UNIVERSITÀ DEGLI STUDI DI MILANO

FACOLTÀ DI SCIENZE E TECNOLOGIE

DIPARTIMENTO DI CHIMICA

Doctoral School in Chemistry - XXX Cycle

MOLECULAR MODELING OF EBOLA VIRUS INHIBITORS

Federico DAPIAGGI

Tutor: Prof. Maurizio SIRONI

Co-Tutor: Dr. Stefano PIERACCINI

Coordinator: Prof. Emanuela LICANDRO

— Academic year 2017/2018 —

Never, I said never, compare with experiment

- Magnus Bergh

No matter how important you are, you are not as important as lunch.

- Randy Pausch

Contents

Acronyms

AMPs	Antimicrobial peptides.
AS	Alanine scanning.
CAS	Computational Alanine Scanning.
CG	Coarse grain.
EBOV	Ebola Virus.
ED	Essential Dynamics.
IFN	interferon.
KPNA	Karyopherin α.
MARV	Marburg Virus.
MD	Molecular Dynamics.
MM-PBSA	Molecular Mechanics-Poisson Boltzmann Surface Area.
NP	Nucleoprotein.
PPIs	Protein-protein interactions.
RDF	Radial Distribution Function.
RMSD	Root Mean Square Deviation.
RMSF	Root Mean Square Fluctuation.
SDF	Solvent Density Function.
VP24	Viral Protein 24.
VP35	Viral Protein 35.

Thesis Overview

Computer science has grown fast in the last few decades and it is going to keep growing. Every year we witness the birth of new softwares, hardwares, OS allowing us to have more and more powerful computers. In many fields of science such as chemistry, physics and biology the theoretical simulation can help shedding light on systems, processes, reactions that are not clear yet. On one hand, computer simulations can help understanding empirical evidences or the mechanism of a reaction, on the other hand simulations can be performed in conditions that could be difficult (or impossible) to have in an actual laboratory. Moreover, simulations are particularly useful in drug design and molecular modeling. Understanding interactions at atomistic level can help to design a potential binder or inhibitor, and thus to avoid experimental tests on a number of molecules.
In this PhD thesis various computational methods are employed, in order to study different biological systems.
The first part of the thesis focuses on the modeling of Ebola virus inhibitors, studying two of the seven viral proteins coded by Ebola genome: Viral Protein 24 and Viral Protein 35.

- Chapter 2 deals with the interactions between known inhibitors of Viral Protein 35 and the protein itself. Ebola virus Viral Protein 35 carries out multiple functions necessary for virus replication and infection, in particular interfering with interferon-α/β signaling. Recently, this protein has been crystallized in complex with small organic molecules able to inhibit its interaction with viral nucleoprotein, thus reducing Ebola infections of cultured cells. In this chapter we study the binding mechanism of these inhibitors both from the energetic and conformational point of view, identifying a set of hot spots common to all the molecules and detecting interesting collective motion of the protein in complex with the ligands.

- In Chapter 3 we study the protein-protein interaction involving Viral Protein 24 and Karyopherin, responsible for the inhibition of the immune response of the host organism. Modeling the VP24–Karyopherin complex allow us to identify the amino acid residues responsible for protein–protein binding and

lead us to the identification of a nonapeptide with VP24 binding potential. Subsequently we assay the ability of this peptide to actually bind VP24 in solution using Saturation Transfer Difference NMR and Circular Dichroism. Finally we compare experimental and molecular modeling data concerning the VP24–peptide complex and we discuss putative peptide binding sites and modes.

- Chapter 4 focuses on the analysis of VP24 pockets. It is known that inhibitors and drugs usually bind in clefts or pockets on the surface of the protein. Moreover, transient pockets are likely to be found on protein surface due to its dynamic character. In this chapter we carry out pocket analysis on VP24 crystallographic structure. Furthermore we perform molecular dynamics simulations in water and more apolar solvents in order to identify transient pockets and to study the evolution and properties of the pockets already present in the crystallographic structure. Once we identify the best pocket we proceed performing a virtual screening of a number of compounds, finding molecules that could be able to bind at the interface with Karyopherin, inhibiting the protein-protein complex formation.

- Chapter 5 handles the behavior of VP24 in presence of osmolytes. We study VP24 in different solutions of either osmoprotectants or denaturants at different temperatures, in order to rationalize the effect of the osmolytes on this protein. The subject of this chapter is not directly related to the modeling of Ebola virus inhibitors however, the study of proteins solutions in different osmolytes represents one of the active research lines of our laboratory and we considered that such a study on VP24 could be of scientific interest.

The second part of the thesis focuses on the coarse grain method, that I learned during my staying in the group of Prof. Siewert Jan Marrink at University of Groningen. The case of study was the interaction of antimicrobial peptides with negatively charged membranes.

Chapter 1

Elements of theory

1.1 Molecular Dynamics

Molecular Dynamics (MD) [1] is a computational technique able to follow the evolution of a system of interacting particles, in our case atoms, in time. In order to perform a MD simulation we need:

- A starting structure containing the position of all atoms.
- A potential able to describe the system we are interested in.
- An integrator.

Given these three elements it is possible to follow the dynamics of the system by the integration of Newton's equations of motion (eq. 1.1).

$$\begin{aligned} \frac{d\vec{v}_i}{dt} &= \frac{1}{m_i}\vec{f}_i \\ \frac{d\vec{r}_i}{dt} &= \vec{v}_i \end{aligned} \tag{1.1}$$

Where $\vec{f}_i = -\vec{\nabla}U$ is the force acting on the i-th atom and m_i is its mass. If we add the initial conditions $\vec{r}_i(0) = \vec{r}_{0i}$ and $\vec{v}_i(0) = \vec{v}_{0i}$ this becomes a Cauchy problem and its solution is the trajectory of the system. The starting structure is usually provided by the Protein Data Bank, containing all the resolved X-ray structures of proteins, nucleic acids and complex assemblies.
The aim of MD is to explore the macroscopic properties of the system (e.g. pressure, energy, enthalpy) through microscopic simulations. The connection between microscopic simulations and macroscopic properties is made by statistical mechanics in which averages related to experiments are defined in terms of ensemble averages [1]. From dynamical trajectories generated at a temperature T one can calculate the

[1] Ensemble average is an average taken over a large number of replicas of the system, considered simultaneously.

ensemble average of any quantity $\bar{A}(\vec{r})$, defined as

$$\langle A \rangle \equiv \frac{\int d\vec{p}^N d\vec{r}^N e^{-\beta H(\vec{p}^N \vec{r}^N)} A(\vec{p}^N \vec{r}^N)}{\int d\vec{p}^N d\vec{r}^N e^{-\beta H(\vec{p}^N \vec{r}^N)}} \tag{1.2}$$

where H is the Hamiltonian of the classical system. Using Birkhoff's ergodic theorem [2] we can write the ensemble average (eq. 1.2) as a time average (eq. 1.3)

$$\langle A \rangle = \lim_{n \to \infty} \frac{1}{n} \sum_{i=0}^{n} A(\vec{r}(t_i)), \tag{1.3}$$

where n is the number of snapshots that builds out the trajectory, taken at times t_i. Therefore, the aim of MD is to generate enough representative conformations such that the ergodic hypothesis is satisfied.

1.1.1 Force Field

The basic idea behind MD simulations is the representation of the potential energy of the system as a function of the atomic coordinates. The potential acting on the system is calculated using a force field, and its usual expression is showed in eq. 1.4.

$$U(r) = \underbrace{U_{str} + U_{bend} + U_{tors} + U_{cross}}_{\text{bonded terms}} + \underbrace{U_{vdW} + U_{elec}}_{\text{non-bonded terms}} \tag{1.4}$$

In this PhD thesis the AMBER force field [3] has been used, and its functional form is showed in eq. 1.5.

$$\begin{aligned} U(r) = \sum_{bonds} k_b(b - b_0)^2 + \sum_{angles} k_\theta(\theta - \theta_0)^2 + \sum_{dihedrals} k_\phi[cos(n\phi + \delta) + 1] \\ + \sum_{atom} \sum_{i} \sum_{i>j} 4\epsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^{6} \right] + \sum_{i} \sum_{i>j} \frac{q_i q_j}{4\pi\epsilon r_{ij}} \end{aligned} \tag{1.5}$$

The first three summation terms are over bonds, angles and dihedrals. The fourth term describes the van der Waals interaction between non-bonded atoms, with a Lennard-Jones (LJ) 12-6 potential and the final term represents the electrostatic interaction between non-bonded atoms. The parameters in eq. 1.5 are optimized in order to reproduce experimental data (X-ray diffraction, NMR and IR spectroscopy) and quantum mechanical calculation on smaller systems.

1.1.2 Integrator

The aforementioned Cauchy problem isn't analytically solvable in our case, because of the nature of the U potential we are using. Under the influence of a continuous

potential, indeed, the motions of all the atoms are coupled together, giving rise to many-body problem that cannot be solved analytically. This is the reason why we have to use a numerical method to solve the problem. The main idea is that the integration can be split up in many small stages, each separated in time by a fixed δt. In this PhD thesis we employed the leapfrog integrator, which evolves position and velocities as shown in eq. 1.6.

$$\begin{aligned} \vec{r}_i(t + \Delta t) &= \vec{r}_i(t) + \vec{v}_i(t + \Delta t/2)\Delta t \\ \vec{v}_i(t + \Delta t/2) &= \vec{v}_i(t - \Delta t/2) + \frac{\vec{f}_i(\vec{r}_i(t))}{m_i}\Delta t \end{aligned} \tag{1.6}$$

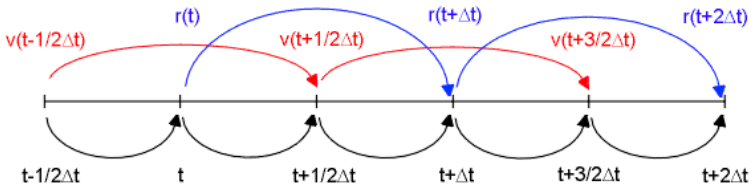


Figure 1.1: The principle of the leapfrog algorithm.[1]

The advancement of positions and velocities is not synchronous. One first advances the positions, then advances the velocities based on the new positions, then advances the positions from the new velocities, and so on.
The choice of the integration step is very important, and it depends on the algorithm chosen to solve the equations of motion. Usually we are not interested in the actual dynamics of the bonded interaction in a molecule, such as bond vibrations which require the time step of the simulation to be smaller than $10^{-15}s$. In order to constrain all the bonds to their equilibrium lengths LINCS algorithm [4] has been used, allowing a time step of 2 fs.

1.1.3 Periodic Boundary Conditions

Periodic boundary conditions enable to perform a simulation of a finite system, composed of a relatively small number of particles, avoiding border effects. To do so, the simulation box is replicated in every direction. Coordinates of particles in neighboring boxes can be computed adding or subtracting integral multiples of the box sides. If a particle, during the MD simulation, leaves the box, it is replaced by an image particle the enters form the opposite side (Figure 1.2). The most time consuming part of a MD simulation is the calculation of the non-bonded energies, that increases as N^2. In principle, non-bonded interactions should be calculated for every pair of atoms in the system, that are infinite if we use the pbc treatment.

[1] Image taken from http://isaacs.sourceforge.net/

To overcome this problem the *minimum image convention* is employed. In this convention each atom can "see" only one image of every other atom in the system. Moreover, a cutoff is employed to truncate the interactions between pairs of atoms whose distance is above the threshold. The cutoff has not to be so large that a particle sees its own image in the neighboring box, it means that the maximum limit of the cutoff value is half the length of the simulation box.

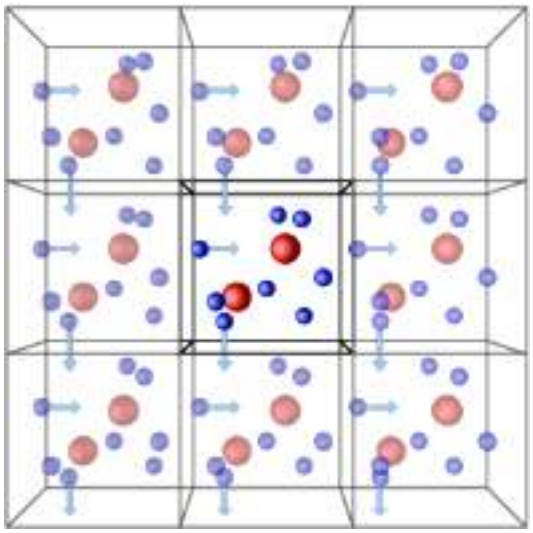


Figure 1.2: Periodic boundary conditions. [1]

1.2 Coarse Grain

The Coarse grain (CG) approach consists of replacing an atomistic description of a biological molecule with a lower-resolution coarse-grained model that averages or smooths away fine details. These approaches have been developed in order to investigate greater time- and length-scale dynamics that are not accessible through atomistic simulations, such as lipid membranes and large proteins (Figure 1.3). CG usually reduces the representation of the system, lowering the degrees of freedom. This allows one to increase the time step of the simulation up to one order of magnitude.

There are different choices to reduce degrees of freedom. If we think about proteins, we can represent an aminoacid with only the heavy atoms or with one or two united atoms per residue. For this scope several different coarse grained force fields exist, with different levels of resolutions. In this PhD thesis we will focus on the MARTINI force field, arguably the most widely used coarse grain model.

[1] Image taken from http://www.chemgapedia.de

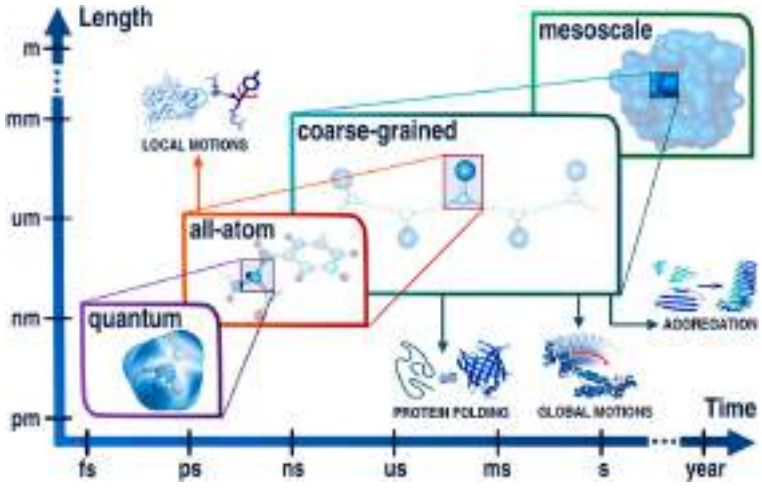


Figure 1.3: Application ranges for MD at different resolutions. Image taken from Ref [5].

1.2.1 MARTINI coarse grain force field for biomolecules

A large variety of CG approaches is available, and typically they are parametrized based on comparison to atomistic simulations. MARTINI force field [6–8] has also been developed on the basis of atomistic simulations. However the philosophy of the force field developed by Marrink *et al.* is a bit different. Instead of focusing on an accurate reproduction of structural details, MARTINI aims at a broader range of applications. The force field is calibrated against experimental data, in particular thermodynamic data such as oil/water partitioning coefficients. In fact, processes as lipid self-assembly, peptide membrane binding and protein-protein recognition depend critically on the degree of partition of the constituents between polar and non-polar environments.

Mapping The MARTINI model is based on a four-to-one mapping, meaning that, on average, 4 heavy atoms plus associated hydrogens are represented by a single interaction center. Mapping of water follows the same principle, so four real water molecules are mapped into a CG water bead. Ions are represented by one CG bead which represents both the ion and its hydration shell. In order to represent ring-like molecules the four-to-one mapping is too coarse, therefore this type of molecules are mapped with an higher resolution of up to two heavy atoms per interaction center. The force field has four main type of particle: polar, non-polar, apolar and charged. Within each type, subtypes are distinguished based on hydrogen bonding capability or degree of polarity. Some examples of MARTINI mapping are shown in Figure 1.4.

Non-bonded interactions Non-bonded interactions are described by a Lennard-Jones 12-6 potential. In addition to the LJ interactions charged groups bear a charge

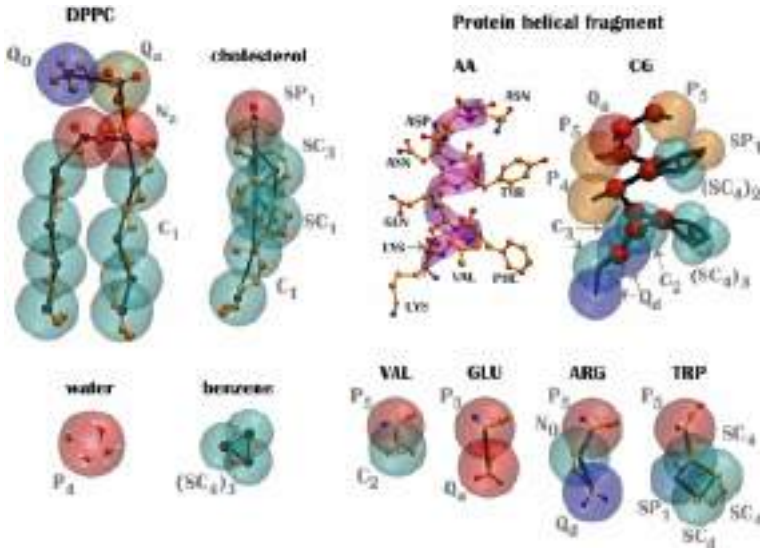


Figure 1.4: Mapping between the chemical structure at the atomistic level (AA) with the coarse-grained Martini model for DPPC, cholesterol, water, benzene, a protein helical fragment, and a few amino acids (valine, glutamic acid, arginine, and tryptophan). The CG beads are shown as transparent vdW spheres. Image taken from Ref [9].

and interact via coulombic energy function. The non-bonded interactions are cut off at a threshold distance. These interactions have been parametrized based on experimental thermodynamic data such as free energy of hydration, vaporization and partitioning free energies between water and a number of organic phases.

Bonded interactions Bonded interactions are described by a set of energy functions analogous to the ones of a classical force field (eq. 1.5). The parametrization was carried out with structural data obtained from atomistic simulations. The Protein Data Bank has been used as an additional source for atomistic geometries for proteins.

Simulations parameters MARTINI simulations, due to the reduced degrees of freedom we are considering, are stable with time steps ranging from 20 to 40 fs.

1.3 Free energy calculations: MM-PBSA approach

The Gibbs free energy (G) provides the most important information about the relationship between the structure, function and stability of biomolecules. Over the past decade several computational approaches have been developed for the calculation of binding free energy. While choosing an approach for free energy calculations, two key points should be considered: accuracy and computational cost of the method. The fastest methods for prediction of binding free energies are the empirical or

knowledge-based approaches, which are based on simple score functions [10, 11]. The main disadvantage of the above mentioned approaches is the lack of conformational sampling which makes this methods faster at the cost of accuracy.
On the other hand, there are the most time-consuming and accurate methods, which are based on force fields. They use MD or Monte Carlo simulations to generate ensemble averages. Among these method free energy perturbation [12] and thermodynamic integration [13] give accurate estimates but require a lot of time, thus it is difficult to estimate binding free energies for large systems.
Since the aforementioned methods are computationally expansive, some more affordable yet reliable methods have been developed. In this thesis we will use the Molecular Mechanics-Poisson Boltzmann Surface Area (MM-PBSA) approach [14], which combines the molecular mechanics and the PBSA model in order to compute $\Delta G_{binding}$ of two species, using structural informations. This method is based on the generation of different structures from an MD simulation in explicit solvent and on the transfer of these structures in an implicit solvent.
Let's consider the binding of a ligand (L) and of a receptor (R) forming a complex (C). In principle we could compute the $\Delta G_{binding}$ in only one passage, using the explicit representation of the solvent. However the major contribution to this calculated energy would arise from the interactions between the solvent molecules, with the resulting fluctuation of the energy value (virtually greater than the $\Delta G_{binding}$ itself). Moreover the time scale associated to this process would be too large to be computationally simulated. In order to bypass this problem we can determine the $\Delta G_{binding}$ exploiting the thermodynamic cycle showed in Figure 1.5, using the implicit solvent model.

Briefly, the binding free energy of a protein to a ligand molecule in solution can be defined as shown in eq. 1.7.

$$\Delta G_{binding} = G_{complex} - (G_{protein} + G_{ligand}) \tag{1.7}$$

MD simulation generates a thermodynamically weighted ensemble of structures. The free energy term is calculated as an average over the considered structures, as shown in eq. 1.8.

$$\langle G \rangle = \langle E_{MM} \rangle + \langle G_{solv} \rangle - T\langle S_{MM} \rangle \tag{1.8}$$

The energetic term E_{MM} is defined as:

$$E_{MM} = E_{int} + E_{coul} + E_{LJ} \tag{1.9}$$

where E_{int} represents bonds, angle and torsional angle energies while E_{coul} and E_{LJ} denote the electrostatic and Lennard-Jones energies respectively.

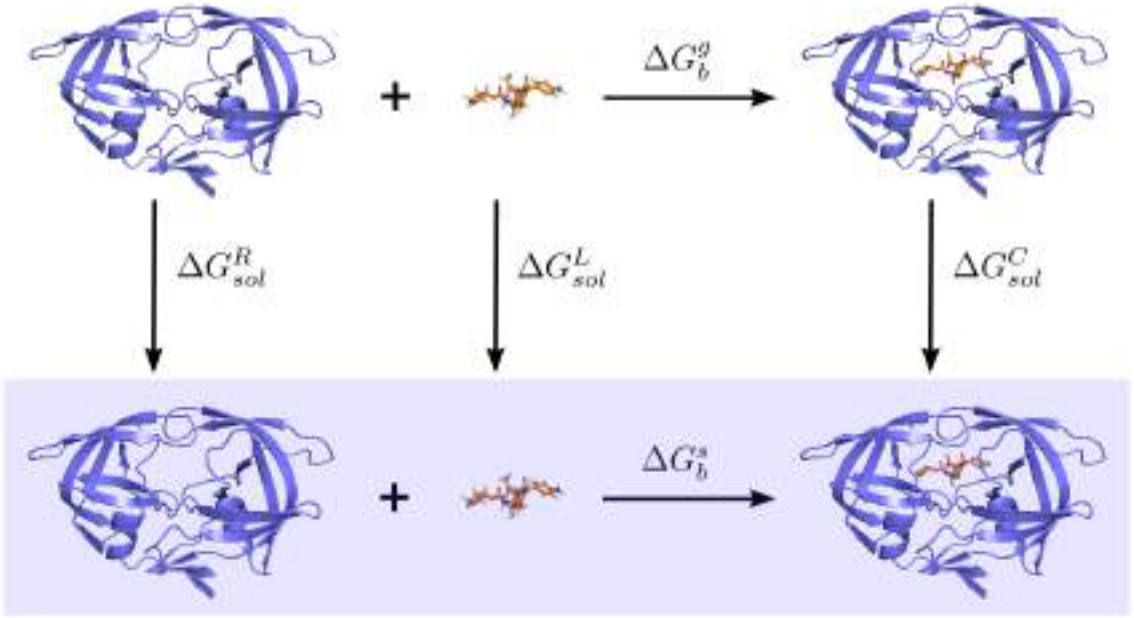


Figure 1.5: The thermodynamic cycle used to calculate $\Delta G_{binding}$ of the complex in solution. The binding free energy is decomposed into binding free energy in vacuum and solvation free energy of the reacting species.[1]

The single trajectory method employed in this thesis to calculate binding energies implies that the trajectories of the protein and the ligand are directly extracted from the complex structure thus zeroing the E_{int} term [15].
The total binding free energy that we want to calculate is the one called ΔG_b^s in Figure 1.5, and we can calculate it using the thermodynamic cycle:

$$\Delta G_S^b = \Delta G_b^g + \Delta G_{SOL}^C - \Delta G_{SOL}^R - \Delta G_{SOL}^L \tag{1.10}$$

1.3.1 Solvation contribution

This contribution is divided in two components: polar and apolar, as showed in eq. 1.11.

$$\Delta G_{solv} = \Delta G_{polar} + \Delta G_{apolar} \tag{1.11}$$

Both contributions in eq. 1.11 can be calculated using the APBS (Adaptive Poisson Boltzmann Solver) software [16].

The thermodynamic cycle shown in Figure 1.6 incorporates several processes into the solvation energy (step 1). Step 2 indicates charging of the solute in solution (e.g., inhomogeneous dielectric, ions present). Step 3 indicates the introduction

[1]Image taken from https://commons.wikimedia.org/

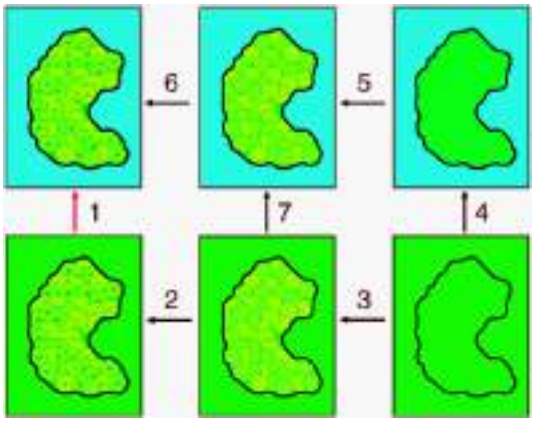


Figure 1.6: The thermodynamic cycle used for the calculation of ΔG_{solv}.[1]

of attractive solute-solvent dispersive interaction interactions (e.g., an integral of Weeks-Chandler-Andersen interactions over the solvent-accessible volume). Step 4 indicates the introduction of repulsive solute-solvent interaction (e.g., cavity formation). Steps 5 and 6 are basically null steps although they could be used to offset unwanted energies added in Steps 3 and 4 above. Finally, Step 6 represents the charging of the solute in a vacuum or homogeneous dielectric environment in the absence of mobile ions.

Polar solvation The full free energy cycle is usually decomposed into polar and nonpolar parts. The polar portion is represented by the charging energies in steps 2 and 6:

$$\Delta\Delta G_{polar} = \Delta G_2 - \Delta G_6 \tag{1.12}$$

Energies returned from APBS electrostatics calculations are charging free energies. Therefore, to calculate the polar contribution to the solvation free energy, we simply need to setup two calculations corresponding to Steps 2 and 6 in the free energy cycle. The electrostatic charging free energies returned by APBS include self-interaction terms. These are the energies of a charge distribution interacting with itself. Such self-interaction energies are typically very large and extremely sensitive to the problem discretization (grid spacing, location, etc.). Therefore, it is very important that the two calculations in Steps 2 and 6 are performed with identical grid spacings, lengths, and centers, in order to ensure appropriate matching (or "cancellation") of self-energy terms. These polar contributions are calculated with the Poisson-Boltzmann equation (PBE):

$$-\nabla \cdot \epsilon(x)\nabla\Phi(x) + \vec{\kappa}^2(x)sinh\Phi(x) = f(x) \tag{1.13}$$

[1]Image taken from https://sites.google.com/a/poissonboltzmann.org/

PBE is a non-linear partial differential equation of the second order. It relates the electrostatic potential (Φ) to the dielectric properties of the solute and solvent (ϵ), to the ionic force of the solution, to the accessibility of ions in solute ($\vec{\kappa}^2$) and to the partial charges distribution of the solute (f). In order to speed up the calculation of this non-linear equation the Linearized PBE approximation has been used, assuming $sinh\Phi(x) \approx \Phi(x)$.

Apolar contribution Referring back to the solvation free energy cycle (Figure 1.6), the nonpolar solvation free energy is usually represented by the energy changes in Steps 3 through 5:

$$\Delta\Delta G_{apolar} = (\Delta G_3 - \Delta G_5) + \Delta G_4 \tag{1.14}$$

where Step 4 represents the energy of creating a cavity in solution and Steps 3-5 represent the energy associated with dispersive interactions between the solute and solvent. The apolar contribution (ΔG_4) is calculated by APBS, using eq. [17]:

$$\Delta G_4 = pV + \gamma A \tag{1.15}$$

where p is the solvent pressure, V is the solute volume, γ is the solvent surface tension and A is the solute surface area.
The dispersion terms (ΔG_3 and ΔG_5) follow a Weeks-Chandler-Anderson framework [18] (eq. 1.16):

$$\Delta G_3 - \Delta G_5 = \vec{\rho} \int_\Omega u^{att}(y)\theta(y)dy \tag{1.16}$$

where $\vec{\rho}$ is the bulk solvent density, Ω is the problem domain, $u^{att}(y)$ is the attractive dispersion interaction between the solute and the solvent at point y and $\theta(y)$ describes the solvent accessibility of point y.

1.3.2 Lennard-Jones contribution

E_{LJ} has been computed using GROMACS package [19]. Lennard-Jones short range and 1-4 interactions have been calculated and added for every snapshot extracted from the MD, obtaining E_{LJ} term for complex, protein and ligand.

1.3.3 Electrostatic contribution

E_coul represent the sum of pairwise Coulombic interactions between all atoms in the molecule (or complex) for a particular uniform dielectric. In order to combine these Coulombic binding energies with the solvation energies described above, we need to make sure consistent dielectric constants are used. In particular, Coulombic

interactions should be calculated using the same uniform dielectric constant as the reference state of the solvation energy above. In both cases a dielectric constant of 2 has been used.
The electrostatic contribution has been computed with coulomb software equipped in APBS package. This software, starting from "*.pqr*" file containing positions, atomic charges and radii applies Coulomb's law (eq. 1.17) to calculate electrostatic energy.

$$U_{coul} = \frac{1}{4\pi\epsilon_0\epsilon_r} \sum_{i,j} \frac{q_i q_j}{r_{ij}} \tag{1.17}$$

1.4 Protein-protein interactions

Proteins are responsible for nearly every molecular transformation of the cellular metabolism. However, it is unlikely for a protein to act alone: approximatively 80% of proteins in a cell don't stand alone, but rather interact with other proteins through protein-protein interactions in order to perform biological functions. Protein-protein interactions (PPIs) play an important role in different cellular functions, raging from complex formation, signalling pathways and biochemical reactions. They are defined as the physical conctacts between two or more proteins. In order to have any biological function these interactions have to be formed with proper partners at the right time and locations [20]. At any time, a human cell may contain about 130k binary interactions between proteins [21]. Any misfunction or alteration in PPIs network can lead to various diseases, including cancer [22]. The detailed understanding of this network of interactions can help to elucidate the molecular basis of different diseases making PPIs a potential target for drug discovery.

1.4.1 Structural characteristics of protein protein interfaces

In principle, protein-protein binding can be driven by both polar and electrostatic interactions. In practice, however, the binding derives from the mutual burying of hydrophobic surfaces. At the interface aromatic residues such as Tyr, Trp, Phe and His are more frequent than in the average protein surface (21% vs 8%). Moreover, interfaces are richer in aliphatic residues such as Leu, Val, Met and Leu and depleted in charged residues such as Glu, Asp and Lys. Surprisingly Arg is the residue making the largest overall contribution to interfaces, 10% [23]. The surface of interaction is usually very large, ranging from 1500 to 3000 $Å^2$ for each protein.
We can classify PPIs as permanent or transient, according to their binding lifetime.

- In permanent PPIs the two proteins bind strongly, and the interface is large and flat. In this case the packing is tighter and very few water molecules are

trapped between monomers. The surface of interaction is mainly hydrophobic. An example is the protein-protein complex between hemoglobin and RNA polymerase [24].

- In transient PPIs the interactions are temporary in nature: the binding partners can associate and dissociate only in particular physiological conditions.

1.4.2 Classes of atoms at protein-protein interface

The interfaces are composed of various types of atoms:

- **Interface atoms** are atoms that lose solvent accessible surface upon the protein-protein binding. Not all the atoms make vdW contacts across the interface.
- **Contact atoms** (Figure 1.7) are the subset of interface atoms that actually make vdW contacts with the binding partner. They can be further classified in fully buried and partially buried contact atoms.
  - Fully buried contact atoms are accessible in the free monomers but have zero surface accessible surface area in the complex.
  - Partially buried contact atoms are partially accessible to solvent.

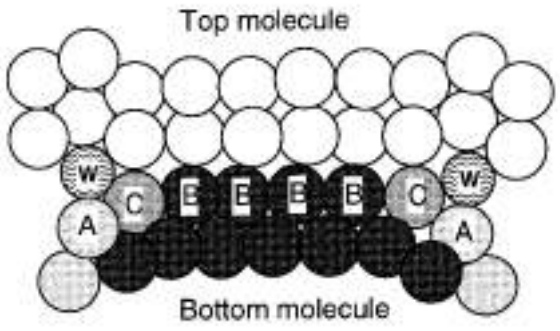


Figure 1.7: Side view of classes of interface atoms. All atoms of the bottom molecules that lose solvent accessibility in the presence of the binding partner are interface atoms. In particular type B are fully buried while type A and C retain partial solvent accessibility. Types B and C are contact atoms, as the make vdW contacts with the top molecules. Image taken from Ref [23].

