S.; Tedesco, G.;

- Wall, I. D.; Woolven, J. M.; Peishoff, C. E.; Head, M. S., A critical assessment of docking programs and scoring functions. *J. Med. Chem.* **2006**, *49*, 5912-5931.
100. Edmondson, D. E.; Binda, C.; Wang, J.; Upadhyay, A. K.; Mattevi, A., Molecular and mechanistic properties of the membrane-bound mitochondrial monoamine oxidases. *Biochemistry* **2009**, *48*, 4220-4230.
101. Lagorce, D.; Reynes, C.; Camproux, A.-C.; Miteva, M. A.; Sperandio, O.; Villoutreix, B. O. In silico ADME/Tox predictions. In *ADMET for Medicinal Chemists*; John Wiley & Sons, Inc.: 2011, pp 29-124.
102. Walker, N.; Stuart, D., An empirical method for correcting diffractometer data for absorption effects. *Acta Crystallographica, Section A* **1983**, *39*, 158-166.
103. Burla, M. C.; Caliendo, R.; Camalli, M.; Carrozzini, B.; Cascarano, G. L.; De Caro, L.; Giacovazzo, C.; Polidori, G.; Spagna, R., SIR2004: an improved tool for crystal structure determination and refinement. *J. Appl. Crystallogr.* **2005**, *38*, 381-388.
104. Sheldrick, G. M., SHELXL97: Program for crystal structure refinement. *Institut für Anorganische Chemie de Universität Göttingen* **1997**.
105. Humphrey, W.; Dalke, A.; Schulten, K., VMD: Visual molecular dynamics. *J. Mol. Graph.* **1996**, *14*, 33-38.
106. Piomelli, D.; Sasso, O., Peripheral gating of pain signals by endogenous lipid mediators. *Nat. Neurosci.* **2014**, *17*, 164-174.
107. Balali-Mood, K.; Bond, P. J.; Sansom, M. S. P., Interaction of monotopic membrane enzymes with a lipid bilayer: a coarse-grained MD simulation study. *Biochemistry* **2009**, *48*, 2135-2145.

108. Halgren, T., Merck molecular force field. II. MMFF94 van der Waals and electrostatic parameters for intermolecular interactions. *J. Comput. Chem.* **1996**, 17, 520-552.
109. Hasel, W.; Hendrickson, T. F.; Still, W. C., A rapid approximation to the solvent accessible surface areas of atoms. *Tetrahedron Computer Methodology* **1988**, 1, 103-116.
110. Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L., Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **1983**, 79, 926-935.
111. Bowers, K. J.; Dror, R. O.; Shaw, D. E., The midpoint method for parallelization of particle simulations. *J. Chem. Phys.* **2006**, 124, 184109-11.
112. Martyna, G. J.; Tobias, D. J.; Klein, M. L., Constant pressure molecular dynamics algorithms. *J. Chem. Phys.* **1994**, 101, 4177-4189.
113. Gibson, D. A.; Carter, E. A., Time-reversible multiple time-scale ab-initio molecular-dynamics. *J. Phys. Chem.* **1993**, 97, 13429-13434.
114. Ryckaert, J.-P.; Ciccotti, G.; Berendsen, H., Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *J. Comput. Phys.* **1977**, 23, 327-341.
115. Darden, T.; York, D.; Pedersen, L., Particle mesh Ewald: An N [center-dot] log(N) method for Ewald sums in large systems. *J. Chem. Phys.* **1993**, 98, 10089-10092.
116. Lipinski, C. A.; Lombardo, F.; Dominy, B. W.; Feeney, P. J., Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv. Drug Delivery. Rev.* **2001**, 46, 3-26.
117. Jorgensen, W. L.; Duffy, E. M., Prediction of drug solubility from structure. *Adv. Drug Delivery. Rev.* **2002**, 54, 355-366.



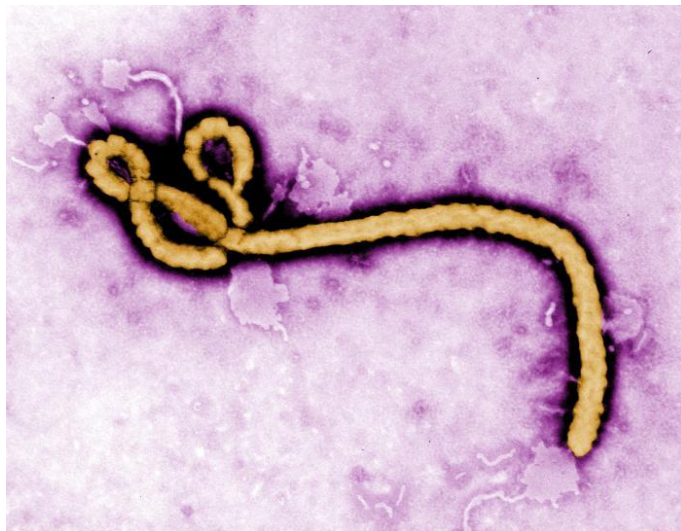
## 4 CHAPTER

### 4.1 INTRODUCTION TO VP35 OF EBOLA VIRUS

EBOV was first discovered in 1976 near the Ebola river in Zaire (now Democratic Republic of the Congo) and has re-emerged in the last decade with an increase of incidence and geographical extension of human outbreaks. In March 2014, largest and most complex outbreak was registered in several countries of West Africa with more than 28000 cases and 11000 deaths as of November 2015.<sup>1</sup>

Ebola virus disease (EVD) is a severe, often fatal illness in humans caused by infection with negative-sense ssRNA viruses of the family *Filoviridae*, genus *Ebolavirus*.<sup>2, 3</sup> This disease causes haemorrhagic fever in humans and non-human primates.

Five *Ebolavirus* species have been identified so far, four of which are known to be pathogenic to humans: *Zaire ebolavirus* (EBOV, formerly ZEBOV), *Sudan ebolavirus* (SUDV, formerly SEBOV), *Tai Forest ebolavirus* (TAFV, formerly *Cote d'Ivoire ebolavirus*, CIEBOV) and *Bundibugyo ebolavirus* (BDBV, formerly BEBOV). The fifth species, *Reston ebolavirus* (RESTV, formerly REBOV), has caused disease in non-human primates but not in humans.<sup>4</sup>



**Figure 1.** Electron microscope image showing ultrastructural morphology assumed by filamentous virion ZEBOV (F.E. Murphy, CDC).

Death from EBOV infection is associated with markedly impaired coagulation and innate immunity cascades, increased production of pro-inflammatory cytokines, profound immune suppression resulting in peripheral T lymphocyte apoptosis and a lack of adaptive immunity.<sup>5</sup> The success of EBOV replication is dependent on viral inhibition of initial innate immune responses to infection. Disarming innate immune responses is a common mechanism employed by highly pathogenic human viruses,<sup>6</sup> and EBOV is one of the most successful in evading innate immune interference.

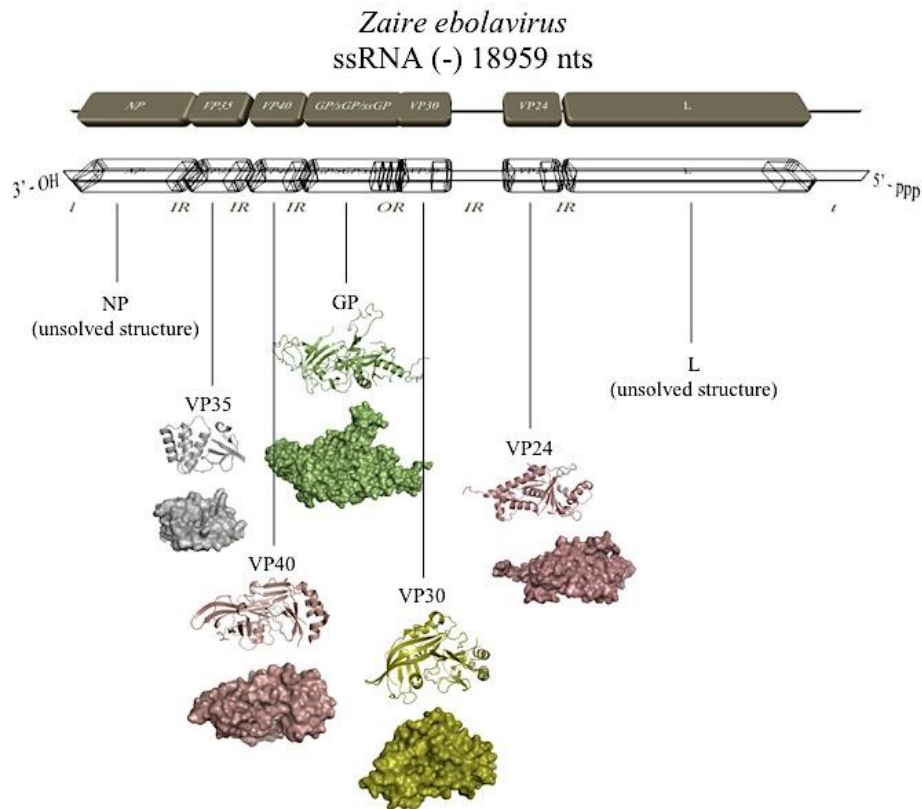
EBOV is introduced into the human population through close contact with the blood, secretions, organs or other bodily fluids of infected animals such as chimpanzees, gorillas, fruit bats, monkeys, followed by subsequent human-to-human transmission.<sup>7</sup> In response to the 2014 EVD outbreak, a human trial campaign has been prompted and is currently underway for several vaccine platforms.<sup>8-11</sup> There is as yet no licensed treatment proven to neutralise the virus but a range of blood, immunological and drug therapies are under development. These include a monoclonal antibody cocktail known as ZMapp, small interfering RNA molecules and phosphorodiamidate morpholino oligomers such as the TKM-Ebola and the AVI-7537 respectively, or nucleotide/nucleoside analogues like Favipiravir and BCX-4430, as well as other small-molecule inhibitors such as Brincidofovir.<sup>12</sup>

In humans, EBOV haemorrhagic fever often results in very high mortality rates, reaching 90% for the most lethal species such as Zaire ebolavirus (ZEBOV). Early events in EBOV infection influence the patient's ability to develop an effective immune response. In fact, survivors to EBOVs infection seem to develop an effective immune response.<sup>13, 14</sup> This suggests that events early in the EBOV infection may influence the patients' ability to activate an effective immune response.<sup>15,16</sup> Hence, the development drug targeting EBOV specific enzymes would be advantageous.

EBOV genome is about 19 kb long and includes 7 genes arranged in tandem 9 coding for viral proteins , including 7 structural and 2 non-structural.

Starting at the 3' to the 5' of the genome we can find the NP gene, coding for the nucleoprotein of the same name; VP35 (Viral Protein 35), encoding for the homonymous polymerase cofactor; VP30, whose gene

product is a transcriptional cofactor; GP , coding for the only surface glycoprotein and, through a process of RNA editing for 2 glycoproteins soluble SGP and ssGP; VP24 and VP40, encoding for the two matrix proteins; L, coding for the catalytic subunit of the complex-RNA-dependent RNA polymerase (RdRp). The 7 genes are separated by intergenic regions transcribed and untranslated, and two regulatory regions leader and trailer which appear respectively at the end of the genome in the 3' and 5'.<sup>17, 18</sup>



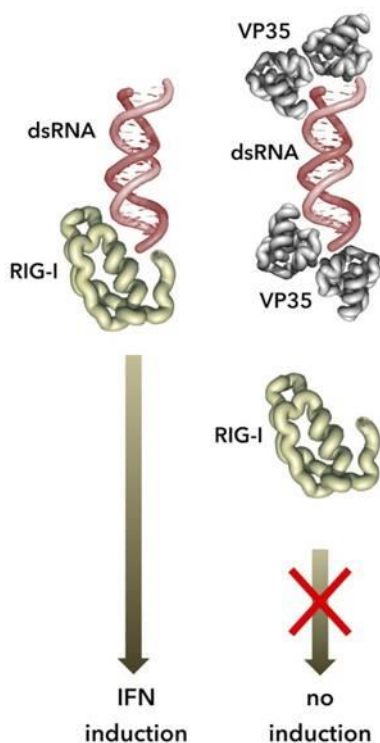
**Figure 2.** VP35 genome organization.

Once inside the cytoplasm of the host cell, the EBOV suppresses the host innate immune system via VP35 and VP24 proteins.<sup>15</sup>

VP35 is a multifunctional protein that is indispensable for EBOVs replication. In fact, VP35 is an essential cofactor of the EBOV RNA polymerase complex,<sup>19-21</sup> it serves as a nucleocapsid viral assembly factor,<sup>22,23</sup> it is a RNAi suppressor in mammalian cells.<sup>24,25</sup>



Furthermore, VP35 is directly involved in neutralizing the host innate immune response by blocking, at many different steps, the production of IFNs- $\alpha/\beta$ . For the latter, VP35 acts, at least in part by sequestering viral dsRNA from recognition by a set of cellular helicases termed RIG-I-like receptors (RLRs)(Figure 3).<sup>26</sup>



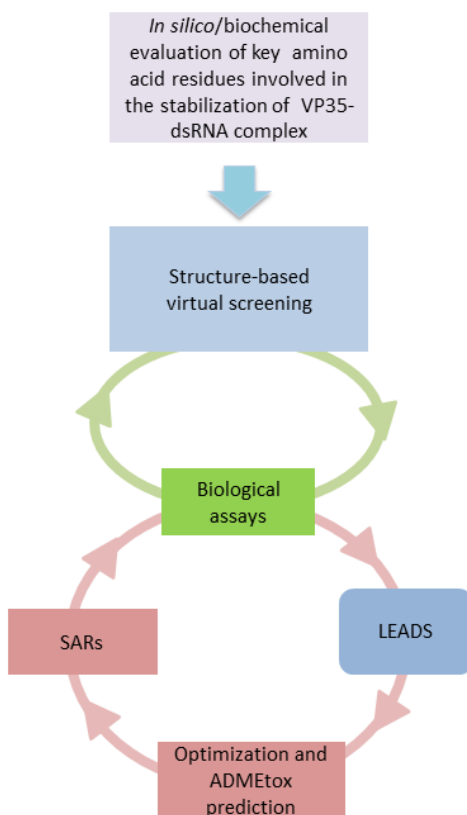
**Figure 3.** Schematic representation of the inhibition of Interferon induced by VP35: VP35 inhibits the binding of ds-RNA to RIG-I and by doing this, it prevents the induction of Interferon and consequently the immune response.

In fact, since cytoplasmic dsRNA has no homologues within cellular nucleic acids, RLRs recognize it as a signature of non-self, promptly activating a signalling cascade that culminates in the production of IFN- $\alpha/\beta$ .<sup>6</sup> However, sequestration of dsRNA by EBOV VP35 acts upstream by subverting the RLRs signalling pathway and ultimately suppressing host IFN- $\alpha/\beta$  production, and consequently immune system activation.<sup>26,6</sup>

Basic and hydrophobic amino acid residues involved in such interactions are highly conserved among all EBOVs and form a positively charged pocket at the VP35-dsRNA interface. Mutation into alanine of

critical residues within this domain, such as R312 and K339, resulted in the loss of VP35 dsRNA binding activity and loss of EBOV IFN suppression capabilities in cellular assays.<sup>26</sup> EBOV variants bearing these amino acid mutations failed to inhibit IFN- $\alpha/\beta$  production and exhibited no virulence in infected animals, further demonstrating the importance of VP35 dsRNA binding-mediated IFN-antagonism for EBOV pathogenesis.<sup>26</sup> Therefore, VP35 is a validated and attractive target for the development of strategies to counter EBOV diseases.

Hence, in order to define their impact on dsRNA end-capping and IFN-antagonism function, we performed here a structure-based alanine scanning mutagenesis of VP35 residues involved in this interaction. Data collected, together with computational studies, allowed the identification of a defined site important for prospective drug design against immune suppression function of EBOV VP35.



**Figure 4.** Project workflow.

## **4.2 DEFINITION KEY RESIDUES IN dsRNA RECOGNITION OF VP35**

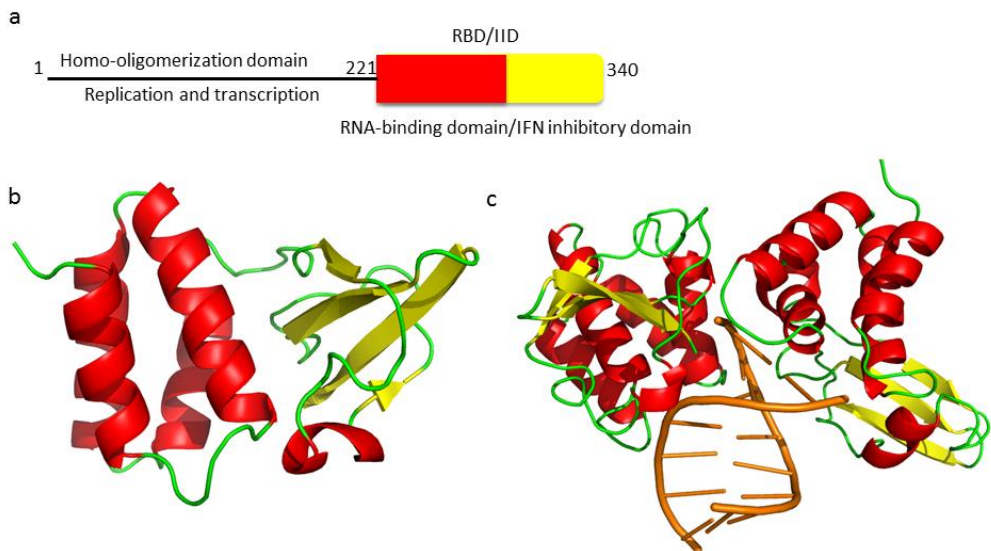
### **4.2.1 dsRNA binding property of Ebola VP35 end-capping mutants**

The analysis of the primary and secondary structure of VP35 has allowed to identify the existence of at least two different regions, an N-terminal region and a C-terminal region.

The N-terminal region comprises amino acids 1-221 and contains a domino coiled-coil involved in the process of homo-oligomerization of the

protein. It is important for viral replication and for the formation of the nucleocapsid (Figure 5a).

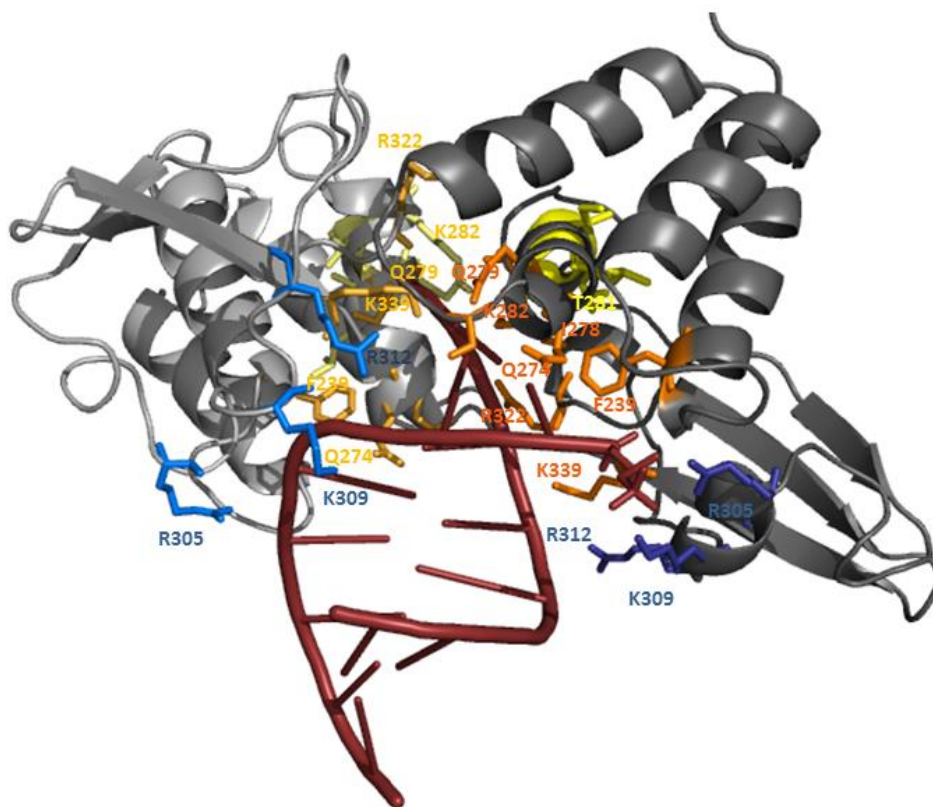
The C-terminal region corresponds instead to approximately one third of the entire amino acid sequence (residues 221-340) and is of great importance as it is responsible for the interaction with the protein NP and with dsRNA.



