Highlights

- The current QSAR study used molecular descriptors and K_i values from a large set of P-gp inhibitors were collated from the original literature, along with the type of the probe substrate used in IC_{50} measurement, and the Km and concentration of the substrate
- P-gp inhibition constants (K_i values) were calculated for 219 compounds from their IC_{50} values using Cheng and Prusoff's equation
- QSAR models were developed for K_i using P-gp docking scores for the inhibitors and molecular descriptors of the inhibitors and the substrates and several predictive models of P-pg inhibition models have been produced
- Parameters of the substrates (used for the measurement of inhibitory activity of the inhibitors) improve the accuracy of the QSAR models for the prediction of K_i values.
- Docking scores could not significantly improve the prediction accuracy of the QSAR models
- Chi-squared (CHAID) model had the lowest K_i prediction error for the external validation set with a mean absolute error of 0.511.
- The QSAR models indicate that potent inhibitors of P-gp have higher lipophilicity and molecular size than lead-like compounds as defined by Oprea and the limiting lipophilicity is logP > 6.85 for this dataset.

Supplementary Table I. Dataset

Inhibitor	data split	Substrate	Substrate code	Cell System	Docking energy	References	Log K_i
4-hydroxycarvedilol	t	vinblastine	9	Caco-2	-13.125806	Neuhoff et al, 2000	2.0111102
5-hydroxycarvedilol	t	vinblastine	9	Caco-2	-12.778502	Neuhoff et al, 2000	2.1792258
Alfentanil	t	Digoxin	4	Caco-2	-15.107859	Ekins et al, 2002	2.0371199
Alprenolol	v	calcein	2	K562-MDR	-10.15932	Richter et al, 2008	1.9566486
Amiodarone 1	t	calcein	2	MDR1 transfected LLC-PK1	-14.167244	Ekins et al, 2002; Cook et al, 2010	0.9147727
Amiodarone 2	t	Daunomycin	3	MDR1 transfected LLC-PK1	-11.255543	Katoh et al, 2000	1.3515496
Amiodarone 3	t	Digoxin	4	MDR1 transfected LLC-PK1 or Caco-2	-14.647169	Cook et al, 2010	1.2649909
Amitriptyline	t	Daunomycin	3	P388 lymphoma cells	-11.625865	Lan et al, 1996	0.876795

Inhibitor	data split	Substrate	Substrate code	Cell System	Docking energy	References	Log K_i
Amprenavir	t	Prazosin	7	MDCK II-MDR1	-14.126471	Rautio et al 2006	2
Amprenavir	t	quinidine	8	MDCK II-MDR1	-14.126471	Lumen et al 2010	0.69897
Avasimibe (cook)	t	Digoxin	4	Caco-2	-15.538904	Cook et al, 2009	2.2954261
Avasimibe 2 (cook)	t	calcein	2	MDCK-MDR1	-15.538904	Cook et al, 2009	1.9862117
Azelastine 1	t	Daunomycin	3	MDR1 transfected LLC-PK1	-12.141804	Katoh et al, 2000	1.2034871
Azelastine 2	t	Digoxin	4	MDR1 transfected LLC-PK1	-12.030252	Katoh et al, 2000	1.4761353
Azithromycin	t	Digoxin	4	Caco-2	-17.220675	Ebrel et al, 2007	1.3263584
Bromocriptine 1	v	calcein	2	MDR1 transfected LLC-PK1	-13.442163	Ekins et al, 2002	0.4487063
Bromocriptine 2	t	vinblastine	9	MDR1 transfected LLC-PK1	-13.535645	Ekins et al, 2002	0.5976952
caffeine	t	Digoxin	4	Caco-2	-9.5322075	Ekins et al, 2002	1.9879019
Captopril 1 (cook)	t	Digoxin	4	Caco-2	-10.234739	Cook et al, 2009	2.9943961
Captopril 2 (cook)	t	calcein	2	MDCK-MDR1	-10.234739	Cook et al, 2009	1.9862117
Carvedilol 1	t	vinblastine	9	Caco-2	-16.581453	Neuhoff et al, 2000	1.0420461
Carvedilol 2 (cook)	t	Digoxin	4	Caco-2	-16.581453	Cook et al, 2009	0.5964561
Carvedilol 3 (cook)	t	calcein	2	MDCK-MDR1	-16.581453	Cook et al, 2009	1.2166606
Chlorpromazine 1	t	Daunomycin	3	P388 lymphoma cells	-10.26	Lan et al, 1996	0.382017
Chlorprothixene 2	t	calcein	2	MDCK II-MDR1	-9.7865658	Matsson & par 2009	2.0725507
Chlorzoxazone	t	Digoxin	4	Caco-2	-7.7284632	Ekins et al, 2002	1.9879019
Cimetidine 1 (cook)	t	Digoxin	4	Caco-2	-11.681512	Cook et al, 2009	2.9943961
Cimetidine 2 (cook)	t	calcein	2	MDCK-MDR1	-11.681512	Cook et al, 2009	1.9862117
Citalopram (cook)	v	Digoxin	4	Caco-2	-12.720747	Cook et al, 2009	1.7578241
Clarithromycin	t	Digoxin	4	Caco-2	-16.730564	Ebrel et al,2007; Cook et al;	1.538707

