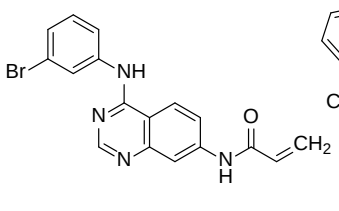
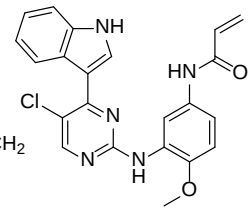
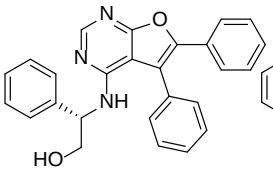


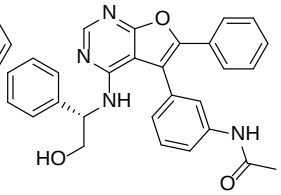
DJK



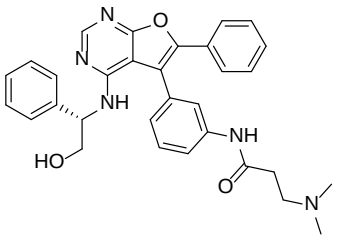
1WY



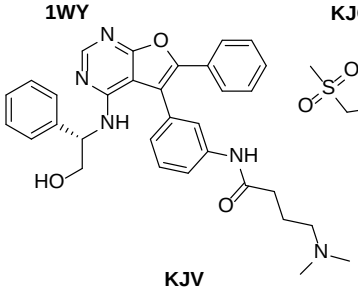
KJQ



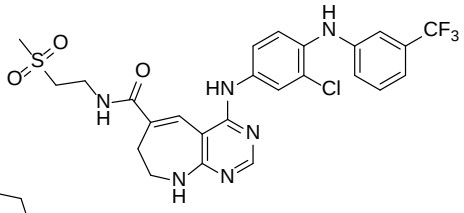
KJR



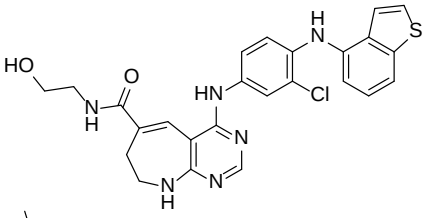
KJ8



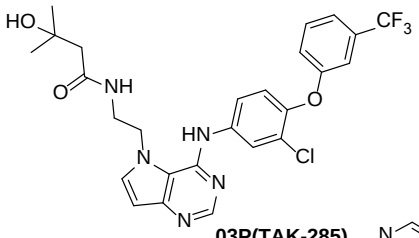
KJV



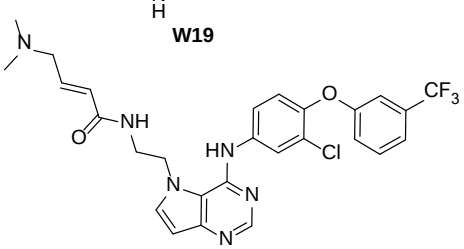
W32



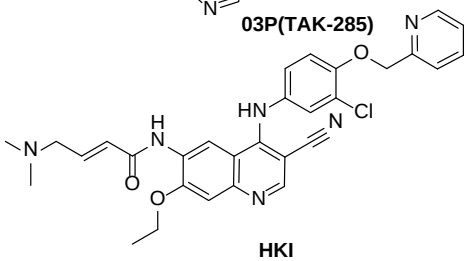
W19



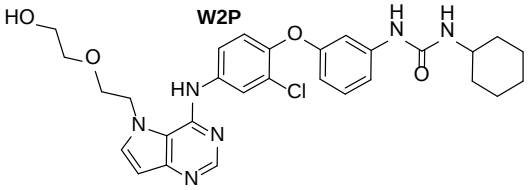
03P(TAK-285)



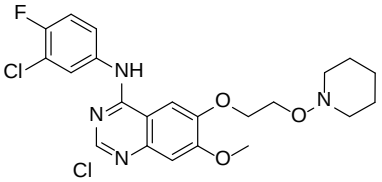
W2P



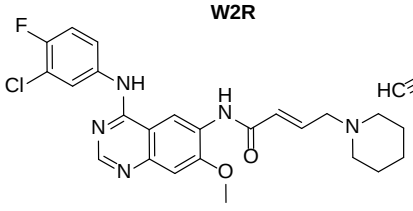
HKI



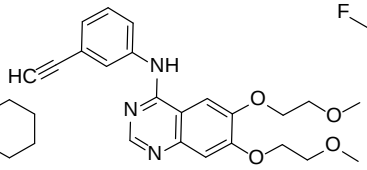
W2R



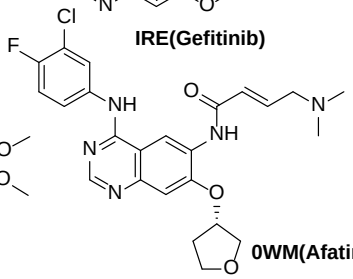
IRE(Gefitinib)



1C9(Dacomitinib)



AQ4(Erlotinib)



0WM(Afatinib)

