Supporting Information

In Silico Design of Miniprotein to Inhibit SARS-CoV-2 Variant Omicron Spike Protein

Jianhua Wu,[†] Hong-Xing Zhang[†,*] and Jilong Zhang,[†,*]

[†] Institute of Theoretical Chemistry, College of Chemistry, Jilin University, Changchun 130023, Jilin, People's Republic of China.

*Emails: zhanghx@jlu.edu.cn; jilongzhang@jlu.edu.cn

Methods

Principal Component Analysis

Principal component analysis (PCA)[1, 2] is a statistical technique for extracting essential information from MD trajectories and is often applied to reduce the dimensionality of data describing protein motion. The observed protein motion is filtered from the largest to the smallest spatial scale by a decomposition process. The translational and rotational motion of all Cα atoms was removed by fitting each frame in the MD simulation trajectories to the mean structure. The 'covar' module of GROMACS was performed to construct a covariance matrix of the Cα atomic positions of the complex MD simulation trajectories. The covariance matrix, C_{ij}, is represented by the following equation

$$C_{ij} = \left\langle \left(r_i - \langle r_i \rangle\right) \times \left(r_j - \langle r_j \rangle\right) \right\rangle \quad (i,j = 1, 2, 3, \ldots, \mathrm{N})$$

where r_i and r_j are the Cartesian coordinates of the i-th and j-th Cα atoms. The eigenvectors and the corresponding eigenvalues are obtained by diagonalizing the covariance matrix based on the Cα atomic fluctuations. The eigenvectors and eigenvalues represent the direction and amplitude of the protein motion, respectively. The eigenvectors were assessed by the 'anaeig' module. The eigenvectors were sorted in descending order according to the eigenvalues, and the eigenvector with the highest eigenvalue, namely, the first principal component (PC1), which is mainly responsible for describing the most important motion of the protein, was used to draw porcupine plots in this

work, to elucidate the RBD^O protein's dominant global motility changes induced by inhibitor binding.

Binding Free Energy Calculation

The binding free energy is an essential indicator for assessing the binding strength of receptors to ligands. The g_mmpbsa tool[3] in the scripts-based APBS program was performed to calculate the binding free energy using the molecular mechanics/Poisson−Boltzmann surface area (MM/PBSA) method,[4-6] which is a popular approach for providing detailed information on the interaction between receptor and ligand. 6000 frames were extracted from MD trajectories with stable conformations to predict the binding free energy between RBD^O protein and ACE2 or inhibitors according to the following equation

$$\Delta G_{bind} = \Delta G_{complex} - (\Delta G_{receptor} + \Delta G_{ligand})$$

here $\Delta G_{complex}$, $\Delta G_{receptor}$, and ΔG_{ligand} refer to the free energy of complex, RBD^O, ACE2 or inhibitors, respectively. ΔG_{bind} can be decomposed into different forms of interactions and is written as

$$\Delta G_{bind} = \Delta E_{MM} + \Delta G_{sol} - T\Delta S$$

$$\Delta E_{MM} = \Delta E_{vdW} + \Delta E_{ele} + \Delta E_{int}$$

where ΔG_{bind} consists of the gas-phase molecular mechanics energy (ΔE_{MM}), solvation free energy (ΔG_{sol}), and conformational entropy upon ligand binding ($-T\Delta S$). Because the goal of the calculation is to obtain the relative free energy and not the absolute value, the entropy with the large computational cost and low prediction accuracy was not considered in the free energy calculation. The ΔE_{MM} can be divided into three terms: van der Waals contribution (ΔE_{vdW}), electrostatic contribution (ΔE_{ele}), and internal energy contribution (ΔE_{int}, sum of the energies of bond, angle, and dihedral). Because the single trajectory strategy was applied between complex, receptor, and ligand, the ΔE_{int} term was completely canceled. The ΔG_{sol} can be broken down into polar (ΔG_{PB}) and nonpolar (ΔG_{SA}) energies.

$$\Delta G_{sol} = \Delta G_{PB} + \Delta G_{SA}$$

$$\Delta G_{SA} = \gamma \cdot \mathrm{SASA} + \beta$$

in which the ΔG_{PB} was computed by using the Poisson−Boltzmann (PB) model with the solute and solvent dielectric constants of 1.0 and 80.0, respectively, and ΔG_{SA} was estimated from solvent

accessible surface area (SASA) model with the parameters γ and β of 0.00542 kcal $mol^{-1}Å^{-2}$ and 0.92 kcal mol^{-1}, respectively.

