**Author name:**

Tawsif Al Arian

[tawsifalarian43@gmail.com](mailto:tawsifalarian43@gmail.com)

<https://orcid.org/0000-0001-5237-3501>

Department of Pharmacy

Jahangirnagar University, Dhaka-1342

**Methodology:**  
**Molecular Dynamics Simulation and molecular mechanics-Poisson Boltzmann surface area (MMPBSA) calculations**

Molecular dynamics is a sophisticated automated simulation approach used for evaluating the degree of stability of protein and protein-ligand complex structure at the minuscule stage through demonstrating the behavioral characteristics, interacting arrangement, fluctuation, physical foundation of function, and structure. The molecular dynamics (MD) simulations were conducted using GROMACS version 2023.1 to investigate the behavior these four complexes [https://doi.org/10.1016/j.softx.2015.06.001]. The protein was parameterized for its protein content using the CHARMM General Force Field. The SwissParam server was utilized to perform the ligand topologies [https://doi.org/10.1002/jcc.21816]. The structures underwent 2500 cycles of vacuum minimization using the steepest descent method in order to mitigate any potential steric issues. The solvation of the structure was accomplished through the utilization of the Simple Point Charge (SPC) water model. Subsequently, the system was rendered neutral through the introduction of Na+ and Cl- ions utilizing the gmx genion tool. This measure was implemented in order to maintain the overall electrical neutrality of the system. After minimization, three steps were conducted in the MD simulation: NVT, NPT, and production. The equilibration of the systems was conducted in two phases. Initially, an NVT equilibration lasting 100 picoseconds was conducted to attain a steady state of the number of particles, volume, and temperature. The purpose of this step was to elevate the system to a temperature of 300 Kelvin. The second step involved conducting a 100 picosecond NPT equilibration, which aimed to maintain the system's pressure and density stability by ensuring an equal number of particles, pressure, and temperature. The simulations involved the imposition of bond constraints on all bonds within the protein, thereby inducing position restraint of the protein group. The constrained conditions of NVT and NPT resulted in the relaxation of water molecules surrounding the protein, leading to a decrease in system entropy. The dynamics were conducted utilizing the Parrinello-Rahman barostat and the v-rescale thermostat [https://doi.org/10.1016/j.molliq.2022.120116]. The relaxation of the barostat and thermostat persisted for a duration of 100 picoseconds. The application of Linear Constraint Solver algorithm was utilized to impose constraints on the covalent bonds. The Particle-Mesh Ewald (PME) method was employed to process the electrostatic interactions. After reaching equilibrium, every system underwent a production run that lasted for 100 nanoseconds (ns) of simulation time.

After the MD simulation, the GROMACS package built-in-tools were deployed to evaluate the trajectories in terms of Principle Component Analysis (PCA) and MMPBSA. The functions “gmx covar” and “gmx anaeig” functions were used for the implementation of PCA. [https://doi.org/10.7717/peerj.14120]

Total decomposition contribution (TDC) was completed via the MMPBSA command prompt to figure out the different protein ligand contacts with other residues.

The MMPBSA approach calculates the interaction of free energy of protein-ligand complexes (potential, polar-solvation and non-polar solvation energies) for molecular dynamics simulation trajectories. The gmx\_MMPBSA (1.5.7) tool was employed for the last 50 ns of the simulation processes [https://doi.org/10.1021/ct300418h]. The trajectories were prepared by removing PBC conditions, and calculation performed using Amberff19SB and GAFF forcefields were used for protein and ligand under the dielectric model (ipb=2) [https://doi.org/10.1021/ct300341d]. Binding free energies were then calculated using either of the following two equations

ΔG (binding) = G(complex) – G(receptor) – G (ligand)

ΔG (binding) = ΔH – TΔS

ΔH corresponds to the enthalpy of binding and – TΔS to the conformational entropy after ligand binding)

In which to calculate each term of the above equation is calculated individually through the application of several procedures to get the energy terms for complex, receptor and ligand as described in the literature [https://doi.org/10.1021/acs.jctc.1c00645s].

**Result**

**Molecular Dynamics Simulation and molecular mechanics-Poisson Boltzmann surface area (MMPBSA) calculations**

**Root Mean Square Deviation (RMSD)**

The Root Mean Square Deviation (RMSD) is used to measure the average change in displacement of a selection of atoms for a particular frame with respect to a reference frame. It is calculated for all frames in the trajectory. The RMSD for frame x is:

where N is the number of atoms in the atom selection; t ref is the reference time, (typically the first frame is used as the reference and it is regarded as time t=0); and r' is the position of the selected atoms in frame x after superimposing on the reference frame, where frame x is recorded at time t x . The procedure is repeated for every frame in the simulation trajectory.

XmGrace package was used to analyze the MD trajectory parameters [<https://doi.org/10.1371/journal.pone.0158939>]. During the MD run, the RMSD is calculated between a defined starting point of the simulation and all succeeding frames [<https://doi.org/10.1089/cmb.2010.0237>]. From the analysis of the result, the average, maximum and minimum RMSD values are in **Table 1** . From the graph, the stability of the RMSD values for all the complexes are in the range of 17 to 72 nanoseconds as shown in **Figure 1(i)** and **Figure 1(ii)**.

The average, maximum and minimum RMSD values can be seen in Table 1. The RMSD is useful in analyzing the time‐dependent motion of a given structure throughout the simulation [<https://doi.org/10.1371/journal.pone.0119264>]. In this way, a plateau of RMSD values indicates that the structure fluctuates around a stable average conformation, which can be observed in all of the performed MD simulations.