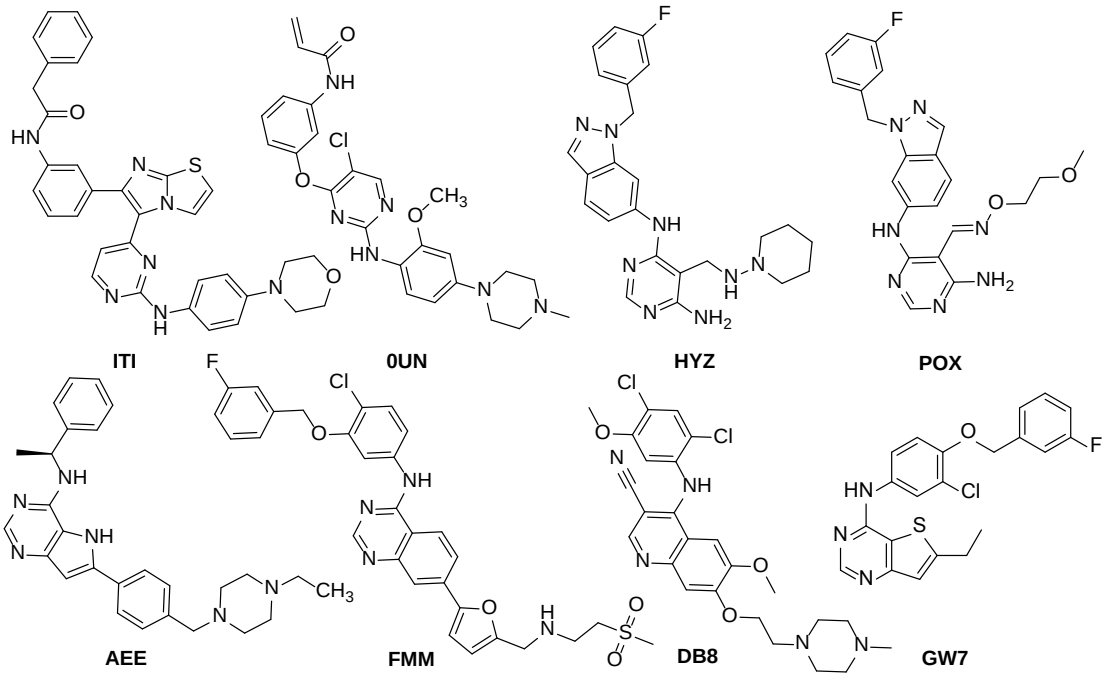
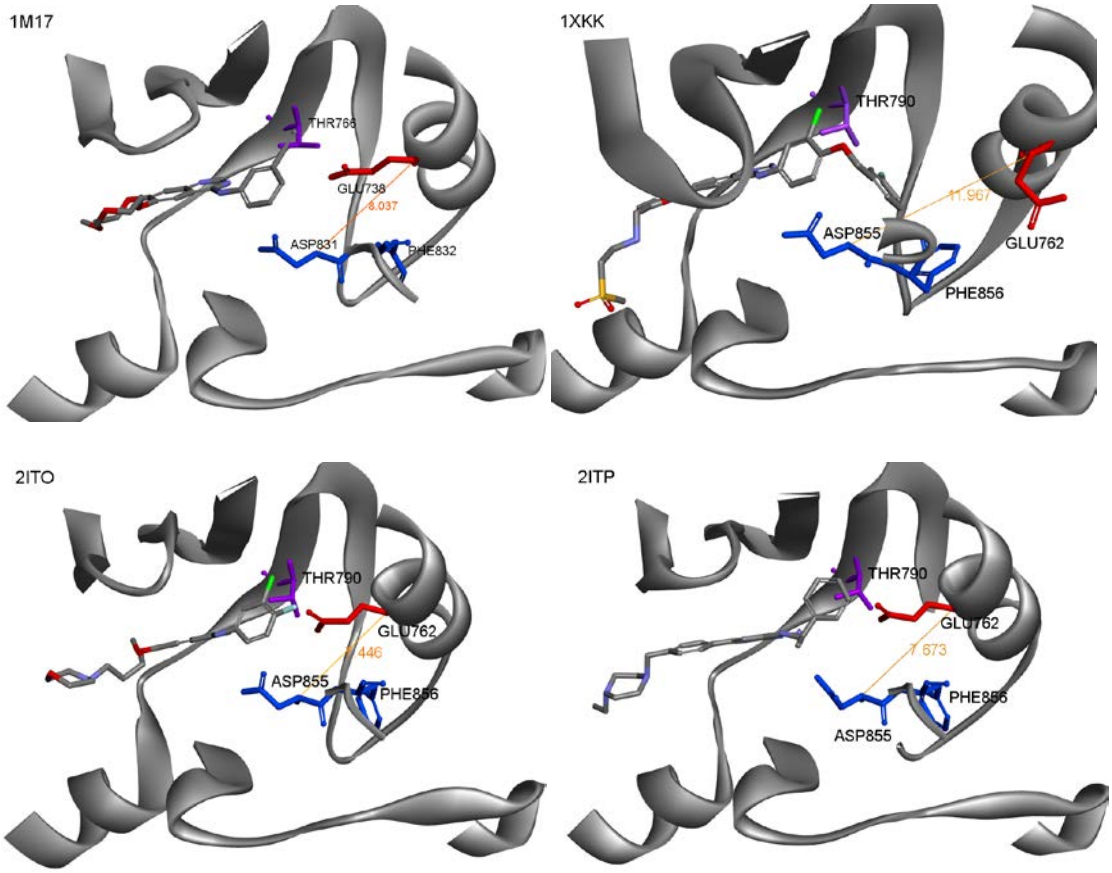
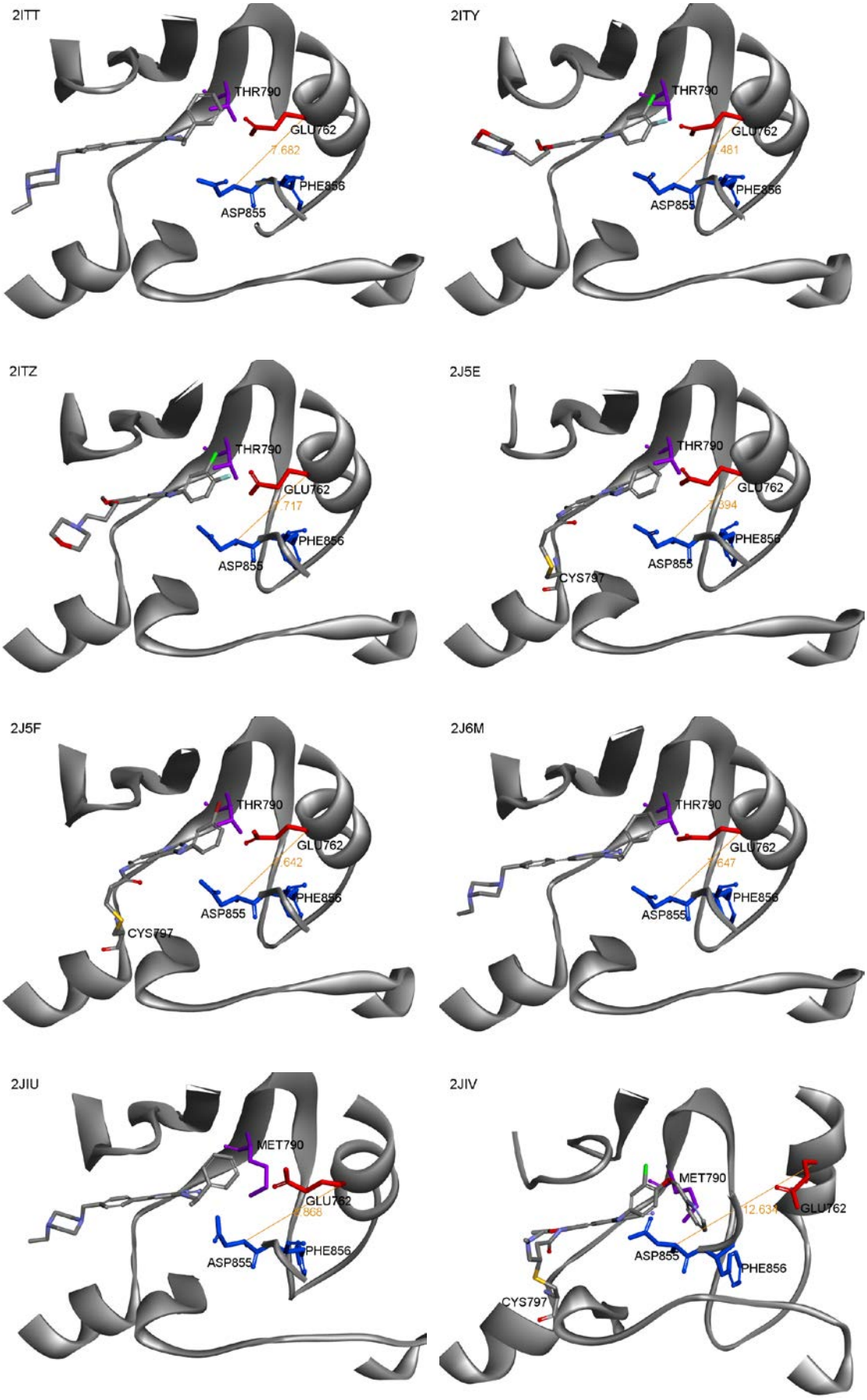