**Figure 5.** EBOV VP35 structure a) schematic representation of VP35 domains; b) VP35 RBD/IID crystal structure (pdb code 3FKE<sup>27</sup>) and asymmetric dimer of VP35 RBD/IID-dsRNA (pdb code 3L25<sup>28</sup>).

The crystallographic structures of EBOV and RESTV VP35 RBD/IID bound to blunt-ended dsRNA show a dimeric complex of two RBD/IID arranged to interact with the nucleic acid through asymmetric dimer with surfaces that describe a backbone-binding monomer and an end-capping one.<sup>29,26, 28</sup> A recent work described the kinetics in which dsRNA binding by the end-capping RBD/IID monomer constitutes the earliest binding event, to which attachment of backbone-binding RBD/IID monomers all along the dsRNA helix follows.<sup>30</sup> Within this picture, a small area interacting with dsRNA terminal bases in the concave surface of the end-capping RBD/IID monomer represents the most suitable target for small molecule inhibitors design. In fact, selected residues laying in this area, like F239, or surrounding it, such as R312, K319, R322 and K339 were found critical for RBD/IID dsRNA binding and IFN inhibition, as their mutation into alanine resulted in a decrease, or even a total loss, of those functions.<sup>26, 28, 31-37</sup>

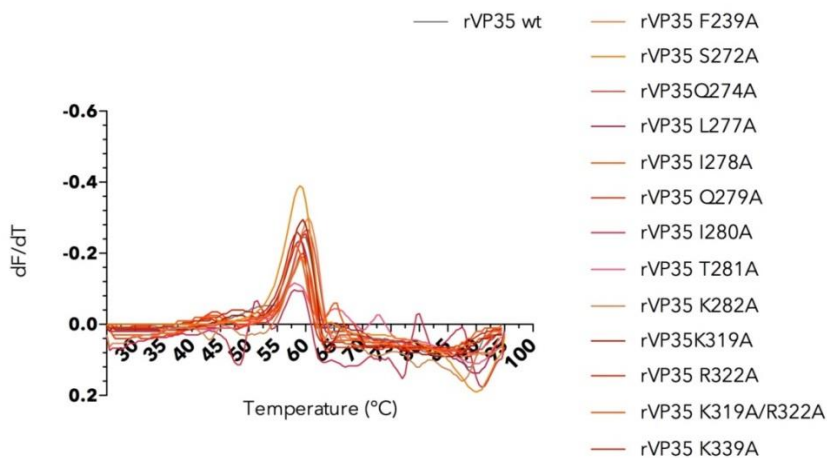
In our previous efforts to characterize a bacterially-expressed full length EBOV rVP35 (recombinant VP35),<sup>38</sup> we also investigated the impact of rVP35 mutants at residues involved in dsRNA backbone interactions, namely R305A, K309A and R312A, on dsRNA binding ability.<sup>37</sup> Here, we wanted to explore the role, relevance and potential effect of the residues involved in dsRNA end-capping and, to this aim, we performed alanine-scanning *in vitro* site-directed mutagenesis of rVP35 residues L277, I278, I280 (hydrophobic), S272, Q279, T281 (polar uncharged), K282 (polar charged) together with the previously studied F239, Q274, K319, R322, K319/R322 and K339 residues.<sup>26, 28, 31-37</sup>



**Figure 6.** *In vitro* site-directed mutagenesis: mutated residues are indicated in different colors: orange: end-capping residues monomer B, pale orange: corresponding backbone residues in monomer A; blue: charged residues interacting with last portion of dsRNA in monomer B; light blue: corresponding residues in monomer A involved in backbone binding ; yellow and pale yellow alpha helix cleft which include residues I278 and K282 that are involved in end capping interactions (L277-K282) respectively in monomer B and A.

Once all rVP35 mutants were expressed and purified as previously described,<sup>38</sup> to assess their proper folding and exclude that any difference in dsRNA binding might have been imputable to major structural perturbations, we performed a differential scanning fluorimetry (DSF) analysis comparing the thermal stability of full length wt and mutants rVP35. Interestingly, all rVP35 mutants showed melting temperature ( $T_m$ ) values comparable to wt rVP35 ( $59.4 \pm 0.6$  °C), which suggests maintenance of proper folding and no structural perturbations beyond minimal local changes (Table 1).

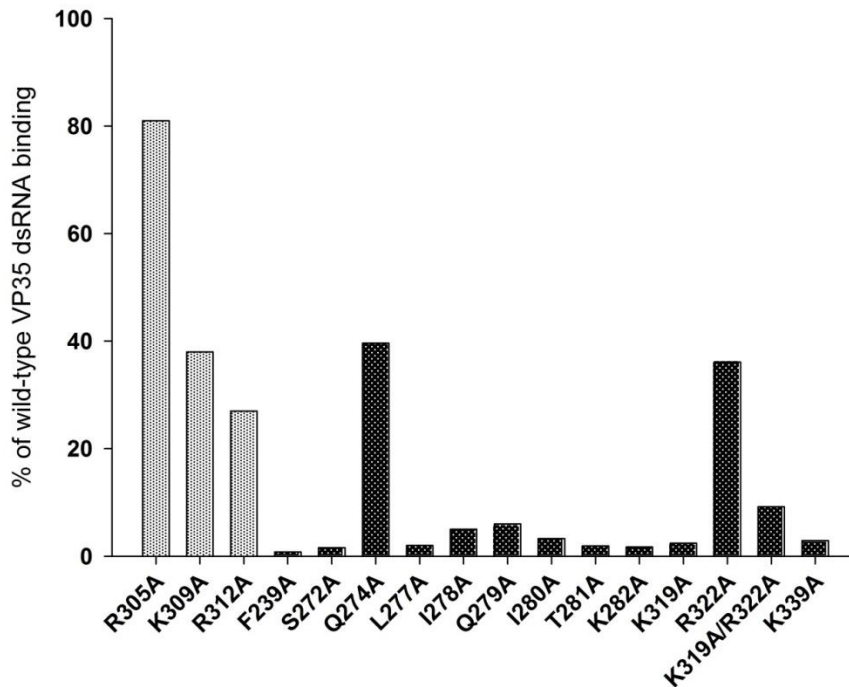
Secondly, by using a previously described magnetic pull down assay,<sup>37</sup> we assessed the dsRNA binding activity of rVP35 mutants through an homologous-competition binding curve titration with an *in vitro* transcribed (IVT) 500 bp dsRNA as ligand. Results showed that EBOV rVP35 end-capping mutants were impaired in dsRNA binding function, having no measurable  $K_d$  values or even poorly detectable activity with percentages ranging from 0.8% to 6% of bound dsRNA with respect to wt rVP35 (Figure 8).



**Figure 7.** Thermal stability of wt and mutants rVP35.

**Table 1.** Thermal stability of wt and mutants rVP35.  $T_m$  values obtained by DSF analysis on rVP35 wt and mutants. Data shown are the mean  $\pm$  SD of three independent experiments performed in triplicate samples.

VP35 protein	$T_m$ [°C]
wt	59.4 $\pm$ 0.6
F239A	59.7 $\pm$ 0.2
S272A	59.0 $\pm$ 0.4
Q274A	60.6 $\pm$ 0.6
L277A	58.8 $\pm$ 0.5
I278A	61.3 $\pm$ 1.2
Q279A	60.0 $\pm$ 0.6
I280A	56.5 $\pm$ 0.9
T281A	58.3 $\pm$ 0.6
K282A	59.4 $\pm$ 0.3
K319A	58.2 $\pm$ 0.6
R322A	60.3 $\pm$ 1.0
K319A/R322A	58.0 $\pm$ 0.4
K339A	59.2 $\pm$ 0.5



**Figure 8.** dsRNA binding ability of EBOV rVP35 backbone and end-capping mutants. Relative dsRNA binding activity of rVP35 backbone (white dotted) and end-capping (black dotted) mutants expressed as percentage of the total amount of dsRNA bound by wt rVP35

Only exceptions were, Q274A, R322A and K319A/R322A rVP35 mutants, which were able to bind the IVT 500 bp dsRNA even though with lower affinity (increased  $K_d$  values) with respect to wt rVP35. In particular,  $K_d$  values for wt, Q274A, R322A and K319A/R322A rVP35 mutants were  $2.6 \pm 0.3$  nM,  $5.23 \pm 1.8$  nM,  $9.13 \pm 1.5$  nM and  $15.62 \pm 2.0$  nM, respectively. According to these data, rVP35 dsRNA binding activity was highly compromised by the change of a single amino acid residue in the end-capping binding surface.

Utilizing for the first time a full-length VP35 protein, binding experiments performed with the magnetic pull down assay, they showed that minimal changes on the surface electrostatic charge in the protein pocket that houses the dsRNA terminus – such as those that may come from the alanine substitution of residues that interact with terminal bases or phosphate groups – may result in significant decrease, if not in total loss, of dsRNA binding activity in biochemical assay. In fact, while Q274 rVP35 mutant showed a 2-fold decreased binding affinity with respect to wt rVP35, for most of the tested end-capping mutants, including F239A, I278A, Q279A, I280A, T281A, K319A and K339A rVP35, the amount of bound dsRNA was barely detectable. Interestingly, and in substantial agreement with what was observed for the backbone binding R312A mutant, the absence of dsRNA binding activity previously reported for the R322A and K319A/R322A VP35 RBD<sup>26,39</sup> was partially restored in our full length rVP35 proteins that bear the same mutations, even though with affinities that were respectively 3-fold and 5-fold lower when compared to wt rVP35. Possibly, this greater efficiency in dsRNA binding displayed by full length VP35s with respect to the RBD alone may reflect the critical contribution of regions lacking in the sole RBD that might have a role in stabilizing or strengthening other domains of the protein during exertion of their function. Previous crystallographic studies on VP35 RBD mutants focused on the effects of mutations in charged residues, showing that the loss of affinity for dsRNA in these mutants is due to changes in the electrostatic surface potential.<sup>26</sup>



#### 4.2.2 Molecular modeling

In order to achieve a deeper understanding of the molecular basis of our observations, we integrated the dsRNA binding data with a computational study with the purpose of identifying a VP35 region suitable as a site for drug design. In particular we wanted to investigate the contribution of these mutations on the two VP35 monomeric binding surfaces (i.e backbone and capping), to select key residues involved in the EBOV VP35 dsRNA binding function. In fact VP35 dimer is asymmetric. Hence the question arisen is if it is possible to weight the importance of mutated residues in end-capping vs backbone ones. This enquiry could not be answered by the biochemical assay.

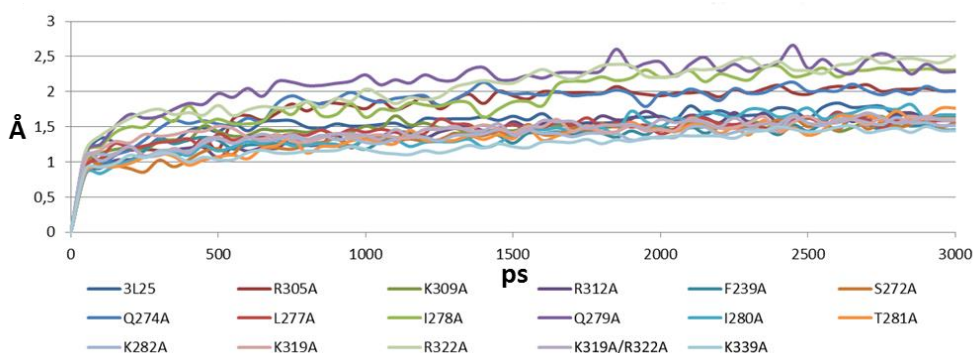
In recent years, molecular dynamic (MD) simulations combined with binding free energy calculations have been applied successfully to understand the behaviour of protein complex structures in solution (protein-ligand, protein-protein, protein-RNA interactions) and guide the drug design.<sup>40</sup>

In light of this, we performed computational studies to understand the molecular basis of the effects of selected VP35 amino acid mutations.

Firstly, we investigated the effect of 5'-ppp portion by adding the triphosphate moiety to the crystallographic model.<sup>26</sup> Once minimized, the complex VP35-dsRNA was subjected to MD simulation, finding the formation of a more stable complex as compared to the original one without 5'-ppp portion (-197 kcal/mol vs -175 kcal/mol). These results are in agreement with previous experimental studies that revealed the importance of 5'-ppp.<sup>6</sup> In fact, in the MD simulation the 5'-ppp appears to be stabilized by an array of hydrophobic, hydrogen bonds and ionic interactions. Once validated this modified model to properly assess EBOV VP35 selectivity for 5'-ppp-dsRNA versus dsRNA, we used it as starting point for all successive MD simulations where single amino acids (except double mutated K319A/R322A) have been mutated in alanine.

During MD simulations, the trajectories were monitored from the convergence of the Root-Mean Square deviation (RMSD) of all atoms in the EBOV VP35-dsRNA complex. All systems resulted stable during MD

simulation (Figure 9), with an overall RMSD fluctuating in the range of 0.5 nm, indicating that the system is indeed equilibrated.



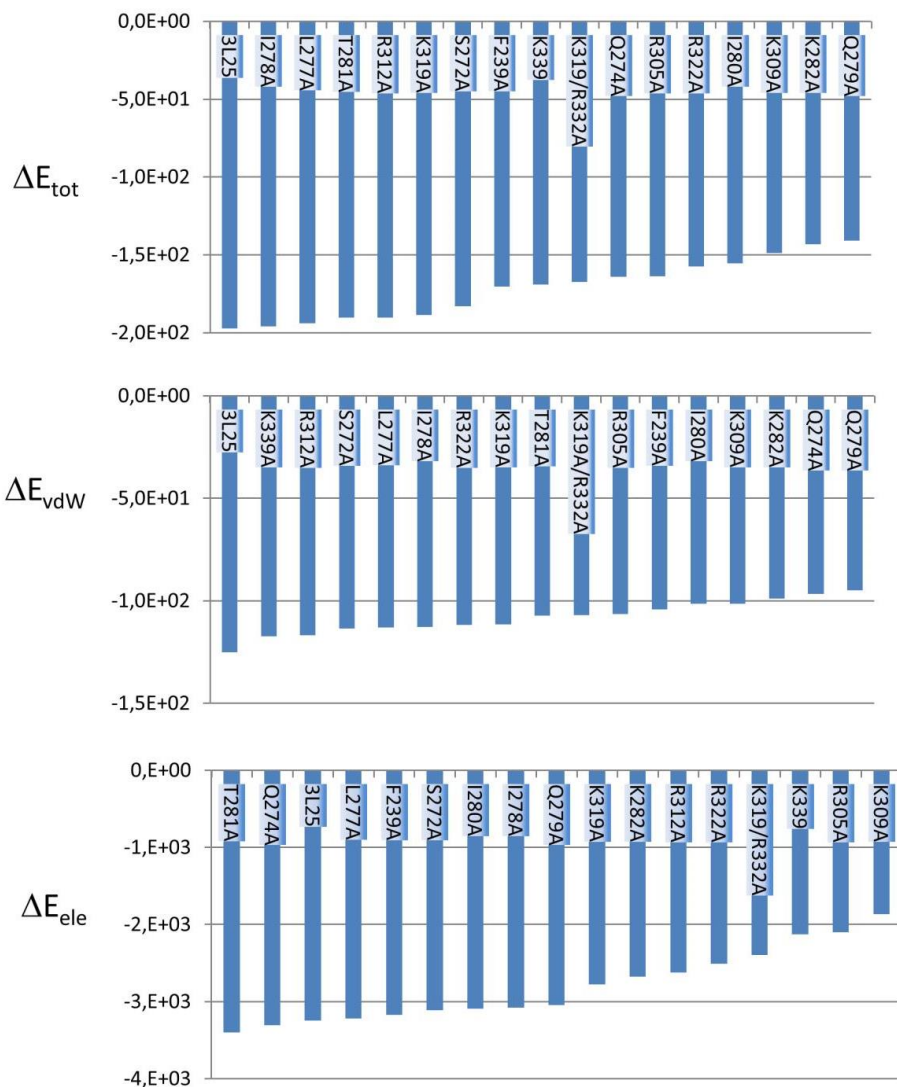
**Figure 9. MD simulations.** Root-Mean Square Deviation (RMSD) of VP35–dsRNA complexes during MD simulations. All systems resulted stable during MD simulation.

All mutants maintained the backbone conformation of wt EBOV VP35, free and in complex with dsRNA. This confirms the observations done by crystallographic experiments reported in PDB [R312A, PDB: 3L27; K339A, PDB:3L28, <sup>26</sup> K319A/R322A, PDB:3L29<sup>39</sup>]. For all VP35 mutants, MD analysis showed a decreased stability of the complexes, given the increased energy with respect to wt (Table 2).

**Table 2. Energetic stability of wt and mutants EBOV VP35-dsRNA complexes.** All mutations showed a decreased energetic stability of the complexes, given the energy increase with respect to wt.

	$\Delta E$ (kcal/mol)	$\Delta E_{mut} - \Delta E_{wt}$
Wild-type	-197,20	
I278A	-195,79	-1,41
L277A	-193,87	-3,32
T281A	-190,23	-6,96
R312A	-190,22	-6,98
K319A	-188,56	-8,64
S272A	-182,84	-14,36
F239A	-170,47	-26,73
K339	-168,92	-28,28
K319A/R332A	-167,48	-29,72
Q274A	-164,11	-33,09
R305A	-163,68	-33,52
R322A	-157,58	-39,62
I280A	-155,31	-41,89
K309A	-148,88	-48,32
K282A	-143,18	-54,02
Q279A	-140,81	-56,39

Decomposing the total energy interaction ( $E_{tot}$ ) into major components we found that electrostatic factor provides the main driving force of binding affinity (Figure 10). This could be easily interpreted since many of the mutated important residues have electrically charged or polar side chains. The only exceptions were F239, L277, I278A, I280A and T281A. These mutants were not associated to particular electrostatic changes and MD simulations showed only subtle differences in the complex side chain conformations.



