Performing effective calculations of protein - ligand binding free energy with the help of molecular dynamics methods

ABSTRACT

Aims: In our study, we aimed to calculate the binding free energies using molecular dynamics methods with the help of biased sampling approach using the binding site residues in protein - ligand complexes.

Methodology: Molecular modeling and dynamics methods such as Steered Molecular Dynamics (SMD) and Umbrella Sampling(US) were used in our studies. Systems to be simulated were prepared using the binding site residues and ligands. In order to perform SMD simulations, the systems were reoriented to allow the ligands to be pulled in the z-axis direction. The poses of the system were recorded while the ligands were drawn in the z-axis direction (reaction coordinate).Using the poses selected from the reaction coordinates, 1 ns molecular dynamics simulations were performed for each selected pose using the umbrella sampling method. The results obtained from US simulations were evaluated through the Gromacs Wham module and the binding free energies of the ligands were calculated.

Results: The binding free energies of the complexes we investigate were calculated as -7.23±1.75 for the 1NFU, -9.73±1.83 for the 2JS4, -9.16±2.53 for the 1FJS, -8.63±0.17 for the 1F0R and -13.76±2.47 kcal/mol for the 1KSN complexes.Experimental values binding free energies of thecomplexes mentioned were reported as -10.63, -10.47, -10.14, -10.51 and +12.90 kcal/mol, respectively. The bias between the binding free energy values we obtained and the experimental data vary between 0.8 kcal/mol (1FJS) and 3.4 kcal/mol (1NFU). In addition, there is a regression coefficient of 0.80 between our calculated binding free energies and the experimental results.

Conclusion: Thehigh regression coefficient between the binding free energies we determined in our study and the experimental results and the low bias of numerical differences of the binding free energies show that our approach gives positive results.

Keywords: Umbrella sampling, molecular dynamics, binding freeenergy, factor Xa

1. INTRODUCTION

An important measure of the interaction between proteins and their ligands is the binding free energy.A high binding free energy is an indication that the interaction between protein and ligand is high, and the dissociation coefficient of the protein and ligand is low [1].Accurate calculation of binding free energy is critical for researchers to achieve high success in a short time in designing and obtaining new inhibitor molecules in drug discovery projects [2]. Molecular docking and various molecular dynamics methods are the two most used molecular modeling approaches to estimate binding free energy [3, 4].Linear interaction energy (LIE) [5], MM/PBSA [6], free energy perturbation [7], alchemical approach [8] and biased sampling methodologies [9] are the most common methods used to calculate

free energy of binding of protein and ligand complexes. Following recent developments in machine learningmethodology, neural network models have begun to be used in estimating binding free energy. [10]

The most important cause of stroke is atrial fibrillation.The risk of embolism because of atrial fibrillation is high, and the use of anticoagulant drugs in patients with this condition reduces the possible fatal risks [11]. Factor Xa is a critical enzyme that plays a role in the blood clotting process. Therefore, using Factor Xa inhibitors to prevent blood clotting is an approach used to prevent possible embolism.The molecule called warfarin, a vitamin K antagonist, has been the most important oral anticoagulant drug used for decades [12]. Although warfarin is widely used, it has problems such as interactions with many drugs, food, and unexpected pharmacokinetic and pharmacodynamic properties [13].For this reason, even today, studies continuethe development of novel Factor Xa inhibitors that will have anticoagulant properties [14, 15].

In molecular dynamics simulations,as the volume of the system or the number of atoms the system contains decreases, the number of calculations that need to be performed also decreases. Thissituation makesit possible to complete the simulations in a shorter time.Also, reduced computational load makes it easier to use workstations or compute clusters with relatively low computational power. In the light of theseinformation, in our study, we aimed to calculate the binding free energies of some Factor Xa inhibitors using molecular dynamics methods with the help of biased sampling approach using the binding site residues in protein - ligand complexes instead of using all amino acids.We evaluated the success of our approach by comparing our results with experimental data and results in the literature.