**Table 1: RMSD Analysis**

|  |  |  |  |
| --- | --- | --- | --- |
| **Value** | **RMSD (nm) in C-α** | | |
| **Average (nm)** | **Minimum (nm)** | **Maximum (nm)** |
| 4QMX (Protein) | 2.428015 | 0.000494 | 4.880086 |
| Hit8 (Ligand) | 0.033533 | 2.428015 | 0.087175 |
| Hit7 (Ligand) | 0.054567 | 2.428015 | 0.090477 |
| Hit1 (Ligand) | 0.043659 | 2.428015 | 0.086057 |
| D1 (Ligand) | 0.144627 | 2.428015 | 0.188523 |

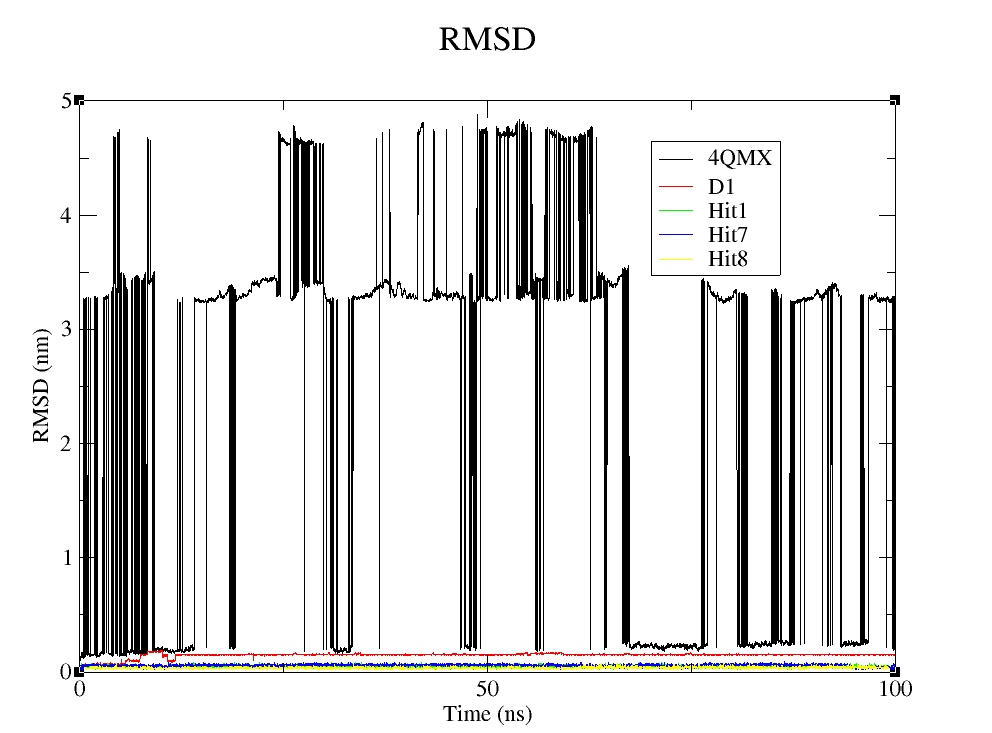


Figure 1: RMSD (Root Mean Square Deviation) of (i) Protein RCSB ID : 4QMX (Black color) and Ligand(s): D1 (Red), Hit1 (Green), Hit7 (Blue), Hit8 (Yellow)

**Root Mean Square Fluctuation (RMSF)**

The local flexibility based on residue displacements during the MD simulation can be defined using RMSF values [<https://doi.org/10.1371/journal.pone.0158939> ]. From the graph, the stability of the RMSF value for all the complexes are in the range of 1000 to 3200 residues as shown in **Figure 2.** The Root Mean Square Fluctuation (RMSF) is useful for characterizing local changes along the protein chain. The RMSF for residue i is:

where T is the trajectory time over which the RMSF is calculated, t ref is the reference time, r i is the position of residue i; r' is the position of atoms in residue i after superposition on the reference, and the angle brackets indicate that the average of the square distance is taken over the selection of atoms in the residue. On this plot, peaks indicate areas of the protein that fluctuate the most during the simulation. Typically, it is observed that the tails (N- and C-terminal) fluctuate more than any other part of the protein.

The average, maximum and minimum RMSF values for the complexes can be seen in **Table 2**. Higher RMSF values represent more flexible movements, whereas lower RMSF values represent movements that are more constrained in regard to average locations during simulation [<https://doi.org/10.1371/journal.pone.0119264>].

**Table 2: RMSF Analysis**

|  |  |  |  |
| --- | --- | --- | --- |
| **Value** | **Radius of Gyration (nm)** | | |
| **Average (nm)** | **Minimum (nm)** | **Maximum (nm)** |
| Hit8-4QMX | 1.066556 | 0.2491 | 2.1544 |
| Hit7-4QMX | 1.167639 | 0.3199 | 2.2164 |
| Hit1-4QMX | 1.848276 | 0.581 | 3.6449 |
| D1-4QMX | 2.058708 | 0.609 | 3.8327 |