1.4.3 Hot-spots

Not all residues at the protein-protein interface equally contribute to the binding free energy. Only a small fraction of them, around 10%, are responsible for the

binding of the proteins and they are called hot-spots [25,26]. Single point mutations of one of these residues often greatly reduce affinity for the partner protein, without changing the overall integrity of the protein [27]. Hot-spots can be experimentally identified using Alanine Scanning. In this approach residues located at the protein-protein interface are mutated into alanine and the difference in binding free energy ($\Delta\Delta G$) is measured. Computationally we use the same kind of approach, explained in Section 1.5.

1.4.4 PPIs in drug design

PPIs are an attractive emerging class of molecular targets as they regulate multiple critical cellular functions such as DNA replication, transcriptional activation, translation and transmembrane signal transduction [28]. In particular, understanding how protein-protein complexes work has allowed to elucidate the molecular basis of different diseases. Aberrant PPIs or the loss of an essential interactions through the formation or stabilization of protein complexes can cause diseases. However PPIs are quite different from enzyme-inhibitor or protein ligand interactions and designing small molecules able to target protein-protein interface is a challenging task. First of all, protein-protein interfaces usually lack of binding pockets. The interface is usually large, approximately 1500-3000 Å^2 and it is flat. [29]. Moreover, unlike enzymes, every PPI has to be considered as a unique case, and the binding sites does not appear to be preserved, even among very similar interfaces [30]. It has been shown that, for the modulation of PPIs the inhibitor, it is not required to bind a large area of the interface: it is enough for the molecule to interact with a small subset of hot-spots.

In principle a modulator of PPIs, to be effective, can act in two different ways: reduce the strength of the PPI or increase the binding of the complex in order to stabilize it. So far the vast majority of the known PPI modulators act as inhibitors [31,32], however in the last years also stabilizing PPI agents are gaining increasing attention [33].

Molecules that can selectively interfere with the formation and/or stabilization of protein-protein complexes serve as powerful research tools in chemical biology. Three are the main classes of PPI modulators which can be identified: humanized monoclonal antibodies, peptides and peptidomimetics and small molecules. Humanized monoclonal antibodies are the vast majority of PPI inhibitors in clinic, due to their high target specificity [34]. Their principal drawbacks are the poor cell and blood–brain barrier permeability properties and the high manufacture cost. Peptide and peptidomimetics derived from protein-protein interface (Chapter 3) are often the starting point for the development of small molecules PPI modulators. Moreover, they can be used as molecular probes in order to investigate the mechanism

of PPIs modulation and for the validation of novel PPIs as targets for drug discovery. However, peptides suffer from poor metabolic stability and low bioavailability. Finally, small-molecules PPI modulators are more amenable from a medicinal chemistry point of view but they are very difficult to obtain.
To identify a small-molecule PPI modulator we can use three main approaches:

- Peptide and peptidomimetics approaches.
- High throughput screening of chemical libraries and natural products.
- Computational approaches based on X-ray crystallographic data.

In this thesis we discuss two of these three approaches. Peptide and peptidomimetics approach consists in selecting an amino-acidic sequence at the protein-protein interface and use it as a lead for the development of novel and improved PPI modulators (Chapter 3). Computational approaches based on X-ray crystallographic data, such as virtual screening (Chapter 4), can lead to the identification of novel PPIs inhibitors through the screening of a large amount of known compounds.

1.5 Computational alanine scanning

Alanine scanning (AS) mutagenesis is an experimental procedure to identify hot-spots at the protein-protein interface [35]. It is generally a slow and labor-intensive procedure since it requires DNA engineering, protein expression and subsequent purification. For these reasons the alanine scanning is applied only to preselected regions; blind AS for explorative purposes is unfeasible. However, we can use the same technique *in silico*: Computational Alanine Scanning (CAS). This technique is able to identify hot-spots evaluating the binding free energy between the subunits making up the complex (ΔG_{WT}). Afterwards the binding energy upon mutation of each of the interfacial amino acid X into an alanine ($\Delta G_{Mut,X}$) is calculated. Finally, for each residue X, a $\Delta\Delta G$ can be obtained as:

$$\Delta\Delta G_X = \Delta G_{Mut,X} - \Delta G_{WT} \tag{1.18}$$

Many protocols exists to perform CAS [15, 36]. They mainly differ for the simulations conditions, such as the use of one or multiple trajectories for wild type and mutated proteins or for the use of implicit/explicit solvent. If we consider the number of MD simulations, we can distinguish three approaches:

- **Single trajectory CAS** uses only the MD simulation of the wild type complex. The structures are then extracted and post-processed in order to obtain the mutated complexes. Monomers are obtained from the same simulation discarding the partner protein from the complex structure.

- **Double trajectory CAS** uses two simulations: one for the wild type and one for the mutated complex. Monomers are obtained as in the former case.
- **Quadruple trajectory CAS** uses four MD simulations: wild type and mutated complex and the two monomers.

In recent studies it has been demonstrated that **single trajectory CAS** generally has better results due to a convenient error cancellation. In multiple trajectory the error due to insufficient sampling strongly affect the results. Moreover using a single trajectory allows to neglect the entropic terms of the $\Delta\Delta G$ because the identity of the structures leads to the mutual cancellation of this term. Regarding the solvation there isn't an agreement about which approach would achieve better results. While, in general, a better description of the solvent is preferred, supporters of implicit solvent claim that, since the calculation of the solvation term in MM-PBSA method is done with implicit solvent, it is better to perform MD in the same conditions. In this work the single trajectory CAS in explicit solvent is carried out, using MM-PBSA method for the calculation of the binding free energy.

Mutating all interfacial amino acids and calculate the $\Delta\Delta G_X$ in the aforementioned way yields a map predicting which residues are important for the ligand (or protein) binding and which are not. This amino acids tend to be grouped into small clusters on the interface and they are classified on the basis of difference in binding free energy between the wild type and the mutant type as "*hot spots*" ($\Delta\Delta G > 4\ kcal\ mol^{-1}$), "*warm spots*" ($2\ kcal\ mol^{-1} < \Delta\Delta G < 4\ kcal\ mol^{-1}$) and "*cold*" or "*null spots*" ($\Delta\Delta G < 2\ kcal\ mol^{-1}$) [37]. Figure 1.8 represents the procedure for the CAS computation.

Some assumptions have been made in the calculation of $\Delta\Delta G$ with the CAS approach, which are follows:

- Point mutations in the protein do not significantly affect the conformation of the mutated protein. The validity of this assumption in computational alanine scanning has been widely confirmed in literature when it has been applied both to protein–protein and to protein–small molecule complexes [15, 38, 39].
- The terms in the energy function used for the calculation of ΔG (Section 1.3) are pairwise additive. This assumption is necessary for the fast computational evaluation of the binding free energy.
- Cofactors, metal ions and water molecules participating in hydrogen bonding at the protein interface are not taken into account
- Proline residues are not mutated into an alanine and thus not considered for CAS. This is because its backbone conformation is different from the one of

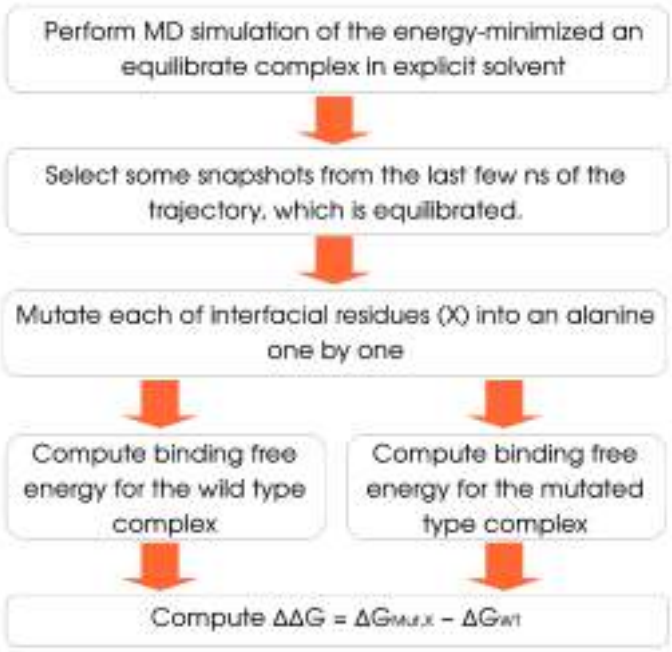


Figure 1.8: Flowchart illustrating the procedure of the computational alanine scanning approach.

alanine, and the mutation would lead to significant conformational changes [40].

1.6 Root mean-square fluctuation

The Root Mean Square Fluctuation (RMSF) (eq. 1.19) is the measure of the deviation between the position of an atom i during the dynamic and its time-averaged position.

$$RMSF = \sqrt{\frac{1}{T}\sum_{t_j=1}^{T}(x_i(t_j) - \bar{x}_i)^2} \tag{1.19}$$

where T is the time over which one wants to average and $\bar{x}_i$ is the time-averaged position of the atom i.

1.7 Essential dynamics

Essential Dynamics (ED) [41] is a very powerful analysis technique which exploits a Principal Component Analysis (PCA) [42] to identify the nature and relative importance of the essential motions of a macromolecule from a MD sampling. The result of an ED analysis is a set of eigenvectors, which describe the nature of motions in

the Cartesian space, and eigenvalues, which represent the amount of variance described by each eigenvector. The ED method is able to separate the few large mainly unharmonic motions (the essential subspace) from the small Gaussian fluctuations (the near-constraints subspace) [43]. The motions in the essential space are linked to the biological functions of the protein.
Given a MD trajectory, the overall translational and rotational motions have to be eliminated. The result is a Cartesian molecular coordinate system in which atomic motions can be expressed. The internal motion is now described by a trajectory $\vec{x}(t)$, where $\vec{x}$ is the 3N-dimensional vector of all atomic coordinates, represented by a column. The correlation between atomic motions can be expressed in the covariance matrix C of the positional deviations, whose element C_{ij} expression is shown in eq. 1.20:

$$C_{ij} = \langle M_{ii}^{\frac{1}{2}}(x_i(t) - \langle x_i \rangle) M_{jj}^{\frac{1}{2}}(x_j(t) - \langle x_j \rangle) \rangle \tag{1.20}$$

where the brackets denote an average over time. The symmetric 3Nx3N covariance matrix can be diagonalized by an orthogonal coordinate transformation R:

$$R^T C R = diag(\lambda_1, \lambda_2, ..., \lambda_{3N}) \tag{1.21}$$

The columns of R are the eigenvectors or principal mode. The eigenvalues λ represent the variance in the direction of the corresponding eigenvector. The original MD can now be projected on the eigenvectors, in order to obtain the principal components p_i, i=1,...,N [44].

$$\vec{p}(t) = R^T M^{\frac{1}{2}}(\vec{x}(t) - \langle \vec{x} \rangle) \tag{1.22}$$

Carrying out the visual inspection of the trajectory projected on the eigenvectors we can highlight the most important collective motions of the protein during the dynamics.

1.8 Molecular Docking

Docking is a computational method able to predict the structure (or structures) of the intermolecular complexes formed by a ligand and a receptor (usually a protein). Docking can be employed to investigate the binding modes of protein inhibitors. The docking algorithm consists of two distinct phases: **posing** and **scoring**. In the posing phase the program generates a large number of possible structures. Once the structures have been generated, we require a mean to score each structure and identify the most lowest energy ones [1].
The are two ways to perform a molecular docking calculation: flexible and rigid. The **flexible docking** involves many degrees of freedom: there are six translational

and rotational degrees of freedom of one molecule relative to the other as well as the conformational degrees of freedom of each molecule. The **rigid docking** considers the ligand and the receptor as rigid objects that cannot change their conformation during the docking process. Various docking algorithms have been developed in the last few decades, and they are more complex the more degrees of freedom they consider. In this PhD thesis the DOCK6 [45] algorithm was employed.

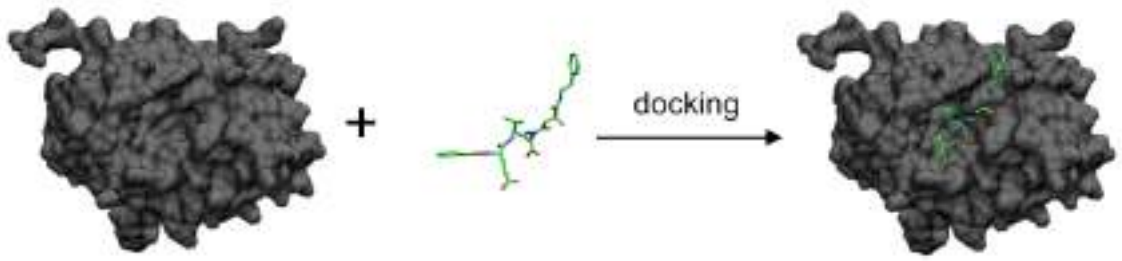


Figure 1.9: Schematic illustration of molecular docking.[1]

1.8.1 DOCK6 algorithm

Posing DOCK is designed to find molecules with high degree shape complementarity to the binding site. First of all the program builds a negative image of the protein binding site, that consists of a collection of overlapping spheres with varying radii. This negative image is built by rolling a ball with the size of a water molecule over the van der Waals surface of the target. Each sphere touches the molecular surface at just two points and has its radius along the surface normal (Figure 1.10). Spheres are calculated over the entire surface and subsequently this very dense representation is filtered to keep only the largest sphere associated with each receptor surface atom. Ligand atoms are then associated to the sphere centers in order to find matching sets in which all the distances between the ligand atoms in the set are equal to the corresponding sphere center-sphere center distances (within a certain tolerance).

To include the internal degrees of freedom of the ligand, DOCK uses the anchor-and-grow incremental construction approach [47]. This kind of approach separates the ligand flexibility in two steps. The largest rigid structure of the ligand (anchor) is identified and rigidly oriented in the active site by matching its heavy atoms centers to the receptor sphere centers. The anchor orientations are optimized using the scoring function. Then, the remaining flexible portions of the ligand are build onto the best anchor orientation within the context of the receptor.

[1]Image taken from https://commons.wikimedia.org/

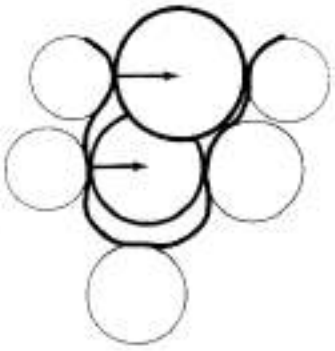


Figure 1.10: Schematic representation of a binding site (thick line) formed from five atoms (thin circles). Image taken from Ref [46].

Scoring DOCK scoring function is based on the AMBER [3] force field. Only the interactions between the ligand and protein are considered, leaving only intermolecular van der Waals and electrostatic components in the function. Since the receptor is considered to be rigid, the receptor contribution to the potential energy can be pre-calculated and stored on a grid. These approximations enable the program to evaluate large libraries of small molecules against a receptor in a reasonable period of time.

1.8.2 Virtual Screening

Virtual screening is a computational technique that consists in the search of libraries of chemical compounds that could be able to bind a drug target, usually proteins or enzymes. There are plenty of free database of commercially available compounds that can be used as a starting point for a virtual screening. Typically there are two major categories of screening technique: ligand based and structure based.

Ligand based If we already know the structure of some ligands that bind to the receptor, their pharmacophore can be obtained. Then, using the pharmacophore model, one can scan the database of compounds in order to find molecules that are compatible with the model.

Structure based Structure based virtual screening involves the docking of a database of compounds in the binding site of the receptor. Every docking pose is evaluated with a scoring function to asses the ability of the ligand to bind in that particular site.

1.9 Preferential Coefficient and Solvent Density Function

The presence of cosolvents usually alters the chemical potential of the protein. They can associate more strongly or more weakly with the protein with respect to water. This phenomenon is called *preferential binding* and it is of a great importance because it has an effect on the physical and chemical properties of proteins. In eq. 1.23 is shown how a consolvent alters the chemical potential of a protein.

$$\begin{aligned}\Delta\mu_P^{tr} &= \int_0^{m_x} \left(\frac{\partial\mu_P}{\partial m_X}\right)_{m_P} dm_X \\ &= -\int_0^{m_x} \left(\frac{\partial\mu_X}{\partial m_X}\right)_{m_P} \left(\frac{\partial m_X}{\partial m_P}\right)_{\mu_X} dm_X\end{aligned} \tag{1.23}$$

where $\Delta\mu_P^{tr}$ is the transfer free energy of the protein (P) from pure water into the water and cosolvent (X) system and m is the molality. The first partial derivative in eq. 1.23 represents the dependence of the protein's chemical potential on the cosolvent molality. The second is the preferential coefficient Γ_{XP}.

$$\Gamma_{XP} \equiv \left(\frac{\partial m_X}{\partial m_P}\right)_{\mu_X} \tag{1.24}$$

The preferential coefficient Γ_{XP} measures the ratio of cosolvent molecules in the neighbourhood of the protein with respect to the bulk. According to Baynes and Trout [48], Γ_{XP} can be evaluated defining two domains (Figure 1.11) as shown in eq. 1.25.

$$\Gamma = \left\langle n_X^{II} - n_W^{II}\left(\frac{n_X^I}{n_W^I}\right)\right\rangle \tag{1.25}$$

Where $n_{X,W}^{I,II}$ is the number of water (W) or osmolyte (X) molecules in the bulk domain (I) and local domain (II). Γ_{XP} is independent of the choice of the boundary between the domains if the boundary is far enough from the protein.
The preferential coefficient is greater than zero if the amount of cosolvent in the neighbourhood of the protein is higher than in the bulk. This behavior is verified for denaturants such as urea. The opposite behavior happens with osmoprotectants.
Solvent Density Function (SDF) describes how the cosolvent is distributed around the protein. The idea behind this function is the same of radial distribution function. The main difference is that, in the case of the SDF, the shape and the volume of the protein are taken into account. The function for a generic

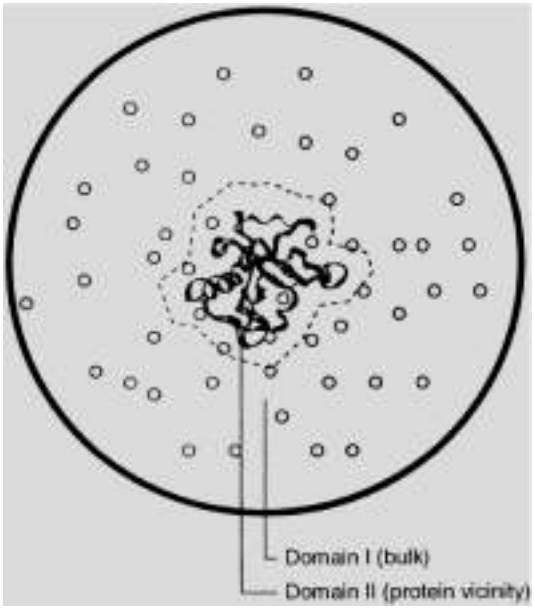


Figure 1.11: The two different domains defined in order to calculate the preferential coefficient. Image taken from Ref [48].

molecule X is computed as shown in eq. 1.26.

$$\rho_X(r) = \frac{X(r, r')}{V(r, r')} \tag{1.26}$$

Where r represent the radius of the solvation shell, $X(r, r')$ is the number of cosolvent molecules found from r to r', and $V(r, r')$ is the volume of the shell r, r'.

Part I

Molecular modeling of Ebola virus inhibitors

Ebola virus

Ebola Virus (EBOV) belongs to the *Filoviridae* family of single stranded, non-segmented negative-sense RNA viruses. The shape of the virus is variable, usually appears as long tubes with many turns and branches (Figure 1.12).

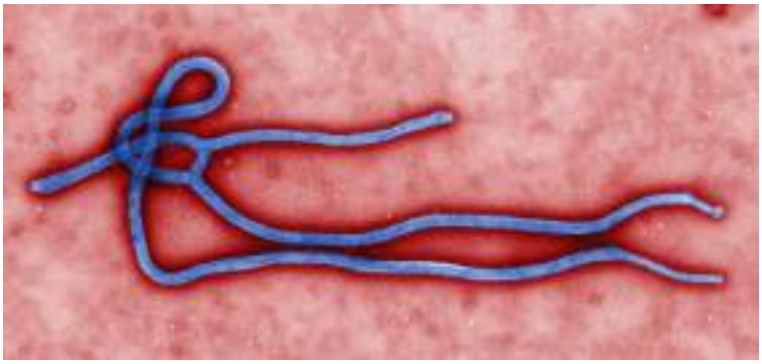


Figure 1.12: Morphology of Ebola virus. [1]

Ebola virus genome codes for seven proteins. The seven genes are for the nucleoprotein, the viral proteins VP24-VP30-VP35-VP40, L (polymerase) and the glycoprotein (Figure 1.13). The surface glycoprotein is expressed in two molecular forms, GP1 and GP2 that are generated by editing the RNA. It has a fundamental role in the pathogenesis of the virus and in the infection, its replication is tightly regulated during virus replication. The nucleoprotein embeds the genetic material and interacts with VP35 and VP30 forming a large protein-protein complex that is involved in the synthesis of virus RNAs. VP40 and VP24 are localized in virus matrix space [49–52].

Only one of these seven protein, the viral large (L) RNA-dependent RNA polymerase protein, has an enzymatic activity, while all the remaining EBOV proteins exert their function through protein–protein interactions, among themselves or with human proteins.

Five species of Ebola virus are known and they are named after the region where they have been identified for the first time: Bundibugyo, Reston, Sudan, Taï Forest and Zaire.

[1] Image taken from https://www.cdc.gov/

Zaire Ebola virus is the causative agent of haemorrhagic fever with a very high human fatality rate near 90% [53]. Recently EBOV has spread in several countries in Africa and a few sporadic cases were registered also in Europe and in the US. The lack of effective therapies or vaccines, combined with this high mortality rate, makes urgent the need to develop antiviral drugs against EBOV.
The virulence and high lethality of this virus are due to different factors, in particular to its ability to inhibit both the innate immune response in the early stages of infection and the subsequent adaptive specific immune responses of the host organism [54, 55]. This is done by different strategies, in particular by the suppression of interferon (IFN)-α/β production and inhibition of interferon induced antiviral activity [56, 57].
In this part we will focus our attention on Viral Protein 35 (VP35) and Viral Protein 24 (VP24).

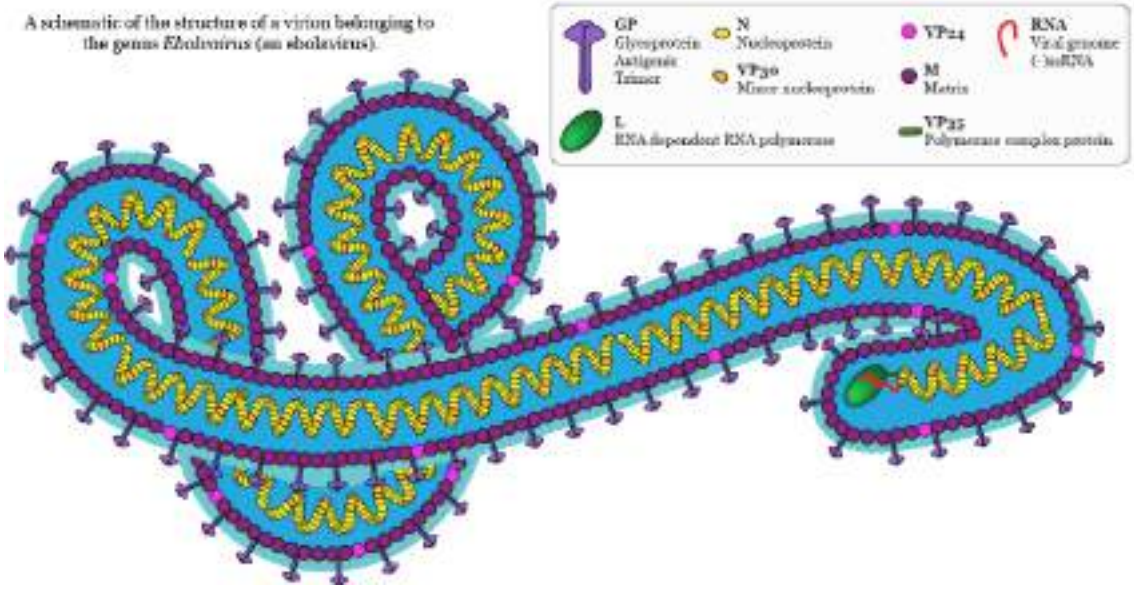


Figure 1.13: Schematic representation of Ebola Virus virion.[1]

[1]Image taken from: http://virologydownunder.blogspot.it/

Chapter 2

In silico study of VP35 inhibitors

2.1 Viral Protein 35

Viral Protein 35 (VP35) is one of the seven proteins coded by Ebola genome and it carries out multiple functions, including the inhibition of RNA dependent protein kinase (PKR) [58,59], suppression of RNA silencing [60] and the inhibition of interferon (IFN)-α/β signaling [61]. The activation of IFN-α/β productions is one of the key steps in the innate immune response to viral infection. In this activation the trascription factor interferon regulatory factor 3 (IRF-3) plays a fundamental role. IRF-3 is a cytoplasmatic protein that becomes hyperphosphorylated on serine and threonine residues, dimerizes, and accumulates in the nucleus, where it participates in initial IFN-α/β gene expression [62–64]. It has been found that VP35 interfers with the activation of IRF-3, thus it is responsible for the inhibition of IFN-α/β production [65]. This makes VP35 a potential therapeutic target.
VP35 consists of a N-terminal coiled-coil domain, required for its oligomerization, and a C-terminal domain, called the interferon inhibitory domain (IID), required for the interaction with the viral Nucleoprotein (NP) and for interferon inhibition. Two basic patches are found in the IID:

- The first basic patch (FBP), important for interactions with EBOV NP and VP35 polymerase cofactor function.
- The central basic patch (CBP), important for VP35 dsRNA binding and IFN inhibition.

The high resolution structure of the C-terminal domain has been recently obtained. It is made up by two different subdomains (Figure 2.1): a N-terminal α-helical subdomain and a C-terminal β-sheet subdomain [66, 67].

In a recent study Brown *et al.* [66] determined the crystallographic structures of VP35 in a complex with nine small molecules able to inhibit the protein–protein

Figure 2.1: VP35 in complex with one of the ligands (GA017). The α-helical subdomain (residues from Ala221 to Arg283) is represented in green and the β-sheet subdomain (residues from Val294 to Ile340) is represented in blue. The binding mode is the same for all ligands.

interaction between VP35 and NP. In this chapter we study at the atomic level the interactions between these inhibitors and VP35 by performing molecular dynamics simulations and computational alanine scanning on these systems, in order to identify the key interactions for the binding of ligands, thus providing hints for the further optimization of active molecules targeting VP35. Then, to better characterize and distinguish the dynamic behavior of the apo-protein from the ligand–protein complexes we measure residues RMSF and perform an essential dynamics analysis. We finally compare the dynamic behavior of the apo-forms of EBOV and Marburg virus VP35.

2.2 Computational details

Molecular dynamics The ligand-protein complexes and the apo-protein structures of EBOV and Marburg Virus (MARV) VP35 were obtained from the Protein Data Bank (PDB codes: 4IBB, 4IBC, 4IBD, 4IBE, 4IBF, 4IBG, 4IBI, 4IBJ, 4IBK [66],3FKE [67],4GH9 [60]). Molecular dynamics simulations were performed with the GROMACS 4.5.3 [19] package using explicit solvent and periodic boundary conditions.
The AMBER99SB-ILDN [68] force field was used for the protein and the Generalized Amber Force Field [69] for ligands. Each complex was solvated with TIP3P [70] waters and neutralized with two chloride ions. The LINCS algorithm [4] was employed to constrain all bonds involving hydrogen to their equilibrium length, allowing a time step of 2 fs. The systems were submitted to 10000 steps of geometry optimization with the steepest descent method. Then they were equilibrated for 200 ps in NVT conditions (T = 300 K) and subsequently for 200 ps in NPT conditions, in order to equilibrate systems density. Then a 25 ns molecular dynamics was performed for every system in NPT conditions (1 bar, 300 K). Temperature and pressure were kept constant to their reference values using the velocity rescale algorithm [71] and the Berendsen barostat respectively [72]. A 14 Å cutoff was applied for non-bonded interactions and the Particles Mesh Ewald algorithm [73] was employed to calculate long range electrostatic interactions.

CAS To perform computational alanine scanning, 250 snapshots were extracted from the last 5 ns of each dynamics (one snapshot every 20 ps). Residues at the ligand–protein interface were selected using Naccess [74]. For each residue at the interface all side chain atoms beyond C_β were removed and the missing hydrogen was added, obtaining an alanine side chain. ΔG of binding was calculated using the MM-PBSA approach (Section 1.3). The Poisson–Boltzmann equation was solved with APBS [16] using a relative dielectric constant of 80 for the region around the

protein and of 2 for the protein interior. The entropic contribution to ΔG is supposed constant in the mutated and wild type structure considering their similarity, so it has not been calculated, as discussed by Kollman *et al.* [36].

Bootstrap root mean-square fluctuations analysis For this analysis we applied the protocol described in the work by Mitra *et al.* [75]. We first determined the number of independent conformations in our MD trajectories by calculating the autocorrelation function of Root Mean Square Deviation (RMSD) of the structures. We found a correlation time of 2.5 ns, corresponding to 10 independent structures in each of our 25 ns trajectories. In order to calculate the RMSF we selected 10 random points with replacement, from every dynamics and repeated it 200 times, giving us a mean RMSF value and a standard error of the mean for each residue in the protein. Using these values we applied a student's t-test to the data and identified contiguous regions that showed significant differences at a significance level of $p < 0.0005$.

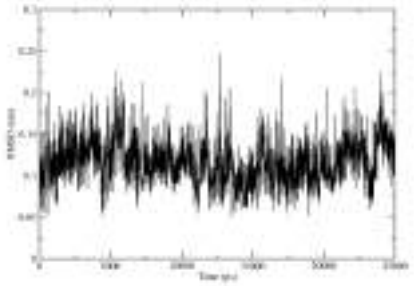
Essential dynamics The covariance matrix and its diagonalization were calculated with the *g_covar* utility of GROMACS 4.5.3, the analysis of the eigenvalues and eigenvectors with *g_anaeig*. The visual inspection of the projection on the first 3 eigenvectors was carried out using the VMD software [76].

2.3 CAS

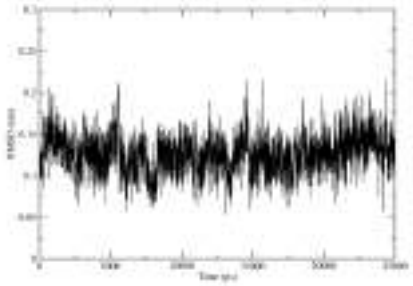
Computational Alanine Scanning on the residues at the protein–ligand interface was carried out in the last 5 ns of the molecular dynamics trajectories, when the system was fully equilibrated in all cases, as shown by RMSD plots (Figure 2.2 and Figure 2.3).

In general, once equilibrium has been reached, trajectories 5 ns long guarantee enough sampling for an accurate binding free energy calculation with the MM-PBSA approach and consequently for computational alanine scanning purposes [77,78]. All the complexes exhibit a negative protein–ligand $\Delta G_{binding}$ during the simulations (Table 2.1).

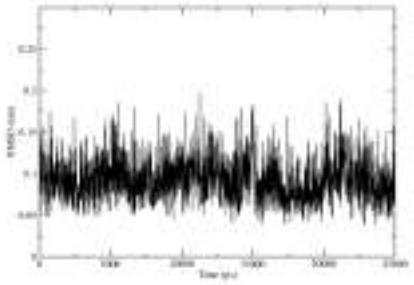
The interface is defined as the ensemble of amino acids whose solvent exposed surface area has a non-zero variation upon complex formation. Ligand molecules (shown in Figure 2.4) share the same binding site and 18 residues in total were found to be at the interface for each molecule. CAS results are summarized in Table 2.2.



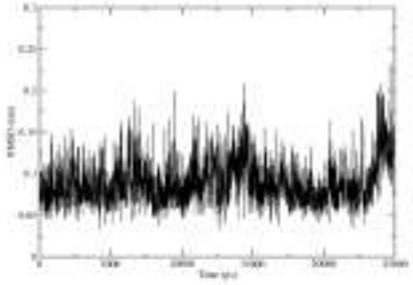
GA017



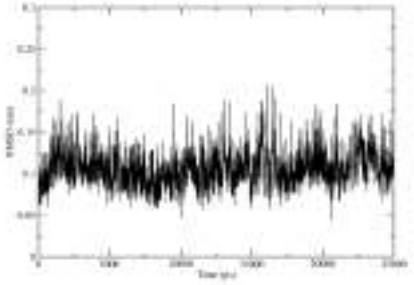
GA246



VPL27



VPL29

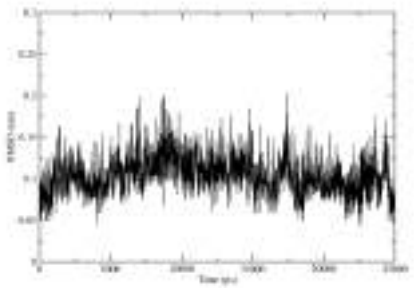


VPL42

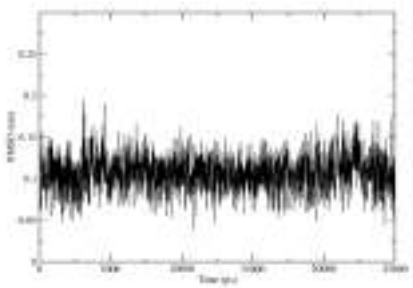
Figure 2.2: RMSD values fitting protein C_α for every protein ligand complex (1/2).

