Electronic Supplementary Material (ESI) for Molecular BioSystems.

Supporting Information

Molecular simulations studies on the binding selectivity of 2-anilino-4-(thiazol-5-yl)-pyrimidines in complexes with CDK2 and CDK7

Tahir Ali Chohan, Hai-Yan Qian, You-Lu Pan, and Jian-Zhong Chen*

College of Pharmaceutical Sciences, Zhejiang University, Hangzhou, Zhejiang 310058, China

* Corresponding Author：email: chjz@zju.edu.cn (J.-Z. Chen)

Table S1. Surflex score of docked ligand CP1 for CDK2 and CP1-3 for CDK7

Docking complex	ΔG_{exp}	CScore [a]	Crash score[b]	Polar score[c]	G score[d]	PMF score[e]	D score[f]	Chem score[g]	Amino acid interaction
CDK2-CP1	-13.66	7.20	-1.02	1.89	-249.57	-62.04	-150.91	-29.01	Leu83, Lys89, Asp86
CDK7-CP1	-8.27	4.49	-1.02	2.15	-136.05	-43.39	-108.54	-24.65	Met94
CDK7-CP2	-9.79	5.41	-1.10	1.68	-200.73	-24.64	-200.73	-34.87	Met94
CDK7-CP3	-11.85	5.55	-0.47	1.82	-110.04	-66.02	-193.54	-25.25	Met94

[a]**CScore** is a consensus scoring which uses multiple types of scoring functions to rank the affinity of ligands, [b]**Crash**-score revealing the inappropriate penetration into the binding site, [c] **Polar** region of the ligand, [d] **G-score** showing hydrogen bonding, complex (ligand-protein), and internal (ligand-ligand) energies, [e] **PMF-score** indicating the Helmholtz free energies of interactions for protein-ligand atom pairs (Potential of Mean Force, PMF), [f] **D-score** for charge and van der Waals interactions between the protein and the ligand, [g] **Chem-score** points for hydrogen bonding, lipophilic contact, and rotational entropy, along with an intercept term.

