

Postdoctoral report: Molecular docking study of the compound Genistein with the membrane-anchored Bcl-xL protein.

Luis Garreta

Advisors:

Dr. Andrés Jaramillo Botero
Dr. Henry Fabian Tobar Tosse

December 14, 2022

Pontificia Universidad Javeriana - Instituto de Investigación en Ciencias Ómicas, iÓmicas

Cali - Colombia

Abstract

The Bcl-2 family of proteins are important regulators of apoptosis and group members with both anti-apoptotic and pro-apoptotic functions. In the case of proteins with anti-apoptotic function, such as Bcl-2 and Bcl-xL, they contain four Bcl-2 homology domains (BH), a flexible loop domain (FLD), and a C-terminal hydrophobic helix that serves as a transmembrane anchor (TMD). The study of these proteins is difficult because the reported structures are incomplete or truncated, either lacking the FLD and TMD domains or containing only the C-terminal helix. However, some works create complete versions of the Bcl-xL protein from these truncated structures to study both in-silico and experimentally. Here, in this work, we used two truncated structures of Bcl-xL to construct a complete membrane-anchored version and study its docking with the compound Genistein, present in sugarcane bagasse. The docking was performed with three tools: Autodock4, Vina, and SwissDock, which generated more than 300 poses of the ligand which were filtered to obtain a single pose that presented the best interactions with the protein. The filtering considered the affinity and localization of the poses, in addition to several conformational parameters evaluated after a 200 ns molecular dynamics simulation. A 1000 ns molecular dynamics was run with the selected pose and conformational parameters as well as protein-ligand interactions and protein motions were evaluated. Finally, it was observed that the ligand was bound to the protein for about 350 ns, time during which the main protein-ligand interactions were hydrogen bonding, hydrophobic, and pi-cation interactions. In addition, the protein in the presence of the ligand exhibited changes in its conformational dynamics showing extended flexible loop domain (FLD) and rotations around the Z-axis of $\sim 220^\circ$ and around the X- and Y-axes of $\sim 50^\circ$. These changes cause the $\alpha 3$ y $\alpha 4$ helices, conforming the hydrophobic groove, to be more exposed and possibly facilitate the coupling of other Bcl-2 family proteins, which activate the anti-apoptotic function of the Bcl-xL protein.

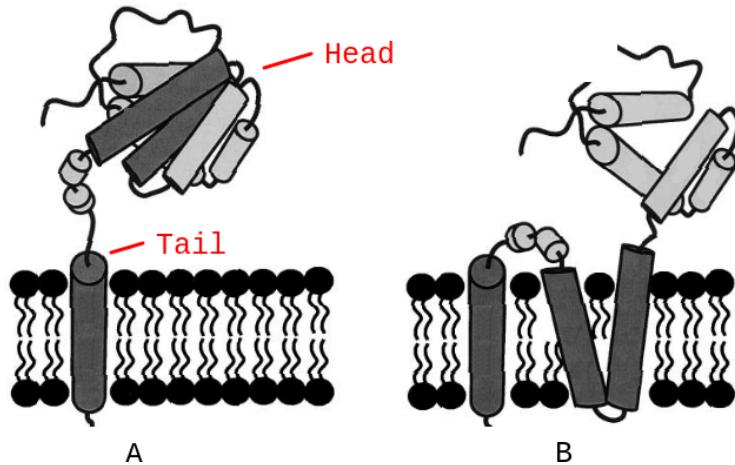
Contents

| | | |
|----------|--|-----------|
| 1 | Introduction | 2 |
| 2 | Materials and Methods | 4 |
| 2.1 | Initial structures | 4 |
| 2.2 | Molecular docking | 5 |
| 2.3 | Selection of docking poses | 5 |
| 2.4 | Molecular dynamics simulations (MD) | 5 |
| 2.5 | Trajectory analysis | 6 |
| 3 | Results and discussion | 6 |
| 3.1 | Molecular docking | 6 |
| 3.2 | Evaluation of selected poses | 9 |
| 3.2.1 | Number of hydrogen bonds | 9 |
| 3.2.2 | Root Mean Squared Deviation (RMSD) | 9 |
| 3.2.3 | Root Mean Square Fluctuation (RMSF) | 10 |
| 3.2.4 | Radius of Gyration (RG) | 10 |
| 3.2.5 | Protein-ligand interactions | 11 |
| 3.3 | Evaluation of the ligand pose in the sw06 system | 12 |
| 3.3.1 | Protein-ligand interactions | 12 |
| 3.3.2 | Root Mean Squared Deviation (RMSD) | 13 |
| 3.3.3 | Root Mean Square Fluctuation (RMSF) | 13 |
| 3.3.4 | Radius of Gyration (RG) | 14 |
| 3.4 | Conformational dynamics | 15 |
| 3.4.1 | Protein rotations | 15 |
| 3.4.2 | Accessibility of the hydrophobic binding groove | 16 |
| 4 | Final discussion | 17 |
| 5 | Availability | 17 |
| 6 | Acknowledgments | 17 |
| | References | 17 |

1 Introduction

This work was originally focused on performing a molecular dynamics (MD) simulation of a membrane-anchored Bcl-2 bound to a compound present in sugarcane bagasse. However, the literature reports difficulties in studying this protein since it is insoluble [Mushtaq et al., 2021] and its complete structure has not yet been resolved [Lan et al., 2020]. In addition, most works have focused on the study of truncated forms of this protein, lacking the main functional characteristics, and on membrane-inserted topologies [Mushtaq et al., 2021, Lan et al., 2020, Caro-Gómez et al., 2019].

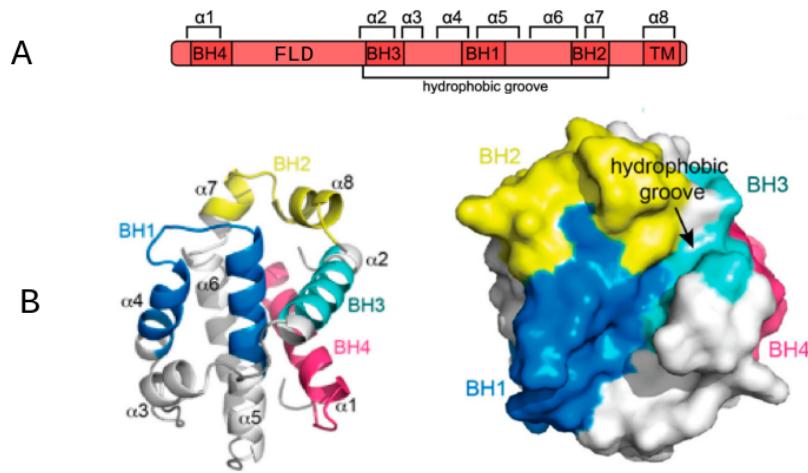
For the above reasons, we decided to work with the membrane-anchored anti-apoptotic protein Bcl-xL (Figure 1.A) for which there is structural information on both the soluble region of the protein, protein head, and the C-terminal helical region, protein tail, which inserts into the membrane [Tyagi et al., 2021, Ryzhov et al., 2020]. In addition, many inhibitors and activators that can bind to Bcl-2 can also bind to Bcl-xL, as they exhibit very similar sequences and structures [Wakui et al., 2018].



Adapted from [Kim et al., 2004]

Figure 1: Two different membrane-integrated Bcl-xL topologies. (A) A membrane-anchored conformation. (B) membrane-inserted conformation.

The structure of Bcl-xL contains 8 alpha helices, with four regions named Bcl-2 homology (BH) domains; an extensive flexible loop domain (FLD) connecting helices α_1 y α_2 ; a hydrophobic C-terminal region that serves as a transmembrane anchor (TM); and a hydrophobic binding groove that is formed primarily by helices α_3 , α_4 , y α_5 , the latter forming the base of the groove [Lee & Fairlie, 2019] (Figure 2).



Adapted from [Renault et al., 2017, Borrás et al., 2020]

Figure 2: Structure of human Bcl-xL protein. (A) Primary structure with 8 helices, 4 BH domains, an FLD loop domain and a transmembrane domain formed by helix 8 (TM). (B) Tertiary structure showing the organization of the secondary structures and a view of the surface highlighting the hydrophobic groove.

Although the structure of Bcl-xL has not been completely determined, truncated structures of the Bcl-xL protein, lacking the FLD domain and the C-terminal tail (e.g. PDB 1lxl) or with only the C-terminal tail known (e.g. PDB 6f46), have been reported. Some of these structures have been used in different works to construct a complete version of the Bcl-xL structure and study it either *in-silico* [Tyagi et al., 2021] or experimentally [Ryzhov et al., 2020].

In this work, our objective was to study *in-silico* the binding interaction between the phenolic compound Genistein, abundant in sugarcane bagasse [Ali et al., 2021], and a membrane-anchored anti-apoptotic Bcl-xL protein. For this, we took the truncated structures of two Bcl-xL proteins and constructed a complete version. Then, we performed molecular docking between the complete Bcl-xL protein and the Genistein compound, selected the best poses, used each pose to construct a protein-ligand system with the protein anchored to a lipid bilayer, and performed 200 ns molecular dynamics simulations for analyzing the interactions of these poses and select the best one. Finally, we run a 1000ns MD simulation on the best interacting pose to study the conformational changes that the protein experimented in the presence of the ligand.

