Electronic Supplementary Material (ESI) for Molecular BioSystems.

Supplementary Information

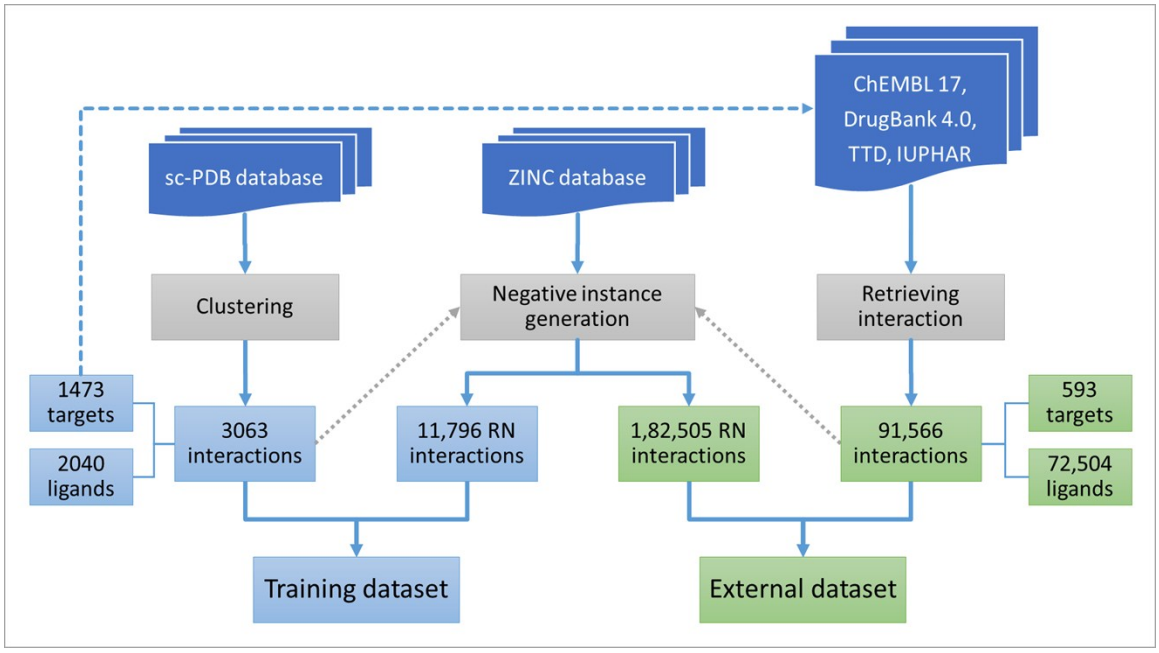


Fig. S1 Compilation of training and external validation set (RN=reliable negative)

Table S1 Selected descriptor set

Descriptors sets	Target Descriptors		Ligand Descriptors		Total
	Number	Type	Number	Type	
2D set	20	PROFEAT descriptors	85	Morgan fingerprint	**105**
3D set	38	FuzCav fingerprint	134	Morgan fingerprint	**172**

Table S2 Effect of presence metal ion in binding site on redock RMSD.

Targets	Count	RMSD (mean)	RMSD (std.)
Without metal ion	2573	1.48	1.18
With metal ion	490	1.47	1.08

Table S3 Enzyme family wise performance of redocking.

EC Number	Name	Count	RMSD (mean)	RMSD (std.)
1	Oxidoreductases	352	1.52	1.45
2	Transferases	1011	1.42	1.04
3	Hydrolases	753	1.56	1.20
4	Lyases	136	1.58	1.06
5	Isomerases	67	1.46	1.09
6	Ligases	67	1.64	1.43
no EC	-	677	1.45	1.11