2. MATERIALS AND METHODS

2.1 Hardware and software

All simulations were carried out on a workstationwith equipped with Intel i7 4770K CPU and two NVIDIA Graphics cards GTX 960 and RTX 1050GPUs running on Ubuntu 18.04 operating system. Atomistic simulations were carried out using Gromacs 2020.6 [16] using Amber99sb [17]. Parameterization of ligands was carried out using Acpype[18] with Anaconda interface [19]. Ligand binding energies were calculated using GromacsWham [20] module. Structure preparations, reorientation and rechaining were carried out using PyMOL, VMD [21,22] and Chimera [23] software. Graphs were created with the help of Python [24] and Microsoft Office Suite.

2.2Preparation of the complexes

The structures we use were downloaded from RCSB data bank with the PDB IDs of 1F0R [25], 1FJS[26], 1KSN [27], 1NFU [28], and 2J34[29].Each structure was loaded into PyMOL one by one and prepared for simulations separately.After loading the structure to PyMOL, water molecules were removed.Then, the amino acids of the binding site located within 5 Å of the ligand were identified and copied to be treated as a separate object.If the amino acids around 5 Å of the ligand were selected as a single structure, neighboring amino acids were also included in the structure and chains consisting of at least three amino acids were obtained.In addition, if one or two amino acid gaps occur between the selected chains, the chains are extended by adding the amino acids in between.The chains consisting of the resulting binding site amino acids were renamed and the structure were saved for the process of directing the ligand towards the z-axis.Later, the structure was loaded into the VMD software, and its coordinates were updated to orient the ligand to be pull on the z-axis direction from the binding site.Following this, the dimensions of the resulting structure were

determined to establish box dimensions of molecular dynamics simulations.Subsequently, the ligand structure was separated from the complex and loaded into the Chimera software, hydrogens were added to the structure and saved. Following this process, Gromacs compatible parameter files of the ligands were created with Acpype software and used in molecular dynamics simulations.The protein structure obtained in the previous step was used as the starting structure for the molecular dynamicssimulations.

2.3 Molecular dynamics simulations

Molecular dynamics simulations were carried out using Amber99sb force field. For steered molecular dynamics (SMD) simulations protein and ligand structures were combined and placed in a rectangular prism water cube whose edges were set to be 1 nm away from the system on x and y axes and four times higher than the dimensions of the complex on z-axis. Complexes were centered on x and y axis and set to be 1 nm close to edge on z-axis(Fig 1). Ion concentration of the cube was set to 0.15 M using Na^+ and Cl^- ions and neutralized.

The energy minimization of the created system was carried out by using the steepest descent minimization method in a maximum of 50000 steps, when the maximum force falls below 10 kJ /mol.The equilibrium process of the energy minimized system was carried out in three steps by using NVT and onestep by using NPTensembles. The positions of the proteins and ligandswerefixed using decreasing constrain values (2500, 1000, 500 kJ mol^{-1} $nm^{-2}$for NVT and 500 kJ mol^{-1} $nm^{-2}$for NPT simulations) during the equilibrium simulations. NTV equilibrium process was continued for 100 ps with time step 2 fs, the temperature of the system was set to 310 K, V-rescale was used as a thermostat. In NPT equilibrium process, time step 2 fs and simulation time was determined as 100 ps similar to NVT. Brendensenbarostatwas used as barostat in NPT and 1 bar was used as reference pressure.Short range electrostatic and van der Waals cut-off values were set to 0.9 nm.

After equilibration ligands were pulled through z-axis with the pulling rate of 0.005 nm/s with the force constant of 1000 kJ mol^{-1} $nm^{-2}$for 500 ps and coordinates and energy values were recorded for every 10 ps (5000 poses for each run). This process was repeated three times for each ligand.In each run, the time-dependent change in the distance between the ligand and the protein structure was recorded. Selected poses of these replicates were used as starting structures in umbrella sampling (US) simulations.