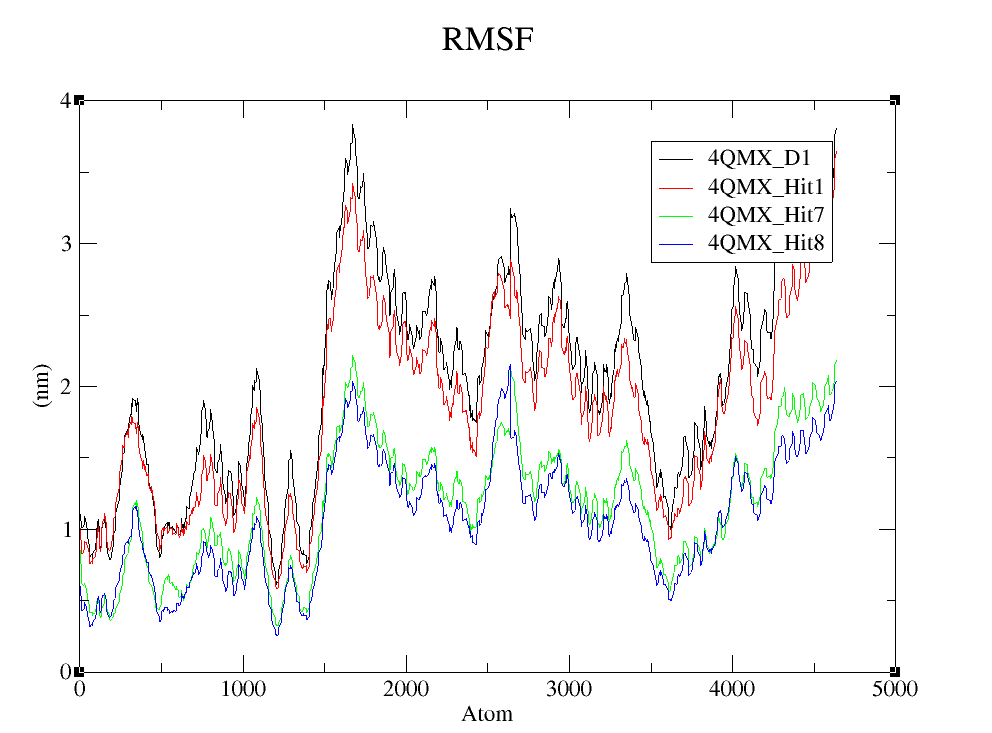


Figure 2: RMSF (Root Mean Square Fluctuation) of 4QMX\_D1 (Black color), 4QMX \_Hit1 (Red), 4QMX\_Hit7 (Green), 4QMX\_Hit8 (Blue)

**Radius of Gyration (Rg)**

The radius of gyration (RG) is commonly described as the root mean square distance of a group of atoms from their shared center of mass, with the added consideration of mass weighting. [<https://doi.org/10.1155/2014/502618>]. From the graph, the stability of the Radius of gyration value for all the complexes are in the range of 15 to 70 nanoseconds as shown in **Figure 3**. The average, maximum and minimum Rg values for the complexes can be seen in **Table 3.**

**Table 3: Radius of Gyration Analysis**

|  |  |  |  |
| --- | --- | --- | --- |
| **Value** | **Radius of Gyration (nm)** | | |
| **Average (nm)** | **Minimum (nm)** | **Maximum (nm)** |
| Hit8-4QMX | 2.267156 | 1.95753 | 4.51784 |
| Hit7-4QMX | 2.294092 | 1.94422 | 4.60393 |
| Hit1-4QMX | 2.932015 | 1.93524 | 5.64843 |
| D1-4QMX | 3.435917 | 1.95693 | 5.62215 |

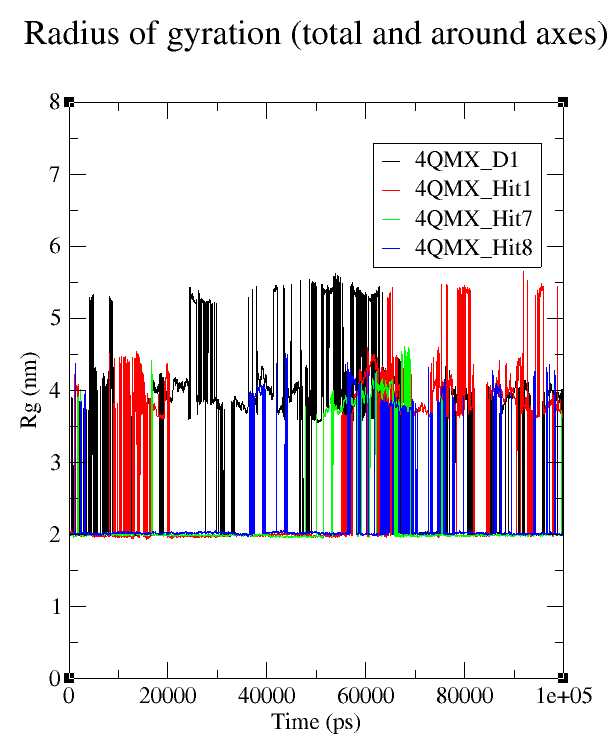


Figure 3: Radius of Gyration of 4QMX\_D1 (Black color), 4QMX \_Hit1 (Red), 4QMX\_Hit7 (Green), 4QMX\_Hit8 (Blue)

**Solvent Accessible Surface Area (SASA)**

The SASA, or solvent-accessible surface area, is widely recognized as the portion of a given protein that is accessible to the surrounding solvent. The SASA analysis methodology offers insights into the ability of a protein to engage in molecular interactions [<https://doi.org/10.1155/2014/502618>]. From the graph, the stability of the SASA value for all the complexes are in the range of 20 to 75 nanoseconds as shown in **Figure 4**. The average, minimum and maximum SASA values for the complexes can be seen in **Table 4.**  The findings indicate that lower value exhibit comparatively lower accessibility in comparison to the complex with higher value, potentially impacting their ability to engage with other molecules.