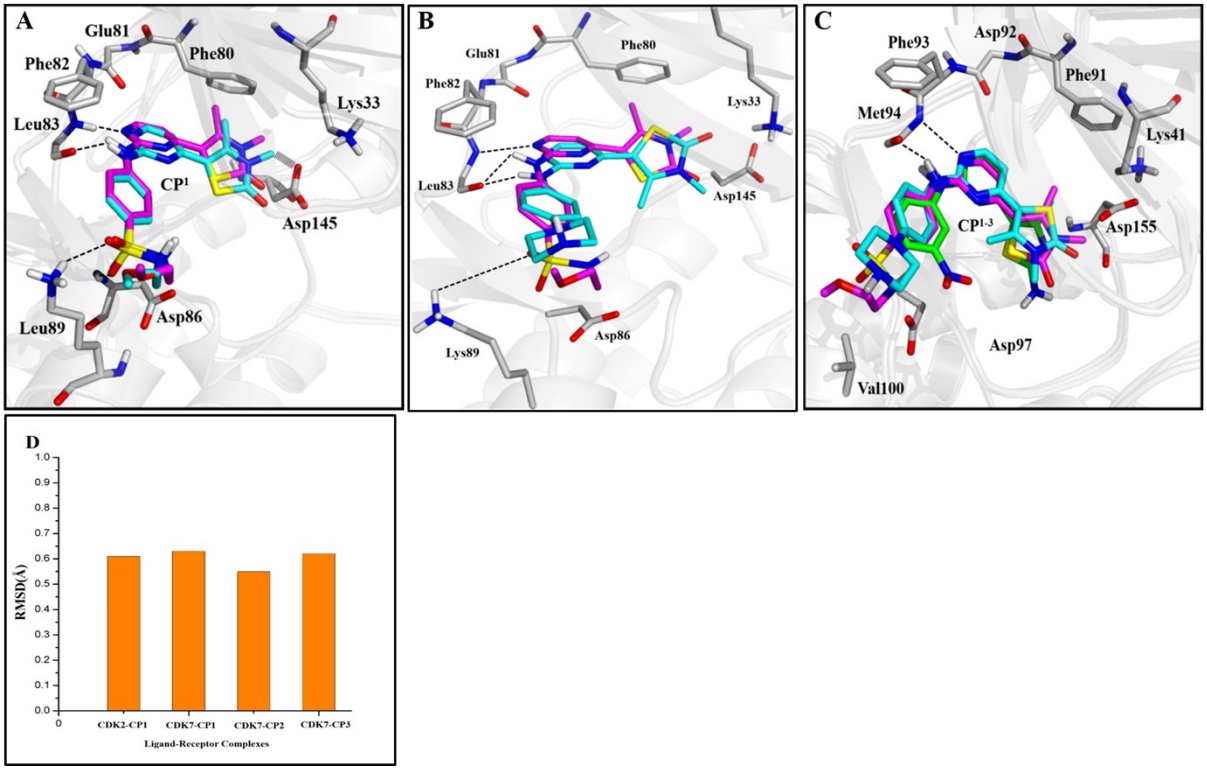

Figure S1. Comparison of binding modes of docked ligand with their starting conformations. **(A)** Superimposition of the docked ligand CP1 (Purple) and experimental conformation of CDK2 (Cyan). **(B)** Superimposition of the docked ligand CP1 (Purple) and experimental conformation of CP3 in the active site of CDK2 (Cyan). **(C)** Alignment of docked compounds CP1 (Purple), CP2 (Green) and CP3 (Cyan) within the active site of CDK7. The main residues constituting the active site are identified with a three letter representation. **(D)** The average means RMSD using bars chat.

```
CDK1     ---------MEDYTKIEKIGEGTYGVVYKGRHKTTGQVVAMKKIRLESE---EEGVPSTA 48
CDK2     ---------MENFQKVEKIGEGTYGVVYKARNKLTGEVVALKKIRLDTE---TEGVPSTA 48
CDK3     FCFPGSSVAMDMFQKVEKIGEGTYGVVYKAKNRETGQLVALKKIRLDLE---MEGVPSTA 57
CDK4     -------MATSRYEPVAEIGVGAYGTVYKARDPHSGHFVALKSVRVPNGGGGGGGLPIST 53
CDK7     -MALDVKSRAKRYEKLDFLGEGQFATVYKARDKNTNQIVAIKKIKLGHRSEAKDGINRTA 59
                .  :  :  :* * :..***.:.  : ..**:*.:::       *:  ::

CDK1     IREISLLKE---LRHPNIVSLQDVLMQ-----DSRLYLIFEFLSMDLKKYLDSIPPGQYM 100
CDK2     IREISLLKE---LNHPNIVKLLDVIHT-----ENKLYLVFEFLHQDLKKFMDASA-LTGI 99
CDK3     IREISLLKE---LKHPNIVRLLDVVHN-----ERKLYLVFEFLSQDLKKYMDSTP-GSEL 108
CDK4     VREVALLRRLEAFEHPNVVRLMDVCATSRTDREIKVTLVFEHVDQDLRTYLDKAP-PPGL 112
CDK7     LREIKLLQE---LSHPNIIGLLDAFGH-----KSNISLVFDFMETDLEVIIKDNS--LVL 109
        :**: **:.   : ***:: * *.       . .: *:*:.:  **.  :.      :

CDK1     DSSLVKSYLYQILQGIVFCHSRRVLHRDLKPQNLLIDDKGTIKLADFGLARAFGIPIRVY 160
CDK2     PLPLIKSYLFQLLQGLAFCHSHRVLHRDLKPQNLLINTEGAIKLADFGLARAFGVPVRTY 159
CDK3     PLHLIKSYLFQLLQGVSFCHSHRVIHRDLKPQNLLINELGAIKLADFGLARAFGVPLRTY 168
CDK4     PAETIKDLMRQFLRGLDFLHANCIVHRDLKPENILVTSGGTVKLADFGLARIYSYQM-AL 171
CDK7     TPSHIKAYMLMTLQGLEYLHQHWILHRDLKPNNLLLDENGVLKLADFGLAKSFGSPNRAY 169
           :*  :   *:*: : * . ::******:*:*:   *.:********: :.    .

CDK1     THEVVTLWYRSPEVLLGSARYSTPVDIWSIGTIFAELATKKPLFHGDSEIDQLFRIFRAL 220
CDK2     THEVVTLWYRAPEILLGSKYYSTAVDIWSLGCIFAEMVTRRALFPGDSEIDQLFRIFRTL 219
CDK3     THEVVTLWYRAPEILLGSKFYTTAVDIWSIGCIFAEMVTRKALFPGDSEIDQLFRIFRML 228
CDK4     TPVVVTLWYRAPEVLLQS-TYATPVDMWSVGCIFAEMFRRKPLFCGNSEADQLGKIFDLI 230
CDK7     THQVVTRWYRAPELLFGARMYGVGVDMWAVGCILAELLLRVPFLPGDSDLDQLTRIFETL 229
        *  *** ***:**:*: :  * . **:*::* *:**:  :  :: *:*: *** :**  :

CDK1     GTPNNEVWPEVESLQDYKNTFPKWKPGSLASHVKNLDENGLDLLSKMLIYDPAKRISGKM 280
CDK2     GTPDEVVWPGVTSMPDYKPSFPKWARQDFSKVVPPLDEDGRSLLSQMLHYDPNKRISAKA 279
CDK3     GTPSEDTWPGVTQLPDYKGSFPKWTRKGLEEIVPNLEPEGRDLLMQLLQYDPSQRITAKT 288
CDK4     GLPPEDDWPRDVSLPR--GAFPPRGPRPVQSVVPEMEESGAQLLLEMLTFNPHKRISAFR 288
CDK7     GTPTEEQWPDMCSLPDYVT-FKSFPGIPLHHIFSAAGDDLLDLIQGLFLFNPCARITATQ 288
        * * :  **   .:     *      .   .    .  .*:  :: ::*  **:.

CDK1     ALNHPYFNDLDNQIKKM----------------------------------------  297
CDK2     ALAHPFFQDVTKPVPHLRL--------------------------------------  298
CDK3     ALAHPYFSSPE-PSPAARQYVLQR----FRH--------------------------  314
CDK4     ALQHSYLHKDEGNPE------------------------------------------  303
CDK7     ALKMKYFSNRPGPTPGCQLPRPNCPVETLKEQSNPALAIKRKRTEALEQGGLPKKLIF  347
```