To calculate the binding energies of the complexes, poses were selected from SMD simulations, consideringthe distances between the protein and ligand structures.For simulations where the distance between protein and ligand was between 0.2 - 1.0 nm,samples were selected forevery 0.05 nm, and between 1.0 - 2.0 nm, samples were selected forevery 0.1 nm.

All the selected poses were equilibratedusing Brendensen as barostat in NPT and 1 bar was used as reference pressure before applying US simulations. US simulations were carried out with the pulling rate of 0 nm/s with the force constant of 500 kJ mol^{-1} $nm^{-2}$to generate poses in the given distance between ligand and protein. At the end of US simulations energy profile of the complex were computed using Gromacs wham module. Binding free energies of the complexes were calculated using differences in energies of binded and unbinded complex.

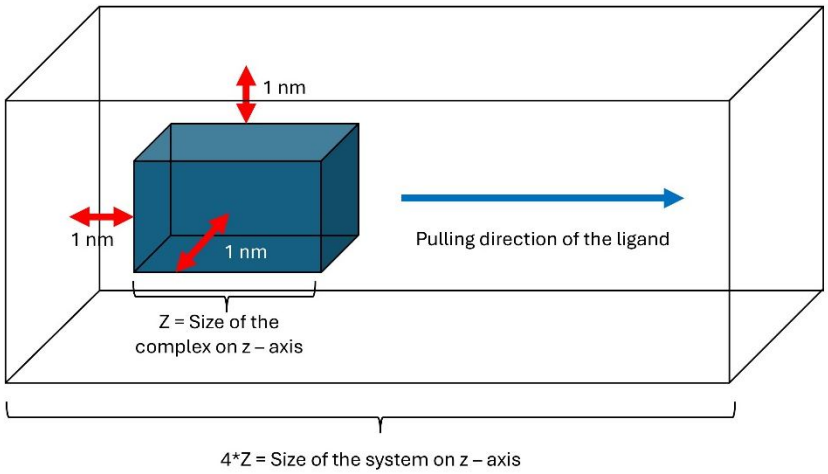

Fig. 1.Dimensions of the systems and pulling direction of the ligands

3. RESULTS AND DISCUSSION

Our aim in thisstudy was to calculate the binding free energies in an accelerated manner using simulations using the residue and ligand structures in the binding region of the complex structures thatwe studied.By focusing on the binding site of the complexes, we aimed to reduce the volume and number of atoms of the system we worked on and to avoid wasting resources on calculations for residues that do not contribute to binding.Table 1 shows the comparison of the sizes of the systems (binding site residuesand full complex) created for 1F0R, one of the complexes we studied, and the number of atoms they contain.When the Table 1 is examined, it can be seen that volume of the system created using the entire complex is approximately seven times (6.93 times) larger than the system created using binding site residues.When the number of atoms in the systems is compared, it is observed that the entire complex contains approximately seven times (7.3 times) more atoms.We also compared simulation performances of equilibrium and SMD simulation of both systems. For binding siteresiduesystem in equilibration phases the averagesimulation rate was 263.80 ns/day and for full complex system this value was 38.28 ns/day. The ratio of simulation rates for equilibration phases is again close to seven with the value of 6.89. Average SMD simulations rates for binding site residuesystem and full complex system were 168.09 and 26.68 ns/day, respectively. The ratio of simulation rates atSMD phase is close to seven with the value of 6.30, too. These numbers show that, , more calculations can be made per unit time by decreasing the volume of the systems and the number of atoms they contain.

Table 1. Dimensions and sizes of the systems for 1F0R complex

	1F0R	1F0R (Full complex)
Size x (nm)	2.1500	3.7990
Size y (nm)	2.4880	4.5930
Size z (nm)	2.1900	7.3938
Box size x (nm)	4.1510	5.7990

Box size y (nm)	4.4880	6.5930
Box size z (nm)	8.7640	29,575
Volume (nm^3)	163,2706	1130.7553
# of total atoms	15753	114791
# of total water molecules	5101	36714

Table 2 contains information about the size and atomic numbers of the systems created by taking into account the binding sites residuesof the complexes we examined in our study.Among these systems, the 1NFU complex has the smallest volume with 125.2538 nm^3, while the largest volume belongs to the 1KSN system with 175.7014 nm^3.While there are 12527 atoms and 4039 water molecules in the 1NFU system, there are 17166 atoms and 5544 water molecules in the 1KSN system.These numbers are the lowest and highest for 1NFU and 1KSN systems, similar to the volume ranking among the complexes studied.