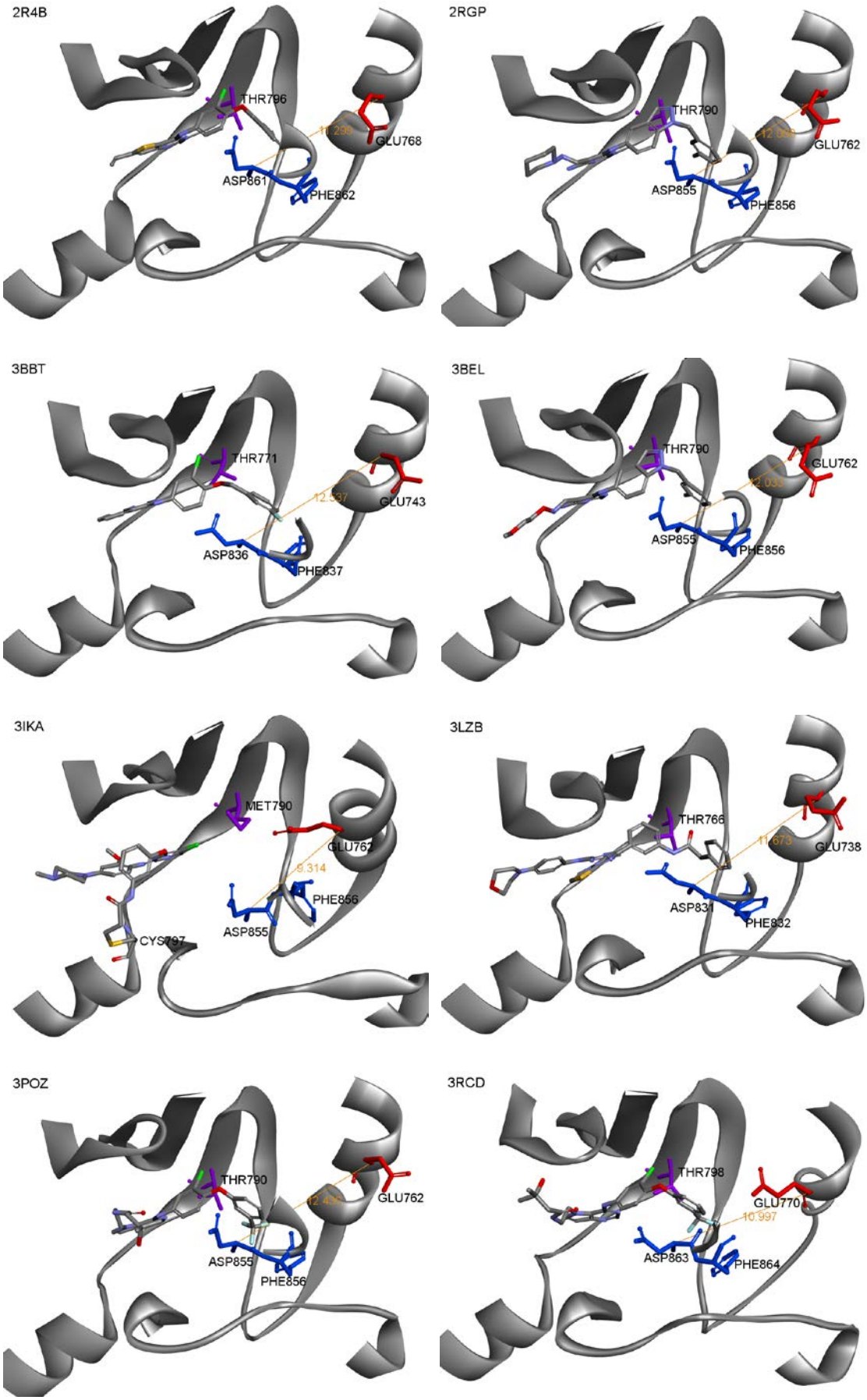
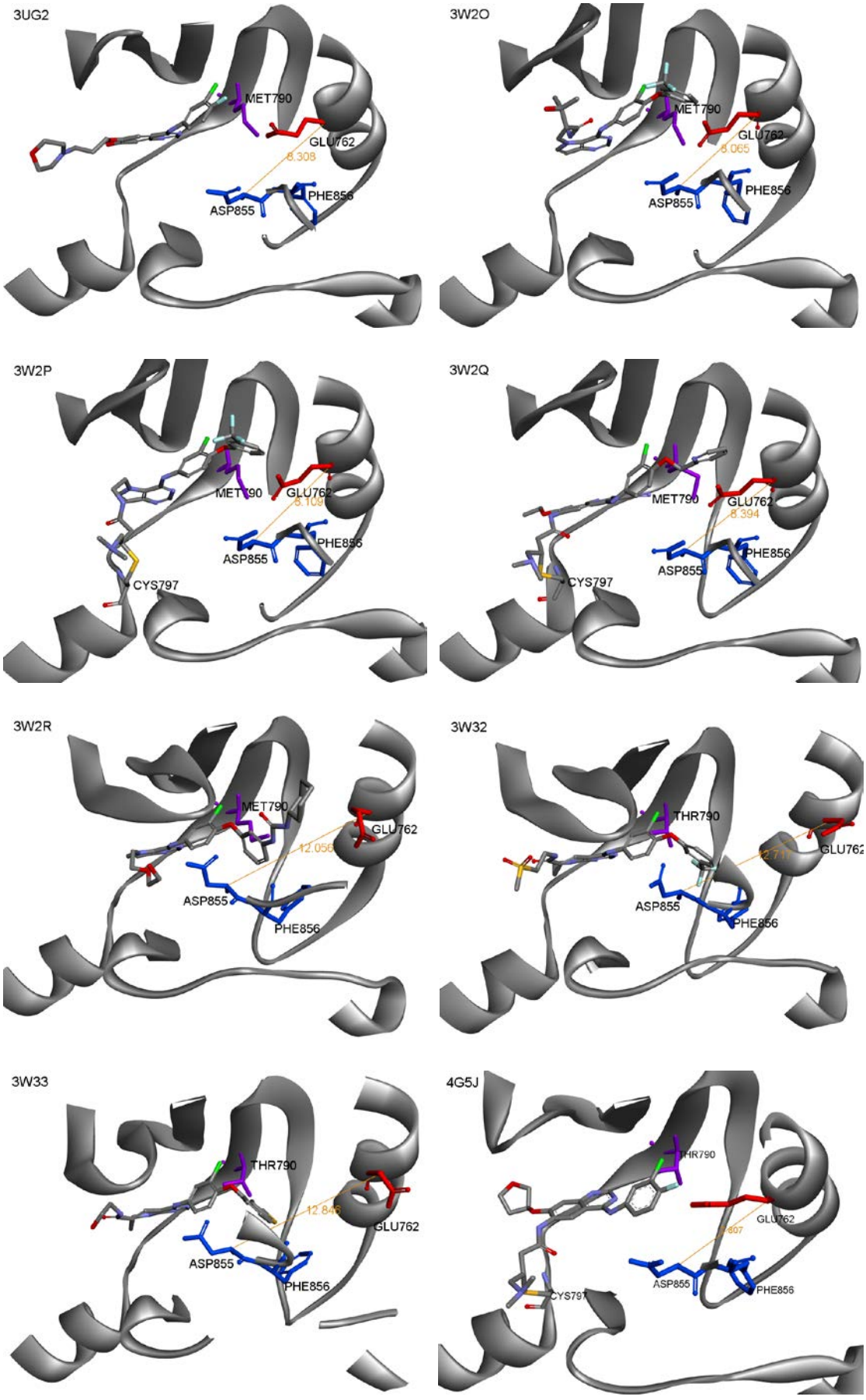


