Supplementary Information (SI)

Molecular dynamics simulations reveal a strong binding of colossolactone H to the EGFR inactive conformation

Duc Toan Truong,^{1,2} Kiet Ho,³ Thai Chinh Tam,⁴ Do Thi Mai Dung,⁴ Nguyen Minh Tam,⁵ Viet Bac T. Phung,⁶ and Minh Tho Nguyen^{1,2,*}

- ¹ Laboratory for Chemical Computation and Modeling, Institute for Computational Science and Artificial Intelligence, Van Lang University, Ho Chi Minh City, 70000 Vietnam
- ² Faculty of Applied Technology, School of Technology, Van Lang University, Ho Chi Minh City, 70000 Vietnam
- ³ Institute of Computational Science and Technology (ICST), Ho Chi Minh City, Vietnam
- ⁴ Hanoi University of Pharmacy, 13-15 Le Thanh Tong, Hoan Kiem 11021, Ha Noi
- ⁵ Faculty of Basic Sciences, University of Phan Thiet, 225 Nguyen Thong, Phan Thiet City, Binh Thuan, Vietnam
- ⁶Center for Environmental Intelligence and College of Engineering and Computer Science, VinUniversity, Gia Lam District, Hanoi 10000, Vietnam

Emails: toan.truongduc@vlu.edu.vn; minhtho.nguyen@vlu.edu.vn

Some of the supporting data for this study are provided via a shared Google Drive folder due to the large file sizes (up to 120 MB). All files are organized to facilitate reuse and independent verification of the study's findings.

This folder includes the following:

- **INPUT files for Steered Molecular Dynamics (SMD):** These contain the pre-aligned pulling directions along the Z-axis for each ligand–protein complex, allowing users to easily reproduce the simulations and visualize the trajectory using PyMOL.
- **Simulation output data:** Time-dependent force profiles, pulling work, and non-equilibrium energies for all ligands.
- Umbrella Sampling data: Free energy profiles, transition state estimations, and corresponding RMSD plots are included to complement the MM-PBSA results and highlight differences in methodological outcomes.

Ν	Index	Gly721	Ala722	Lys745	Thr790	Gln791	Leu792	Met793	Cys797	Arg841	Leu858
1	Ι	0.50	0.84	0.23	0.23	0.48	0.47	0.40	0.24	0.39	0.82
2	II	0.51	0.87	0.22	0.22	0.49	0.48	0.40	0.25	0.41	0.78
3	III	0.49	0.79	0.21	0.22	0.50	0.48	0.39	0.24	0.38	0.81
4	IV	0.25	0.43	0.27	0.31	0.48	0.50	0.42	0.24	0.33	0.84
5	V	0.27	0.42	0.27	0.44	0.54	0.36	0.35	0.24	0.28	0.76
6	VI	0.25	0.31	0.31	0.30	0.49	0.34	0.31	0.23	0.26	0.76
7	VII	0.27	0.43	0.39	0.28	0.43	0.41	0.36	0.25	0.29	0.95
8	VIII	0.37	0.74	0.23	0.24	0.34	0.28	0.25	0.23	0.49	0.94
9	А	0.42	0.67	0.23	0.24	0.52	0.36	0.30	0.23	0.40	0.83
10	В	0.40	0.67	0.23	0.24	0.51	0.33	0.28	0.24	0.40	0.82
11	С	0.52	0.64	0.66	0.61	0.71	0.67	0.60	0.52	0.53	1.04
12	D	0.57	0.97	0.23	0.26	0.39	0.29	0.25	0.24	0.50	0.81
13	Е	0.34	0.71	0.23	0.26	0.40	0.32	0.28	0.23	0.46	0.83
14	F	0.45	0.55	0.32	0.48	0.44	0.25	0.21	0.19	0.22	0.77
15	G	0.58	0.96	0.25	0.25	0.49	0.24	0.23	0.21	0.33	0.86
16	Н	0.58	0.92	0.25	0.24	0.48	0.25	0.22	0.21	0.38	0.84

Table S1: Minimum distance (nm) between 16 ligands and 10 key residues forming the ATP binding region (active state). Data was averaged from the last 20ns of 100ns MD simulations method.