2 Materials and Methods

2.1 Initial structures

Receptor:

We constructed a model of the 3D structure of the complete Bcl-xL protein by taking two truncated structures of the Bcl-xL protein [Tyagi et al., 2021, Ryzhov et al., 2020]. Residues 1 to 222 of the structure with PDB ID 1lx1, which lacks the C-terminal helix, were taken as the head of the protein; while residues 209 to 231 of the structure with PDB ID 6f46 were taken as the tail (Figure 3.A).

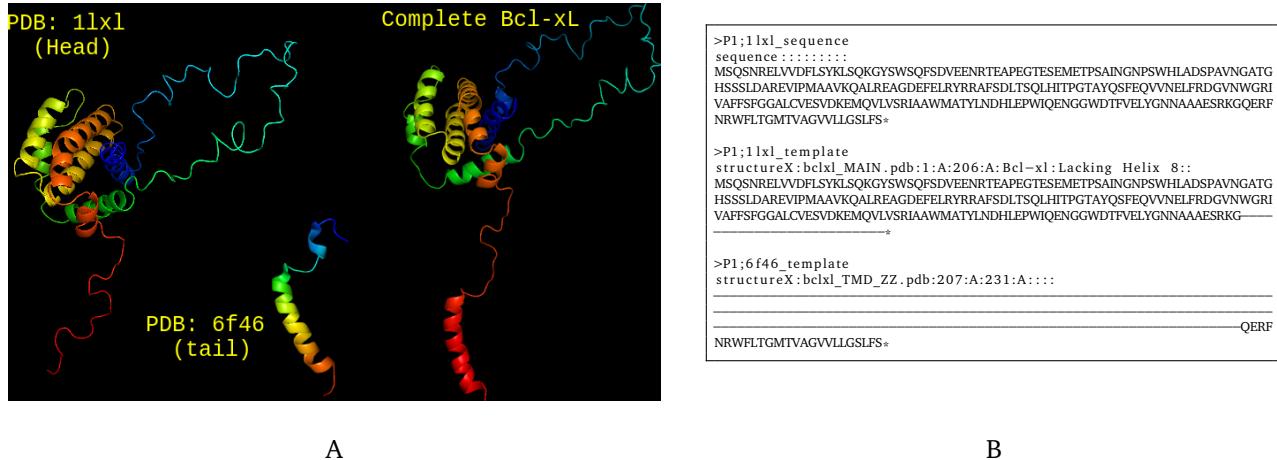


Figure 3: Construction of the complete Bcl-xL protein. (A) Structures used for construction. (B) Alignment file used by MODELLER for construction.

The model was constructed with the MODELLER software for comparative modeling [Eswar et al., 2006] using an alignment file in which the complete amino acid sequence of the Bcl-xL protein (PDB: 1lx1) was aligned with two templates: the head protein structure (amino acids 1 to 206) and the tail protein structure (amino acids 207 to 231) previously aligned along the negative Z-axis (Figure 3.B).

As a result, 10 models were obtained, ranked with the MODELLER scoring function, of which the one with the best score was taken as the complete structure of Bcl-xL (Figure 3.C). Finally, the entire structure was oriented with respect to the negative normal of the membrane (Z-axis by definition) using as reference the pre-oriented coordinates of the helical structure obtained from the membrane protein database (OPM) (<https://opm.phar.umich.edu/>).

Ligand:

For the ligand, the chemical structure of the compound Genistein (PDB ID: GEN) was taken from the PDB database (<https://www.rcsb.org/ligand/GEN>), from which the optimized coordinates were downloaded in the ideal version and SDF format (Figure 4).

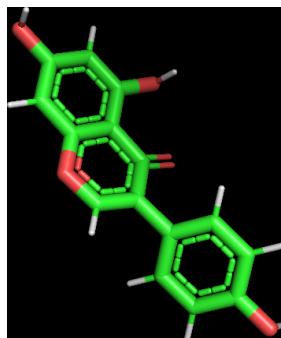


Figure 4: 3D structure for the Genistein ligand.

2.2 Molecular docking

Three tools were used to dock the Bcl-xL protein with the Genistein ligand: Autodock4 [Morris et al., 2009], Autodock Vina [Trott & Olson, 2010] and SwissDock [Grosdidier et al., 2011]. The first two were executed via command line scripts, and the third used the SwissDock web server interface (<http://www.SwissDock.ch/>).

The final structure of the complete Bcl-xL described above was taken as the receptor after a 10 ns MD simulation in a protein-membrane system as described below. While the ligand was taken from the optimized structure of the Genistein compound described above.

The docking performed was rigid and blind. Rigid because only the ligand and not the receptor is allowed to be flexible, and blind because the entire protein surface is scanned without specifying a particular site or region. However, to perform a smaller scan, we used a cubic grid box centered on the protein head without including the large FLD loop and the membrane-inserted tail (Figure 5).

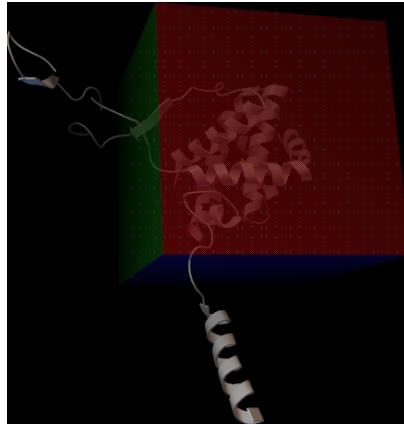


Figure 5: Cubic docking grid box centered on the protein head. The box does not include the extended flexible loop domain (FLD) or the protein tail.

2.3 Selection of docking poses

An iterative selection procedure was carried out to filter out the different poses resulting from the three docking tools. Initially, 326 ligand poses were docked into the protein: 50 from Autodock4, 20 from Vina and 256 from SwissDock. From these, the 30 best poses (10 for each tool) were pre-selected based on their affinity score assigned by each tool. The 30 poses were reduced to 10 by hierarchical clustering based on their coordinates, and the resulting 10 poses were filtered to 6 based on three criteria: their free binding energy (FBE), their proximity to the other poses, and their proximity to helices rather than to loop domains. The FBE was calculated by re-scoring the docking poses with the Autodock4 and Vina scoring functions [Morris et al., 2009, Trott & Olson, 2010] after a 1ns MD simulation in a protein-ligand system without membrane or solvent.

Finally, the last 6 poses were evaluated after a 200 ns MD simulation in a protein-ligand system prepared with each pose (described in the next section) and the pose with the best interactions was selected to run an extended 1000 ns MD simulation and analyze its effect on protein conformational dynamics.

2.4 Molecular dynamics simulations (MD)

MD simulations were used to analyze the conformational dynamics of the protein bound to the ligand, specifically bound to the different ligand poses resulting from the docking process. The results of the simulations were described by different conformational parameters, such as the number of hydrogen bonds (HB), root mean square deviation (RMSD), root mean square fluctuation (RMSF), radius of gyration (RG), intramolecular interactions, protein rotations and hydrophobic binding groove accessibility.

In total, nine systems were prepared for simulation, as shown in the Table 1. In system s1, a 10 ns simulation was performed to minimize and equilibrate the protein in a membrane-anchored system.

In systems s2 to s7, six 200 ns protein-ligand simulations were performed with the membrane-anchored protein, where the last protein structure of system s1 described above was taken as the receptor, and six selected poses from the docking process with the compound Genistein were taken as ligands.

In system s8 a 1000 ns protein-ligand simulation was performed with the protein anchored to the membrane and the pose with the best interactions among the six previous ones as ligand. And in system s9 a simulation was performed as in s8 but without ligand, in order to compare the dynamics of the protein with and without the presence of ligand.

| System | Tool | Description | Simulation time |
|--------|-----------|----------------------------------|-----------------|
| s1 | | Protein + Membrane | 10 ns |
| s2 | Autodock4 | Protein + Membrane + Ligand ad22 | 200 ns |
| s3 | Vina | Protein + Membrane + Ligand av01 | 200 ns |
| s4 | SwissDock | Protein + Membrane + Ligand sw06 | 200 ns |
| s5 | SwissDock | Protein + Membrane + Ligand sw14 | 200 ns |
| s6 | SwissDock | Protein + Membrane + Ligand sw22 | 200 ns |
| s7 | SwissDock | Protein + Membrane + Ligand sw25 | 200 ns |
| s8 | SwissDock | Protein + Membrane + Ligand sw06 | 1000 ns |
| s9 | | Protein + Membrane | 1000 ns |

Table 1: Systems prepared for molecular dynamics simulations.

Input files for the simulations were prepared using the membrane builder module of the CHARMM-GUI web server [Lee et al., 2016, Jo et al., 2007] and molecular dynamics simulations were run with the NAMD 2.14 package (with NVIDIA CUDA acceleration)[Phillips et al., 2020]. For all systems the CHARMM36m force field with CMAP corrections [Huang & MacKerell, 2013], was used, for ligands the CGenFF force field [Vanommeslaeghe & MacKerell, 2012] was used, and a TIP3P water model [Mark & Nilsson, 2001] neutralized with KCl ions was used as solvent. The protein was anchored to a lipid bilayer-type membrane composed of 100% palmitoylphosphatidyl-choline (POPC) with 412 molecules in the upper layer and 411 in the lower layer. All side chains of the acidic and histidine residues of the protein were protonated.