**Table 4: SASA analysis**

|  |  |  |  |
| --- | --- | --- | --- |
| **Value** | **SASA (nm2)** | | |
| **Average (nm2)** | **Minimum (nm2)** | **Maximum (nm2)** |
| Hit8-4QMX | 156.0108 | 144.054 | 168.349 |
| Hit7-4QMX | 151.7943 | 144.064 | 160.864 |
| Hit1-4QMX | 149.2164 | 140.825 | 162.479 |
| D1-4QMX | 152.9101 | 143.21 | 161.829 |

A graph of a graph

Description automatically generated with medium confidence

Figure 4: SASA (Solvent Accessible Surface Area) of 4QMX\_D1 (Black color), 4QMX \_Hit1 (Red), 4QMX\_Hit7 (Green), 4QMX\_Hit8 (Blue)

**Hydrogen Bonds**

The study of hydrogen bonds can direct the alteration of a lead molecule to increase its activity and provide important insights into the stability of a ligand-protein complex. (H-bonds) play a significant role in ligand binding. Consideration of the number of hydrogen bonds formed is important because of their strong influence on drug specificity, metabolization and adsorption. Hydrogen bonds between a protein and a ligand can be further broken down into four subtypes: backbone acceptor; backbone donor; side-chain acceptor; side-chain donor. The current geometric criteria for protein-ligand H-bond is distance of 3.5 Å between the donor and acceptor atoms (D—H···A); a donor angle of ≥120° between the donor-hydrogen-acceptor atoms (D—H···A); and an acceptor angle of ≥90° between the hydrogen-acceptor-bonded atom atoms (H···A—X).