Ν	Index	Gly721	Ala722	Lys745	Thr790	Gln791	Leu792	Met793	Cys797	Arg841	Leu858
1	Ι	0.46	0.57	0.44	0.25	0.30	0.29	0.21	0.25	0.24	1.08
2	II	0.60	0.79	0.37	0.25	0.38	0.28	0.26	0.24	0.27	1.19
3	III	0.56	0.77	0.40	0.23	0.30	0.25	0.21	0.27	0.22	1.08
4	IV	0.42	0.63	0.77	1.00	1.09	0.76	0.71	0.20	0.33	1.40
5	V	0.29	0.40	0.63	0.95	0.92	0.64	0.58	0.19	0.25	1.41
6	VI	0.48	0.66	0.53	0.77	0.89	0.58	0.64	0.23	0.26	1.30
7	VII	0.54	0.67	0.49	0.30	0.32	0.28	0.22	0.23	0.26	1.15
8	VIII	0.50	0.72	0.23	0.24	0.47	0.39	0.40	0.21	0.23	1.02
9	А	0.28	0.49	0.41	0.85	0.97	0.79	0.93	0.25	0.26	1.18
10	В	0.28	0.47	0.41	0.67	0.75	0.56	0.50	0.22	0.19	0.99
11	С	0.63	0.89	0.24	0.23	0.50	0.39	0.40	0.22	0.23	1.03
12	D	0.76	0.98	0.44	0.24	0.27	0.26	0.20	0.23	0.29	1.28
13	Е	0.43	0.66	0.90	1.25	1.35	0.90	1.04	0.56	0.57	1.59
14	F	0.47	0.67	0.43	0.58	0.67	0.54	0.53	0.22	0.25	1.22
15	G	0.42	0.60	0.72	0.80	0.80	0.52	0.49	0.25	0.27	1.34
16	Н	0.47	0.68	0.42	0.38	0.45	0.37	0.48	0.25	0.23	1.08

Table S2: Minimum distance (nm) between 16 ligands and 10 key residues forming the ATP binding region (inactive state). Data was averaged from the last 20ns of 100ns MD simulations method.

Ν	Index	Lys745	Ile759	Glu762	Ala763	Met766	Cys775	Leu777	Leu788	Thr790	Thr854	Asp855	Phe856
1	Ι	0.28	0.21	0.24	0.37	0.22	0.27	0.23	0.23	0.28	0.31	0.24	0.23
2	II	0.22	0.21	0.23	0.35	0.23	0.48	0.26	0.22	0.24	0.30	0.25	0.42
3	III	0.36	0.21	0.26	0.38	0.22	0.28	0.26	0.23	0.33	0.33	0.25	0.22
4	IV	0.23	0.22	0.23	0.42	0.23	0.33	0.25	0.25	0.27	0.34	0.24	0.35
5	V	0.28	0.22	0.23	0.45	0.29	0.87	0.53	0.26	0.71	0.69	0.40	0.29
6	VI	0.29	0.23	0.24	0.26	0.28	0.71	0.34	0.24	0.59	0.52	0.26	0.39
7	VII	0.24	0.24	0.44	0.69	0.66	1.11	0.73	0.33	0.81	0.82	0.54	0.81
8	VIII	0.23	0.22	0.22	0.28	0.22	0.37	0.23	0.22	0.24	0.25	0.24	0.26
9	Α	0.25	0.22	0.24	0.44	0.24	0.70	0.38	0.25	0.56	0.50	0.21	0.33
10	В	0.26	0.21	0.23	0.39	0.22	0.25	0.25	0.23	0.25	0.30	0.23	0.26
11	С	0.21	0.22	0.23	0.42	0.22	0.28	0.27	0.23	0.29	0.32	0.22	0.23
12	D	0.25	0.22	0.24	0.35	0.22	0.25	0.23	0.23	0.25	0.27	0.23	0.32
13	Е	0.27	0.21	0.24	0.26	0.23	0.27	0.22	0.22	0.22	0.38	0.27	0.27
14	F	0.28	0.20	0.23	0.25	0.22	0.26	0.22	0.22	0.22	0.36	0.28	0.27
15	G	0.27	0.22	0.24	0.26	0.22	0.25	0.22	0.22	0.23	0.35	0.23	0.30
16	Н	0.26	0.22	0.22	0.28	0.22	0.23	0.23	0.22	0.23	0.27	0.26	0.27

Table S3: Minimum distance (nm) between ligand and residues locating in the allosteric pocket. Data was averaged from the last 20ns of 100ns MD simulations method

Table S4: Distribution of number of contact, electrostatic potential energy (kcal/mol) and Van der Walls potential energy (kcal/mol) between *colossolactone* – H and EGFR protein as well as *colossolactone* – H and 40 residues locating surround the allosteric region.

