Characterization of the binding pockets of TgCDPK1 protein

Natalia V. Karimova

1. Structure Selection

This study's target protein is *Toxoplasma Gondii* calcium-dependent protein kinase 1 (TgCDPK1) protein. The protein structure was downloaded from the Protein Data Bank (PDB) with an ID 6BFA (it came with the inhibitor UW1553). The structure 6BFA was selected because it was received with a good resolution (2.80 Å) and has fewer missing fragments and no iso-peptide bonds [1]. This structure has less than 1% of missing atoms in the structure: 20 missing residues and five incomplete sidechains (**Fig.1**). The missing residues are related to the "broken end" (the same fragment is missing in all other structures presented in the RCSB PDB Databank).

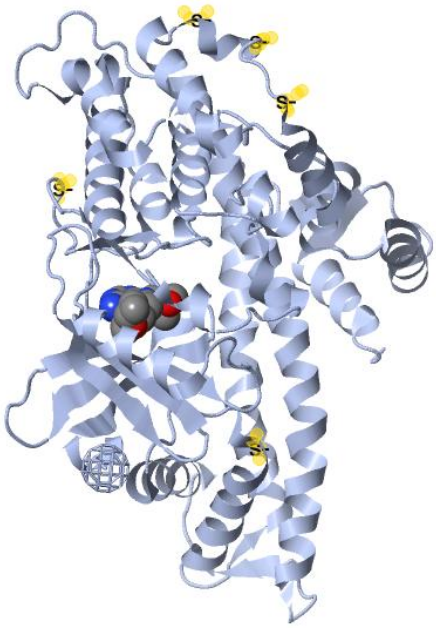

Figure 1. TgCDPK1 protein with inhibitor UW1553 (structure 6BFA). An empty spherical bucket is used to mark the place of missing residues. The incomplete sidechains (with missing atoms) are marked with yellow halos (to visualize the missing parts, the tool "FirstGlance in Jmol" was used [2].

2. Pocket Detection and Characterization

The protein pockets were identified using FpocketWeb application [3, 4]. The analysis identified 42 possible pockets, with total scores ranging from 0.211 to -0.115, indicating a range of potential

binding sites from highly favorable to less probable. Based on comparative analysis, it was found that Pocket 1 is the most promising binding site (**Fig. 2A**):

(i) **Volume**: This pocket has the largest volume, with a high alpha sphere density, indicating a well-defined shape. The max distance of 21.922 Å suggests that the pocket has an elongated or dispersed structure.

(ii) **Polarity**: The pocket is moderately polar, as 40.278% of its atoms are polar. The polarity score of 10 suggests a strong potential for hydrophilic interactions, making it suitable for binding polar ligands. Additionally, the charge score of -3 implies a negative electrostatic environment, which might favor positively charged ligands.

(iii) **Hydrophobicity**: The hydrophobicity score of 41.310 and apolar Solvent Accessible Surface Area (SASA) of 170.830 Å² indicate a significant hydrophobic character. The pocket has a good balance between polar and apolar regions, making it potentially versatile for binding hydrophilic and hydrophobic ligands.

Additionally, the most important is that Pocket 1 has the highest druggability score (0.945), suggesting it is the most viable target for ligand binding.

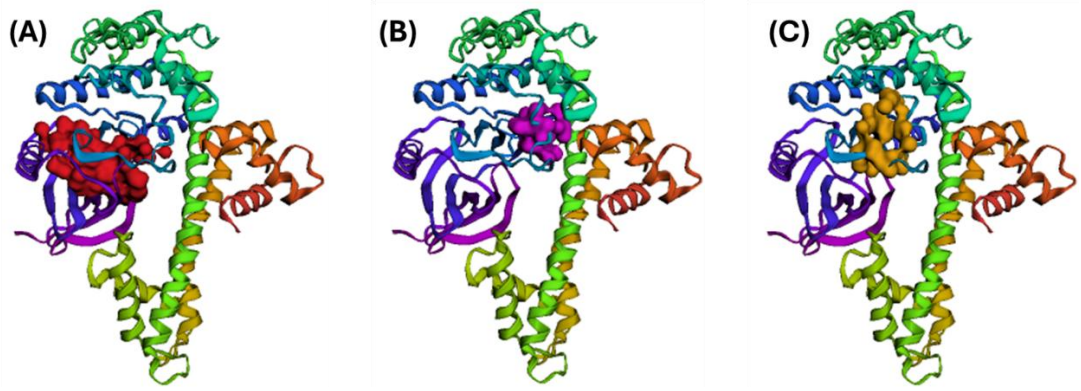

Figure 2. Visualization of the top three identified pockets: (**A**) Pocket 1 (red), (**B**) Pocket 2 (purple), and (**C**) Pocket 3 (orange). Images obtained using FpocketWeb.