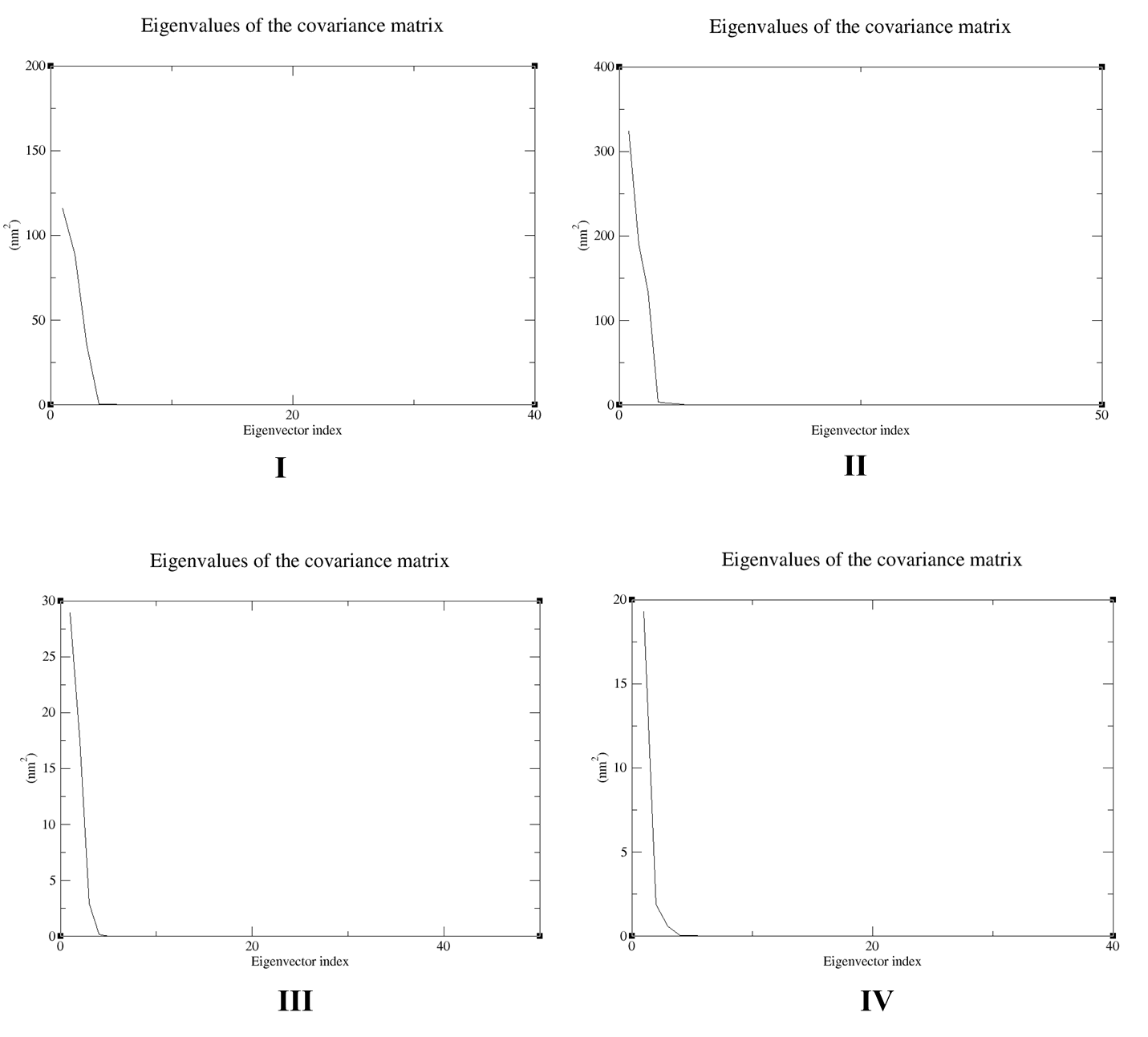
A diagram of a graph

Description automatically generated

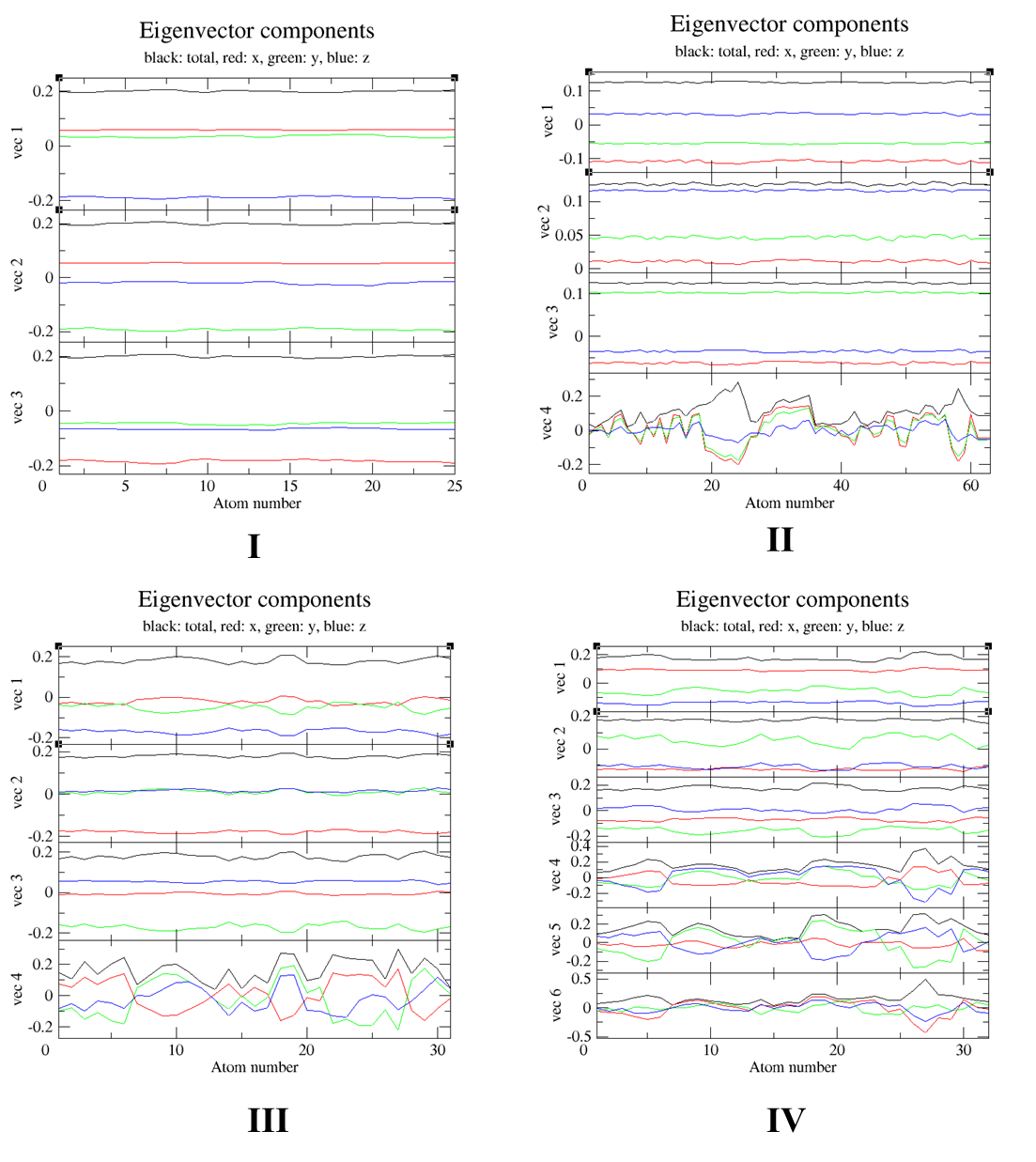
Figure 5: Number of Hydrogen Bond of 4QMX\_D1 (Black color), 4QMX \_Hit1 (Red), 4QMX\_Hit7 (Green), 4QMX\_Hit8 (Blue)

**Principle Component Analysis (PCA)**

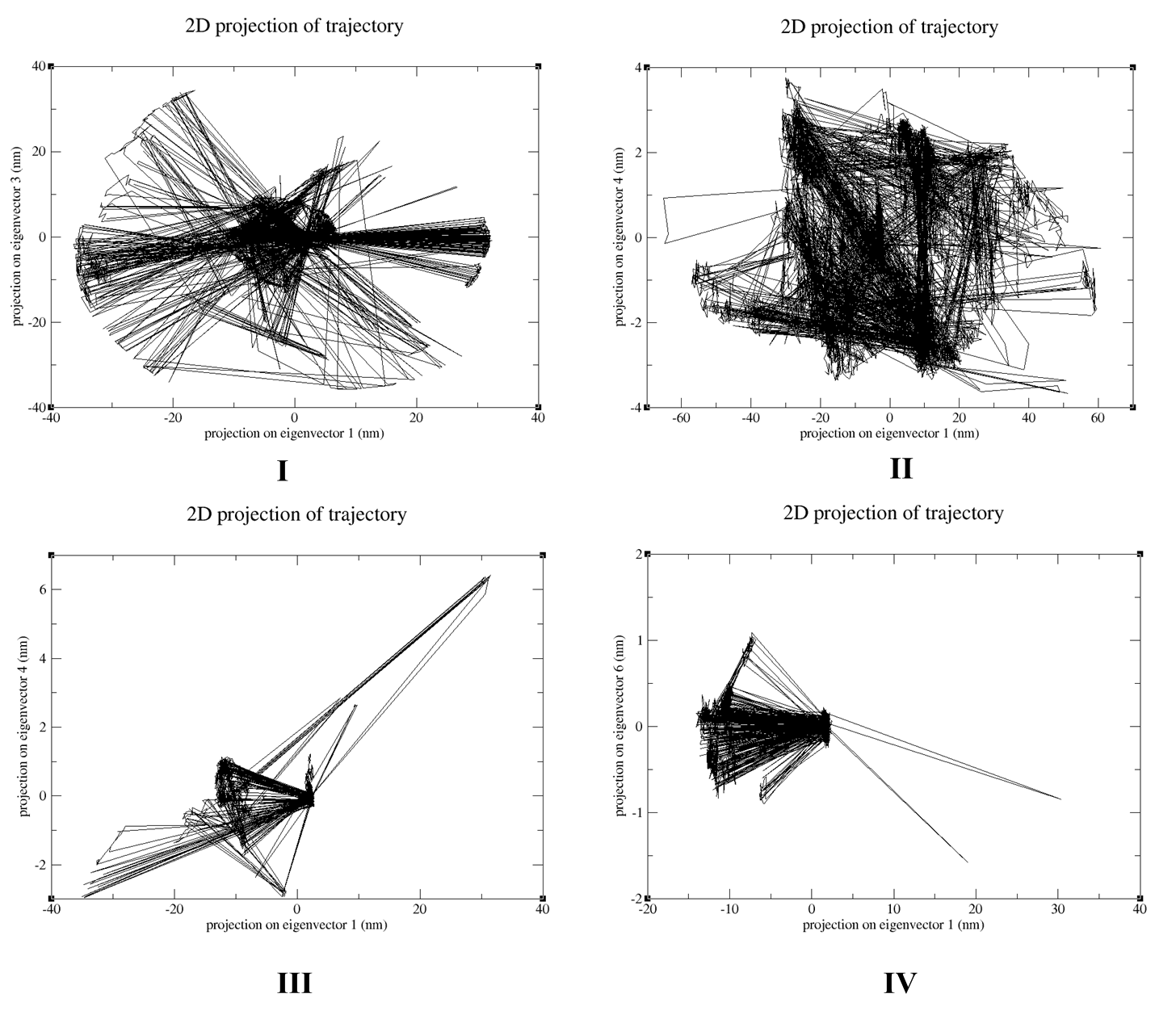
Throughout the simulation, PCA was used to evaluate the linked motions. Both eigenvalues and eigenvectors were determined as projections of the protiens’ C-alpha atoms, as shown in **Figure 6 and Figure 7**. Eigenvectors define motion orientation, whereas eigenvalues represent motion magnitude. The projected motions of protein when complexed with the ligand express high similarity, as shown by **Figure 8**. The eigenvector suggests that when the ligand attaches to the protein, the protein act similarly, creating persistent clusters and occupying the same subspace, indicating the stability of the created complex, as shown in Eigenrmsf value in **Figure 9**. Individual examination of the correlated motion of the first four eigenvectors (which account for 85% of the variations) reveals that the ligand showed moderate relative motion with residues, as shown in **Figure 9.**

