

Supporting Information

Steered molecular dynamics simulations for studying protein-ligand interaction in cyclin-dependent kinase 5

Jagdish Suresh Patel, Anna Berteotti, Simone Ronsisvalle,

Walter Rocchia, Andrea Cavalli

Contents

Steered Molecular Dynamics (SMD) details	S12
Potential of Mean Force Calculation in SMD	S12
Calculation of average force profiles	S14
Details about the choice of the pulling velocities, spring constant, pulling length	S14
Ligand solvation as a function of SMD simulation time	S18
Calculation of total number of hydrogen bonds (H-bonds)	S19
Stability of the binding poses during the MD simulation	S113
References	S115

Steered Molecular Dynamics (SMD) details

In SMD simulations, a time-dependent external force is applied to the ligand to facilitate its unbinding from the protein, which usually cannot be achieved by standard MD simulation. In particular, in SMD the transition between two states, here the bound and unbound ones, is achieved by adding to the standard Hamiltonian a harmonic time-dependent potential $U(r,t)$ acting on a descriptor $s(r)$ (e.g., the protein-ligand distance), which holds the following time dependency:

$$U(\mathbf{r}, t) = \frac{k}{2} [s(\mathbf{r}) - s_0(t) - vt]^2 \quad (1)$$

where s_0 is the value of the descriptor in the initial state, t is the time, and k is a constant representing the strength of the applied force. After a predetermined amount of time, the harmonic constraint will be centered in its final position:

$$s_1 = s_0 + vt_1. \quad (2)$$

Therefore, at constant pulling velocity, if the spring constant k is large enough (stiff-spring regime), it is reasonable to assume that, at the final time t_1 , the system has approximately reached the point s_1 . During this transition, the value of the exerted force F is calculated using the following equation:

$$F(t) = -k[s(\mathbf{r}) - s_\lambda(t) - vt]. \quad (3)$$

The external work ΔW performed on the system was calculated by integrating the power along the entire transition time:

$$\Delta W = v \int_{t_0}^{t_1} F(t) dt. \quad (4)$$

Potential of Mean Force Calculation in SMD

Jarzynski^{1,2} developed an equality connecting the free energy difference between two states with a suitable average of the out-of-equilibrium work profiles obtained in SMD simulations. Indeed, he provided the connection between equilibrium free energy differences and the work done through many realizations of non-equilibrium processes:

$$e^{-\beta\Delta A} = \langle e^{-\beta W_{i,0 \rightarrow t}} \rangle, \quad \beta = \frac{1}{k_B T} \quad (5)$$

where T is the temperature of the system and k_B the Boltzmann constant. Here, the ensemble average is taken over a number of realizations of non-equilibrium work done on the system during pulling simulations. The Jarzynski equality thus provides an estimate for ΔA for a set of N work values given by

$$\Delta A_{Jarz} = -\frac{1}{\beta} \ln \left[\frac{1}{N} \sum_{i=1}^N e^{-\beta W_{i,0 \rightarrow t}} \right]. \quad (6)$$

In practice, its direct application is hampered by the limited number of collectable trajectories along with the complex nature of the biological systems, which often result in a standard deviation of the work (σ_w) several times higher than $k_B T$.³ The Jarzynski equality requires a weighted average of the work values with an exponential set of weights. The net effect of this averaging is that only trajectories with significantly low work, corresponding to the left tail of Gaussian work distribution, give a relevant contribution to the average.³ Moreover, SMD pulling trajectories often sample the region around the peak rather than the tails of the Gaussian work distribution. In the recent years, this statistical uncertainty has been tackled in various studies giving rise to more effective applications of the Jarzynski's equality.^{4,5} By assuming a Gaussian distribution for the work values, eq. 5 can be simplified using the second order cumulant expansion.⁴ This approach (eq. 7) has been widely applied⁴⁻⁷ for reconstructing the free energy profile of bio molecular processes. This was here used to compare the free energy profiles of all the ligands in this study.

$$\Delta A = \langle W \rangle - \frac{1}{2} \beta \sigma_w^2, \quad \sigma_w^2 = \langle W^2 \rangle - \langle W \rangle^2. \quad (7)$$

The above equation gives the free energy differences (ΔA) between two states from the mean work $\langle W \rangle$ over all the trajectories and subtracting the variance σ_w^2 contribution. The idea is that highly irreversible work trajectories are likely to have higher σ_w^2 so that, in theory, one can obtain the same ΔA value using above equation with different amounts of dissipated work (i.e., at different pulling velocities).