Table S1. Alanine mutation energy and effect of mutation based on inhibitor AHB2 (kcal/mol). Residues are sorted from high to low according to mutation energy and effect.

Mutation	Energy	Effect of Mutation	Mutation	Energy	Effect of Mutation
HIS18→ALA	2.48	DESTABILIZING	LEU69→ALA	0.05	NEUTRAL
ASP11→ALA	2.16	DESTABILIZING	LEU16→ALA	0.04	NEUTRAL
LEU44→ALA	1.45	DESTABILIZING	LEU20→ALA	0.04	NEUTRAL
GLU41→ALA	1.34	DESTABILIZING	ILE55→ALA	0.04	NEUTRAL
TRP37→ALA	1.29	DESTABILIZING	VAL13→ALA	0.03	NEUTRAL
LEU21→ALA	1.13	DESTABILIZING	SER14→ALA	0.03	NEUTRAL
ASN36→ALA	0.97	DESTABILIZING	LEU46→ALA	0.03	NEUTRAL
GLU3→ALA	0.94	DESTABILIZING	ILE58→ALA	0.03	NEUTRAL
GLU45→ALA	0.90	DESTABILIZING	ILE65→ALA	0.03	NEUTRAL
GLU30→ALA	0.88	DESTABILIZING	VAL6→ALA	0.02	NEUTRAL
GLU15→ALA	0.81	DESTABILIZING	GLN12→ALA	0.02	NEUTRAL
GLU19→ALA	0.71	DESTABILIZING	VAL9→ALA	0.01	NEUTRAL
GLU54→ALA	0.71	DESTABILIZING	LEU66→ALA	0.01	NEUTRAL
GLU4→ALA	0.68	DESTABILIZING	LEU72→ALA	0.01	NEUTRAL
GLU61→ALA	0.67	DESTABILIZING	ALA17→ALA	0.00	NEUTRAL
ASP50→ALA	0.66	DESTABILIZING	ALA33→ALA	0.00	NEUTRAL
MET7→ALA	0.65	DESTABILIZING	ALA62→ALA	0.00	NEUTRAL
MET43→ALA	0.64	DESTABILIZING	ALA73→ALA	0.00	NEUTRAL

		G			
GLU59→ALA	0.64	DESTABILIZING	ALA32→ALA	-0.01	NEUTRAL
HIS22→ALA	0.59	DESTABILIZING	ALA39→ALA	-0.01	NEUTRAL
ASP51→ALA	0.58	DESTABILIZING	SER49→ALA	-0.01	NEUTRAL
GLU57→ALA	0.58	DESTABILIZING	MET42→ALA	-0.02	NEUTRAL
GLU27→ALA	0.56	DESTABILIZING	PHE35→ALA	-0.03	NEUTRAL
GLU52→ALA	0.54	DESTABILIZING	LEU24→ALA	-0.04	NEUTRAL
GLU60→ALA	0.53	DESTABILIZING	THR25→ALA	-0.07	NEUTRAL
GLU70→ALA	0.53	DESTABILIZING	GLY26→ALA	-0.16	NEUTRAL
GLU1→ALA	0.52	DESTABILIZING	ARG31→ALA	-0.43	NEUTRAL
GLU67→ALA	0.49	NEUTRAL	LYS75→ALA	-0.43	NEUTRAL
GLU71→ALA	0.47	NEUTRAL	ARG74→ALA	-0.46	NEUTRAL
GLU28→ALA	0.42	NEUTRAL	LYS23→ALA	-0.48	NEUTRAL
LEU29→ALA	0.39	NEUTRAL	ARG53→ALA	-0.48	NEUTRAL
ILE47→ALA	0.37	NEUTRAL	HIS68→ALA	-0.55	STABILIZING
THR40→ALA	0.22	NEUTRAL	ARG56→ALA	-0.57	STABILIZING
TYR34→ALA	0.21	NEUTRAL	ARG64→ALA	-0.57	STABILIZING
LEU2→ALA	0.19	NEUTRAL	ARG63→ALA	-0.61	STABILIZING
LEU10→ALA	0.10	NEUTRAL	LYS48→ALA	-0.68	STABILIZING
GLN5→ALA	0.08	NEUTRAL	HIS8→ALA	-0.70	STABILIZING
TRP38→ALA	0.06	NEUTRAL			