Inhibitor	data split	Substrate	Substrate code	Cell System	Docking energy	References	Log K_i
						2010	
Clarithromycin 2 (cook)	t	calcein	2	MDCK-MDR1	-16.730564	Cook et al, 2009	1.7420866
clotrimazole 1	v	calcein	2	MDR1 transfected LLC-PK1	-10.805565	Ekins et al, 2002	1.6434527
clotrimazole 2	t	vinblastine	9	MDR1 transfected LLC-PK1	-10.54844	Ekins et al, 2002	1.4759616
colchicine	t	Digoxin	4	Caco-2	-11.941282	Ekins et al, 2002	1.9879019
Conivaptan 1 (cook)	t	Digoxin	4	Caco-2	-13.845804	Cook et al, 2009	1.5854607
Conivaptan 2 (cook)	v	calcein	2	MDCK-MDR1	-13.845804	Cook et al, 2009	1.3286344
Cortisol	t	Digoxin	4	Caco-2	-11.844773	Ekins et al, 2002	1.7282646
CP100356	t	Digoxin	4	Caco-2	-15.003446	Wandel et al, 1999	-0.9551933
CP101556	t	Digoxin	4	Caco-2	-13.905317	Wandel et al, 1999	-0.2339469
CP114769	t	Digoxin	4	Caco-2	-13.585456	Wandel et al, 1999	-0.5349769
CP117227	t	Digoxin	4	Caco-2	-13.143414	Wandel et al, 1999	-1.1428664
CP12379	t	Digoxin	4	Caco-2	-12.3411	Wandel et al, 1999	-0.1670001
CP147478	t	Digoxin	4	Caco-2	-14.031638	Wandel et al, 1999	-0.8659701
CP69042	t	Digoxin	4	Caco-2	-16.359205	Wandel et al, 1999	0.3496297
CP99542	v	Digoxin	4	Caco-2	-14.624609	Wandel et al, 1999	0.5676855
Cyclosporin 2	t	Digoxin	4	Caco-2	-20.379795	Noguchi et al, 2009; Ekins et al, 2002; Cook et al, 2010	0.2889196
Cyclosporin 3	t	Daunomycin	3	P388 lymphoma cells	-20.379795	Lan et al, 1996	-1.4202164
Cyclospori	t	calcein	2	MDR1	-20.379795	Ekins et al,	0.4087

Inhibitor	data split	Substrate	Substrate code	Cell System	Docking energy	References	Log K_i
ne 1				transfected LLC-PK1		2002 ; IC50 from Cook et al	91
Cyclosporine 4	t	Prazosin	7	MDCK II-MDR1	-20.379795	Rautio et al 2006	-0.0087739
D-703	t	Digoxin	4	Caco-2	-9.8488	Pauli-Magnus et al, 2000	0.1920219
Daunomycin	t	Digoxin	4	average of Caco-2 from Ekins et al, and Caco-2 and MDCK-MDR1 from Tang et al	-18.165121	Ekins et al, 2002; Tang et al, 2002	1.9077499
Debrisoquine	t	Digoxin	4	Caco-2	-9.2348728	Ekins et al, 2002	1.9879019
Desethylamiodarone 1	t	Daunomycin	3	MDR1 transfected LLC-PK1	-15.265665	Katoh et al, 2000	1.1868878
Desethylamiodarone 2	t	Digoxin	4	MDR1 transfected LLC-PK1	-13.810134	Katoh et al, 2000 and Kakumoto et al, 2002	1.1205972
desipramine	t	calcein	2	K562-MDR	-13.140268	Richter et al, 2008	2.0433623
Desmethylazelastine 1	t	Daunomycin	3	MDR1 transfected LLC-PK1	-13.500105	Katoh et al, 2000	1.0712491
Desmethylazelastine 2	t	Digoxin	4	MDR1 transfected LLC-PK1	-12.200897	Katoh et al, 2000	1.6201904
Desmethylcarvedilol	v	vinblastine	9	Caco-2	-14.946373	Neuhoff et al, 2000	1.8938253
Diacetolol	v	vinblastine	9	Caco-2	-12.302096	Neuhoff et al, 2000	3.4509182
digoxin	t	calcein	2	K562-MDR	-17.8	Richter et al, 2008	1.9754318
dihydroergocristine 1	t	calcein	2	MDR1 transfected LLC-PK1	-14.04105	Ekins et al, 2002	2.7084209
dihydroergocristine 2	t	vinblastine	9	MDR1 transfected LLC-PK1	-14.04105	Ekins et al, 2002	1.20412
dihydroergocryptine	t	calcein	2	MDR1 transfected LLC-PK1	-13.441021	Ekins et al, 2002	2.5569053