Name	$\Delta G_b / kcal\ mol^{-1}$	$\sigma / kcal\ mol^{-1}$
GA246	-47.32	5.69
GA017	-41.46	2.36
VPL60	-40.72	5.28
VPL27	-40.42	4.48
VPL48	-40.40	4.98
VPL58	-39.23	4.27
VPL42	-39.15	5.05
VPL57	-38.15	5.28
VPL29	-37.00	4.72

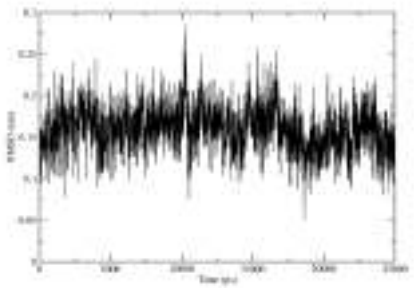
Table 2.1: Protein-ligands $\Delta G_{binding}$ energies.



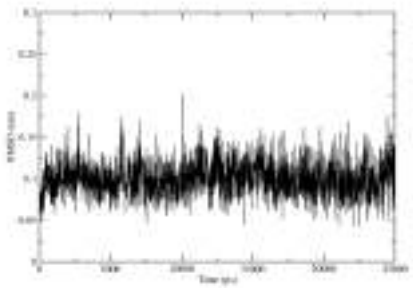
VPL48



VPL57

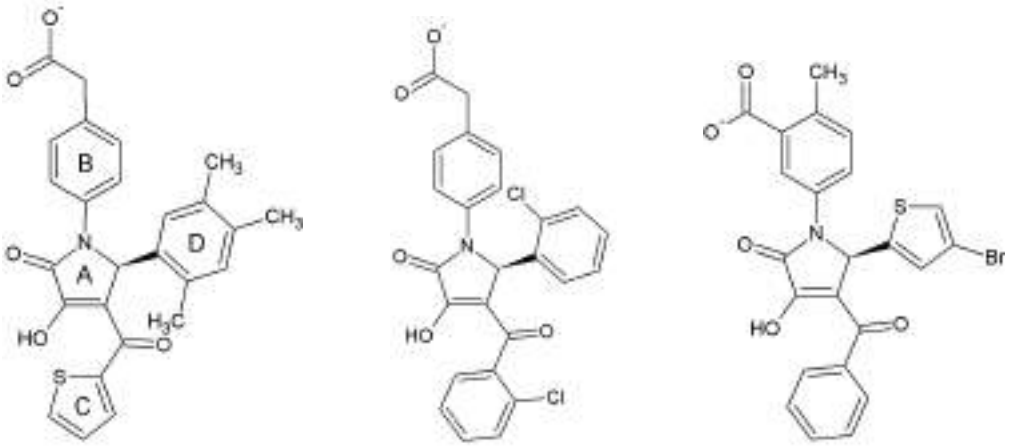


VPL58

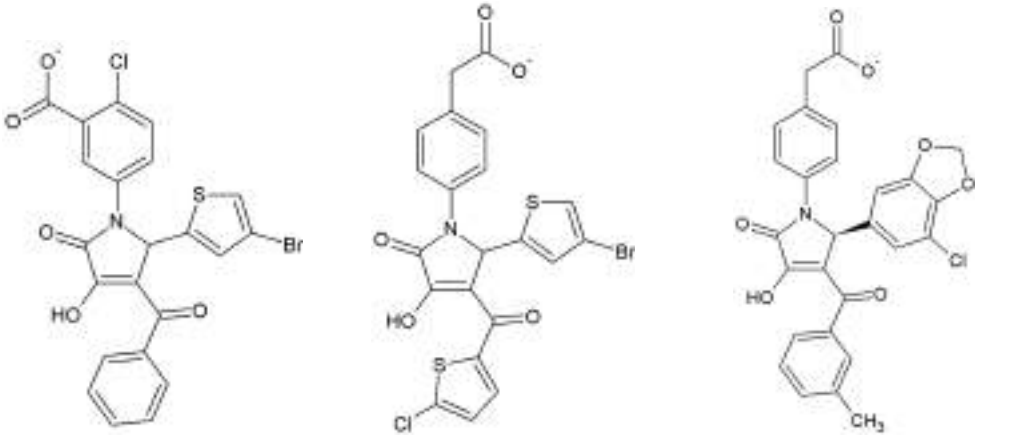


VPL60

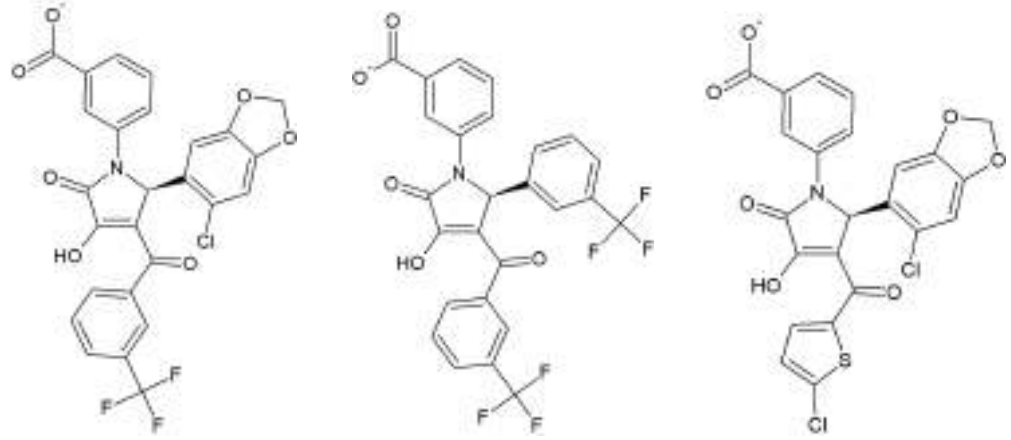
Figure 2.3: RMSD values fitting protein C_α for every protein ligand complex (2/2).



GA017 GA246 VPL27



VPL29 VPL42 VPL48



VPL57 VPL58 VPL60

Figure 2.4: Molecular structures of simulated ligands.

Residue number	$\Delta\Delta G/kcal\ mol^{-1}$								
	GA017	GA246	VPL27	VPL29	VPL42	VPL48	VPL57	VPL58	VPL60
Lys 222	**9.80**	**4.61**	**4.53**	**3.35**	**4.41**	**5.14**	**3.39**	**2.98**	**3.27**
Arg 225	**9.19**	**17.28**	**6.22**	**4.80**	**7.47**	**8.33**	**5.59**	**6.27**	**6.24**
Gln 241	0.24	1.10	0.65	0.28	-0.91	0.72	0.80	-0.23	-0.26
Gln 244	**2.75**	**3.26**	**4.56**	1.96	**4.29**	**2.77**	**4.54**	**4.14**	**4.07**
Val 245	0.51	0.61	0.77	0.72	0.74	0.65	0.63	0.62	0.69
Cys 247	0.21	0.35	0.12	0.32	-0.01	0.14	0.21	0.05	0.21
Lys 248	**9.91**	**8.73**	**12.02**	**14.83**	**11.16**	**9.42**	**14.39**	**12.63**	**14.54**
Leu 249	0.24	0.29	0.60	0.56	0.51	0.67	0.64	0.18	0.62
Lys 251	**12.96**	**11.22**	**8.00**	**11.14**	**8.92**	**10.61**	**8.43**	**10.04**	**10.19**
Asp 252	-1.28	-0.85	-2.48	-4.73	-3.51	-2.23	-5.53	-1.97	-4.28
Asp 289	1.09	1.26	1.05	1.04	1.11	1.05	0.99	1.04	0.98
Val 294	0.01	0.03	0.04	0.04	0.03	0.04	0.04	0.06	0.03
Ile 295	**3.93**	**5.29**	**4.78**	**4.08**	**4.49**	**5.24**	**4.97**	**4.70**	**4.76**
His 296	-0.13	-0.19	-0.18	-0.10	-0.10	0.04	-0.05	-0.18	-0.16
Ile 297	0.23	0.96	0.80	0.78	0.73	1.13	0.87	1.10	0.87
Asp 302	-0.83	-0.97	-1.07	-0.81	-1.03	-0.41	-1.05	-0.93	-1.27
Ile 303	0.54	0.22	0.06	0.25	0.22	0.29	0.38	0.32	0.15
Phe 328	0.38	1.09	0.77	0.83	0.85	1.23	0.67	0.69	0.84

Table 2.2: CAS values for protein–ligand complexes. The hot-spots are in bold.

Residue numbering and domain names are taken from Leung *et al.* [67]. In Figure 2.1 the two subdomains of VP35 are shown. The α-helical subdomain consists of residues Ala221–Arg283 and the β-sheet subdomain is formed by residues Val294–Ile340. The loop Val284–Pro293 connects the two subdomains. The binding site is the same for all the ligands and it is located between the α-helical and the β-sheet subdomains.
Six amino acids are found to be hot-spots in all the simulated complexes, namely Lys222, Arg225, Gln244, Lys248, Lys251 and Ile295 (Figure 2.5).

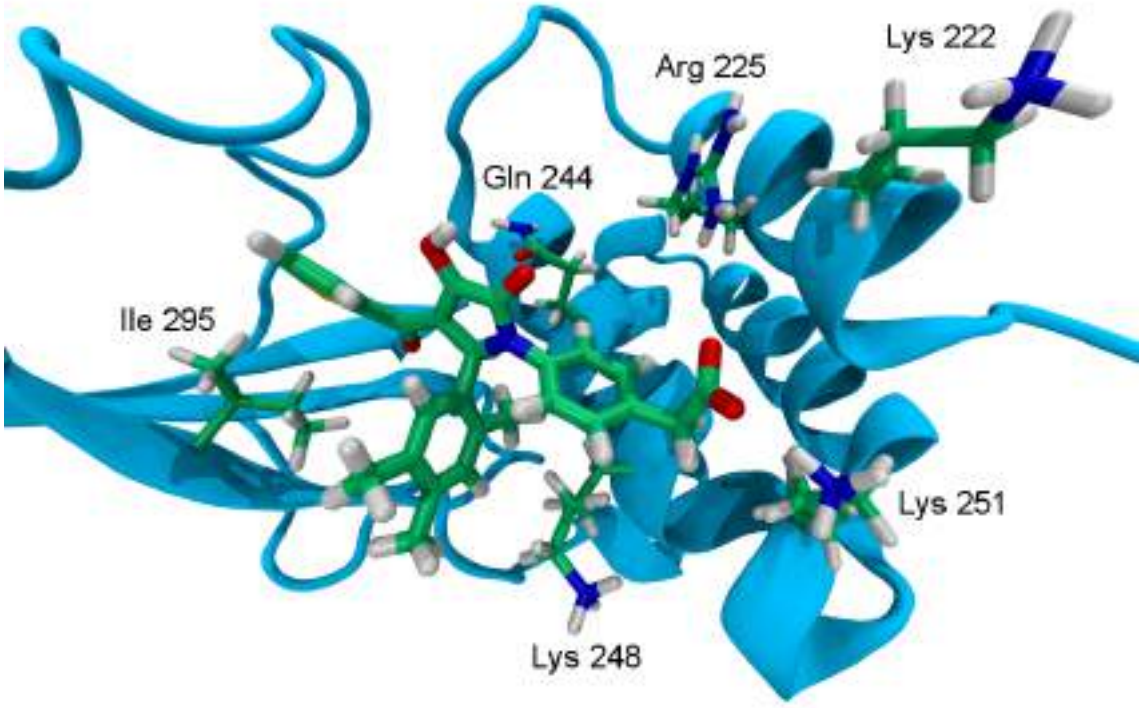


Figure 2.5: An overview of the hot-spot residues in the GA017–VP35 complex. These residues are common to all the complexes studied.

Brown *et al.* [66] performed experimental mutation of Lys248 or Ile295 into alanine, showing that these two residues are of paramount importance for ligands binding, as their mutation resulted in a near complete loss of binding affinity for all the tested ligands. Our computational results mirror the available experimental ones, furthermore they led to the identification of four new hot-spots residues exerting a critical function in the protein–ligand interaction. All the molecules considered in the present study share an almost equal orientation within the binding pocket, with the common carboxylic group oriented toward a region of the protein containing a cluster of positive charged residues, such as Lys222, Lys251, Lys248 and Arg225, which are all hot-spots, the most prominent two being Lys248 and Lys251. Lys251 forms a strong salt bridge with the carboxylic group of the ligands,

which is observed in all the simulations. The loss of this interaction upon mutation into alanine explains the high value of $\Delta\Delta G$ of the residue in all the complexes. In some cases, such as the GA017, GA246 and VPL29 complexes, the average distance between the carboxylic carbon and the charged lysine nitrogen is below 4 Å, the interaction being almost continuously present during the MD run. In other cases, for example the VPL27 complex, the average distance is around 5 Å, resulting in a lower $\Delta\Delta G$. Lys248, although bearing a positive charge, also forms hydrophobic interactions through its aliphatic chain with the B and D rings of the ligands (Figure 2.2). A coulombic interaction between opposite charges of Lys248 and the carboxylic group of the ligands is likely to be also present. Analysing the geometries of the complexes during the simulations and measuring the average distance between the carboxylic carbon and the charged lysine nitrogen, it is clear that the contribution of the hydrophobic interaction is constant for all the complexes, while the different values of Lys248 $\Delta\Delta G$ are modulated by electrostatic interactions during the dynamics. For VPL29 and VPL57, showing the higher $\Delta\Delta G$ value, the mean distance is around 5.5 Å, instead for VPL48 and GA246 this value is over 7 Å. Lys222 and Arg225 do not form stable salt bridges during the dynamics, but they are affected by the proximity of the ligand's carboxylic group. Their $\Delta\Delta G$ values are lower than those of the aforementioned positive charged residues. Ile295 interacts with the D and C aromatic rings of the ligands; these interactions are mainly hydrophobic. Interestingly, Asp252 exhibits a negative $\Delta\Delta G$ in all the complexes. This is probably due to its proximity to the negatively charged carboxylic group of the ligands. Substituting this residue with an alanine results in the loss of this disadvantaged interaction. Asp302 exhibits the same behavior, albeit to a lower extent (see Table 2.2).

2.4 Root mean-square fluctuations bootstrap analysis

To further characterize the dynamic behavior of EBOV VP35 and how it is affected by ligand binding, C_α root mean-square fluctuations were calculated for the apo-form and for each protein–ligand complex. As we can notice in Figure 2.6, there are five protein segments exhibiting a fluctuation larger than average, corresponding to residues His231–Phe235, Ile246–Ser253, Ser266–Ser272, Pro293–Ile297 and Phe328–Leu336.

Segments comprising residues His231–Phe235 and Ser266–Ser272 correspond to loop regions connecting α-helices, which have an intrinsically larger mobility than neighboring residues, so we focused our analysis on the Ile246–Ser253, Pro293–Ile297 and Phe328–Leu336 segments (Figure 2.7), which all show significant differences in

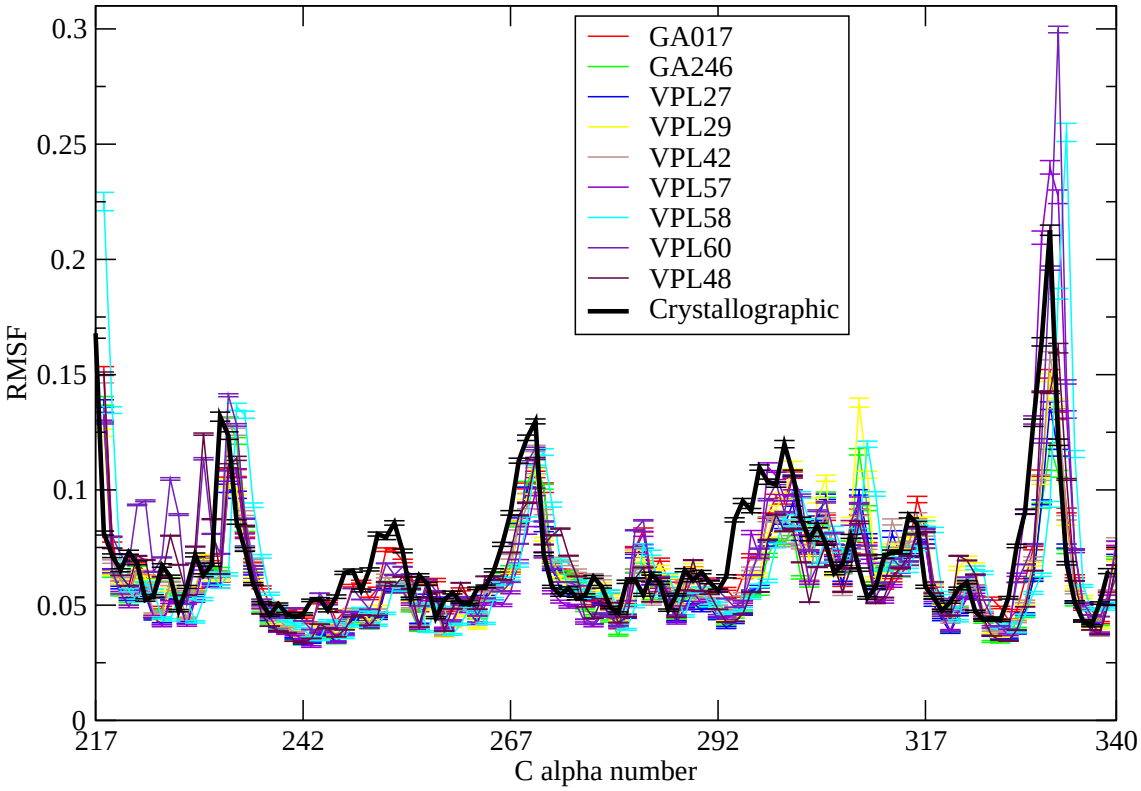


Figure 2.6: RMSF bootstrap analysis. Three protein segments have been further analyzed, due to their larger flexibility: Ile246–Ser253, Pro293–Ile297 and Phe328–Leu336 (see text).

RMSF between the apo-protein and the complexes, according to student's t-test.

Figure 2.6 shows that residues ranging from Pro293 to Ile297, belonging to the β-sheet subdomain, exhibit a higher RMSF value in the apo-form than in all the complexes. A similar behavior is also evident in residues ranging from Ile246 to Ser253 belonging to the α-helical subdomain. The presence of any ligand reduces the fluctuation of these two regions which compose the binding site, and they are thus less flexible upon ligand binding. Moreover, RMSF exhibits a sharp peak corresponding to the region between residues Phe328 and Leu336. Here RMSF reaches values around 0.2 nm for the apo-form and even greater for the VPL60, VPL57 and VPL58 complexes. These ligands seem to exalt the flexibility of these residues. Finally, root mean square fluctuation analysis was performed on MARV VP35. On the whole, more flexible loops correspond to those observed in the EBOV protein (Figure 2.8), as expected due to the high homology and great structural similarity between the two proteins.

2.5 Essential dynamics

In order to better understand the results arising from RMSF analysis we performed essential dynamics (ED) calculations on the apo-protein and on all the simulated

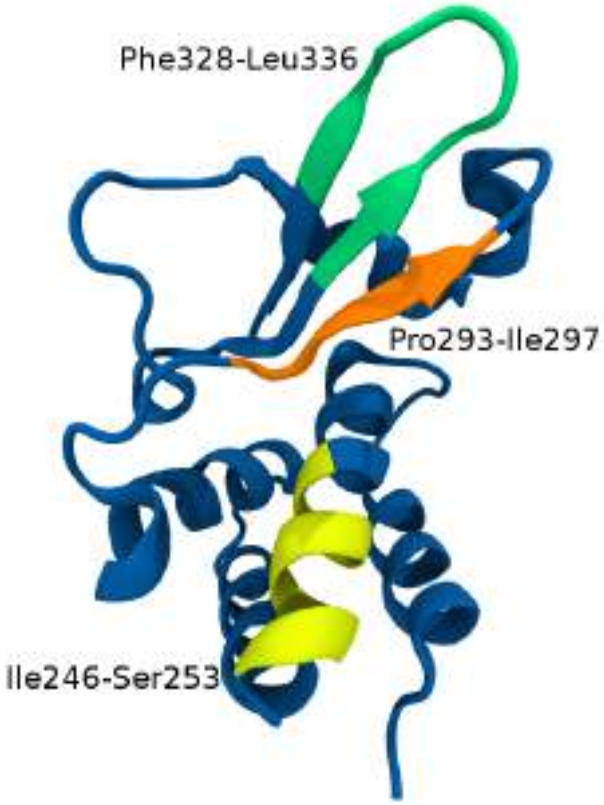


Figure 2.7: Flexible domains of VP35.

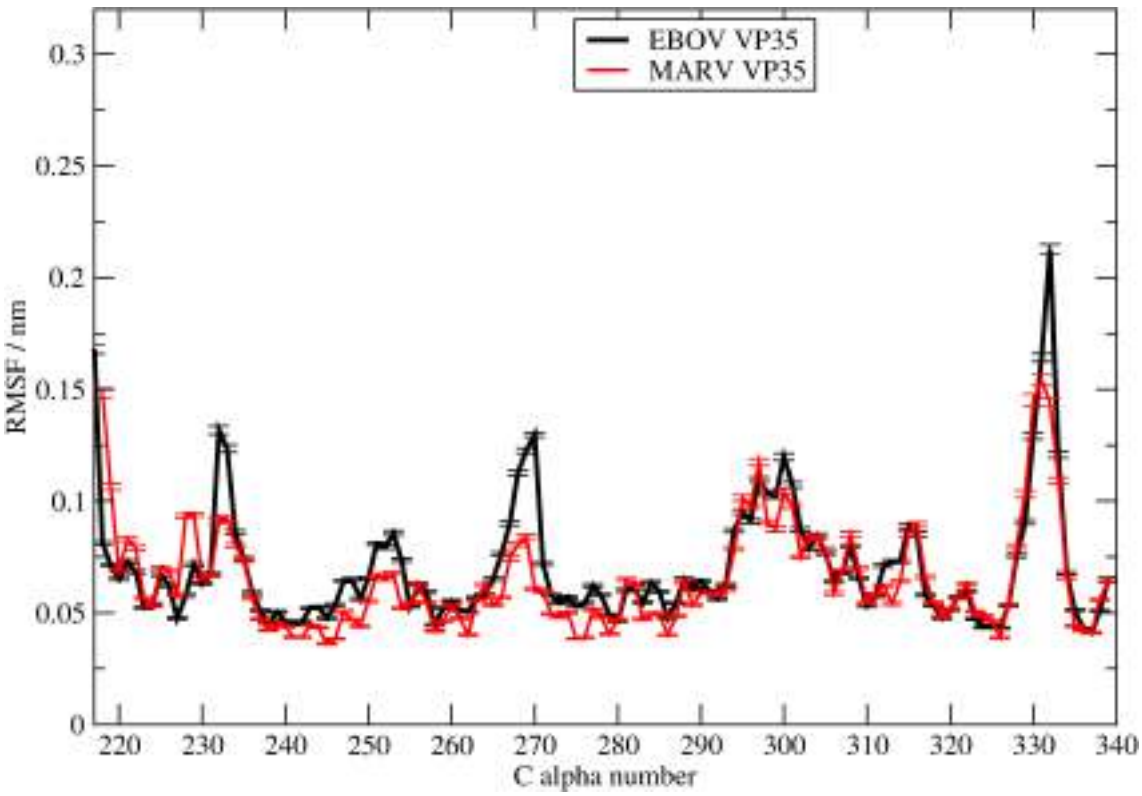
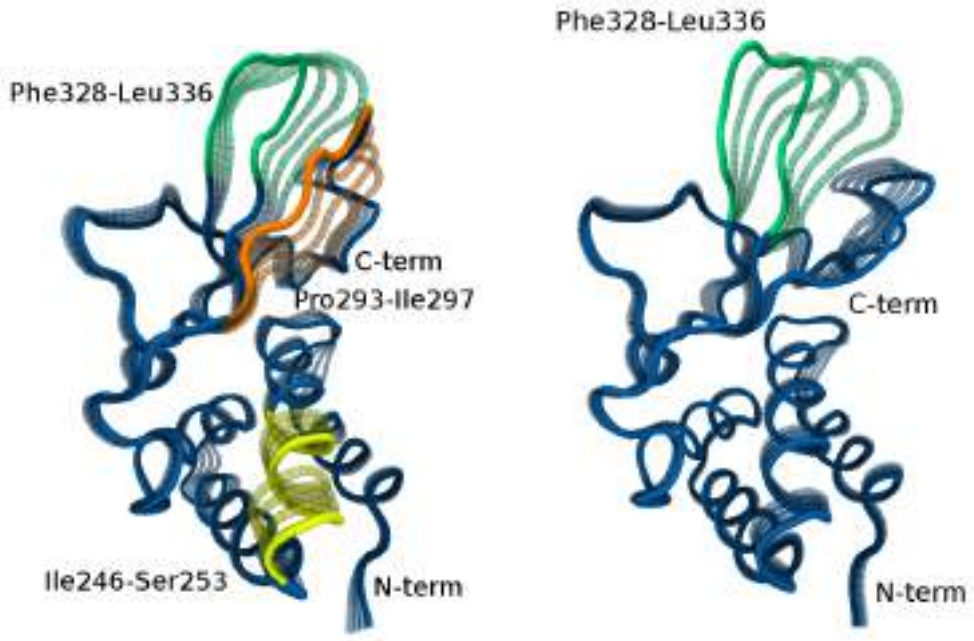


Figure 2.8: Comparison between RMSF values of EBOV and MARV VP35.

complexes. The MD trajectories were processed by principal components analysis considering the protein C_α for the construction of the covariance matrix. For each system, the projection of the trajectory on the three eigenvectors corresponding to the greater eigenvalues were visualized. In all the systems the first three eigenvectors represent at least 30% of the global motion.

The ED analysis of the apo-form of VP35 highlights two kinds of motions. The projection along the first two eigenvectors describes a closing motion of the β-sheet subdomain on the α-helical subdomain. In particular loops Pro293–Ile297 and Phe328–Leu336, both belonging to the β-sheet subdomain, appear to bend towards the residues Ile246–Ser253 of the α-helical subdomain (Figure 2.9). The projection on the 3rd eigenvector shows the motion of the Phe328–Leu336 loop alone. These results strictly mirror the observations from the RMSF analysis, where the aforementioned groups exhibited the higher mobility. For the apo-form the first three eigenvectors describe more than 50% of the collective motion of the protein.

The presence of some ligands, such as GA017, GA246 and VPL29, hinders the closing motion observed in the apo-form; the projection on the first eigenvector of the GA017 complex is the same as the apo-form, but the amplitude of the motion is smaller. Also in this case we confirm our previous observation from the RMSF analysis, as RMSF values for residues Ile246–Ser253 and Pro293–Ile297 of the complexes were lower than for the apo-protein.



A B

Figure 2.9: **A panel**: schematic representation of the projection of the MD simulation on the 1st eigenvector extracted from the ED of the apo-protein. Highlighted in orange and green we can see the closing motion of the β-sheet subdomain towards residues Ile246–Ser253 belonging to the α-helical subdomain, highlighted in yellow.
B panel: The projection of the MD simulation on the 1st eigenvector of the VPL57–VP35 complex. Highlighted in green we can see the intense motion of the Phe328–Leu336 loop. The characterizing closing motion of the apo-protein is not present in this case.

Interestingly, the RMSF analysis suggested that ligands VPL57, VPL58 and VPL60 could exalt the motion of residues between Phe328 and Leu336. The projections on the first three eigenvectors for the VPL60 complex trajectory show an intense movement of these residues. In this case the characteristic closing motion observed in the apo-form was not detected, while it was significantly reduced in VPL57 and VPL58 complexes. For the VPL60 complex the first three eigenvectors describe about 48% of the global motion. For the VPL57 complex the projections on the 1st and 2nd eigenvector show a very large motion of the loop Phe328–Leu336 (Figure 2.9) while the projection on the 3rd eigenvector describes the closing motion of the β-sheet subdomain towards the α-helical subdomain.
Finally, VPL58 has a similar behavior, the projection of the trajectory on the 1st eigenvector resulting in a large motion of the Phe328–Leu336 loop and the projections on the 2nd and 3rd eigenvectors representing the closing motion of the α-helical subdomain on the β-sheet subdomain. The first three eigenvectors of both VPL57 and VPL58 describe almost 40% of the collective motion of the complex.

2.6 Conclusions

We studied EBOV VP35 and its inhibitors from a computational point of view. We performed both MD simulations of apo-protein and complexes starting from the crystallographic structures and CAS, obtaining a very good agreement with the structural and biochemical experimental results. We deepened the study of this systems from the point of view of the intermolecular interactions, identifying new hot-spots in addition to those experimentally determined. Then we performed the RMSF bootstrap analysis, in order to identify the most mobile residues and to highlight the difference between the apo-protein and the complexes. In particular we identified three protein segments whose behavior was different in the apo-protein compared to the complexes.

To characterize the RMSF results, in order to analyze the global motion of the systems, we performed ED analysis, obtaining the collective and most important motion for every system. We noticed that ligands hinder the typical closing motion of the β sheet-subdomain towards the α helical-subdomain of the apo-form. The binding of some ligands, like VPL57 VPL58 and VPL60, exalts the movement of the loop Phe328–Leu336.

Chapter 3

Computer aided design of an oligopeptide targeting VP24

3.1 Introduction

Viral Protein 24 (VP24), acts as an inhibitor of interferon (IFN) signaling. Immune response of the organism to a viral infection is due to interferons, through the phosphorylation of the STAT (signal transducer and activator of transcription) protein operated by the Janus kinase [79]. The phosphorylated STAT is recognized by a subset of Karyopherin α (KPNA) family of nuclear transport factor. This STAT-KPNA complex is transported in the cell nucleus where it induces the expression of ISG gene, that has an antiviral activity [80,81]. VP24 acts in a cell intrinsic manner inhibiting IFN signaling and making cells refractory to IFN as it forms a complex with KPNA, thus making it unavailable for the STAT transportation into the cell nucleus (Figure 3.1) [82].

Inhibition of VP24-Karyopherin interaction is a possible therapeutic strategy to reduce Ebola virulence. Targeting Protein-protein interactions (PPIs) is a difficult task, due to the inherent features of protein-protein contact surfaces, which are often flat and lack well defined pockets, and to their variability (Section 1.4) [30, 83]. Nevertheless, in recent years many encouraging results have been obtained, showing that exploiting PPIs as therapeutic targets is a viable, albeit challenging option [84,85] and several PPIs targeting molecules, both of peptidic [28,86,87] and non peptidic nature [88–90], have been described. Among the molecular modeling techniques aimed to study PPIs, computational alanine scanning [36] (Section 1.5) is a widely used one. This kind of approach has been successfully used to identify peptides acting as PPIs inhibitors and corresponding to protein subsets located at the protein-protein interface [91], as well as to describe the interactions between protein surface and small molecule binders [92].