Table S2. Mutation energy of single mutation based on the inhibitor AHB2 (kcal/mol). The values of favorable residue mutation energy are bolded.

	E3	E4	M7	D11	E15	H18	E19	L21	E30	N36	W37	E41	M43	L44	E45	D50	E54	E59	E61
ALA	0.94	0.68	0.65	2.16	0.81	2.48	0.71	1.13	0.88	0.97	1.29	1.34	0.64	1.45	0.90	0.66	0.71	0.64	0.67
ARG	1.46	0.61	0.62	2.86	1.42	1.8	1.34	0.04	0.74	**-1.69**	1.56	1.57	0.45	1.89	1.76	1.32	1.48	1.24	1.37
ASN	0.67	0.68	0.18	2.13	0.76	2.00	0.74	0.29	0.48	0.00	0.49	0.54	0.44	1.18	0.87	0.68	0.74	0.61	0.68
ASP	-0.15	**-0.65**	**-1.00**	0.03	0.04	2.31	0.19	0.49	-0.09	-0.34	**-0.6**	-0.20	-0.50	0.23	0.03	0.09	0.03	0.00	0.04
CYS	0.84	0.60	0.59	1.86	0.76	2.35	0.67	1.04	0.73	0.62	1.01	1.37	0.55	1.27	0.93	0.71	0.72	0.64	0.67
GLN	0.67	0.74	-0.14	2.18	0.78	0.88	0.56	-0.27	0.72	**-0.81**	1.22	0.91	-0.15	0.64	0.84	0.70	0.70	0.63	0.66
GLU	0.00	-0.2	**-1.48**	1.29	0.05	1.82	0.28	0.06	0.04	**-1.19**	-0.40	0.03	**-1.00**	-0.28	-0.05	0.06	0.02	0.00	0.00
GLY	1.09	0.77	1.05	2.32	0.87	2.83	0.75	1.43	1.35	1.41	1.85	1.44	0.83	1.92	0.84	0.72	0.71	0.64	0.67
HIS	1.65	-0.29	**-0.59**	1.13	0.61	0.49	0.84	-0.07	1.27	-0.38	-0.33	0.24	**-1.48**	-0.30	0.74	0.71	0.80	0.61	0.61
ILE	0.61	0.23	0.08	1.59	0.55	0.74	0.61	-0.09	0.53	0.37	0.98	0.77	0.02	0.94	0.88	0.70	0.71	0.62	0.65
LEU	0.48	0.02	-0.47	1.69	0.55	1.47	0.66	-0.01	0.04	**-0.72**	0.65	1.03	-0.07	-0.07	0.86	0.74	0.70	0.61	0.65
LYS	1.46	1.56	0.70	3.29	1.63	0.82	1.35	0.59	1.79	**-0.98**	2.39	1.87	0.90	1.70	1.58	1.33	1.38	1.29	1.34
MET	1.04	0.32	-0.04	2.05	0.64	1.25	0.84	-0.15	0.50	0.35	1.17	0.55	-0.08	0.38	0.78	0.71	0.72	0.62	0.65
PHE	0.81	-0.08	**-1.44**	1.44	0.13	0.94	0.83	-0.26	-0.14	-0.16	-0.4	0.94	-0.26	0.06	0.62	0.62	0.77	0.61	0.62
PRO	0.86	0.42	0.44	1.84	0.46	1.94	0.49	0.80	0.20	0.72	0.62	0.65	0.37	0.86	0.64	0.66	0.71	0.60	0.64
SER	0.93	0.90	0.92	2.23	0.79	2.80	0.67	1.21	1.40	0.87	2.42	1.24	0.65	1.92	0.88	0.69	0.72	0.63	0.68
THR	0.81	0.87	0.31	1.95	0.69	2.28	0.61	0.71	1.31	0.26	2.13	1.16	0.46	1.26	0.90	0.70	0.73	0.61	0.67
TRP	1.09	**-1.24**	**-1.12**	0.59	0.04	0.55	0.61	**-0.87**	**-0.63**	**-0.58**	0.05	1.07	**-1.63**	**-1.38**	0.84	0.57	0.72	0.52	0.61
TYR	-0.01	**-0.6**	-0.08	1.53	0.08	-0.04	0.69	-0.11	-0.14	**-1.07**	0.75	1.14	**-0.54**	**-1.18**	0.66	0.57	0.77	0.62	0.64
VAL	0.70	0.36	0.92	1.82	0.62	1.89	0.85	0.59	0.70	0.28	0.26	0.86	0.33	0.72	0.89	0.71	0.71	0.62	0.66