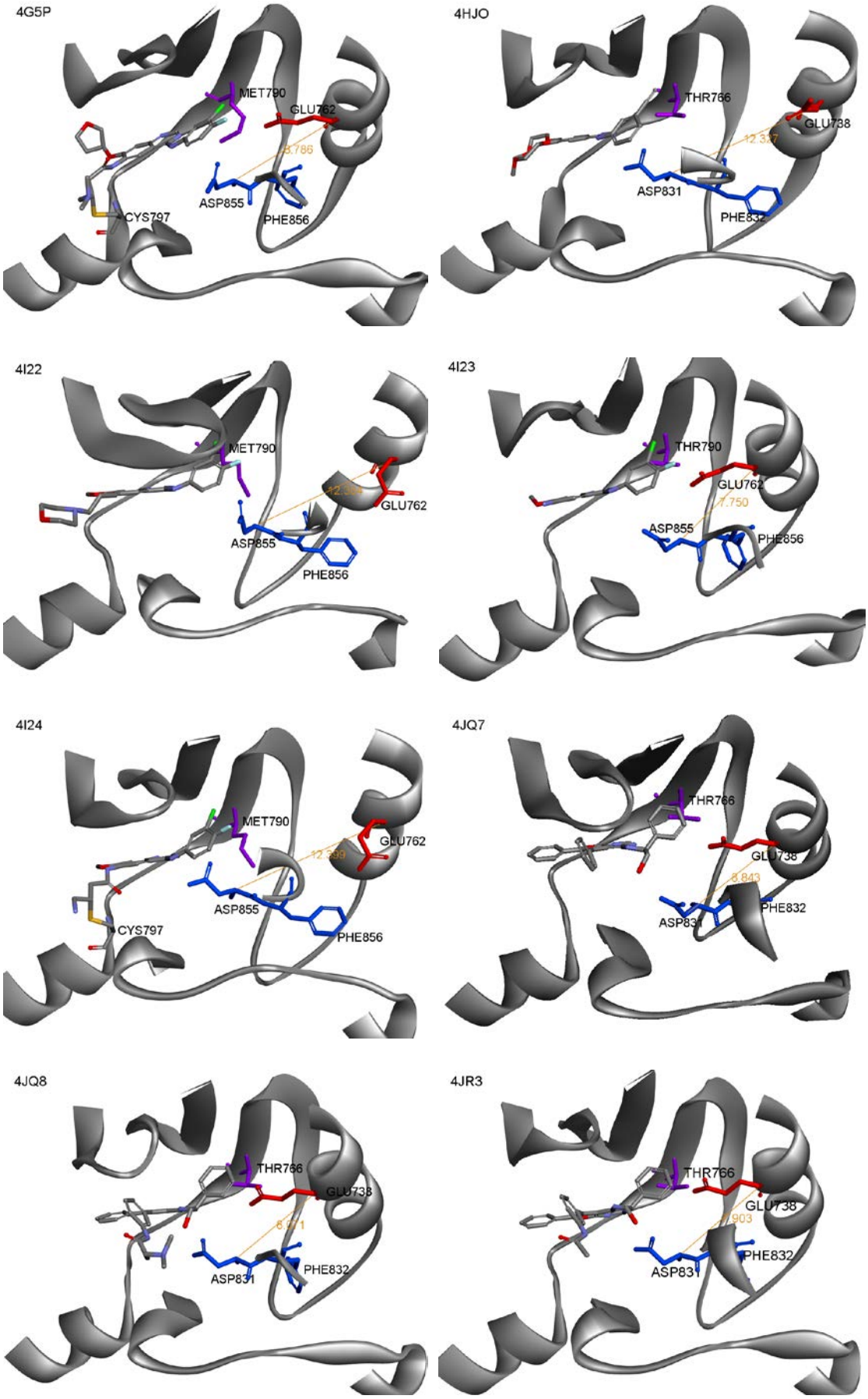
Figure S1. Chemical structures of all the small molecules within the EGFR protein crystal complexes.