The systems were minimized and equilibrated using the default CHARMM-GUI protocol [Jo et al., 2007], with a 10000-step minimization and a 1 ns gradual and constrained equilibration through six stages, where the first two used NVT dynamics (250 ps, constant volume and temperature), while the remaining stages used NPAT dynamics at 310 K (750 ps, constant pressure, area and temperature). For production simulations, an integration step of 2 fs was defined and coordinates were saved every 100 ps.

2.5 Trajectory analysis

All trajectory analyses were performed with in-house scripts. Conformational parameters were calculated by calling VMD plugins using scripts in TCL. Protein-ligand interactions were calculated using the Prolif package written in Python [Bouysset & Fiorucci, 2021]. Protein rotations were determined from the moment of inertia of alpha carbons in protein structures from the simulation trajectory, where Euler angles were calculated from their initial to final structure. The accessibility of the hydrophobic binding groove was calculated by measuring the angle of deviation between the two normals of the binding groove region (helices $\alpha 3$ and $\alpha 4$) and the upper region of the membrane, where the normals were determined from the principal axes of each region using the VMD orient plugin. Finally, data management and plotting of the results were programmed using R scripts.

3 Results and discussion

3.1 Molecular docking

Initially, the three docking tools produced 326 poses. Next, 30 poses-10 per tool-were selected (Figures 6.A, 6.B, and 6C), which were then filtered to 10 by hierarchical clustering (Figure 6. D). Finally, only 6 poses (Figure 6.E) were retained for further study according to their binding affinity and localization on the protein (see Subsection 2.3).

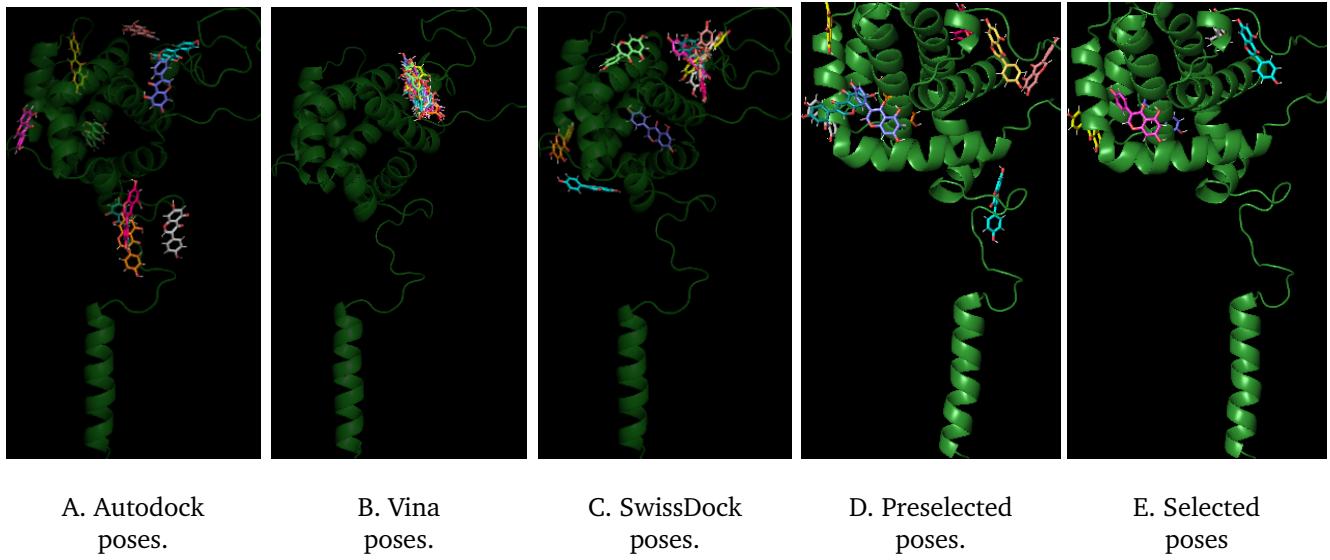


Figure 6: Selected docking poses. (A) The 10 poses selected for Autodock. (B) The 10 poses selected for Vina. (C) The 10 poses selected for SwissDock. (D) The 10 preselected poses after hierarchical grouping. (E) The 6 poses selected for evaluation.

Six membrane-anchored systems were constructed with each of the 6 selected poses from the docking process (Figure 7.A). The constructed systems were named ad22, av01, sw06, sw14, sw22 and sw25, (see Table 1), where the prefixes ad, av and sw refer to Autodock4, Vina and SwissDock, respectively, the tools that generated the poses (Figure 7.B).

A 200 ns MD simulation was run on these systems with the aim to evaluate different conformational parameters and select the best pose bound to the protein (see Subsection 2.4), which is shown in the next Subsection.

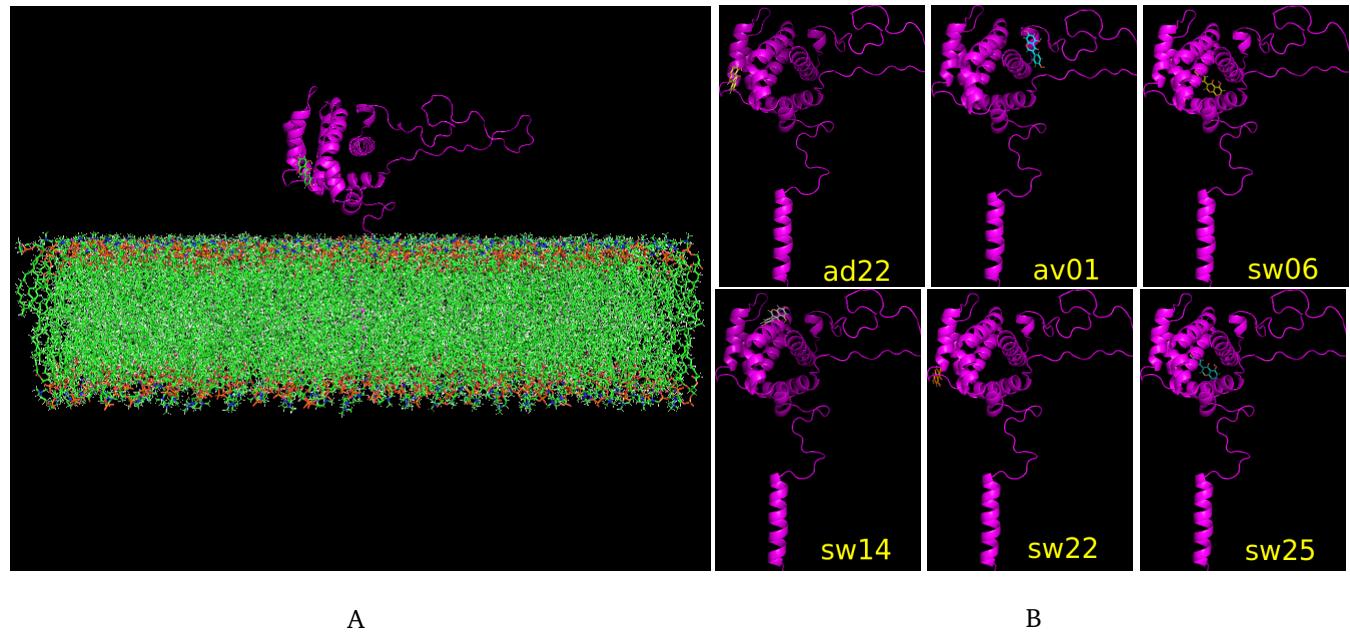


Figure 7: Protein-ligand systems for the selected poses. (A) One of the protein-ligand systems with the membrane-anchored protein, where the solvent and counterions have been hidden to facilitate better observation of the system. (B) The poses and their final location on the receptor protein.

Figure 8 shows some snapshots of the trajectory produced by each system, the simulation time is shown at the top and the ligand pose name in the center. The protein is shown in silver blue and the ligand in light green. In the following, each of the six systems will be named according to the ligand pose identifier, i.e. results are presented for systems ad22, av01, sw06, sw14, sw22 and sw25.

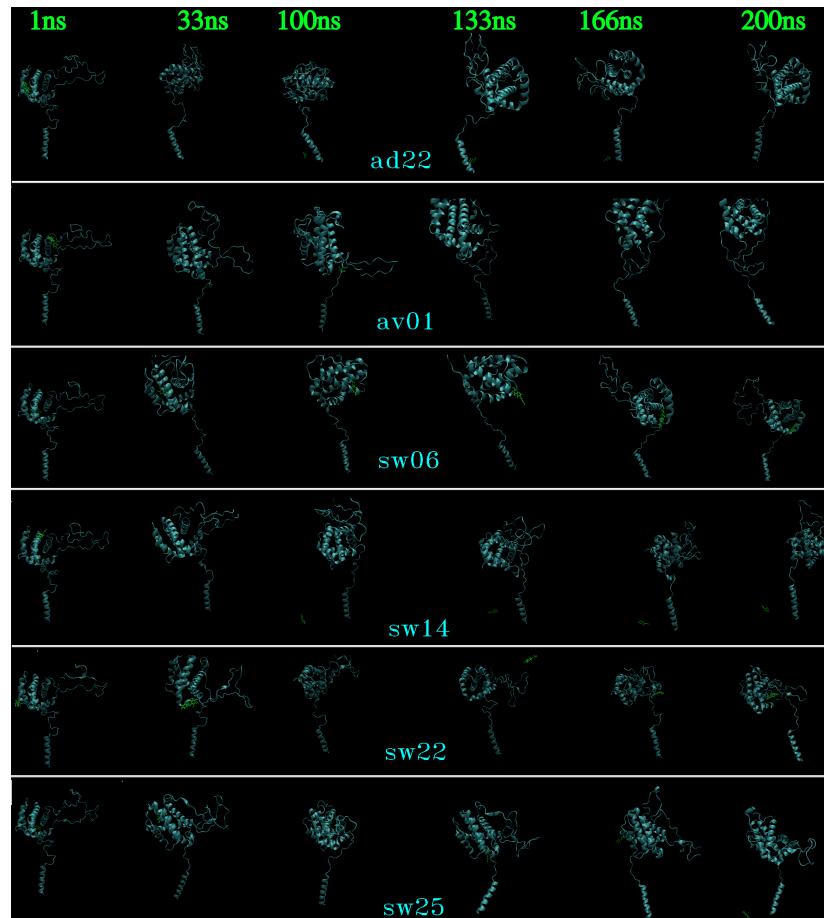


Figure 8: Snapshots of the trajectories of the DM simulations of the six ligand-protein systems.