Mutation	Mutation Energy
M7E + E4D/W/Y	-1.77, -2.54, 2.45
M7E + L21W	-2.15
M7E + E30W	-2.11
M7E + N36R/Q/E/L/K/W/Y	-2.63, -1.67, 2.58, -2.03, -2.08, 0.61, -1.89
M7E + W37D	-2.02
M7E + M43E/H/W/Y	-2.86, -3.2, **-3.44**, **-4.33**
M7E + L44W/Y	-3.14, -2.7

Table S3. Mutational energy of M7E-based double point mutation based on the inhibitor AHB2 (kcal/mol).

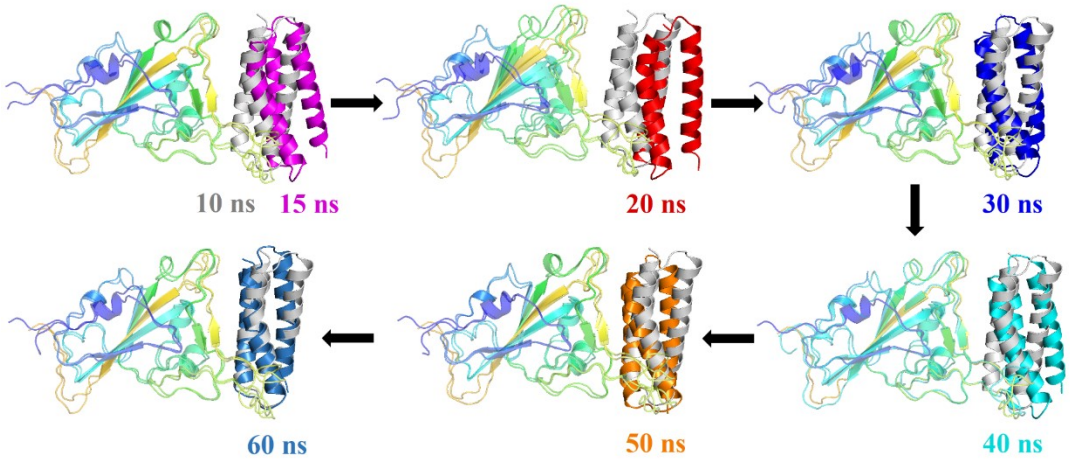

Fig S1. Superimposed 3D structures of the inhibitor M7E+M43Y with its the structure (t = 10 ns) at different MD simulation periods (t = 15, 20, 30, 40, 50, 60 ns). The triple helix inhibitor is represented by different colors at different MD simulation times, and the RBD^{O} region is represented as transparent.

