Table S1. The relationship between the anchors and moieties of 15 co-crystallized ligands for TK

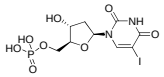
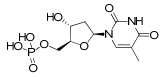
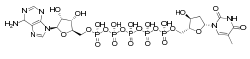
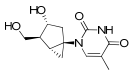
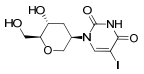
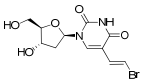
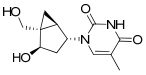
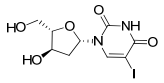
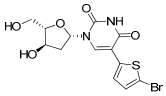
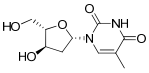
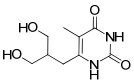
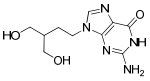
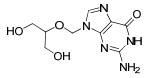
PDB code	Compound structure	Anchor			
		E1	H1	H2	V1
3vtk			R-OH		
1vtk			R-OH		
1p7c			R-OH		
1of1			R-OH		
1ki6			R-OH		
1ki8			R-OH		
1e2k			R-OH		
1ki7			R-OH		
1ki4			R-OH		
1kim			R-OH		
1e2p			R-OH		
1ki3			R-OH		Aromatic moiety
1ki2			R-OH		Aromatic moiety
1qhi			R-OH	R-NH_2	Aromatic moiety
2ki5			R-OH	R-NH_2	Aromatic moiety

Table S2. Comparing SiMMap with other methods on thymidine kinase and estrogen receptor by false-positive rates

True positive (%)	Thymidine kinase (TK)				Estrogen receptor (ER)			
	SiMMap	DOCK[a]	FlexX[a]	GOLD[a]	SiMMap	DOCK[a]	FlexX[a]	GOLD[a]
80	6.3[b]	23.4	8.8	8.3	1.1	13.3	57.8	5.3
90	6.8	25.5	13.3	9.1	1.1	17.4	70.9	8.3
100	6.8	27	19.4	9.3	7.5	18.9	NA	23.4

[a] Summarized from Bissantz *et al*. (20)

[b] The false-positive rate (%) is obtained based on 1000 compounds.

Table S3. The mapping between the anchors and 10 active and some typical compounds for TK

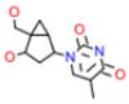
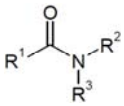
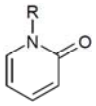
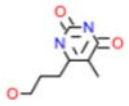
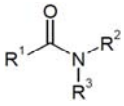
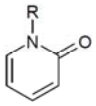
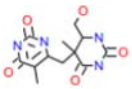
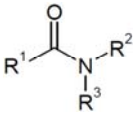
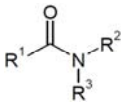
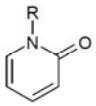
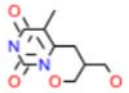
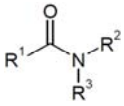
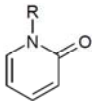
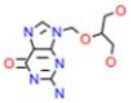
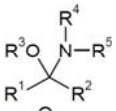
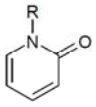
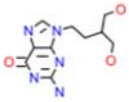
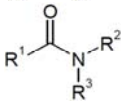
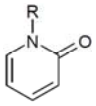
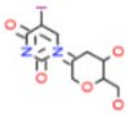
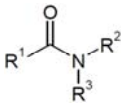
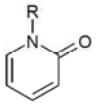
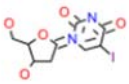
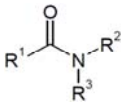
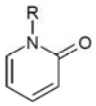
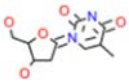
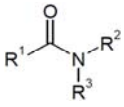
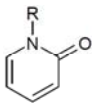
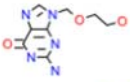
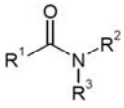
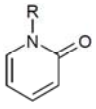
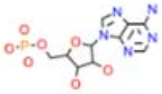
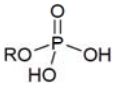
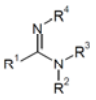
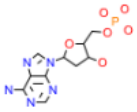
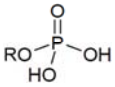
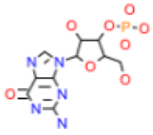
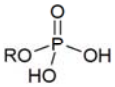
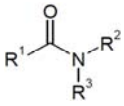
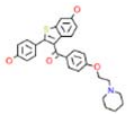
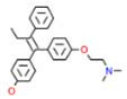
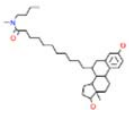
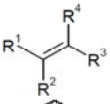
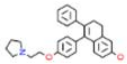
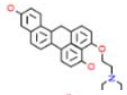
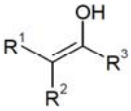
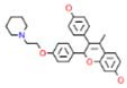
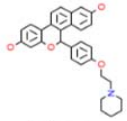
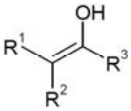
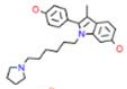
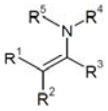
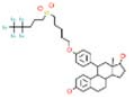
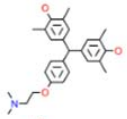
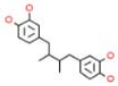
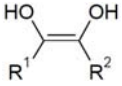
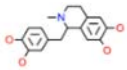
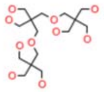
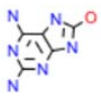
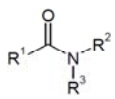
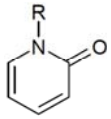
Compound	Structure	GEMDOCK rank	SiMMap rank	SiMMap score	E1	H1	H2	V1
TK_01		89	57	3.253		R—OH		
TK_02		182	76	3.233		R—OH		
TK_03		127	67	3.243				
TK_04		122	65	3.244		R—OH		
TK_05		46	39	3.262		R—OH		
TK_06		42	36	3.263		R—OH		
TK_07		75	53	3.255		R—OH		
TK_08		68	52	3.256		R—OH		
TK_09		181	77	3.233		R—OH		
TK_10		148	70	3.239		R—OH		
MFCD00005750		8	1	4.283		R—OH		Aromatic moiety
MFCD00005753		19	3	4.271		R—OH		Aromatic moiety
MFCD00005763		87	9	4.253		R—OH		Aromatic moiety
MFCD00011393		4	384	1.295				
MFCD00003569		9	118	2.281				Other