3.2 Evaluation of selected poses

3.2.1 Number of hydrogen bonds

Figure 9 shows the number of hydrogen bonds formed between the ligand atoms and the protein atoms for the six systems. We observed that the only poses that preserved its bonds until the end of the simulation were sw06 and sw22, suggesting that in these two systems the ligand is bound to the protein while in the others it has already unbonded. However, the sw22 system has a period of time where it completely lost its bonds (~ 70 to ~ 125 ns), suggesting that the ligand may be separating from the protein as in the other systems. This partial separation of the sw22 pose can be seen in the trajectory images for this pose, sw22, in Figure 8.

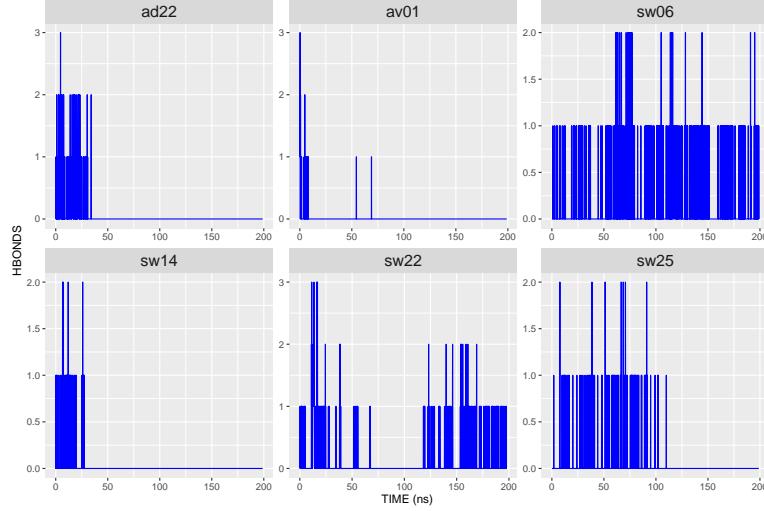


Figure 9: Number of hydrogen bonds of the six systems. The name of each system is indicated at the top of each table.

3.2.2 Root Mean Squared Deviation (RMSD)

RMSD on alpha carbons was used to evaluate protein stability in the simulations. Figure 10 presents the RMSD values of the six systems, calculated between the alpha carbons of the initial protein structure and those of the protein structures from the beginning to the end of the simulation. Figure 10 shows that almost all systems start to stabilize towards the middle of the simulation (~ 100 ns). However, in the sw06 system, in which we assume that the ligand remains bound to the protein, it is observed that after increasing the RMSD, up to almost 100 ns, this value decays steeply ($\sim 20\text{\AA}$), indicating that the protein in the presence of the ligand becomes unstable.

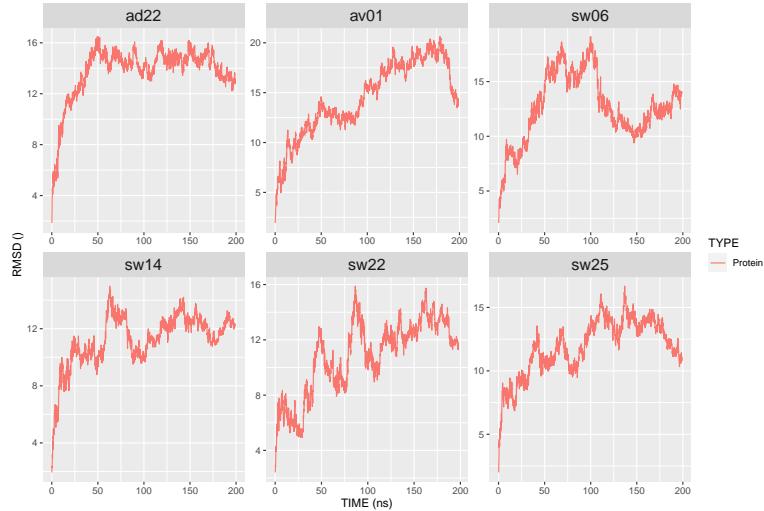


Figure 10: RMSDs for the six systems. The RMSD is calculated on the alpha carbons of the protein structures of the trajectory using the initial structure as a reference..

3.2.3 Root Mean Square Fluctuation (RMSF)

Figure 11 shows the RMSF values calculated from the protein alpha carbons for the six systems along with the protein and its numbered residues. For all systems, a region of large fluctuation is generally observed around AA 50, which corresponds to the largest flexible loop domain of the protein, FLD, (AA 20 to 84, turquoise blue to dark green colors). Furthermore, the average RMSF of the sw06 system was higher (7.3) than that of the others (5.7, 5.3, 6.7, 5.6 and 5.9). Suggesting that this is the only ligand posture that affected the protein, as the other postures are probably no longer bound to the protein, as suggested by the above analyses.

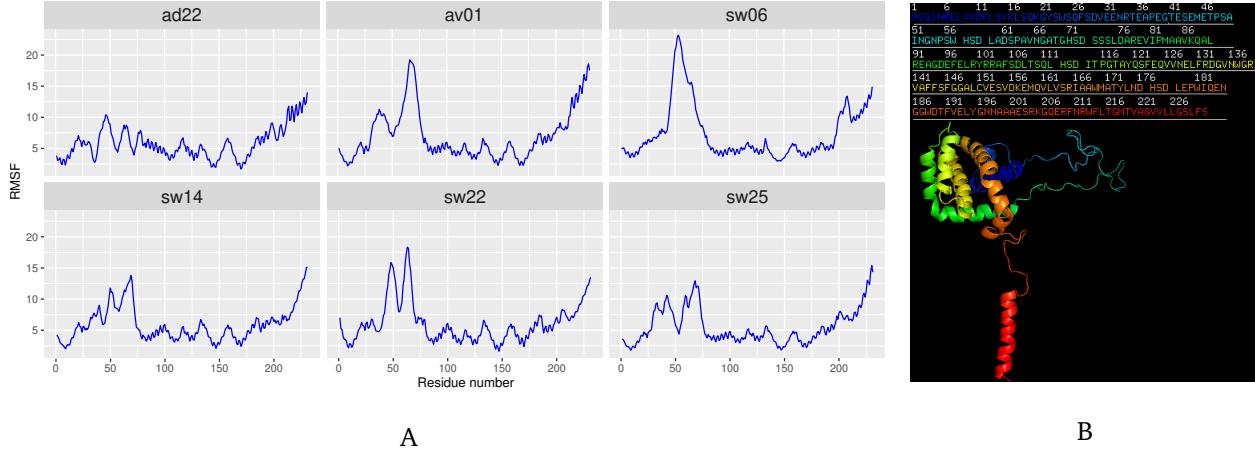


Figure 11: RMSFs for the six systems. (A) RMSF values of each system (B) Protein structure with its residues numbered and colored according to the representative color given to the protein structures.

3.2.4 Radius of Gyration (RG)

We used the RG to compare protein compactness in both the protein-ligand complex and the protein alone (red and blue colored lines in Figure 12, respectively). From the Figure 12, we observed that in all systems the RG value in both cases fluctuates, so the systems were unstable up to that time of the simulation.

Furthermore, we observed that in systems where the ligand is docked into the protein, such as sw06, the RG values were quite similar (lines overlap), while in the other systems the RG values of the protein in the complex are marginally higher than those of the protein alone (red line on blue line). This would indicate that the compactness of the protein in the complex was similar to that of the protein alone and, therefore, there was a good interaction between the ligand and the protein.

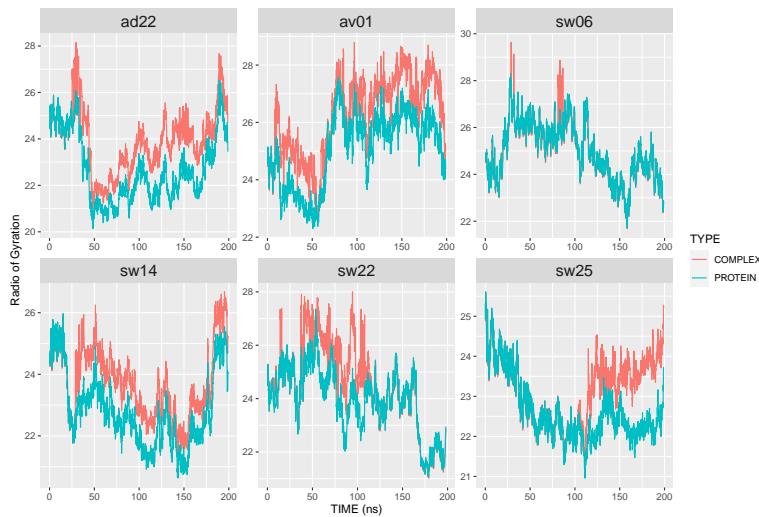


Figure 12: Radius of Gyration for the six systems. The red line corresponds to the RG for the protein in the protein-ligand complex, while the blue line corresponds to the RG calculation for the protein alone.