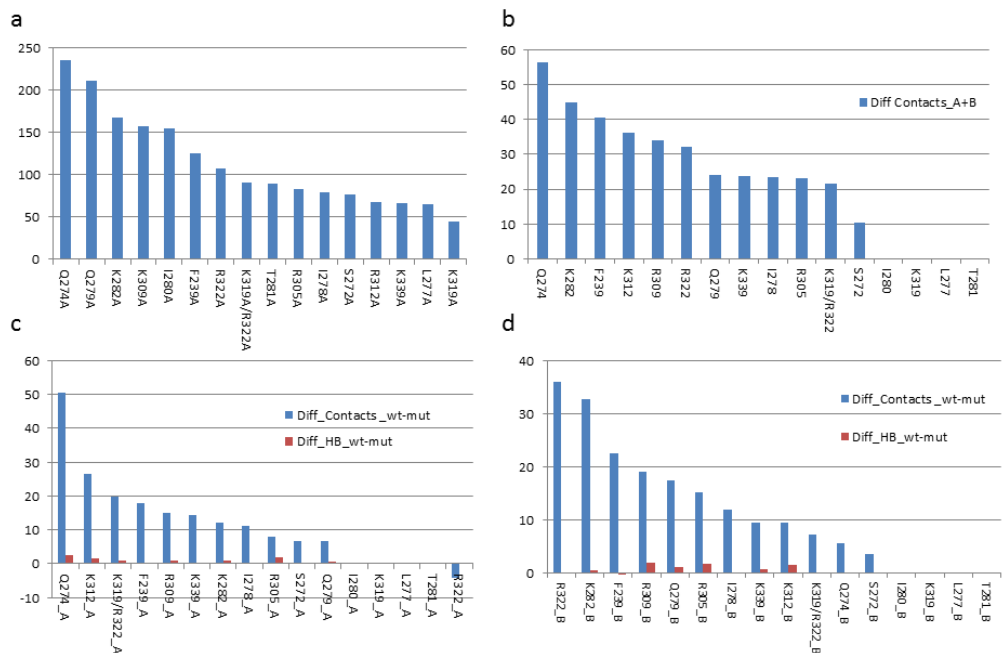
**Figure 10.** Calculated averages of  $\Delta E$  and its major components ( $\Delta E_{\text{vdw}}$  and  $\Delta E_{\text{ele}}$ ) expressed in kcal/mol of the wild type(3L25) and mutated VP35–dsRNA complexes, based on the frames of MD trajectories.

Therefore their importance should probably be attributed to the loss of contacts with dsRNA. Hence, we investigated *i)* the loss of contacts between the mutant VP35s and the whole dsRNA (Figure 11a), *ii)* the loss of contacts between the single mutated amino acid in both backbone binding monomer (chain A) and end-capping monomer (chain B) together (Figure

11b) and *iii*) for single chain A (Figure 11c) and chain B separately (Figure 11d). The analysis of the overall loss of contacts showed that all mutations are associated with fewer contacts between VP35 and the whole dsRNA (Figure 11a).

The decomposition analysis of contacts loss of backbone vs end-capping binding monomers did not show a univocal behaviour for all residues. However, for K282A and R322A VP35 mutants the loss of affinity for dsRNA is due exclusively to the contacts loss between these two residues in the end-capping binding monomer and dsRNA (Figure 11d), for other residues the loss of contacts is higher for end-capping monomer B: Q279A, R305A, F239A, S272A, I278A.

In addition, MD analysis showed that some residues that impaired VP35 affinity for dsRNA such as K319, I280 and T281 were not involved directly with interactions with dsRNA. However, their localization in chain B is closer to dsRNA with respect to chain A, therefore their mutations in the end-capping position appear to be more important than that in backbone-binding. Overall, we conclude that the effects of mutation of the B-chain residues involved in end-capping binding is more important than the effect of the same residues when they are involved in backbone binding.



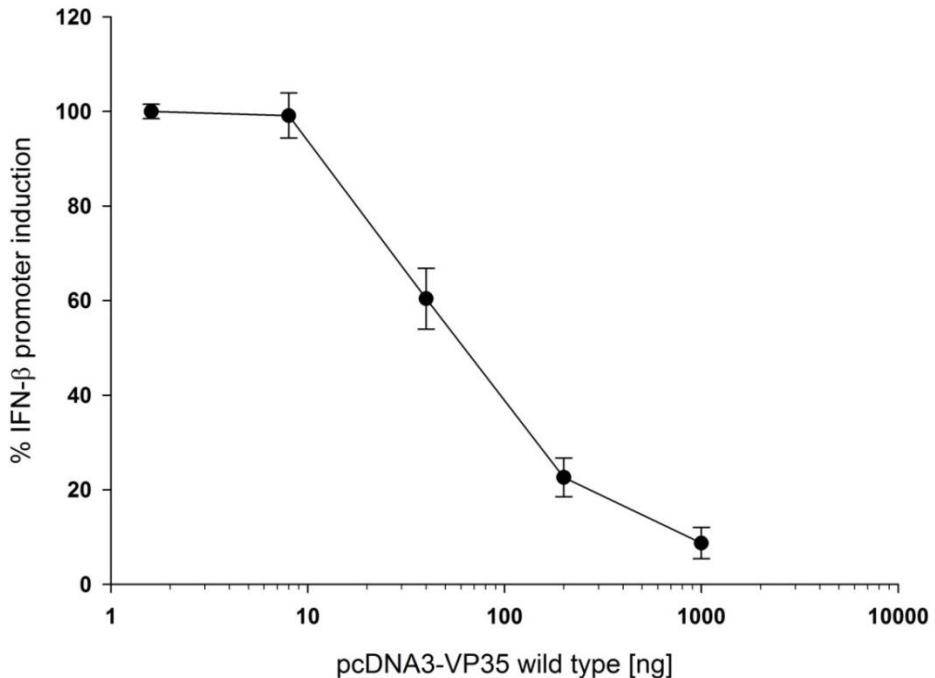
**Figure 11. Number of contacts loss between the mutated enzyme and dsRNA.** a) Loss of good contacts between VP35 RBD and the whole dsRNA; b) Loss of good contacts between the mutated residue (both chains A and B) and dsRNA; c) VP35 RBD loss of good contacts specific residues mutated in chain A (backbone-binding) and dsRNA; d) VP35 RBD loss of good contacts between specific residues mutated in chain B (end-capping-binding backbone-binding).

To summarize computational results, firstly we have confirmed the possibility of modeling the VP35 interaction with dsRNA by showing the importance of 5'-ppp portion for a more stable VP35:dsRNA complex compared to the crystal (without 5'-ppp) and, secondly, we have shown that - investigating the contacts loss between the mutant VP35s and the whole dsRNA - the effect upon mutation on the monomer involved in end-capping binding seems to be more important than the one involved in backbone binding monomer. Moreover, it was determined that the electrostatic component provides the main driving force of binding affinity. In fact, most of the important mutated residues have electrically charged or polar side chains. We can then define a region that could be targeted by small molecules delimited by hot spots (determinant) amino acids. But first we investigated the effect of mutants in RIG-I-mediated signaling cascade

measuring the IFN- $\beta$  inhibition to confirm the key residues also with cellular assay.

#### **4.2.3 IFN- $\beta$ inhibition property of EBOV VP35 mutants**

EBOV VP35 interferes at various levels the RIG-I-mediated signaling cascade that leads to the type I IFN production, ultimately allowing uncontrolled viral replication.<sup>32</sup> To assess the effect of end-capping mutations on the ability of EBOV VP35 to suppress type I IFN induction in cellular systems, a reporter gene assay was carried out that uses Influenza A virus (IAV) dsRNA transfection as a stimulus for RIG-I signaling pathway culminating in the activation of the IFN- $\beta$  promoter.<sup>41</sup> When A549 cells were co-transfected with IFN- $\beta$  reporter vector and increasing amounts of plasmid expressing EBOV VP35 wt or mutants, we observed a dose-dependent inhibition of the dsRNA-stimulated, RIG-I-mediated IFN- $\beta$  production by EBOV VP35 wt (Figure 12). Hence, we tested the IFN-antagonism ability of EBOV VP35 end-capping mutants transfecting the VP35 carrying plasmids at different concentration to assess the relative impairment of the mutant VP35s IFN-antagonism function. Since the assay herein used is different from those previously described, we firstly assayed VP35 mutants R305A, K309A and R312A to properly compare our results with those previously reported.<sup>32,34,35,27</sup>

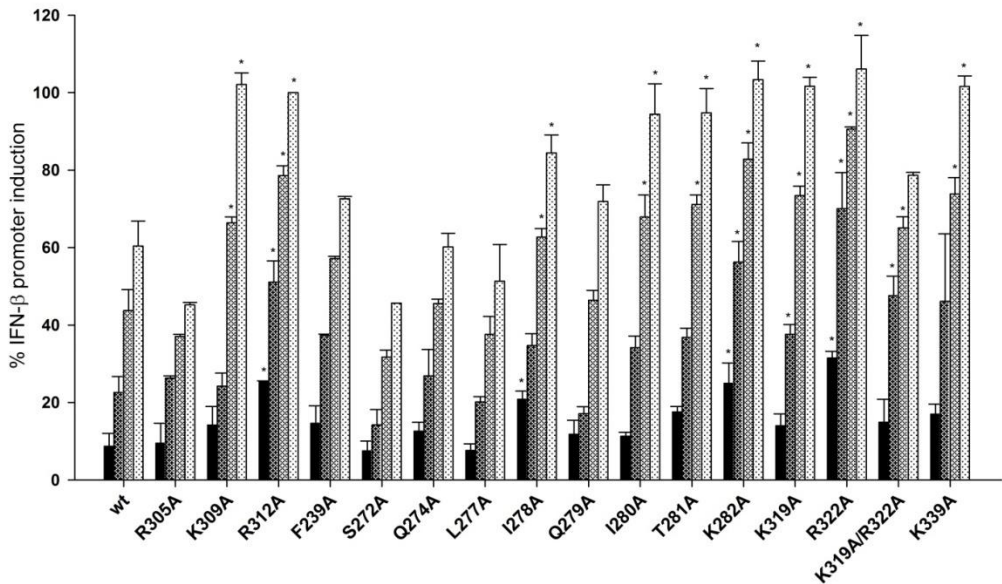


**Figure 12. The dose dependent inhibition of the vRNA RIG-I pathway activation by EBOV VP35 wild type.** A549 cells were co-transfected with 250 ng of pGL IFN-β luc and various amount (1.6, 8, 40, 200 and 1000 ng) of pcDNA3-ZEBOV-VP35 wild type. 24 hours post transfection cells were additionally transfected with 250 ng of IAV RNA and after additional 6 hours cells were lysed and luciferase activity was measured. Inhibition of luciferase expression was indicated as percentage of induced control. The error bars indicate standard deviation from three independent experiments.

The obtained results on these three mutants showed that K309A have a moderate and R312A a significant reduction in IFN-antagonist activity, while R305A does not affect this function, confirming previous reports (Figure 13).<sup>32,35</sup> Secondly, we tested all other VP35 mutants showing that F239A, S272A, Q274A, L277A and Q279A VP35 mutants substantially maintained an IFN-antagonist ability comparable to the one showed by wt VP35 (Figure 13). Differently, I278A, I280A, T281A, K319A, K339A mutants exhibited a strong reduction in IFN-antagonism ability. The reduction was statistically significant at lower VP35 mutant plasmid concentrations, suggesting that these VP35 mutants can compensate dsRNA binding affinity reduction at higher protein concentrations. Most relevantly, similarly to R312A VP35, K282A and R322A VP35s showed the strongest reduction in



VP35 IFN-antagonist ability, statistically significant even at the highest tested VP35 plasmid concentration (Figure 13).



**Figure 13. Luciferase reporter RIG-I-mediated IFN-β activation gene assay inhibition of EBOV VP35 wild type and mutants.** A549 cells were co-transfected with 250 ng of pGL IFN-β luc and 1000 (black), 200 (black dotted), 100 (grey dotted) and 40 ng (white dotted) of pcDNA3-ZEBOV-VP35 wild type and end-capping mutants. 24 hours post transfection cells were additionally transfected with 250 ng of IAV RNA and after additional 6 hours cells were lysed and luciferase activity was measured. Inhibition of luciferase expression was indicated as percentage of induced control. Data are average of three independent experiments, with error bars representing standard deviation (\* p value < 0.05).

An exception case is the one of K319A/R322A VP35 mutant that showed a reduction of the IFN-antagonist ability only at the intermediate plasmid concentrations. The dose dependent inhibition of the vRNA RIG-I pathway activation allowed to determine the plasmid concentration required to reduce the IFN-β promoter induction by 50% (IC<sub>50</sub>) value, for all tested VP35s. Compared to the wt VP35 IC<sub>50</sub> value of 110 ± 4.9 ng, I278A, I280A, T281A, K319A, and K339A VP35s exhibited a significant increase in their IC<sub>50</sub> values comprised in the range of 150-170 ng (p < 0.05), while R312A, K282A and R322A mutants exhibited higher IC<sub>50</sub> values (517 ± 28.2

ng,  $584 \pm 49.4$  ng and  $663 \pm 60.3$  ng, respectively, with  $p < 0.05$ ), demonstrating the criticality of these three single amino acid mutations for the inhibitory activity of the protein (Table 3). It is worth to note that 2 out of 3 VP35 mutants that showed the strongest reduction in IFN-antagonist efficacy in cell-based assays, R312A and R322A, maintained an although reduced capability of binding dsRNA in biochemical assays (Figure 8).

**Table 3.** Wt and mutant EBOV VP35s IFN- $\beta$  promoter induction inhibition

VP35 protein	<sup>a</sup> IC <sub>50</sub> [ng]
wt	110 ± 4,9
R305A	88 ± 15,5
K309A	146 ± 4,8
R312A	517 ± 28,2
F239A	143 ± 1,9
S272A	59 ± 8,1
Q274A	89 ± 1,8
L277A	85 ± 9,4
I278A	151 ± 0,5
Q279A	100 ± 8,2
I280A	157 ± 11,4
T281A	166 ± 6,2
K282A	584 ± 49,4
K319A	169 ± 6,3
R322A	663 ± 60,3
K319A/R322A	177 ± 16,8
K339A	173 ± 17,3

In contrast, other VP35 mutants such as F239A, S272A, Q274A, L277A, I278A, I280A, T281A, K319A, and K339A that showed no or limited reduction of their IFN-antagonist efficacy in cell-based assays almost completely lost their dsRNA binding ability in biochemical assays (Figure 8). A number of reasons could be involved to explain these observations: *i*) some of the mutated VP35 amino acid residues, such as F239, Q274, I278 and K339, participate in both end-capping and backbone-binding VP35 monomers; *ii*) it is not clear at the moment if the magnetic binding assay with rVP35 involves monomeric, dimeric, multimeric (or a mixture of them) VP35 forms, and whether these forms have different responses to alanine scanning mutagenesis; *iii*) some of the mutated VP35 amino acid residues participate

in both protein–RNA and protein–protein interactions with cellular proteins involved in the RIG-I pathway and the loss of one of the two functions is not sufficient to impair the VP35 inhibition.

Comparing the biochemical and cellular data, it is clear that the loss of dsRNA binding capacity by some mutants is not necessarily manifested in a loss of VP35 IFN-antagonism. In fact, it is known that VP35 also interacts with PACT (PKR activator), leading to the inhibition of PACT-induced RIG-I ATPase activity,<sup>42</sup> it interacts with the TBK-1/IKK- $\epsilon$  kinases complex suppressing the activation of IRF-3,<sup>43</sup> and it was reported that, when the IFN- $\beta$  production is activated by a dsRNA-independent mechanism, the mutated R312A, R322A and K339A VP35 showed reduced suppression of IFN- $\beta$  production.<sup>32,26</sup> On the contrary, the mutated F239A VP35, which lost the dsRNA-binding function, preserved the dsRNA-independent IFN-inhibitory property.<sup>26</sup> These results suggest that the loss of the dsRNA binding ability by some amino acid substitutions may not impair the protein-protein interactions among others components of RIG-I signalling pathway, which explains the discrepancy between our biochemical and cellular data. Overall, therefore, we can classify these VP35 mutants in 3 categories:

- i)* VP35 mutants that show a reduced dsRNA binding ability in biochemical assay but retain IFN-antagonism in cell culture assay such as F239A, S272A, Q274A, L277A and Q279A, probably retaining a dsRNA-binding independent inhibition mechanism;
- ii)* VP35 mutants that show a reduced dsRNA binding ability in biochemical assay as well as a reduced IFN-antagonism in cell culture assay at lower level of protein expression but they retain IFN-antagonism functions at higher level of protein expression, such as I278A, I280A, T281A, K309A, K319A, K319A/R322A and K339A, suggesting that they can compensate dsRNA binding affinity reduction at higher protein concentrations;
- iii)* VP35 mutants that show a reduced dsRNA binding ability in biochemical assay and show a reduced IFN-antagonism in cell culture assay even at higher level of protein expression such

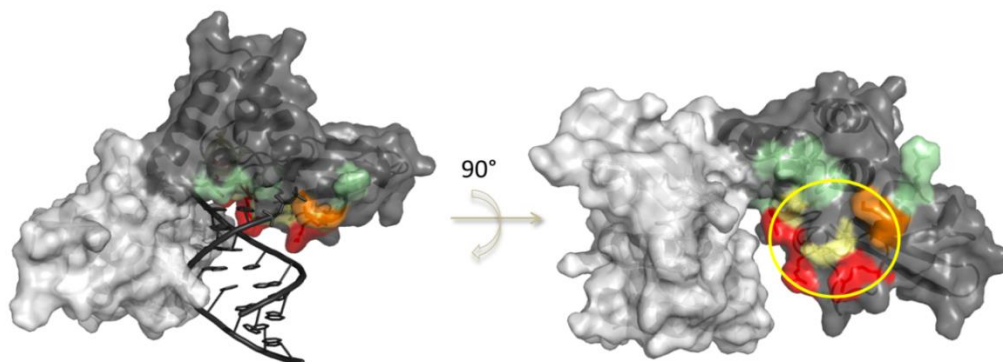
as R312A, K282A and R322A, proving to be involved also in dsRNA-binding independent inhibition mechanism.

In any case, the overall biochemical and cellular analysis of the results showed that a number of the selected VP35 amino acid residues such as I278A, I280A, T281A, K319A, K319A/R322A and K339A, and in particular R312A, K282A and R322A, significantly reduced VP35 ability to inhibit vRNA induced RIG-I pathway activation.

#### 4.2.4 Conclusions

The multifunctional viral protein EBOV VP35 is a validate drug target since it is one of the most potent weapons that EBOV uses to evade the innate immune antiviral response.<sup>44</sup> In the context of EBOV infection, loss of VP35 dsRNA binding and IFN-antagonist functions results in severely impaired virus replication in cells capable of mounting an IFN- $\alpha/\beta$  response and also fully attenuates the virus *in vivo*.<sup>33,39</sup> Moreover, point mutations in VP35 that impaired dsRNA binding abolished VP35–PACT interaction and VP35 became unable to disrupt PACT–RIG-I interaction. This generated a loss of VP35 inhibition of PACT-facilitated activation of RIG-I.<sup>44,42</sup>

Previous studies reported that some residues comprised in the end-capping domain have critical roles in dsRNA binding wt EBOV VP35<sup>26,27</sup> and mutational studies have confirmed their implication in loss of dsRNA binding ability<sup>32,27,26,36,37</sup> as well as in suppression of IFN- $\beta$  induction.<sup>32,33,34,26,27,36</sup> Overall, the set of data obtained with mutational studies with the three methods (biochemical, computational and cellular) allowed us to highlight the importance of some residues involved in end-capping binding and are localized in a delimited area (Figure 14), where the most important residues appear to be the R312, K282 and R322. Noteworthy, the overall data analysis suggests that this site could be involved not only in VP35 dsRNA binding but also in VP35 interaction with other cellular proteins involved in the RIG-I pathway. Therefore, this site seems to be an attractive target to develop drugs that can impede the VP35 impairment of the IFN activation and could be used in future studies in order to find small molecules able to inhibit EBOV VP35.



**Figure 14. Hot spots of VP35 surface.** Residues are coloured in order of importance: red (R312, K282 and R322) > orange (I278, I280, T281, and K339) > yellow (K309 and K319) > green (R305, F239, S272, Q274, L277 and Q279) according to p value. It is possible to highlight a binding site useful for drug design delimited by yellow circle.