Table 1 presents a detailed comparison of the calculated parameters for the three highest-scoring pockets, highlighting the distinct characteristics that set Pocket 1 apart. This comparison underscores the significant differences in volume, polarity, hydrophobicity, and druggability, demonstrating why Pocket 1 is the most favorable binding site. These top three pockets were visualized and are presented in **Fig. 2**.

Table 1. Fpocket parameters for the first three pockets with higher Score.

Parameter	Pocket 1	Pocket 2	Pocket 3	Pocket 1(Eq)
Score	0.211	0.148	0.136	0.474
Druggability	0.945	0.001	0.067	0.720
Volume, (Å3)	1026.9	214.6	414.6	2658.394
Volume Score	4.172	4.615	4.385	4.175
Alpha Sphere Density	7.926	2.813	4.94	13.941
Center of Mass, Alpha Sphere Max Distance (Å)	21.922	6.79	11.203	34.979
Polar SASA, (Å2)	120.158	23.638	41.986	369.449
Polarity Score	10	8	9	34
Proportion of Polar Atoms (%)	40.278	42.857	40.541	41.618
Charge Score	-3	0	1	2
Apolar SASA, (Å2)	170.830	36.229	88.157	351.242
Apolar Alpha Sphere Proportion	0.536	0.07	0.367	0.431
Mean Local Hydrophobic Density	41.821	2	19.273	32.283
Hydrophobicity Score	41.310	13.692	17.923	21.603

3. Molecular Dynamics (MD) Simulations

3.1.Protein Preparation

The selected structure (6BFA) contained a missing terminal residue ("broken end") and missing atoms in some side chains. The "broken end" residues were not filled, as they are often unresolved due to high flexibility in experimental structures. A preliminary test confirmed that the absence of terminal residues does not affect the stability of this protein. Additionally, since these regions are unlikely to influence pocket functionality and are not essential for pocket analysis, their reconstruction will be omitted in this trial case.
The solvent molecules and ligands were removed, and the missing side chain atoms were restored using Chimera software [5].

3.2.Computational Methodology

The GROMACS 2022.6 software package [6] was employed to perform all simulations, following a well-established workflow for system preparation, equilibration, and production MD. The OPLS-AA/L all-atom force field [7]
was employed to model molecular interactions, while the SPC/E water model [8]
was used to represent the solvent environment.
Step 1 (Topology generation): The system's molecular topology was generated using GROMACS' pdb2gmx utility with the appropriate force field. Hydrogens were added according to the default protonation state at physiological pH.
Step 2 (System Solvation): The system was placed in a cubic simulation box with a minimum solute-box distance of 1.0 nm. The SPC216 water model [9] was used to solvate the protein in an explicit solvent environment.

Step 3 (System Ionization): To neutralize the system and achieve physiological ionic strength, Na^+ and Cl^- ions were added to the solvated structure. First, energy parameters were preprocessed (grompp), generating a portable binary run input (ions.tpr). Then, ions were added using genion, replacing water molecules (SOL) to maintain charge neutrality. The results indicated that only Na^+ ions were required to neutralize the system, suggesting an initial net negative charge in the solvated structure.

Step 4 (Energy Minimization): To remove unfavorable steric clashes and optimize the system's initial geometry, energy minimization was performed until the maximum force was below the convergence threshold.

Step 5 (Equilibration):

(a) NVT Equilibration - The canonical ensemble (NVT) equilibration was performed to stabilize the system at a constant temperature of 300 K using the velocity-rescaling thermostat (V-rescale). A position restraint was applied to the protein to allow the solvent to equilibrate around it. The leap-frog integrator was used with a 2 fs time step for a total of 100 ps (50,000 steps). Electrostatic interactions were handled using Particle Mesh Ewald (PME) with a 1.0 nm cutoff, while van der Waals interactions were truncated at 1.0 nm with a dispersion correction applied for energy and pressure. The system was divided into two coupling groups (Protein and Non-Protein) to improve accuracy in temperature regulation. Periodic boundary conditions were applied in all three dimensions, and initial velocities were assigned from a Maxwell distribution at 300 K to initiate the equilibration process.