VP24 adopts a compact, single domain, α/β structure of novel fold. The overall

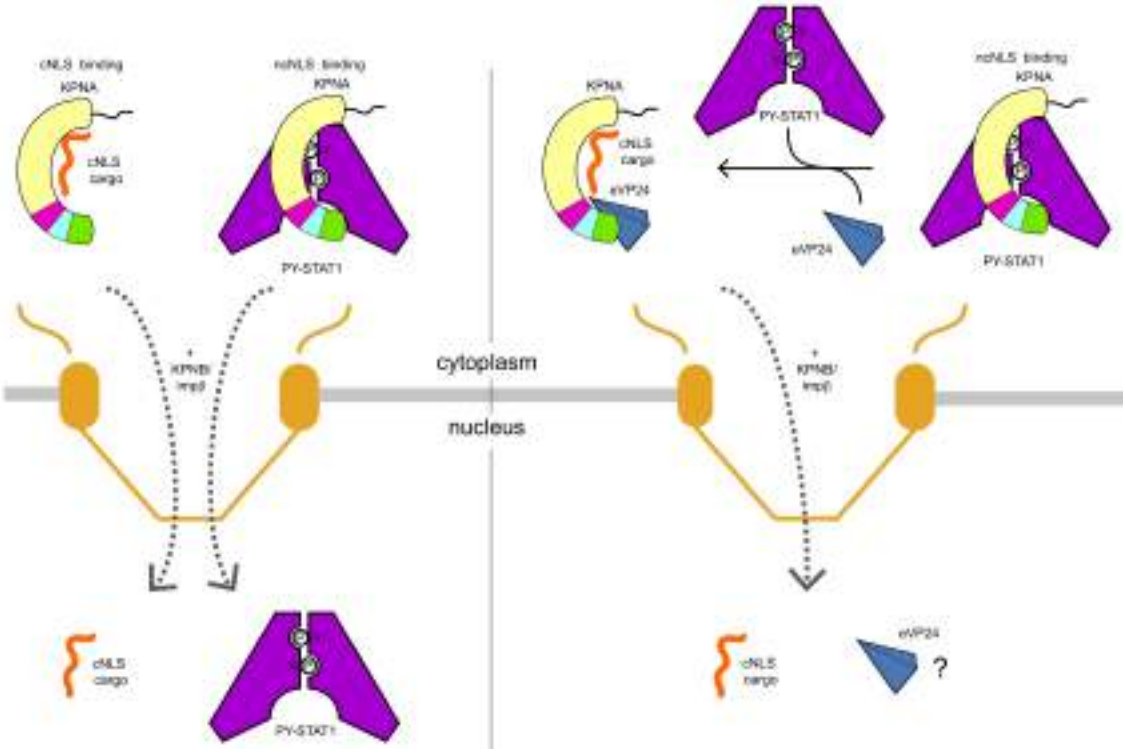


Figure 3.1: **Left panel:** In a healthy organism the C-terminal domain of KPNA binds to the STAT forming a protein-protein complex able to reach the cell nulceus.
Right panel: VP24 binds to the C-terminal domain of KPNA, making it unavaiable to transport the STAT in the cell nucleus. Image taken from Ref [82].

shape of VP24 resembles a triangular pyramid, the three faces of the pyramid are numbered 1, 2 and 3 (Figure 3.2). A collection of α helices (α1 and α5-10) and a small, three-stranded, antiparallel β sheet (β1-3) form the top of the pyramid with the N-terminus at the apex. A five-stranded antiparallel β sheet (β4-8) forms the center, while a second collection of α helices (α2-4) forms the base [93].

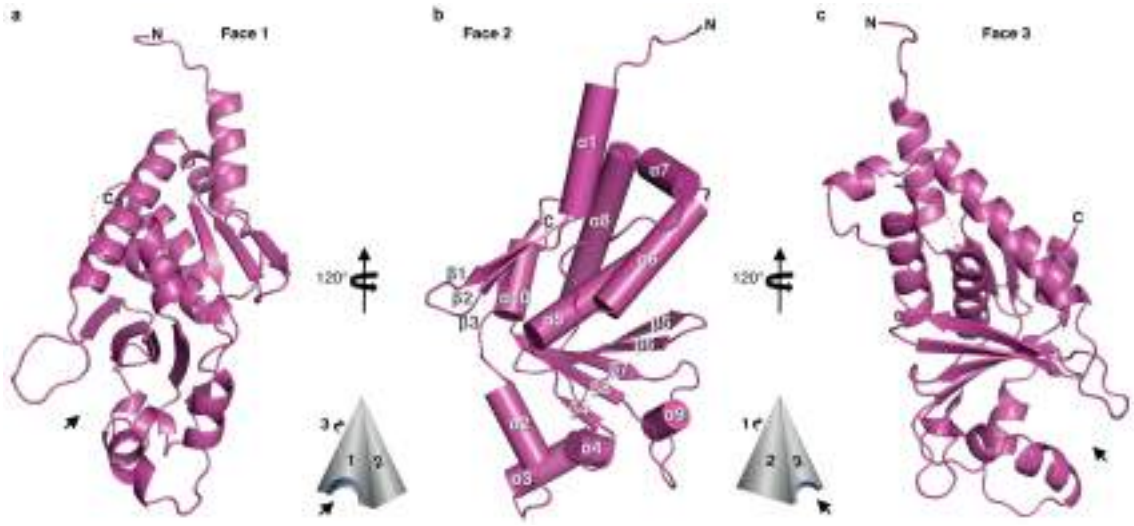


Figure 3.2: VP24 pyramidal structure. Image taken from Ref [93].

In this work we performed an atomic level mapping of the interactions between VP24 and KPNA, employing molecular dynamics simulations in explicit solvent and free energy calculations, and identified a small subset of residues at the VP24-KPNA interface responsible for protein complex formation. In particular, such critical residues belonging to Karyopherin were used as the basis for the design of a nine residues long peptide potentially able to interact with VP24, thus competitively inhibiting its interaction with Karyopherin. To evaluate whether the identified peptide could retain its ability to interact with VP24 even when extracted from its protein environment, molecular dynamics simulation of the peptide-VP24 complex was performed, as well as CAS, in order to evaluate the structural and energetic behavior of the peptide. The interaction between VP24 and the identified peptide was then experimentally investigated through advanced NMR techniques. The interactions between the Ebola receptor and the nonapeptide ligand are governed by weak forces including Van der Waals interactions, hydrogen bonding and hydrophobic associations. In this case, the ligand-based NMR approach represent an optimal angle of observation, because it permits the analysis of interactions governed by weak equilibrium dissociation constants (in the micromolar to the millimolar range), focusing on the ligand. We confirmed the binding of the peptide with VP24 protein and mapped the interaction epitope by determining the ligand regions in contact with the receptor.

3.2 Computational details

Molecular Dynamics The crystallographic structure of the complex VP24-KPNA was obtained from the Protein Data Bank (PDB code: 4U2X [82]), and the residue numbering used here is referred to such structure. Molecular dynamics simulations were carried out with the GROMACS 4.5.3 [19] package using explicit solvent and periodic boundary conditions. The AMBER99SB-ildn [68] force field was used. Every system was solvated with TIP4P [94] waters and neutralized with Cl^- or Na^+ ions to reach neutrality. The LINCS algorithm [4] was employed to constraint all bonds to their equilibrium length, allowing a time step of 2 fs. The system was submitted to 10000 steps of geometry optimization with the steepest descent method. Afterwards it was equilibrated for 200 ps in NVT conditions (T=300 K) and subsequently for 200 ps in NPT conditions, in order to equilibrate system density. For the VP24-KPNA complex a 50 ns molecular dynamics was performed in NPT conditions (1 bar, T=300 K). For each of the peptide-VP24 complexes, a 100 ns MD simulation was run in the aforementioned conditions. Temperature and pressure were kept constant at their reference value using the velocity rescale algorithm [71] and the Berendsen barostat [72], respectively. A 14 Å cutoff was applied for non-bonded interactions and the Particles Mesh Ewald algorithm [73] was employed to calculate long range electrostatic interactions.

Computational alanine scanning 500 snapshots were extracted from the last 20 ns of the dynamics of the VP24-KPNA complex (one snapshot every 40 ps). Solvent accessible surface area (SASA) of each residue was calculated using Naccess [74]. The ΔG of binding was calculated with the MM-PBSA approach as implemented in the GMXPBSA 2.0 suite [95, 96]. This protocol implicitly assumes that point mutations in the protein do not significantly affect its conformation. The validity of this assumption in computational alanine scanning has been widely confirmed in the literature, when applied to protein-protein interactions [97, 98]. A dielectric constant of 2 was chosen for protein interior. The same protocol was employed for the computational alanine scanning for each of the VP24-peptide complexes.

3.3 Computational Alanine Scanning results

A molecular dynamics of 50 ns was performed on the VP24-KPNA complex. Computational Alanine Scanning (CAS) was carried out in the last 20 ns of the dynamics when the system was fully equilibrated as shown by RMSD plot (Figure 3.3).

In general, a few nanoseconds long trajectories allow a sufficient sampling for an accurate binding free energy calculation with the MM-PBSA approach and consequently for CAS purposes [77, 78]. The protein-protein interface is defined as the

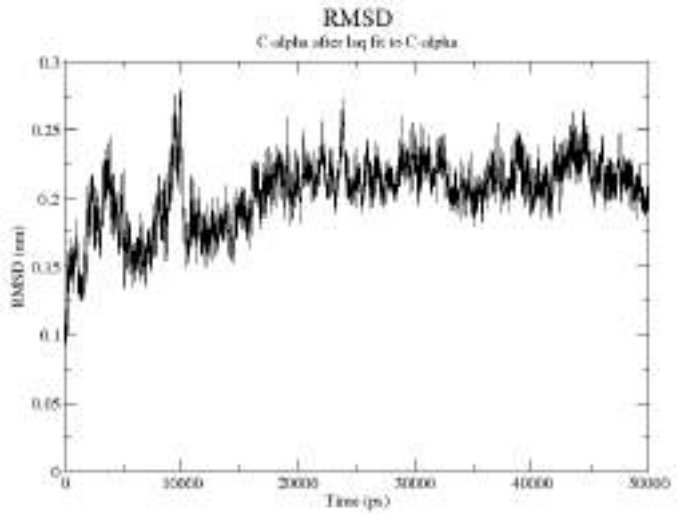


Figure 3.3: RMSD versus time for VP24-K complex C_α atoms considering the crystal structure as reference.

ensemble of amino acids whose solvent exposed surface area has a nonzero variation upon complex formation and it is composed of 56 residues. Residues whose mutation led to a variation in binding free energy ($\Delta\Delta G$) greater than 2 $kcalmol^{-1}$ are defined as hot-spots. CAS results are summarized in Table 3.1. Hot-spots are evenly distributed between VP24 and KPNA, which contain 10 and 11 hot-spots respectively (Figure 3.4).

The two proteins exhibit a high degree of electrostatic complementarity, as can be inferred from electrostatic potential maps (Figure 3.5). A more detailed insight into the network of interactions at the interface that contribute to VP24-KPNA complex formation can be achieved by the analysis of the local environment of the hot-spots during the simulation.

As we can see from Table 3.4, since most of the hot-spots are either charged or polar residues, most of the interactions within the VP24-KPNA complex are salt bridges and hydrogen bonds, although hydrophobic hot-spots and non-polar interactions are also observed. In particular, Arg140 and Glu475' (residues belonging to KPNA are labeled with a prime on their sequence number) form a salt bridge that is populated along 88% of the MD trajectory, while salt bridges between Glu203 and Lys399', between Asp205 and Arg396' and between Asp124 and Lys481' are populated during 85%, 78% and 64% of the trajectory respectively (Figure 3.6).

Residue Asp124 is also involved in a hydrogen bond between its carboxyl group and Thr434' hydroxyl, which is observed during 75% of the simulation, while Arg137 and Asp480' backbones form a hydrogen bond for 74% of the trajectory. A small hydrophobic cluster, comprising Leu121, Val141 and Tyr477', Phe484' and the

Mutation	$\Delta\Delta G/kcal\ mol^{-1}$	$err/kcal\ mol^{-1}$	Mutation	$\Delta\Delta G/kcal\ mol^{-1}$	$err/kcal\ mol^{-1}$
GLU88	-10.99	0.79	LEU390	0.10	0.79
GLU113	**6.16**	**0.79**	GLN391	0.41	0.79
LEU121	**6.28**	**0.79**	GLU394	-13.16	0.79
ASP124	**4.35**	**0.79**	PHE395	1.43	0.76
TRP125	0.96	0.79	**ARG396**	**21.04**	**0.79**
LEU127	0.14	0.79	**ARG398**	**18.80**	**0.74**
THR128	-0.64	0.79	**LYS399**	**23.07**	**0.84**
THR129	1.53	0.79	**LYS427**	**16.67**	**0.79**
ASN130	0.24	0.79	ASP431	-14.81	0.79
THR131	-0.21	0.79	LEU433	0.50	0.79
PHE134	1.84	0.79	**THR434**	**6.09**	**0.79**
ASN135	**3.30**	**0.79**	VAL435	1.19	0.79
MET136	1.62	0.79	**MET436**	**3.63**	**0.79**
ARG137	**23.84**	**0.81**	ASP437	-13.57	0.74
THR138	-2.89	0.79	**GLU474**	**7.91**	**0.79**
GLN139	0.29	0.79	GLU475	1.98	0.79
ARG140	**9.94**	**0.79**	**TYR477**	**6.31**	**0.79**
VAL141	**2.65**	**0.79**	LEU479	0.62	0.79
GLN184	-1.55	0.79	**ASP480**	**17.56**	**0.79**
ASN185	1.15	0.74	**LYS481**	**28.90**	**0.81**
HIS186	0.14	0.79	GLU483	0.79	0.79
LEU201	1.27	0.79	**PHE484**	**6.02**	**0.79**
GLN202	-0.07	0.79	LEU485	0.55	0.79
GLU203	**10.37**	**0.72**	SER487	-1.53	0.79
ASP205	**16.31**	**0.76**	HIS488	1.15	0.79
SER207	0.24	0.79	GLU489	-4.44	0.76
ASN210	-0.38	0.79	ILE501	-0.45	0.79
LYS218	**11.13**	**0.79**	PHE505	1.00	0.79

Table 3.1: Computational Alanine Scanning data with standard errors. Left column refers to VP24 residues, right column to KPNA residues. Hot-spots are reported in bold.

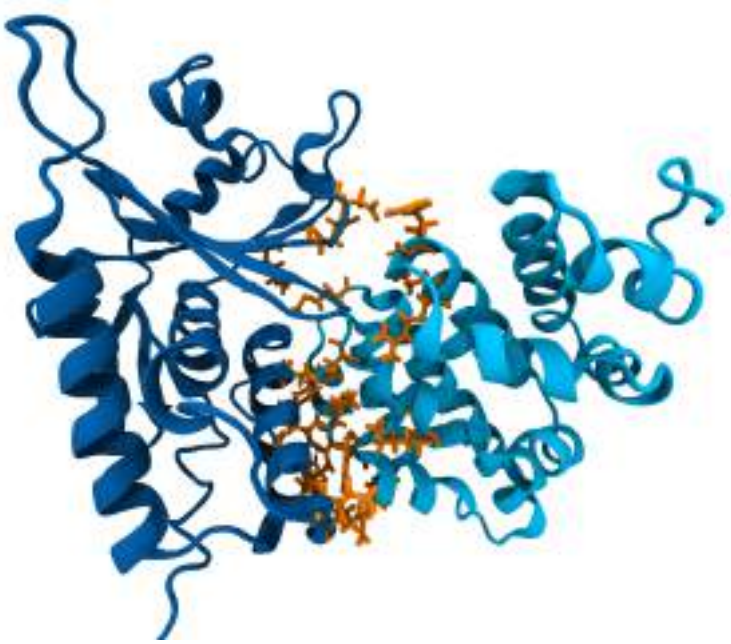
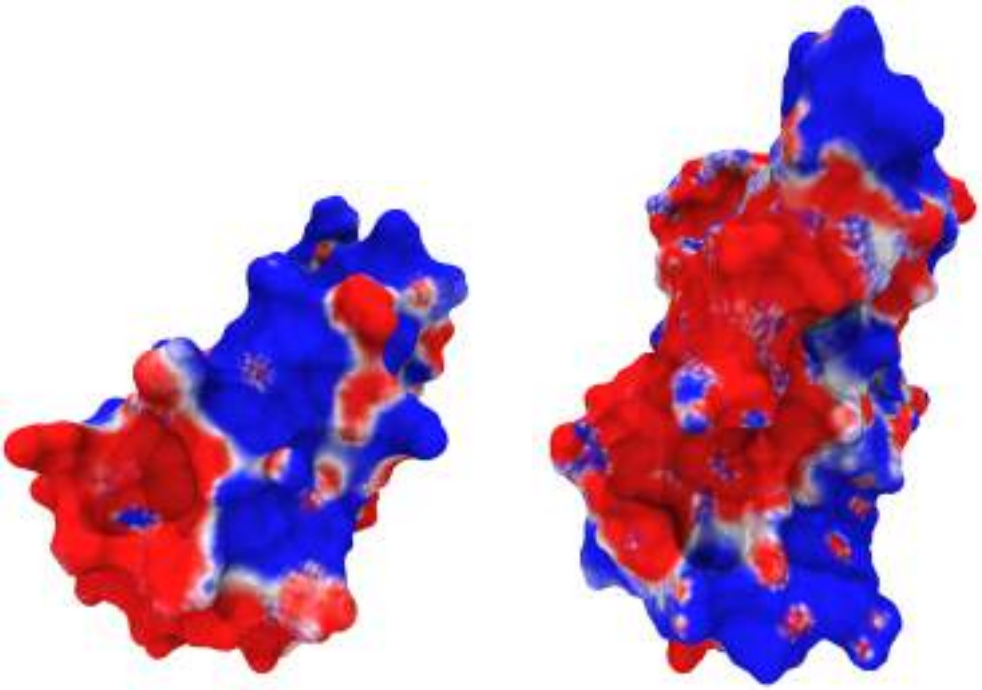


Figure 3.4: VP24 (blue)-KPNA (cyan) complex. The hot-spots arising from the CAS analysis are highlighted in orange.



KPNA VP24

Figure 3.5: **Left panel:** Electrostatic potential of Karyopherin at the interface with VP24. Values ranging from $-k_bT/e$ (red) e to $+k_bT/e$ (blue).
Right panel: Electrostatic potential of VP24 at the interface with Karyopherin. Values ranging from $-k_bT/e$ (red) to $+k_bT/e$ (blue).

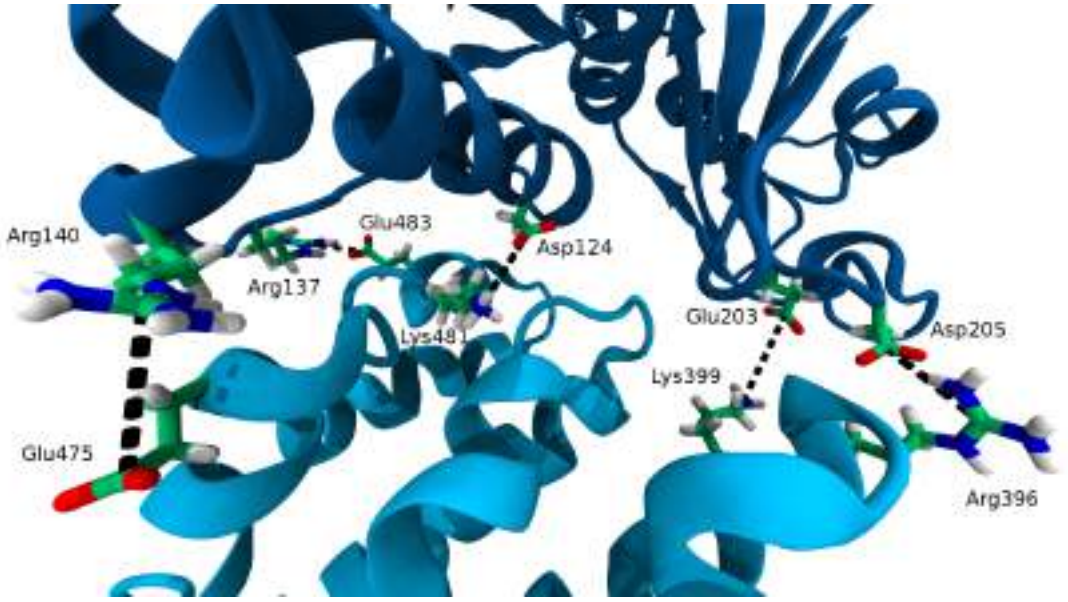


Figure 3.6: Overview of the five salt bridges at the VP24-K interface.

aliphatic portion of the sidechain of Lys481' has been also observed and is steadily formed during the whole simulation (Figure 3.7).

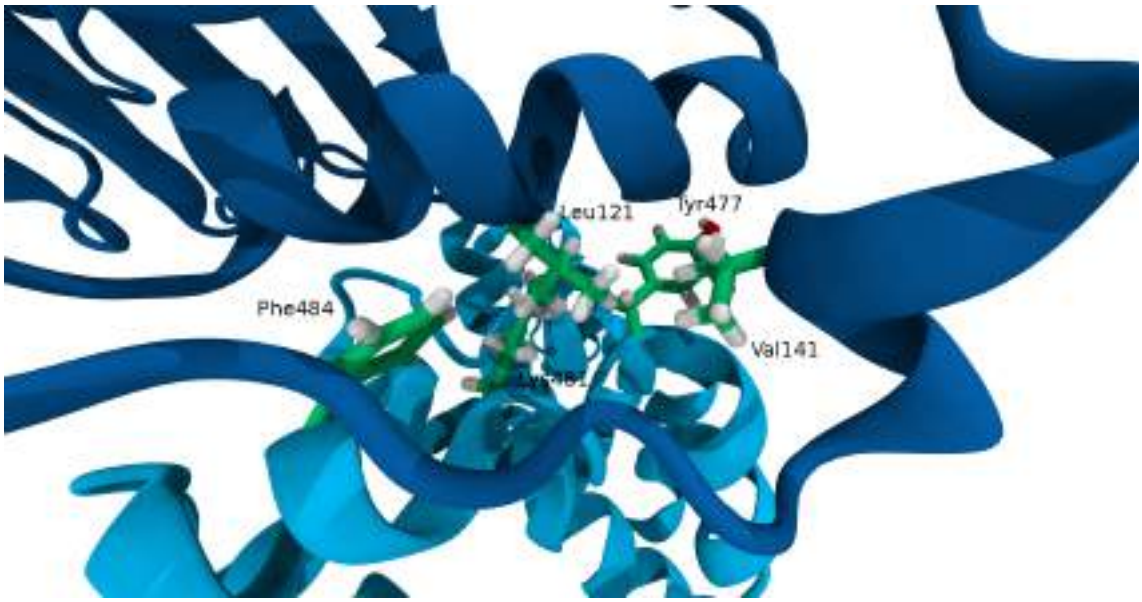


Figure 3.7: The hydrophobic cluster at the VP24-KPNA interface.

Table 3.4 also shows the presence of some residues exhibiting a largely negative $\Delta\Delta G$ (i.e. residues whose mutation into alanine increase the binding energy between the proteins). These are Glu88, Glu394', Asp431' and Asp437'. Inspecting the electrostatic potential on VP24 and KPNA surface, it may be observed (Figure 3.8) that all these negatively charged residues, when the complex is formed, actually face a negative potential area on the binding partner. This suggest a somehow

suboptimal affinity of VP24 for KPNA, which deserves further attention and should be investigated by molecular biologists.

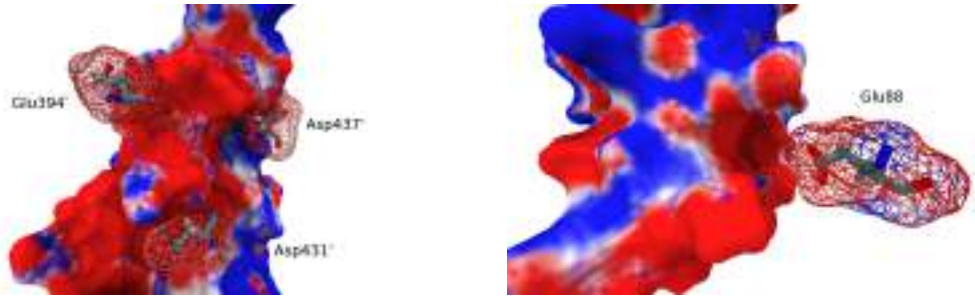


Figure 3.8: **Left panel:** Electrostatic potentials of Glu394', Asp431' and Asp 437' (mesh grid) facing VP24 interface (surface representation). Values ranging from $-k_bT/e$ (red) e to $+k_bT/e$ (blue).
Right panel: Electrostatic potential of Glu88 (mesh grid) facing Karyopherin interface (surface representation). Values ranging from $-k_bT/e$ (red) e to $+k_bT/e$ (blue).

3.4 Design of the RS peptide

Stemming from these geometric and energetic features of the VP24-KPNA interface, we identified two KPNA segments comprising a set of hot-spot residues that are close to one another not only in the protein tridimensional structure, but also in its sequence. We inferred that a peptide sequence corresponding to one of these KPNA subsequences could bind to VP24, thus interfering with VP24-KPNA complex formation. The binding capability of an isolated peptide is not guaranteed in principle, even if it contains several hot-spots, because it may undergo major structural rearrangements when isolated from the parent protein and because of the complexity of the interaction networks that lead to protein complex formation. The first subsequence we considered, comprising the hot-spot residues Arg396', Arg398' and Lys399', ranges from Ala393' to Glu400'. A 100ns MD simulation of this peptide in complex with VP24 shows large conformational fluctuations (Figure 3.10), leading to a partial detachment of the peptide from the protein. In particular, the salt bridge interactions involving the former hot-spots are completely lost. For these reasons the peptide seems unable to interact with VP24, thus it was not further considered.
The second subsequence comprises residues ranging from Glu474' to Phe484'. A 100 ns MD simulation and a CAS of the peptide-VP24 complex were performed in order to evaluate the binding capability of the selected sequence. Both Glu474' (hot-spot in VP24-KPNA complex) and Glu475' show negative $\Delta\Delta G$ values (Table

Mutation	$\Delta\Delta G/kcal\ mol^{-1}$	
	VP24-KPNA	VP24-Glu474' Phe484
GLU474	**7.91**	-2.13
GLU475	1.98	-6.26
TYR477	**6.31**	**6.57**
LEU479	0.62	0.43
ASP480	**17.56**	**19.34**
LYS481	**28.90**	**27.60**
ILE482	-	-0.1
GLU483	0.79	**9.08**
PHE484	**6.02**	**8.70**

Table 3.2: Comparison between the $\Delta\Delta G$ values caluclated for VP24-KPNA complex and for VP24-Glu474' Phe484 complex.

3.2), while all the remaining hot-spots are conserved. Based on these results we decided to shorten the sequence discarding Glu474' and Glu475', obtaining a peptide with a smaller net charge (-1 instead of -3) and maintaining a very high density of hot-spots, namely Tyr477', Asp480', Lys481' and Phe484'.

The selected peptide (from now on named RS, Figure 3.9) involves residues ranging from Ala476' to Phe484' and has the sequence AYGLDKIEF. The same protocol applied in the two aforementioned subsequences was employed. We ran a 100 ns long molecular dynamics simulations of the VP24–RS complex, in order to investigate whether the identified peptide could retain its ability to bind to VP24, even when extracted from its protein environment. In contrast to what happened for the two former sequences, the protein-peptide complex showed a good structural stability during the whole simulation, with protein binding site and peptide binding mode well conserved (Figure 3.10). On the other hand, its helical secondary structure was progressively relaxed and eventually lost during the simulation (Figure 3.11).

Figure 3.9: Structure of the RS peptide.

Subsequently, a CAS was carried out in order to verify if the previously identi-

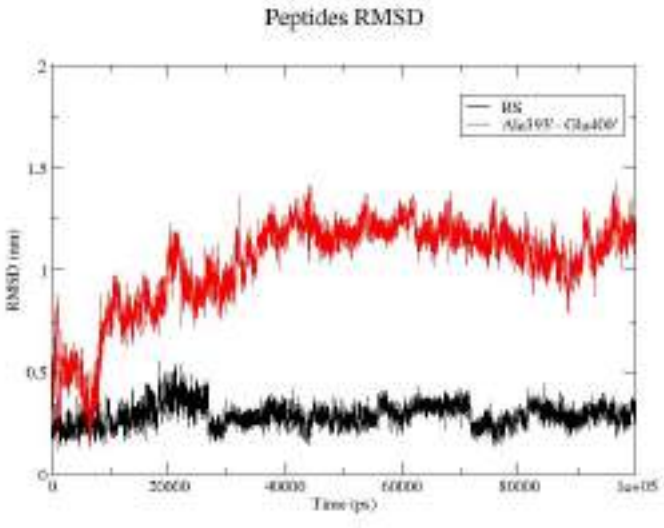


Figure 3.10: RMSD versus time for VP24-K complex C_α atoms considering the crystal structure as reference.

Figure 3.11: The RS peptide in complex with VP24. In the left panel the starting complex structure is shown, in the right panel the VP24-RS structure after MD simulation. RS, while keeping its interaction with the protein, loses its helical shape.

Mutation	$\Delta\Delta G/kcal\ mol^{-1}$	$\sigma/kcal\ mol^{-1}$
TYR477	**7.68**	**0.60**
LEU479	0.76	0.60
ASP480	**18.28**	**0.57**
LYS481	**19.71**	**0.60**
ILE482	1.67	0.62
GLU483	**10.67**	**0.55**
PHE484	**6.99**	**0.60**

Table 3.3: $\Delta\Delta G$ values caluclated for VP24-RS complex.

fied KPNA hot-spots were conserved also in the VP24-RS complex. As shown in Figure 3.12 and in Table 3.3, not only all the hot-spots are conserved, but also a new one, i.e. Glu483', can be identified. This residue forms a salt bridge with Arg137 which is observed in 90% of the trajectory of the VP24-RS complex. Interestingly, this salt bridge, albeit present also in the VP24-KPNA complex, was observed only in 31% of the trajectory.

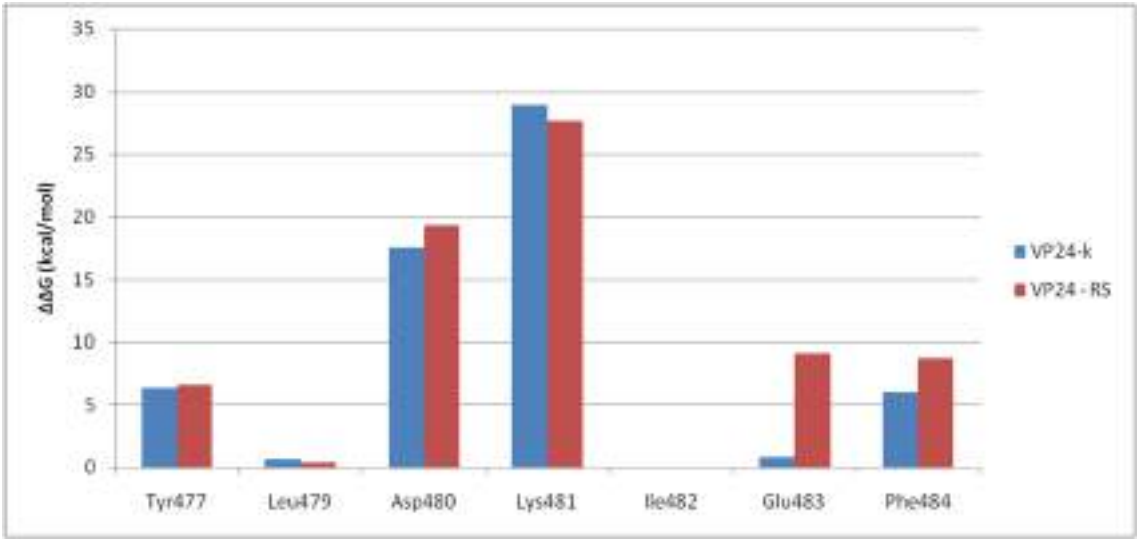


Figure 3.12: Computational Alanine Scanning results for residues ranging from Tyr 477' to Phe 484' belonging to Karyopherin (blue bars) or to the RS peptide (red bars).

3.5 Experimental evidences: NMR and CD

3.5.1 NMR

In order to experimentally verify if RS can actually bind VP24 in solution, NMR experiments were performed. First, the capped-peptide was analyzed in terms of conformational features by NOESY (500ms) and ROESY (200ms), observing no medium and long-range correlations. Only sequential correlations could be found, indicating that the peptide exists in random conformation. tr-NOESY and tr-ROESY experiments showed no definite secondary structure in the peptide even upon addition of VP24. These observations mirror the behavior of RS in MD simulations, both alone and in complex with VP24 (Figure 3.11). In this work, STD technique was used as an epitope mapping device to describe the VP24-peptide interactions. The method is based on the transfer of saturation from the protein to the bound peptide which in turn, by exchange, is moved into solution where it is detected. During the period of saturation, the magnetization gradually moves from the protein to the protons of the peptide when the ligand binds to the target. For transient interactions with rapid exchange [99], the ligand polarization in the bound state is transferred to the free state where the saturation accumulates during the irradiation time of the experiment. The STD spectrum shows only the signals of peptide protons that were in close contact with the protein (Figure 3.13).

Figure 3.14 shows the absolute STD percentage (grouped in two intensity ranges) for protons which forms the epitope that reflects the relative proximity of the atoms to the binding site. In fact, the ligand protons nearest to the protein are most likely to be saturated to the highest degree, and therefore have the strongest signal in the mono-dimensional STD spectrum. In particular, the more strictly interacting moieties are the aromatic protons of Tyr477' and Phe484' (1.5 and 1.6 % absolute STD respectively), the amidic proton of Asp480' (2 % absolute STD) and the C-terminal amide protons (1.6 % absolute STD). Although CAS analysis and STD point out different aspect of peptide protein interaction (CAS is focused on the energetic of side chain interactions highlighted by their mutation into alanine, while STD supplies information on the proximity of a set of hydrogen atoms of the peptides with respect to the protein) we may notice that, by and large, the two approaches give mutually compatible and sound results.