### 4.3 VIRTUAL SCREENING

VP35 is a validate drug target and we have already remarked the strategic role in evading the innate immune antiviral response.<sup>44</sup> Thus the inhibition of VP35 could lead to restore the ability of human immunity system to fight against EVD. Probably this would also means to reduce the mortality caused by EBOV infection.

Hence there is an urgent need of weapons to win against this target.

Therefore we carried out a Virtual screening aiming at finding drug candidates able to inhibit VP35 and in particular the formation of the complex with dsRNA.

Recently, another group published a virtual screening targeting VP35 but the protein-protein interaction with NP protein. In this work some compounds were found to inhibit VP35–NP interaction with affinities (Kd ) <100  $\mu$ M.<sup>45, 46</sup>

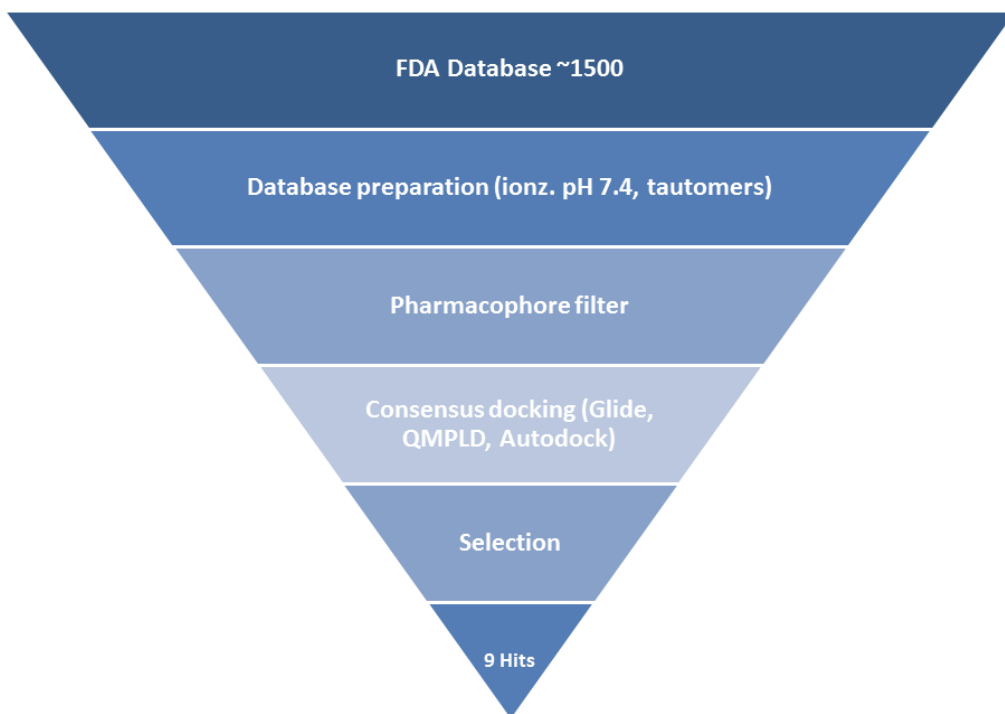
Our strategy was different. Previous studies highlighted that end-capping RBD/IID monomer constitutes the earliest binding event, to which attachment of backbone-binding RBD/IID monomer follows.<sup>30</sup> Furthermore in the previous paragraph we have studied several mutations and, with the support of computational studies, we have demonstrated the importance of end-capping residues (Figure 14).

The emergence of the outbreak registered in several countries of West Africa last year obligates a rapid response, thus we decided to screen the FDA approved drugs Database.<sup>47</sup> Drug repositioning is an interesting approach with more and more success stories available in literature.<sup>48</sup> In fact, the finding that many drugs interact with more than one target provided the rationale behind the selective optimization of side activities (SOSA) approach recently developed.<sup>49, 50</sup>

First we applied a pharmacophore filter by means of LigandScout program.<sup>51, 52</sup> This approach, as it was highlighted in the first chapter offers several advantages and it is a fast tool useful to perform rapid and reliable VS.<sup>53, 54</sup> Furthermore in literature there are plenty of success stories where structure-based pharmacophore was applied alone or in combination.<sup>55-61</sup> Since we aim to find molecules that bind the end-capping region to avoid the dsRNA binding, structure-based pharmacophore was generated considering the terminal portion of dsRNA as “core molecule”. In particular the lowest energetic frame of MD simulation of wt 5'-pppVP35-dsRNA complex was selected as starting structure. The pharmacophore was applied to screen the prepared FDA Database containing all the approved drugs available in the market. Selected compounds were then minimized and docked considering three different protocols: GlideXP,<sup>62</sup> QMPLD<sup>63, 64</sup> and Autodock.<sup>65, 66</sup> Best pose of each compound was then subjected to a post-docking procedure considering eMBrAcE.<sup>67</sup> In this way we have rescored the docking poses obtained with different docking programs using the same procedure. eMBrAcE applies multiple minimizations, during which each of the specified pre-positioned ligand is minimized with the receptor. By calculating a consensus score we ranked the compounds and selected 9 hits. These were purchased and tested in cells assay for their ability to inhibit VP35.

Preliminary data show that two of the compounds are active in the low micromolar range.

Due to the novelty of this target and the importance of the results obtained we cannot show their structures. In fact, up to now, these compounds are the only ones acting with this mechanism and showing to inhibit VP35 at low concentration. For the same reason the virtual screening is not described in details.



**Figure 15.** VS workflow

#### **4.3.1 Conclusions**

In conclusion we have demonstrated the ability of the combination of two structure-based approaches in selecting good hits. We are going to perform other biological tests to confirm the activity and the mechanism.

We will then perform a further selection of similar compounds to confirm the importance of the scaffolds and to derive structure activity relationships to guide the optimization process. Starting from approved drugs the process for their development should be easier and less expensive.

## **4.4 EXPERIMENTAL PART**

### **4.4.1 EBOV VP35 wt and mutant plasmids**

Cloning of the EBOV VP35 gene into pET45b(+) vector (Novagen) to obtain the pET45b-ZEBOV-VP35 plasmid was previously described.<sup>38</sup> The QuickChange II Site-Directed Mutagenesis kit (Agilent technologies) was used to obtain corresponding mutant plasmids by introducing in the EBOV cDNA VP35 gene sequence previously described single point mutations R305A, K309A, R312A<sup>37</sup> and new mutations F239A, S272A, Q274A, L277A, I278A, Q279A, I280A, T281A, K282A, K319A, R322A, K319A/R322A and K339A.

### **4.4.2 EBOV VP35 wt and mutants mammalian expression plasmid**

To introduce its sequence into a mammalian expression vector, we amplified cDNA of the EBOV VP35 gene (GenBank: NC) previously cloned in the pET45b-ZEBOV-VP35 vector<sup>38</sup> (Zinzula et al., 2009), by polymerase chain reaction (PCR). Primers were designed to amplify the gene and subclone it into the pcDNA3 mammalian expression plasmid (Invitrogen) by BamHI and NotI restriction enzymes. Primer sequences were: 5'-TCAGCAGAGGATCCGATAATGCATCACCACCACCATCAC-3' and 5'-GTACTAATATGCGGCCGCTCAAATTTTGAGTCCAAGTGT-3'. PCR reaction was carried out in a mixture containing: pET45b-ZEBOV-VP35 plasmid (100 ng), each primer (400 µM), MgCl<sub>2</sub> (1.5 mM), each dNTP (0.2 mM) and 0.025 µl FidelityTaq™ DNA Polymerase (Usb). PCR mixtures were filled with nuclease-free water to a final volume of 50 µL and PCR cycle consisted of: an initial denaturation at 94 °C for 2 min, 35 cycles of denaturation at 94 °C for 30 sec, annealing at 55 °C for 30 sec, extension at 68 °C for 2 min, and a final extension at 68 °C for 5 min. The amplified EBOV VP35 gene and the pcDNA3 plasmid (Invitrogen) were digested by BamHI and NotI restriction enzymes (New England BioLabs), linear pcDNA3 plasmid was dephosphorylated with Antarctic Phosphatase (New England BioLabs) and fragments (50 ng of linear pcDNA3 and 20 ng of EBOV VP35 insert) were



ligated by T4 DNA ligase (New England BioLabs) to obtain the pcDNA3-EBOV-VP35 plasmid, which was used to transform *E. coli* TOP 10 cells (Invitrogen) by standard heat shock protocol at 42 °C for 90 sec. Plasmid was extracted and sequenced for control. The same procedure was used to produce the mutants pcDNA3-EBOV-VP35/R305A, pcDNA3-EBOV-VP35/K309A, pcDNA3-EBOV-VP35/R312A, pcDNA3-EBOV-VP35/K319A/R322A, pcDNA3-EBOV-VP35/S272A, pcDNA3-EBOV-VP35/Q274A, pcDNA3-EBOV-VP35/F239A, pcDNA3-EBOV-VP35/L277A, pcDNA3-EBOV-VP35/I280A, pcDNA3-EBOV-VP35/I278A, pcDNA3-EBOV-VP35/Q279A, pcDNA3-EBOV-VP35/T281A, pcDNA3-EBOV-VP35/K282A, pcDNA3-EBOV-VP35/K319A, pcDNA3-EBOV-VP35/R322A and pcDNA3-EBOV-VP35/K339A.

#### **4.4.3 Expression and purification of full-length wt and mutants EBOV rVP35**

Full-length, wt and mutants, bacterially-expressed rVP35s were obtained as previously described.<sup>38</sup> Briefly, protein expression was carried out on transformed *E. coli* BL21AI (Invitrogen) cultured in LB media at 37 °C and induced at an optical density of 0.6 OD at 600 nm with 0.4% L-arabinose (Sigma-Aldrich). rVP35s were IMAC purified with Ni-Sepharose High Performance (GE Healthcare) beads by using a BioLogic LP FPLC system (Biorad), and dialyzed in desalting buffer (50 mM sodium phosphate pH 7.5, 300 mM NaCl, 10% glycerol, 0.014%  $\beta$ -mercaptoethanol). As previously published,<sup>38</sup> rVP35 full-length proteins were purified at  $\approx$ 95% homogeneity, their integrity was assessed by PAGE analysis and their concentration was calculated using the Protein Quantification kit-Rapid (Fluka).

#### **4.4.4 Differential scanning fluorimetry analysis**

Differential scanning fluorimetry (DSF) analysis was carried out in a MiniOpticon real-time PCR instrument (BioRad) by using the Protein Thermal Shift Dye kit (Life technologies) according to manufacturer's instructions. Measurements were performed using excitation  $\lambda = 470$ -505

nm and emission  $\lambda = 540\text{-}700$  nm and data were acquired on a temperature gradient from 25 °C to 95 °C with increments of 0.5 °C. Samples contained 2  $\mu\text{g}$  rVP35 protein, 1X SYPRO Orange (Life technologies), 50 mM sodium phosphate pH 7.5, 150 mM NaCl and 20 mM  $\text{MgCl}_2$  in a 20  $\mu\text{l}$  final volume. Fluorescence data were analysed, and the derivative of the curve representing melting temperature ( $T_m$ ) of wild type and mutants rVP35 was obtained by using the CFX manager software v.2.1 (BioRad) tool for protein thermal shift assays protocols.

#### **4.4.5 Magnetic pull down assay**

Heterologous 500 bp dsRNA and 500 bp radio-labeled dsRNA production was previously described.<sup>37</sup> The rVP35-dsRNA complex formation was assessed exploiting the properties of the TALON paramagnetic Dynabeads (Invitrogen). Firstly, 1  $\mu\text{g}$  of rVP35 was conjugated to 50  $\mu\text{L}$  TALON beads in a volume of 700  $\mu\text{L}$  of binding/washing buffer (50 mM sodium phosphate pH 7.5, 150 mM NaCl, 0.05% Tween-20) for 15 min at 23 °C under gentle rotating agitation (20 rpm). Unconjugated rVP35 was removed by magnetic field application, supernatant removal and further washing with binding/washing buffer. Pellets with conjugated rVP35 were re-suspended in a 100  $\mu\text{l}$  volume of binding/washing buffer containing 20 mM  $\text{MgCl}_2$  and 1.5 nM 500 bp  $^3\text{H}$ -dsRNA (0.1 Ci/mmoles), then incubated for 60 min at 37 °C (20 rpm). Unbound  $^3\text{H}$ -dsRNA was separated by the conjugated rVP35-dsRNA complex by magnetic field application and supernatant removal. A further washing step was performed to completely remove unbound  $^3\text{H}$ -dsRNA. Elution of VP35-dsRNA elution was performed by incubation of the beads pellet in 300  $\mu\text{l}$  elution buffer (binding/washing buffer plus 1 M imidazole pH 7.5) for 10 min at 23 °C (20 rpm), subsequent magnetic field application and supernatant removal. The supernatant was transferred to vials and radioactivity was determined with a Beckman LS 6500 beta-counter (Beckman-Coulter).

#### 4.4.6 Molecular modeling

**Complex preparation.** The coordinates of VP35-dsRNA homodimer structure were taken from the RCSB Protein Data Bank (PDB ID: 3L25).<sup>26</sup> The protein was prepared by using the Maestro GUI Protein Preparation Wizard module (Schrödinger LLC. 2014. Maestro GUI, New York, NY, USA). Bond orders and formal charges were added for hetero groups, and all the hydrogen atoms were added in the structure. After preparation, the structures were refined to optimize the hydrogen-bond network by using OPLS2005 force field.<sup>68</sup> The minimization was terminated once the energy converged or the RMSD reached a maximum cut off of 0.30Å. The wt model was generated introducing 5'-ppp dsRNA as terminal portion using build module in Maestro GUI available in Schroedinger Suite 2014 (Schrödinger LLC. 2014. Maestro GUI, New York, NY, USA). The protein was then minimized with the Polak-Ribiere conjugate gradient minimization allowing 10000 iterations and a convergence threshold of 0.01 in GB/SA implicit water.<sup>69</sup> Since the model misses the coil-coil tail portion and dsRNA is likely to be open during the simulation, force constraints were applied around residue Asp218 and dsRNA terminal bases (C1-G8). The rest of the complex was left free to move. Minimized complex was then used as initial structure for MD simulation, where the same constraints were maintained. Mutants were generated based on wt structure by mutating the corresponding residues.

**Molecular dynamic simulations and MMGBSA calculations.** MacroModel of Schrödinger Suite Software<sup>70</sup> was applied to carry out molecular dynamic simulations (3ns) including the energy minimization and equilibration protocols (200ps). The production phase runs 3ns. The binding free energy calculations were performed by MM/GBSA<sup>71</sup> method using extracted frames from equilibrated trajectories. Binding free energy was computed:  $\Delta G_{\text{bind}} = G_{\text{complex}} - G_{\text{protein}} - G_{\text{RNA}}$ . The free energy G, can be calculated using the scheme as follows:  $\Delta G = \Delta E_{\text{MM}} + \Delta G_{\text{sol}} - T\Delta S$ ,  $\Delta G_{\text{sol}}$  and  $\Delta S$  are considered approximately similar for all similar complex therefore  $\Delta G$  is proportional to  $\Delta E$ :  $\Delta E_{\text{MM}} = \Delta E_{\text{int}} + \Delta E_{\text{ele}} + \Delta E_{\text{vdw}}$ . The average energies of MD simulation are listed in Table 4 while Figure 10 shows  $\Delta E$  and the most important contributions to the energy of binding: electrostatic ( $\Delta E_{\text{ele}}$ ) and

van der Walls ( $\Delta E_{vdw}$ ). The resulting complexes were considered for the binding modes graphical analysis with Maestro (Schrödinger LLC. 2014. Maestro GUI, New York, NY, USA) and Pymol.<sup>72</sup>

## VIRTUAL SCREENING

**Pharmacophore generation.** The frame with the lowest energy of wt 3P-VP35 coming from the MD was considered for pharmacophore generation. Only end capping interacting monomer was selected. The two terminal bases and the triphosphate portion were considered as core molecule and the pharmacophore was automatically generated by Ligandscout. This original pharmacophore was modified with the introduction of hydrophobic feature in correspondence of Guanine aromatic portion and used to screen the database allowing 2 omitting features.

**Database preparation.** FDA Drugs .sdf file was downloaded from IPMC (Institut de Pharmacologie Moléculaire et Cellulaire).<sup>47</sup> This file contains one conformer for each drug with indication of chiral centres or list of possible diastereoisomers if the drug possess more than one chiral center and it is given as a mixture.

Compounds were prepared with Ligprep using the right ionization at 7.4 and allowing possible tautomers.<sup>73, 74</sup> This file was then converted in a multiconformer .ldb database with Ldbgen by LigandScout and screened with the pharmacophore.

**Docking.** Selected compounds were then minimized and docked considering 3 docking protocols: GlideXP,<sup>62</sup> QMPL<sup>63</sup> and Autodock.<sup>65, 66</sup>

Grid was centred considering the residue F239 located in the middle of end-capping surface and it was wide enough to contain the whole monomer (36 x36 x36)Å. For Glide and QPLD, extra-precision (XP) mode was selected. All the other settings were left as default as well as for Autodock, where the files were generated with Raccoon.<sup>75</sup>

The best poses of each compound were then subjected to a post-dock eMBrAcE protocol.<sup>67</sup> For each ligand, the protein-ligand complex, the free protein, and the free ligand were all subjected to energy minimization in implicit solvent (generalized Born).<sup>76</sup> It uses a traditional molecular mechanics (MM) method to calculate ligand–receptor interaction energies, with a Gaussian smooth dielectric constant function method for

electrostatic part of solvation energy and solvent-accessible surface for the nonpolar part of solvation energy. A conjugate gradient minimization protocol was used in all the performed minimizations. The compounds with the best consensus scoring were then analysed by visual inspection. 9 compounds were purchased.

#### **4.4.7 Cell line**

A549 cells were grown in Dulbecco's modified Eagle's medium (DMEM) (Gibco) supplemented with 10% fetal bovine serum (FBS) (Gibco), 1% Penicillin-Streptomycin (Pen/Strept) (EuroClone). Cells were incubated at 37 °C in a humidified 5% CO<sub>2</sub> atmosphere

#### **4.4.8 EBOV VP35 luciferase reporter gene inhibition assay**

A549 cells ( $5 \times 10^4$  per well) were co-transfected in 48-well plates with T-Pro P-Fect Transfection Reagent (T-Pro Biotechnology) with the pGL IFN- $\beta$  luc plasmid, kindly provided by Prof. Stephan Ludwig (Institute of Molecular Virology, Münster, Germany), and the pcDNA3-ZEBOV-VP35. Twenty-four hours after transfection cells were additionally transfected with Influenza A Virus/Puerto-Rico/8/34 viral RNA (IAV PR8 vRNA) and incubated for further 6 hours at 37 °C with 5% CO<sub>2</sub>. Cells were harvested with lysis buffer (50 mM Na-MES pH 7.8, 50 mM Tris-HCl pH 7.8, 1 mM dithiothreitol, 0.2% Triton X-100). The crude cell lysates were cleared by centrifugation and 50  $\mu$ L of cleared lysates were added to 50  $\mu$ L of luciferase assay buffer (125 mM Na-MES pH 7.8, 125 mM Tris- HCl pH 7.8, 25 mM magnesium acetate, 2.5 mg/ml ATP) in a white 96-well plate. Immediately after addition of 50  $\mu$ L 1 mM D-luciferin (Gold Biotechnology) into each well luminescence was measured in a Victor<sup>3</sup> luminometer (Perkin Elmer). The relative light units (RLU) were normalized as the fold activity of the unstimulated control. Inhibition of luciferase expression was indicated as percentage of induced control; each assay was carried out in triplicate.