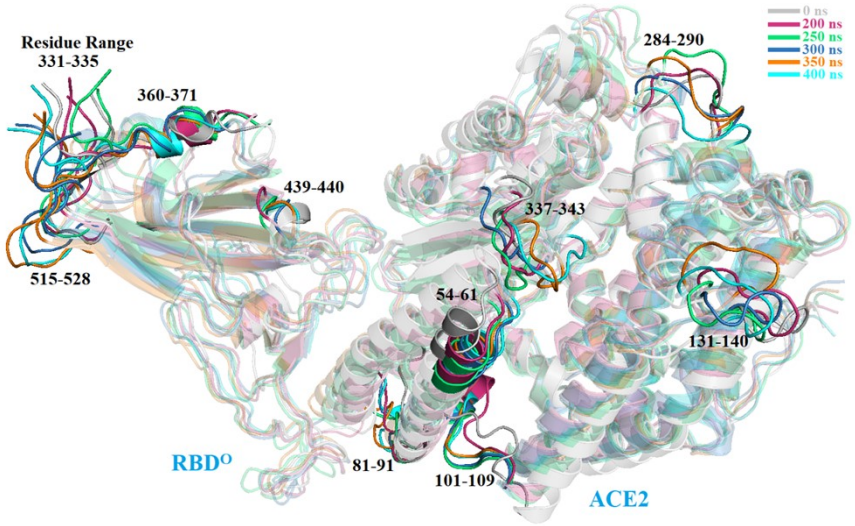

Fig S2. Superimposed 3D structures of the RBD^{O}-ACE2 complex (Run_1 MD simulation) with its initial structure (t = 0 ns) at different MD simulation periods (t = 200, 250, 300, 350, and 400 ns). Regions with large RMSD fluctuations are represented as solids, while other regions are represented as transparent.

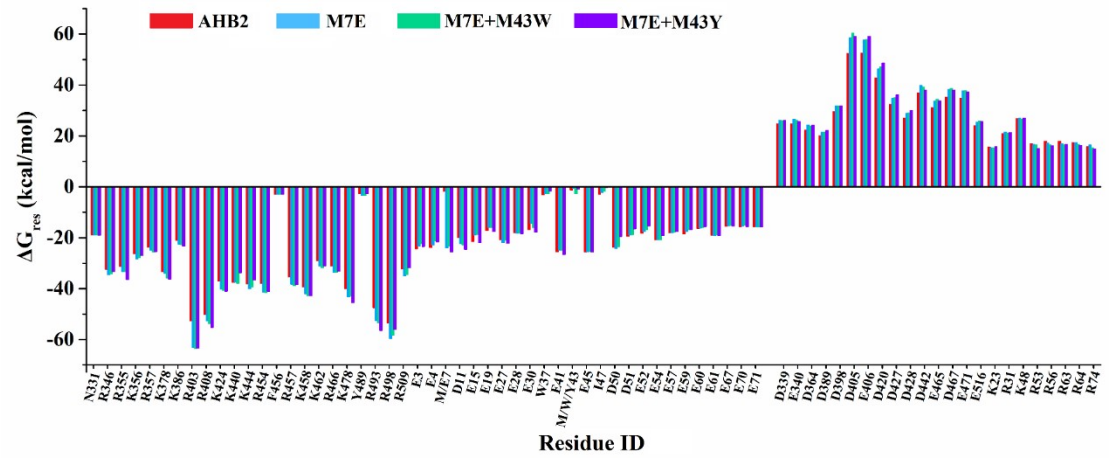

Fig S3. Residue-based energy decomposition on critical residues for AHB2 (red), M7E (blue), M7E+M43W (green), and M7E+M43W (violet) systems. ΔG_{res} < -2.0 kcal/mol and ΔG_{res} > 2.0 kcal/mol are considered as crucial residues.