Table S4 Predictive performances of different dataset

Model	Dataset	ROC_auc	Recall	Specificity	Precision	Accuracy	f1_score	MCC	No. of Positive instances	No. of Negative instances
3d_rf	train	1.000	1.000	1.000	1.000	1.000	1.000	1.000	2756	10616
	test	0.962	0.927	0.998	0.993	0.984	0.959	0.949	306	1179
	external	0.909	0.912	0.905	0.828	0.908	0.868	0.800	91566	182505
	external1	0.899	0.902	0.896	0.766	0.898	0.829	0.761	41161	108858
3d_svm	train	0.961	0.923	0.998	0.994	0.983	0.957	0.948	2756	10616
	test	0.959	0.920	0.998	0.992	0.982	0.955	0.945	306	1179
	external	0.909	0.914	0.905	0.828	0.908	0.869	0.800	91566	182505
	external1	0.900	0.901	0.899	0.771	0.899	0.831	0.765	41161	108858
2d_rf	train	1.000	1.000	1.000	1.000	1.000	1.000	1.000	2756	10616
	test	0.966	0.937	0.996	0.983	0.984	0.960	0.950	306	1179
	external	0.902	0.926	0.878	0.792	0.894	0.854	0.777	91566	182505
	external1	0.888	0.915	0.862	0.714	0.876	0.802	0.726	41161	108858
2d_svm	train	0.945	0.894	0.997	0.986	0.975	0.938	0.924	2756	10616
	test	0.944	0.891	0.996	0.985	0.975	0.935	0.921	306	1179
	external	0.907	0.874	0.940	0.879	0.918	0.877	0.815	91566	182505
	external1	0.901	0.864	0.938	0.841	0.918	0.853	0.796	41161	108858

external1 represents clustered external dataset.

Table S5 Recall for external dataset based on number of common bit-pairs

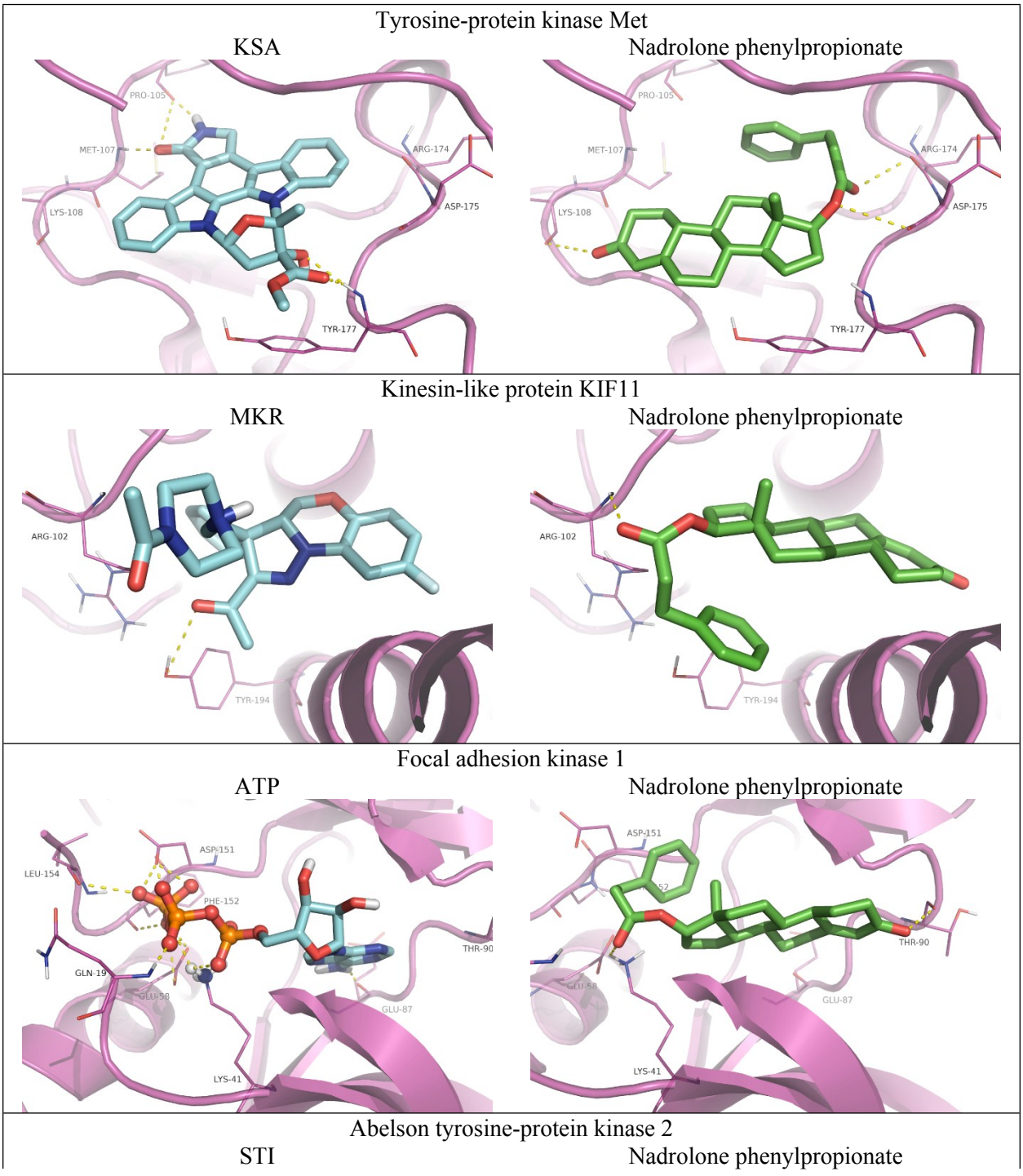
Number of Common Bit-pairs	Count	3d-rf	2d-rf
0	13	0.00	0.00
1-25	570	0.24	0.23
26-50	2012	0.53	0.54
51-75	3954	0.70	0.73
76-100	6763	0.80	0.83
101-125	9949	0.87	0.89
126-150	6170	0.90	0.93
151-175	14118	0.93	0.95
176-200	7235	0.95	0.96
201-225	7015	0.96	0.97
226-250	6778	0.97	0.98
251-275	7082	0.98	0.98
276-300	8202	0.99	0.99
301-325	3445	0.99	0.99
326-350	814	0.99	0.99
351-375	1765	1.00	1.00
376-400	1532	1.00	1.00
401-425	1326	1.00	0.99