## 4.5 REFERENCES

1. <http://www.cdc.gov/vhf/ebola/outbreaks/2014-west-africa/case-counts.html>.
2. Barrette, R. W.; Xu, L.; Rowland, J. M.; McIntosh, M. T., Current perspectives on the phylogeny of Filoviridae. *Infect. Genet. Evol.* **2011**, 11, 1514-9.
3. Kuhn, J. H.; Becker, S.; Ebihara, H.; Geisbert, T. W.; Johnson, K. M.; Kawaoka, Y.; Lipkin, W. I.; Negrodo, A. I.; Netesov, S. V.; Nichol, S. T.; Palacios, G.; Peters, C. J.; Tenorio, A.; Volchkov, V. E.; Jahrling, P. B., Proposal for a revised taxonomy of the family Filoviridae: classification, names of taxa and viruses, and virus abbreviations. *Arch. Virol.* **2010**, 155, 2083-103.
4. Feldmann, H.; Geisbert, T. W., Ebola haemorrhagic fever. *Lancet* **2011**, 377, 849-62.
5. Ansari, A. A., Clinical features and pathobiology of Ebolavirus infection. *J. Autoimmun.* **2014**, 55, 1-9.
6. Zinzula, L.; Tramontano, E., Strategies of highly pathogenic RNA viruses to block dsRNA detection by RIG-I-like receptors: hide, mask, hit. *Antiviral Res.* **2013**, 100, 615-35.
7. Gire, S. K.; Goba, A.; Andersen, K. G.; Sealfon, R. S.; Park, D. J.; Kanneh, L.; Jalloh, S.; Momoh, M.; Fullah, M.; Dudas, G.; Wohl, S.; Moses, L. M.; Yozwiak, N. L.; Winnicki, S.; Matranga, C. B.; Malboeuf, C. M.; Qu, J.; Gladden, A. D.; Schaffner, S. F.; Yang, X.; Jiang, P. P.; Nekoui, M.; Colubri, A.; Coomber, M. R.; Fonnies, M.; Moigboi, A.; Gbakie, M.; Kamara, F. K.; Tucker, V.; Konuwa, E.; Saffa, S.; Sellu, J.; Jalloh, A. A.; Kovoma, A.; Koninga, J.; Mustapha, I.; Kargbo, K.; Foday, M.; Yillah, M.; Kanneh, F.; Robert, W.; Massally, J. L.; Chapman, S. B.; Bochicchio, J.; Murphy, C.; Nusbaum, C.;

Young, S.; Birren, B. W.; Grant, D. S.; Scheiffelin, J. S.; Lander, E. S.; Happi, C.; Gevao, S. M.; Gnirke, A.; Rambaut, A.; Garry, R. F.; Khan, S. H.; Sabeti, P. C., Genomic surveillance elucidates Ebola virus origin and transmission during the 2014 outbreak. *Science* **2014**, 345, 1369-72.

8. Kibuuka, H.; Berkowitz, N. M.; Millard, M.; Enama, M. E.; Tindikahwa, A.; Sekiziyivu, A. B.; Costner, P.; Sitar, S.; Glover, D.; Hu, Z.; Joshi, G.; Stanley, D.; Kunchai, M.; Eller, L. A.; Bailer, R. T.; Koup, R. A.; Nabel, G. J.; Mascola, J. R.; Sullivan, N. J.; Graham, B. S.; Roederer, M.; Michael, N. L.; Robb, M. L.; Ledgerwood, J. E., Safety and immunogenicity of Ebola virus and Marburg virus glycoprotein DNA vaccines assessed separately and concomitantly in healthy Ugandan adults: a phase 1b, randomised, double-blind, placebo-controlled clinical trial. *Lancet* **2015**, 385, 1545-54.

9. Ledgerwood, J. E.; DeZure, A. D.; Stanley, D. A.; Novik, L.; Enama, M. E.; Berkowitz, N. M.; Hu, Z.; Joshi, G.; Ploquin, A.; Sitar, S.; Gordon, I. J.; Plummer, S. A.; Holman, L. A.; Hendel, C. S.; Yamshchikov, G.; Roman, F.; Nicosia, A.; Colloca, S.; Cortese, R.; Bailer, R. T.; Schwartz, R. M.; Roederer, M.; Mascola, J. R.; Koup, R. A.; Sullivan, N. J.; Graham, B. S., Chimpanzee Adenovirus Vector Ebola Vaccine - Preliminary Report. *N. Engl. J. Med.* **2014**.

10. Rampling, T.; Ewer, K.; Bowyer, G.; Wright, D.; Imoukhuede, E. B.; Payne, R.; Hartnell, F.; Gibani, M.; Bliss, C.; Minhinnick, A.; Wilkie, M.; Venkatraman, N.; Poulton, I.; Lella, N.; Roberts, R.; Sierra-Davidson, K.; Krahling, V.; Berrie, E.; Roman, F.; De Ryck, I.; Nicosia, A.; Sullivan, N. J.; Stanley, D. A.; Ledgerwood, J. E.; Schwartz, R. M.; Siani, L.; Colloca, S.; Folgori, A.; Di Marco, S.; Cortese, R.; Becker, S.; Graham, B. S.; Koup, R. A.; Levine, M. M.; Moorthy, V.; Pollard, A. J.; Draper, S. J.; Ballou, W. R.; Lawrie, A.; Gilbert, S. C.; Hill, A. V., A Monovalent Chimpanzee Adenovirus Ebola Vaccine - Preliminary Report. *N. Engl. J. Med.* **2015**.

11. Sarwar, U. N.; Costner, P.; Enama, M. E.; Berkowitz, N.; Hu, Z.; Hendel, C. S.; Sitar, S.; Plummer, S.; Mulangu, S.; Bailer, R. T.; Koup, R. A.; Mascola, J. R.; Nabel, G. J.; Sullivan, N. J.; Graham, B. S.; Ledgerwood, J. E., Safety and immunogenicity of DNA vaccines encoding Ebola virus and Marburgvirus wild-type glycoproteins in a phase I clinical trial. *J. Infect. Dis.* **2015**, 211, 549-57.
12. Bishop, B. M., Potential and emerging treatment options for Ebola virus disease. *Ann. Pharmacother.* **2015**, 49, 196-206.
13. Becquart, P.; Wauquier, N.; Mahlakoiv, T.; Nkoghe, D.; Padilla, C.; Souris, M.; Ollomo, B.; Gonzalez, J. P.; De Lamballerie, X.; Kazanji, M.; Leroy, E. M., High prevalence of both humoral and cellular immunity to Zaire ebolavirus among rural populations in Gabon. *PLoS One* **2010**, 5, e9126.
14. Wauquier, N.; Becquart, P.; Gasquet, C.; Leroy, E. M., Immunoglobulin G in Ebola outbreak survivors, Gabon. *Emerg. Infect. Dis.* **2009**, 15, 1136-7.
15. Basler, C. F.; Amarasinghe, G. K., Evasion of interferon responses by Ebola and Marburg viruses. *J. Interferon Cytokine Res.* **2009**, 29, 511-20.
16. Mohamadzadeh, M., Potential factors induced by filoviruses that lead to immune suppression. *Curr. Mol. Med.* **2009**, 9, 174-85.
17. Ascenzi, P.; Bocedi, A.; Heptonstall, J.; Capobianchi, M. R.; Di Caro, A.; Mastrangelo, E.; Bolognesi, M.; Ippolito, G., Ebola virus and Marburgvirus: Insight the Filoviridae family. *Mol. Asp. Med.* **2008**, 29, 151-185.
18. Lai, K. Y.; Ng, W. Y. G.; Cheng, F. F., Human Ebola virus infection in West Africa: a review of available therapeutic agents that target different steps of the life cycle of Ebola virus. *Infectious Diseases of Poverty* **2014**, 3, 43.



19. Boehmann, Y.; Enterlein, S.; Randolph, A.; Muhlberger, E., A reconstituted replication and transcription system for Ebola virus Reston and comparison with Ebola virus Zaire. *Virology* **2005**, 332, 406-17.
20. Muhlberger, E.; Lotfering, B.; Klenk, H. D.; Becker, S., Three of the four nucleocapsid proteins of Marburg virus, NP, VP35, and L, are sufficient to mediate replication and transcription of Marburg virus-specific monocistronic minigenomes. *J. Virol.* **1998**, 72, 8756-64.
21. Muhlberger, E.; Weik, M.; Volchkov, V. E.; Klenk, H. D.; Becker, S., Comparison of the transcription and replication strategies of marburg virus and Ebola virus by using artificial replication systems. *J. Virol.* **1999**, 73, 2333-42.
22. Huang, Y.; Xu, L.; Sun, Y.; Nabel, G. J., The assembly of Ebola virus nucleocapsid requires virion-associated proteins 35 and 24 and posttranslational modification of nucleoprotein. *Mol. Cell* **2002**, 10, 307-16.
23. Johnson, R. F.; McCarthy, S. E.; Godlewski, P. J.; Harty, R. N., Ebola virus VP35-VP40 interaction is sufficient for packaging 3E-5E minigenome RNA into virus-like particles. *J. Virol.* **2006**, 80, 5135-44.
24. Haasnoot, J.; de Vries, W.; Geutjes, E. J.; Prins, M.; de Haan, P.; Berkhout, B., The Ebola virus VP35 protein is a suppressor of RNA silencing. *PLoS Pathog.* **2007**, 3, e86.
25. Fabozzi, G.; Nabel, C. S.; Dolan, M. A.; Sullivan, N. J., Ebolavirus proteins suppress the effects of small interfering RNA by direct interaction with the mammalian RNA interference pathway. *J. Virol.* **2011**, 85, 2512-23.
26. Leung, D. W.; Prins, K. C.; Basler, C. F.; Amarasinghe, G. K., Ebolavirus VP35 is a multifunctional virulence factor. *Virulence* **2010**, 1, 526-31.

27. Leung, D. W.; Ginder, N. D.; Fulton, D. B.; Nix, J.; Basler, C. F.; Honzatko, R. B.; Amarasinghe, G. K., Structure of the Ebola VP35 interferon inhibitory domain. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, 106, 411-6.
28. Leung, D. W.; Prins, K. C.; Borek, D. M.; Farahbakhsh, M.; Tufariello, J. M.; Ramanan, P.; Nix, J. C.; Helgeson, L. A.; Otwinowski, Z.; Honzatko, R. B.; Basler, C. F.; Amarasinghe, G. K., Structural basis for dsRNA recognition and interferon antagonism by Ebola VP35. *Nat. Struct. Mol. Biol.* **2010**, 17, 165-172.
29. Kimberlin, C. R.; Bornholdt, Z. A.; Li, S.; Woods, V. L., Jr.; MacRae, I. J.; Saphire, E. O., Ebolavirus VP35 uses a bimodal strategy to bind dsRNA for innate immune suppression. *Proc. Natl. Acad. Sci. U. S. A.* **2010**, 107, 314-9.
30. Bale, S.; Julien, J. P.; Bornholdt, Z. A.; Krois, A. S.; Wilson, I. A.; Saphire, E. O., Ebolavirus VP35 coats the backbone of double-stranded RNA for interferon antagonism. *J. Virol.* **2013**, 87, 10385-8.
31. Bale, S.; Julien, J. P.; Bornholdt, Z. A.; Kimberlin, C. R.; Halfmann, P.; Zandonatti, M. A.; Kunert, J.; Kroon, G. J.; Kawaoka, Y.; MacRae, I. J.; Wilson, I. A.; Saphire, E. O., Marburg virus VP35 can both fully coat the backbone and cap the ends of dsRNA for interferon antagonism. *PLoS Pathog.* **2012**, 8, e1002916.
32. Cardenas, W. B.; Loo, Y. M.; Gale, M., Jr.; Hartman, A. L.; Kimberlin, C. R.; Martinez-Sobrido, L.; Saphire, E. O.; Basler, C. F., Ebola virus VP35 protein binds double-stranded RNA and inhibits alpha/beta interferon production induced by RIG-I signaling. *J. Virol.* **2006**, 80, 5168-78.
33. Hartman, A. L.; Bird, B. H.; Towner, J. S.; Antoniadou, Z. A.; Zaki, S. R.; Nichol, S. T., Inhibition of IRF-3 activation by VP35 is critical for the high level of virulence of ebola virus. *J. Virol.* **2008**, 82, 2699-704.

34. Hartman, A. L.; Ling, L.; Nichol, S. T.; Hibberd, M. L., Whole-genome expression profiling reveals that inhibition of host innate immune response pathways by Ebola virus can be reversed by a single amino acid change in the VP35 protein. *J. Virol.* **2008**, *82*, 5348-58.
35. Hartman, A. L.; Towner, J. S.; Nichol, S. T., A C-terminal basic amino acid motif of Zaire ebolavirus VP35 is essential for type I interferon antagonism and displays high identity with the RNA-binding domain of another interferon antagonist, the NS1 protein of influenza A virus. *Virology* **2004**, *328*, 177-84.
36. Prins, K. C.; Binning, J. M.; Shabman, R. S.; Leung, D. W.; Amarasinghe, G. K.; Basler, C. F., Basic residues within the ebolavirus VP35 protein are required for its viral polymerase cofactor function. *J. Virol.* **2010**, *84*, 10581-91.
37. Zinzula, L.; Esposito, F.; Pala, D.; Tramontano, E., dsRNA binding characterization of full length recombinant wild type and mutants Zaire ebolavirus VP35. *Antiviral Res.* **2012**, *93*, 354-63.
38. Zinzula, L.; Esposito, F.; Muhlberger, E.; Trunschke, M.; Conrad, D.; Piano, D.; Tramontano, E., Purification and functional characterization of the full length recombinant Ebola virus VP35 protein expressed in *E. coli*. *Protein Expr. Purif.* **2009**, *66*, 113-9.
39. Prins, K. C.; Delpeut, S.; Leung, D. W.; Reynard, O.; Volchkova, V. A.; Reid, S. P.; Ramanan, P.; Cardenas, W. B.; Amarasinghe, G. K.; Volchkov, V. E.; Basler, C. F., Mutations abrogating VP35 interaction with double-stranded RNA render Ebola virus avirulent in guinea pigs. *J. Virol.* **2010**, *84*, 3004-15.
40. Liu, H.; Yao, X., Molecular basis of the interaction for an essential subunit PA-PB1 in influenza virus RNA polymerase: insights from molecular

dynamics simulation and free energy calculation. *Mol. Pharm.* **2010**, *7*, 75-85.

41. Cannas, V.; Daino, G. L.; Corona, A.; Esposito, F.; Tramontano, E., A Luciferase Reporter Gene Assay to Measure Ebola Virus Viral Protein 35-Associated Inhibition of Double-Stranded RNA-Stimulated, Retinoic Acid-Inducible Gene 1-Mediated Induction of Interferon beta. *J. Infect. Dis.* **2015**, *212* Suppl 2, S277-81.

42. Luthra, P.; Ramanan, P.; Mire, C. E.; Weisend, C.; Tsuda, Y.; Yen, B.; Liu, G.; Leung, D. W.; Geisbert, T. W.; Ebihara, H.; Amarasinghe, G. K.; Basler, C. F., Mutual antagonism between the Ebola virus VP35 protein and the RIG-I activator PACT determines infection outcome. *Cell Host Microbe* **2013**, *14*, 74-84.

43. Prins, K. C.; Cardenas, W. B.; Basler, C. F., Ebola virus protein VP35 impairs the function of interferon regulatory factor-activating kinases IKKepsilon and TBK-1. *J. Virol.* **2009**, *83*, 3069-77.

44. Basler, C. F., Innate immune evasion by filoviruses. *Virology* **2015**, *479-480*, 122-30.

45. Brown, C. S.; Lee, M. S.; Leung, D. W.; Wang, T.; Xu, W.; Luthra, P.; Anantpadma, M.; Shabman, R. S.; Melito, L. M.; MacMillan, K. S.; Borek, D. M.; Otwinowski, Z.; Ramanan, P.; Stubbs, A. J.; Peterson, D. S.; Binning, J. M.; Tonelli, M.; Olson, M. A.; Davey, R. A.; Ready, J. M.; Basler, C. F.; Amarasinghe, G. K., In Silico Derived Small Molecules Bind the Filovirus VP35 Protein and Inhibit Its Polymerase Cofactor Activity. *J. Mol. Biol.* **2014**, *426*, 2045-2058.

46. Dapiaggi, F.; Pieraccini, S.; Sironi, M., In silico study of VP35 inhibitors: from computational alanine scanning to essential dynamics. *Mol. BioSyst.* **2015**, *11*, 2152-2157.

47. Pihan, E.; Colliandre, L.; Guichou, J.-F.; Douguet, D., e-Drug3D: 3D structure collections dedicated to drug repurposing and fragment-based drug design. *Bioinformatics* **2012**, *28*, 1540-1541.
48. Ashburn, T. T.; Thor, K. B., Drug repositioning: identifying and developing new uses for existing drugs. *Nat. Rev. Drug Discov.* **2004**, *3*, 673-683.
49. Wermuth, C. G., Selective optimization of side activities: the SOSA approach. *Drug Discov. Today* **2006**, *11*, 160-4.
50. Langer, T.; Wermuth, C.-G. Selective Optimization of Side Activities (SOSA): A Promising way for Drug Discovery. In *Polypharmacology in Drug Discovery*; John Wiley & Sons, Inc.: 2012, pp 227-243.
51. IntelLigand Software GmbH *LigandScout 3.0*, Maria Enzersdorf, Austria.
52. Wolber, G.; Langer, T., LigandScout: 3-D Pharmacophores Derived from Protein-Bound Ligands and Their Use as Virtual Screening Filters. *J. Chem. Inf. Model.* **2005**, *45*, 160-169.
53. Langer, T., Pharmacophores in drug research. *Molecular Informatics* **2010**, *29*, 470-475.
54. Mannhold, R.; Kubinyi, H.; Folkers, G.; Langer, T.; Hoffmann, R. D., *Pharmacophores and pharmacophore searches*. John Wiley & Sons: 2006; Vol. 32.
55. Berry, M.; Fielding, B. C.; Gamielien, J., Potential Broad Spectrum Inhibitors of the Coronavirus 3CLpro: A Virtual Screening and Structure-Based Drug Design Study. *Viruses* **2015**, *7*, 6642-60.
56. Markt, P.; Feldmann, C.; Rollinger, J. M.; Raduner, S.; Schuster, D.; Kirchmair, J.; Distinto, S.; Spitzer, G. M.; Wolber, G.; Laggner, C.; Altmann, K. H.; Langer, T.; Gertsch, J., Discovery of novel CB2 receptor ligands by a

pharmacophore-based virtual screening workflow. *J. Med. Chem.* **2009**, *52*, 369-78.

57. Markt, P.; McGoohan, C.; Walker, B.; Kirchmair, J.; Feldmann, C.; De Martino, G.; Spitzer, G.; Distinto, S.; Schuster, D.; Wolber, G.; Laggner, C.; Langer, T., Discovery of novel cathepsin S inhibitors by pharmacophore-based virtual high-throughput screening. *J. Chem. Inf. Model.* **2008**, *48*, 1693-705.

58. Markt, P.; Petersen, R. K.; Flindt, E. N.; Kristiansen, K.; Kirchmair, J.; Spitzer, G.; Distinto, S.; Schuster, D.; Wolber, G.; Laggner, C.; Langer, T., Discovery of novel PPAR ligands by a virtual screening approach based on pharmacophore modeling, 3D shape, and electrostatic similarity screening. *J. Med. Chem.* **2008**, *51*, 6303-17.

59. Polishchuk, P. G.; Samoylenko, G. V.; Khristova, T. M.; Krysko, O. L.; Kabanova, T. A.; Kabanov, V. M.; Korniylov, A. Y.; Klimchuk, O.; Langer, T.; Andronati, S. A.; Kuz'min, V. E.; Krysko, A. A.; Varnek, A., Design, Virtual Screening, and Synthesis of Antagonists of  $\alpha$ IIb $\beta$ 3 as Antiplatelet Agents. *J. Med. Chem.* **2015**, *58*, 7681-94.