****

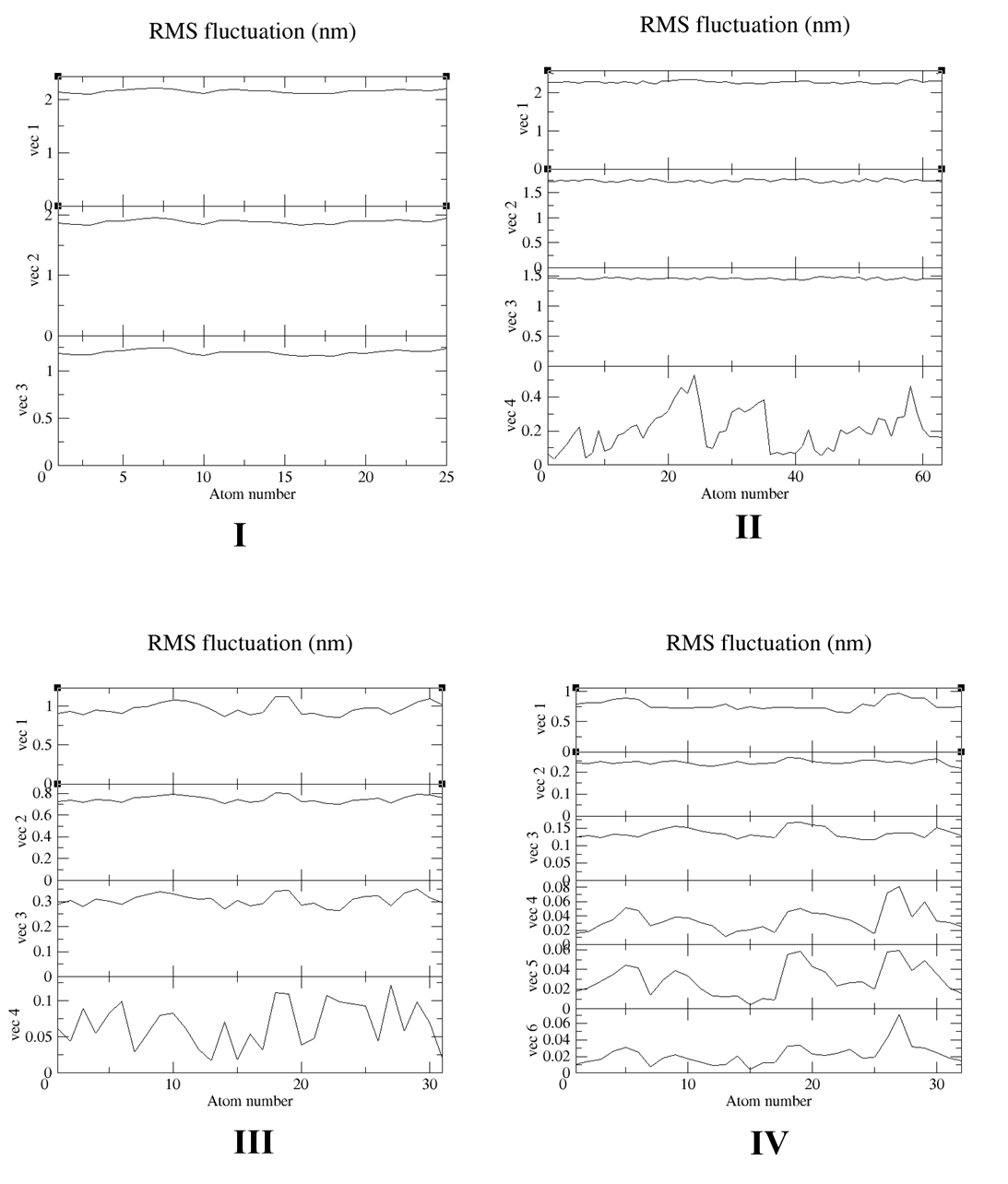
**Figure 6:** Eigenvalue of the covariant matrix of **(I) 4QMX\_D1 (II) 4QMX\_Hit1 (III) 4QMX\_Hit7 (IV) 4QMX\_Hit8**