The key role of the aromatic residues Tyr477' and Phe484', which take part in hydrophobic interactions with VP24, are highlighted both by CAS and by NMR analysis. Moreover, the role of charged residues, in particular Lys481' and Asp480', while somehow overestimated by CAS, as often reported in literature [15], has been observed by STD analysis. In order to perform a negative control, we analyzed the linear octapeptide A (LGYGFVNY) in presence of VP24 protein by STD exper-

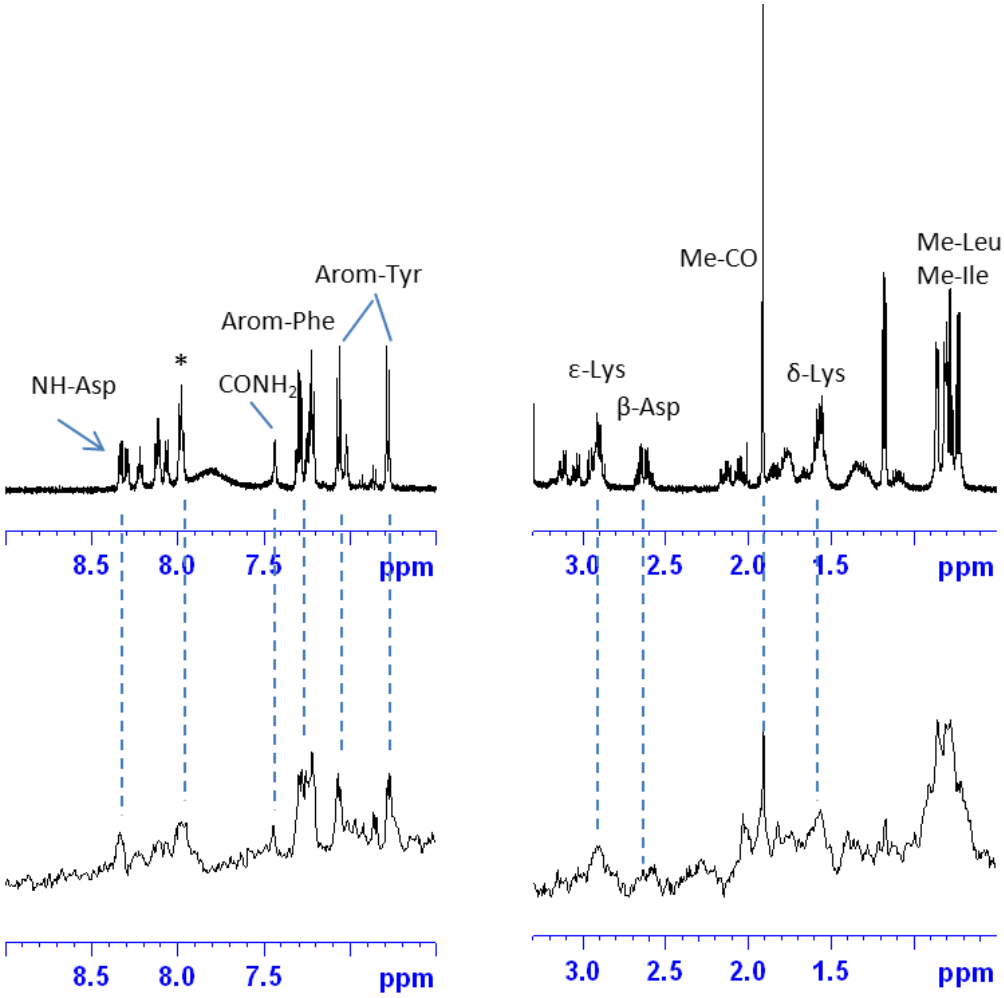


Figure 3.13: **Upper panel:** selected regions of1H-NMR spectrum of 1.1 mM peptide in the presence of 22 μM VP24 protein in phosphate buffer, resulting in a 50:1 ligand-target ratio. The signal labeled with an asterisk is due to the Lys, Leu and Ile amide protons which show the same chemical shift. **Bottom panel:** The same region of the STD spectrum.

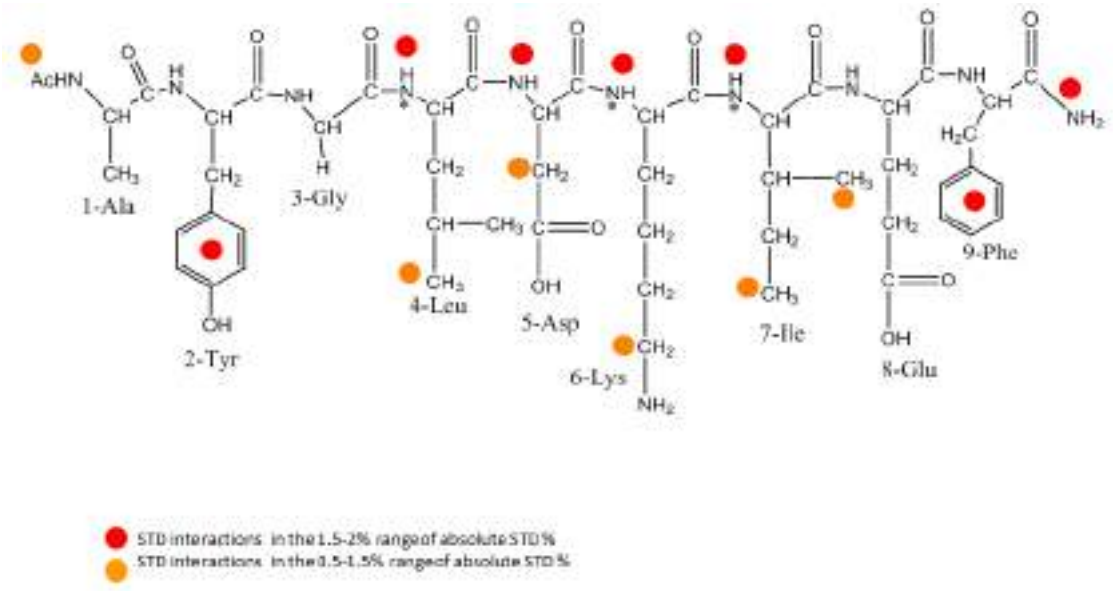


Figure 3.14: Absolute STD% values, grouped in two intensity ranges, for peptide interacting with VP24 toxins. The NH-amides of Leu, Lys and Ile, labeled with an asterisk, show the same chemical shift.

iment. The A peptide has six residues in common with peptide RS, nevertheless no interaction with the protein is observed. The observation that RS can actually form a complex with VP24 in solution and the overall agreement between NMR and computational data, support our initial hypothesis that the selected peptide could bind to VP24 and suggest the plausibility of the proposed binding site.

3.5.2 Circular Dichroism

As a further confirmation of the interaction between RS and VP 24, far-UV circular dichroism (CD) experiments were performed. This technique is, indeed, particularly useful for studying protein folding and interactions [100]. In 50 mM PBS at pH 7.4, RS peptide failed to adopt an ordered conformation, exhibiting a random coil structure (Figure 3.15, red line). In the same condition, VP24 is characterized by two negative bands at 210 and 225 nm (amide $n \rightarrow \pi^*$ transition), that are typical of α-helix (Figure 3.15, black line). On the other hand, the absence of a positive band below 200 nm (amide $\pi \rightarrow \pi^*$ transition) indicated that VP24 is present in solution as a mixture different conformations. VP24-RS interaction was then evaluated. A solution containing a mixture of RS and VP24 in 50 mM PBS was incubated at room temperature and then CD spectrum was recorded. A comparison between the spectrum of experimental mixture (Figure 3.15, blue line) and the arithmetic sum of the two separated components (Figure 3.15, green line) highlighted that, actually, an interaction occurred. Furthermore, the decrease in intensity of the negative Cotton effects at 210 and 225 nm in the experimental mixture indicated a loss of the helical

content of the protein, upon this interaction.

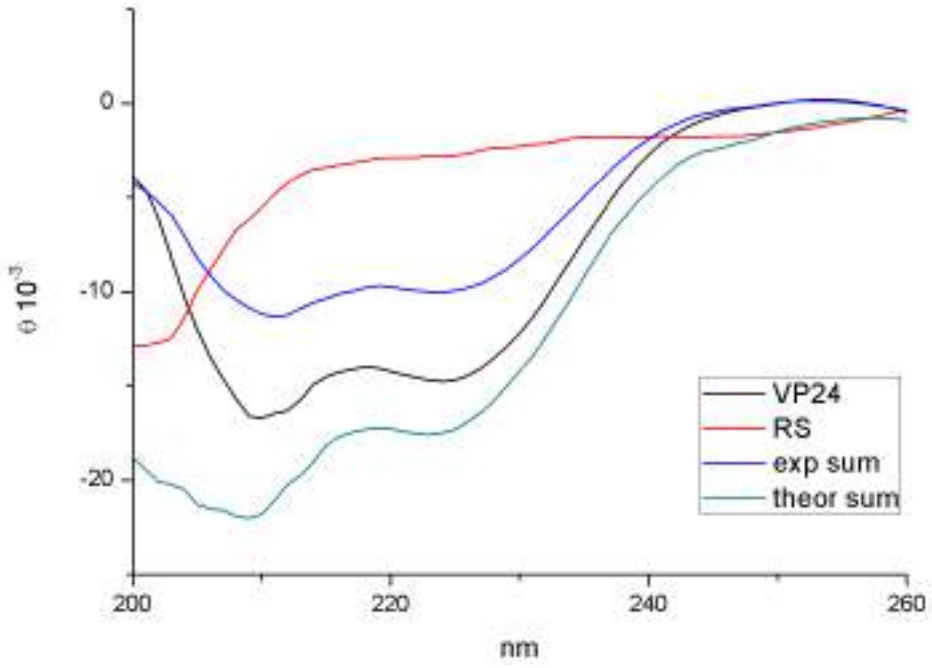


Figure 3.15: CD spectra of VP24 (black line), of peptide RS (red line), and of VP24/RS experimental mixture (blue line). The green line is the arithmetic sum of VP24 and RS single spectra.

3.6 Stapled peptides

As confirmed by CD and MD simulations, the RS peptide is unstructured, showing no secondary structure features. However, when inserted in parent protein KPNA, RS peptide exhibit an α-helical structure (Figure 3.11), that could play an important role in the interaction with VP24. For this purpose, starting from its original sequence, we designed two stapled peptides. The term stapling refers to bond covalently the sidechains of two amino acids to permanently reinforce a particular cyclic restricted structure. This gives the peptide enhanced biophysical properties, such as conformational stability and specificity for their biological target [101].
In Table 3.4 the sequences of the stapled peptides are shown. To build Mork peptide we replaced Gly478' and Ile482' with an aspartate and an ornitine respectively, while for Mindy we analogously substituted Leu479' and Glu483'. As can be noticed, the replaced residues are in position *i, i+4*. Moreover we capped the peptides with an acetyl group (ACE) and a NH_2 (NHE). Ornitine and aspartate sidechains were bound with an amidic bond as shown in Figure 3.16.

RS	Mork	Mindy
ACE	ACE	ACE
ALA	ALA	ALA
TYR	TYR	TYR
GLY	GLY	**ORC**
LEU	**ORC**	LEU
ASP	ASP	ASP
LYS	LYS	LYS
ILE	ILE	**ASC**
GLU	**ASC**	GLU
PHE	PHE	PHE
NHE	NHE	NHE

Table 3.4: Peptides sequences. In bold we have the two residues whose sidechains are linked through an amidic bond. ORC denotes an ornitine while ASC is an aspartate.

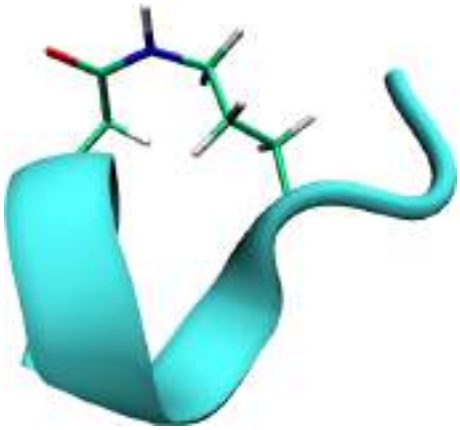


Figure 3.16: In cartoon representation the backbone of Mindy. In licorice the stapling. The aspartate and ornitine are bound through and amidic bond.

In order to assay the ability of the stapled peptides to maintain their secondary structure features with respect to RS, we performed a 100ns molecular dynamics simulation of Mork and Mindy in water monitoring their secondary structure. RS and Mork compleately lose their alpha helical character (Figures 3.17, 3.18), showing respectively no secodary structure and turn structure. This may be due to the fact that the Mork stapling involves the substitution of Glu483', with an aspartate. As shown in Section 3.4, Glu483' becomes an hot-spot when the peptide is extracted from KPNA (Figure 3.12). This substitution can affect its ability to recover its helical structure. On the other hand Mindy initially retains its alpha helical structure (residues ranging from Gly478' to Glu483') but after a few nanoseconds the peptide gains a 3_{10} helical structure. 3_{10} helix differs in its hydrogen bonds pattern from the α-helix: the carbonyl-amide hydrogen bonds are between residues i and $i+3$ in the former and between residues i and $i+4$ in the latter. Moreover 3_{10} helices have been proposed to be intermediates in the folding/unfolding of α-helices [102].

3.7 Conclusions

In this work we analyzed at the atomic level the network of interactions responsible for the formation of the VP24-KPNA complex. The protein-protein interface has been characterized, identifying the hot-spots residues, that give a major contribution to binding energy. From this analysis, a nonapeptide ranging from Ala476' to Phe484' and comprising four hot-spots close to one another both in the tertiary and primary structure ok KPNA has been identified. The ability of the selected peptide to actually interact with VP24 has been assayed with Saturation Transfer Difference NMR, which allowed to map the interaction epitope by determining the peptide re-

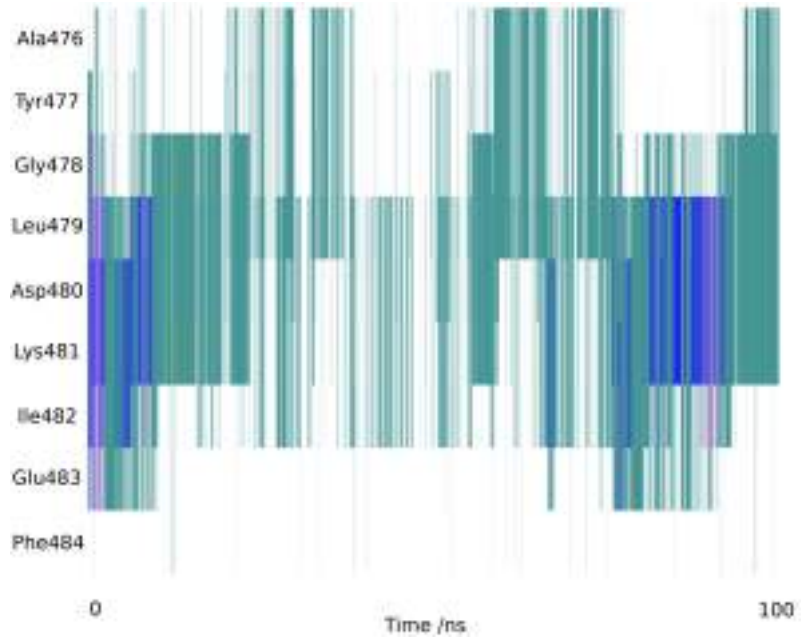
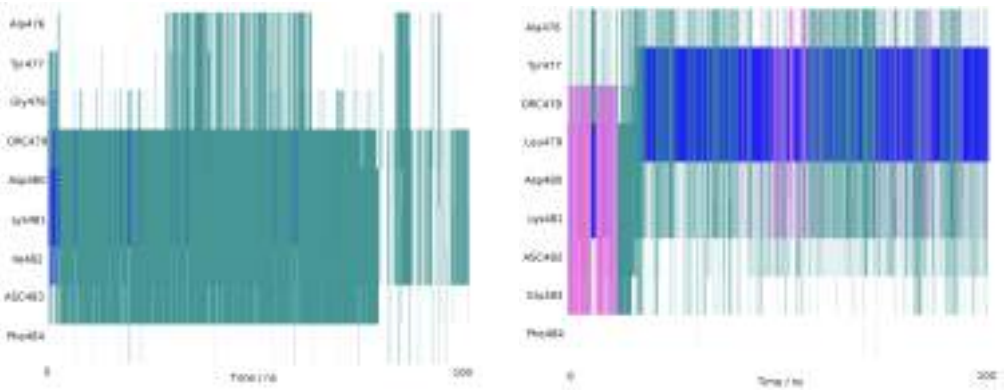


Figure 3.17: Secondary structure analysis for the RS peptide. The colour code represents the belonging of residues to a certain type of secondary structure.
White: no secondary structure. Pink: alpha helix. Green: turn region. Blue: 3_{10} helix.



Mork Mindy

Figure 3.18: Secondary structure analysis of the stapled peptides. In the left panel Mork peptide didn't show an helical secondary structure, the cyan colour indentify a turn region indeed. On the right panel Mindy peptide loses its alpha-helical structure at the beginning of the dynamics (pink region) and gains a 3_{10} helical structure (blue region).

gions in contact with the protein receptor. On the whole, the epitope sketched by NMR is consistent with the binding mode proposed by molecular modeling. The interaction between the selected RS peptide and VP24 was assessed also by far-UV CD experiments. Finally, we designed two stapled peptides in order to stabilize RS helical secondary structure even when extracted from its protein environment. One of them, Mindy, partially succeed in this task, showing a 3_{10} helical structure during most of the dynamics. VP24 is a potential target to reduce Ebola virulence and lethality, but its structure, lacking well defined pockets and grooves is far from being easily druggable. The identification of the RS peptide, which interacts with VP24 through its surface, can be a useful starting point for further development of VP24 targeting active molecules.

Chapter 4

VP24 Pocket analysis and virtual screening

4.1 The importance of pockets in proteins

Nowadays, a large amount of proteins molecular structures are available, due to the advances in experimental (X rays and NMR) and computational methodologies, in hardware and instrumentation [103]. In computational chemistry it is quite common to use protein structure as a starting point for the drug design. In particular, most of the drugs that target disease-related proteins binds in a site usually located inside a concave pocket which allows the ligand to make many favorable contacts with the target protein [104]. The crystal structures of protein-protein complexes, however, often lack deep pockets or clefts at the interface region. The surface at the protein-protein interface is, indeed, wide and almost flat. Therefore the structure-based design of potential proteins inhibitors is considered quite difficult. Moreover, using a static structure for the design of a potential inhibitor is often rather inadequate. First of all, the structure of a protein depends to some extent on its environment, so its structure in the crystal and in solution can show significant differences, in particular concerning solvent exposed loops [105]. Second, the geometry of the binding site in the holo-structures can be biased by the presence of the drug itself. This can favour the binding of ligands with steric properties similar to the co-crystallized inhibitor [106]. Finally, the crystal structures represent a conformation averaged on thermal fluctuations. [107]. So the actual dynamics of the protein itself has to be considered while designing a potential binder. The variability of the protein structure can be taken into account using a large number of snapshots taken from a molecular dynamics simulation. Moreover, this could allow the identification of transient pockets that may not be present in the crystallographic structures. Because there are strong evidences for the high plasticity of protein surfaces [108], we can assume that transient pockets that are large and deep enough to bind small-molecule

inhibitors may open from time to time. It has been shown that this happens on the nanoseconds time scale, which is accessible with MD simulations [107, 109]

4.2 Aim of the work

As discussed in Section 3.1 blocking the interaction between VP24 and Karyopherin could be a good therapeutic strategy to reduce Ebola virulence. One possible way to inhibit the formation of the protein-protein complex is to design a small molecule able to bind on the surface of interaction between VP24 and KPNA. This can be done exploiting the formation of transient pockets on the surface of VP24. Starting from the crystallographic structure we carried out a pocket analysis of VP24 using MD simulations in different solvents, in order to identify the best pocket in terms of volume, polarity and position.
During the last two decades, plenty of computational methods have been developed in order to predict the position, shape, and interaction properties of a ligand binding pockets starting from protein structures [110, 111]. In this work we employ the $EPOS^{BP}$ (Ensemble of Pockets on Protein Surfaces with BALLPass) software in which the PASS [112] (Putative Active Sites with Spheres) algorithm is re-implemented using the BALL [113] (Biochemical Algorithms Library) library. This software allows us to identify also dynamic pockets and analyze their properties, such as relative frequency, and their average, minimum and maximum pocket properties, namely volume, polarity, and depth. Now we will discuss, more in detail, the geometrical basis of the PASS algorithm.

PASS PASS algorithm is designed to fill cavities in a protein structure with a set of spheres located at the centers of binding pockets. Initially a coating of probe spheres is calculated with the protein as substrate, then additional layers of probes are placed onto the previously found probe spheres. Only probes with low solvent exposure are retained and the routine finishes when an accretion layer produces no new buried probe spheres. This is a purely geometrical algorithm. Different steps of the algorithm are showed in Figure 4.1.

Once the best pocket has been identified, we performed a virtual screening (Subsection 1.8.2) of a database of known compounds to asses their ability to bind VP24 thus inhibiting its interaction with KPNA.

4.3 Computational details

Molecular dynamics VP24 crystallographic structure was obtained from the complex VP24-KPNA deposited in the Protein Data Bank (PDB ID: 4U2X [82]).

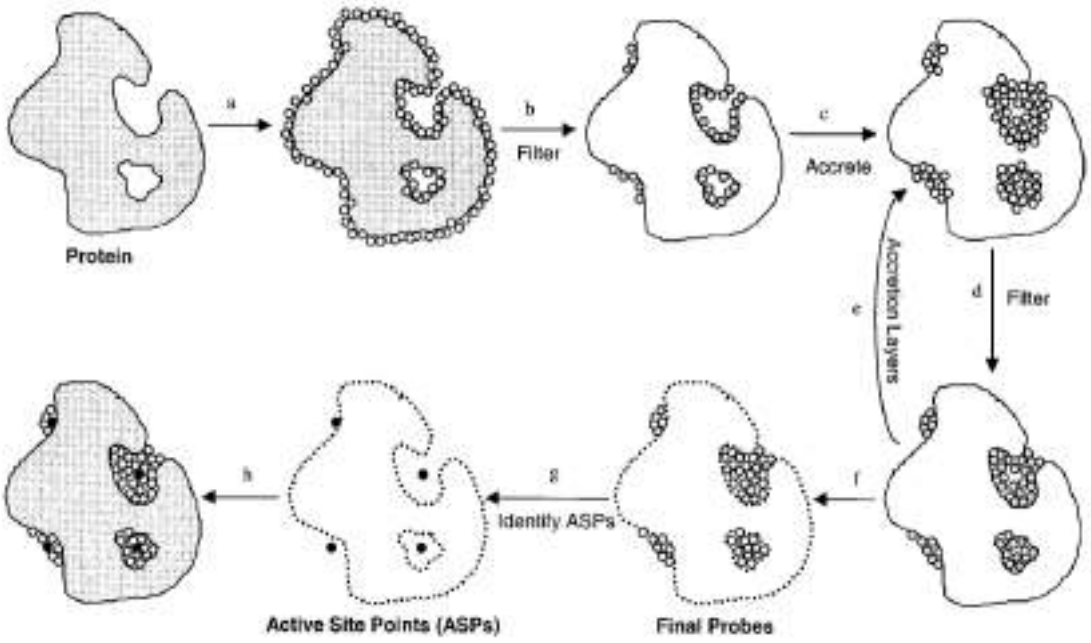


Figure 4.1: PASS algorithm. (a) PASS uses three-point geometry to coat the protein with an initial layer of spherical probes. (b) These probes are filtered to eliminate those that (i) clash with the protein, (ii) are not sufficiently buried, and (iii) lie within 1 Å of a more buried probe. (c) A new layer of spheres (white) is accreted onto a scaffold consisting of all previously identified probes (shaded). (d) The probes are filtered as described in step b. (e) Accrete a new layer of spheres onto the existing probes, as in step c. (f) Accretion and filtering (steps e and d) are repeated until a layer is encountered in which no newly found probes survive the filters. This leaves the final set of probe spheres. (g) Probe weights (PW) are computed for each sphere and active site points (ASPs) are identified from amongst the final probes. (h) The final PASS visualization is produced. By default, the final probe spheres are first smoothed, leaving only clusters of four or more. Image taken from Ref [112].

The residues numbers used here are referred to this structure. Molecular dynamics simulations were carried out using GROMACS 4.5.3 [19] package using explicit solvent (TIP4P water [94]) and periodic boundary conditions. AMBER99SB-ildn [68] force field was used. The overall charge of the systems was neutralized with Cl^- or Na^+ ions. The LINCS algorithm [4] was employed to constraint all bonds to their equilibrium length, allowing a time step of 2 fs. The system was submitted to 10000 steps of geometry optimization with the steepest descent method. Afterwards it was equilibrated for 200 ps in NVT conditions (T=300 K) and subsequently for 200 ps in NPT conditions, in order to equilibrate systems density. The cosolvents (methanol and 2-methyl-2-propanol) were parametrized using the Generalized Amber Force Field [69]. The productive phase of MD was performed in NPT conditions (1 bar, T=300 K). Temperature and pressure were kept constant at their reference value using the velocity rescale algorithm [71] and the Berendsen barostat [72], respectively. A 14 Å cutoff was applied for non-bonded interactions and the Particles Mesh Ewald algorithm [73] was employed to calculate long range electrostatic interactions. The cluster analysis was carried out with the utility *g_cluster* using the GROMOS clustering method [114].

Pockets identification and virtual screening The EPOS software was employed for identification of the pockets on the protein surface. The system was set up for the virtual screening using Chimera Dock Prep tool [115]. This tool deletes solvent and ions, adds missing hydrogens and adds charges to the system, using the Amber force field [3]. DOCK6 spheres of radius between 1.4 and 4.0 Å were used to map the chosen pocket, centered on residue Leu201. In order to describe the pocket we selected spheres at a maximum distance of 5.5 Å from Leu201. Ligands were taken from the ZINC database [116]. Docking calculations were carried out with DOCK6 [45] using the Grid Score scoring function, using a grid of spacing 0.4 Å.

4.4 Pocket analysis

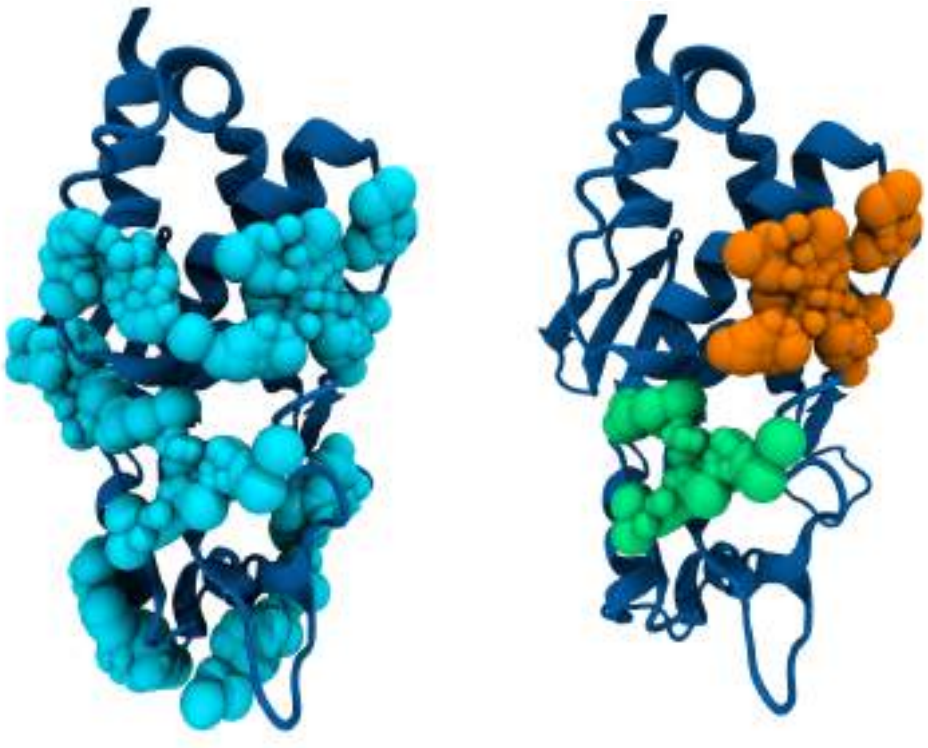
4.4.1 Crystallographic structure

Starting from the crystallographic structure of VP24 we performed a pocket analysis using the EPOS software. The 7 pockets identified are listed in Table 4.1. Two of them, labelled with Q and Y, attracted our attention because they are located at the interface with KPNA (Figure 4.2). Moreover, they have the largest volume compared to the others.

- Residues involved in Y pocket: 86 92 127 128 185 186 201 202 203 204 216 217 218.

PID	Volume / Å^3	Polarity	Depth / Å
Y	444.520	0.351	7.427
Q	648.177	0.354	3.848
2	247.232	0.304	7.077
3	305.199	0.385	5.496
4	532.360	0.343	6.558
5	338.725	0.351	7.462
6	385.095	0.365	8.184

Table 4.1: 7 pockets found in the crystallographic structure.



A B

Figure 4.2: A panel: The 7 pockets identified for VP24 crystallographic structure.
B panel: The Y pocket (highlighted in green) and the Q pocket (highlighted in orange).

- Residues involved in Q pocket: 99 100 103 107 109 110 111 113 116 117 119 120 121 123 124 127 128 137 138 140 181 183 185 186 188.

These are static pockets, obtained from the analysis of the crystallographic structure. As we are interested in the evolution of the protein structure and of pocket shape and properties with time we performed a MD simulation in water and in more apolar solvents: methanol and 2-methyl-2-propanol.

4.4.2 MD in water

We performed a 100 ns molecular dynamics simulation, in order to obtain a statistical ensemble of conformations to evaluate how shape and properties of the Q and Y pockets change during the MD and to evaluate the formation of possible transient pockets. We extracted 10000 frames from the trajectory and we performed a cluster analysis using the GROMOS method with a 0.1 nm cutoff on the C_α RMSD, obtaining 24 clusters. We analysed only the central structure of each cluster. Unfortunately we didn't identify any transient pocket. Anyway we analyzed the variation in volume, depth and polarity of the Q and Y pockets. These informations are collected in Table 4.2. Y and Q pockets are present during alomost the whole dynamics, maintaining a quite large mean volume and depth.

4.4.3 MD in methanol and 2-methyl-2-propanol

To asses the stability of Q and Y pockets in more apolar solvents and further try to identify transient pockets, we performed two 100 ns MD simulations in methanol and in a 1M water solution of 2-methyl-2-propanol. This was done to mimic the dynamic of the interface of VP24 in complex with KPNA, as the protein-protein interaction surface is usually apolar. We extracted 10000 frames from each trajectory and we performed a cluster analysis using the GROMOS method with a 0.1 nm cutoff obtaining 31 clusters for methanol and 10 clusters for the 1M solution of 2-methyl-2-propanol. We analysed the central structures of all the clusters but we didn't identify any interesting transient pocket. We then monitored the properties of the pockets Q and Y (Table 4.2). Again, both pockets can be identified in the vast majority of structures.

MD in water - 24 clusters

Name	freq[%]	mean vol[Å^3]	min vol[Å^3]	max vol[Å^3]	mean pol	min pol	max pol	mean depth[Å]	min depth[Å]	max depth[Å]
Y	100	428.086	122.018	701.815	0.345	0.188	0.398	5.64	3.97	7.26
Q	91.7	469.819	241.406	684.39	0.353	0.306	0.406	7.07	3.55	11.70

MD in methanol - 31 clusters

Name	freq[%]	mean vol[Å^3]	min vol[Å^3]	max vol[Å^3]	mean pol	min pol	max pol	mean depth[Å]	min depth[Å]	max depth[Å]
Y	100	482.079	247.487	828.073	0.355	0.289	0.385	6.67	3.75	9.85
Q	93.5	526.459	300.240	822.167	0.340	0.279	0.396	7.07	1.23	11.04

MD in 2-methyl-2-propanol 1M - 10 clusters

Name	freq[%]	mean vol[Å^3]	min vol[Å^3]	max vol[Å^3]	mean pol	min pol	max pol	mean depth[Å]	min depth[Å]	max depth[Å]
Y	100	392.647	236.025	546.192	0.333	0.306	0.358	6.07	3.66	7.60
Q	90	528.34	343.007	659.117	0.353	0.325	0.380	7.16	5.90	9.81

Table 4.2: Comparison between Q and Y pockets in different solvents

Unfortunately this pocket analysis didn't lead to the identification of any transient pockets suitable for virtual screening. Anyway we found out that both Q and Y pocket are conserved during the dynamics, in water and in more apolar solvents. Based on this results we decided to use the crystallographic Y pocket (present in the 100% of the dynamics) as the binding site for a virtual screening.