(b) NPT Equilibration - After NVT equilibration, the system underwent NPT equilibration to stabilize pressure at 1.0 bar using the Parrinello-Rahman barostat with isotropic coupling. The V-rescale thermostat maintained the temperature at 300 K, with separate coupling for Protein and Non-Protein groups. A 2 fs time step was used for 100 ps (50,000 steps), allowing solvent density relaxation. PME handled electrostatics, while van der Waals interactions were truncated at 1.0 nm with dispersion correction. LINCS constraints were applied to hydrogen bonds, and periodic boundary conditions ensured system stability. Since NPT followed NVT, no new velocities were assigned (gen_vel = no). This step ensured proper density and pressure equilibration before production MD.

Step 6 (Production Molecular Dynamics (MD) Simulation): A 100 ns production MD simulation was performed using the leap-frog integrator with a 2 fs time step. The system continued from the NPT-equilibrated state, with LINCS constraints applied to hydrogen bonds to ensure stability. Electrostatics were handled using PME with a 1.0 nm cutoff, while van der Waals interactions were truncated at the same distance. The V-rescale thermostat maintained 300 K for Protein and Non-Protein groups, and the Parrinello-Rahman barostat regulated pressure at 1.0 bar with isotropic coupling. To optimize storage, compressed coordinates (.xtc) were saved every 10 ps, while energy and log files were updated at the same interval. Only one trajectory was generated.

Notes: (a) Steps 1 to 5 were executed exclusively on CPU processors, while Step 6 (production MD) was conducted on a hybrid GPU/CPU system to enhance computational efficiency. (b) All GROMACS commands used for these calculations can be found in **Appendix 1**.

To analyze the trajectory, the following parameters were computed: Root Mean Square Deviation (RMSD), Solvent Accessible Surface Area (SASA), Radius of Gyration (Rg), Pocket Volume, and Root Mean Square Fluctuation (RMSF).

3.3. Results

The analysis of Pocket 1 across different stages—its presence in the crystal structure, post-equilibration (Steps 1-5), and during the MD production trajectory—provides valuable insights into the effects of ligand removal on protein structure and pocket stability. Initially, the X-ray structure contained a bound ligand in Pocket 1 (**Fig.1**), which was removed before the protein equilibration. Before molecular dynamics (MD) simulations, the ligand was removed, and the protein was subjected to preliminary equilibration through short NVT and NPT trajectories. At the start of the MD trajectory (T = 0 ns), Pocket 1 exhibited a significantly larger volume than its original state in the X-ray structure (**Fig. 3A** and **Tables 1, 2**), suggesting an initial relaxation and expansion in the absence of the ligand.

Table 2. Fpocket parameters for the Pocket 1 of protein structures obtained at t=0 and 100 ns MD trajectory.

	T = 0 ns	**T=50 ns**	
Parameter	**Pocket 1**	**Pocket 1a**	**Pocket 1b**
Score	0.474	0.083	0.066
Druggability	0.720	0.445	0.511
Volume, (Å3)	2658.394	1661.759	336.749
Volume Score	4.175	4.304	4.917
Alpha Sphere Density	13.941	10.792	3.774
Center of Mass, Alpha Sphere Max Distance (Å)	34.979	24.408	9.067
Polar SASA, (Å2)	369.449	278.074	24.688
Polarity Score	34	20	6
Proportion of Polar Atoms (%)	41.618	43.902	20
Charge Score	2	1	2
Apolar SASA, (Å2)	351.242	233.389	64.004
Apolar Alpha Sphere Proportion	0.431	0.343	0.759
Mean Local Hydrophobic Density	32.283	25.59	39.122
Hydrophobicity Score	21.603	32.196	37.833

Note: All pockets obtained for structure at 100 ns have draggability close to zero.

After 50 ns, Pocket 1 was split into two smaller pockets: Pocket 1a and 1b, see **Table 2** and **Fig.3B**).