		Numcount	Elec	Vdw
1	Protein	3270	-5.65	-52.15
2	ALA722	4	-0.02	-0.14
3	PHE723	82	0.07	-1.51
4	ALA743	4	-0.03	-0.08
5	LYS745	223	1.78	-2.51
6	<i>LEU747</i>	264	0.03	-3.33
7	ARG748	4	0.02	-0.1
8	<i>GLU749</i>	83	0.05	-1.47
9	ALA750	4	-0.05	-0.16
10	THR751	5	0	-0.12
11	ALA755	15	-0.05	-0.22
12	<i>GLU758</i>	57	-0.64	-0.74
13	ILE759	304	-0.1	-4.23
14	<i>LEU760</i>	8	-0.03	-0.2
15	<i>GLU762</i>	220	-0.13	-3.16
16	ALA763	118	0.08	-1.45
17	MET766	259	0.62	-4.28
18	ALA767	5	0.05	-0.16
19	VAL769	3	-0.02	-0.11
20	CYS775	70	-0.05	-1.1
21	ARG776	54	-2.27	-1.74
22	<i>LEU777</i>	231	0.18	-3.52
23	<i>LEU778</i>	4	0.01	-0.14
24	<i>ILE780</i>	9	0	-0.16
25	VAL786	7	0.01	-0.22
26	<i>LEU788</i>	311	-0.24	-3.47
27	ILE789	10	-0.02	-0.29
28	THR790	125	-0.05	-1.44
29	GLN791	4	-0.02	-0.09
30	HIS835	5	-0.05	-0.11
31	ASP837	4	0.05	-0.14
32	<i>LEU844</i>	19	0.02	-0.16
33	ILE853	3	0	-0.09
34	THR854	78	-0.52	-0.95
35	ASP855	212	-1.71	-4.5
36	PHE856	74	-0.31	-1.65
37	GLY857	50	-0.06	-1.03
38	<i>LEU858</i>	103	-0.8	-2.54
39	ALA859	153	-0.03	-2.37
40	LYS860	30	-1.29	-0.81
41	LEU861	42	-0.11	-0.5

		Numcount			ļ	Elec				Vdw		
		<i>G1</i>	G2	<i>G3</i>		<i>G1</i>	G2	G3		<i>G1</i>	G2	G3
1	Protein	878	1356	1071		-2.44	-0.06	-4.16	ĺ	-14.35	-19.21	-19.23
2	ALA722	4	0	0		-0.02	0	0		-0.14	0	0
3	<i>PHE723</i>	67	15	0	ĺ	0.06	0	0		-1.18	-0.34	-0.01
4	ALA743	0	0	3		0	0	0.02		0	0	-0.08
5	LYS745	9	124	76		0.02	0.16	0.55		-0.16	-1.31	-1.06
6	<i>LEU747</i>	165	113	1		-0.07	0.07	0		-2.21	-1.27	-0.07
7	ARG748	4	0	0		0.02	0	0]	-0.1	0	0
8	GLU749	83	0	0		0.05	0	0]	-1.46	-0.01	0
9	ALA750	4	0	0		-0.05	0	0]	-0.16	0	0
10	THR751	5	0	0		0	0	0]	-0.12	0	0
11	ALA755	15	2	0		-0.05	-0.06	0		-0.22	-0.02	0
12	GLU758	52	21	0		-0.59	-0.6	0		-0.62	-0.25	0
13	ILE759	209	142	2		0.08	-0.02	-0.02		-2.82	-1.89	-0.07
14	<i>LEU760</i>	2	8	0		-0.01	-0.03	0		-0.06	-0.15	-0.02
15	GLU762	75	158	21		-0.07	-0.09	-0.01		-0.99	-2.17	-0.36
16	ALA763	7	75	41		0.01	0.05	0.03		-0.09	-0.76	-0.66
17	<i>MET766</i>	1	111	145		0	0.03	0.5]	-0.07	-1.3	-2.92
18	ALA767	0	0	5		0	0	0.05		0	-0.01	-0.15
19	VAL769	0	0	3		0	0	-0.02		0	0	-0.11
20	CYS775	0	0	70		0	0	-0.05		0	0	-1.1
21	ARG776	0	0	53	Į	0	0	-2.27	ļ	0	-0.01	-1.74
22	<i>LEU777</i>	0	25	194		0	0.01	-0.15	ļ	-0.01	-0.29	-3.21
23	<i>LEU778</i>	0	0	3		0	0	-0.02	ļ	0	0	-0.14
24	ILE780	0	7	3		0	0	0	ļ	-0.01	-0.08	-0.08
25	VAL786	1	7	0		0	0.01	0	ļ	-0.08	-0.13	-0.02
26	<i>LEU</i> 788	16	165	120		0	0.06	-0.81		-0.22	-1.5	-1.84
27	ILE789	0	0	4		0	0	0.16		0	-0.01	-0.28
28	THR790	0	0	114		0	0	-0.12		0	-0.03	-1.4
29	GLN791	0	0	4		0	0	-0.02		0	0	-0.09
30	HIS835	0	5	0		0	-0.03	0		0	-0.1	0
31	ASP837	0	4	0		0	0.11	0	ļ	0	-0.11	0
32	<i>LEU844</i>	0	0	19		0	0	0.02	ļ	0	0	-0.16
33	ILE853	0	0	3		0	0	0		0	0	-0.09
34	THR854	0	3	75		0	-0.05	-0.5		0	-0.14	-0.79
35	ASP855	0	107	93		0	0.52	-1.44		-0.06	-2.36	-1.67
36	PHE856	0	52	16		0	0.33	-0.06		-0.03	-0.9	-0.5
37	GLY857	2	42	0		0.01	0.04	0		-0.08	-0.8	-0.02
38	LEU858	51	51	0		-0.67	0.2	0		-1.31	-1.19	-0.01
39	ALA859	71	78	0		0.12	-0.45	0		-1.17	-1.17	-0.01
40	LYS860	28	2	0		-1.28	-0.03	0		-0.7	-0.11	0
41	LEU861	4	32	0		0	-0.19	0		-0.08	-0.36	-0.01