Table 2. Dimensions and sizes of the systems for binding site residues

	1NFU	**2JS4**	**1FJS**	**1F0R**	**1KSN**
Size x (nm)	2.0880	2.4390	2.7820	2.1500	2.5430
Size y (nm)	2.2930	2.3580	2.3600	2.4880	2.5760
Size z (nm)	1,8270	1.7980	1.7940	2.1900	2.1190
Box size x (nm)	4.0880	4.4390	4.7820	4.1510	4.5430
Box size y (nm)	4.2930	4.3580	4.3600	4.4880	4.5760
Box size z (nm)	7.3080	7.1920	7.1760	8.7640	8.4760
Volume (nm^3)	125.2538	139.1304	149.6162	163,2706	175.7014
# of total atoms	12527	13609	14682	15753	17166
# of total water molecules	4039	4391	4722	5101	5544

Simulation rates of the binding site residuessystems we studied were also examined and summarized in Table 3. As seen in Table 3, it was observed that the simulation speeds decreased as the volumes of the systems and the number of atoms they contain increased,

Table 3. Simulation rates of binding site residuessystems

	1NFU	**2JS4**	**1FJS**	**1F0R**	**1KSN**
Average simulation rate in equilibrium phase (ns/day)	296.22	269.51	275.88	263.80	251.88
Average simulation rate in SMD phase (ns/day)	241.22	213.59	214.88	168.09	188.74
Average simulation rate in US phase (ns/day)	355.30	331.75	303.61	301.36	274.46

As mentioned in the material method section, US simulations were repeated three times with the help of poses selected from the SMD simulations. The results of three replicates are summarized in Table 4 and showed in Fig 2.Simulated binding freeenergy for **1NFU**binding site residuescomplex wascalculated as **-7.23±1.75**, for **2JS4**binding site residuescomplex as **-9.73±1.83**, for **1FJS**binding site residuescomplex as-**9.16±2.53**, for **1F0R**binding site residuescomplex as-**8.63±0.17,** for **1KSN**binding site residuescomplex as **-13.76±2.47** kcal/moles.When the results obtained were evaluated, it was observed that the simulations with the highest reproducibilitywasin the 1F0R binding site residuessystem with 0.17 standard deviation value and the lowest in the 1FJS binding site system with 2.53 standard deviation value.

Table 4. Simulated binding energies of binding ~~pocket~~ site residuessystems

	1NFU	2JS4	1FJS	1F0R	1KSN
Simulated binding freeenergy for first US simulations (kcal/mol)	-8.06	-11.71	- 9.02	- 8.71	-16.60
Simulated binding freeenergy for second US simulations (kcal/mol)	-8.41	-9.39	- 11.75	- 8.43	- 12.61
Simulated binding freeenergy for third US simulations (kcal/mol)	-5.22	-8.09	- 6.70	- 8.74	- 12.07
Average simulated binding freeenergy of US simulations (kcal/mol)	-7.23	-9.73	- 9.16	- 8.63	- 13.76
Standard deviations of simulated binding freeenergy of US simulations (kcal/mol)	1.75	1.83	2.53	0.17	2.47