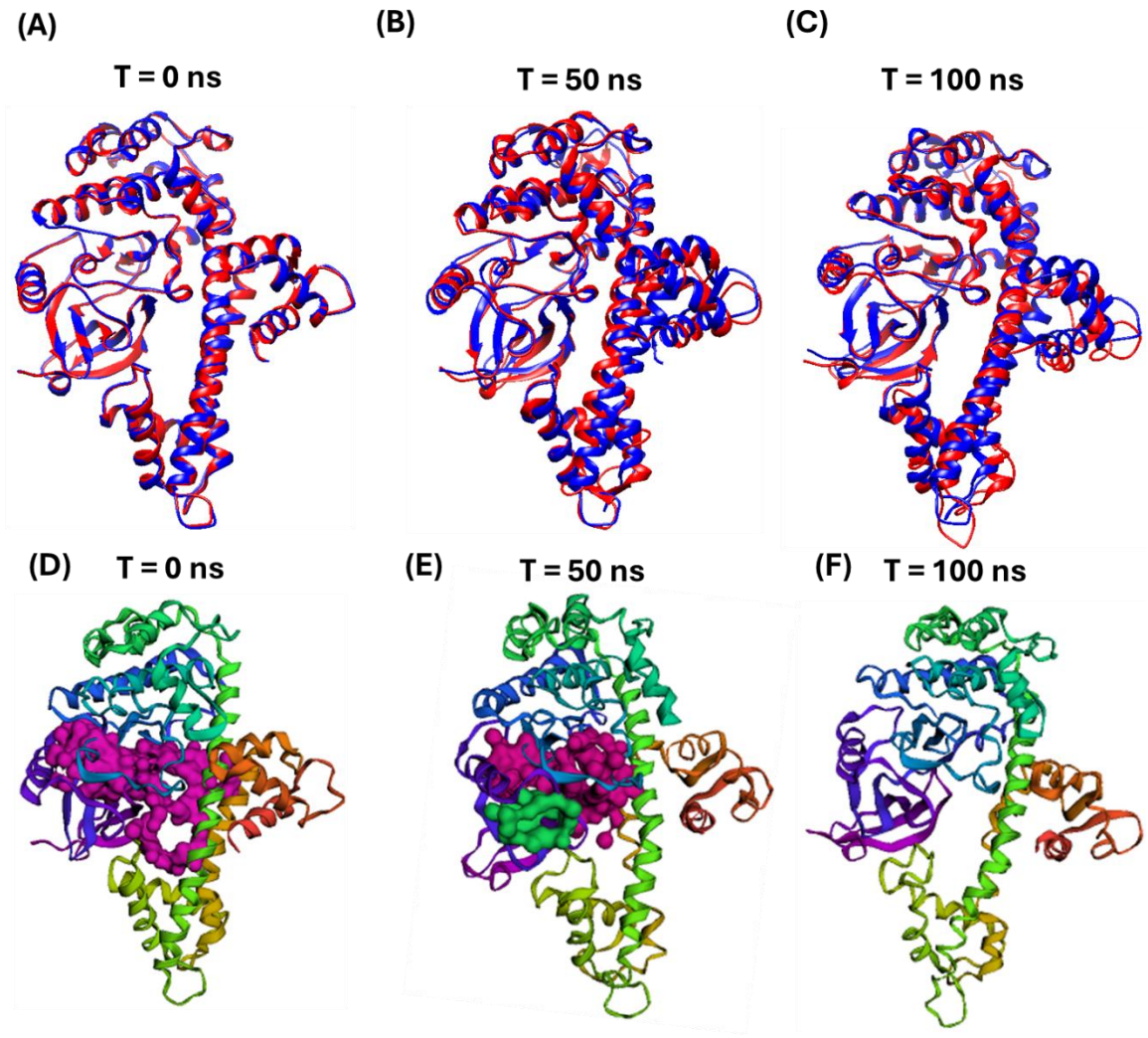

Figure 3. Comparison of X-ray and MD Structures: Conformational Changes: (A) at the beginning of MD trajectory right after equilibration and (B) at the end of MD trajectory at t=100 ns. Pocket 1 position: (C) at t=0 ns, and (D) at t=100 ns (FpocketWeb). All pockets obtained for structure at 100 ns have draggability close to zero.