Figure S2. The sequence alignments of CDK1, CDK2, CDK3, CDK4 and CDK7 generated by Clustal W. at http://www.ebi.ac.uk/Tools/msa/clustalw2/. In the sequences, an asterisk (*) indicates an identical or conserved residue; a colon (:) indicates conserved substitutions; a stop (.) indicates semi-conserved substitutions.

Table S2. Result of binding site comparison of CDK2 and CDK7, based on EasyMIFs using CMET probes.

Rank	Total energy (kcal·mol^{-1})		Volume (Å)		Residues contributing to each cluster (identified by subsite)	
	CDK2	CDK7	CDK2	CDK7	CDK2	CDK7
1	-1622.97	-955.83	123	75	I10, G11, G13, V18, K33, V64, F80, Q131, N132, A144, D145	E20, T96, D97, E99, V100, K139, P140, N141
2	-797.57	-841.43	66	57	E8, K9, I10, G11, K20, F82, H84	L18, K41, F91, D92, F93, M94, E95, T96, D97, L144, A154, D155
3	-125.81	-561.12	8	50	I10, E81, F82, L83, A31, L134	F93, E95, T96, E147

Table S3. Quantum chemical descriptors based upon DFT calculations used for MESP for compounds CP1, CP2, and CP3.

Quantum descriptors	Gas phase			Solvent phase (Aqueous)		
	CP1	**CP2**	CP3	CP1	CP2	CP3
E_{LUMO} (eV)	-0.080	-0.105	-0.065	-0.071	-0.117	-0.067
E_{HOMO} (eV)	-0.222	-0.215	-0.190	-0.217	-0.215	-0.188
Total dipole moment μ (D)	11.52	4.41	5.19	15.75	5.85	6.85