Calculation of average force profiles

The average force profiles along the time shown in the Figure 1 and 2 for each ligand were calculated in the following way:

$$\langle F(t) \rangle = \frac{1}{N} \sum_{i=1}^N F_i(t). \quad (8)$$

Details about the choice of the pulling velocities

The choice of the pulling velocities was based on a compromise between accuracy and speed, which is a classical tradeoff in drug design. In particular, four different pulling velocities have been tested: 0.005 Å/ps, 0.006 Å/ps, 0.007 Å/ps and 0.025 Å/ps, keeping all the other parameters identical. The ligand is outside the protein and fully solvated after 5 ns, 4 ns, 3 ns, and 1 ns, respectively. A comparison between the two regimes is reported in Figure S1, where the average force profiles are reported for compounds **1-6** for the four velocities. The slow regime using a velocity of 0.007 Å/ps was chosen because the correlation between the average force profile and the biological activity is better as compared to the fast regime (0.025 Å/ps) and it needs reasonable computer time. However, given the fact that our protocol was not able to discriminate different chemical compounds, see the force profile for the Series II, we were aware that a possible explanation for this failure could be related to the SMD pulling velocities, which should be very low to allow a proper reconstruction of accurate PMF profiles, which can then be quantitatively correlated with experimental data. This can be seen in the Figure S1 where the Force profiles as a function of simulation time are reported also for lower velocities. The main atomistic features of the unbinding process are conserved through the all 4 velocities regime, in fact the forces' profiles are very similar in all the 4 velocities regimes, see the position of the peaks. The ability of SMD to discriminate between molecules with very similar IC₅₀ values is increased with lower SMD pulling velocities. At lower velocities, see for example 0.006 and 0.005 Å/ps the decay of first force peak is very well distinguishable for molecules **1, 2, 3, 4**, while in case of fast velocity regime is not possible to distinguish between molecules **1, 2, 3, 4**. It is possible to envision that the use of very

low velocity will improve the ability of SMD, but in the compromise between simulation time and results, higher velocities regime can still be useful for drug design purpose. However, as mentioned, discrimination between binders and non-binders can still be achieved using high pulling velocities, which require simulation resources in line with the drug discovery requirements of speed and cost-effectiveness.