At T = 50 ns, Pocket 1 gradually split into two smaller pockets. The distance from the center of the original pocket to its outer edge decreased from 34.979 Å (Pocket 1) to 24.408 Å (Pocket 1a) and 9.067 Å (Pocket 1b), indicating that the binding site was becoming more compact and fragmented. This suggests that the ligand helped stabilize and keep the pocket open in the original X-ray structure. However, after the ligand was removed, by T = 50 ns, the pocket reorganized and split into two smaller cavities, making it less accessible for new molecules to bind. Additionally, increased hydrophobicity and decreased druggability suggest that the pocket's environment became less favorable for ligand binding as the simulation progressed.

By the end of the trajectory (at 100 ns), no pockets could accommodate the original ligand, and druggability scores were close to zero for all obtained pockets. This suggests that the binding site collapsed or became highly unstable in the absence of the ligand, further confirming the significant structural rearrangement observed throughout the simulation.

To analyze the trajectory, the following parameters were computed: Root Mean Square Deviation (RMSD), Solvent Accessible Surface Area (SASA), Radius of Gyration (Rg), Pocket Volume (MDPocket), and Root Mean Square Fluctuation (RMSF).

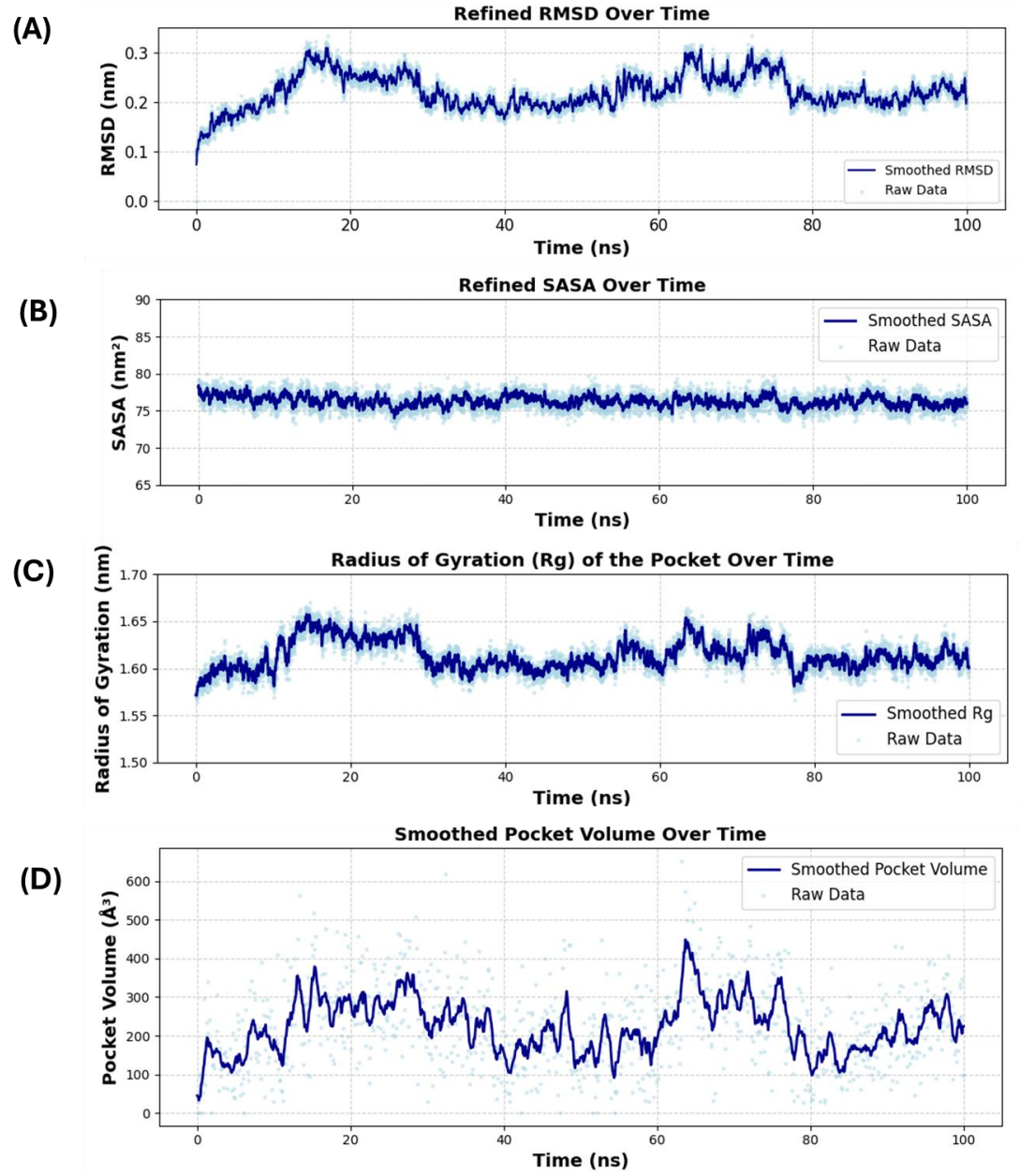

Figure 4. Structural and dynamic properties of the pocket over time: (**A**) RMSD, (**B**) SASA, (**C**) Radius of Gyration, and (**D**) Pocket Volume (MDPocket).