Table S5: Distribution of number of contact, electrostatic potential energy (kcal/mol) and Van der Walls potential energy (kcal/mol) between fragments of *colossolactone* – H and EGFR protein as well as fragments of *colossolactone* – H and 40 residues locating surround the allosteric region.

Figure S1: HOMO-LUMO representation of 16 colossolactones, obtained at

B3LYP/6-31+G(p,d) computations.



 $E_{HOMO} = -6.168 \text{ eV}$

 $E_{LUMO} = -1.429 \text{ eV}$

Colossolactone I



Colossolactone II



Colossolactone V



Colossolactone VIII







Colossolactone F



 $E_{HOMO} = -6.276 \text{ eV}$

 $E_{LUMO} = -2.362 \text{ eV}$

Colossolactone G



Colossolactone H

Figure S2: Initial conformations obtained by docking 16 ligands (active state) into the ATP binding region of EGFR wide type.



Colossolactone I



Colossolactone II



Colossolactone III



Colossolactone IV



Colossolactone V



Colossolactone VI



Colossolactone VII



Colossolactone VIII





Colossolactone B



Colossolactone C



Colossolactone D



Colossolactone E



Colossolactone F





Colossolactone H

Figure S3: Initial conformations obtained by docking 16 ligands into the ATP binding region of EGFR wide type (inactive state).



Colossolactone I



Colossolactone II



Colossolactone III



Colossolactone IV



Colossolactone V



Colossolactone VI



Colossolactone VII


Colossolactone VIII



Colossolactone A



Colossolactone B



Colossolactone C



Colossolactone D



Colossolactone E



Colossolactone F



Colossolactone G



Colossolactone H

Figure S4: Initial conformations obtained by docking ligands into the allosteric binding region of EGFR wide type.



Colossolactone I



Colossolactone II



Colossolactone III



Colossolactone IV



Colossolactone V

.



Colossolactone VI



Colossolactone VII



Colossolactone VIII



Colossolactone A



Colossolactone B



Colossolactone C



Colossolactone D



Colossolactone E



Colossolactone F



Colossolactone G



Colossolactone H

Figure S5: Representative illustration for 16 colossolactones in ATP-binding site (active state). The Gibbs energy landscape between SASA and the contact count is plotted in the left-hand side. The interaction between protein and ligand in representative configuration is described in the right-hand side.



Colossolactone I



Colossolactone II



Colossolactone III





Colossolactone V



Colossolactone VI



Colossolactone VII



Colossolactone VIII



Colossolactone A



Colossolactone B



Colossolactone C


Colossolactone D



Colossolactone E



Colossolactone F



Colossolactone G



Colossolactone H

Figure S6: Representative illustration for 16 colossolactones in ATP-binding site (inactive state). The Gibbs energy landscape between SASA and the contact count is plotted in the left-hand side. The interaction between protein and ligand in representative configuration is described in the right-hand side.





Colossolactone II



Colossolactone III





Colossolactone V



Colossolactone VI



Colossolactone VII



Colossolactone VIII



Colossolactone A



Colossolactone B



Colossolactone C



Colossolactone D



Colossolactone E



Colossolactone F



Colossolactone G



Figure S7: Representative illustration for 16 colossolactones in allosteric binding site. The left hand side plot is Gibbs energy landscape between SASA and the contact count. The right hand side describes the interaction between protein and ligand in representative configuration.



Gibbs Energy Landscape



Colossolactone II



Colossolactone III



Colossolactone IV



Colossolactone V







Colossolactone VIII



Colossolactone A



Colossolactone B



Colossolactone C



Colossolactone D



Colossolactone E



Colossolactone F



Colossolactone G


##