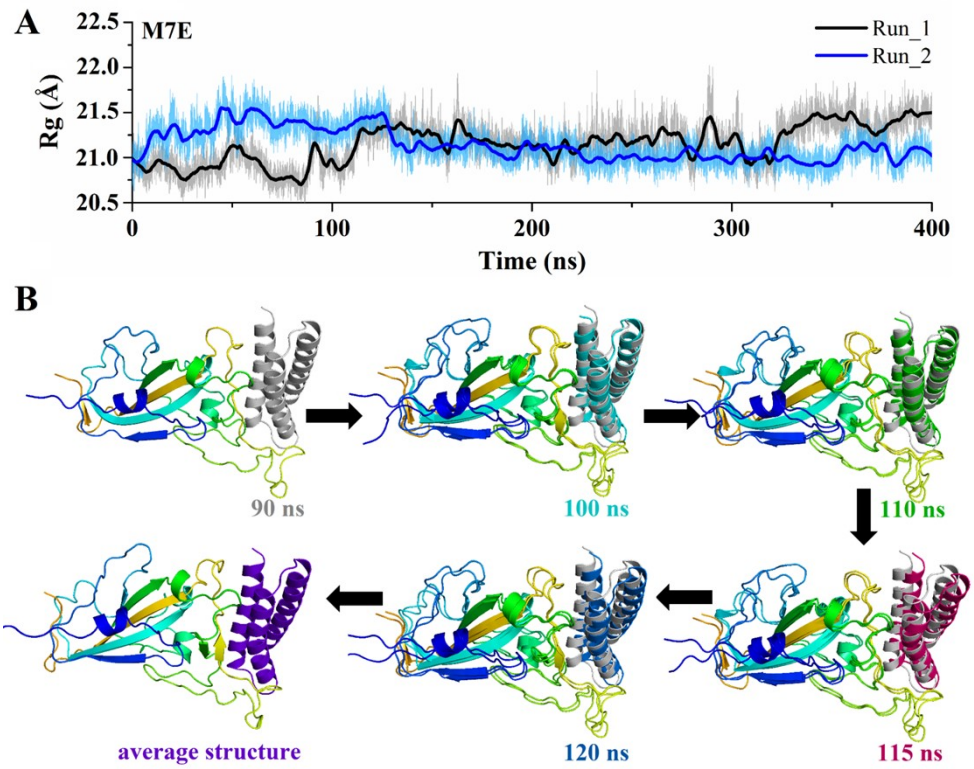

Fig S4. (A) Rg plot of the M7E complex for two parallel MD simulations. The transparent curves are the Rg value fluctuations and the smooth solid curves represent the average Rg value. The average Rg curves were plotted by the Savitzky–Golay method with the adjacent 1000 data points. (B) Superimposed 3D structures of the M7E complex with its the structure (t = 90 ns) at different MD simulation periods (t = 100, 110, 115, and 120 ns). The triple helix inhibitor is represented by different colors at different MD simulation times.