4.5 Virtual Screening

4.5.1 Database choice

With our computational resources we are able to perform the docking of a maximum of 500k compounds. Knowing this limitation we decided to select some subcategories of the ZINC database. We started docking Natural products databases, comprising: AfroDb Natural Products, AnalytiCon Discovery NP, IBScreen NP, Indofine Natural Products, Specs Natural Products, UEFS Natural Products and TCM Database @ Taiwan. Then we moved to ZINC Purchasable databases, in particular: 3B Scientific Corporation, BioBlocks BB, ChemDiv BuildingBlocks, Sigma Aldrich (Building Blocks). Finally we docked the database Edrugs-3D, that contains FDA (Food & Drug administration) approved drugs and active metabolites. The docking was performed with DOCK6 using the *Grid score* scoring function for the energy evaluation. We *a priori* discarded molecules with a MW $< 150\ g\ mol^{-1}$ and MW $> 500\ g\ mol^{-1}$ and with a formal charge greater than +4 and lower than -4. To analyze the docked structures we used an energetic criterion, selecting molecules with at least a 2.5 $kcal\ mol^{-1}$ binding energy per heavy atom. Moreover we discarded molecules with more than 15% of electrostatic contribution to the overall energy.

4.5.2 Docking results

3B Scientific Corporation The database has more than 74000 molecules. The molecular docking led to the identification of two major classes of compounds, showed in panels A and B of Figure 4.3. Compound in Panel A is a triazine with three substituents pointing towards different direction within the binding site of VP24. Panel B compound is called olopatine and it is sold as a prescription eye drop. The binding energy of both compounds is around $-50\ kcal\ mol^{-1}$.

AfroDb Natural Products The database contains 954 compounds. Among the docked compounds only one met the filter criteria in this database. It is shown in Panel C in Figure 4.3. The binding energy of this compound is $-42\ kcal\ mol^{-1}$.

AnalytiCon Discovery NP The database has 5154 molecules. Only one docked compound met the filter criteria and it is shown in Panel D of Figure 4.3. This compound is called haematoxylin, it is used as a stain and has a binding energy of $-41\ kcal\ mol^{-1}$. The polycyclic motif is similar to the one found in the database 3B Scientific Corporation (Panel B).

BioBlocks BB This database contains 5840 compounds. Three molecules met the filter criteria we applied. All of them have the same polycyclic structure as the one shown in Panel E of Figure 4.3. The binding energy is around $-43\ kcal\ mol^{-1}$.

ChemDiv BuildingBlock The database has 25952 molecules. From the docking results analysis we obtained only one compound that met the filter criteria and it is shown in Panel F of Figure 4.3. The binding energy is $-41\ kcal\ mol^{-1}$.

Edrugs-3D The dataset has 1632 molecules. The most interesting molecules arising from the docking calculations are tioconazole, econazole (Panel G and H in Figure 4.3) and other molecules with similar structure. These molecules are used as antifungal medications and belong to the imidazole class. The average binding energy is around $-38\ kcal\ mol^{-1}$. Moreover also olopatadine, already emerging from the analysis of database 3B Scientific Corporation (Panel B), was found among the best molecules for this database.

IBScreen NP The database has 84215 molecules. We observed two major classes of molecules. The first class comprise molecules similar to the one showed in Panel I of Figure 4.3. The binding energy is lower compared to the other selected molecules, $-47\ kcal\ mol^{-1}$. The second class has polycyclic compounds with various substituents (Panel E) similar to the ones found in the BioBlocks BB database. Also from the energetic point of view we obtain similar energies, around $-43\ kcal\ mol^{-1}$.

Sigma Aldrich (Building Blocks) The database contains 41718 compounds. Interestingly, the most intresting compounds result to be the same antifungal molecules we found in Edrugs-3D database (Panel G and H in Figure 4.3).

Specs Natural Products, TCM Database @ Taiwan, UEFS Natural Products These databases contain respectively 1489, 36043 and 473 molecules. After the application of filters we didn't find any molecule that met the filter criteria.

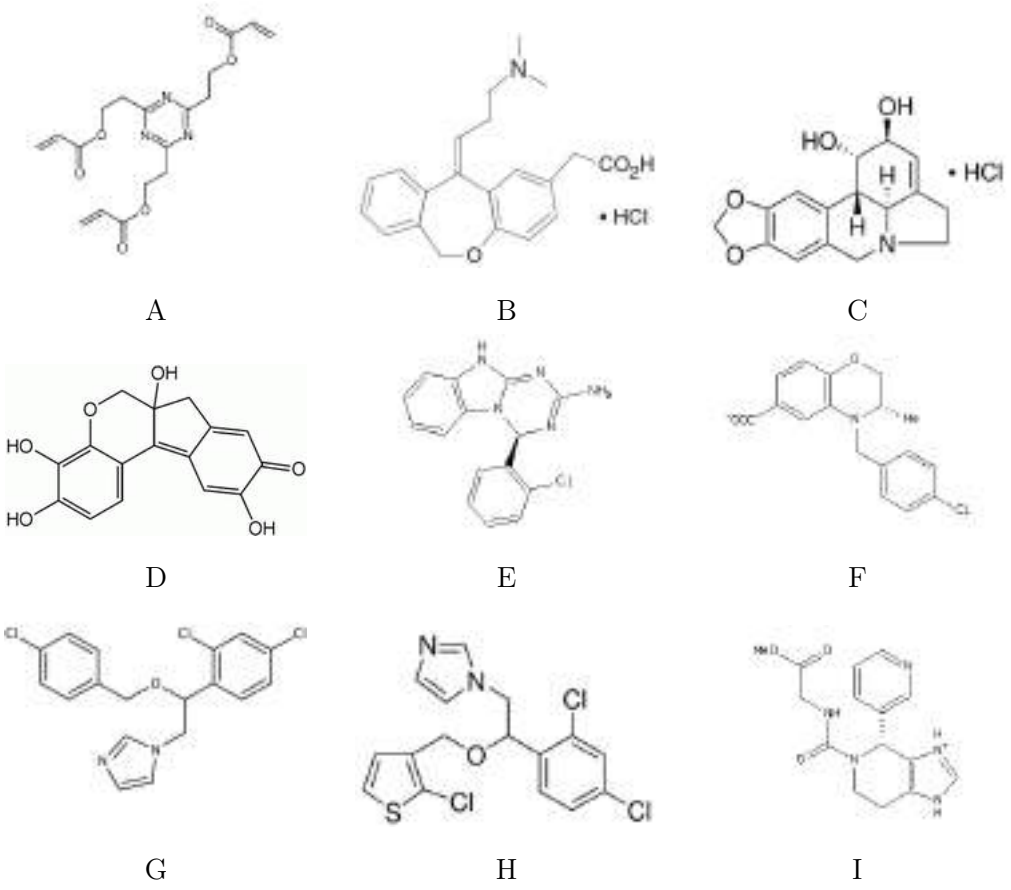


Figure 4.3: Representative compounds arising from the analysis and selection of the virtual screening results.

4.5.3 Discussion

Dealing with the analysis of a huge amount of data is not easy, and we had to choose a criterion to filter all the docked structure coming from the virtual screening. We already decided to filter the compounds in input on the basis of the MW. We kept molecules with $150g\ mol^{-1} < MW < 500g\ mol^{-1}$ and with a formal charge between -4 and +4. To filter the molecules arising form the docking calculations we used an energetic criterion, based on the energy per heavy atom: we decided to keep molecules with a binding energy of 2.5 $kcal\ mol^{-1}$ per heavy atom or more. Finally we discarded compounds with an electrostatic contribution to the overall energy of 15% or more. As we can see from Figure 4.3 the majority of the selected compounds are polycyclic (Panels B C D E F I), and some of them appear to be very similar (Panel B and D). Moreover the analysis on two different databases (Sigma Aldrich (Building Blocks) and Edrugs-3D) led to the identification of the same class of compounds: the antifungals tioconazole and econazole. Evenly, the analysis of both 3B Scientific Corporation and Edrugs-3D databases identified the same compound: olopatadine (Panel B). The three non-polycyclic compounds (Panels A G H) have a feature in common: all three of them have three substituents pointed towards different directions within the binding site of VP24, as shown in Figure 4.4.

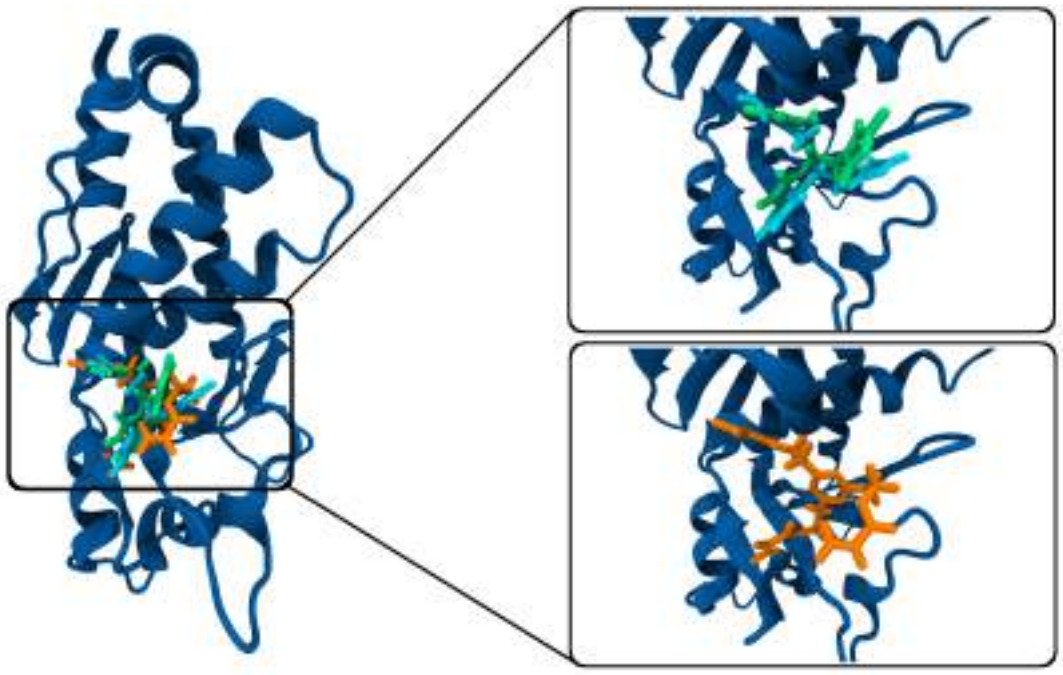


Figure 4.4: VP24 in complex with the docking poses of tioconazole (green), econazole (cyan) and triazine based compound (orange)

4.6 Conclusions

In this study we carried out a pocket analysis on VP24 and we monitored their evolution in time. We identified a small number of compounds that could be able to bind at the interface with KPNA. The pocket analysis on the crystallographic structure highlighted the presence of two interesting pockets at the interface VP24-KPNA: Q and Y pockets. Properties of these pockets were studied with MD simulations in water and in two more apolar solvents, showing that the pockets are conserved during the whole dynamics. Unfortunately MD didn't highlight the presence of any transient pocket. Crystallographic Y pocket was then chosen as the binding site for a virtual screening on approximately 250k molecules. The Virtual screening results were finally filtered with an energetic criterion and led to the identification of a small number of compounds. These compounds have to be assayed biologically in order to test if they are able to inhibit the formation of the protein-protein complex VP24-KPNA.

Chapter 5

VP24 behavior in presence of osmoprotectants

5.1 Osmolytes

Osmolytes are solutes able to affect cell osmosis. In nature there are plenty of strategies in order to allow plants and animals to survive in severe conditions, such as high pressures and temperatures or high salinity. One of them is the accumulation of highly soluble molecules, able to reduce osmotic stress of the cells, in cellular fluids. A well known example is the resurrection plant (Figure 5.1) which is able to "die" in the dry season and "resurrect" in the rainy one [117, 118].

Figure 5.1: The resurrection plant, *Selaginella lepidophylla*. [1]

Many different small organic molecules are known to serve as osmolytes. They can be grouped in a few major chemical categories: small carbohydrates including sugars (trehalose), polyols (glycerol, inositols, sorbitol, etc.) and derivatives (such as o-methyl-inositol); amino acids (glycine, proline, taurine, etc.) and derivatives (e.g.

[1] Image taken from: http://www.sketchstationery.com/

ectoine); methylamines (such as N-trimethylamine oxide (TMAO) and glycine betaine) and methylsulfonium solutes including dimethylsulfonopropionate (DMSP); and urea. [119, 120]. In addition to their ability to control cell water loss or gain, some osmolytes are also able to stabilize the native, folded structure of proteins against either thermal or chemical denaturation [121]. These osmolytes are called osmoprotectants or compatible solutes due to the high concentration they can reach in living organisms without toxic effects. On the other hand, other osmolytes such as urea, stabilize the unfolded structure of proteins and they are called denaturants. Regardless the protein, the stability of the folded/unfolded state is linearly dependent on the concentration of osmolyte/denaturant [122].
Besides their biological functions, osmoprotectants play an important role in different fields:

- Agriculture: Plants able to synthesize this kind of compounds could be more drought resistant and able to survive with less water for irrigation.
- Cosmetics: The introduction of osmoprotectants in creams can bring many benefits. For instance epithelial cells will be less likely to undergo dehydration phenomena.
- Ophthalmology: Osmoprotectants have recently found application in the treatment of dry eye, to compensate extracellular hyperosmolarity without interfering with cellular metabolic processes [123].

Although the diffusion of osmoprotectants in different fields, their effect on the protein folded state is still unknown. There are different hypotheses that try to rationalize the effect of osmoprotectants and denaturants. One ascribes the effect of osmolytes to their **interaction with the protein backbone**, that varies the stability of the protein itself [124–127]. The second hypothesis deals with the **role of the solvent**. The water structure is weakened by denaturants, increasing the solvation of the protein hydrophobic residues. Osmoprotectants strengthen water hydrogen bond network leading to the opposite effect [128–131]. A third hypothesis was proposed by Bolen [132]. In proteins native state the percentage of backbone exposed to the solvent is lower with respect to the denatured state. Bolen purposes an **osmophobic effect**, in which the driving force of the osmoprotection is the direct, unfavorable interaction between osmoprotectants and protein backbone, which leads to an exclusion of osmoprotectants from the protein domain. Osmoprotectants destabilize both the native and the denaturated structure, however they destabilize more the denaturated one, because it exposes more backbone residues.
In this chapter we study EBOV Viral Protein 24 in water and in presence of osmolytes, in order to understand how they influence the protein structure. The choice

of VP24 as a target was due to its clean circular dichroism spectrum. The effect of osmoprotectants on the secondary structure of the protein can be easily seen by a circular dichroism analysis. In this study we used urea (URE) as denaturant and ectoine (ECT), glycine betaine (GBE), taurine (TAU) and trimethylamine N-oxide (TMAO) as osmoprotectants. The osmolytes used in this work are shown in Figure 5.2. Analysis on the local solvent order, hydrogen bonds, diffusion coefficients were carried out.

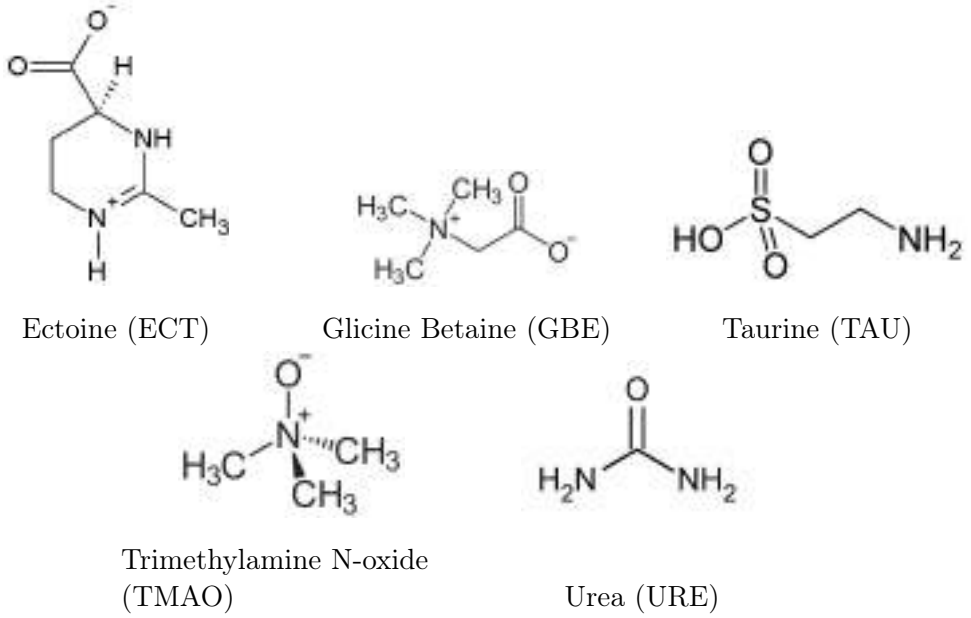


Figure 5.2: Molecular structures of osmolytes.

5.2 Computational details

The crystallographic structure of VP24 was obtained from the Protein Data Bank (PDB ID: 4M0Q [133]). Six different systems were set up: VP24 in water and VP24 in a 3 M solution of every osmolyte shown in Figure 5.2. Molecular dynamics simulations were carried out with the GROMACS 4.5.3 [19] package using explicit solvent and periodic boundary conditions. The AMBER99SB-ildn [68] force field was used. Every system was solvatd with TIP4P [94] waters and neutralized with Cl^- or Na^+ ions to reach neutrality. The choice of TIP4P water, which has a better description of the electrostatic potential around the molecule, is due to the fact that the aim of the work is to study solvent properties. The LINCS algorithm [4] was employed to constraint all bonds to their equilibrium length, allowing a time step of 2 fs. The system was submitted to 10000 steps of geometry optimization with the steepest descent method. Afterwards it was equilibrated for 200 ps in NVT conditions and subsequently for 200 ps in NPT conditions, in order to equilibrate

systems density. Two sets of 50 ns molecular dynamics were performed in NPT conditions, one at 300 K and the second at 330 K. Temperature and pressure were kept constant at their reference value using the velocity rescale algorithm [71] and the Berendsen barostat [72], respectively. A 14 Å cutoff was applied for non-bonded interactions and the Particles Mesh Ewald algorithm [73] was employed to calculate long range electrostatic interactions.

5.3 Results and Discussion

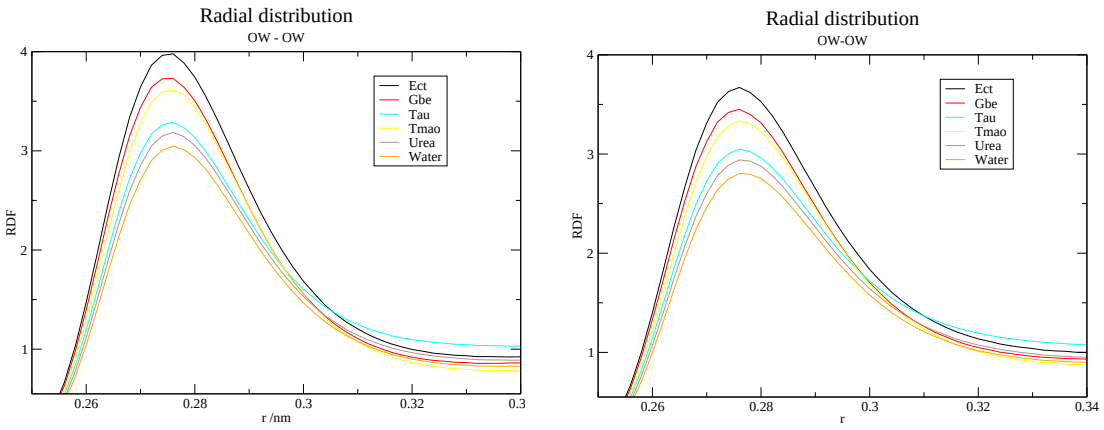
Molecular dynamics simulation of VP24 in water and in different solutions of osmolytes were performed. A 3M concentration of osmolytes was used, and all the dynamics were performed at two different temperature: 300 K and 330 K. In order to understand how the presence of osmolytes influences the behavior of the protein, general properties of VP24 were monitored for each simulation. In Table 5.1 water self diffusion coefficients were computed using g_msd utility.

Solvent	D / $10^{-5} cm^2\ s^{-1}$	
	300 K	330 K
H_2O	3.74	6.33
Tau 3M	1.93	3.18
Ect 3M	0.94	1.86
Gbe 3M	1.37	2.46
Tmao 3M	1.75	3.30
Urea 3M	3.19	5.32

Table 5.1: Water self-diffusion coefficients for all the studied systems.

We can notice that all osmolytes reduce the water self-diffusion coefficients at 300 and 330 K. However, osmoprotectants dramatically reduce the self-diffusion coefficients compared to denaturants such as urea. At 300 K the presence of ectoine reduces the water self-diffusion coefficient by 75%. For urea the reduction is roughly of the 15%. We can notice the same behavior at 330 K. Ordering the structure of water (namely decreasing its diffusion coefficient) entails an entropic penalty for water itself to act as primary denaturant, thus making the denaturation of the protein more difficult even at high temperatures.
We can clearly see the effect of osmoprotectants on the order of water molecules also from the analysis of the Radial Distribution Function (RDF) of water shown in Figure 5.3.

The analysis was carried out at 300 and 330 K for all the systems studied. As we can notice, all osmolytes induce an higher degree of order in water structure. In particular osmoprotectants show an much higher RDF value with respect to water



300 K 330 K

Figure 5.3: Radial distribution functions. We focus our analysis on the first peak of the function.

and urea. Moreover the RDF peak of the systems simulated at 330 K is less intense, due to the higher temperature that induces more kinetic energy in water molecules, resulting in a more disordered structure.

The average number of hydrogen bonds between protein-osmolyte, protein-water and osmolyte-water was calculated and the results are reported in Table 5.2. The average number of hydrogen bonds between protein and urea is one order of magnitude greater with respect to the osmoprotectants. This supports the hypothesis of urea acting as backbone replacer. Urea, indeed, interacts more with the protein, leading to a crowding of urea molecules in the proximity of the protein, as will be discussed later. From Table 5.2 we can also notice that TMAO has the tendency to form less hydrogen bonds compared to the other osmoprotectants; it has, indeed, only one acceptor of hydrogen bond (Figure 5.2).

Solvent	Average Number of Hydrogen bonds					
	300 K			330K		
	Protein-osmolyte	Protein-H_2O	Osmolyte-H_2O	Protein-osmolyte	Protein-H_2O	Osmolyte-H_2O
H_2O	-	452.77	-	-	440.24	-
Tau 3M	73.17	366.94	17541.00	73.54	361.78	17188.00
Ect 3M	71.59	362.84	15273.53	73.97	349.50	15047.20
Gbe 3M	36.66	384.36	11593.70	39.50	354.309	11393.40
Tmao 3M	18.15	400.25	8845.26	15.82	387.71	8498.81
Urea 3M	145.51	332.35	15273.53	130.25	332.43	13802.11

Table 5.2: Average number of hydrogen bonds in the studied systems.

The analysis on the preferential coefficient (Section 1.9) was carried out for the systems at 300 K, the results are shown in Figure 5.4. The preferential coefficient shows how many molecules of cosolvent are present at a certain distance from the system, with respect to the bulk. If the preferential coefficient is lower than zero it means that the cosolvent is excluded from the protein domain; this behavior is verified for all the osmoprotectants. On the other hand if the value is greater than zero, it shows a crowding of the cosolvent near the system, and this happens with urea. This results support the osmophobic effect theory proposed by Bolen [132].

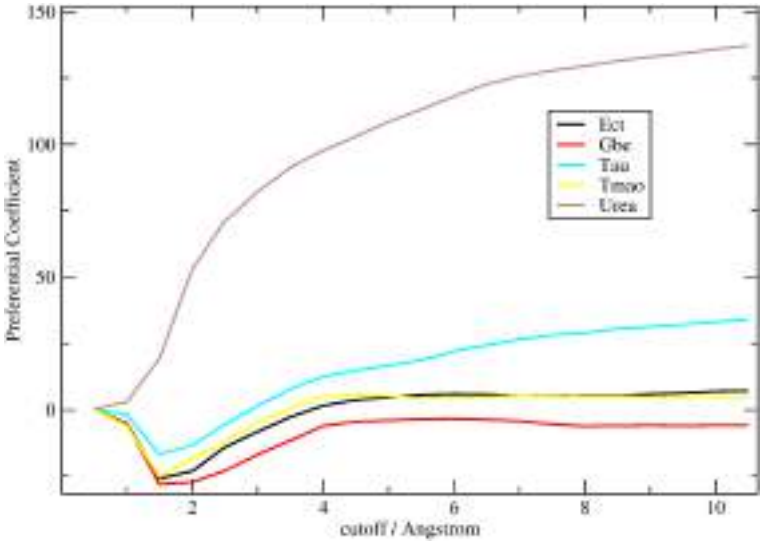


Figure 5.4: Preferential Coefficient

Finally the solvent density function was computed (Section 1.9). This function describes how the molecules are distributed around the protein. It is, in principle, equivalent to the radial distribution function, but it takes into account the shape and the volume of the protein itself. The results are shown in Figure 5.5. The absence, for all osmoprotectants, of a well-defined and prominent peak, shows that they are excluded form the protein neighborhood. On the other hand, urea shows a very intense peak supporting, once again, its tendency to act as a backbone replacer and stabilize the denaturated state of the protein, which exposes more backbone.

5.4 Conclusions

In this work we studied the effect of osmolytes on VP24. The obtained values of preferential coefficient and solvent density function show that osmoprotectants and urea have opposite effects on the protein. Osmoprotectants are excluded from the protein domain, while urea molecules crowd in the proximity of the protein. Both these effects support the osmophobic effect theory [132]: the interaction between the osmoprotectants and the protein backbone is unfavorable, so they are excluded

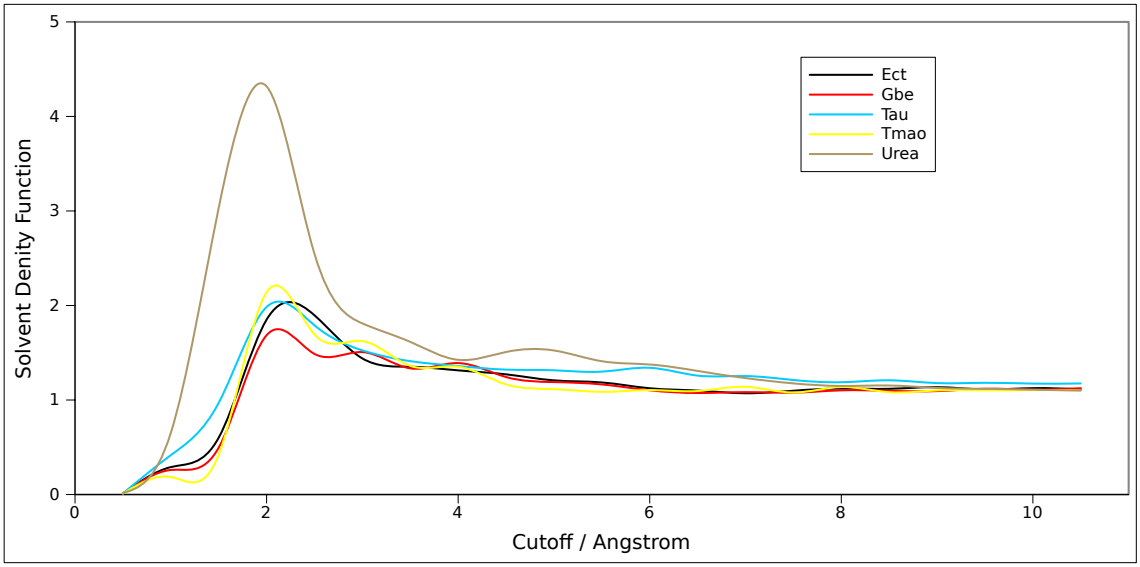


Figure 5.5: Solvent density function

from the protein domain. Moreover the analysis on hydrogen bonds, RDF and diffusion coefficients highlight how osmoprotectants can influence the order of the solvent, namely increasing the degree of order of water. In this case osmoprotectants prevent water to act as primary denaturant, ordering its structure.
In future $\Delta G_{unfolding}$ of VP24 has to be calculated and circular dichroism analysis on these system have to be performed, in order to asses the effect of osmolytes on the unfolding temperature of VP24.

Concluding remarks and future perspectives

In this PhD thesis different inhibition strategies for Ebola virus have been explored, targeting two of the seven viral protein coded by Ebola genome: VP24 and VP35. The interaction between VP35 C-terminal interferon inhibitory domain and known inhibitors of this protein is studied at the atomistic level in **Chapter 2**, highlighting new important key interactions between the inhibitors and the protein. Essential dynamics analysis pointed out an interesting closing motion that is characteristic of the apo-form, and it is reduced in amplitude when the protein is bound to inhibitors. The functional role of these motions in the interaction with viral NP is not clear yet. Recent literature [134, 135] shows that the N-terminal domain of VP35 is actually interacting with viral nucleoprotein. The collective motions highlighted in this study can, therefore, play an relevant role at the allosteric level. This subject deserves further attention from molecular biologist.
VP24 was studied in Chapters 3, 4 and 5.
Chapter 3 dealt with the study of VP24-KPNA protein-protein complex. A computational alanine scanning allowed us to identify a KPNA subsequence with a high density of hot-spots. A nonapeptide was designed starting from this sequence and its ability to bind VP24 was tested both computationally and experimentally through NMR and CD. It is largely known, however, that peptides are not the ideal candidates to inhibit protein protein interactions *in vivo*, due to their poor metabolic stability and low bioavailability. Although, a peptide like this can serve as a starting point for the design organic inhibitors sharing the same pharmacophore with the peptide itself. Starting from the original sequence, stapled peptides were also desiged, as they can be less prone to metabolic degradation and, moreover, one of them maintains a 3_{10} helical structure that could improve the binding to VP24. We are, however, waiting for them to be synthesized, in order to test them on VP24 through NMR and CD.
In **Chapter 4** a pocket analysis on VP24 was carried out, which identified 2 interesting pockets at the interface with KPNA. Geometric and physical properties of these pockets were studied in different solvents through MD simulations, highlighting their

presence during the whole MD. A subset of ZINC database was then docked into one of these pockets and ranked through an energetic criterion using DOCK 6 scoring function. This led to the identification of a small number of potential VP24 ligands. It is known that DOCK scoring functions doesn't provide an accurate estimate of the binding energy however, due the large number of docked compounds (over 200k), this scoring function was the only fast and accessible method to energetically rank the compounds. For the most interesting molecules, more accurate analysis are in progress, including MD simulations and subsequent MM-PBSA calculation of the binding energies.

Chapter 5 dealt with the analysis of the folding stability of VP24 by explicit solvent MD both in the presence and absence of osmolytes. The analysis of different properties was carried out, including water self diffusion coefficient, water radial distribution function, average number of hydrogen bonds, preferential coefficient and solvent density function. Based on these results it is clear how osmoprotectants and urea have opposite effects on the protein, the former stabilizing the folded state and the latter shifting the equilibrium to the denatured state. As a further development experimental tests on melting temperature of VP24 in presence of these osmolytes through CD have to be carried out.

Part II

Coarse grain simulations of antimicrobial peptides

Chapter 6

Investigating SVS-1 pore formation mechanism

6.1 Antimicrobial peptides

Antimicrobial peptides (AMPs) are widely distributed in nature, being produced by mammals, birds, amphimians, insects, plants and microorganisms [136]. Although they form a diverse group of peptides as judged by their primary structures, they are often cationic and amphiphilic molecules with opposing hydrophobic and policationic faces. Their mechanism of action entails the disruption of cancer cell membranes. Indeed, during the apoptosis process the membrane undergoes both biochemical and structural changes. The morphological modifications include cell shrinkage, chromatin condensation, nuclear fragmentation and formation of dynamic plasma membrane protrusions called blebs, which eventually detach, culminating in fragmentation into apoptotic bodies [137]. From the biochemical point of view the membrane lipid distribution in cancer cells is different from the one in normal cells. Normal mammalian cells have an asymmetric lipid distribution of phospholipid type between the two leaflets of the lipid bilayer. The outer leaflet contains mostly zwitterion lipids, displaying an overall neutral charge. The inner leaflet contains mainly phosphatidylethanolamine and phosphatidylserine (PS), resulting in an overall negative charge [138, 139]. Loss of lipid asymmetry, due to translocation of PS in the outer leaflet, is found to occur in a number of tumor cells, allowing this negatively charged surface to be used as a target in cancer therapy (Figure 6.1). The cationic face of AMPs is responsible for engaging the negatively charged surface of the cancer cell through hydrogen bonds and electrostatic interactions, the amphiphilic face enables membrane permeabilization.