**Figure 7:** Eigenvector components of **(I) 4QMX\_D1 (II) 4QMX\_Hit1 (III) 4QMX\_Hit7 (IV) 4QMX\_Hit8**



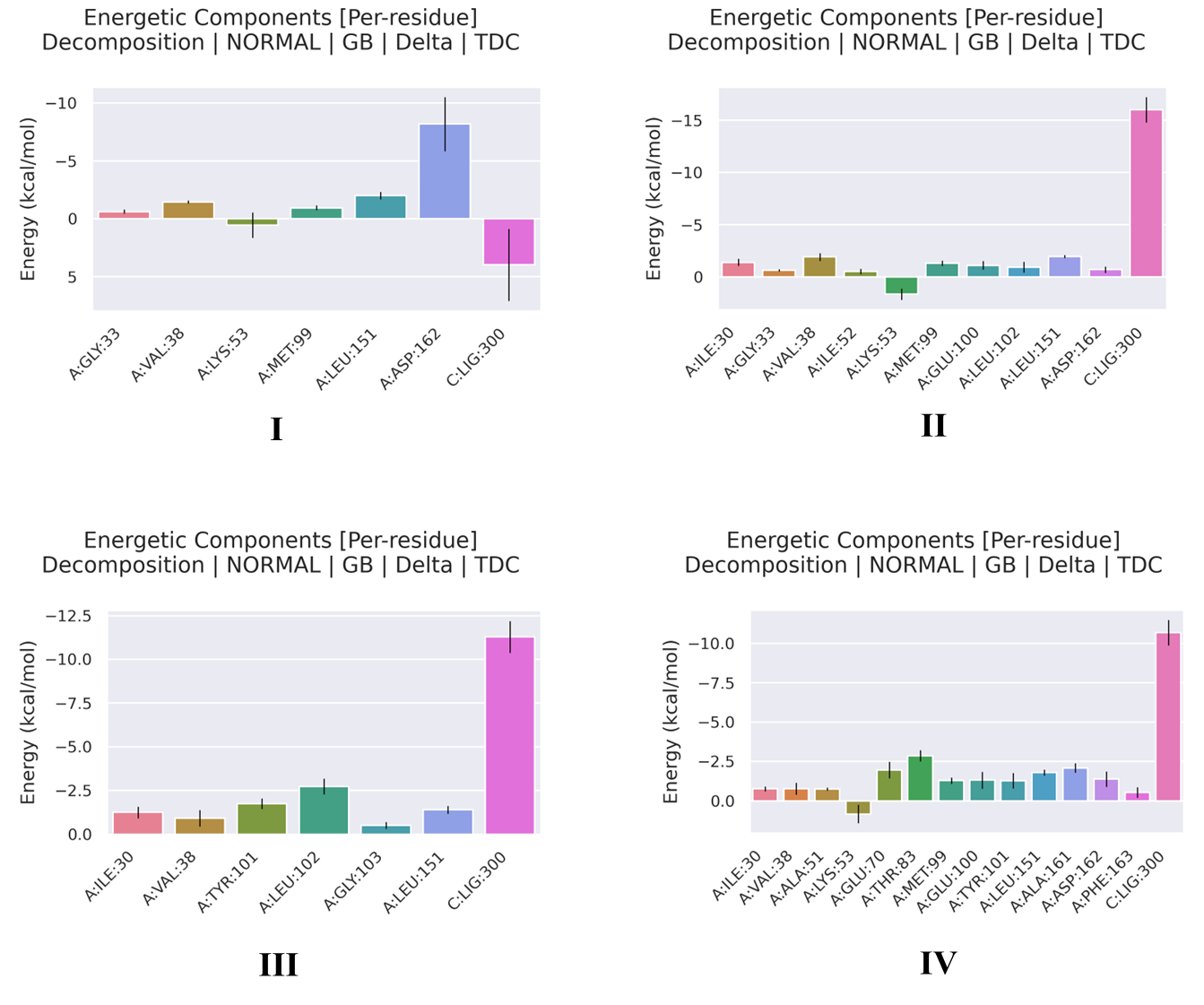
**Figure 8:** 2D projected motion of **(I) 4QMX\_D1 (II) 4QMX\_Hit1 (III) 4QMX\_Hit7 (IV) 4QMX\_Hit8**



**Figure 9:** Individual examination of the correlated motion of the first four eigenvectors of ligand showing moderate relative motion with residues of **(I) 4QMX\_D1 (II) 4QMX\_Hit1 (III) 4QMX\_Hit7 (IV) 4QMX\_Hit8**