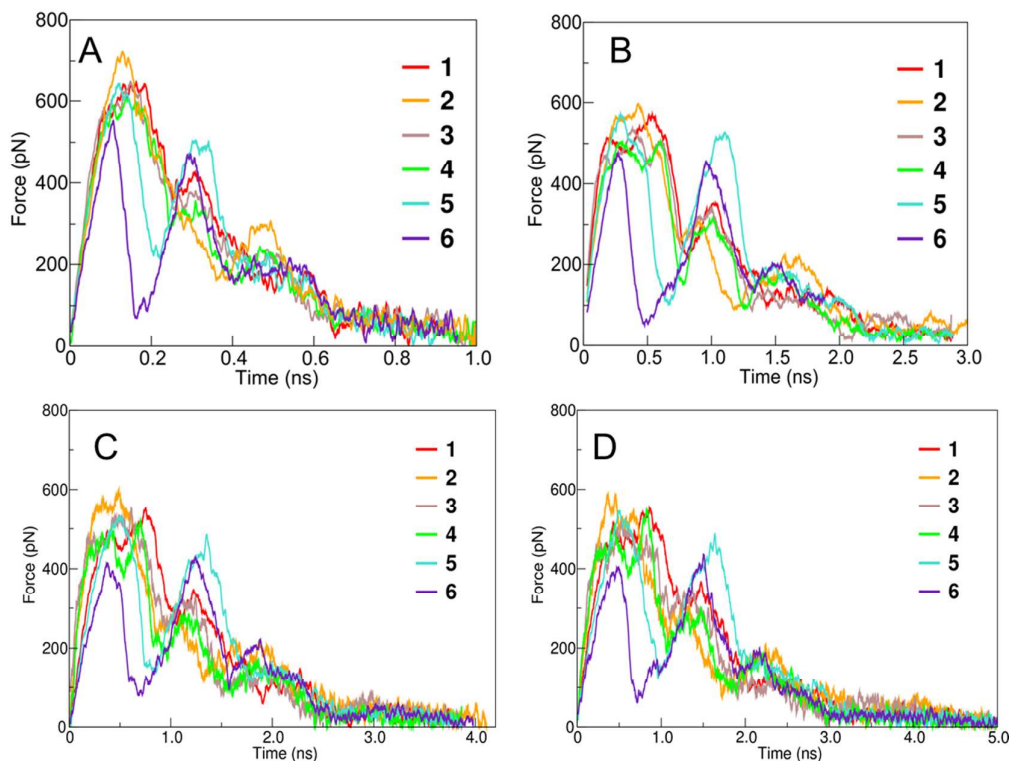


Figure S1. Average force profile as a function of time for compounds 1-6 in the four regimes: 0.025 Å/ps (A), 0.007 Å/ps (B), 0.006 Å/ps (C) and 0.005 Å/ps (D).

Details about the choice of the spring constant

The spring constant was chosen hard enough in order to steer the system to the target point (see Figure S2). From the plot in fact it can be seen that the distance closely follows the constraint center of the harmonic constraining potential in line with ref.^{4,8}.

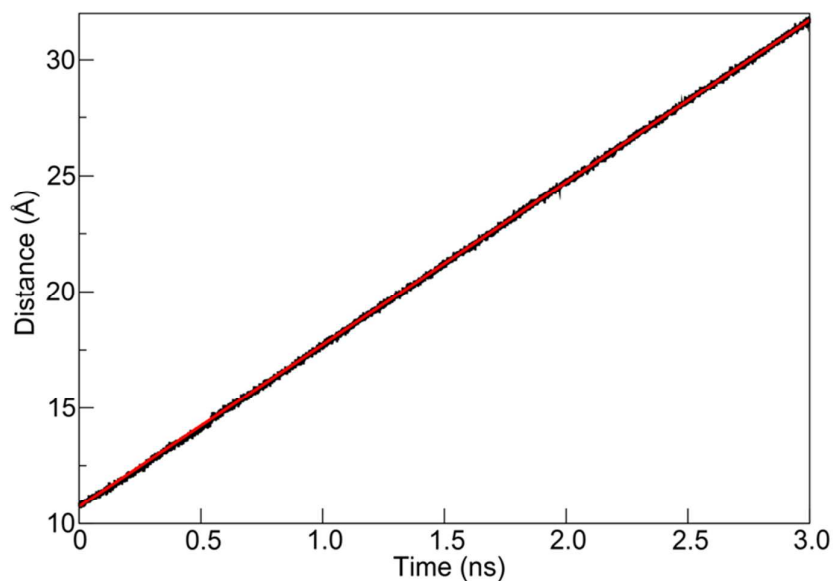


Figure S2. Line in black displays the behavior of the system along the SMD simulation of compound **2**, using a spring constant of $100 \text{ kcal}/(\text{\AA}^2 \cdot \text{mol})$. Red line represents the center of the moving harmonic constraining potential.

We also checked that the chosen force constant together with the adopted velocity regime was not able to disrupt the protein structure. In Figure S3, the RMSD of the $C\alpha$ of the binding pocket as a function of the SMD simulation time is plotted for compound **2**. The RMSD is calculated on the $C\alpha$ atoms belonging to the following residues: Ile10, Gly11, Glu12, Gly13, Phe80, Glu81, Phe82, Cys83, Asp84, Gln85, Asp86, Ala143 and Asp144. As it can be seen, there are no dramatic changes in the conformation of the binding pocket.