The results show that the pocket changed shape significantly after removing the ligand. At the beginning of the simulation, the pocket was more open and flexible, as seen in the high fluctuations in RMSD and Radius of Gyration (Rg) during the first 20 ns. Over time, the pocket adjusted to the absence of the ligand, becoming more stable but still shifting its structure (**Fig. 4A,C**).

The Solvent-Accessible Surface Area (SASA) mainly stayed the same, meaning the pocket did not fully close or become completely exposed to the solvent. However, the pocket volume decreased and fluctuated significantly in the first 50 ns, before settling into a smaller and more compact form. This suggests that the pocket tried to adjust but eventually shrunk and became less accessible (**Fig. 4B**).

As the pocket became smaller and more rigid, its ligand binding ability decreased. The higher local hydrophobic density and reduced druggability scores indicate that the pocket became tighter and less suitable for binding. The final pocket volume was much smaller than at the start, showing that removing the ligand partially caused the pocket to collapse or split into smaller cavities (**Fig. 4D**). Additionally, the analysis of RMSF for pocket residues reveals that residues are primarily located within loop regions and beta-sheets of the protein structure (**Fig.5**).

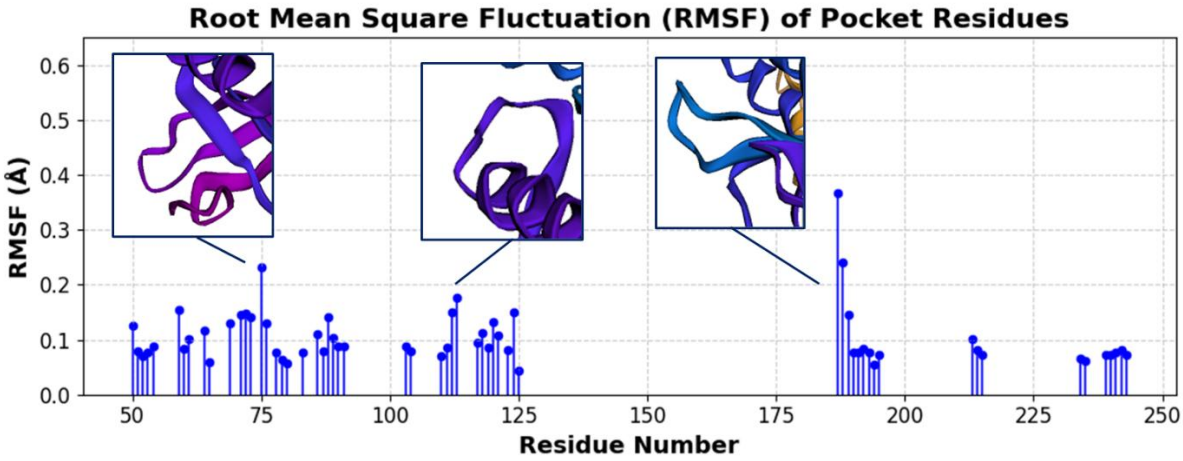

Figure 5. Root Mean Square Fluctuation (RMSF) Analysis of Pocket Residues.

Conclusions:

The pocket showed instability throughout the simulation, with significant changes in volume, RMSD, RMSF, and druggability. Several factors could explain this:

(i) Short Trajectory—The simulation may be too short to determine whether the pocket stabilizes over time.
(ii) Single Trajectory – One simulation may not be enough to understand pocket flexibility fully. Running more simulations could give a clearer picture.

(iii) Broken Terminus – A nearby broken end in the structure might cause unwanted movement, making the pocket unstable.
(iv) Force Field and Solvent Effects – The simulation settings, including the force field and water model, could affect pocket stability, especially if the pocket is hydrophobic.