60. Schuster, D.; Kowalik, D.; Kirchmair, J.; Laggner, C.; Markt, P.; Aebischer-Gumy, C.; Strohle, F.; Moller, G.; Wolber, G.; Wilckens, T.; Langer, T.; Odermatt, A.; Adamski, J., Identification of chemically diverse, novel inhibitors of 17 $\beta$ -hydroxysteroid dehydrogenase type 3 and 5 by pharmacophore-based virtual screening. *J. Steroid Biochem. Mol. Biol.* **2011**, *125*, 148-61.

61. Sengupta, S.; Roy, D.; Bandyopadhyay, S., Structural insight into Mycobacterium tuberculosis maltosyl transferase inhibitors: pharmacophore-based virtual screening, docking, and molecular dynamics simulations. *J. Biomol. Struct. Dyn.* **2015**, *33*, 2655-66.

62. Friesner, R. A.; Murphy, R. B.; Repasky, M. P.; Frye, L. L.; Greenwood, J. R.; Halgren, T. A.; Sanschagrin, P. C.; Mainz, D. T., Extra Precision Glide: Docking and Scoring Incorporating a Model of Hydrophobic Enclosure for Protein–Ligand Complexes. *J. Med. Chem.* **2006**, *49*, 6177-6196.
63. Chung, J. Y.; Hah, J.-M.; Cho, A. E., Correlation between Performance of QM/MM Docking and Simple Classification of Binding Sites. *J. Chem. Inf. Model.* **2009**, *49*, 2382-2387.
64. Schrödinger LLC. *QMPolarized protocol*, 2012; New York, NY, USA.
65. Morris, G. M.; Goodsell, D. S.; Halliday, R. S.; Huey, R.; Hart, W. E.; Belew, R. K.; Olson, A. J., Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function. *J. Comput. Chem.* **1998**, *19*, 1639-1662.
66. Morris, G. M.; Huey, R.; Lindstrom, W.; Sanner, M. F.; Belew, R. K.; Goodsell, D. S.; Olson, A. J., AutoDock4 and AutoDockTools4: Automated Docking with Selective Receptor Flexibility. *J. Comput. Chem.* **2009**, *30*, 2785-2791.
67. Guvench, O.; Weiser, J.; Shenkin, P.; Kolossváry, I.; Still, W. C., Application of the frozen atom approximation to the GB/SA continuum model for solvation free energy. *J. Comput. Chem.* **2002**, *23*, 214-221.
68. Banks, J. L.; Beard, H. S.; Cao, Y.; Cho, A. E.; Damm, W.; Farid, R.; Felts, A. K.; Halgren, T. A.; Mainz, D. T.; Maple, J. R.; Murphy, R.; Philipp, D. M.; Repasky, M. P.; Zhang, L. Y.; Berne, B. J.; Friesner, R. A.; Gallicchio, E.; Levy, R. M., Integrated Modeling Program, Applied Chemical Theory (IMPACT). *J. Comput. Chem.* **2005**, *26*, 1752-80.
69. Still, W. C.; Tempczyk, A.; Hawley, R. C.; Hendrickson, T., Semianalytical treatment of solvation for molecular mechanics and dynamics. *J. Am. Chem. Soc.* **1990**, *112*, 6127-6129.

70. Mohamadi, F.; Richards, N. G. J.; Guida, W. C.; Liskamp, R.; Lipton, M.; Caufield, C.; Chang, G.; Hendrickson, T.; Still, W. C., Macromodel—an integrated software system for modeling organic and bioorganic molecules using molecular mechanics. *J. Comput. Chem.* **1990**, *11*, 440-467.
71. Kollman, P. A.; Massova, I.; Reyes, C.; Kuhn, B.; Huo, S.; Chong, L.; Lee, M.; Lee, T.; Duan, Y.; Wang, W.; Donini, O.; Cieplak, P.; Srinivasan, J.; Case, D. A.; Cheatham, T. E., 3rd, Calculating structures and free energies of complex molecules: combining molecular mechanics and continuum models. *Acc. Chem. Res.* **2000**, *33*, 889-97.
72. Delano, W. L. The PyMOL Molecular Graphics System. <http://www.pymol.org>
73. Greenwood, J. R.; Calkins, D.; Sullivan, A. P.; Shelley, J. C., Towards the comprehensive, rapid, and accurate prediction of the favorable tautomeric states of drug-like molecules in aqueous solution. *J. Comput. Aided Mol. Des.* **2010**, *24*, 591-604.
74. Shelley, J. C.; Cholleti, A.; Frye, L. L.; Greenwood, J. R.; Timlin, M. R.; Uchimaya, M., Epik: a software program for pK( a ) prediction and protonation state generation for drug-like molecules. *J. Comput. Aided Mol. Des.* **2007**, *21*, 681-91.
75. Forli, S., Raccoon|AutoDock VS: an automated tool for preparing AutoDock virtual screenings <http://autodock.scripps.edu/resources/raccoon>
76. Qiu, D.; Shenkin, P. S.; Hollinger, F. P.; Still, W. C., The GB/SA Continuum Model for Solvation. A Fast Analytical Method for the Calculation of Approximate Born Radii. *The Journal of Physical Chemistry A* **1997**, *101*, 3005-3014.



## 5 CHAPTER

### 5.1 INTRODUCTION TO COVALENT DOCKING WITH AUTODOCK

The aim of this work was to apply and compare these two methods of covalent docking developed in Olson's laboratory at the Scripps Research Institute, in order to find the best procedure to be implemented in Autodock platform.

There has recently been a resurgence of interest in inhibitors that bind covalently to their biomolecular targets.<sup>1-3</sup> These compounds have the advantage of very tight binding, allowing to design of compounds with small molecular mass but with high potency. Potential problems with selectivity have been a concern for development of covalent inhibitors. However one third of currently-approved drugs act through covalent mechanisms. The major approach to reducing toxicity is to improve the selectivity of the compounds, both by optimizing the non-covalent interactions within the binding site and by tailoring the chemistry of reaction with a specific site of alkylation.

Computational docking provides an effective way to evaluate the interaction of a trial compound with a target. With the recent increased interest in covalent inhibitors, a variety of methods have been reported for using computational docking to predict the binding of covalently-bound compounds.<sup>4-7</sup> In general, these methods search the conformations available to the ligand in its covalently-attached state, and evaluate the energetics of interaction with the target binding site. More detailed methods, such as quantum mechanical analysis, complement these docking analyses to address the chemical reactivity of the compound.

AutoDock is an automated procedure for predicting the interactions of ligands with macromolecular targets. It is a two steps process where first the maps and the interactions of each atom type of the ligand and the receptor are pre-calculated in a grid surrounding the binding region. Then the docking simulation takes place, in which grid interactions energies are used as look-up table to speed ligand energy evaluation. Autodock uses a Lamarckian genetic algorithm where a population of trial conformations is created, and

then in successive generations these individuals mutate, exchange conformational parameters, and compete in a manner analogous to biological evolution, ultimately selecting individuals with lowest binding energy.<sup>8,9</sup>

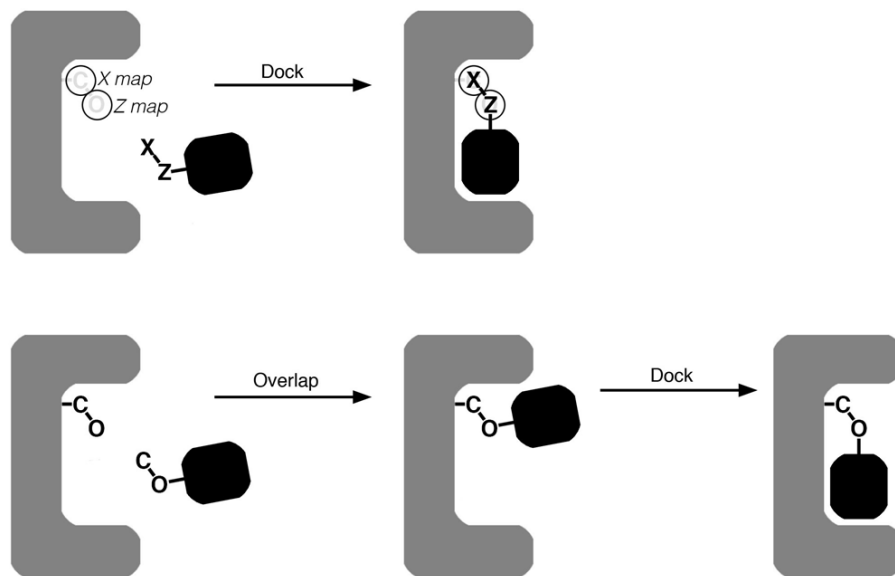
We have studied two different covalent methods to modify the distributed version of AutoDock<sup>9</sup> to evaluate covalent complexes, by extending its standard functionality with custom potentials and new atom types. The specification and architecture of AutoDock was designed to allow this type of user modification for specialized systems, and has been used previously to model protein flexibility,<sup>10</sup> flexible rings in ligands,<sup>11</sup> zinc interaction,<sup>12</sup> and selective hydration.<sup>13</sup>

The need to compute automated docking for covalent complexes led to the development of two methods: the double point attractor method and the flexible sidechain method.<sup>9</sup> In this regard, Autodock has proven to be extremely flexible and reliable. These two methods differ in the way the ligand is modeled (Figure 1). With the *two-point attractor method*, the alkylating molecule is modeled as a free ligand, and a custom potential is used to bring together the covalently-bound portions of ligand and alkylated residue.

With the *flexible sidechain method*, the ligand is joined in an arbitrary conformation with the target, and the attached ligand is modeled as a fully flexible side chain in the AutoDock simulation, in a similar fashion as other docking programs (GOLD<sup>14</sup> and FlexX<sup>15</sup>). Previously reported results suggest that AutoDock flexible residue method outperformed GOLD in a dataset of 76 complexes.<sup>4</sup>

With both methods, it is possible to model the most frequently covalently modified residues, such as cysteine, lysine, threonine, and serine.

These two methods were tested on a set of 20 diverse systems in order to compare their relative performance and identify pros and cons of each.



**Figure 1.** Schematic diagram of the two covalent docking approaches, with the protein in gray and the ligand in black. a) The two-point attractor method calculates two energetically-attractive maps (shown schematically with circles) at the site of covalent attachment on the protein, and uses two dummy atom types (X and Z) to target the ligand to the site. b) The flexible sidechain method overlaps the ligand, then uses AutoDock to optimize the conformation.

## 5.2 METHODS

### 5.2.1 Two point attractor method

The two-point attractor method uses modified interaction maps and modified atom types. The target residue side chain is clipped to remove two terminal atoms (i.e.  $C\beta$  and  $O\gamma$  for serine). These two atoms are attached to the alkylating ligand at the appropriate location with ideal chemical geometry, and assigned two special atom types (X and Z). Two specialized interaction maps are created for these atom types, with a Gaussian potential (termed here the Z-potential) centered on the location of the original atoms in the receptor structure, with a negative value close to the desired location and rising to zero at distant locations. The Z-potential

penalizes poses where Z or X atoms are outside their covalently-attached location, pulling the ligand in the proper pose (see Figure 1a).

The Gaussian Z-potential is defined with two parameters: width ( $\delta$ , unit: Å) and amplitude ( $\varepsilon$ , unit: Kcal/Mol). Optimal values for these parameters need to provide a good compromise between atom placement accuracy and search accessibility. The combination of large  $\delta$  and small  $\varepsilon$  results in a narrow and deep potential, whose energetic minimum will be accurately located on the atom position, but will be much harder to be found by the search algorithm. Conversely, small  $\delta$  and large  $\varepsilon$  would be easier to be reached during the search, but it would provide lower precision in reproducing the covalent geometry.

Prior to performing dockings on the whole dataset with the two-point attractor method, we identified the optimal the values of Gaussian potential coefficients  $\delta$  and  $\varepsilon$  using the complex of a DD-peptidase with penicillin G (PDB entry 1pwc<sup>16</sup>) which was also used as test case for the first implementation of the method.<sup>9</sup> We tested a range of values (0.5-100 Å for  $\delta$  and 10-50 kcal/mol for  $\varepsilon$ , Table 1). Most values provided satisfying results, with several combinations achieving RMSD below 1.0 Å. As expected, extreme value pairs (i.e. high- $\delta$ /low- $\varepsilon$ , low- $\delta$ /high- $\varepsilon$ ) corresponded to the worst results with higher deviations from the experimental poses. To identify the optimal values, the following criteria were adopted: *a*) the largest  $\delta$  was chosen that still provides proper atom placement, to improve the search effectiveness; *b*) the smallest magnitude of  $\varepsilon$  was chosen that results in proper placement, to ensure that the contribution of the Z-potential does not swamp out the energetics of interaction of the non-covalent portion of the molecule. The optimal values selected were  $\delta=3$  Å, and  $\varepsilon = 10$  Kcal/mol.

**Table 1.** Calibration results of the double point attractor method using the crystal structure 1pwc. Values show the influence of different values  $\delta$  and  $\epsilon$  on docking accuracy (RMSD (Å) with respect to the experimental structure) and binding free energy ( B. E. ) (Kcal/mol).

Widt h( $\delta$ )	RMSD - $\epsilon$ -10	RMSD - $\epsilon$ -20	RMSD - $\epsilon$ -30	RMSD - $\epsilon$ -40	RMSD - $\epsilon$ -50	B. E.- $\epsilon$ -10	B. E.- $\epsilon$ -20	B. E.- $\epsilon$ -30	B. E.- $\epsilon$ -40	B. E. - $\epsilon$ -50
0.50	1.07	2.13	0.7	1.29	2.36	-2.92	0.39	2.04	3.07	5.38
1	1.05	1.17	1.14	2.49	0.67	-5.48	-5.02	-2.52	-2.14	-3.67
3	0.73	0.61	0.69	0.85	0.67	-6.23	-6.19	-5.95	-6.02	-6.08
5	0.7	0.73	1.04	0.69	0.65	-6.37	-6.25	-6.28	-6.5	-6.13
10	0.96	0.74	1.18	0.72	1.37	-6.38	-6.35	-6.37	-6.6	-6.34
25	0.87	1.49	0.95	1.49	0.96	-6.35	-6.51	-6.36	-6.5	-6.27
50	1.79	1.99	0.81	0.89	0.82	-6.61	-6.58	-6.89	-6.29	-6.6
75	1.14	1.07	1.02	1.1	0.9	-6.27	-6.29	-6.29	-6.01	-6.51
100	1.41	2.5	3.08	0.8	1.12	-6.1	-5.78	-6.39	-6.3	-6.03

### 5.2.2 Flexible sidechain method

In the flexible sidechain method, a ligand coordinate file is modified by connecting the two target residue atoms at the site of alkylation, with ideal chemical geometry. These two ligand atoms are then overlapped with the matching atoms in the receptor structure to establish the covalent bond with the residue before running the docking. Then, during the docking, the complex is treated as a fully flexible side chain, using the existing AutoDock method for modeling selected receptor flexibility.  $C\alpha$  and  $C\beta$  atoms are fixed in space, while all torsions of both ligand and residue (including  $C\alpha$ - $C\beta$ ) are allowed to rotate (Figure 1b).

### 5.2.3 Data set

In order to compare results obtained with the two methods, we filtered the PDB to obtain a representative data set using the following criteria:

- covalent ligand binding;
- X-ray diffraction resolution  $\leq 2.65$  Å;
- structurally diverse ligands, with torsional degrees of freedom between 5 and 22;
- Ser or Cys covalent residue;
- no co-factors in the binding site;

The final data set includes 20 structures from 19 different protein families. Ligands in these complexes present a wide range of structural complexity, with torsional degrees of freedom ranging from 6 to 20, and molecular weight from 369 to 738 Daltons.

### 5.2.4 Coordinate preparation and docking

The two covalent protocols require different input preparation protocols. Ligand and receptor structures were retrieved from the Protein Data Bank and subsequently a single protein subunit has been selected and all waters and co-factors removed. AutoDockTools (ADT)<sup>9</sup> was used to add hydrogens, calculate Gasteiger charges and generate PDBQT files.

Regarding the two-point attractor method, the ligand was extracted from the PDB file and modified it by adding two extra atoms X and Z, corresponding to the two side chain terminal atoms (i.e., serine C $\beta$ , O $\gamma$ ). Hydrogens and charges were added, and the resulting PDBQT coordinates were randomized (orientation, translation and torsions) using AutoDock Vina.<sup>17</sup> The grid parameter file (GPF) was modified to include the definition of the two Z-potentials for atoms X and Z. The potential centers were defined on the original coordinates of the residue atoms, and grid map calculated following the standard AutoDock protocol for ligand docking.

In the flexible sidechain method, the ligand file was created by adding two receptor atoms to the ligand coordinates in ideal chemical geometry, and then this file was used to superimpose the ligand on the appropriate residue in the target protein (Figure 1b). ADT was used to add hydrogens, calculate Gasteiger charges and generate a modified flexible ligand, using default methods. The resulting sidechain-ligand structure is treated as flexible during the docking simulation, sampling torsional degrees of freedom to optimize the interaction of the tethered ligand with the rest of the protein. Grid maps were calculated following the standard AutoDock protocol for flexible side chains.<sup>9</sup> The torsions of the flexible residue have been randomized using AutoDock Vina<sup>17</sup>.

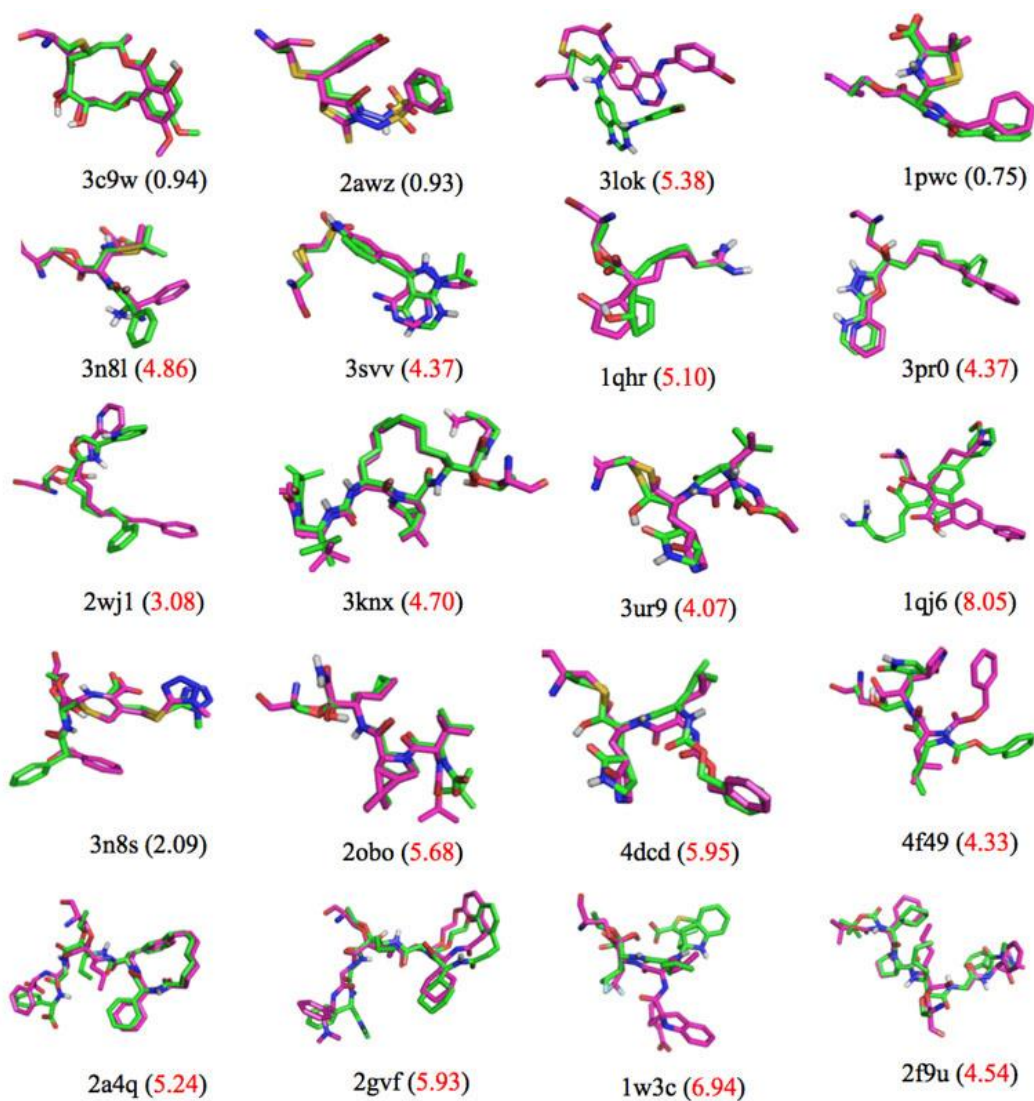
Docking simulations were performed using the Lamarckian Genetic Algorithm. For the flexible residue method, dockings on the whole dataset were run with default LGA settings (50 poses generated). Due to the higher complexity of untethered two-point attractor method, results obtained with the default LGA settings (50 poses) were compared with extended search parameters ( $ga\_population = 300$ ,  $ga\_num\_evals = 10^6$ , 100 poses generated). Final results were clustered with AutoDock with 2.0 Å RMSD tolerance.