**Total Decomposition Contribution**

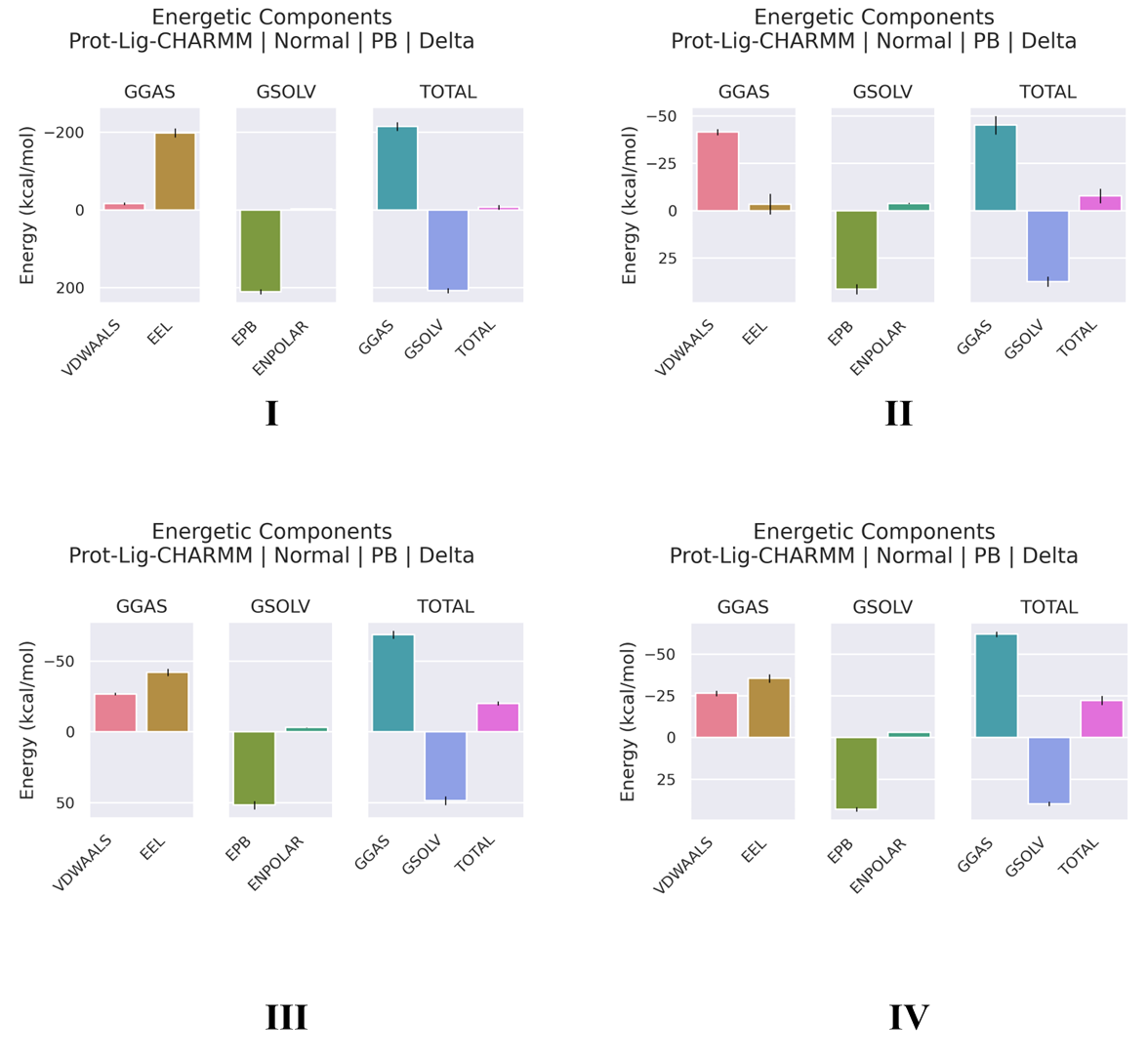
In molecular dynamics simulations, there are energetic contributions of different components (e.g., residues, ligands) to the overall binding of free energy. Energy decomposition analysis helps dissect the total energy into individual terms, such as van der Waals interactions, electrostatic interactions, and solvation effects. The TDC represents the contribution of each residue (or other components) to the overall binding free energy.It combines various energy terms (van der Waals, electrostatic, polar solvation, non-polar solvation) associated with a specific residue. The TDC can be calculated for the complex, receptor, ligand, or their differences (delta), as shown in **Figure 11**.



**Figure 10:** Total decomposition contribution of **(I) 4QMX\_D1 (II) 4QMX\_Hit1 (III) 4QMX\_Hit7 (IV) 4QMX\_Hit8**

**molecular mechanics-Poisson Boltzmann surface area (MMPBSA)**

Molecular Mechanics Poisson-Boltzmann Surface Area (MMPBSA) is one of the most often used methods for calculating binding free energy, as shown in **Figure 12**. The lower the predicted binding free energy of a ligand-protein complex, the more stable that complex is expected to be and the higher the ligand’s activity and potency. The energy values of the ligand produced consistent values, which are shown in **Table 2.**



**Figure 11: MMPBSA energetic components (I) 4QMX\_D1 (II) 4QMX\_Hit1 (III) 4QMX\_Hit7 (IV) 4QMX\_Hit8**

Table 2: MMPBSA calculations of protein-ligand complex

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | D1 | Hit1 | Hit7 | Hit8 |
| ΔEVDWAALS | -16.19±3.37 | -41.64±1.61 | -26.78±0.89 | -26.53±1.69 |
| ΔEEL | -198.22±11.28 | -3.61±5.68 | -41.92±2.55 | -35.49±2.52 |
| ΔEPB | 210.27±6.62 | 41.33±2.62 | 51.62±2.94 | 43.02±1.3 |
| ΔENPOLAR | -2.48±0.05 | -3.99±0.25 | -3.08±0.08 | -3.22±0.01 |
| ΔGSOLV | 207.8±6.66 | -37.34±2.6 | 48.55±3.01 | 39.8±1.3 |
| ΔGGAS | -214.4±11.07 | -45.5±4.9 | -68.7±2.84 | -62.02±1.66 |
| TOTAL | -22.22±2.73 | -7.91±3.74 | -20.15±1.48 | -22.22±1.37 |

\* EVDWAALS (Vander Waals), EEL (Electrostatic energy), EPB (Polar solvation energy), ENPOLAR (Non-polar solvation energy), GSOLV (Total solvation free energy), GGAS (Total gas phasefreeenergy), and Total