3.2.5 Protein-ligand interactions

In addition to the above analyses, Figure 13 shows the main interactions of the protein residues with the atoms of the different poses of the ligand. We observed that the only poses that presented interactions until the end of the simulation were sw06 and sw22, although the interactions of the latter are isolated. All this confirms that the sw06 pose is the only one that retained significant interactions until the end of the 200 ns simulation.

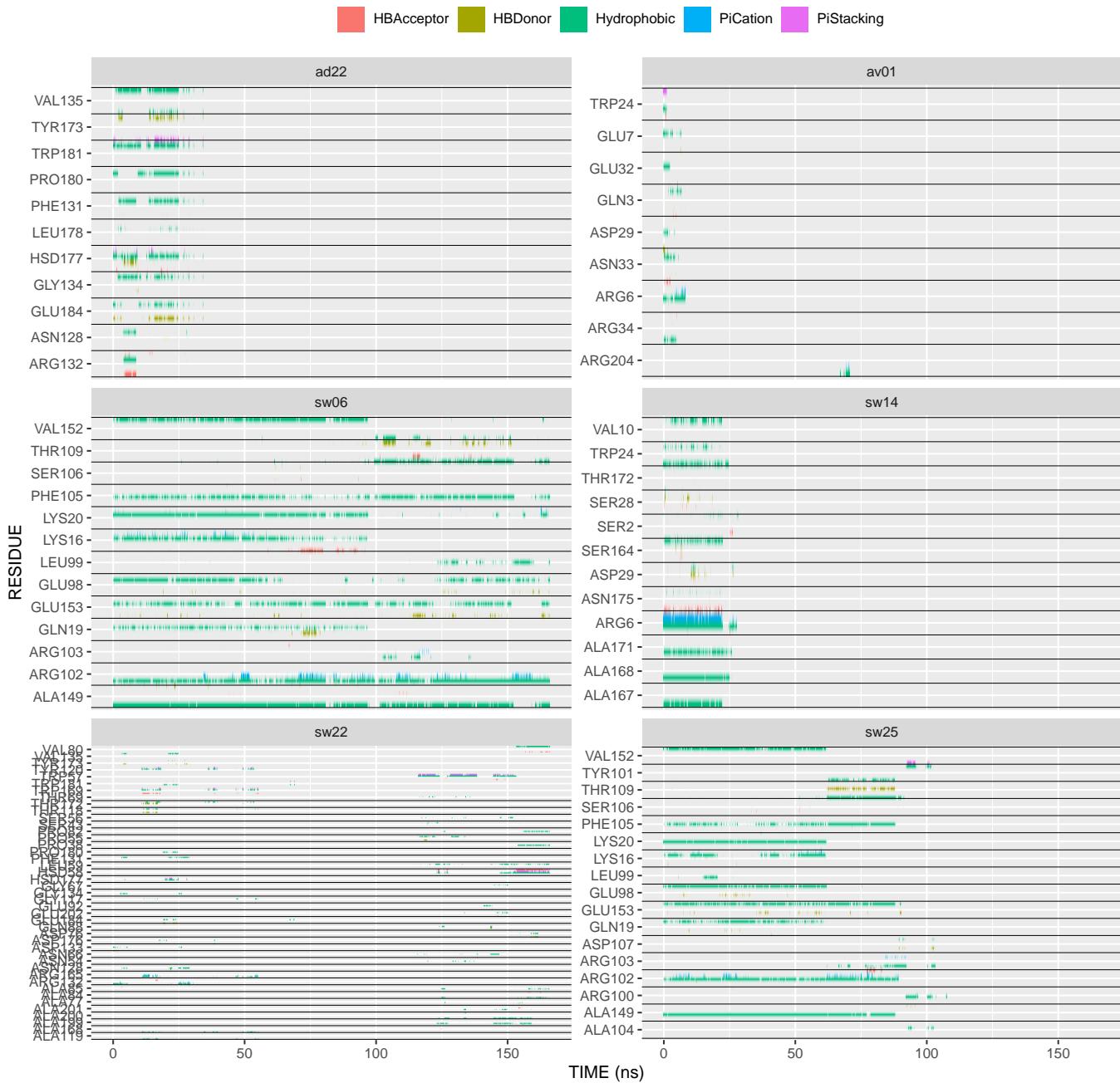


Figure 13: Major interactions between protein residues and the six poses of the ligand.

3.3 Evaluation of the ligand pose in the sw06 system

From the above analysis of the six poses, we conclude that the sw06 pose was the best coupled to the protein and retained the interactions until the end of the 200 ns simulation. Now, we present the analysis of protein binding to this pose after a 1000 ns MD simulation. At the same time, these results are compared with the 1000 ns simulation of the protein alone, without the presence of the ligand, in order to better identify the conformational changes that the protein underwent.

3.3.1 Protein-ligand interactions

Figure 14] presents the most frequent non-covalent interaction forces in the protein-ligand interaction during the time of simulation. The results showed that the interactions were present during the first 350 ns of simulation and then were lost. Furthermore, the forces that contributed most to the binding of the ligand with the protein were first the hydrophobic forces (green lines), second the hydrogen bonds, (red lines for acceptors and ochre for givers), and third the pi-Cation interactions (blue lines).

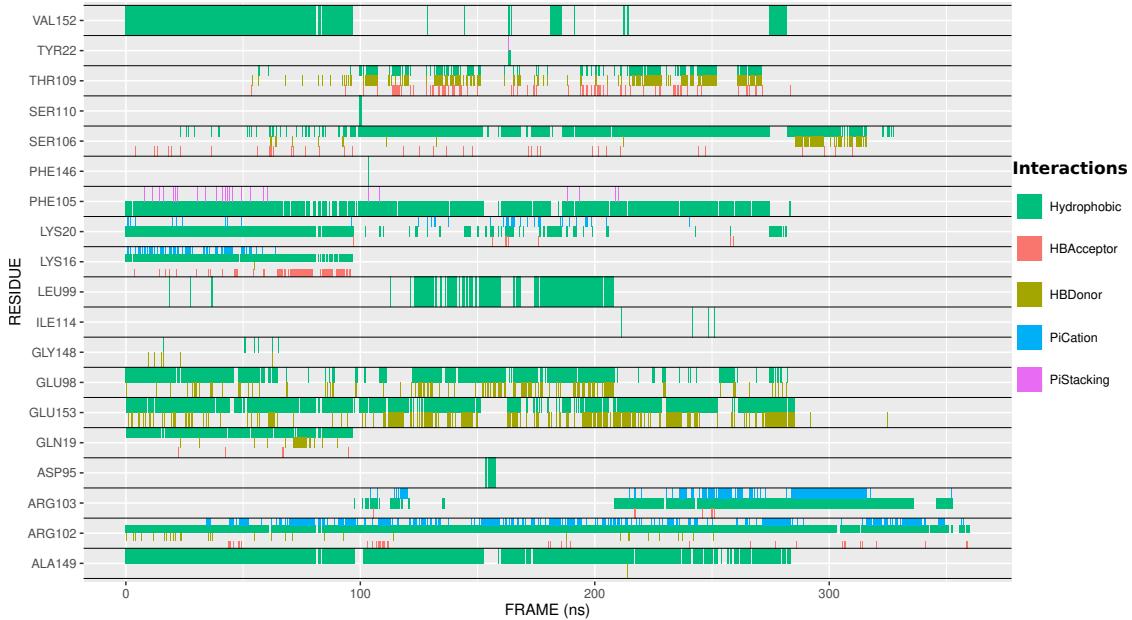


Figure 14: Non-covalent interaction energies between protein residues and atoms of the sw06 pose.

When specifically evaluating hydrogen bonds, we observed that one to two hydrogen bonds were counted between the protein and the sw06 ligand during the first ~ 350 ns of the simulation, which were then lost and consequently the ligand unbonded from the protein (Figure 15]). However, all of these forces contributed to maintain the ligand bound to the protein during this time so it may have affected the conformational dynamics of the protein, which is what will be shown Subsection 3.4.

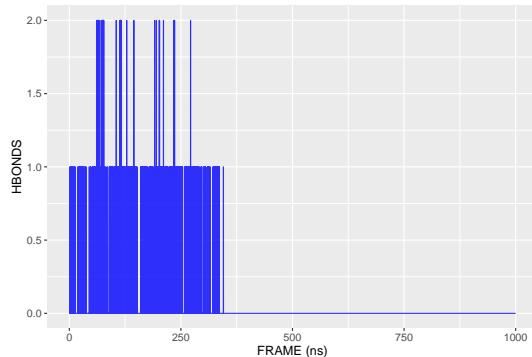


Figure 15: Number of hydrogen bonds in the protein-ligand complex of the sw06 system.