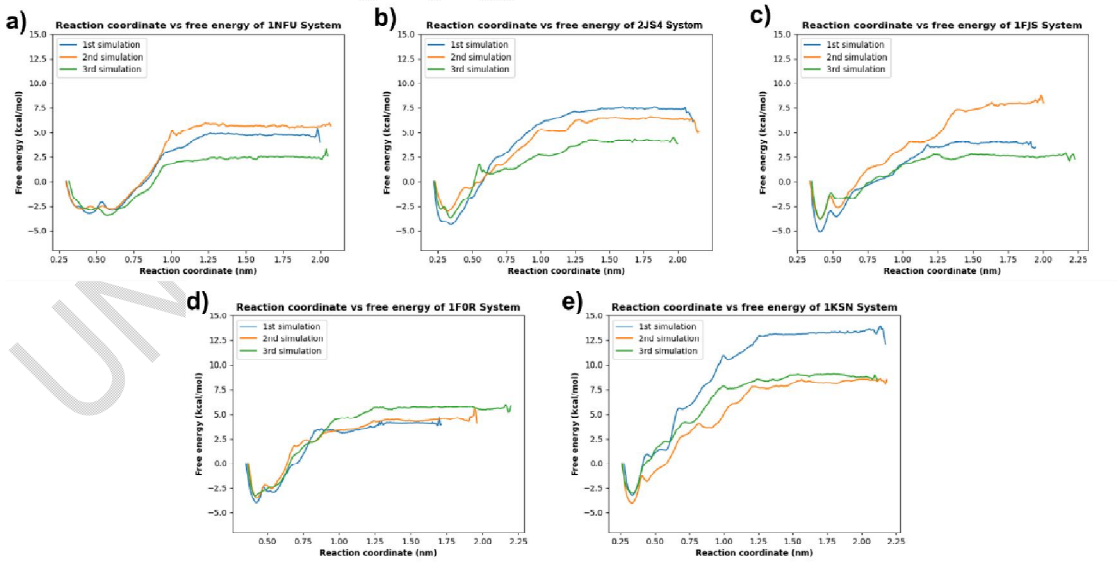

Fig.2.Simulated binding freeenergy graphs of a) 1NFU, b) 2JS4, c) 1FJS, d) 1F0R and e) 1KSN systems

When the results summarized above were compared with the literature information, it was observed that promising values were obtained. Ngo and colleagues worked on four different class of target complexes – cathepsin K (CSTK), type II dehydroquinase (DHQase), heat shock protein 90 (HSP90) and factor Xa (FXa) – on calculation of binding energies using umbrella sampling method [30]. In their paper experimental binding energies for **1NFU** is reported as **-10.63**, for **2JS4** as **-10.74**, for**1FJS** as **-10.14,** for**1F0R** as **-10.51 and** for **1KSN** as **-12.90 kcal/mol.** The simulated binding energies obtained in our study, which we summarize in Table 4, and the experimental results are very close to each other.For example, the difference between the simulated binding freeenergy and experimental binding energies in **1KSN** complex was determined as **0.86 kcal/mol**, which are quite close to each other.The largest difference between the simulated binding freeenergy and experimental binding energies was obtained in the **1NFU** complex with a difference of **3.40 kcal/mol**.When the binding freeenergy differences for other complexes were examined, the difference was determined to be **0.98** for **1FJS** complex, **1.01** for **2JS4** and **1.88 kcal/mol** for **1F0R**.For the complexes we examined in our study, information is given about the simulated binding energies using the full complex in the work conducted by Ngo and colleagues.Accordingly, the simulated binding freeenergy for the **1NFU** complex is stated as **-15.43**, **-16.06** for **2JS4**, **-13.48** for **1FJS**, **-12.65** for **1F0R** and finally **-24.45 kcal/mol** for **1KSN**.The differences between the results we obtained in our study and the experimental results are lower than the differences obtained in the study by Ngo and colleagues.In addition, the simulated binding energies calculated in our study are lower than experimental results in all except the 1KSN complex, while they are higher in all complexes in the study by Ngo and colleagues.All the comparisons mentioned above are summarized in Table 5.Regression analyzes were also performed with the differences in the numbers between simulated binding energies and experimental binding energies.As a result of the analysis, it was determined that there was a regression of **0.80** between the binding energies.To calculate the experimental binding energies of the studied ligands according to the obtained formula, the formula$\mathbf{\Delta G_{EXP} = 0.400\ \Delta G_{US} - 7.1031}$ should be used (Fig 3).