426-450	834	1.00	0.99
451-475	628	1.00	1.00
476-500	396	1.00	1.00
501-525	52	1.00	1.00
526-550	260	1.00	1.00
551-575	202	1.00	1.00
576-600	102	1.00	1.00
601-625	7	1.00	1.00
626-650	61	1.00	1.00
651-675	60	1.00	1.00
676-700	5	1.00	1.00
701-725	61	1.00	1.00
726-750	39	1.00	1.00
751-775	5	1.00	1.00
776-800	13	1.00	1.00
801-825	17	1.00	1.00
826-850	0	nan	nan
851-875	10	1.00	1.00
876-900	4	1.00	1.00
901-925	1	1.00	1.00
926-950	3	1.00	1.00
951-975	0	nan	nan
976-1000	5	1.00	1.00
1001-1025	0	nan	nan
1026-1050	32	1.00	1.00
1051-1075	0	nan	nan
1076-1100	2	1.00	1.00
1101-1125	1	1.00	1.00
1126-1150	22	1.00	1.00
1151-1175	0	nan	nan
1176-1200	0	nan	nan
1201-1225	1	1.00	1.00

DOCKING POSES

Interaction for docking poses were identified using Binana 1.0.2 script.[1] Docking poses were generated using PyMOL.[2] All types of interaction were shown in the docking poses with special colour coding. Docked ligands were shown in green colour. π–π staking interaction were shown in orange colour, π–T interactions were shown in yellow colour, salt bridges were shown in pink

colour and cation–π interactions were shown in magenta colour. All these interactions were shown in ball and stick representation.