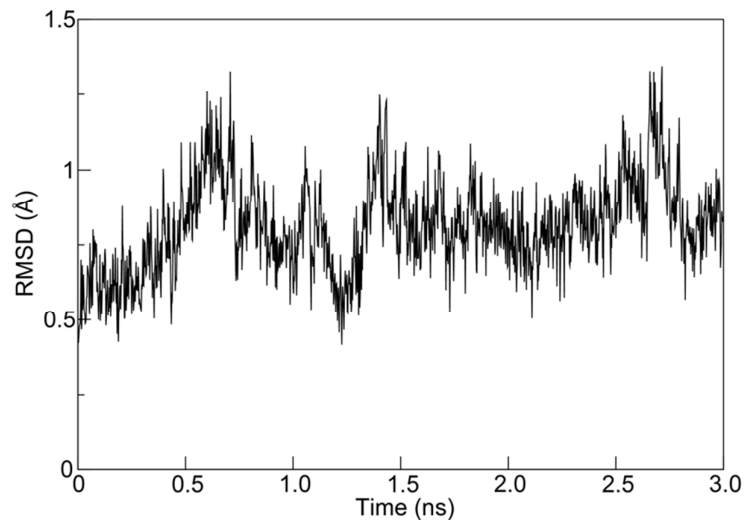


Figure S3. α atoms RMSD of the binding pocket during the SMD simulation run for compound **2**.

Details about the choice of the pulling length

The choice of the pulling length was based on the ligand solvation. Basically, from the solvation plot reported in Figure S2 it is reasonable to establish the total pulling length to 20 Å. At a pulling distance of 25 Å the ligand is completely out of the binding site and fully solvated.

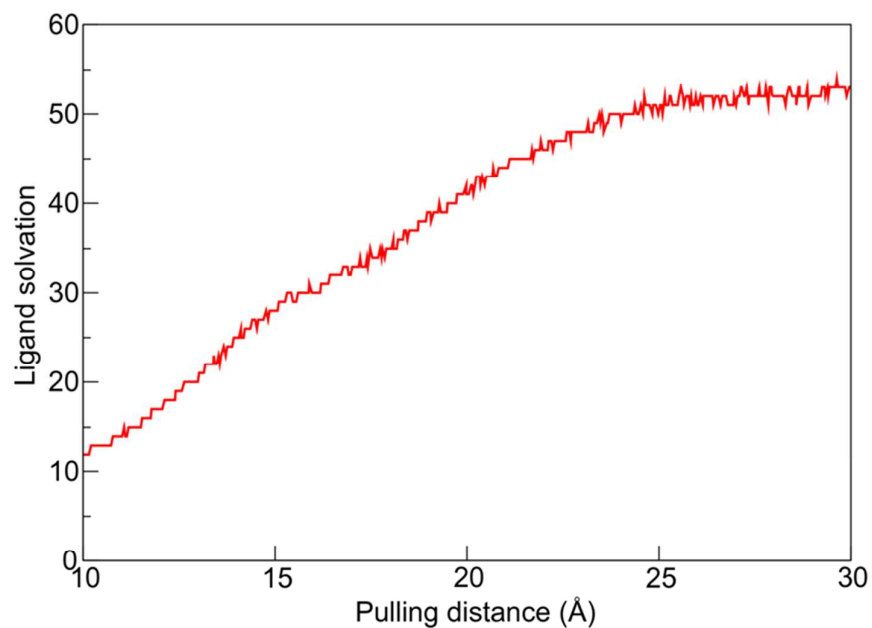


Figure S4. The solvation of compound **1** as a function of SMD pulling distance. At a pulling distance of 25 Å the ligand is fully solvated and the number of water molecules around the ligand after this point remains constant over time.

Details about the pulling distance and the cylindrical harmonic restraint

The pulling distance was defined between the center of mass of selected atoms of the ligand (see Table S1) and the center of mass of the following atoms of the protein: the C α atoms of Val30, Ala31, Leu78, the nitrogen backbone atoms of Lys33, Met185, the oxygen backbone atoms of Lys33, Leu78 and Asp184, and the carbon atom belonging to the carbonyl backbone group of Leu78 and Leu142.