Table 5. Comparison of simulated binding energies obtained in our study and those performed by Ngo and colleagues.

	ΔG_{exp} $(kcal/mol)$	ΔG_{US*} $(kcal/mol)$	ΔG_{US**} $(kcal/mol)$	$\Delta\Delta G_{exp-*}$ $(kcal/mol)$	$\Delta\Delta G_{exp-**}$ $(kcal/mol)$
1NFU	-10.63	- 7.23	- 15.43	3.40	4.80
2JS4	- 10.74	- 9.73	- 16.06	1.01	5.32
1FJS	- 10.14	- 9.16	- 13.48	0.98	3.34
1F0R	- 10.51	- 8.63	- 12.65	1.88	2.14
1KSN	- 12.90	- 13.76	- 24.45	0.86	11.55

**Results of this study*
*** Results of Ngo and colleagues*

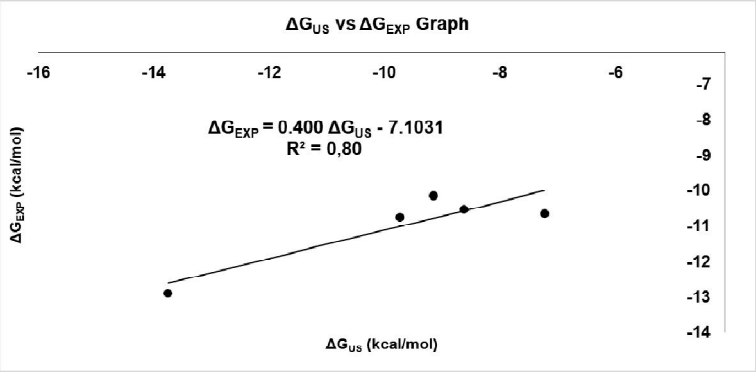

Fig.3.Regression graph of ΔG_{US} vs ΔG_{EXP}

4. CONCLUSION

Our aim in carrying out this study was to develop a fast and effective approach to calculate the binding energies of protein - ligand complexes.To achieve our goal, it is aimed to bypass the molecular dynamics calculations for residues that are not in the binding site in protein-ligand complexes and perform operations for fewer atoms and smaller system volumes. It wasobserved that, by selecting only binding site residues,we were able to reduce the volume of the systems and the number of atoms in the systems by approximately seven timeand accelerated calculation rate approximately 6.5 times compared to full residues systems.

As a result of our simulations and calculations, we observed differences of 0.86 to 3.40 kcal/mol between the simulated and experimental binding energies.The fact that these values are lower than the values of 2.14 (lowest difference) and 11.55 kcal/mol (highest difference) obtained by Ngo and colleagues [30], who examined the same complexes in the literature, indicates that our approach has some advantages. In addition, the regression between our simulated binding free energies and experimental binding free energies was also examined.Accordingly, while the regression coefficient we obtained in our study was determined as 0.80, this value was stated as 0.95 in the study conducted by Ngo and colleagues [30].Although the differences in binding free energies were lower than the literature, the regression results fell behind the literature.

Success has also been achieved in increasing the calculation speed, which wass one of the main goals of our work. We were able to reach simulation rates higher than 300 ns/day even with hardware that have disadvantageous - like low ram capacity and calculation capabilities -compared to the modern, powerful CPUs and GPUs available today.With the use of modern GPUs available today, it is possible to increase simulation rates much higher than 300 ns/day without sacrificing accuracy of the results.