3.3.2 Root Mean Squared Deviation (RMSD)

Figure 16.A compares the RMSD values for protein in complex versus protein alone (red and green lines, respectively). We observed that the RMSD values changed increasingly for the two systems during the first ~50 ns of the simulation and gradually stabilized from ~100 ns onwards. However, sharp changes in the values were possibly due to the continuous fluctuation of the two large protein loops (see section 2.1). To verify the above, Figure 16].B shows the RMSDs for the region of the soluble part of the protein, excluding the two large loops and the transmembrane helix. From Figure 16.B we observed that the RMSD values of the protein in both cases fluctuate slightly and are closely similar, confirming that the main changes in the RMSD values are due to the fluctuation of the loops and not to major structural changes.

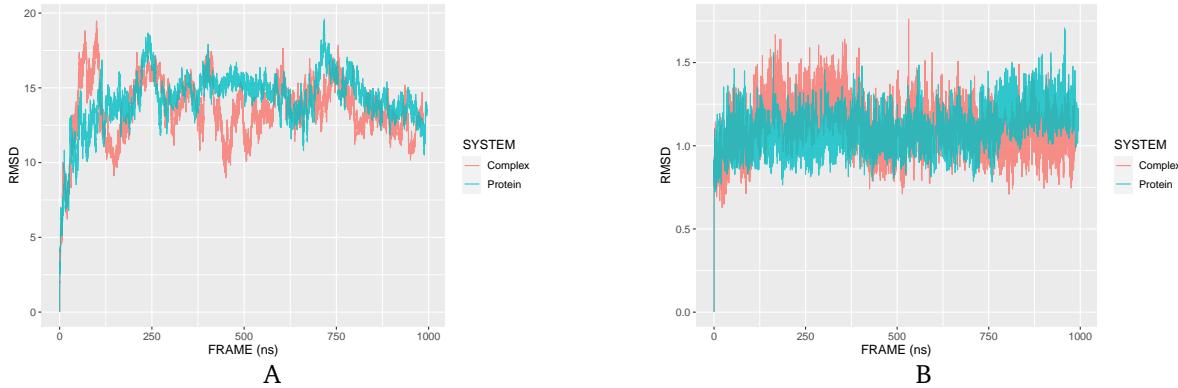


Figure 16: RMSD values for the protein in the sw06 system. RMSDs calculated for the alpha carbons of both the protein in the complex and the protein alone. (A) RMSDs for the complete protein. (B) RMSDs only for the soluble part of the protein, lacking the two large loops and the transmembrane helix.

3.3.3 Root Mean Square Fluctuation (RMSF)

Figure 17.A shows the local structural fluctuations of the protein-ligand complex and protein alone (red and green lines, respectively), both calculated on protein alpha carbons. We observed that the AAs Gly-40, Ser-62, and Glu-208 of the protein-ligand complex showed significant increases in their overall flexibility compared to that of the protein alone. The first two AAs, Gly-40 and Ser-62, belong to the flexible loop domain (FLD), whereas the last AA, Glu-208, belongs to the loop connecting the membrane-anchored C-terminal helix (red color, 17.B).

The length of these two loops makes them very flexible, especially the FLD which contains 60 AA and is located on the opposite side of the hydrophobic binding groove. Furthermore, Figure 17.A also shows that the flexibility of these loops increased with the presence of ligand, which may lead to conformational changes in the protein structure or in its orientation with respect to the membrane.

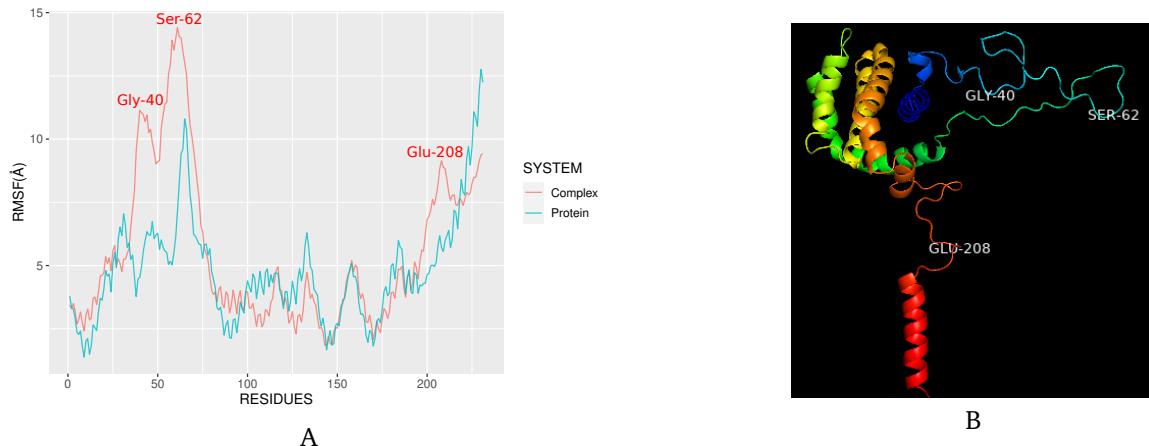


Figure 17: RMSF values for protein in the sw06 system. (A) RMSF plot for protein in the complex (red line) and protein alone (green line) with the names of the most fluctuating AAs. (B) Protein structure showing the location of the most fluctuating AAs..

3.3.4 Radius of Gyration (RG)

We measured the RG to assess the compactness of the protein structure in the two cases: with and without ligand. As shown in Figure 18, the RG values for the two systems fluctuated throughout the simulation, which should be an indication that the protein still lacks compactness. However, these RG values are mainly the result of fluctuations in the two large loops, and especially in the FLD loop, which is extensive and very flexible.

It can also be seen that the RG values for the two cases were very close until ~ 360 ns, the time at which the ligand becomes unbound from the protein. Thereafter, the fluctuations are more pronounced for the protein in the complex than for the protein alone. In the former, the RG of the protein fluctuated considerably, increasing up to 29 Å and then decreasing to 23 Å, whereas in the latter this value decreased to 21 Å. This suggests that the presence of the ligand confers more flexibility to the structure, specifically to the long FLD loop, while the protein structure without the ligand starts to improve its compactness.

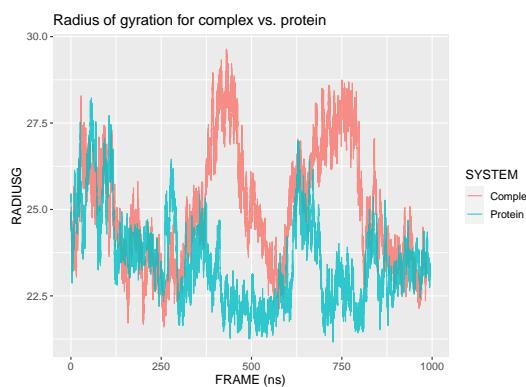


Figure 18: RG values for protein in the sw06 system.

3.4 Conformational dynamics

3.4.1 Protein rotations

Figura 19 shows the rotations undergone by the protein in the presence of ligand and by the protein alone. While in both systems the rotations around the X and Y axes (red and green lines in Figure Figura 19) were smaller, in the range from -50 to 50°, the rotations around the Z axis (blue line in Figure Figura 19) were larger. In the protein-ligand complex they started to stabilize very early (~125 ns, -220° on average), whereas in the protein alone, they started to stabilize at the end of the simulation (~620 ns, -100° on average).

As a consequence of this ~220° turn, induced by the presence of the ligand, the hydrophobic groove formed by helices 3 and 4 (Figure 19.B, yellow colored region of the protein) is observed more distant from the membrane and thus more exposed to other BCL-2 family proteins (e.g. Bak) binding more readily and potentially activating the anti-apoptotic function of the Bcl-xL protein.