2ITT
THR790
GLU762
7.682
PHE856
ASP855
2ITY
THR790
GLU762
7.481
PHE856
ASP855
2ITZ
THR790
GLU762
7.717
PHE856
ASP855
2J5E
THR790
GLU762
7.694
PHE856
ASP855
CYS797
2J5F
THR790
GLU762
7.642
PHE856
ASP855
CYS797
2J6M
THR790
GLU762
7.647
PHE856
ASP855
2JIU
MET790
GLU762
7.868
PHE856
ASP855
2JIV
MET790
12.634
GLU762
ASP855
PHE856
CYS797


2R4B
THR796
GLU768
ASP861
PHE862
2RGP
THR790
GLU762
ASP855
PHE856
3BBT
THR771
GLU743
ASP836
PHE837
3BEL
THR790
GLU762
ASP855
PHE856
3IKA
MET790
GLU762
PHE856
ASP855
CYS797
3LZB
THR766
GLU738
ASP831
PHE832
3POZ
THR790
GLU762
ASP855
PHE856
3RCD
THR798
GLU770
ASP863
PHE864


3UG2
MET790
GLU762
8.308
ASP855
PHE856
3W2O
MET790
GLU762
8.065
ASP855
PHE856
3W2P
MET790
GLU762
8.109
ASP855
PHE856
CYS797
3W2Q
MET790
GLU762
8.394
ASP855
PHE856
CYS797
3W2R
MET790
GLU762
12.056
ASP855
PHE856
3W32
THR790
GLU762
12.711
ASP855
PHE856
3W33
THR790
12.848
GLU762
ASP855
PHE856
4G5J
THR790
GLU762
8.807
ASP855
PHE856
CYS797


4G5P
MET790
GLU762
8.786
ASP855
PHE856
CYS797
4HJO
THR766
GLU738
12.317
ASP831
PHE832
4I22
MET790
12.504
GLU762
ASP855
PHE856
4I23
THR790
GLU762
7.750
ASP855
PHE856
4I24
MET790
GLU762
12.239
ASP855
CYS797
PHE856
4JQ7
THR766
GLU738
8.843
PHE832
ASP831
4JQ8
THR766
GLU738
8.971
ASP831
PHE832
4JR3
THR766
GLU738
8.903
PHE832
ASP831