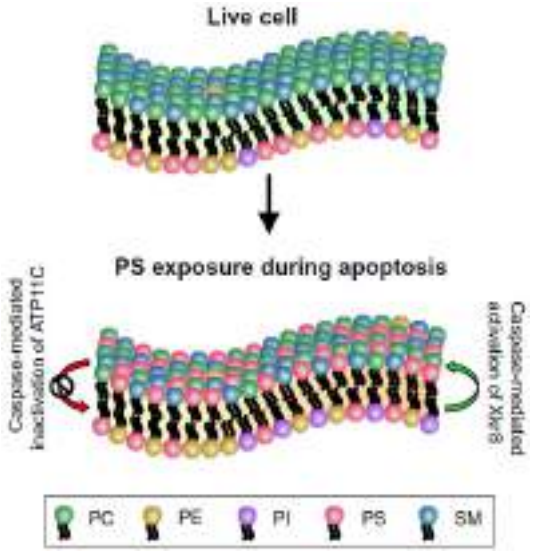


Figure 6.1: Different lipid distribution in a normal cell membrane (upper panel) and in a cancer cell membrane (bottom panel). PC: phosphatidylcholine. PE: phosphatidylethanolamine. PI: phosphatidylinositol. PS: phosphatidylserine. SM: Sphingomyelin. Image taken from Ref [137].

6.2 SVS-1

Most of cationic AMPs are unstructured in solution and gain their secondary structure, usually helical, at the cell surface. The driving force of the folding is the favorable interactions between the membrane surface and the peptide [140]. Even if the biological active conformation of most anticancer AMPs is helical, there are few that adopt a β-structure. However, unlike helical peptides, these β rich AMPs are folded prior to engaging the membrane.

In 2010 Schneider *et al.* reported, for the first time, the design of a peptide (from now on called SVS-1) that folds in a bioactive β-structure at the surface of cancer cells, adopting an amphiphilic β hairpin structure capable of membrane disruption [141].

SVS-1 is a 18 residues long peptide. Its sequence is:

$$KVKVKVKV^DP^LPTKVKVKVK - NH_2$$

It contains N- and C- terminal strand regions of alternating valine and lysine residues connected by a tetrapeptide ($-V^DP^LPT-$) designed to mimic the β turn of the β hairpin structure. In solution the peptide adopts an ensemble of random coil conformations, where the turn region of the peptide is structured and the two strand regions aren't. In absence of any compensatory interactions, the high positive charge density from lysine side chains results in intrastrand repulsion, making the C- and N-

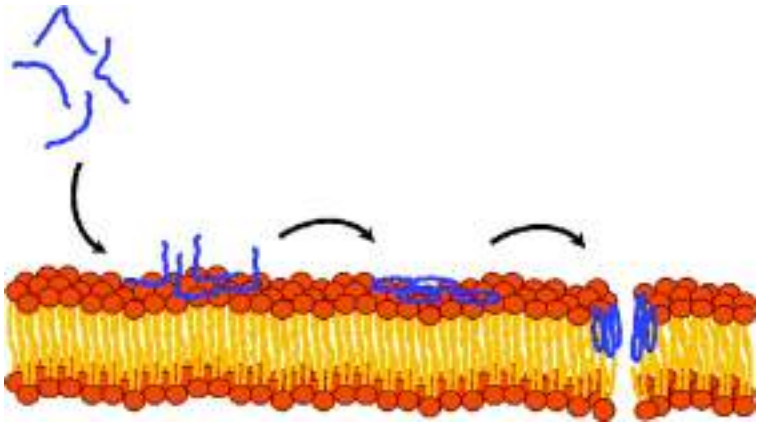


Figure 6.2: SVS-1 in solution adopts a random coil conformation. When it engages the negatively charged membrane surface, SVS-1 folds into a bioactive, β hairpin conformation. Image taken from Ref [141].

terminal strands interaction impossible. However, when SVS-1 encounters the negatively charged surface of the cell membrane, the lysine rich face interacts with it, allowing the terminal strands to collapse and form the β-hairpin. This surface-induced folding is made possible due the thermodynamic balance between the electrostatic strand repulsion and the formation of favorable membrane-peptide ion pairs. In the folded state SVS-1 should exhibit one lysine-rich face which engages the cell surface and an opposing valine-rich hydrophobic face. Schneider *et al.* proved the folding of the peptide on the membrane and showed that the peptide is able of disrupting the membrane integrity forming pores. However the mechanism of formation of these pores is still unknown. During my stay in the group of Prof. S. J. Marrink in University of Groningen I tried to elucidate this mechanism using a coarse grain representation of the system.

6.3 Computational details

The Martini CG force field [6–8] version 2.2 ElNeDyn was employed to describe the system at the coarse grain level. The ElNeDyn approach uses a global elastic network between the backbone beads to conserve the conformation in proteins or peptides. Molecular dynamics simulations were carried out with GROMACS 5.1.x package [142]. All the membranes were built with the *insane* tool [143]. The *martinize* tool was used to convert the peptide from the all atom representation to the coarse grained one. Systems were solvated with polarizable water model. Ions were added in order to reach the neutrality. Every system was submitted to 10000 steps of geometry optimization with the steepest descent method. Then the systems have been submitted to two different equilibration phases in NPT condition: the first one with a time step of 10 fs for 15 ns, the second one with a time step of 20 fs for 30 ns.

The production phase of the molecular dynamics was performed with a 20 fs time step. Temperature and pressure were kept constant to their reference values using the velocity rescale algorithm [71] and the Berendsen barostat respectively [72]. A 11 Å cutoff was applied for non-bonded interactions and the Particles Mesh Ewald algorithm [73] was employed to calculate long range electrostatic interactions.

6.4 All atom simulation of SVS-1

The SVS-1 peptide was built with the Pymol *build* utility [144]. An all atom molecular dynamics simulation in explicit solvent of 50 ns was performed on the peptide, using *distance restraints* between the carbonyl and NH group of the backbone in order to force a β-hairpin conformation. This simulation was performed in order to obtain a statistical ensemble of conformations of the SVS-1 peptide in the β-hairpin structure. At the very beginning of the dynamics (first few ns) the peptide assumes the β-hairpin conformation, as shown in Figure 6.3.

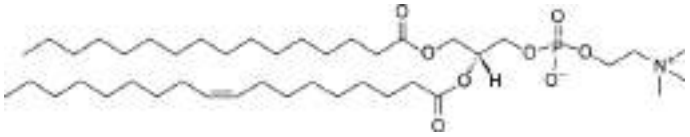
A B

Figure 6.3: **A panel**: Starting structure for the all atom dynamics of SVS-1 peptide. **B panel**: One snapshot extracted from the MD simulation. The backbone in orange cartoon representation, lysine residues in blue and valine residues in green.

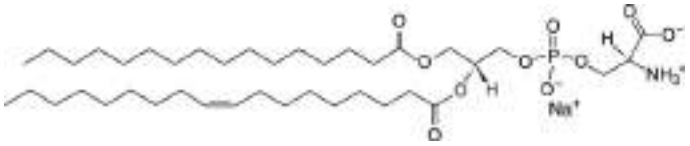
Once the conformation was extracted from the MD simulation, we coarse grained the peptide using the *martinize* tool.

6.5 Plain MD simulations

There are several choices of lipids in order to build up a negatively charged membrane in Martini. As in the work of Schneider *et al.* [141], for the experimental tests they used a POPS POPC 1:1 membrane, we used a membrane of the same composition. The structures of the lipids are shown in Figure 6.4. The simulation box is 14x14x15 nm solvated with polarizable water model. We built two different systems composed of the same membrane with different number of peptides adsorbed on it, that is:



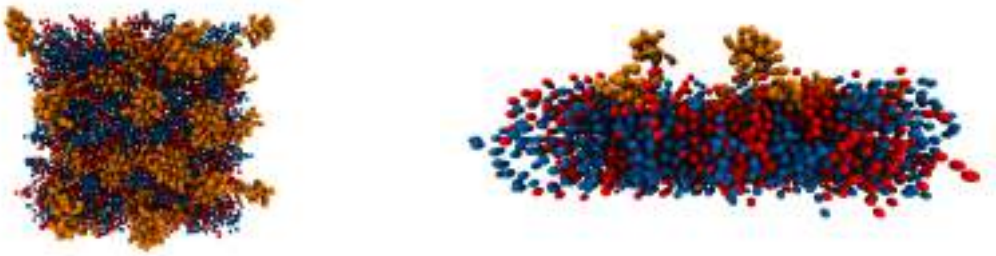
(a) POPC



(b) POPS

Figure 6.4: Structures of lipids composing the membrane.

- A low density peptide system, with 1/32 peptide/lipids (P/L) ratio.
- A high density peptide system, with 1/20 peptide/lipids ratio. (Figure 6.5)

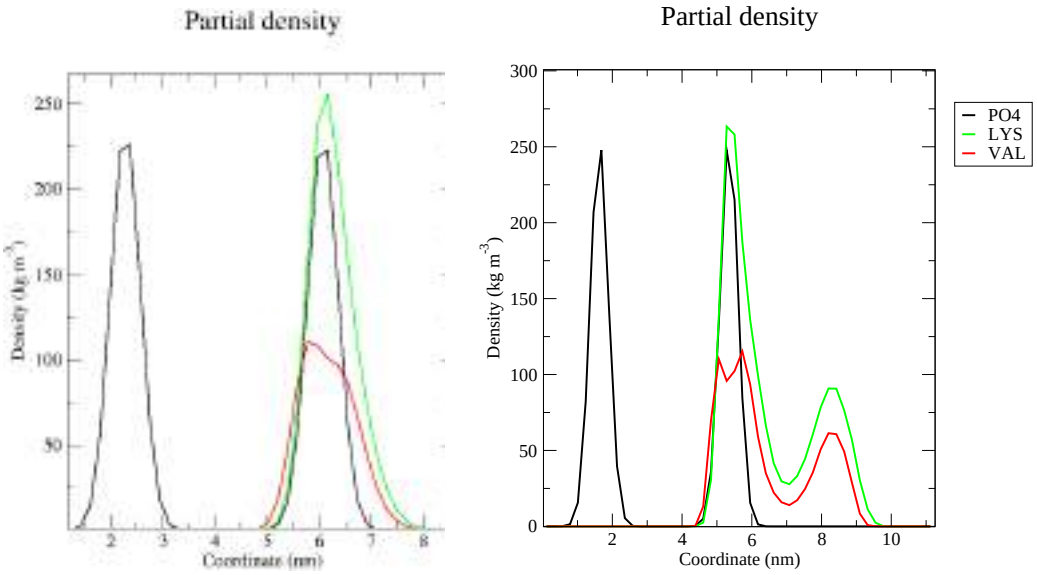


A B

Figure 6.5: In orange SVS-1 peptides, in blue POPC and in red POPS. Waters are hidden for sake of clarity. **A panel**: A view of the xy plane of the high density system.
B panel: A view of the xz plane of the high density system. As it can be seen peptides are placed only on one side of the membrane.

Peptides were placed only on one side of the membrane because we want to simulate the process of disruption of the membrane itself. What happens *in vivo* indeed is that the antimicrobial peptides approach the membrane only from the outer leaflet. We performed 7 μs MD simulations for both systems, observing no pore formation. In both systems the peptides stick to the membrane and diffuse on it. The analysis of the depth of insertion of the peptides with a density plot (Figure 6.6) shows that valine has a slightly higher depth of insertion than lysine, as expected. Lysine residues, in fact, interact with the charged phosphate groups on the surface of the membrane, while valine residues tend to interact more with the GL beads located deeper in the membrane. We can also notice some differences between the density plots of the two systems. Namely, in the 1/20 P/L system, we observe two density peaks both for valine and lysine, which are missing in the 1/32

P/L one. This means that there is no more space for the peptides to adsorb on the membrane, so they dispose on two different layers.



1/32 P/L 1/20 P/L

Figure 6.6: Density plot for phosphate groups (black), valine residues (red) and lysine residues (green). In the 1/20 P/L system the peptides dispose themselves on two different layers.

The number of contacts shows a clear preference for the peptides to interact with POPS with respect to POPC (Figure 6.7). This is not surprising, as each peptide bears a +8 charge and POPS has an anionic headgroup. If we look deeper in the number of contacts and compute valine and lysine contacts separately we can see that lysine interacts definitely more with the membrane than valine (Figure 6.8). This, again, is due to the positive charge of lysine that interacts with the phosphate group. This analysis shows the same results for both of the systems. For sake of brevity only the results on the 1/32 P/L system are shown. The pore formation could occur in a larger timescale that is not simulable in our case.

6.6 Induced pore formation

Considering that we don't observe any pore formation during the MD simulation and this could be due to kinetics, we decided to induce the pore formation in the membrane using an inverted potential. This type of potential is implemented since GROMACS 5.1 version and is shown in Figure 6.9. It is a type of flat-bottomed potential, used to restrain particles to a particular part of the simulation volume. No force acts on the restrained particles within the flat-bottomed region of the

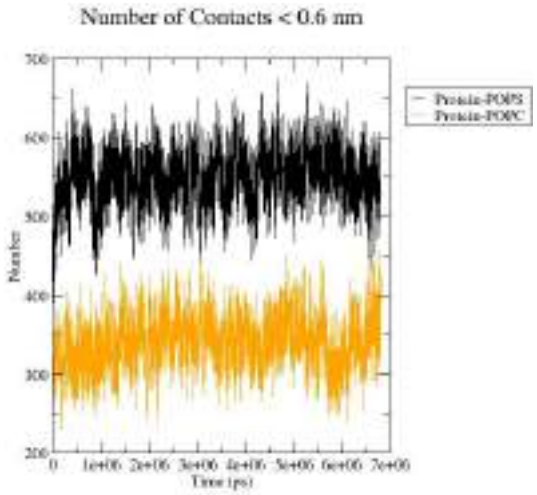
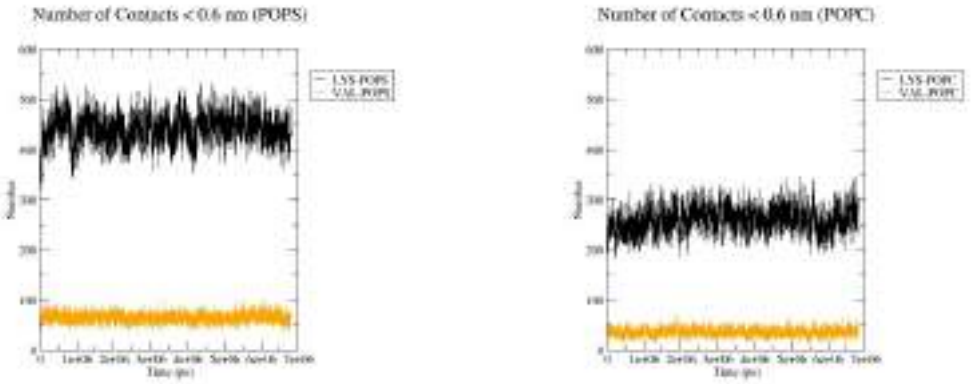


Figure 6.7: Number of conctats between peptides and the two different lipid types of the membrane.



POPS POPC

Figure 6.8: Number of conctacts of valine and lysine residues with POPS (left panel) and POPC (right panel).

potential. When the restrained particles move out of the flat-bottomed region, a harmonic force acts on them to move particles to the flat-bottomed region (eq. 6.1).

$$V_{fb}(\vec{r}_i) = \frac{1}{2} k_{fb}[d_g(\vec{r}_i; \vec{R}_i) - r_{fb}]^2 H[d_g(\vec{r}_i; \vec{R}_i) - r_{fb}] \tag{6.1}$$

$\vec{R}_i$ is the reference position, r_{fb} is the distance from the center with a flat potential, k_{fb} is the force constant and H is the Heaviside step function. The distance $d_g(\vec{r}_i; \vec{R}_i)$ from the reference position depends on the geometry of the flat-bottomed potential. In GROMACS it is possible to invert the restrained region with the unrestrained region, leading to a potential that acts to keep the particles outside the defined volume (Figure 6.9, panel B). This inverted potential is shown in eq. 6.2.

$$V^{inv}_{fb}(\vec{r}_i) = \frac{1}{2} k_{fb}[d_g(\vec{r}_i; \vec{R}_i) - |r_{fb}|]^2 H[d_g(\vec{r}_i; \vec{R}_i) - |r_{fb}|] \tag{6.2}$$

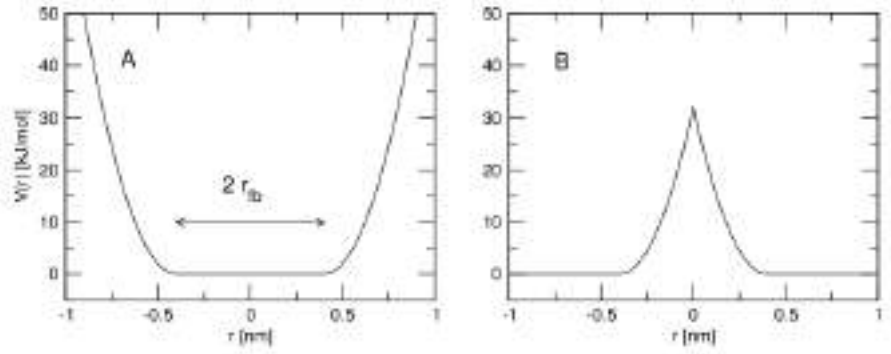
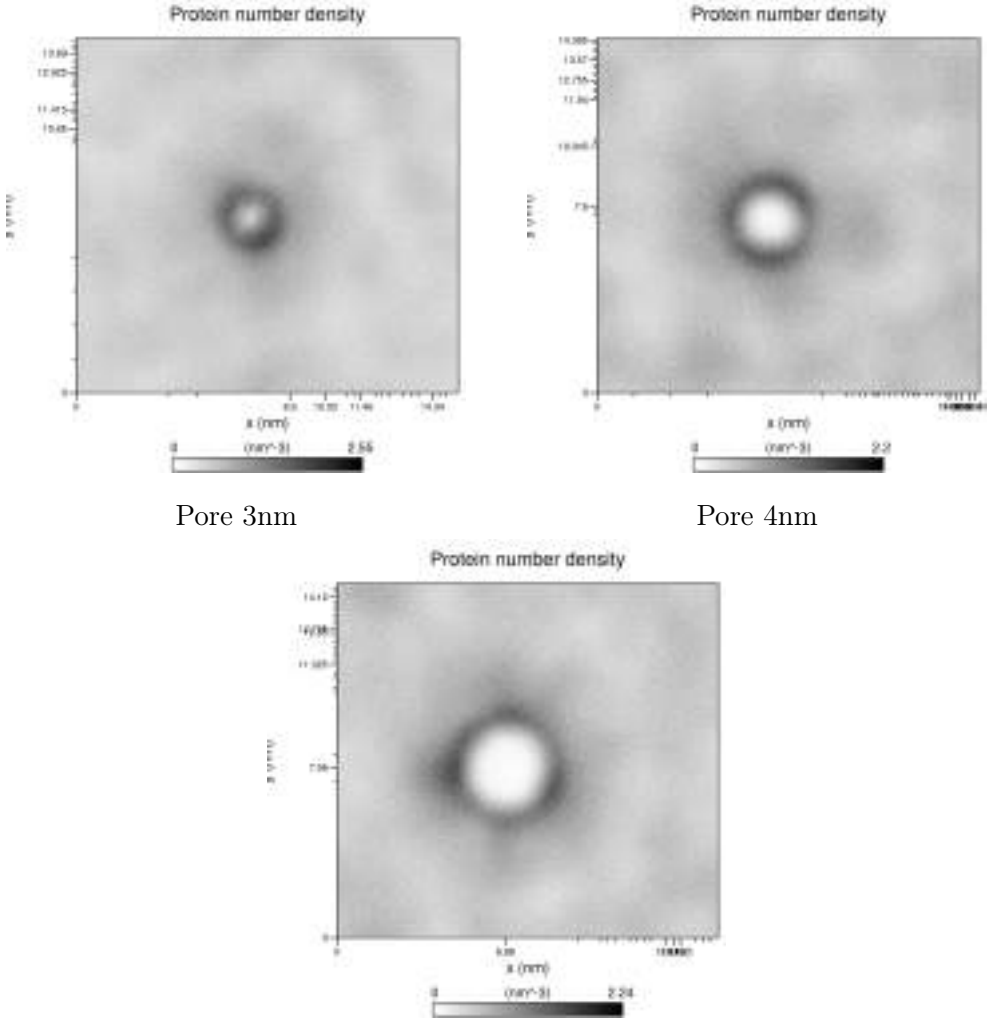


Figure 6.9: Flat-bottomed and inverted potential implemented in GROMACS.

We selected a point in the center of the membrane and we applied a 1000 $kJmol^{-1}nm^{-1}$ repulsive potential acting on the GL beads of POPS and POPC. We ran MD simulations on both the 1/20 and 1/32 P/L systems forming, for every system, three pores of different dimensions: 3nm, 4nm and 5 nm. The MD simulations were 7 μs long for the 1/32 P/L system and 5.5 μs long for the 1/20 P/L system. In this case the peptides are allowed to pass on the other side of the membrane exceeding the pore itself. From the density analysis of peptides on xy plane performed with the *g_densmap* utility we can notice that peptides density is higher on the edge of the pores (Figure 6.10 and Figure 6.11). This happens for both of the simulated systems and for all the pore dimensions. The peptides diffuse on both sides of the membrane passing through the pore.

After a few μs of sampling we selected some snapshots taken from different systems and pore dimensions in which one or more peptides were positioned through the pore and we removed the inverted potential creating the pore. In total we simulated the potential removal in 10 systems. The typical starting structure is

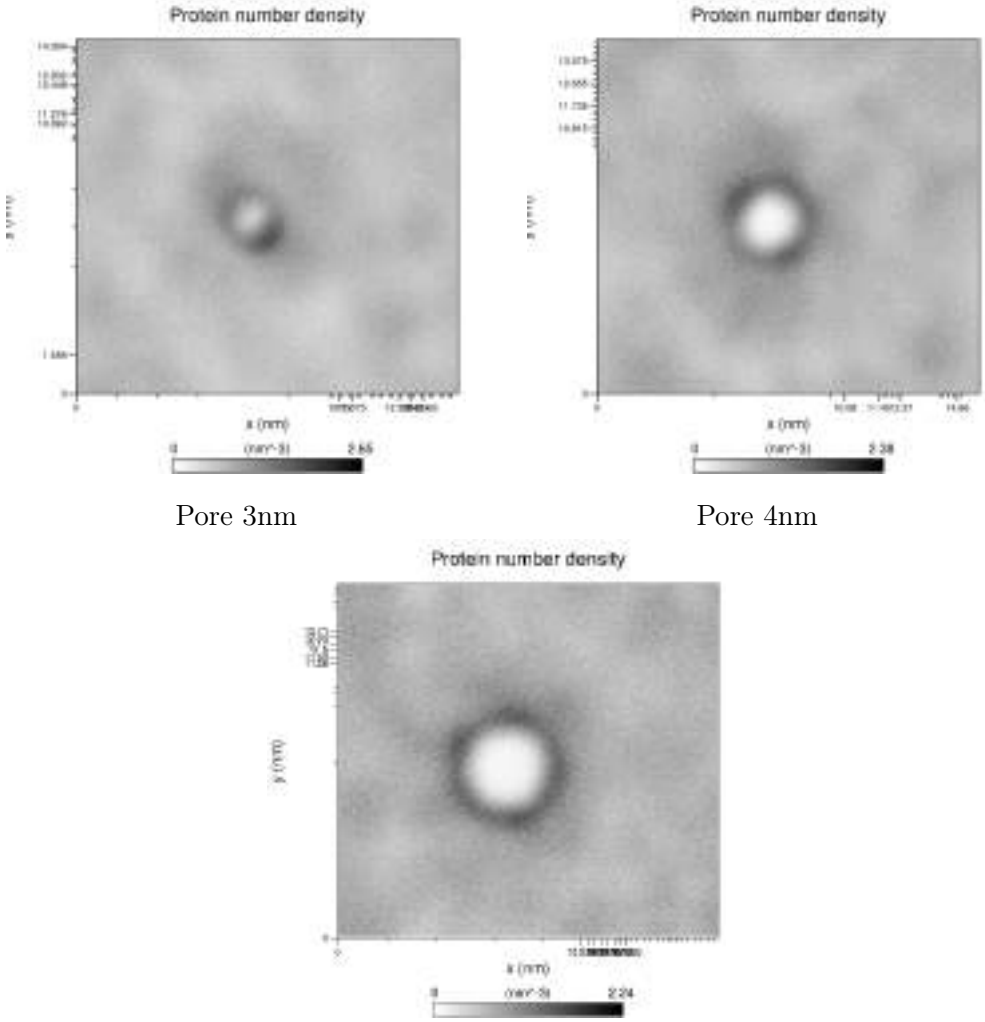


Pore 3nm

Pore 4nm

Pore 5nm

Figure 6.10: Density plot of the peptides for the 1/32 P/L system in the xy plane. Peptides density is higher on the edge of the pore.



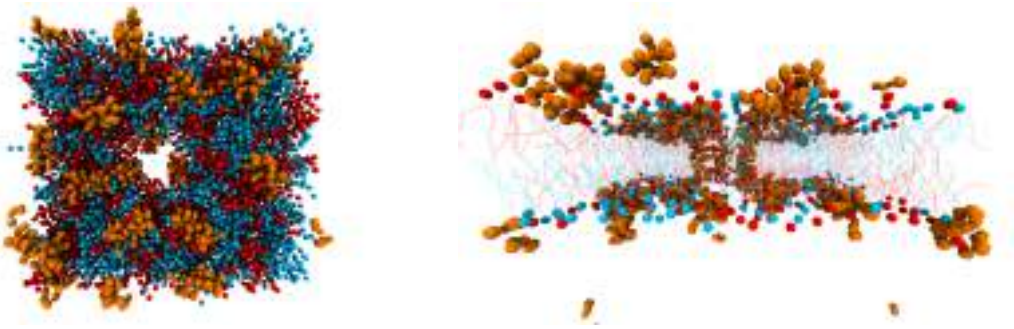
Pore 3nm

Pore 4nm

Pore 5nm

Figure 6.11: Density plot of the peptides for the 1/20 P/L system in the xy plane. Peptides density is higher on the edge of the pore.

shown in Figure 6.12.



xy plane xz plane

Figure 6.12: The starting structure for the potential removal of one of the simulated systems.

In the majority of the cases (7 out of 10) the pore closes in few tents of nanoseconds (20-40 ns). In 3 cases the pore closes but one peptide remains stuck through the membrane, although it is spitted out after 1 μs. The induced pore formation seems not to help the peptide inserting through the membrane, moreover the peptides that manage to stay into the membrane after the inverted potential removal don't succeed to remain inside the membrane itself.

6.7 Slow potential removal

Based on the results arising from the former dynamics we decided to try to remove the inverted potential slowly, decreasing step by step its value from 1000 to 500, 200, 100, 0 $kJmol^{-1}nm^{-1}$ and performing short 100 ns long MD simulations for each value of the potential. The pore remains in its position till the potential is completed shut down. With no potential acting on lipids the pore slowly closes and we obtain something very similar to what we observed in the Section 6.6.

6.8 Conclusions

It is clear from the MD results that these peptide are not able to form pores. One of the reasons could be the intrinsic stability of the lipid bilayer in the MARTINI force field. MARTINI, indeed, was originally created as a CG force field for lipids, and the interactions among the apolar beads governing the formations of the bilayer were tuned to form the bilayer itself.
We also tried to simulate a less thick membrane (DLPC/DLPS) but no pore formation was observed. For sure, in order to simulate system bearing big charges like these, polarizable water model and PME treatment of electrostatic interactions are

strongly recommended. In our opinion a better parametrization of the electrostatic interaction of the force field (Q beads) could improve the description of the mechanism of pore formation in MARTINI. Marrink *et al.* are already working on the development of MARTINI 3.0 where reformulation of the Q-particles will be carried out.

Acknowledgements

Sembra incredibile, ma sono giunto al termine di questo lungo percorso che prende il nome di Dottorato di Ricerca. Tre anni ricchi di soddisfazioni, divertimento e tanto lavoro. Proprio per questo motivo è assolutamente necessario che io ringrazi le persone che mi sono state vicino durante questi tre anni e senza le quali difficilmente avrei completato questo percorso. *In primis* grazie al Prof. **Maurizio Sironi**, gli devo molto. Grazie per avermi dato la possiblità di svolgere il PhD nel suo gruppo, per la sua disponibilità e professionalità. Grazie a **Stefano Pieraccini**, che mi ha seguito da vicino in tutti questi 3 anni, consigliandomi e stimolandomi a fare sempre meglio ("*Dobbiamo essere più produttivi*" cit.). Grazie al mio gruppo di ricerca: **Davide**, **Rosario** ed al mio compagno di ufficio **Froc**, personificazione di buon umore e calma zen. Grazie anche a tutti i tirocinanti e tesisti passati da noi nel corso degli anni: **Henreich**, **Giuseppe**, **Anna**, **Giacomo**. Grazie agli irriducibili compagni del pranzo e delle pause: **Francesca**, **Nicholas**, **Ambra**. Grazie a **Gabas** per i quotidiani caffè del buongiornissimo. Grazie ai miei storici compagni di università che ho ritrovato anche durante il PhD: queLL'orcagna deL **FRONTZ** e **Raffa**. Many thanks to Prof. **Siewert Jan Marrink** and to my Dutch research group. I will always remember my staying in the Netherlands with a smile on my face and this is because of you. Grazie alla **Vale** per aver condiviso con me numerosi viaggi in metro, la sua passione d'altronde. Grazie a quel poverino di **Leeno**, che l'Europa risieda in te, sempre. Grazie ai **G4BBIANI**, per le numerose serate passate insieme on the line. Grazie agli amici di Lodi, in particolare a **Gnu**, e al **Tower** fonte inesauribile di povertà cinematografica. Grazie al maestro **Lale** e ad **Arianna** per le numerose sessioni di D&D. Grazie a **Cuccu**, ti voglio bene. Grazie al **Devis**, anche se i chilometri sono tanti noi ce ne sbattiamo. Grazie a **Gimbo** e **Laura** grandi amici e ormai anche grandi genitori. Grazie soprattutto a te, **Marta**, mi hai sostenuto, supportato e sopportato sempre. Ti amo. Infine grazie a tutta la mia famiglia. So di essere una persona estremamente fortunata ad avere al mio fianco persone come la mia **mamma** e il mio **papà**.

E mentre succede tutto questo, chissà **IlC**...

Bibliography

[1] Andrew R Leach. *Molecular modelling: principles and applications*. Pearson education, 2001.

[2] George D Birkhoff. Proof of the ergodic theorem. *Proceedings of the National Academy of Sciences*, 17(12):656–660, 1931.

[3] DA Case, TA Darden, TE Cheatham III, CL Simmerling, J Wang, RE Duke, R Luo, RC Walker, W Zhang, KM Merz, et al. Amber 11, 2010. *University of California, San Francisco*, 5, 2010.

[4] Berk Hess, Henk Bekker, Herman JC Berendsen, Johannes GEM Fraaije, et al. Lincs: a linear constraint solver for molecular simulations. *Journal of computational chemistry*, 18(12):1463–1472, 1997.

[5] Sebastian Kmiecik, Dominik Gront, Michal Kolinski, Lukasz Wieteska, Aleksandra Elzbieta Dawid, and Andrzej Kolinski. Coarse-grained protein models and their applications. *Chem. Rev*, 116(14):7898–7936, 2016.

[6] Siewert J Marrink, H Jelger Risselada, Serge Yefimov, D Peter Tieleman, and Alex H De Vries. The martini force field: coarse grained model for biomolecular simulations. *The journal of physical chemistry B*, 111(27):7812–7824, 2007.

[7] Siewert J Marrink, Alex H De Vries, and Alan E Mark. Coarse grained model for semiquantitative lipid simulations. *The Journal of Physical Chemistry B*, 108(2):750–760, 2004.

[8] Luca Monticelli, Senthil K Kandasamy, Xavier Periole, Ronald G Larson, D Peter Tieleman, and Siewert-Jan Marrink. The martini coarse-grained force field: extension to proteins. *Journal of chemical theory and computation*, 4(5):819–834, 2008.

[9] Xavier Periole and Siewert-Jan Marrink. The martini coarse-grained force field. *Biomolecular Simulations: Methods and Protocols*, pages 533–565, 2013.

[10] Hans-Joachim Böhm. The development of a simple empirical scoring function to estimate the binding constant for a protein-ligand complex of known three-dimensional structure. *Journal of computer-aided molecular design*, 8(3):243–256, 1994.

[11] Ajay N Jain. Scoring noncovalent protein-ligand interactions: a continuous differentiable function tuned to compute binding affinities. *Journal of computer-aided molecular design*, 10(5):427–440, 1996.

[12] Peter Kollman. Free energy calculations: applications to chemical and biochemical phenomena. *Chemical reviews*, 93:2395–2395, 1993.

[13] Hiroaki Gouda, Irwin D Kuntz, David A Case, and Peter A Kollman. Free energy calculations for theophylline binding to an rna aptamer: Comparison of mm-pbsa and thermodynamic integration methods. *Biopolymers*, 68(1):16–34, 2003.