Inhibitor	data split	Substrate	Substrate code	Cell System	Docking energy	References	Log K_i
1							
dihydroergocryptine 2	t	vinblastine	9	MDR1 transfected LLC-PK1	-12.909486	Ekins et al, 2002	1.2971037
dihydroergotamine 1	t	calcein	2	MDR1 transfected LLC-PK1	-14.209814	Ekins et al, 2002	3
dihydroergotamine 2	v	vinblastine	9	MDR1 transfected LLC-PK1	-12.999072	Ekins et al, 2002	2.0770043
Dilevalol	t	vinblastine	9	Caco-2	-15.486352	Neuhoff et al, 2000	2.9780939
Diltiazem 1	t	Daunomycin	3	P388 lymphoma cells	-12.050211	Lan et al, 1996	0.7331973
Diltiazem 2 (cook)	v	Digoxin	4	Caco-2	-12.050211	Cook et al, 2009	1.5506986
Diltiazem 3 (cook)	v	calcein	2	MDCK-MDR1	-12.050211	Cook et al, 2009	1.463333
Dipyridamole 1	v	Digoxin	4	MDR1 transfected LLC-PK1 and LLC-GA5-COL150 Cells	-13.658478	Kakumoto et al, 2002; Cook et al, 2010	1.516315
Dipyridamole 2	v	Daunomycin	3	P388 lymphoma cells	-13.658478	Lan et al, 1996	-0.2839967
Elacridar 1 (GF120918)	t	calcein	2	MDCK II-MDR1	-15.056113	Matsson P. et al, 2009	-0.5642714
Elacridar 2 (GF120918)	t	Prazosin	7	MDCK II-MDR1	-16.170345	Rautio et al 2006	-1.30103
Elacridar 3 (GF120918)	t	Digoxin	4	Caco-2	-16.167965	Tang et al 2002	-0.3819519
Elacridar 4 (GF120918) - efflux	t	Irinotecan (cpt -11)	6	MDCK II-MDR1	-16.167965	Luo et al, 2002	-0.506498
ergocornine 1	v	calcein	2	MDR1 transfected LLC-PK1	-14.900295	Ekins et al, 2002	2.0220157
ergocornine 2	t	vinblastine	9	MDR1 transfected LLC-PK1	-14.975269	Ekins et al, 2002	1.3891661
ergocristine 1	t	calcein	2	MDR1 transfected LLC-PK1	-14.826789	Ekins et al, 2002	1.6314438
ergocristin	t	vinblastine	9	MDR1	-14.826789	Ekins et al,	1.1248