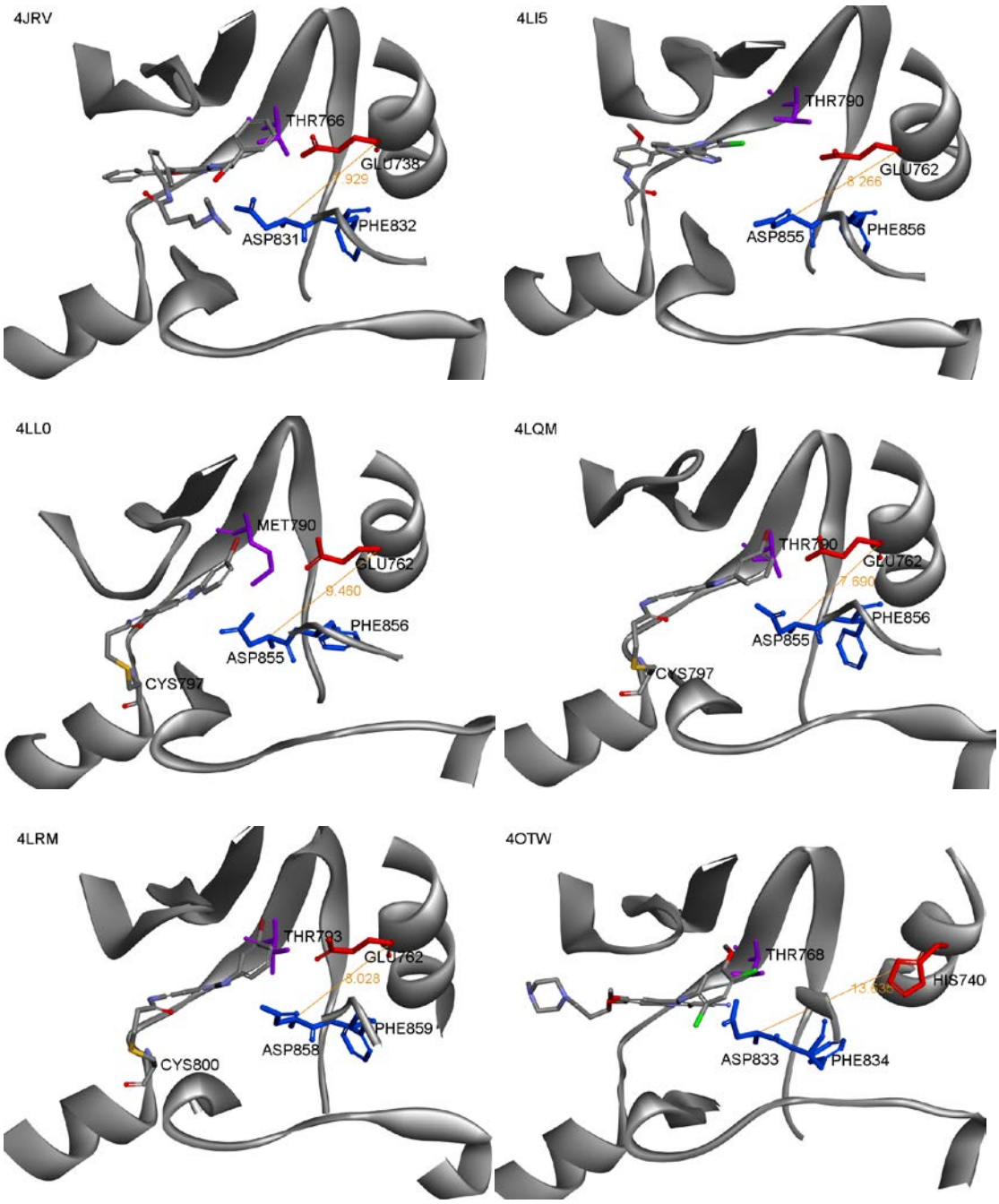


Figure S2. All the protein-ligand binding modes within 42 EGFR protein crystal complexes.

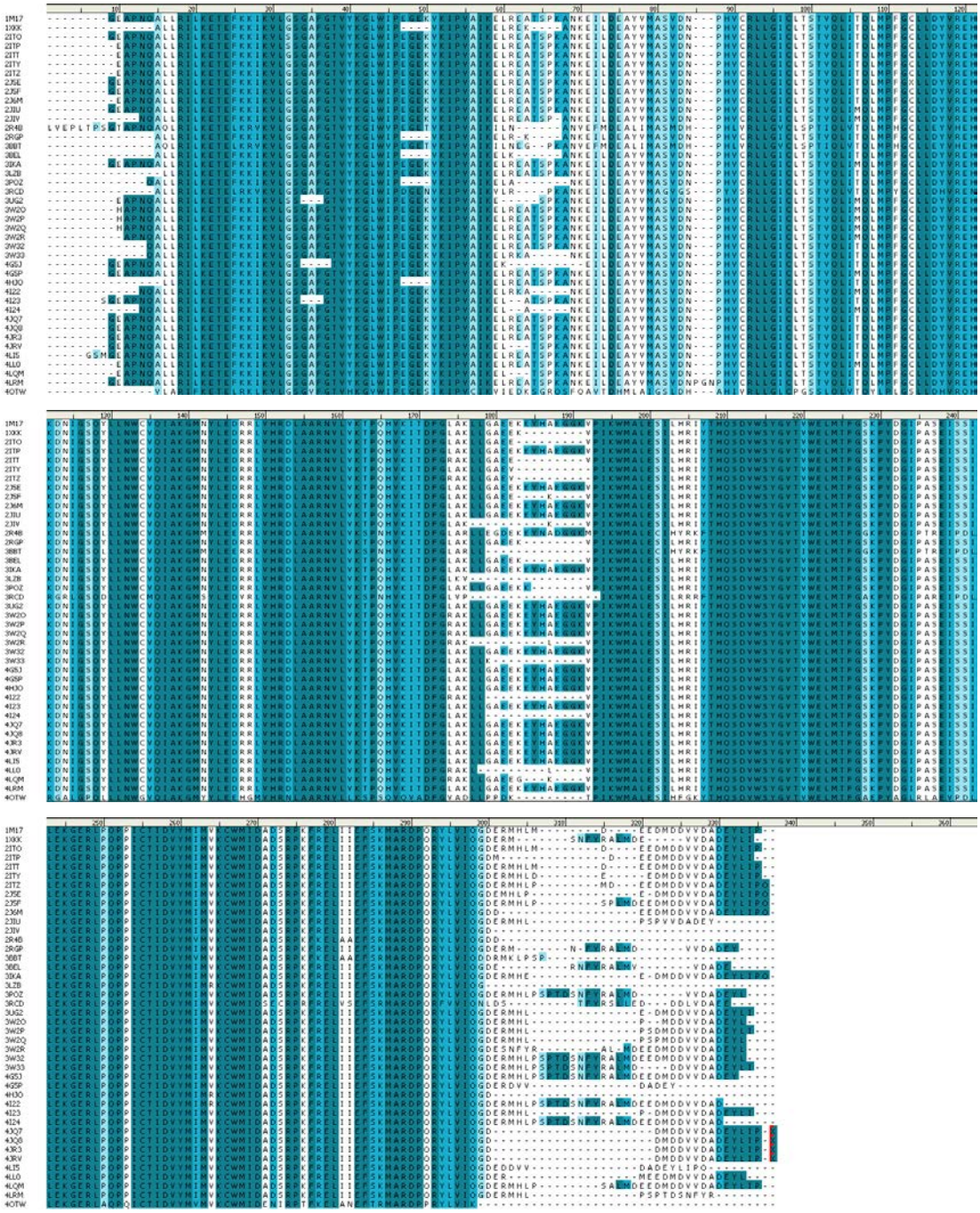


Figure S3. Sequence alignment within 42 EGFR protein crystal complexes.