As explained in the Methods section of the paper, we used a cylindrical harmonic restraint that hinders the interaction of the ligand with the surface of the protein while the former is out of the binding site. This restraint is implemented as the distance of a point to the axis of the cylinder. The point is the center of mass of selected atoms of the ligand while the axis connects the center of mass of two atom groups of the protein. One of these two groups is the same used to define the pulling distance for the ligand to be pulled out of the binding pocket. The second group is the center of mass of the following residues: the C α atoms of Ile10, Ala31, Gln85, Asp86, Leu87, Leu133, Asn135, Ile191, the nitrogen backbone atoms of Val30 and Asp86, the carbon atom belonging to the carbonyl backbone group of Lys9, Phe19, Val30, Asp84 and the oxygen atom of the carbonyl backbone group of Phe19.

Ligand solvation as a function of SMD simulation time

Here below we report the ligand **1** solvation as a function of SMD simulation time as representative case for all the 9 ligands. The ligand solvation has been calculated by means of ProDy⁹ tool through all the 50 SMD trajectories and the reported values are average values. The solvation is calculated as the number of water molecules around the ligand within a cut-off of 4 Å.

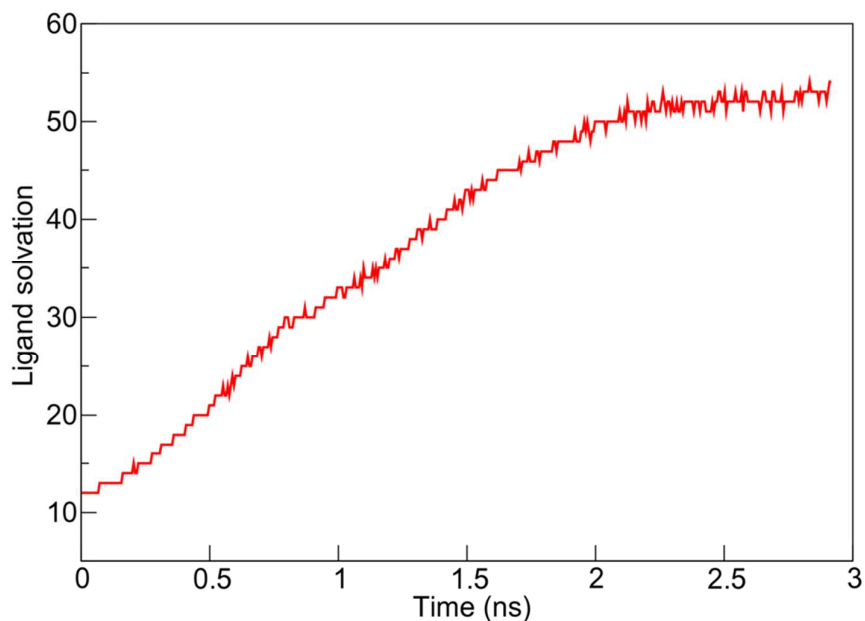


Figure S5. The solvation of compound **1** as a function of SMD simulation time. After 2 ns the ligand is completely solvated and the number of water molecules around the ligand after this time remains constant.

Calculation of total number of hydrogen bonds (H-bonds)

The total number of H-bonds shown in Figures 5 and 6 in the Results section are calculated as the sum of contacts between all the H-bond acceptor (A) and donor (D) atoms of the ligand and all the H-bond acceptor and donor atoms of the binding pocket, using the following equation:

$$R = \sum_{i \in D} \sum_{j \in A} R_{ij}. \quad (9)$$

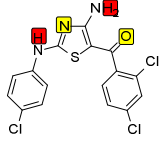
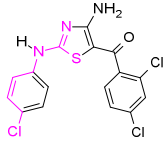
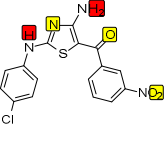
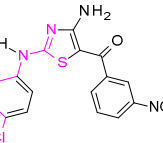
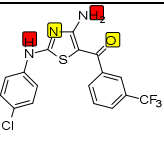
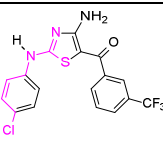
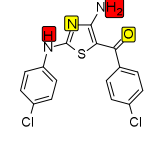
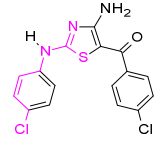
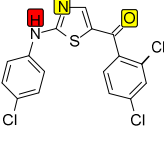
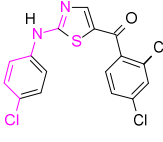
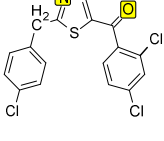
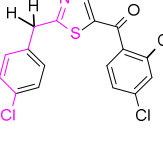
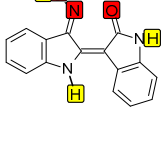
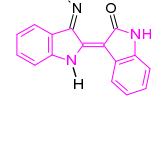
The sum is done over all the pairs of atoms with $i \in D$ and $j \in A$. The individual contributions R_{ij} are calculated using a switching function, which is given by:

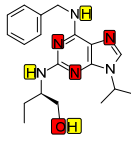
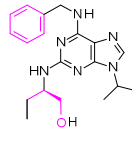
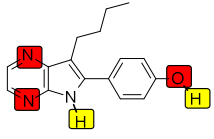
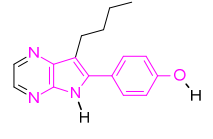
$$R_{ij} = \frac{1 - \left(\frac{r_{ij}}{r_0}\right)^p}{1 - \left(\frac{r_{ij}}{r_0}\right)^q}, \quad (10)$$

where r_{ij} is the distance between the i -th and j -th atoms, r_0 is a cut-off distance, taken to be 2.5 Å, p and q are set to 60 and 100 respectively. The atoms used as A and D in the sum of eq. 10 are reported in the table **S1** for each compound. For the protein, the acceptor atoms used in the sum have been the oxygen atom belonging to the backbone carbonyl group of Glu81, Phe82, Cys83, Asp84, Gln85, Ile191 and the oxygen atoms of the carboxylic group belonging to the side chain of residues Asp84 and Asp86. The donor atoms of the protein have been the hydrogen atom bond to the nitrogen of the backbone of residues Cys83, Asp84, Gln85, Asp86 and the hydrogen atoms bond to the nitrogen atom of the side chain of Gln85 and Lys89.

After having obtained R for all the 50 runs for each ligand, the mean H-bond value along the time series was obtained in the similar way as mentioned in the eq. 8 by replacing the force with R .

Table S1. In the first column, the selected acceptor and donor hydrogen bond atoms used in the calculation of the number of H-bond between protein and ligand are reported; in the second column, the atoms used in the definition of the center of mass used for the pulling distance are highlighted in pink.

Compound	Acceptor hydrogen atoms in yellow and donor hydrogen atoms in red	The atoms chosen for the definition of the center of mass for the pulling distance are in pink
1		
2		
3		
4		
5		
6		
7		