[14] Irina Massova and Peter A Kollman. Combined molecular mechanical and continuum solvent approach (mm-pbsa/gbsa) to predict ligand binding. *Perspectives in drug discovery and design*, 18(1):113–135, 2000.

[15] Shuanghong Huo, Irina Massova, and Peter A Kollman. Computational alanine scanning of the 1: 1 human growth hormone–receptor complex. *Journal of computational chemistry*, 23(1):15–27, 2002.

[16] Nathan A Baker, David Sept, Simpson Joseph, Michael J Holst, and J Andrew McCammon. Electrostatics of nanosystems: application to microtubules and the ribosome. *Proceedings of the National Academy of Sciences*, 98(18):10037–10041, 2001.

[17] Jason A Wagoner and Nathan A Baker. Assessing implicit models for nonpolar mean solvation forces: the importance of dispersion and volume terms. *Proceedings of the National Academy of Sciences*, 103(22):8331–8336, 2006.

[18] Emilio Gallicchio, Linda Yu Zhang, and Ronald M Levy. The sgb/np hydration free energy model based on the surface generalized born solvent reaction field and novel nonpolar hydration free energy estimators. *Journal of computational chemistry*, 23(5):517–529, 2002.

[19] Berk Hess, Carsten Kutzner, David Van Der Spoel, and Erik Lindahl. Gromacs 4: Algorithms for highly efficient, load-balanced, and scalable molecular simulation. *Journal of chemical theory and computation*, 4(3):435–447, 2008.

[20] Woojin Jung, Hyun-Hwan Jeong, and KiYoung Lee. Protein interactome and its application to protein function prediction. In *Protein-Protein Interactions-Computational and Experimental Tools*. InTech, 2012.

[21] Laura Bonetta. Protein-protein interactions: Interactome under construction. *Nature*, 468(7325):851–854, 2010.

[22] Ratna Rajesh Thangudu, Stephen H Bryant, Anna R Panchenko, and Thomas Madej. Modulating protein–protein interactions with small molecules: the importance of binding hotspots. *Journal of molecular biology*, 415(2):443–453, 2012.

[23] Loredana Lo Conte, Cyrus Chothia, and Joël Janin. The atomic structure of protein-protein recognition sites. *Journal of molecular biology*, 285(5):2177–2198, 1999.

[24] Michael D Wendt. Protein-protein interactions as drug targets. In *Protein-Protein Interactions*, pages 1–55. Springer, 2012.

[25] Andrew A Bogan and Kurt S Thorn. Anatomy of hot spots in protein interfaces. *Journal of molecular biology*, 280(1):1–9, 1998.

[26] Yanay Ofran and Burkhard Rost. Protein–protein interaction hotspots carved into sequences. *PLoS computational biology*, 3(7):e119, 2007.

[27] Shane Atwell, Mark Ultsch, Abraham M De Vos, and James A Wells. Structural plasticity in a remodeled protein-protein interface. *Science*, 278(5340):1125–1128, 1997.

[28] Giovanna Zinzalla and David E Thurston. Targeting protein–protein interactions for therapeutic intervention: a challenge for the future. 2009.

[29] Tom L Blundell, David F Burke, Dimitri Chirgadze, Venugopal Dhanaraj, Marko Hyvönen, C Axel Innis, Emilio Parisini, Luca Pellegrini, Muhammed Sayed, and B Lynn Sibanda. Protein-protein interactions in receptor activation and intracellular signalling. *Biological chemistry*, 381(9-10):955–959, 2000.

[30] Patrick Chène. Drugs targeting protein–protein interactions. *ChemMedChem*, 1(4):400–411, 2006.

[31] Michelle R Arkin and Adrian Whitty. The road less traveled: modulating signal transduction enzymes by inhibiting their protein–protein interactions. *Current opinion in chemical biology*, 13(3):284–290, 2009.

[32] Asher Mullard. Protein–protein interaction inhibitors get into the groove: drug developers are getting closer to tapping an unmined gold reserve of protein–protein interaction targets. *Nature reviews Drug discovery*, 11(3):173–176, 2012.

[33] Philipp Thiel, Markus Kaiser, and Christian Ottmann. Small-molecule stabilization of protein–protein interactions: An underestimated concept in drug discovery? *Angewandte Chemie International Edition*, 51(9), 2012.

[34] Peter J Hudson and Christelle Souriau. Engineered antibodies. *Nature medicine*, 9(1):129–134, 2003.

[35] Brian C Cunningham and James A Wells. High-resolution epitope mapping of hgh-receptor interactions by alanine-scanning mutagenesis. *Science*, 244(4908):1081–1085, 1989.

[36] Irina Massova and Peter A Kollman. Computational alanine scanning to probe protein- protein interactions: a novel approach to evaluate binding free energies. *Journal of the American Chemical Society*, 121(36):8133–8143, 1999.

[37] Jaume Pons, Arvind Rajpal, and Jack F Kirsch. Energetic analysis of an antigen/antibody interface: alanine scanning mutagenesis and double mutant cycles on the hyhel-10/lysozyme interaction. *Protein Science*, 8(5):958–968, 1999.

[38] Stefano Pieraccini, Riccardo De Gonda, and Maurizio Sironi. Molecular modeling of the inhibition of protein–protein interactions with small molecules: The il2–il2rα case. *Chemical Physics Letters*, 517(4):217–222, 2011.

[39] Irina S Moreira, Pedro A Fernandes, and Maria J Ramos. Computational alanine scanning mutagenesis—an improved methodological approach. *Journal of computational chemistry*, 28(3):644–654, 2007.

[40] Tim Clackson, Mark H Ultsch, James A Wells, and Abraham M de Vos. Structural and functional analysis of the 1: 1 growth hormone: receptor complex reveals the molecular basis for receptor affinity. *Journal of molecular biology*, 277(5):1111–1128, 1998.

[41] Andrea Amadei, Antonius Linssen, and Herman JC Berendsen. Essential dynamics of proteins. *Proteins: Structure, Function, and Bioinformatics*, 17(4):412–425, 1993.

[42] Svante Wold, Kim Esbensen, and Paul Geladi. Principal component analysis. *Chemometrics and intelligent laboratory systems*, 2(1-3):37–52, 1987.

[43] A Amadei, ABM Linssen, BL De Groot, DMF Van Aalten, and HJC Berendsen. An efficient method for sampling the essential subspace of proteins. *Journal of Biomolecular Structure and Dynamics*, 13(4):615–625, 1996.

[44] Berk Hess. Similarities between principal components of protein dynamics and random diffusion. *Physical Review E*, 62(6):8438, 2000.

[45] P Therese Lang, Scott R Brozell, Sudipto Mukherjee, Eric F Pettersen, Elaine C Meng, Veena Thomas, Robert C Rizzo, David A Case, Thomas L James, and Irwin D Kuntz. Dock 6: Combining techniques to model rna–small molecule complexes. *Rna*, 15(6):1219–1230, 2009.

[46] Irwin D Kuntz, Jeffrey M Blaney, Stuart J Oatley, Robert Langridge, and Thomas E Ferrin. A geometric approach to macromolecule-ligand interactions. *Journal of molecular biology*, 161(2):269–288, 1982.

[47] Demetri T Moustakas, P Therese Lang, Scott Pegg, Eric Pettersen, Irwin D Kuntz, Natasja Brooijmans, and Robert C Rizzo. Development and validation of a modular, extensible docking program: Dock 5. *Journal of computer-aided molecular design*, 20(10-11):601–619, 2006.

[48] Brian M Baynes and Bernhardt L Trout. Proteins in mixed solvents: a molecular-level perspective. *The Journal of Physical Chemistry B*, 107(50):14058–14067, 2003.

[49] Gopi S Mohan, Ling Ye, Wenfang Li, Ana Monteiro, Xiaoqian Lin, Bishu Sapkota, Brian P Pollack, Richard W Compans, and Chinglai Yang. Less is more: Ebola virus surface glycoprotein expression levels regulate virus production and infectivity. *Journal of virology*, 89(2):1205–1217, 2015.

[50] Shinji Watanabe, Takeshi Noda, Peter Halfmann, Luke Jasenosky, and Yoshihiro Kawaoka. Ebola virus (ebov) vp24 inhibits transcription and replication of the ebov genome. *Journal of Infectious Diseases*, 196(Supplement 2):S284–S290, 2007.

[51] T Hoenen, N Biedenkopf, F Zielecki, S Jung, A Groseth, H Feldmann, and S Becker. Oligomerization of ebola virus vp40 is essential for particle morphogenesis and regulation of viral transcription. *Journal of virology*, 84(14):7053–7063, 2010.

[52] L Falasca, C Agrati, N Petrosillo, A Di Caro, MR Capobianchi, G Ippolito, and M Piacentini. Molecular mechanisms of ebola virus pathogenesis: focus on cell death. *Cell Death & Differentiation*, 22(8):1250–1259, 2015.

[53] Heinz Feldmann and Thomas W Geisbert. Ebola haemorrhagic fever. *The Lancet*, 377(9768):849–862, 2011.

[54] Mike Bray and Thomas W Geisbert. Ebola virus: the role of macrophages and dendritic cells in the pathogenesis of ebola hemorrhagic fever. *The international journal of biochemistry & cell biology*, 37(8):1560–1566, 2005.

[55] Catharine M Bosio, M Javad Aman, Case Grogan, Robert Hogan, Gordon Ruthel, Diane Negley, Mansour Mohamadzadeh, Sina Bavari, and Alan Schmaljohn. Ebola and marburg viruses replicate in monocyte-derived dendritic cells without inducing the production of cytokines and full maturation. *The Journal of infectious diseases*, 188(11):1630–1638, 2003.

[56] Lawrence W Leung, Amy L Hartman, Osvaldo Martinez, Megan L Shaw, Caroline Carbonnelle, Viktor E Volchkov, Stuart T Nichol, Christopher F Basler, et al. Ebola virus vp24 binds karyopherin α1 and blocks stat1 nuclear accumulation. *Journal of virology*, 80(11):5156–5167, 2006.

[57] Charalampos Valmas, Osvaldo Martinez, Freddy Mauricio Sanchez, Christopher F Basler, et al. Ebola virus vp24 proteins inhibit the interaction of npi-1 subfamily karyopherin α proteins with activated stat1. *Journal of virology*, 81(24):13469–13477, 2007.

[58] Zongdi Feng, Melissa Cerveny, Zhipeng Yan, and Bin He. The vp35 protein of ebola virus inhibits the antiviral effect mediated by double-stranded rna-dependent protein kinase pkr. *Journal of virology*, 81(1):182–192, 2007.

[59] Michael Schümann, Thorsten Gantke, and Elke Mühlberger. Ebola virus vp35 antagonizes pkr activity through its c-terminal interferon inhibitory domain. *Journal of virology*, 83(17):8993–8997, 2009.

[60] Shridhar Bale, Jean-Philippe Julien, Zachary A Bornholdt, Christopher R Kimberlin, Peter Halfmann, Michelle A Zandonatti, John Kunert, Gerard JA Kroon, Yoshihiro Kawaoka, Ian J MacRae, et al. Marburg virus vp35 can both fully coat the backbone and cap the ends of dsrna for interferon antagonism. *PLoS pathogens*, 8(9):e1002916, 2012.

[61] Washington B Cárdenas, Yueh-Ming Loo, Michael Gale, Amy L Hartman, Christopher R Kimberlin, Luis Martínez-Sobrido, Erica Ollmann Saphire, and Christopher F Basler. Ebola virus vp35 protein binds double-stranded rna and inhibits alpha/beta interferon production induced by rig-i signaling. *Journal of virology*, 80(11):5168–5178, 2006.

[62] Rongtuan Lin, Christophe Heylbroeck, Paula M Pitha, and John Hiscott. Virus-dependent phosphorylation of the irf-3 transcription factor regulates nuclear translocation, transactivation potential, and proteasome-mediated degradation. *Molecular and cellular biology*, 18(5):2986–2996, 1998.

[63] Mitsutoshi Yoneyama, Wakako Suhara, and Takashi Fujita. Review: control of irf-3 activation by phosphorylation. *Journal of interferon & cytokine research*, 22(1):73–76, 2002.

[64] Mitsutoshi Yoneyama, Wakako Suhara, Yukiko Fukuhara, Makoto Fukuda, Eisuke Nishida, and Takashi Fujita. Direct triggering of the type i interferon system by virus infection: activation of a transcription factor complex containing irf-3 and cbp/p300. *The EMBO journal*, 17(4):1087–1095, 1998.

[65] Washington B Cárdenas, Christopher F Basler, et al. Homo-oligomerization facilitates the interferon-antagonist activity of the ebolavirus vp35 protein. *Virology*, 341(2):179–189, 2005.

[66] Craig S Brown, Michael S Lee, Daisy W Leung, Tianjiao Wang, Wei Xu, Priya Luthra, Manu Anantpadma, Reed S Shabman, Lisa M Melito, Karen S MacMillan, et al. In silico derived small molecules bind the filovirus vp35 protein and inhibit its polymerase cofactor activity. *Journal of molecular biology*, 426(10):2045–2058, 2014.

[67] Daisy W Leung, Nathaniel D Ginder, D Bruce Fulton, Jay Nix, Christopher F Basler, Richard B Honzatko, and Gaya K Amarasinghe. Structure of the ebola vp35 interferon inhibitory domain. *Proceedings of the National Academy of Sciences*, 106(2):411–416, 2009.

[68] Kresten Lindorff-Larsen, Stefano Piana, Kim Palmo, Paul Maragakis, John L Klepeis, Ron O Dror, and David E Shaw. Improved side-chain torsion potentials for the amber ff99sb protein force field. *Proteins: Structure, Function, and Bioinformatics*, 78(8):1950–1958, 2010.

[69] Junmei Wang, Romain M Wolf, James W Caldwell, Peter A Kollman, and David A Case. Development and testing of a general amber force field. *Journal of computational chemistry*, 25(9):1157–1174, 2004.

[70] William L Jorgensen, Jayaraman Chandrasekhar, Jeffry D Madura, Roger W Impey, and Michael L Klein. Comparison of simple potential functions for simulating liquid water. *The Journal of chemical physics*, 79(2):926–935, 1983.

[71] Giovanni Bussi, Davide Donadio, and Michele Parrinello. Canonical sampling through velocity rescaling. *The Journal of chemical physics*, 126(1):014101, 2007.

[72] Herman JC Berendsen, JPM van Postma, Wilfred F van Gunsteren, ARHJ DiNola, and JR Haak. Molecular dynamics with coupling to an external bath. *The Journal of chemical physics*, 81(8):3684–3690, 1984.

[73] Tom Darden, Darrin York, and Lee Pedersen. Particle mesh ewald: An n log (n) method for ewald sums in large systems. *The Journal of chemical physics*, 98(12):10089–10092, 1993.

[74] Simon J Hubbard and Janet M Thornton. Naccess. *Computer Program, Department of Biochemistry and Molecular Biology, University College London*, 2(1), 1993.

[75] Arpita Mitra and David Sept. Taxol allosterically alters the dynamics of the tubulin dimer and increases the flexibility of microtubules. *Biophysical journal*, 95(7):3252–3258, 2008.

[76] William Humphrey, Andrew Dalke, and Klaus Schulten. Vmd: visual molecular dynamics. *Journal of molecular graphics*, 14(1):33–38, 1996.

[77] Tingjun Hou, Junmei Wang, Youyong Li, and Wei Wang. Assessing the performance of the mm/pbsa and mm/gbsa methods: I. the accuracy of binding free energy calculations based on molecular dynamics simulations. *Journal of chemical information and modeling*, 51(1):69, 2011.

[78] Irina S Moreira, Pedro A Fernandes, and Maria J Ramos. Unravelling hot spots: a comprehensive computational mutagenesis study. *Theoretical Chemistry Accounts: Theory, Computation, and Modeling (Theoretica Chimica Acta)*, 117(1):99–113, 2007.

[79] S Goodbourn, L Didcock, and RE Randall. Interferons: cell signalling, immune modulation, antiviral response and virus countermeasures. *Journal of General Virology*, 81(10):2341–2364, 2000.

[80] Xiaomin Chen, Uwe Vinkemeier, Yanxiang Zhao, David Jeruzalmi, James E Darnell, and John Kuriyan. Crystal structure of a tyrosine phosphorylated stat-1 dimer bound to dna. *Cell*, 93(5):827–839, 1998.

[81] Kevin M McBride, Gregg Banninger, Christine McDonald, and Nancy C Reich. Regulated nuclear import of the stat1 transcription factor by direct binding of importin-α. *The EMBO journal*, 21(7):1754–1763, 2002.

[82] Wei Xu, Megan R Edwards, Dominika M Borek, Alicia R Feagins, Anuradha Mittal, Joshua B Alinger, Kayla N Berry, Benjamin Yen, Jennifer Hamilton, Tom J Brett, et al. Ebola virus vp24 targets a unique nls binding site on karyopherin alpha 5 to selectively compete with nuclear import of phosphorylated stat1. *Cell host & microbe*, 16(2):187–200, 2014.

[83] Michelle R Arkin and James A Wells. Small-molecule inhibitors of protein–protein interactions: progressing towards the dream. *Nature reviews Drug discovery*, 3(4):301–317, 2004.

[84] Duncan E Scott, Andrew R Bayly, Chris Abell, and John Skidmore. Small molecules, big targets: drug discovery faces the protein-protein interaction challenge. *Nature Reviews Drug Discovery*, 15(8):533–550, 2016.

[85] Monika Raj, Brooke N Bullock, and Paramjit S Arora. Plucking the high hanging fruit: A systematic approach for targeting protein–protein interactions. *Bioorganic & medicinal chemistry*, 21(14):4051–4057, 2013.

[86] Rudi Fasan, Ricardo LA Dias, Kerstin Moehle, Oliver Zerbe, Jan W Vrijbloed, Daniel Obrecht, and John A Robinson. Using a β-hairpin to mimic an α-helix: cyclic peptidomimetic inhibitors of the p53–hdm2 protein–protein interaction. *Angewandte Chemie*, 116(16):2161–2164, 2004.

[87] Sara Pellegrino, Nicola Ferri, Noemi Colombo, Edoardo Cremona, Alberto Corsini, Roberto Fanelli, Maria Luisa Gelmi, and Chiara Cabrele. Synthetic peptides containing a conserved sequence motif of the id protein family modulate vascular smooth muscle cell phenotype. *Bioorganic & medicinal chemistry letters*, 19(22):6298–6302, 2009.

[88] James A Wells and Christopher L McClendon. Reaching for high-hanging fruit in drug discovery at protein–protein interfaces. *Nature*, 450(7172):1001–1009, 2007.

[89] Michelle R Arkin, Mike Randal, Warren L DeLano, Jennifer Hyde, Tinh N Luong, Johan D Oslob, Darren R Raphael, Lisa Taylor, Jun Wang, Robert S McDowell, et al. Binding of small molecules to an adaptive protein–protein interface. *Proceedings of the National Academy of Sciences*, 100(4):1603–1608, 2003.

[90] Thorsten Berg. Modulation of protein–protein interactions with small organic molecules. *Angewandte Chemie International Edition*, 42(22):2462–2481, 2003.

[91] Stefano Pieraccini, Giorgio Saladino, Graziella Cappelletti, Daniele Cartelli, Pierangelo Francescato, Giovanna Speranza, Paolo Manitto, and Maurizio

Sironi. In silico design of tubulin-targeted antimitotic peptides. *Nature chemistry*, 1(8):642–648, 2009.

[92] Domingo González-Ruiz and Holger Gohlke. Targeting protein-protein interactions with small molecules: challenges and perspectives for omputational binding epitope detection and ligand finding. *Current medicinal chemistry*, 13(22):2607–2625, 2006.

[93] Adrianna PP Zhang, Zachary A Bornholdt, Tong Liu, Dafna M Abelson, David E Lee, Sheng Li, Virgil L Woods Jr, and Erica Ollmann Saphire. The ebola virus interferon antagonist vp24 directly binds stat1 and has a novel, pyramidal fold. *PLoS Pathog*, 8(2):e1002550, 2012.

[94] William L Jorgensen, Jayaraman Chandrasekhar, Jeffry D Madura, Roger W Impey, and Michael L Klein. Comparison of simple potential functions for simulating liquid water. *The Journal of chemical physics*, 79(2):926–935, 1983.

[95] Cristina Paissoni, Dimitrios Spiliotopoulos, Giovanna Musco, and Andrea Spitaleri. Gmxpbsa 2.0: A gromacs tool to perform mm/pbsa and computational alanine scanning. *Computer Physics Communications*, 185(11):2920–2929, 2014.

[96] Dimitrios Spiliotopoulos, Andrea Spitaleri, and Giovanna Musco. Exploring phd fingers and h3k4me0 interactions with molecular dynamics simulations and binding free energy calculations: Aire-phd1, a comparative study. *PloS one*, 7(10):e46902, 2012.

[97] Sílvia A Martins, Marta AS Perez, Irina S Moreira, Sérgio F Sousa, MJ Ramos, and PA Fernandes. Computational alanine scanning mutagenesis: Mm-pbsa vs ti. *Journal of chemical theory and computation*, 9(3):1311–1319, 2013.

[98] Richard T Bradshaw, Bhavesh H Patel, Edward W Tate, Robin J Leatherbarrow, and Ian R Gould. Comparing experimental and computational alanine scanning techniques for probing a prototypical protein–protein interaction. *Protein Engineering, Design & Selection*, 24(1-2):197–207, 2010.

[99] Donatella Potenza, Laura Belvisi, Francesca Vasile, Elisabetta Moroni, Federica Cossu, and Pierfausto Seneci. A nmr and computational study of smac mimics targeting both the bir2 and bir3 domains in xiap protein. *Organic & biomolecular chemistry*, 10(16):3278–3287, 2012.

[100] Sharon M Kelly, Thomas J Jess, and Nicholas C Price. How to study proteins by circular dichroism. *Biochimica et Biophysica Acta (BBA)-Proteins and Proteomics*, 1751(2):119–139, 2005.

[101] David P Fairlie and Aline Dantas de Araujo. Review stapling peptides using cysteine crosslinking. *Peptide Science*, 106(6):843–852, 2016.

[102] Glenn L Millhauser. Views of helical peptides: a proposal for the position of 310-helix along the thermodynamic folding pathway. *Biochemistry*, 34(12):3873–3877, 1995.

[103] Shuchismita Dutta, Kyle Burkhardt, Jasmine Young, Ganesh J Swaminathan, Takanori Matsuura, Kim Henrick, Haruki Nakamura, and Helen M Berman. Data deposition and annotation at the worldwide protein data bank. *Molecular biotechnology*, 42(1):1–13, 2009.

[104] Daria B Kokh, Stefan Richter, Stefan Henrich, Paul Czodrowski, Friedrich Rippmann, and Rebecca C Wade. Trapp: a tool for analysis of transient binding pockets in proteins, 2013.

[105] James S Fraser, Henry van den Bedem, Avi J Samelson, P Therese Lang, James M Holton, Nathaniel Echols, and Tom Alber. Accessing protein conformational ensembles using room-temperature x-ray crystallography. *Proceedings of the National Academy of Sciences*, 108(39):16247–16252, 2011.

[106] Pietro Cozzini, Glen E Kellogg, Francesca Spyrakis, Donald J Abraham, Gabriele Costantino, Andrew Emerson, Francesca Fanelli, Holger Gohlke, Leslie A Kuhn, Garrett M Morris, et al. Target flexibility: an emerging consideration in drug discovery and design. *Journal of medicinal chemistry*, 51(20):6237–6255, 2008.

[107] Susanne Eyrisch and Volkhard Helms. Transient pockets on protein surfaces involved in protein-protein interaction. *Journal of Medicinal Chemistry*, 50(15):3457–3464, 2007.

[108] Volkhard Helms. Protein dynamics tightly connected to the dynamics of surrounding and internal water molecules. *ChemPhysChem*, 8(1):23–33, 2007.

[109] Christopher D Wassman, Roberta Baronio, Özlem Demir, Brad D Wallentine, Chiung-Kuang Chen, Linda V Hall, Faezeh Salehi, Da-Wei Lin, Benjamin P Chung, G Wesley Hatfield, et al. Computational identification of a transiently open l1/s3 pocket for reactivation of mutant p53. *Nature communications*, 4:1407, 2013.

[110] Stefan Henrich, Outi MH Salo-Ahen, Bingding Huang, Friedrich F Rippmann, Gabriele Cruciani, and Rebecca C Wade. Computational approaches to identifying and characterizing protein binding sites for ligand design. *Journal of Molecular Recognition*, 23(2):209–219, 2010.

[111] R Laurie, T Alasdair, and Richard M Jackson. Methods for the prediction of protein-ligand binding sites for structure-based drug design and virtual ligand screening. *Current Protein and Peptide Science*, 7(5):395–406, 2006.

[112] G Patrick Brady and Pieter FW Stouten. Fast prediction and visualization of protein binding pockets with pass. *Journal of computer-aided molecular design*, 14(4):383–401, 2000.

[113] Oliver Kohlbacher and Hans-Peter Lenhof. Ball—rapid software prototyping in computational molecular biology. *Bioinformatics*, 16(9):815–824, 2000.

[114] Xavier Daura, Karl Gademann, Bernhard Jaun, Dieter Seebach, Wilfred F van Gunsteren, and Alan E Mark. Peptide folding: when simulation meets experiment. *Angewandte Chemie International Edition*, 38(1-2):236–240, 1999.

[115] Eric F Pettersen, Thomas D Goddard, Conrad C Huang, Gregory S Couch, Daniel M Greenblatt, Elaine C Meng, and Thomas E Ferrin. Ucsf chimera—a visualization system for exploratory research and analysis. *Journal of computational chemistry*, 25(13):1605–1612, 2004.

[116] John J Irwin and Brian K Shoichet. Zinc- a free database of commercially available compounds for virtual screening. *Journal of chemical information and modeling*, 45(1):177–182, 2005.

[117] Antonella Furini, Csaba Koncz, Francesco Salamini, and Dorothea Bartels. High level transcription of a member of a repeated gene family confers dehydration tolerance to callus tissue of craterostigma plantagineum. *The EMBO Journal*, 16(12):3599–3608, 1997.

[118] Xin Deng, Jonathan Phillips, Annemarie H Meijer, Francesco Salamini, and Dorothea Bartels. Characterization of five novel dehydration-responsive homeodomain leucine zipper genes from the resurrection plant craterostigma plantagineum. *Plant molecular biology*, 49(6):601–610, 2002.

[119] George N Somero, Charles B Osmond, and Carla L Bolis. *Water and life: comparative analysis of water relationships at the organismic, cellular, and molecular levels*. Springer Science & Business Media, 2012.

[120] Paul H Yancey. Organic osmolytes as compatible, metabolic and counteracting cytoprotectants in high osmolarity and other stresses. *Journal of Experimental Biology*, 208(15):2819–2830, 2005.

[121] David AC Beck, Brian J Bennion, Darwin OV Alonso, and Valerie Daggett. Chapter twenty-two-simulations of macromolecules in protective and denaturing osmolytes: Properties of mixed solvent systems and their effects on water

and protein structure and dynamics. *Methods in enzymology*, 428:373–396, 2007.

[122] Raymond F Greene and C Nick Pace. Urea and guanidine hydrochloride denaturation of ribonuclease, lysozyme, α-chymotrypsin, and β-lactoglobulin. *Journal of Biological Chemistry*, 249(17):5388–5393, 1974.

[123] P.H. Yancey, M.E. Clark, S.C. Hand, R.D. Bowlus, and G.N. Somero. Living with water stress: Evolution of osmolyte systems. *Science*, 217(4566):1214–1222, 1982. cited By 2295.

[124] C.Y. Hu, H. Kokubo, G.C. Lynch, D.W. Bolen, and B.M. Pettitt. Backbone additivity in the transfer model of protein solvation. *Protein Science*, 19(5):1011–1022, 2010. cited By 47.

[125] Martin C. Stumpe and Helmut Grubmüller. Interaction of urea with aminoacids: Implications for urea-induced protein denaturation. *Journal of the American Chemical Society*, 129(51):16126–16131, 2007.

[126] Lan Hua, Ruhong Zhou, D. Thirumalai, and B. J. Berne. Urea denaturation by stronger dispersion interactions with proteins than water implies a 2-stage unfolding. *Proceedings of the National Academy of Sciences*, 105(44):16928–16933, 2008.

[127] Matthew Auton, Luis Marcelo F. Holthauzen, and D. Wayne Bolen. Anatomy of energetic changes accompanying urea-induced protein denaturation. *Proceedings of the National Academy of Sciences*, 104(39):15317–15322, 2007.

[128] Haiyan Wei, Yubo Fan, and Yi Qin Gao. Effects of urea, tetramethyl urea, and trimethylamine n-oxide on aqueous solution structure and solvation of protein backbones: A molecular dynamics simulation study. *The Journal of Physical Chemistry B*, 114(1):557–568, 2010.

[129] Brian J. Bennion and Valerie Daggett. Counteraction of urea-induced protein denaturation by trimethylamine n-oxide: A chemical chaperone at atomic resolution. *Proceedings of the National Academy of Sciences of the United States of America*, 101(17):6433–6438, 2004.

[130] Qin Zou, Brian J. Bennion, Valerie Daggett, and Kenneth P. Murphy. The molecular mechanism of stabilization of proteins by tmao and its ability to counteract the effects of urea. *Journal of the American Chemical Society*, 124(7):1192–1202, 2002.

[131] Sandip Paul and G. N. Patey. Hydrophobic interactions in urea-trimethylamine-n-oxide solutions. *The Journal of Physical Chemistry B*, 112(35):11106–11111, 2008.

[132] DW Bolen and Ilia V Baskakov. The osmophobic effect: natural selection of a thermodynamic force in protein folding. *Journal of molecular biology*, 310(5):955–963, 2001.

[133] Megan R Edwards, Britney Johnson, Chad E Mire, Wei Xu, Reed S Shabman, Lauren N Speller, Daisy W Leung, Thomas W Geisbert, Gaya K Amarasinghe, and Christopher F Basler. The marburg virus vp24 protein interacts with keap1 to activate the cytoprotective antioxidant response pathway. *Cell reports*, 6(6):1017–1025, 2014.

[134] Daisy W Leung, Dominika Borek, Priya Luthra, Jennifer M Binning, Manu Anantpadma, Gai Liu, Ian B Harvey, Zhaoming Su, Ariel Endlich-Frazier, Juanli Pan, et al. An intrinsically disordered peptide from ebola virus vp35 controls viral rna synthesis by modulating nucleoprotein-rna interactions. *Cell reports*, 11(3):376–389, 2015.

[135] Robert N Kirchdoerfer, Dafna M Abelson, Sheng Li, Malcolm R Wood, and Erica Ollmann Saphire. Assembly of the ebola virus nucleoprotein from a chaperoned vp35 complex. *Cell reports*, 12(1):140–149, 2015.

[136] Maria Papagianni. Ribosomally synthesized peptides with antimicrobial properties: biosynthesis, structure, function, and applications. *Biotechnology advances*, 21(6):465–499, 2003.

[137] Linda Julian and Michael F Olson. Apoptotic membrane dynamics in health and disease. *Cell Health and Cytoskeleton*, 2015(7):133–142, 2015.

[138] AJ Verkleij, RFA Zwaal, B Roelofsen, P Comfurius, D Kastelijn, and LLM Van Deenen. The asymmetric distribution of phospholipids in the human red cell membrane. a combined study using phospholipases and freeze-etch electron microscopy. *Biochimica et Biophysica Acta (BBA)-Biomembranes*, 323(2):178–193, 1973.

[139] Alain Zachowski. Phospholipids in animal eukaryotic membranes: transverse asymmetry and movement. *Biochemical Journal*, 294(Pt 1):1, 1993.

[140] David W Hoskin and Ayyalusamy Ramamoorthy. Studies on anticancer activities of antimicrobial peptides. *Biochimica et Biophysica Acta (BBA)-Biomembranes*, 1778(2):357–375, 2008.

[141] Chomdao Sinthuvanich, Ana Salomé Veiga, Kshitij Gupta, Diana Gaspar, Robert Blumenthal, and Joel P Schneider. Anticancer β-hairpin peptides: membrane-induced folding triggers activity. *Journal of the American Chemical Society*, 134(14):6210, 2012.

[142] Mark James Abraham, Teemu Murtola, Roland Schulz, Szilárd Páll, Jeremy C Smith, Berk Hess, and Erik Lindahl. Gromacs: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX*, 1:19–25, 2015.

[143] Tsjerk A Wassenaar, Helgi I Ingolfsson, Rainer A Bockmann, D Peter Tieleman, and Siewert J Marrink. Computational lipidomics with insane: a versatile tool for generating custom membranes for molecular simulations. *Journal of chemical theory and computation*, 11(5):2144–2155, 2015.

[144] Schrödinger, LLC. The PyMOL molecular graphics system, version 1.8. November 2015.