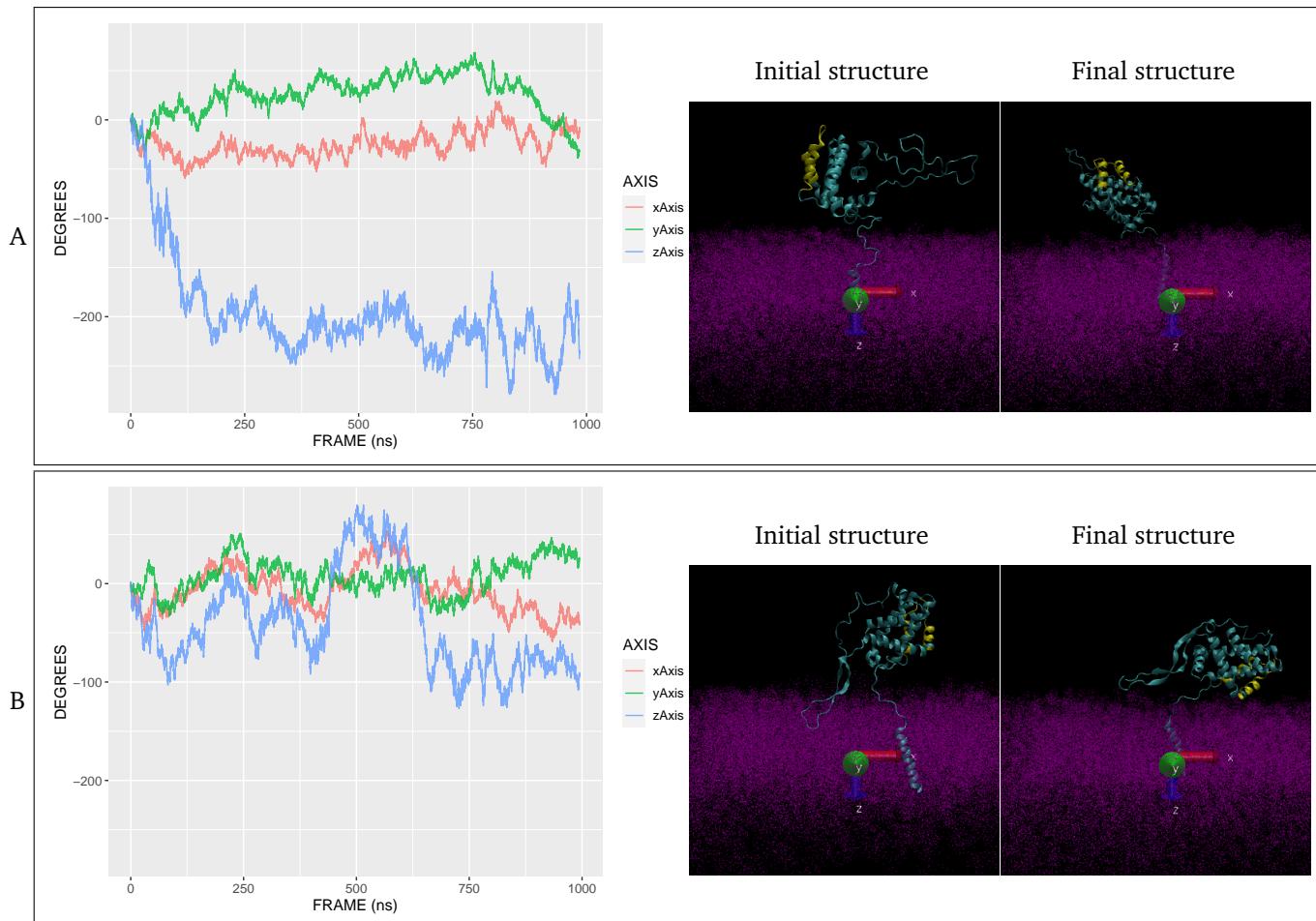


Figure 19: Rotational movements of the protein in the sw06 system. (A) Protein in the protein-ligand complex. (B) Protein alone. The region colored yellow corresponds to the helices α_3 and α_4 , which form the hydrophobic binding groove..

3.4.2 Accessibility of the hydrophobic binding groove

To quantify the accessibility of the hydrophobic groove, the deviation angle formed between its normal vector and the negative normal of the membrane was measured (long red and green arrows in the figure 20, respectively). Depending on the orientation of the vectors, deviation angles less than 90° suggest that the hydrophobic groove is partially hidden and more difficult to access for other proteins. In contrast, angles greater than 90° expose more of the hydrophobic groove and thus facilitate its access for binding to other proteins.

Considering the above, we observed that in the protein-ligand complex the angle between the hydrophobic groove and the membrane increased from 104° to 145° at the end of the simulation (Figure 20.A). Whereas in the protein alone the angle was reduced from 94° to 57° (Figure 20.B). This suggests that the protein in the presence of ligand exposed more of the hydrophobic groove, and this occurred very early in the simulation (~ 100 ns, Figure 20.A) and remained so until the end.

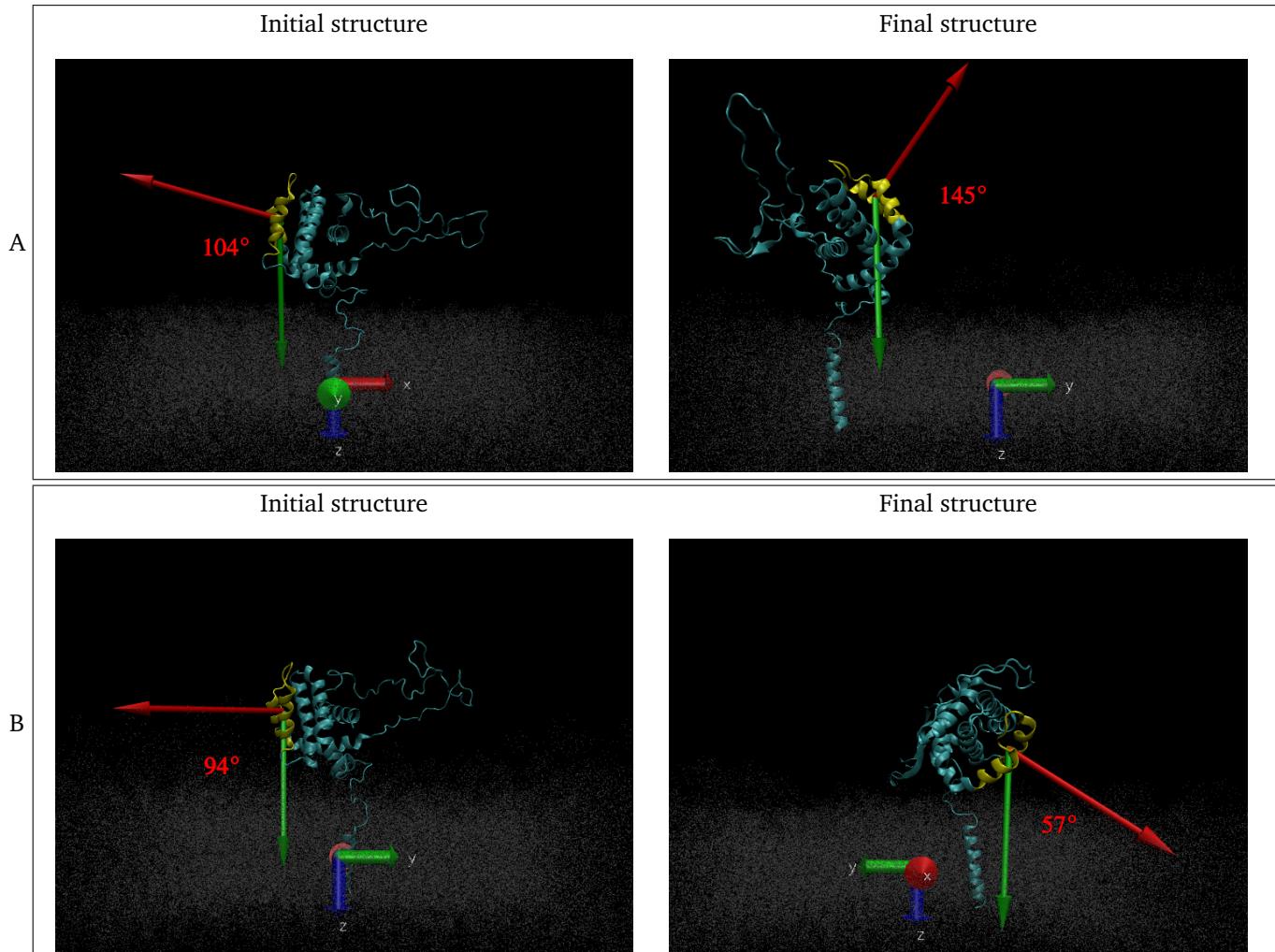


Figure 20: Deviation angles between the hydrophobic groove and the membrane in the sw06 system. (A) Protein in the protein-ligand complex. (B) Protein alone. The yellow-colored region corresponds to the helices α_3 and α_4 , which form the hydrophobic binding groove. The red vector is the normal to the plane formed by helices α_3 and α_4 of the hydrophobic groove and the green vector is the normal vector of the membrane.

4 Final discussion

A better exploration of the ligand conformational space was achieved by using the three molecular docking tools Autodock4, Vina and SwissDock in this work. The results could have been biased by using only one, as was the case for the poses produced with the Vina tool (see Figure 6.B), all of which were located in the same region, very close to the flexible loop domain (FLD). Based on the literature, it was found that the Vina tool is not very suitable for blind docking, which was the type of docking performed in this work, and that is why there are alternative versions of the algorithm that seek to improve this type of exploration [Hassan et al., 2017].

Now, to reduce the conformational space that would have to be explored by blind docking, a rectangular box was defined around the globular part of the protein without including most of the flexible loop domain (FLD) and consequently without the exploration of possible ligand poses in these regions (see Figure 5). However, despite the reduced search space, a large number of ligand poses were docked to the protein which were filtered using different techniques and criteria (see Subsection 2.3). In the end, only one pose was selected for further studies, sw06, as the other poses were unbound from the protein before the 200 ns of simulation were completed (see Figures 13 and 9).

The conformational dynamics of the protein in the presence of ligand was measured by different conformational parameters along with measurements of protein orientation and hydrophobic groove accessibility. Measurements of protein RMSD values (see Figure 16) showed large fluctuations that could be interpreted as changes in protein structure. Similarly, protein compactness had large variations, as observed in the RG values (see Figure 18). However, as RMSF analyses showed, the regions in which the protein fluctuates the most are in the two long loops of the protein: the flexible loop domain (FLD) and the transmembrane helix-loop-helix-loop-helix (see Section 1), mainly the FLD (60 AA), which is the most extensive and flexible domain of the protein.

On the other hand, the observed rotational motions of the protein in the protein-ligand complex show a large rotation of $\sim 220^\circ$ about the Z axis and smaller rotations ($\sim 50^\circ$) about the X and Y axes that, taken together, could be exposing more of the hydrophobic groove region (helices $\alpha 3$ and $\alpha 4$ in yellow, Figure 19.B) and consequently making it more accessible to other Bcl-2 proteins (e.g. Bax), which could more readily dock and thus activate its anti-apoptotic function.