Inhibitor	data split	Substrate	Substrate code	Cell System	Docking energy	References	Log K_i
e 2				transfected LLC-PK1		2002	301
ergocryptine 1	t	calcein	2	MDR1 transfected LLC-PK1	-14.927869	Ekins et al, 2002	1.0863598
ergocryptine 2	t	vinblastine	9	MDR1 transfected LLC-PK1	-14.927869	Ekins et al, 2002	0.808211
ergometrine 1	t	calcein	2	MDR1 transfected LLC-PK1	-12.537419	Ekins et al, 2002	2.062582
ergometrine 2	t	vinblastine	9	MDR1 transfected LLC-PK1	-12.546585	Ekins et al, 2002	2
ergotamine 1	t	calcein	2	MDR1 transfected LLC-PK1	-12.368077	Ekins et al, 2002	1.9951963
ergotamine 2	t	vinblastine	9	MDR1 transfected LLC-PK1	-12.368077	Ekins et al, 2002	1.1538149
Erlotinib	t	vincristine	10	K562-MDR	-13.533154	Richter et al, 2008	0.2521814
Erythromycin 1	t	calcein	2	MDR1 transfected LLC-PK1	-14.585213	Richter et al 2008 ; Ekins et al, 2002	2.7263196
Erythromycin 2	t	Digoxin	4	Caco-2	-14.198096	Eberl et al, 2007; Cook et al, 2010	2.3128383
Erythromycin 3	t	vinblastine	9	MDR1 transfected LLC-PK1	-14.590425	Ekins et al, 2002	1.5773769
Etoposide 1	t	Digoxin	4	average of Caco-2 and MDCK-MDR1	-16.388716	Tang et al 2002	2.7250945
Etoposide 2 efflux	v	Irinotecan (cpt -11)	6	MDCK II-MDR1	-16.388716	Luo et al, 2002	2.7891527
Felodipine 1 (cook)	t	Digoxin	4	Caco-2	-11.028463	Cook et al, 2009	1.4567941
Felodipine 2 (cook)	t	calcein	2	MDCK-MDR1	-11.028463	Cook et al, 2009	1.1903317
Fentanyl	t	Digoxin	4	Caco-2	-12.234112	Ekins et al, 2002	0.8008152
fexofenadine	v	Digoxin	4	Caco-2	-14.202948	Ekins et al, 2002	1.9879019
Fluconazole 1	t	calcein	2	MDR1 transfected LLC-PK1	-11.205918	Ekins et al, 2002	3
Fluconazole 2	t	vinblastine	9	MDR1 transfected LLC-PK1	-11.556297	Ekins et al, 2002	2.60206

Inhibitor	data split	Substrate	Substrate code	Cell System	Docking energy	References	Log K_i
Fluoxetine	v	Digoxin	4	Caco-2	-12.101847	Ekins et al, 2002	0.9879019
Fluphenazine	t	Daunomycin	3	P388 lymphoma cells	-13.385862	Lan et al, 1996	0.7419391
fluvastatin	t	calcein	2	K562-MDR	-12.218078	Richter et al, 2008	1.877947
Gallopamil	v	vinblastine	9	Caco-2	-16.382755	Neuhoff et al, 2000	0.1165631
Gemcabene 1 (cook)	t	Digoxin	4	Caco-2	-13.465055	Cook et al, 2009	2.9943961
Gemcabene 2 (cook)	t	calcein	2	MDCK-MDR1	-13.465055	Cook et al, 2009	1.162303
guanabenz	t	calcein	2	K562-MDR	-9.5497828	Richter et al, 2008	2.09691
Haloperidol	t	calcein	2	MDCK II-MDR1	-15.387538	Matsson P. et al, 2009	1.5496719
imipramine	t	calcein	2	K562-MDR	-10.665337	Richter et al, 2008	1.9542425
Indinavir 1	t	Digoxin	4	Caco-2	-12.465014	Choo et al, 2000	1.6313546
Indinavir 2	t	Prazosin	7	MDCK II-MDR1	-12.465014	Rautio et al 2006	1.69897
Isradipine 1 (cook)	t	Digoxin	4	Caco-2	-13.454189	Cook et al, 2009	1.1704874
Isradipine 3 (cook)	t	calcein	2	MDCK-MDR1	-13.454189	Cook et al, 2009	1.4775734
Itraconazole 1	t	Digoxin	4	Caco-2	-13.988383	Cook et al, 2009	0.5964561
Itraconazole 2 (cook)	t	calcein	2	MDCK-MDR1	-13.988383	Cook et al, 2009	-0.235637
Ivermectin	t	Digoxin	4	Caco-2	-16.138742	Ekins et al, 2002	0.9879019
Ketoconazole 1	v	calcein	2	MDR1 transfected LLC-PK1	-14.031419	Ekins et al, 2002 ; Cook et al, 2010	1.1930811
Ketoconazole 2	t	Prazosin	7	MDCK II-MDR1	-12.8391	Rautio et al 2006	0.376577
Ketoconazole 3	t	Digoxin	4	Caco-2	-12.8391	Cook et al, 2009 ; Ekins et al, 2002	0.4554584
Labetalol	t	vinblastine	9	Caco-2	-14.0572	Neuhoff et al, 2000	3.2456121
Lansoprazole	t	calcein	2	K562-MDR	-13.582537	Richter et al, 2008	1.6379898
Levofloxacin (cook)	v	Digoxin	4	Caco-2	-11.785893	Cook et al, 2009	2.6933661
Loperamid	v	quinidine	8	MDCK II-	-13.761807	Lumen et al	-1