In addition to the approach, we mentioned in our study allowing fast and high-accuracy calculation of binding free energy, other application areas can be used. The increase in the rates of calculations allows more samples to be made per unit time. Our approach outlined in our study can be used for systems that require more sampling or for systems that need to be examined in a wider range of reaction coordinates. For example, the protein-ligand binding is a dynamic process and calculation of binding free energies with higher accuracy would be

possible by evaluating the many reaction coordinates of binding – unbinding of complex.Compared to systems where all residues were used, the increase in simulation speed obtained in systems where residues in the binding site were used will make it possible to sample the ligand in reaction coordinates to be created in line with other vectors. In this way, more detailed information about the binding and unbinding process will be obtained.Another area of use or advantage of our approach is that it allows more sampling at selected poses along the reaction coordinate. Increasing the number of samples will provide more information for calculating the binding free energy. Finally, in the US method, poses selected at certain intervals on the reaction coordinate are used as the initial configuration. Decreasing the spacing or increasing the number of selected poses will allow a more detailed examination of how the binding free energies of the protein and ligand change along the reaction coordinate.

CONSENT

It's not applicable.

ETHICAL APPROVAL

It's not applicable.

DISCLAIMER (ARTIFICIAL INTELLIGENCE)

Author(s) hereby declare that NO generative AI technologies such as Large Language Models (ChatGPT, COPILOT, etc.) and text-to-image generators have been used during the writing or editing of this manuscript.

REFERENCES

1. Borea, P. A., Varani, K., Gessi, S., Gilli, P., & Dalpiaz, A. (1998). Receptor binding thermodynamics as a tool for linking drug efficacy and affinity. Il Farmaco, 53(4), 249-254.

2. Cournia, Z., Allen, B., & Sherman, W. (2017). Relative binding free energy calculations in drug discovery: recent advances and practical considerations. *Journal of chemical information and modeling*, *57*(12), 2911-2937.

3. Guedes, I. A., de Magalhães, C. S., & Dardenne, L. E. (2014). Receptor–ligand molecular docking. *Biophysical reviews*, *6*, 75-87.

4. Deng, Y., & Roux, B. (2009). Computations of standard binding free energies with molecular dynamics simulations. *The Journal of Physical Chemistry B*, *113*(8), 2234-2246.

5. de Amorim, H. L. N., Caceres, R. A., & Netz, P. A. (2008). Linear interaction energy (LIE) method in lead discovery and optimization. *Current drug targets*, *9*(12), 1100-1105.

6. Hou, T., Wang, J., Li, Y., & Wang, W. (2011). Assessing the performance of the MM/PBSA and MM/GBSA methods. 1. The accuracy of binding free energy calculations based on molecular dynamics simulations. *Journal of chemical information and modeling*, *51*(1), 69-82.

7. Wang, J., Deng, Y., & Roux, B. (2006). Absolute binding free energy calculations using molecular dynamics simulations with restraining potentials. *Biophysical journal*, *91*(8), 2798-2814.

8. Ngo, S. T., Thai, Q. M., Nguyen, T. H., Tuan, N. N., Pham, T. N. H., Phung, H. T., & Quang, D. T. (2024). Alchemical approach performance in calculating the ligand-binding free energy. *RSC advances*, *14*(21), 14875-14885.

9. Kästner, J. (2011). Umbrella sampling. *Wiley Interdisciplinary Reviews: Computational Molecular Science*, *1*(6), 932-942.

10. Akkus, E., Tayfuroglu, O., Yildiz, M., & Kocak, A. (2022). Accurate binding free energy method from end-state MD simulations. *Journal of Chemical Information and Modeling*, *62*(17), 4095-4106.

11. McCarty, D., & Robinson, A. (2016). Factor Xa inhibitors: a novel therapeutic class for the treatment of nonvalvular atrial fibrillation. *Therapeutic advances in cardiovascular disease*, *10*(1), 37-49.

12. Yeh, C. H., Fredenburgh, J. C., & Weitz, J. I. (2012). Oral direct factor Xa inhibitors. *Circulation research*, *111*(8), 1069-1078.

13. Pinto, D. J., Smallheer, J. M., Cheney, D. L., Knabb, R. M., & Wexler, R. R. (2010). Factor Xa inhibitors: next-generation antithrombotic agents. Journal of medicinal chemistry, 53(17), 6243-6274.