Table S4. The mapping between the anchors and 10 active and some typical compounds for ER

Compound	Structure	GEMDOCK rank	SiMMap rank	SiMMap score	H1	V1	V2	V3
EST_01		2	1	4.239	R—OH	Aromatic moiety	OH	OH
EST_02		32	19	4.216	R—OH	Aromatic moiety	OH	OH
EST_03		28	16	4.217	R—OH		OH	OH
EST_04		8	5	4.226	R—OH	Aromatic moiety	OH	Aromatic moiety
EST_05		6	4	4.228		Aromatic moiety	OH	OH
EST_06		3	2	4.238	R—OH	Aromatic moiety	OH	OH
EST_07		21	13	4.218		Aromatic moiety	OH	Other
EST_08		10	7	4.225	R—OH	Other		Other
EST_09		30	20	4.216	R—OH	Aromatic moiety	OH	Aromatic moiety
EST_10		246	84	4.193	R—OH	Aromatic moiety	OH	OH
MFCD00002206		4	3	4.232		Other	OH	OH
MFCD00012748		17	11	4.221	R—OH	Other	OH	OH
MFCD00004690		5	154	3.23	R—OH		Other	Other
MFCD00013089		25	617	2.218				

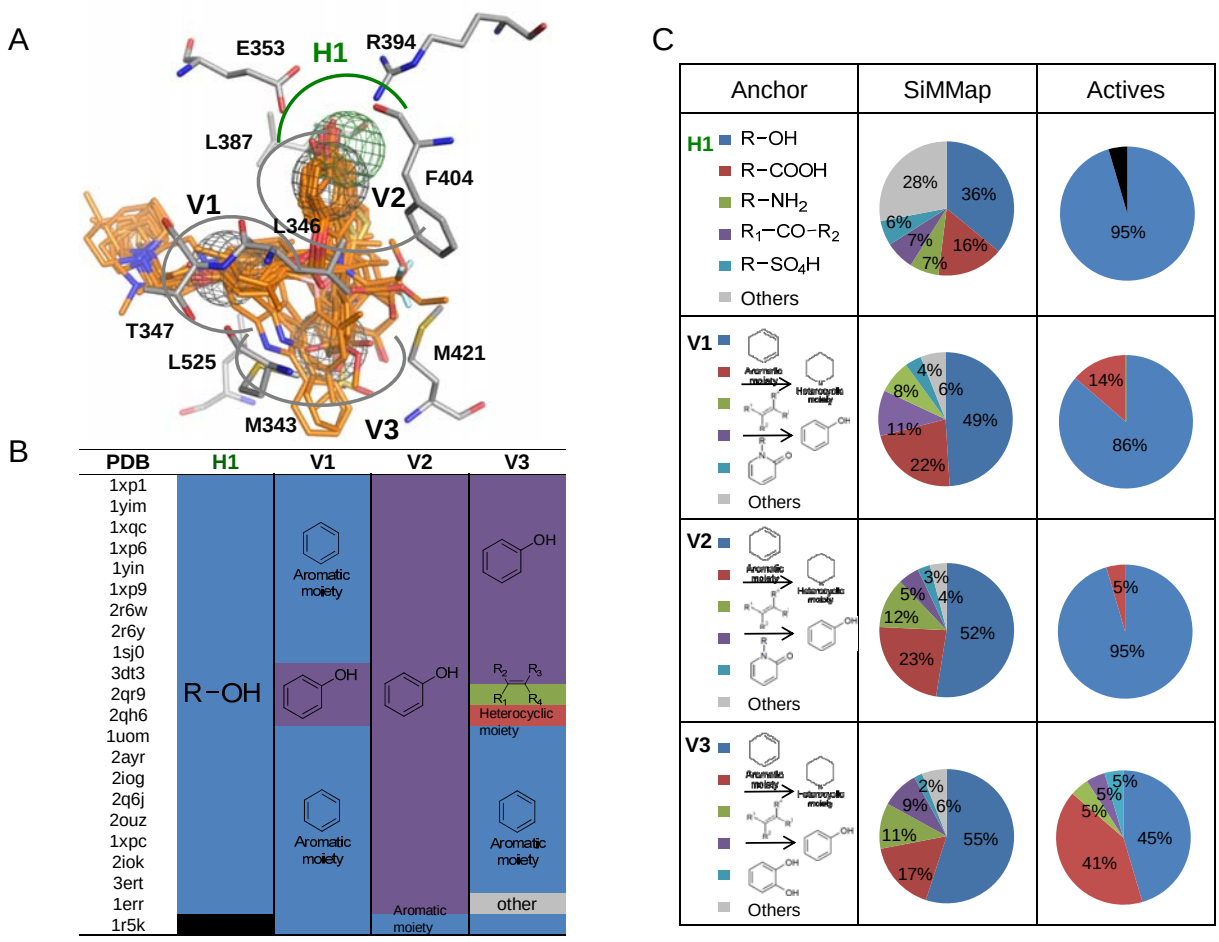


Figure S1. The relationships between the site-moiety map and 22 co-crystallized ligands of ER. (A) The mapping between four inferred anchors (binding pocket with conserved interacting residues) and these 22 ligands in the active site. (B) The moieties of these 22 ligands in each anchor. Black cell presents that the moiety of the compound does not agree with the anchor H1. (C) The moiety compositions of 1000 docked compounds (SiMMap) and these 22 ligands.

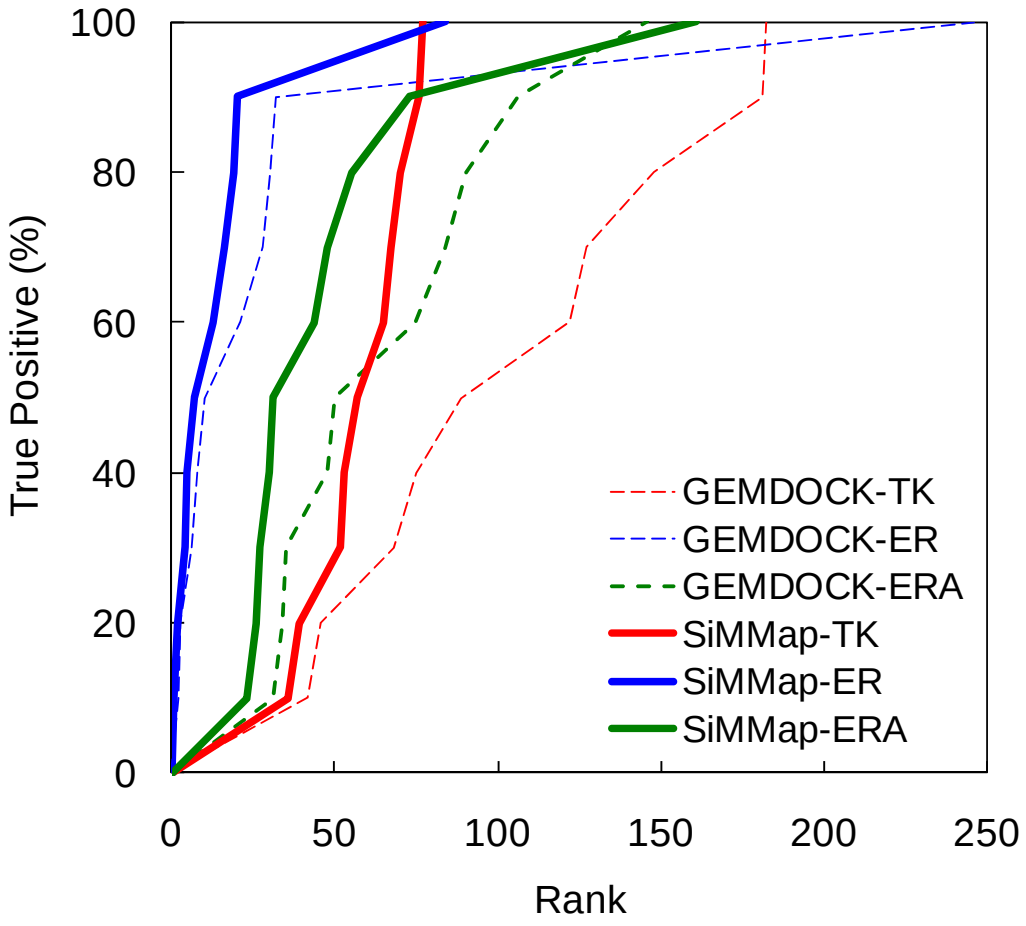


Figure S2. Comparison of SiMMap with GEMDOCK. The SiMMap server (solid lines) consistently outperforms GEMDOCK (dash lines) on three targets: TK (red), ER (blue) and ERA (green).