8		
9		

Stability of the binding poses during the MD simulation

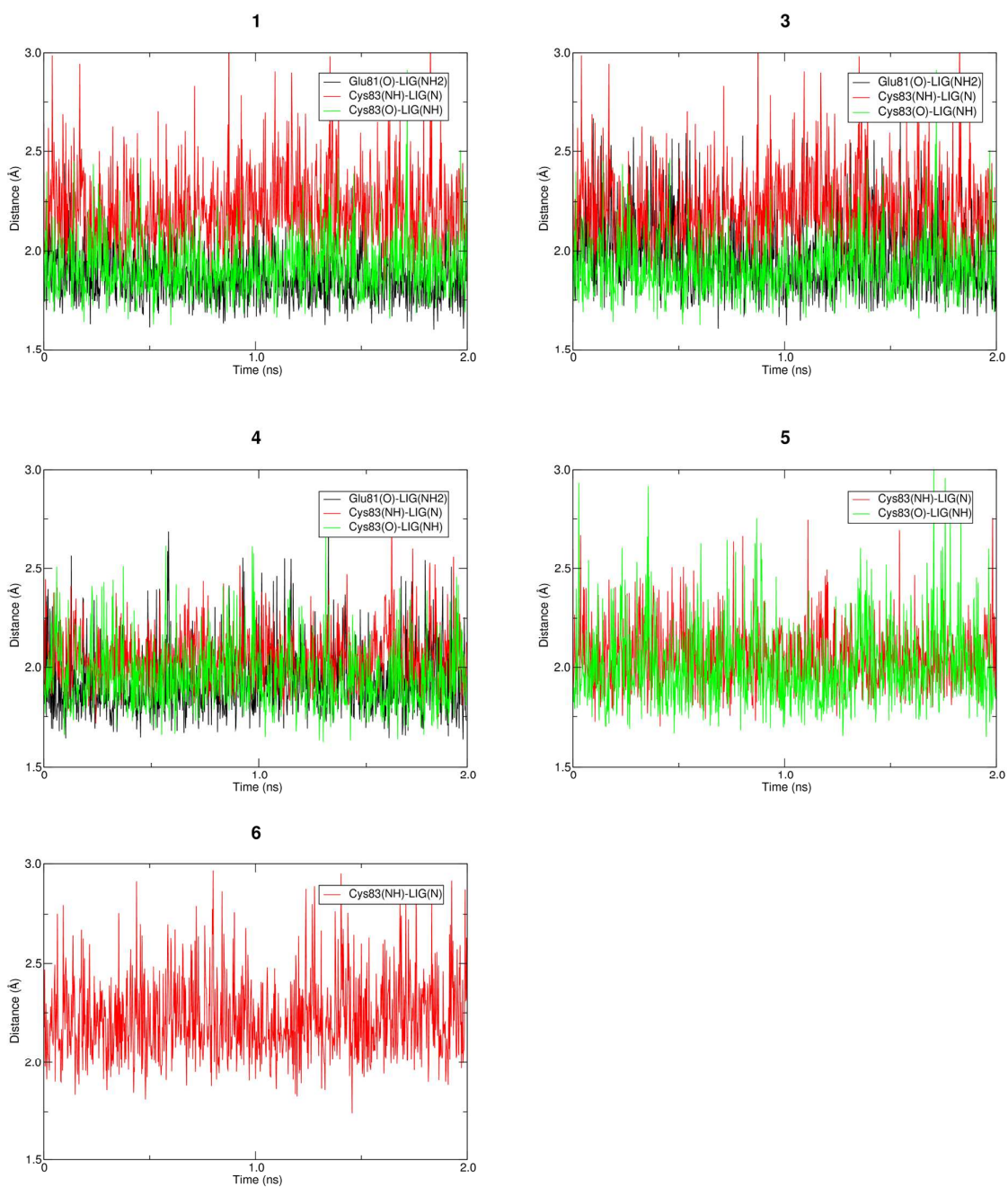


Figure S6. Distances between backbone atoms of Cys83 and Glu81 with the interacting atoms of the ligands.

Table S2. Ligand RMSD average value during the 2 ns of MD compared to the dock structure and standard deviation after Cys83 and Glu81 protein backbone atoms alignment to the dock structure.