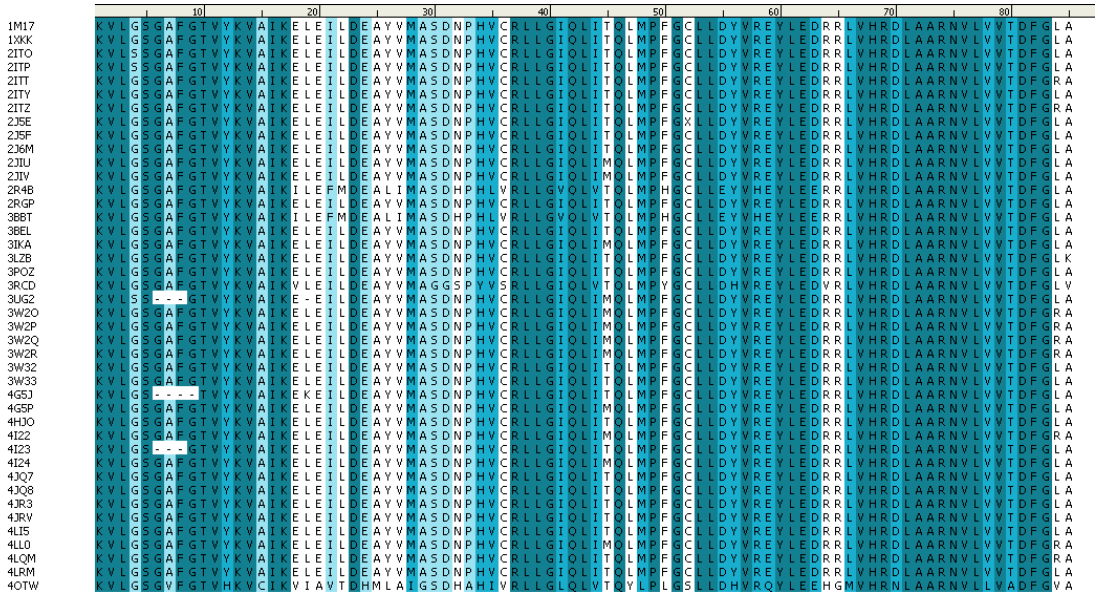
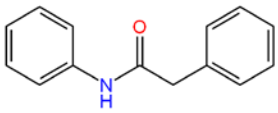
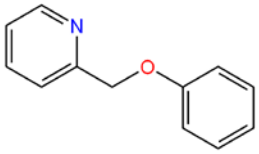
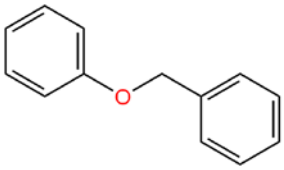
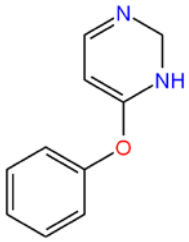
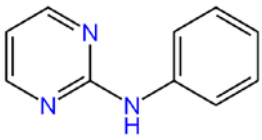
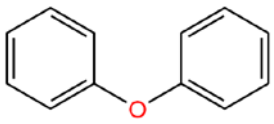
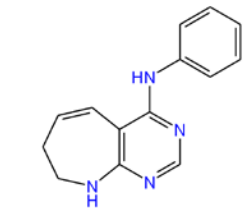
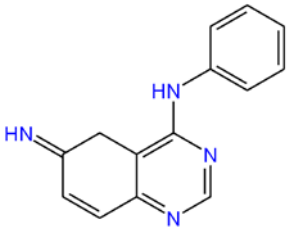
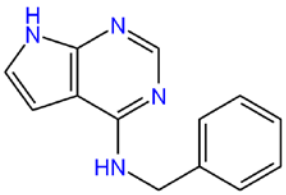
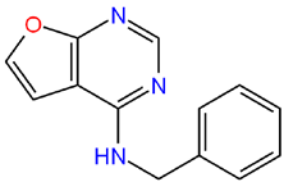
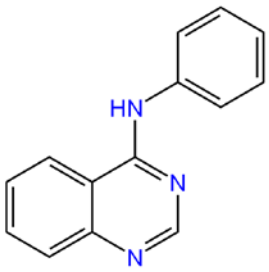
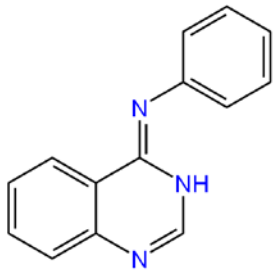
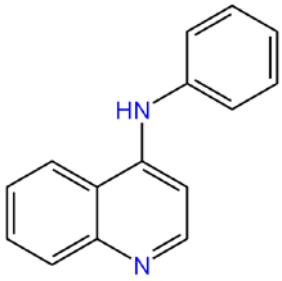
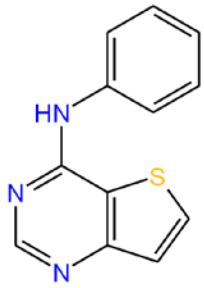
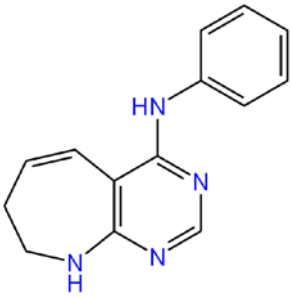
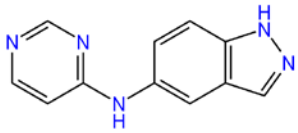


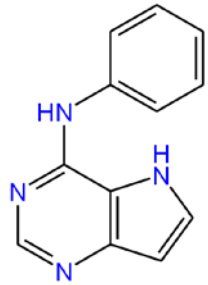
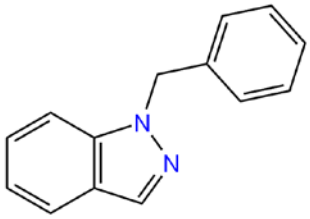
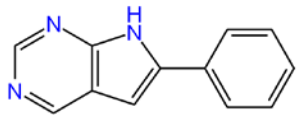
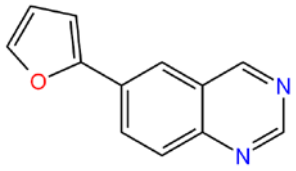
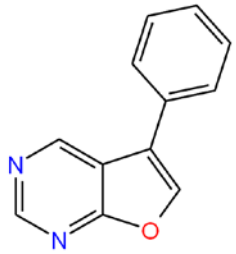
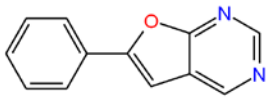
Figure S4. Sequence alignment of 85 conserved residues in the binding pockets of 42 EGFR protein crystal complexes.

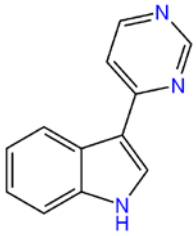
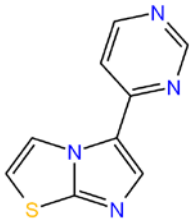
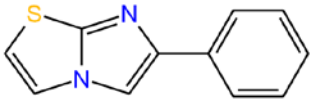
Table S1. The detailed information of 2D fragments decomposed from the 42 HER protein crystal complexes.