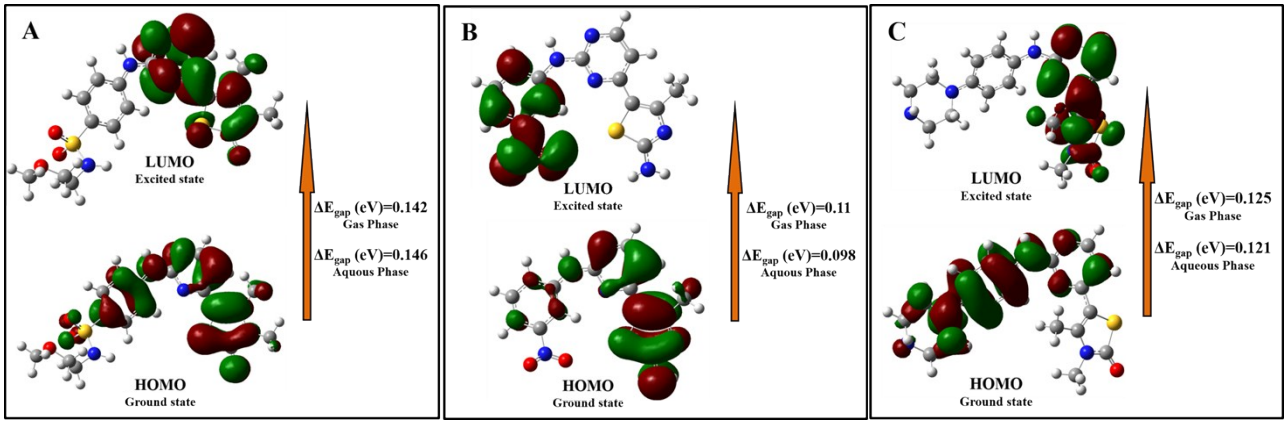

Figure S3. Molecular orbital for the HOMO–LUMO plot of **(A)** CP1, **(B)** CP2, and **(C)** CP3 with B3LYP/6-31G (d,p)