Future Directions:

The combination of these factors suggests that the pocket is not well-defined without the ligand. Further simulations with longer timescales, additional replicas, and structure refinement (such as fixing terminal residues) would be necessary to confirm whether this instability is a genuine feature or a simulation artifact.

References:

[1] https://doi.org/10.1021/acs.jmedchem.6b00760

[2] Tool "FirstGlance in Jmol" / https://www.bioinformatics.org/firstglance/fgij/.

[3] https://doi.org/10.1186/s13321-022-00637-0

[4] FpocketWeb https://durrantlab.pitt.edu/fpocketweb/

[5] Chimer https://www.cgl.ucsf.edu/chimera/

APPENDIX – 1

Note: Steps 1 to 5 were executed exclusively on CPU processors, while Step 6 (production MD) was conducted on a hybrid GPU/CPU system to enhance computational efficiency.

GROMACS Commands (CPU):

Step 1 - Generate Topology

echo 15 | gmx_mpi pdb2gmx -f 6bfa_noWater_noLigand_fullSideChains_Chimera.pdb -o test_str_processed.gro -water spce

Step 2 - Solvation

Step 2.1: Creating simulation box.
gmx_mpi editconf -f test_str_processed.gro -o str_newbox.gro -c -d 1.0 -bt cubic

Step 2.2: Adding water molecules.
gmx_mpi solvate -cp str_newbox.gro -cs spc216.gro -o str_solv.gro -p topol.top

Step 3 - Ionization

Step 3.1: Preparing system for ion addition.
gmx_mpi grompp -f ions.mdp -c str_solv.gro -p topol.top -o ions.tpr

Step 3.2: Adding ions to neutralize the system. (Selection 13 = SOL)
gmx_mpi genion -s ions.tpr -o str_solv_ions.gro -p topol.top -pname NA -nname CL -neutral

Step 4 - Energy Minimization

Step 4.1: Preparing input for energy minimization.
gmx_mpi grompp -f minim.mdp -c str_solv_ions.gro -p topol.top -o em.tpr

Step 4.2: Running energy minimization.
srun gmx_mpi mdrun -v -deffnm em

Step 5 - Equilibration (NVT & NPT)

Step 5.1: Running NVT equilibration.
gmx_mpi grompp -f nvt.mdp -c em.gro -r em.gro -p topol.top -po mdout.mdp -o nvt.tpr
srun gmx_mpi mdrun -deffnm nvt -v

Step 5.2: Extracting total energy from NVT.
cho -e "16\n0" | gmx_mpi energy -f nvt.edr -o temperature.xvg

Step 5.3: Running NPT equilibration.
gmx_mpi grompp -f npt.mdp -c nvt.gro -r nvt.gro -t nvt.cpt -p topol.top -o npt.tpr
srun gmx_mpi mdrun -deffnm npt -v
Step 5.4: Extracting density from NPT.

echo -e "24\n0" | gmx_mpi energy -f npt.edr -o pressure.xvg

Step 6 - Final MD (100 ns)

gmx_mpi grompp -f md.mdp -c npt.gro -t npt.cpt -p topol.top -o md_0_1.tpr

GROMACS Command (GPU/CPU)

srun gmx_mpi mdrun -nb gpu -pme cpu -bonded cpu -deffnm md_0_1 -v

APPENDIX – 3

MD Analysis

Step-1: Bringing Protein in the center of box:

gmx trjconv -s md_0_1.tpr -f md_0_1.xtc -o md_0_1_noPBC.xtc -pbc mol -center

Step-2: Calculation of RMSD

gmx rms -s md_0_1.tpr -f md_0_1_noPBC.xtc -o rmsd.xvg -tu ns

Step-3: Calculation of RMSF

gmx rmsf -s md_0_1.tpr -f md_0_1_noPBC.xtc -o rmsf.xvg -res

Step-4: Calculation of Radius of Gyration

gmx gyrate -s md_0_1.tpr -f md_0_1_noPBC.xtc -o gyrate.xvg

Step-5: Calculation of Total Number of Hydrogen bonds

gmx hbond -s md_0_1.tpr -f md_0_1_noPBC.xtc -num hydrogen-bonds.xvg

Step-6: Calculation of Total Solvent Accessible Surface Area

gmx sasa -s md_0_1.tpr -f md_0_1_noPBC.xtc -o area.xvg -tu ns