Chemical structure	Fragment name	Source structure	Pocket position	FLE[a]
	**1-1**	3LZB	BP-I, II, III, IV	3.27
	**1-2-1**	2JIV*	BP-I, II, III, IV	3.7
	1-2-2	3W2Q*	K, BP-II	3.7
	**1-3-1**	1XKK	BP-I, II, III, IV	4.86
	1-3-2	2R4B	BP-I, II, III, IV	3.7
	1-3-3	3BBT	BP-I, II, III, IV	3.34
	**1-4**	3IKA*	A, R	2.92
	**1-5-1**	3LZB	A	2.62
	1-5-2	4LI5	A, R	3.43
	1-5-3	3IKA*	A, R	3.65
	**1-6-1**	3W2O*	BP-I, II	3.21
	1-6-2	3W2P*	BP-I, II	2.98
	1-6-3	3W2R*	BP-I, II	4.33
	1-6-4	3W32	BP-I, II	4.87
	1-6-5	3W33	BP-I, II	5.59
	1-6-6	3POZ	BP-I, II	4.83
	1-6-7	3RCD	BP-I, II	4.91
	1-7-1	3W32	A, K	4.36

	**1-7-2**	3W33	A, K	5.01
	**1-8**	4LL0*	A, K	4.49
	**1-9-1**	2J6M	A, K	3.04
	1-9-2	2JIU*	A, K	2.88
	1-9-3	2ITP	A, K	3.03
	1-9-4	2ITT	A, K	3.35
	**1-10-1**	4JQ7	A, K	2.95
	1-10-2	4JQ8	A, K	4.21
	1-10-3	4JR3	A, K	3.65
	1-10-4	4JRV	A, K	3.67
	**1-11-1**	1XKK	A, K	4.98
	1-11-2	4G5P*	A, K	7.34
	1-11-3	4HJO	A, K	5.74
	1-11-4	4I22*	A, K	3.79
	1-11-5	4I23	A, K	5.69
	1-11-6	4I24	A, K	5.18
	1-11-7	4LRM	A, K	7.38
	1-11-8	2J5E	A, K	7.38
	1-11-9	3BBT	A, K	3.42
	1-11-10	4G5J	A, K	8.49
	1-11-11	1M17	A, K	5.74
	1-11-12	3UG2*	A, K	3.79
	1-11-13	2ITO	A, K	4.61
	1-11-14	2ITY	A, K	4.85

	1-11-15	2ITZ	A, K	5.72
	**1-12-1**	2ITY	A, K	4.07
	1-12-2	3BBT	A, K	2.87
	**1-13-1**	2JIV*	A, K	5.36
	1-13-2	3W2Q*	A, K	5.36
	1-13-3	4OTW	A, K	-
	**1-14**	2R4B	A, K	3.49
	**1-15**	3W33	A, K	5.01
	**1-16-1**	2RGP	A, BP-I, II	2.87
	1-16-2	3BEL	A, BP-I, II	3.83
	1-17-1	3W2O*	A, K	2.22
	1-17-2	3W2P*	A, K	2.07

	**1-17-3**	3W2R*	A, K	3
	1-17-4	3POZ	A, K	3.35
	1-17-5	3RCD	A, K	3.4
	**1-18-1**	2RGP	BP-I, II, III, IV	5.78
	1-18-2	3BEL	BP-I, II, III, IV	7.73
	**1-19-1**	2J6M	A, R	2.16
	1-19-2	2JIU	A, R	2.9
	1-19-3	2ITP	A, R	3.05
	1-19-4	2ITT	A, R	3.37
	**1-20**	1XKK	A, R	2.97
	**1-21-1**	4JQ7	A, R	2.46
	1-21-2	4JQ8	A, R	3.51
	1-21-3	4JQ3	A, R	3.05
	1-21-4	4JRV	A, R	3.07
	**1-22-1**	4JQ7	A, R	2.96
	1-22-2	4JQ8	A, R	4.23
	1-22-3	4JQ3	A, R	3.67
	1-22-4	4JRV	A, R	3.69

	**1-23**	4LI5	A, K	2.87
	**1-24**	3LZB	A, K	1.41
	**1-25**	3LZB	A, K	3.08

*EGFR Protein containing T790M mutation. [a]Fragment lipophilicity efficiency, FLE = LipE*AlogP$_{(fragment)}$.

Table S2. The obtained results of docking-based virtual screening.

CDOCKE[a]	3W2R (L858R/T790M)		3W33 (Wild Type)	
	compounds	CDOCKER_Interaction_Energy (kcal/mol)	compounds	CDOCKER_Interaction_Energy (kcal/mol)
Ten top molecules ranked by docking scores	**A-2**	-57.896	**A-10**	-70.2457
	A-3	-37.3167	**A-2**	-70.9749
	B-3	-25.4889	**A-5**	-66.752
	No refined pose	-[c]	**A-8**	-68.1622
			A-9	-75.3367
			B-10	-66.7921
			B-3	-66.5094
			W19	-66.3401
			A-4	-66.245
			B-2	-66.177
Glide[b] (sp)	3W2R (L858R/T790M)		3W33 (Wild Type)	
	compounds	Glide score (kcal/mol)	compounds	Glide score (kcal/mol)
Ten top molecul	**B-1**	-14.531775	**A-10**	-13.534213
	A-10	-12.001108	**A-3**	-13.310207

es ranked by docking scores	**A-2**	-11.442017	**A-1**	-12.569783
	A-3	-11.146296	**A-2**	-11.641528
	A-8	-9.333602	**A-8**	-11.011472
	A-5	-8.991333	**A-7**	-10.336488
	A-9	-8.87509	**A-4**	-9.825871
	A-4	-8.35916	**A-5**	-9.816869
	A-6	-8.250604	**A-9**	-8.633209
	B-2	-8.143607	**A-6**	-7.025011
Glide[b] (xp)	3W2R (L858R/T790M)		3W33 (Wild Type)	
	compounds	Glide score (kcal/mol)	compounds	Glide score (kcal/mol)
Ten top molecules ranked by docking scores	**B-1**	-13.532223	**A-2**	-13.178289
	A-10	-12.889075	**A-10**	-12.988253
	W2R	-12.821827	**W19**	-12.057136
	A-3	-12.009716	**A-5**	-10.109834
	A-9	-11.92504	**A-3**	-10.051887
	A-1	-11.745627	**A-9**	-10.00585
	A-2	-11.528264	**A-7**	-9.949243
	B-4	-10.28542	**A-8**	-9.232259
	B-3	-8.963894	**A-1**	-8.606092
	A-7	-8.867099	**A-4**	-8.171397

[a] Performed on the Discovery studio 3.5 platform. [b] Performed on the Schrodinger 2012 suite; sp, standard precision; xp, extra precision. [c] No data.