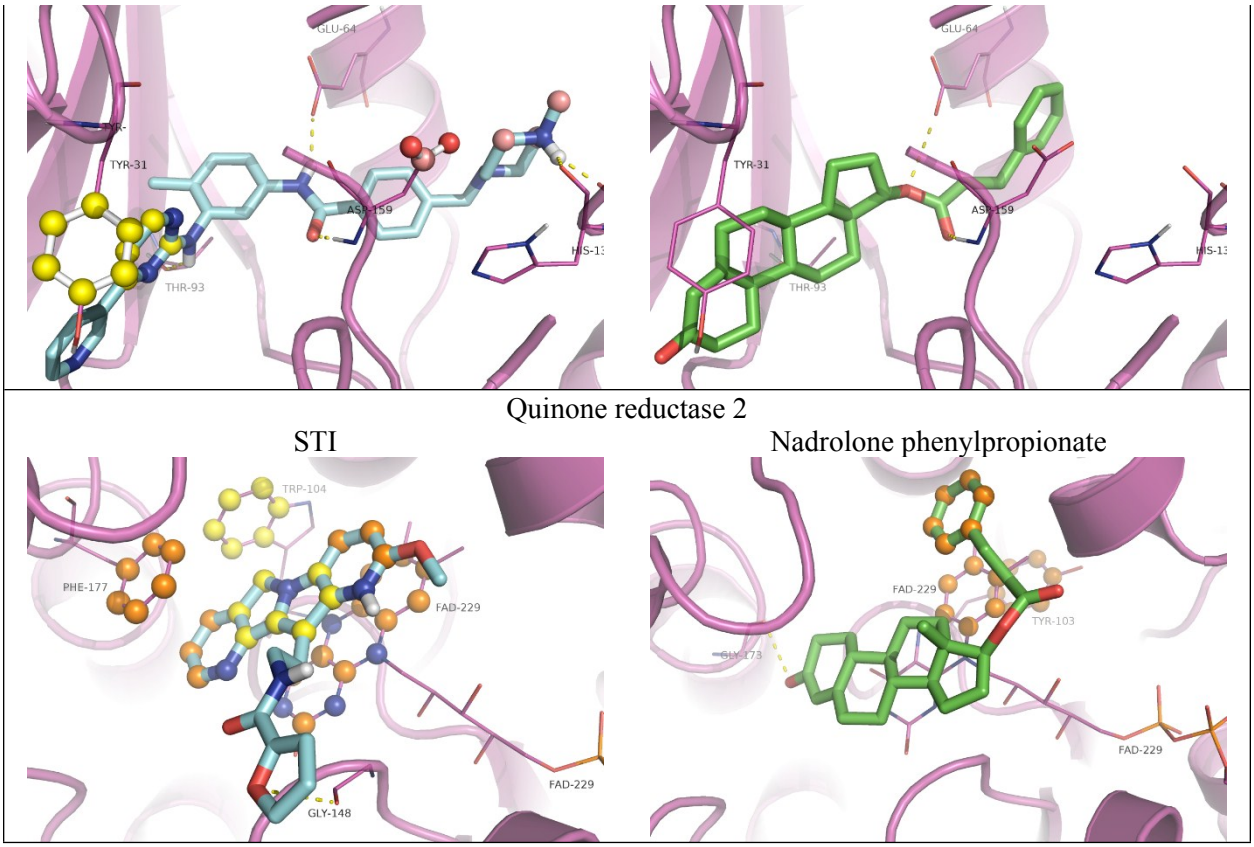
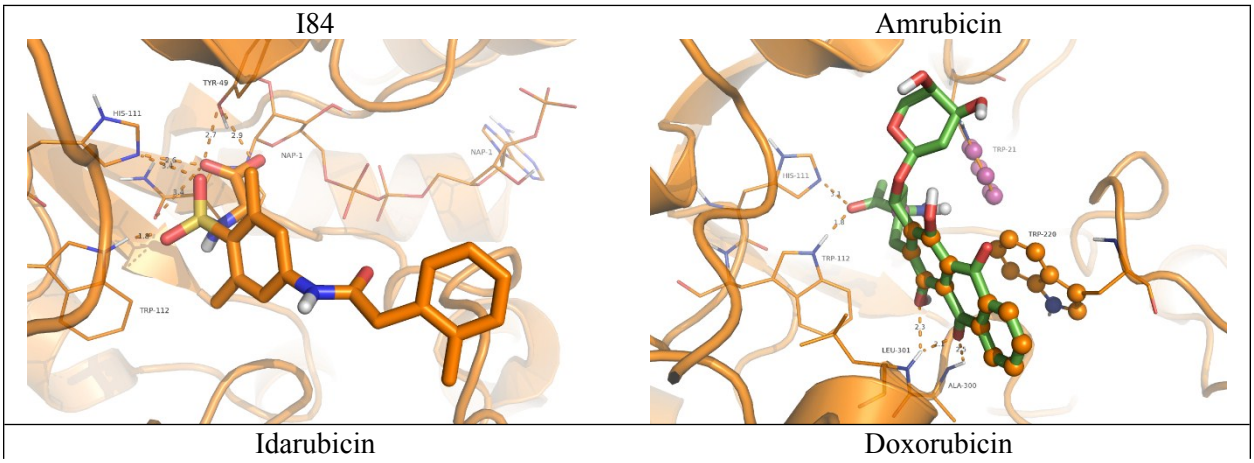