14. Zheng, W., Dai, X., Xu, B., Tian, W., & Shi, J. (2023). Discovery and development of Factor Xa inhibitors (2015–2022). *Frontiers in Pharmacology*, *14*, 1105880.

15. Zhao, Y., Liu, Q., Du, J., Meng, Q., Sun, L., & Zhang, L. (2024). Accelerating factor Xa inhibitor discovery with a de novo drug design pipeline. *Chinese Journal of Chemical Engineering*.

16. Abraham MJ, Murtola T, Schulz R, Páll S, Smith JC, Hess B, et al. (2015) GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. SoftwareX, 1, 19- 25.

17. Hornak V, Abel R, Okur A, Strockbine B, Roitberg A, Simmerling C. (2006). Comparison of multiple Amber force fields and development of improved protein backbone parameters. Proteins, 65(3), 712-725.

18. Da Silva AWS, Vranken WF. (2012). ACPYPE Antechamber python parser interface. BMC Res Notes, 5, 1-8.

19. Anaconda Software Distribution. Computer software. Vers. 2-2.4.0. Anaconda, Nov. 2016. Web. <https://anaconda.com>.

20. Hub, J. S., De Groot, B. L., & Van Der Spoel, D. (2010). g_wham A Free Weighted Histogram Analysis Implementation Including Robust Error and Autocorrelation Estimates. Journal of chemical theory and computation, 6(12), 3713-3720.

21. Schrödinger, L., & DeLano, W. (2020). PyMOL. Retrieved from http://www.pymol.org/pymol

22. Humphrey, W., Dalke, A., & Schulten, K. (1996). VMD: visual molecular dynamics. Journal of molecular graphics, 14(1), 33-38.

23. Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., & Ferrin, T. E. (2004). UCSF Chimera—a visualization system for exploratory research and analysis. *Journal of computational chemistry*, *25*(13), 1605-1612.

24. Van Rossum, G., & Drake, F. L. (2009). Python 3 Reference Manual. Scotts Valley, CA: CreateSpace.

25. Maignan, S., Guilloteau, J. P., Pouzieux, S., Choi-Sledeski, Y. M., Becker, M. R., Klein, S. I., ... & Mikol, V. (2000). Crystal structures of human factor Xa complexed with potent inhibitors. *Journal of medicinal chemistry*, *43*(17), 3226-3232.

26. Adler, M., Davey, D. D., Phillips, G. B., Kim, S. H., Jancarik, J., Rumennik, G., ... & Whitlow, M. (2000). Preparation, characterization, and the crystal structure of the inhibitor ZK-807834 (CI-1031) complexed with factor Xa. *Biochemistry*, *39*(41), 12534-12542.

27. Guertin, K. R., Gardner, C. J., Klein, S. I., Zulli, A. L., Czekaj, M., Gong, Y., ... &Pauls, H. W. (2002). Optimization of the β-Aminoester class of factor Xa inhibitors. part 2: Identification of FXV673 as a potent and selective inhibitor with excellent In vivo anticoagulant activity. *Bioorganic & medicinal chemistry letters*, *12*(12), 1671-1674.

28. Maignan, S., Guilloteau, J. P., Choi-Sledeski, Y. M., Becker, M. R., Ewing, W. R., Pauls, H. W., ... & Mikol, V. (2003). Molecular structures of human factor Xa complexed with ketopiperazine inhibitors: preference for a neutral group in the S1 pocket. *Journal of medicinal chemistry*, *46*(5), 685-690.

29. Senger, S., Convery, M. A., Chan, C., & Watson, N. S. (2006). Arylsulfonamides: a study of the relationship between activity and conformational preferences for a series of factor Xa inhibitors. *Bioorganic & medicinal chemistry letters*, *16*(22), 5731-5735.

30. Ngo, S. T., Vu, K. B., Bui, L. M., & Vu, V. V. (2019). Effective estimation of ligand-binding affinity using biased sampling method. *ACS omega*, *4*(2), 3887-3893.