### 5.3 RESULTS AND DISCUSSION

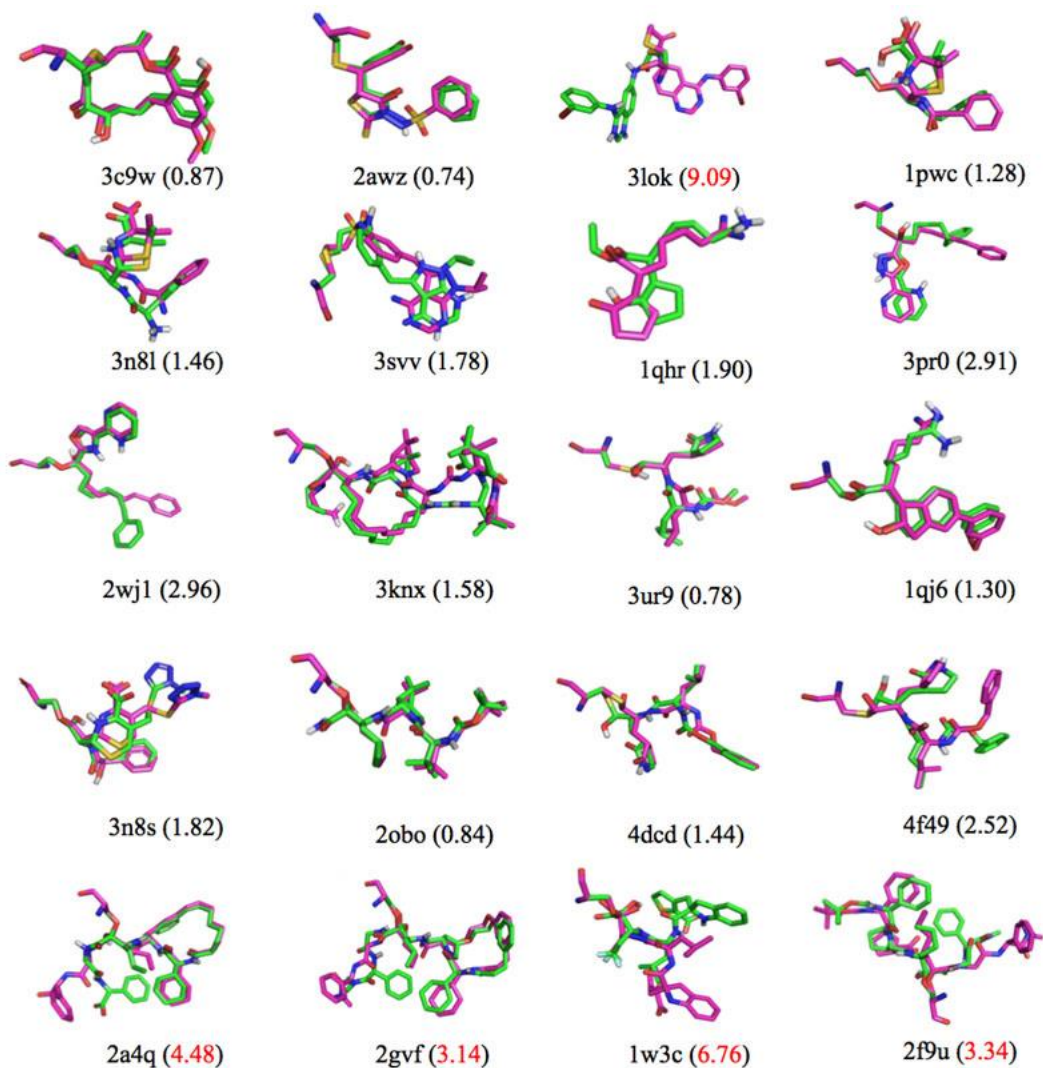
The aim of this work was to provide a suitable method for the docking calculations of covalent complexes by modifying AutoDock. For this purpose we compared and evaluated the reliability of two different methods: the two-point attractor method and the flexible sidechain method. Accuracy was assessed by measuring the RMSD between the lowest energy pose and the experimental coordinates, and results with deviations below 3.0 Å were considered successful. The total number of result clusters (2.0 Å RMSD tolerance) and the number of poses in the lowest energy cluster were used to assess the reliability in finding the correct pose. For the two-point attractor method, the RMSD between the C $\alpha$  and C $\beta$  of the docked poses and the experimental coordinates were all less than 0.09 Å, so the method is effective for forming the covalent bond. RMSD results are included in

Figures 2 and 3 along with images of the best docked conformation for each complex. Data for RMSD and cluster size, and their relationship to flexibility of complexes, are presented graphically in Figures 2 and 3 and Tables 2-4.





**Figure 2.** Double point attractor method at 100 runs: best pose of compounds with the covalently bound residues (purple) compared with the crystallographic pose (green) of the ligands. RMSD values are given in parentheses.

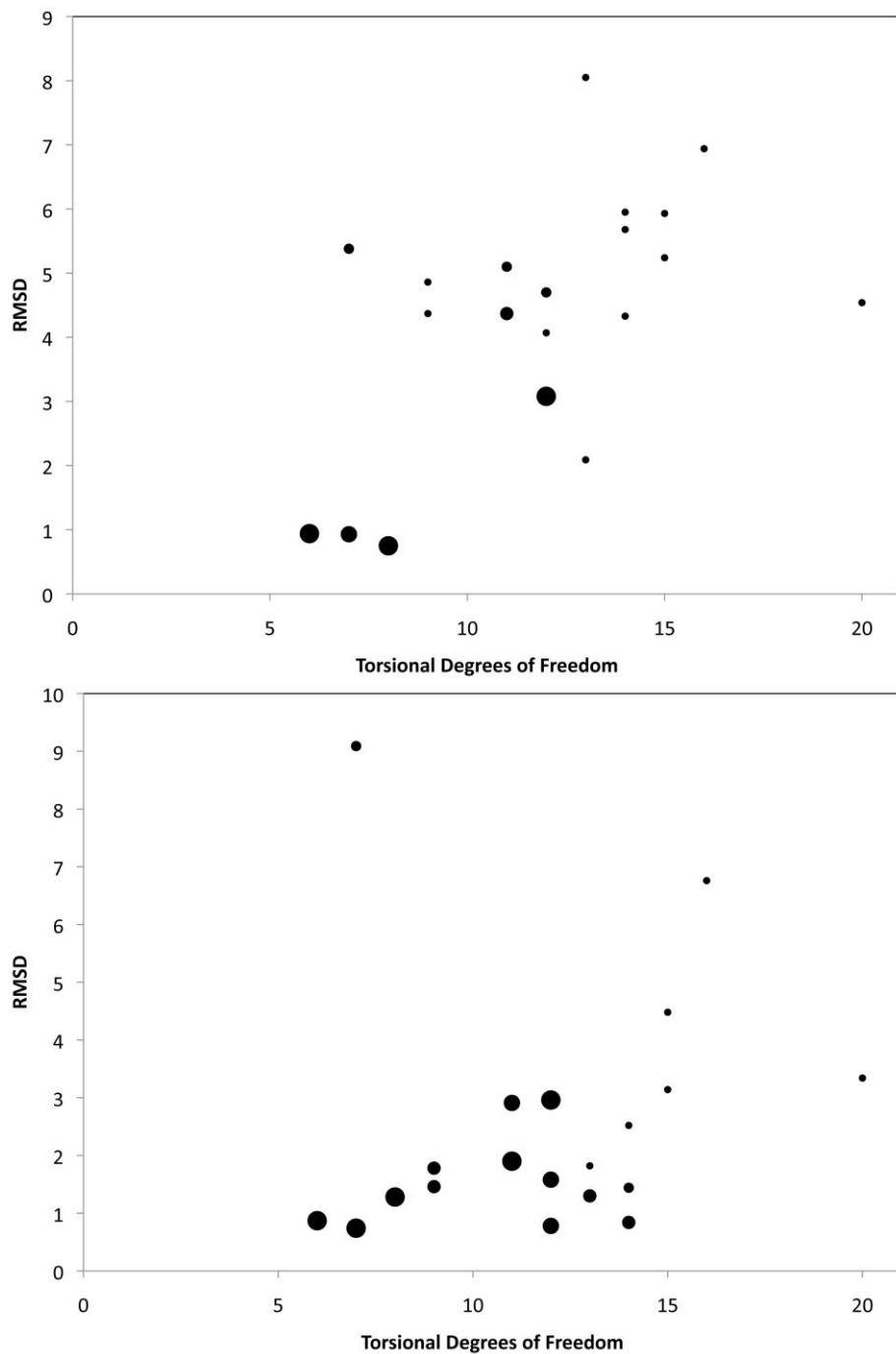


**Figure 3.** Flexible sidechain method at 50 runs: best pose of compounds with the covalently bound residues compared with the crystallographic pose of the ligands.

Overall, the two-point attractor method performed poorly, reproducing the experimental coordinates in 4 systems out of 20 (20% success rate). Energies improved slightly when going from default LGA settings (Table 3) to extended settings (Table 4), but no significant variations were found in the overall success rate.

Moreover, accuracy appears to be correlated with the number of rotatable bonds of the ligand (Figure 4a), with only one successful result with ligands over 9 torsional degrees of freedom. Accuracy of the flexible

side chain method is considerably higher, reproducing experimental coordinates in 15 out of 20 systems (75% success rate). As expected, this method is less sensitive to the torsional complexity of the ligands, obtaining successful results up to 14 torsional degrees of freedom (Figure 4b).



**Figure 4.** Correlation between the torsional degrees of freedom and results RMSD: *a)* two-point attractor method; *b)* flexible side chain method. The size of data points is roughly proportional to the number of conformations in the cluster of best predicted energy.

Overall, highly flexible ligands showed poorer reliability in finding the correct result, as observed in the larger number of scarcely populated clusters for very flexible ligands. Since the ligand is docked untethered with the two-point attractor method, it results in larger numbers of degrees of freedom, making the search implicitly more complex. Therefore, it is more prone to finding false positive poses that can be ranked higher than the correct pose, resulting in RMSD values as high as 6 Å. As expected, search efficiency is more effective with the flexible residue method than with the two-point attractor method. Both methods are successful with low torsional degrees of freedom (6, 7 and 8 for 2awz<sup>18</sup>, 3c9w<sup>19</sup>, 1pwc<sup>16</sup>, respectively), but they both fail with complex 3lok<sup>20</sup> (7 torsions). This is due to the absence of a coordinating water between the ligand and residue Glu339 (data not shown). Also, neither method is successful in re-docking the complex 1w3c,<sup>21</sup> due to the large fraction of the ligand (i.e., thiophene and indole rings) exposed to the solvent.

**Table 2.** Double point attractor method obtained with default LGA settings (50 poses)

PDB code	Resolution (Å)	Number of Clusters	Poses in best cluster	Lowest Energy (Kcal mol <sup>-1</sup> )	RMSD of lowest energy (Å)	Energy of lowest RMSD	Lowest RMSD found	RMS D C $\alpha$ and C $\beta$	Rotatable bonds
3c9w	2.5	2	47	-6.7	1.07	-4.68	0.76	0.025	6
2awz	2.15	11	29	-6.38	1	-5.84	0.3	0.125	7
3lok	2.48	19	14	-7.89	5.55	-5.43	2.19	0.005	7
1pwc	1.1	6	22	-6.23	0.73	-5.41	0.52	0.085	8
3n8l	1.4	18	2	-4.86	5.71	-3.85	3.39	0.545	9
3svv	2.2	36	3	-7.56	4.48	-4.33	4.26	0.09	9
1qhr	2.2	14	7	-5.9	4.85	-4.75	4.65	0.09	11
3pr0	2.2	16	8	-9.44	4.33	-8.88	3.63	0.25	11
2wj1	1.84	10	16	-12.72	3.14	-9.03	1.61	0.17	12
3knx	2.65	27	12	-11.28	4.69	-9.05	4.81	0.445	12
3ur9	1.65	37	1	-4.54	4.45	-3.46	4.03	0.17	12
1qj6	2.2	27	4	-12.23	1.63	-12.12	1.1	0.04	13
3n8s	2	22	8	-4.37	2.31	-4.22	1.73	0.1	13
2obo	2.6	37	2	-6.05	5.43	-4.14	5.07	0.65	14
4dcd	1.69	43	1	-5.64	5.74	-4.11	5.32	0.37	14
4f49	2.25	37	6	-6.56	5.47	-4.41	3.01	0.05	14
2a4q	2.45	42	5	-9.31	5.23	-9.14	4.8	1.3	15
2gvf	2.5	46	1	-6.46	6.17	-1.91	5.54	0.325	15
1w3c	2.30	45	1	-10.71	6.12	-9.00	2.10	0.485	16
2f9u	2.6	49	1	-6.92	4.97	-4.44	4.7	0.745	20

**Table 3.** Double point attractor method obtained with extended LGA settings (100 poses)

PDB code	Res (Å)	N of Clusters	Poses in best cluster	Lowest Energy (Kcal mol <sup>-1</sup> )	RMSD of lowest energy (Å)	Energy of lowest RMSD	Lowest RMSD found	Rotatable bonds
3c9w	2.5	2	99	-6.7	0.94	-6.39	0.6	6
2awz	2.15	10	56	-6.27	0.93	-5.53	0.4	7
3lok	2.48	18	30	-7.97	5.38	-6.6	2.64	7
1pwc	1.1	6	62	-6.4	0.75	-5.92	0.55	8
3n8l	1.4	28	16	3.68	4.86	-1.08	3.21	9
3svv	2.2	42	9	-7.79	4.37	-6.87	3.67	9
1qhr	2.2	29	27	-5.74	5.1	-5.16	4.5	11
3pr0	2.2	13	45	-10.2	4.37	-9.12	3.66	11
2wj1	1.84	11	96	-12.88	3.08	-11.84	1.94	12
3knx	2.65	37	36	-11.83	4.7	-8.29	4.61	12
3ur9	1.65	46	10	-5.44	4.07	-4.16	3.88	12
1qj6	2.2	26	13	-12.55	8.05	-12.33	0.64	13
3n8s	2	30	11	-5.04	2.09	-1.42	1.48	13
2obo	2.6	45	15	-6.52	5.68	-4.04	5.02	14
4dcd	1.69	67	2	-6.38	5.95	-3.96	5.28	14
4f49	2.25	66	2	-7.7	4.33	-6.68	2.79	14
2a4q	2.45	72	18	-10.37	5.24	-9.58	4.63	15
2gvf	2.5	80	1	-9.08	5.93	-2.41	5.51	15
1w3c	2.30	65	1	-11.05	6.94	-9.41	2.31	16
2f9u	2.6	96	1	-7.19	4.54	-7.19	4.54	20

**Table 4.** Flexible docking results with standard LGA settings (50 poses)

PDB code	Res. (Å)	N of Clusters	Poses in best cluster	Lowest Energy (Kcal mol <sup>-1</sup> )	RMSD of lowest energy (Å)	Energy of lowest RMSD	Lowest RMSD found	Rotatable bonds
3c9w	2.5	1	50	-10.17	0.87	-10.17	0.86	6
2awz	2.15	1	50	-10.82	0.74	-10.81	0.5	7
3lok	2.48	5	17	-8.36	9.09	-7.58	5.31	7
1pwc	1.1	6	42	-14.09	1.28	-11.98	0.95	8
3n8l	1.4	5	21	-35	1.46	-12.2	1.38	9
3svv	2.2	10	25	-12.25	1.78	-14.96	0.94	9
1qhr	2.2	4	47	-13.64	1.9	-13.07	0.71	11
3pr0	2.2	2	34	-15.2	2.91	-14.96	2.38	11
2wj1	1.84	3	47	-15.67	2.96	-14.33	2.59	12
3knx	2.65	4	31	-17.18	1.58	-16.2	1.11	12
3ur9	1.65	9	33	-12.83	0.78	-12.82	0.75	12
1qj6	2.2	8	29	-15.35	1.3	-15.18	0.7	13
3n8s	2	9	9	-15.6	1.82	-14.52	1.59	13
2obo	2.6	10	30	-16.89	0.84	-16.59	0.54	14
4dcd	1.69	11	14	-15.13	1.44	-14.91	0.6	14
4f49	2.25	11	7	-15.19	2.52	-14.81	1.47	14
2a4q	2.45	25	8	-21.82	4.48	-19.79	1.71	15
2gvf	2.5	22	5	-20.86	3.14	-18.75	2.52	15
1w3c	2.30	23	6	-18.32	6.76	-17.20	2.87	16
2f9u	2.6	36	1	-21.65	3.34	-20.65	0.93	20

## 5.4 CONCLUSIONS

We compared two methods for performing covalent dockings with AutoDock: a two-point attractor method and a modified protocol of the flexible receptor residues. The methods were tested on a diverse set of complexes in order to assess their relative performances. Overall, the flexible residue method provided better accuracy (75% success rate) than the two-point attractor method (20%). The success rate of both methods showed a correlation with the complexity of the ligands, decreasing accuracy with the increase of torsional degrees of freedom. The flexible



residue method provides the most effective approach in simulating covalent ligands.

The two-point attractor method, on the other hand, allows incorporation of a certain degree of approximation in the position of the covalent attachment, which may be important for modeling significant motion in residues at the site--we are currently assessing this possibility. Finally, both methods provide examples of how AutoDock suite parameters can be customized to implement new methods and extend basic functionalities. Both methods are implemented in the current version of AutoDock (v.4.2.6), and tools and scripts for preparing input files are included in latest release of AutoDockTools (v.1.5.7-latest), all available at <http://autodock.scripps.edu>.