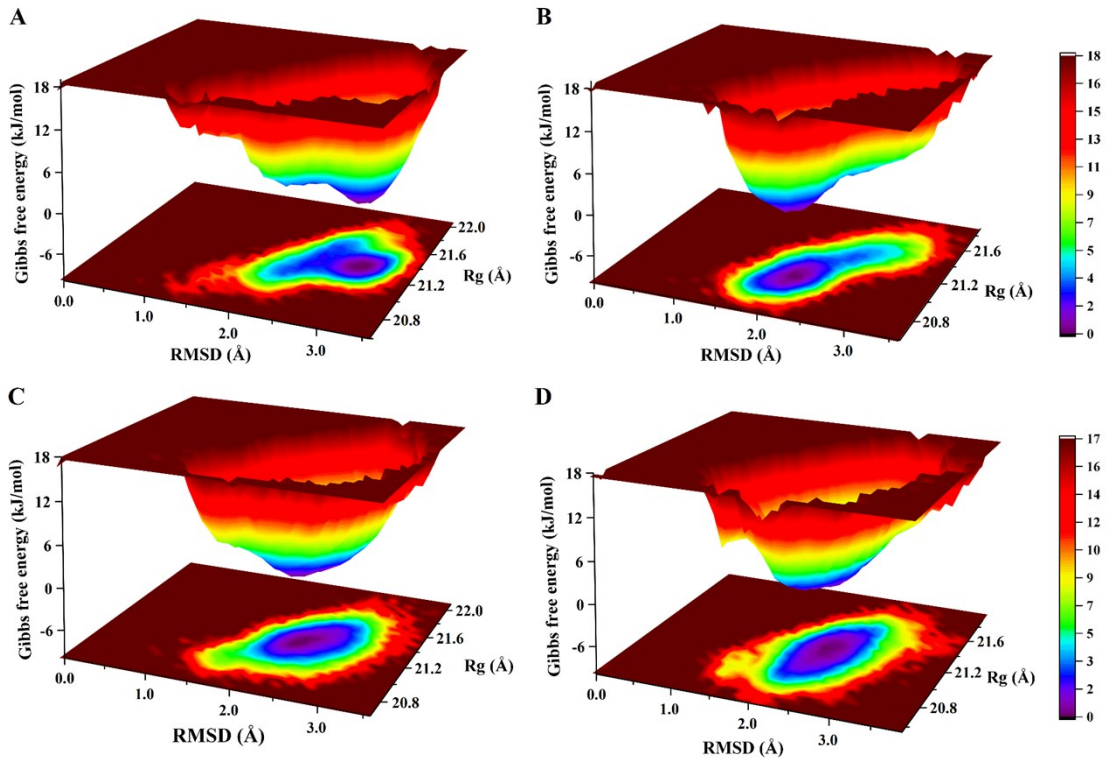

Figure S5. Free energy landscapes of (A) AHB2, (B) M7E, (C) M7E+M43W, and (D) M7E+M43Y complexes during replica MD simulation.

References

1. Amadei, A.; Linssen, A. B.; Berendsen, H. J., Essential dynamics of proteins. *Proteins* **1993,** 17 (4), 412−425.

2. Sun, H.; Li, Y.; Li, D.; Hou, T., Insight into Crizotinib Resistance Mechanisms Caused by Three Mutations in ALK Tyrosine Kinase using Free Energy Calculation Approaches. *J. Chem. Inf. Model.* **2013,** 53 (9), 2376−2389.

3. Kumari, R.; Kumar, R.; Lynn, A.; Open Source Drug Discovery, C., g_mmpbsa-A GROMACS Tool for High-Throughput MM-PBSA Calculations. *J. Chem. Inf. Model.* **2014,** 54 (7), 1951−1962.

4. Genheden, S.; Ryde, U., The MM/PBSA and MM/GBSA methods to estimate ligand-binding affinities. *Expert Opin. Drug Discov.* **2015,** 10 (5), 449−461.

5. Sun, H.; Li, Y.; Shen, M.; Tian, S.; Xu, L.; Pan, P.; Guan, Y.; Hou, T., Assessing the performance of MM/PBSA and MM/GBSA methods. 5. Improved docking performance using high solute dielectric constant MM/GBSA and MM/PBSA rescoring. *Phys. Chem. Chem. Phys.* **2014,** 16 (40), 22035−22045.

6. Wang, E.; Weng, G.; Sun, H.; Du, H.; Zhu, F.; Chen, F.; Wang, Z.; Hou, T., Assessing the performance of the MM/PBSA and MM/GBSA methods. 10. Impacts of enhanced sampling and variable dielectric model on protein-protein Interactions. *Phys. Chem. Chem. Phys.* **2019,** 21 (35), 18958−18969.