	RMSD (Å)	SD
1	1.1	0.3
3	1.4	0.3
4	1.1	0.4
5	1.3	0.5
6	1.1	0.3

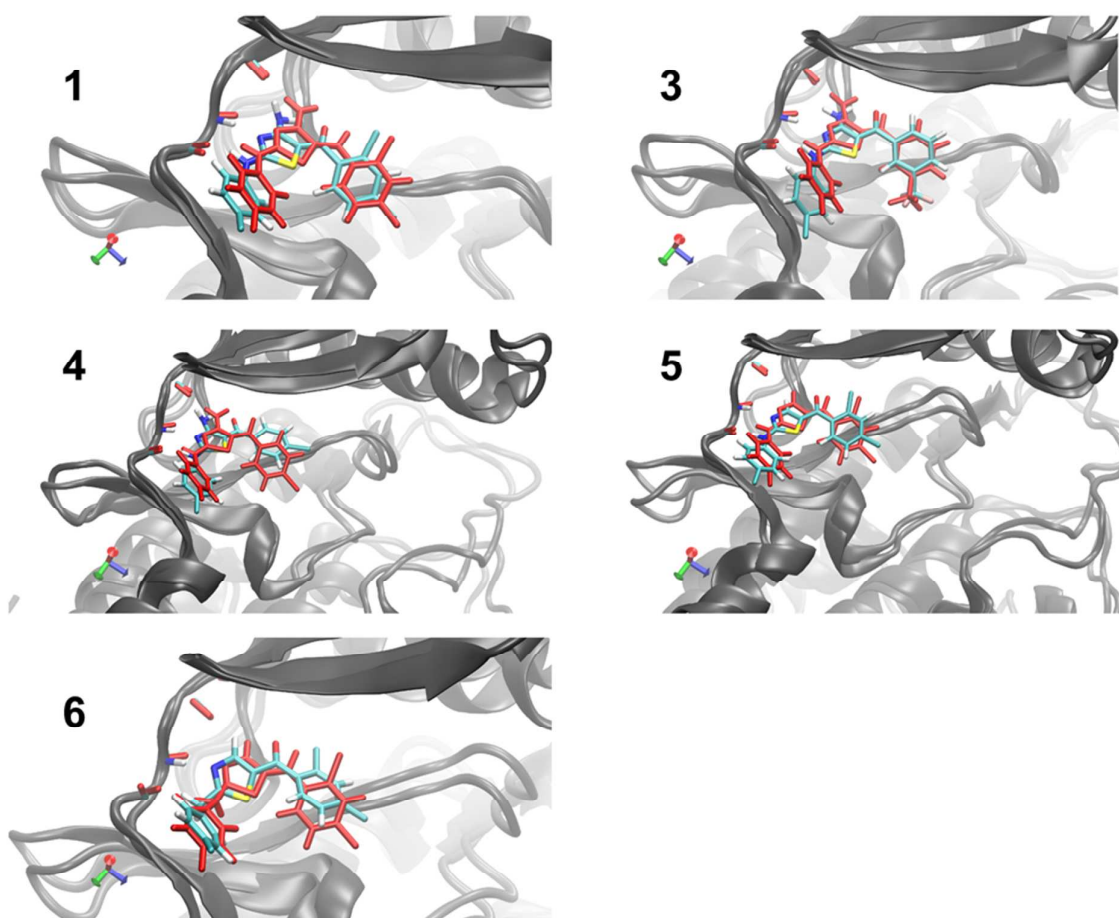


Figure S7: Superimposition between the dock structure in red and the final structure of 2 ns MD simulation for the compounds **1**, **3**, **4**, **5**, **6**. The backbone atoms of the hinge region interacting with the ligands are represented as licorice.

REFERENCES

1. Jarzynski, C., Equilibrium free-energy differences from nonequilibrium measurements: A master-equation approach. *Phys. Rev. E* **1997**, *56*, 5018-5035.
2. Jarzynski, C., Nonequilibrium Equality for Free Energy Differences. *Phys. Rev. Lett.* **1997**, *78*, 2690-2693.
3. Xiong, H.; Crespo, A.; Marti, M.; Estrin, D.; Roitberg, A. E., Free energy calculations with non-equilibrium methods: applications of the Jarzynski relationship. *Theor. Chem. Acc.* **2006**, *116*, 338-346.
4. Park, S.; Schulten, K., Calculating potentials of mean force from steered molecular dynamics simulations. *J. Chem. Phys.* **2004**, *120*, 5946-5961.
5. Jensen, M. O.; Park, S.; Tajkhorshid, E.; Schulten, K., Energetics of glycerol conduction through aquaglyceroporin GlpF. *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99*, 6731-6736.
6. Ytreberg, F. M., Computational study of small molecule binding for both tethered and free conditions. *J. Phys. Chem. B* **2010**, *114*, 5431-5434.
7. Vashisth, H.; Abrams, C. F., Ligand escape pathways and (un)binding free energy calculations for the hexameric insulin-phenol complex. *Biophys. J.* **2008**, *95*, 4193-4204.
8. Park, S.; Khalili-Araghi, F.; Tajkhorshid, E.; Schulten, K., Free energy calculation from steered molecular dynamics simulations using Jarzynski's equality. *J. Chem. Phys.* **2003**, *119*, 3559-3566.
9. Bakan, A.; Meireles, L. M.; Bahar, I., ProDy: protein dynamics inferred from theory and experiments. *Bioinformatics* **2011**, *27*, 1575-1577.