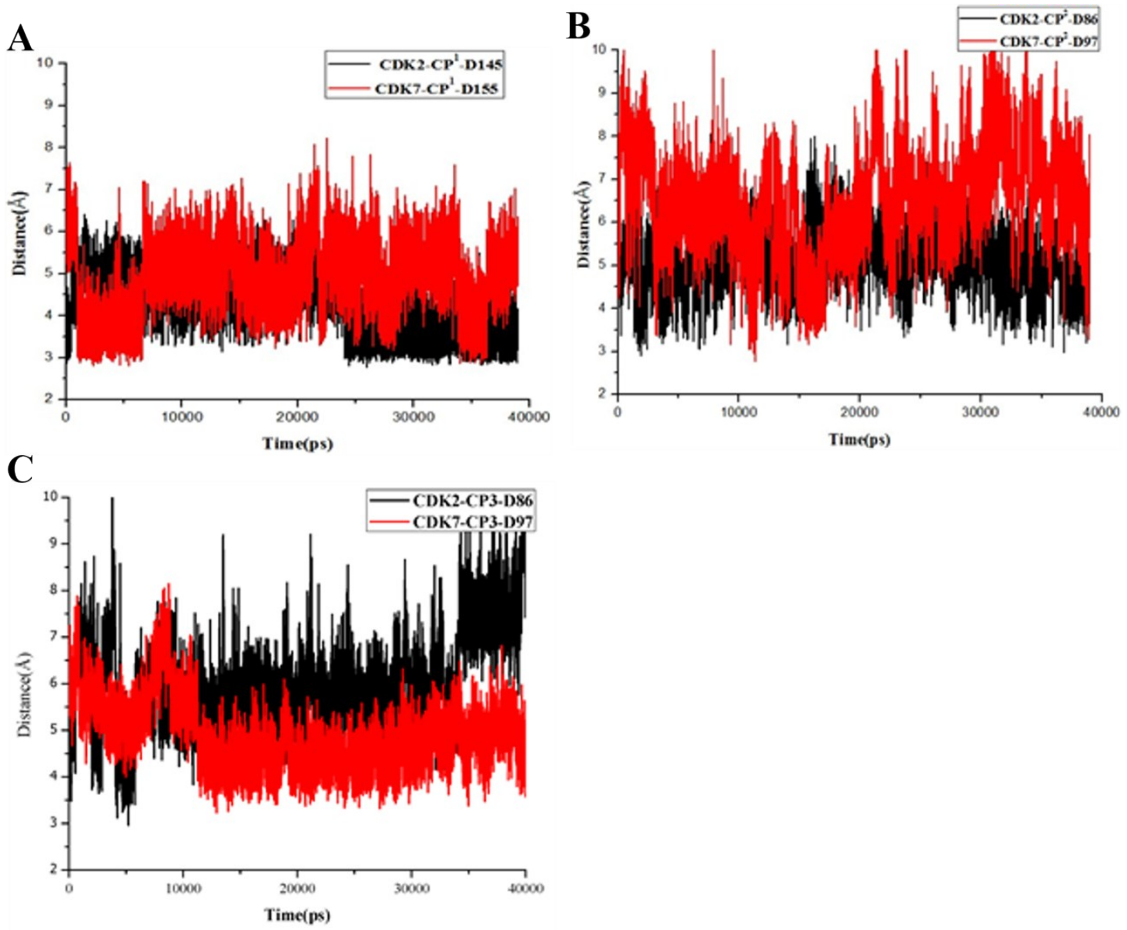

Figure S4. (A) The distance of CDK2-Asp145 (OD1) and CDK7-Asp155 (OD1) from C27 atom of inhibitor CP1, plotted as a function of time in CDK2-CP1 and CDK7-CP1. **(B)** The distance of CDK2-Asp86 (OD1) and CDK7-Asp97 (OD1) from N_{17} atom of inhibitor CP2, plotted as a function of time in CDK2-CP2 and CDK7-CP2. **(C)** The distance of CDK2-Asp86 (CB) and CDK7-Asp97 (CB) from N_6 atom of inhibitor CP3, plotted as a function of time in CDK2-CP3 and CDK7-CP3.

Table S4. Hydrogen bonds analyses from MD trajectories [a]

System	Donor[a]	Acceptor[a]	Occupancy(%)[b]	Distance(Å)[c]	Angle(°)[d]
CDK2-CP1	Leu83 N-H	CP1 N1	98.48	3.003 ± 0.12	24.90 ± 13.12
	CP1 N1-H9	Leu83 O	73.16	2.999 ± 0.17	27.15 ± 8.12
	Lys89 NZ-HZ3	CP1 O2	41.69	3.234 ± 0.45	33.05 ± 15.12
	Lys89 NZ-HZ2	CP1 O2	31.69	3.202 ± 0.60	33.53 ± 14.87
CDK7-CP1	CP1 N1-H9	MET94 O	99.97	2.952 ± 0.15	23.29 ± 10.12
	Met94 N-H	CP1 N1	99.28	3.064 ± 0.14	32.19 ± 14.55
	Lys44 NZ-HZ2	CP1 O3	36.00	2.87 ± 0.70	154.31 ± 14.73
CDK2- CP2	CP2 N7-H8	Leu83 O	99.81	3.001 ± 0.16	21.96 ± 11.08
	Leu83 N-H	CP2 N1	97.83	3.099 ± 0.15	29.01 ± 13.13
	CP2 N7A-H5	Asp145 OD1	46.28	3.011 ± 0.28	21.36 ± 11.00
	CP2 N7A-H4	Asp145 OD2	40.40	3.454 ± 0.59	32.67 ± 14.69
	Lys89 NZ-HZ2	CP2 O8B	12.80	3.498 ± 0.78	34.19 ± 14.49
CDK7- CP2	CP2 N7-H8	Met94 O	99.30	2.961 ± 0.15	19.31 ± 11.30
	Met94 N-H	CP2 N1	98.37	3.106 ± 0.15	28.58 ± 12.97
	CP2 N7A-H4	Asp155 OD2	30.59	3.241 ± 0.75	27.72 ± 15.58
	CP2 N7A-H4	Asp155 OD1	28.21	4.142 ± 0.72	37.56 ± 13.74
CDK2-CP3	CP3 N7-H6	Leu83 O	100	3.026 ± 0.16	27.77 ± 13.07
	Leu83 N-H	CP3 N1	99.80	3.054 ± 0.13	21.74 ± 10.26
CDK7- CP3	CP3 N7-H6	Leu83 O	99.94	3.038 ± 0.21	19.55 ± 10.12
	MET94 N-H	CP3 N1	99.51	3.023 ± 0.13	31.79 ± 13.87
Q85T- CP1	Leu83 N-H	CP1 N1	85.90	3.090 ± 0.16	21.70 ± 9.69
K89L- CP1	Leu83 N-H	CP1 N1	82.97	3.123 ± 0.16	21.65 ± 9.79
	CP1 N8-H	Glu228 O	100	2.63 ± 0.11	158.97 ± 9.88
	CP1 N16-H	Asp292 OD2	30.00	4.31 ± 0.43	144.23 ± 13.75
D145A- CP1	Leu83 N-H	CP1 N1	79.69	3.167 ± 0.19	21.65 ± 9.79

[a] The listed donor and acceptor pairs satisfy the criteria for the hydrogen bond over 30.0% of the time during the 40 ns of simulation. [b] Occupancy is in unit of percentage of the investigated time period. [c] The average distance with standard error (SE = standard deviation/$N_{1/2}$) between hydrogen acceptor atom and proton on hydrogen donor atom in the investigated time period. [d] The average angle with standard error (SE = standard deviation/$N_{1/2}$) in parentheses for hydrogen bond in the investigated time period.