Fig. S2. Docking poses for Nadrolone phenylpropionate and native ligands (Cyan) in predicted targets



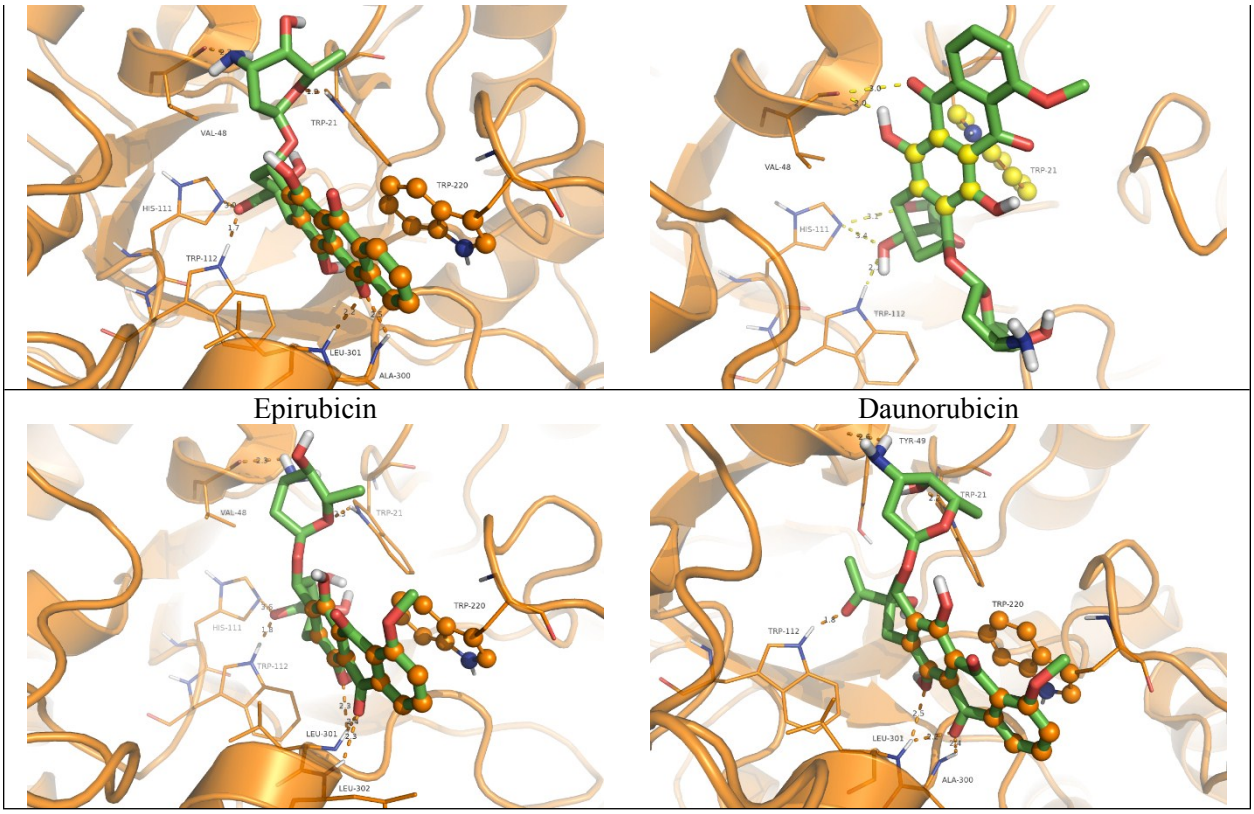


Fig. S3. Docking poses of anthracyclines and native ligand (Orange) in Aldose reductase 1B1

REFERENCES

1. Durrant, J. D.; McCammon, J. A., BINANA: A novel algorithm for ligand-binding characterization. *J. Mol. Graph. Model.* **2011**, 29, 888-893.
2. DeLano, W. L., The PyMOL molecular graphics system. **2002**.