Finally, it should be noted that there are other elements that must be taken into account in the interaction of this protein with any ligand. In this work, a cubic grid box was defined that does not include most of the FLD in order to limit the docking search space. Nevertheless, some papers focused on the FLD domain report that apart from being an inherently disordered and flexible region, the FLD plays an important role in the stability and regulation of the anti-apoptotic activity of this type of proteins [Lan et al., 2020, Caro-Gómez et al., 2019, Raghav et al., 2012] also facilitating the binding of different ligands [Caro-Gómez et al., 2019]. Additionally, it is necessary to consider that the MD simulations performed in this work used only the protonated protein anchored to a membrane composed of 100% palmitoylphosphatidylcholine (POPC). However, it has been reported that both the type of protonation and the membrane composition can influence the conformational dynamics of the Bcl-xL protein [Tyagi et al., 2021] and this may influence the interaction with the ligand.

5 Availability

All data and scripts from the execution and analysis of the molecular docking and molecular dynamics simulations are located in the repository at <https://github.com/luisgarreta/dockingBCL2>.

The data on preparation of the systems for simulation, the simulation run scripts and the resulting trajectories are located on the omicas server (IP: 192.168.170.227, "/users/legarreta").

6 Acknowledgments

The author thanks the Pontificia Universidad Javeriana de Cali for funding this postdoctoral study through the project "Utilización de compuestos fenólicos y productos de la reacción de Maillard derivados de la caña de azúcar para la obtención de productos con valor agregado: Potencial antioxidante y anticancerígeno de las melazas y vinazas" with code 125180864199 and contract number 859-2018 corresponding to the call 808 of 2018 of Colciencias, funded by MINCIENCIAS and led by Professor Fabián Tobar.

In addition, the author thanks the Instituto de Investigación en Ciencias Ómicas (iÓmicas) of the Pontificia Universidad Javeriana de Cali and especially its director Dr. Andrés Jaramillo Botero for his valuable assistance and for providing the computational infrastructure and facilities for the development of this project.

References

- [Ali et al., 2021] Ali, S. E., Yuan, Q., Wang, S., & Farag, M. A. (2021). More than sweet: A phytochemical and pharmacological review of sugarcane (*Saccharum officinarum* L.). *Food Bioscience*, 44(PB), 101431.
- [Borrás et al., 2020] Borrás, C., Mas-Bargues, C., Román-Domínguez, A., Sanz-Ros, J., Gimeno-Mallench, L., Inglés, M., Gambini, J., & Viña, J. (2020). BCL-xL, a Mitochondrial Protein Involved in Successful Aging: From *C. elegans* to Human Centenarians. *International journal of molecular sciences*, 21(2), 418.
- [Bouyssset & Fiorucci, 2021] Bouyssset, C. & Fiorucci, S. (2021). ProLIF: a library to encode molecular interactions as fingerprints. *Journal of Cheminformatics*, 13(1), 1–9.
- [Caro-Gómez et al., 2019] Caro-Gómez, L. A., Rosas-Trigueros, J. L., Mixcoha, E., Vique-Sánchez, J. L., Gasperin-Sánchez, H., Benítez-Cardoza, C. G., & Zamorano-Carillo, A. (2019). Exploring the Conformational Space of Bcl-2 Protein Variants: Dynamic Contributions of the Flexible Loop Domain and Transmembrane Region. *Molecules (Basel, Switzerland)*, 24(21), 3896.
- [Eswar et al., 2006] Eswar, N., Webb, B., Marti-Renom, M. A., Madhusudhan, M. S., Eramian, D., Shen, M.-y., Pieper, U., & Sali, A. (2006). Comparative Protein Structure Modeling Using Modeller. *Current Protocols in Bioinformatics*, 15(1), 5.6.1–5.6.30.
- [Grosdidier et al., 2011] Grosdidier, A., Zoete, V., & Michelin, O. (2011). SwissDock, a protein-small molecule docking web service based on EADock DSS. *Nucleic Acids Research*, 39(SUPPL. 2), 270–277.
- [Hassan et al., 2017] Hassan, N. M., Alhossary, A. A., Mu, Y., & Kwoh, C.-K. (2017). Protein-Ligand Blind Docking Using QuickVina-W With Inter-Process Spatio-Temporal Integration. *Scientific Reports*, 7(1), 15451.
- [Huang & MacKerell, 2013] Huang, J. & MacKerell, A. D. J. (2013). CHARMM36 all-atom additive protein force field: validation based on comparison to NMR data. *Journal of computational chemistry*, 34(25), 2135–2145.
- [Jo et al., 2007] Jo, S., Kim, T., & Im, W. (2007). Automated Builder and Database of Protein/Membrane Complexes for Molecular Dynamics Simulations. *PLOS ONE*, 2(9), 1–10.
- [Kim et al., 2004] Kim, P. K., Annis, M. G., Dlugosz, P. J., Leber, B., & Andrews, D. W. (2004). During Apoptosis Bcl-2 Changes Membrane Topology at Both the Endoplasmic Reticulum and Mitochondria. *Molecular Cell*, 14(4), 523–529.
- [Lan et al., 2020] Lan, Y.-J. J., Yeh, P.-S. S., Kao, T.-Y. Y., Lo, Y.-C. C., Sue, S.-C. C., Chen, Y.-W. W., Hwang, D. W., & Chiang, Y.-W. W. (2020). Anti-apoptotic BCL-2 regulation by changes in dynamics of its long unstructured loop. *Communications biology*, 3(1), 668.
- [Lee & Fairlie, 2019] Lee, E. F. & Fairlie, W. D. (2019). The Structural Biology of Bcl-x(L). *International journal of molecular sciences*, 20(9), 2234.
- [Lee et al., 2016] Lee, J., Cheng, X., Swails, J. M., Yeom, M. S., Eastman, P. K., Lemkul, J. A., Wei, S., Buckner, J., Jeong, J. C., Qi, Y., Jo, S., Pande, V. S., Case, D. A., Brooks, C. L. I. I. I., MacKerell, A. D. J., Klauda, J. B., & Im, W. (2016). CHARMM-GUI Input Generator for NAMD, GROMACS, AMBER, OpenMM, and CHARMM/OpenMM Simulations Using the CHARMM36 Additive Force Field. *Journal of Chemical Theory and Computation*, 12(1), 405–413.
- [Mark & Nilsson, 2001] Mark, P. & Nilsson, L. (2001). Structure and Dynamics of the TIP3P, SPC, and SPC/E Water Models at 298 K. *The Journal of Physical Chemistry A*, 105(43), 9954–9960.
- [Morris et al., 2009] Morris, G. M., Huey, R., Lindstrom, W., Sanner, M. F., Belew, R. K., Goodsell, D. S., & Olson, A. J. (2009). AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. *Journal of computational chemistry*, 30(16), 2785–2791.
- [Mushtaq et al., 2021] Mushtaq, A. U., Ådén, J., Clifton, L. A., Wacklin-Knecht, H., Campana, M., Dingeldein, A. P. G., Persson, C., Sparrman, T., & Gröbner, G. (2021). Neutron reflectometry and NMR spectroscopy of full-length Bcl-2 protein reveal its membrane localization and conformation. *Communications Biology*, 4(1), 507.

- [Phillips et al., 2020] Phillips, J. C., Hardy, D. J., Maia, J. D. C., Stone, J. E., Ribeiro, J. V., Bernardi, R. C., Buch, R., Fiorin, G., Hénin, J., Jiang, W., & Others (2020). Scalable molecular dynamics on CPU and GPU architectures with NAMD. *The Journal of chemical physics*, 153(4), 44130.
- [Raghav et al., 2012] Raghav, P. K., Verma, Y. K., & Gangenahalli, G. U. (2012). Molecular dynamics simulations of the Bcl-2 protein to predict the structure of its unordered flexible loop domain. *Journal of Molecular Modeling*, 18(5), 1885–1906.
- [Renault et al., 2017] Renault, T. T., Dejean, L. M., & Manon, S. (2017). A brewing understanding of the regulation of Bax function by Bcl-xL and Bcl-2. *Mechanisms of Ageing and Development*, 161, 201–210.
- [Ryzhov et al., 2020] Ryzhov, P., Tian, Y., Yao, Y., Bobkov, A. A., Im, W., & Marassi, F. M. (2020). Conformational States of the Cytoprotective Protein Bcl-xL. *Biophysical journal*, 119(7), 1324–1334.
- [Trott & Olson, 2010] Trott, O. & Olson, A. J. (2010). AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *Journal of computational chemistry*, 31(2), 455–461.
- [Tyagi et al., 2021] Tyagi, V., Vasquez-Montes, V., Freites, J. A., Kyrychenko, A., Tobias, D. J., & Ladokhin, A. S. (2021). Effects of Cardiolipin on the Conformational Dynamics of Membrane-Anchored Bcl-xL.
- [Vanommeslaeghe & MacKerell, 2012] Vanommeslaeghe, K. & MacKerell, A. D. J. (2012). Automation of the CHARMM General Force Field (CGenFF) I: Bond Perception and Atom Typing. *Journal of Chemical Information and Modeling*, 52(12), 3144–3154.
- [Wakui et al., 2018] Wakui, N., Yoshino, R., Yasuo, N., Ohue, M., & Sekijima, M. (2018). Exploring the selectivity of inhibitor complexes with Bcl-2 and Bcl-XL: A molecular dynamics simulation approach. *Journal of Molecular Graphics and Modelling*, 79, 166–174.