## 5.5 REFERENCES

1. Johnson, D. S.; Weerapana, E.; Cravatt, B. F., Strategies for discovering and derisking covalent, irreversible enzyme inhibitors. *Future medicinal chemistry* **2010**, 2, 949-64.
2. Singh, J.; Petter, R. C.; Baillie, T. A.; Whitty, A., The resurgence of covalent drugs. *Nature reviews. Drug discovery* **2011**, 10, 307-17.
3. Mah, R.; Thomas, J. R.; Shafer, C. M., Drug discovery considerations in the development of covalent inhibitors. *Bioorganic & medicinal chemistry letters* **2014**, 24, 33-9.
4. Ouyang, X.; Zhou, S.; Su, C. T.; Ge, Z.; Li, R.; Kwoh, C. K., CovalentDock: automated covalent docking with parameterized covalent linkage energy estimation and molecular geometry constraints. *Journal of computational chemistry* **2013**, 34, 326-36.
5. Zhu, K.; Borrelli, K. W.; Greenwood, J. R.; Day, T.; Abel, R.; Farid, R. S.; Harder, E., Docking covalent inhibitors: a parameter free approach to pose prediction and scoring. *Journal of chemical information and modeling* **2014**, 54, 1932-40.

6. London, N.; Miller, R. M.; Krishnan, S.; Uchida, K.; Irwin, J. J.; Eidam, O.; Gibold, L.; Cimermancic, P.; Bonnet, R.; Shoichet, B. K.; Taunton, J., Covalent docking of large libraries for the discovery of chemical probes. *Nature chemical biology* **2014**, 10, 1066-72.
7. Scholz, C.; Knorr, S.; Hamacher, K.; Schmidt, B., DOCKTITE-A Highly Versatile Step-by-Step Workflow for Covalent Docking and Virtual Screening in the Molecular Operating Environment. *Journal of chemical information and modeling* **2015**, 55, 398-406.
8. Morris, G. M.; Goodsell, D. S.; Halliday, R. S.; Huey, R.; Hart, W. E.; Belew, R. K.; Olson, A. J., Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function. *J. Comput. Chem.* **1998**, 19, 1639-1662.
9. Morris, G. M.; Huey, R.; Lindstrom, W.; Sanner, M. F.; Belew, R. K.; Goodsell, D. S.; Olson, A. J., AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. *Journal of computational chemistry* **2009**, 30, 2785-91.
10. Osterberg, F.; Morris, G. M.; Sanner, M. F.; Olson, A. J.; Goodsell, D. S., Automated docking to multiple target structures: incorporation of protein mobility and structural water heterogeneity in AutoDock. *Proteins: Struct. Funct. Genet.* **2001**, in press.
11. Forli, S.; Botta, M., Lennard-Jones potential and dummy atom settings to overcome the AUTODOCK limitation in treating flexible ring systems. *Journal of chemical information and modeling* **2007**, 47, 1481-92.
12. Santos-Martins, D.; Forli, S.; Ramos, M. J.; Olson, A. J., AutoDock4(Zn): an improved AutoDock force field for small-molecule docking to zinc metalloproteins. *Journal of chemical information and modeling* **2014**, 54, 2371-9.

13. Forli, S.; Olson, A. J., A force field with discrete displaceable waters and desolvation entropy for hydrated ligand docking. *Journal of medicinal chemistry* **2012**, 55, 623-38.
14. Verdonk, M. L.; Cole, J. C.; Hartshorn, M. J.; Murray, C. W.; Taylor, R. D., Improved protein-ligand docking using GOLD. *Proteins* **2003**, 52, 609-23.
15. Kramer, B.; Rarey, M.; Lengauer, T., Evaluation of the FLEXX incremental construction algorithm for protein-ligand docking. *Proteins* **1999**, 37, 228-41.
16. Silvaggi, N. R.; Josephine, H. R.; Kuzin, A. P.; Nagarajan, R.; Pratt, R. F.; Kelly, J. A., Crystal structures of complexes between the R61 DD-peptidase and peptidoglycan-mimetic beta-lactams: a non-covalent complex with a "perfect penicillin". *Journal of molecular biology* **2005**, 345, 521-33.
17. Trott, O.; Olson, A. J., AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *Journal of computational chemistry* **2010**, 31, 455-461.
18. Powers, J. P.; Piper, D. E.; Li, Y.; Mayorga, V.; Anzola, J.; Chen, J. M.; Jaen, J. C.; Lee, G.; Liu, J.; Peterson, M. G.; Tonn, G. R.; Ye, Q.; Walker, N. P.; Wang, Z., SAR and mode of action of novel non-nucleoside inhibitors of hepatitis C NS5b RNA polymerase. *Journal of medicinal chemistry* **2006**, 49, 1034-46.
19. Rastelli, G.; Rosenfeld, R.; Reid, R.; Santi, D. V., Molecular modeling and crystal structure of ERK2-hypothemycin complexes. *Journal of structural biology* **2008**, 164, 18-23.
20. Kluter, S.; Simard, J. R.; Rode, H. B.; Grutter, C.; Pawar, V.; Raaijmakers, H. C.; Barf, T. A.; Rabiller, M.; van Otterlo, W. A.; Rauh, D., Characterization of irreversible kinase inhibitors by directly detecting

covalent bond formation: a tool for dissecting kinase drug resistance.

*Chembiochem : a European journal of chemical biology* **2010**, 11, 2557-66.

21. Ontoria, J. M.; Di Marco, S.; Conte, I.; Di Francesco, M. E.; Gardelli, C.; Koch, U.; Matassa, V. G.; Poma, M.; Steinkuhler, C.; Volpari, C.; Harper, S., The design and enzyme-bound crystal structure of indoline based peptidomimetic inhibitors of hepatitis C virus NS3 protease. *Journal of medicinal chemistry* **2004**, 47, 6443-6.

## CONCLUDING REMARKS

The application of structure-based methods to the different targets object of my PhD work allowed to reach important results that I am going to briefly summarize in few points:

- I followed the investigation on dual Inhibitors of RT discovered previously by virtual screening approach. These compounds are able to inhibit both RT catalytic functions causing a complete block of RT. The mechanism of some new inhibitors has been examined in deep by means of biochemical and computational methods. These studies highlighted new and interesting mechanism. In particular these compounds seem to bind in two different sites opening the road to new allosteric RT inhibitors with a favourable resistance profile. This could lead to a reduction of combined drugs used in therapy and of toxic side effects.

- Monoamine oxidase B is a target involved in many neurodegenerative disorders and in the production of reactive oxygen species (ROS). Therefore the increase number of elderly people affected by these disorders has led to an increased interest towards selective, potent and reversible MAO inhibitors. The information derived by computational studies guided the optimization of active inhibitors allowing to obtain compounds active in low nM range. I also validated both protocols of docking and MD obtaining good correlation between predicted and experimental binding energy. The protocol can be applied to other series of compounds to rationalize their activity or to find new scaffolds.

- The emergence of Ebola virus outbreak prompted the development of a vaccine in the immediate. However, it would offer great advantage to find inhibitors able to block the virus and to reduce the mortality. Therefore, since our collaborators were able to obtain and purify the whole VP35 protein, I first investigated the effect of single mutations in the dsRNA binding area and then applied a virtual screening workflow to find possible VP35 inhibitors. The VS protocol includes pharmacophore and consensus docking methods. We have found two compounds which show activity against this target in the low micromolar range. Although the studies are still in progress these results are encouraging and offer a base for further

investigation and optimization. Furthermore the same VS protocol can be applied to larger databases.

• Computational tools are widely used by academy and pharmaceutical companies. Limitations on their application have to be considered and if possible overcome. Until few years ago docking approaches could not be applied to predict the binding of covalent species but only show the step before binding: the approaching process to the amino acid involved. Recently many companies and groups, involved in computational software development, invest energy in tools able to predict covalent binding event.

One is the group of Prof Arthur Olson at the Scripps Institute. I was involved in the validation of two covalent docking protocols of the software Autodock and now these methods are available for the scientific community.

Many of the results presented in this thesis were object of multidisciplinary scientific publications, oral and poster presentations.

## PUBLICATIONS

- Meleddu, R.; Cannas, V.; Distinto, S.; Sarais, G.; Del Vecchio, C.; Esposito, F.; Bianco, G.; Corona, A.; Cottiglia, F.; Alcaro, S.; Parolin, C.; Artese, A.; Scalise, D.; Fresta, M.; Arridu, A.; Ortuso, F.; Maccioni, E.; Tramontano, E., *Design, synthesis, and biological evaluation of 1,3-diarylpropenones as dual inhibitors of HIV-1 reverse transcriptase*. ChemMedChem **2014**, 9, 1869-1879.

- Meleddu, R.; Distinto, S.; Corona, A.; Bianco, G.; Cannas, V.; Esposito, F.; Artese, A.; Alcaro, S.; Matyus, P.; Bogdan, D.; Cottiglia, F.; Tramontano, E.; Maccioni, E., *(3Z)-3-(2-[4-(aryl)-1,3-thiazol-2-yl]hydrazin-1-ylidene)-2,3-dihydro-1H-indol-2-one derivatives as dual inhibitors of HIV-1 reverse transcriptase*. European Journal of Medicinal Chemistry **2015**, 93, 452-460.

- Meleddu, R.; Maccioni, E.; Distinto, S.; Bianco, G.; Melis, C.; Alcaro, S.; Cottiglia, F.; Ceruso, M.; Supuran, C. T., *New 4-[(3-cyclohexyl-4-aryl-2,3-dihydro-1,3-thiazol-2-ylidene)amino]benzene-1-sulfonamides, synthesis and*

*inhibitory activity toward carbonic anhydrase I, II, IX, XII.* Bioorganic and Medicinal Chemistry Letters, **2015**, 25, 3281-3284.

- Bianco, G.; Forli, S.; Goodsell, D. S.; Olson, A. J., *Covalent docking using autodock: Two-point attractor and flexible side chain methods.* Protein Science, **2016**, 25, 295-301.

- Distinto, S.; Meleddu, R.; Yanez, M.; Cirilli, R.; Bianco, G.; Sanna, M. L.; Arridu, A.; Cossu, P.; Cottiglia, F.; Faggi, C.; Ortuso, F.; Alcaro, S.; Maccioni, E., *Drug design, synthesis, in vitro and in silico evaluation of selective monoaminoxidase B inhibitors based on 3-acetyl-2-dichlorophenyl-5-aryl-2,3-dihydro-1,3,4-oxadiazole chemical scaffold.* European Journal of Medicinal Chemistry **2016**, 108, 542-552.

- Meleddu, R.; Distinto, S.; Maccioni, E.; Arridu, A.; Melis, C.; Bianco, G.; Matyus, P.; Cottiglia, F.; Sanna, A.; De Logu, A., *Exploring the thiazole scaffold for the identification of new agents for the treatment of fluconazole resistant Candida.* Journal of Enzyme Inhibition and Medicinal Chemistry **2016**, 1-6.

- Corona, A.; Meleddu, R.; Esposito, F.; Distinto, S.; Bianco, G.; Masaoka, T.; Maccioni, E.; Menéndez-Arias, L.; Alcaro, S.; Le Grice S.F.J.; Tramontano, E. *Ribonuclease H/DNA Polymerase HIV-1 Reverse Transcriptase Dual Inhibitor: Mechanistic Studies on the Allosteric Mode of Action of Isatin-Based Compound RMNC6.* PLoS ONE, **2016** in press DOI: 10.1371/journal.pone.0147225.

- Cannas, V.; Zinzula, L.; Distinto, S.; Fadda, V.; Daino, G. L.; Bianco, G.; Corona, A.; Esposito, F.; Taylor, G.; Alcaro, S.; Maccioni, E.; Tramontano, E.; *Definition of key residues in dsRNA recognition and IFN-antagonism of Ebola Virus VP35 for drug development. Submitted.*

## ORAL PRESENTATIONS

- Bianco G., Distinto S., Meleddu R., Cannas C., Del Vecchio C., Cottiglia F., Alcaro S., Parolin C., Tramontano E., Maccioni .*New dual inhibitors of RT associated functions: investigation on their mechanism and mode of action.* In: Book of abstract-VII edizione del Meeting "Nuove Prospettive in Chimica Farmaceutica". Savigliano (CN), 29<sup>th</sup> 31<sup>st</sup> May 2013.

-Bianco G. Identification and application of docking and MD protocol for ligand-MAO-B complexes study. 4th Meeting of the Paul Ehrlich MedChem Euro-PhD Network. 20<sup>th</sup> – 22<sup>nd</sup> June 2014, Hradec Králové

-Bianco G., Distinto S., Meleddu R., Arridu A., Melis C., Alcaro S., Cottiglia F., Tramontano E., Maccioni E. Synthesis, biological evaluation and investigation on the mechanism of action of a new class of indolinones as dual inhibitors of HIV-1 reverse transcriptase. 5<sup>o</sup> meeting of the Paul Ehrlich MedChem Euro-PhD Network, 3<sup>rd</sup>-5<sup>th</sup> July, 2015, Cracow

## POSTERS

-Dunkel P., Bianco G, Deme R., Maccioni E., Mátyus P.: *Synthesis of semicarbazide-sensitive amino-oxidase inhibitors via Tert-amino Effect cyclization*. In: Book of abstract-1st meeting of the Paul Ehrlich MedChem Euro-Phd Network; 7th meeting of the European Network of Doctoral Studies in Pharmaceutical Sciences-, Madrid, 13<sup>th</sup> 15<sup>th</sup> July 2011.

-Bianco G., Czompa A., Deme R., Schlich M., Maccioni E., Mátyus P.: *Synthesis of novel semicarbazide-sensitive amine oxidase inhibitors via Tert-amino Effect*. In: Book of abstract-XXII<sup>nd</sup> International symposium on medicinal Chemistry (EFMC-ISM 2012) Berlino, 2<sup>nd</sup> 6<sup>th</sup> September 2012.

-Bianco G., Distinto S., Meleddu R., Cannas V., Del Vecchio C., Cottiglia F., Alcaro S., Parolin C., Tramontano E., and Maccioni E.: *Investigation on the mechanism of action of new dual inhibitors of HIV-1 RT*. In: Book of abstract-IX EWDD Ninth European Workshop in Drug Design-Certosa di Pontignano (Siena-Italy), 19<sup>th</sup> 25<sup>h</sup> May 2013.

-Meleddu R., Distinto S., Sissi C., Musetti C., Arridu A., Bianco G., Parrotta L., Ortuso F., Alcaro S., Maccioni E. *Design, synthesis and biological activity evaluation of c-myc proto-oncogene binders as potential antitumor agents*. In: Book of abstracts-VII edizione del Meeting "Nuove Prospettive in Chimica Farmaceutica" (NPCF7) . Savigliano (CN), 29<sup>th</sup> – 31<sup>st</sup> May 2013.

-Meleddu R., Arridu A., Bianco G., Tramontano E., Distinto S., Matyus P., Maccioni E. *Towards the total block of HIV-1 Reverse Transcriptase: the dual-inhibitors approach*. In: *Book of abstracts- From medicine to bionics* 1st European PhD conference. Budapest, 13<sup>rd</sup>-15<sup>th</sup> June 2013.



-Arridu A., Meleddu R., Bianco G., Distinto S., Bogdan D., Matyus P., Maccioni E. *In search of new antitumor agents through the combination of individually active moieties: construction of a [3,5-diaryl-(4,5-dihydropyrazol-1-yl)-4-oxo-1,3-thiazol-5-ylidene]- 1H-indol-2-ones library* . In: Book of abstracts- 3rd Meeting of the Paul Ehrlich MedChem Euro PhD Network . Santa Margherita di Pula (CA), 27<sup>th</sup>-29<sup>th</sup> September 2013.

-Bianco G., Cossu P., Distinto S., Meleddu R., Arridu A., Ortuso F., Alcaro S., Maccioni E. *MD protocol validation for the study of MAOB-inhibitor complexes*. In: Book of abstracts-3<sup>rd</sup> Meeting of the Paul Ehrlich MedChem Euro-PhD Network- Santa Margherita di Pula (CA), 27<sup>th</sup>-29<sup>th</sup> September 2013.

- Bianco G., Distinto S., Meleddu R., Arridu A., Melis C., Alcaro S., Cottiglia F., Tramontano E., Maccioni E. *Synthesis, biological evaluation and investigation on the mechanism of action of a new class of indolinones as dual inhibitors of HIV-1 reverse transcriptase*. 5<sup>th</sup> meeting of the Paul Ehrlich MedChem Euro-PhD Network, 3<sup>rd</sup> - 5<sup>th</sup> July, 2015

### **CONFERENCE PROCEEDINGS ABSTRACTS**

-Meleddu R., Distinto S., Corona A., Bianco G., Arridu A., Cannas V., Tramontano E., Alcaro S., Matyus P., Maccioni E. *Synthesis and biological activity evaluation of new dual-inhibitors of both associated functions of HIV-1 RT* . In: Book of abstracts-3<sup>rd</sup> Meeting of the Paul Ehrlich MedChem Euro-PhD Network . Pula, 27<sup>th</sup>-29<sup>th</sup> September 2013.

-Corona A., Meleddu R., Esposito F., Distinto S., Bianco G., Maccioni E., Le Grice S., Tramontano E. *Site directed mutagenesis studies on HIV-1 reverse transcriptase (RT) shed light on the mechanism of action of a new Ribonuclease H/ DNA polymerase RT dual inhibitor*. In: *Retrovirology*. vol. 10, Churchill College, Cambridge University, 16<sup>th</sup> – 18<sup>th</sup> September 2013, doi:10.1186/1742-4690-10-S1-P19.

### **PARTECIPATION AT SCHOOLS**

-Approaches for identification of Antiviral agents summer school (IAAASS)-Santa Margherita di Pula, 30<sup>th</sup> September-4<sup>th</sup> October 2012.

- IX EWDD Ninth European Workshop in Drug Design-Certosa di Pontignano (Siena-Italy), 19<sup>th</sup>-25<sup>th</sup> May 2013.

### **GRANTS and AWARDS**

22/06/2014 GlobusDoc grant for a three months fellowship at the Molecular Graphics Laboratory, department of Molecular Biology of the Scripps Research Institute, La Jolla, California.

18/06/2015 ErasmusPlaceDoc grant for a three months fellowship at the Molecular Modeling laboratory, department of Biomedical Sciences of the University of Alcalá, Madrid, Spain.

05/07/2015 Certificate of distinction. 5th meeting of the Paul Ehrlich MedChem Euro-PhD Network, 3rd - 5th July, 2015.

## ACKNOWLEDGEMENTS

First of all I would like to thank my mentor Simona Distinto. Not only did she guide me through the hard work that she carries on every day with passion, but also she taught me how to be a better person. Her generosity and reliability are the most important qualities that she imparted to me during these three years of my PhD and that I will cherish forever.

I would also like to thank my professor Elias Maccioni, who believed in me from the beginning and who has always been a reliable guide for me.

I would also like to thank Dr. Rita Meleddu who was there since the very beginning of my adventure and who has always supported me.

A big thanks goes to my friends and colleagues Antonella Arridu and Claudia Melis with whom I share this incredible life and work experience. We supported and helped each other and I will never forget this.

A huge thanks goes to professor Arthur Olson and all his collaborators of the Scripps Research Institute of La Jolla (San Diego) since he gave me the opportunity to work there for three months. These three months have been unforgettable for me, because of their professionalism and humanity. In particular I would like to thank professor David Goodsell for teaching me a new way of working with big patience and attention.

In addition, a special mention goes to Dr. Stefano Forli who made me feel at home from the beginning of this experience and which great help was fundamental to turn this work experience in a precious outcome for my career.

I would also like to thank professor Federico Gago of the University of Alcalá, who gave me the great opportunity to work in his laboratory for three months. His kindness, competence and intelligence made me think about computational chemistry in a more practical way.

His collaborator Dr. Pedro Sánchez-Murcia taught me different techniques that I would not have the possibility to learn otherwise. I would like to thank him for his patience and expertise.

Moreover, I would like to thank the section I.S.M.O.K.A. of the University of Cagliari since they gave me the opportunity to go abroad twice during these years.

Last but not least, I am deeply convinced that without my family and the serenity, love and support they constantly show me, I would have not been able to work and achieve results. Therefore a big thanks goes to my unique family.