Inhibitor	data split	Substrate	Substrate code	Cell System	Docking energy	References	Log K_i
e				MDR1		2010	
Loperamide 1	t	calcein	2	MDCK II-MDR1	-12.898769	Matsson P. et al, 2009	1.3735807
Loperamide 2	t	Digoxin	4	Caco-2	-13.761807	Ekins et al, 2002	0.4192656
Losartan 1 (cook)	t	Digoxin	4	Caco-2	-12.777067	Cook et al, 2009	2.1527586
Losartan 2 (cook)	t	calcein	2	MDCK-MDR1	-12.777067	Cook et al, 2009	1.9862117
Lovastatin	t	Digoxin	4	Caco-2	-12.66326	Ekins et al, 2002	0.9879019
LY335979 - 1 (Zosuquidar)	t	Digoxin	4	Caco-2	-17.302494	Choo et al, 2000	-1.6318869
LY335979 - 2 (Zosuquidar)	t	Abacavir	1	MDCK II-MDR1	-13.445078	Shaik et al 2007	-1.30103
Mefloquine	t	Daunomycin	3	P388 lymphoma cells	-12.642743	Lan et al, 1996	-0.3665315
Meloxicam (cook)	t	Digoxin	4	Caco-2	-12.929171	Cook et al, 2009	2.6933661
Mibefradil 1	t	Digoxin	4	Caco-2	-13.578841	Ekins et al, 2002; Cook et al, 2010	0.6563124
Mibefradil 2 (cook)	v	calcein	2	MDCK-MDR1	-13.578841	Cook et al, 2009	0.764363
Miconazole 1	t	calcein	2	MDR1 transfected LLC-PK1	-11.482579	Ekins et al, 2002	1.744293
Miconazole 2	t	vinblastine	9	MDR1 transfected LLC-PK1	-11.826444	Ekins et al, 2002	1.4209454
Midazolam 1	t	Digoxin	4	Caco-2	-11.014441	Ekins et al, 2002	1.7282646
Midazolam 2	t	calcein	2	K562-MDR	-11.014441	Richter et al, 2008	1.5676144
Milameline 1 (cook)	t	Digoxin	4	Caco-2	-8.0218182	Cook et al, 2009	2.9943961
Milameline 2 (cook)	v	calcein	2	MDCK-MDR1	-8.0218182	Cook et al, 2009	1.9862117
Mitomycin C 1	t	Digoxin	4	Caco-2	-12.410781	Ekins et al, 2002	0.9879019
MK571	v	calcein	2	MDCK II-MDR1	-13.198287	Matsson P. et al, 2009	1.3735807
Montelukast (cook)	v	Digoxin	4	Caco-2	-13.465913	Cook et al, 2009	0.8974861

Inhibitor	data split	Substrate	Substrate code	Cell System	Docking energy	References	Log K_i
Morphine	v	Digoxin	4	Caco-2	-10.813716	Ekins et al, 2002	1.9879019
Nelfinavir	t	Digoxin	4	Caco-2	-13.952493	Choo et al, 2000	0.1340299
Nicardipine 1 (cook)	t	Digoxin	4	Caco-2	-14.032031	Cook et al, 2009	0.8974861
Nicardipine 2 (cook)	t	calcein	2	MDCK-MDR1	-14.032031	Cook et al, 2009	0.609461
Nifedipine 1 (cook)	t	Digoxin	4	Caco-2	-11.731437	Cook et al, 2009	1.718672
Nifedipine 2 (cook)	t	calcein	2	MDCK-MDR1	-11.731437	Cook et al, 2009	1.6583096
Nitrendipine 1 (cook)	v	Digoxin	4	Caco-2	-11.945253	Cook et al, 2009	1.1405242
Nitrendipine 2 (cook)	t	calcein	2	MDCK-MDR1	-11.945253	Cook et al, 2009	1.5989956
Norgallopamil	t	vinblastine	9	Caco-2	-15.350445	Neuhoff et al, 2000	0.6415682
Norverapamil 1	t	vinblastine	9	Caco-2	-15.751464	Neuhoff et al, 2000	0.5317414
Norverapamil 2	v	Digoxin	4	Caco-2	-15.751464	Pauli-Magnus et al, 2000	-0.5349769
Omeprazole 1	t	vinblastine	9	Caco-2	-13.01254	Neuhoff et al, 2000	1.8537655
Omeprazole 2 (cook)	t	Digoxin	4	Caco-2	-13.01254	Cook et al, 2009	1.923815
Omeprazole 3 (cook)	t	calcein	2	MDCK-MDR1	-13.01254	Cook et al, 2009 ; Richter et al, 2008	1.6653934
Orlistat (cook)	t	Digoxin	4	Caco-2	-13.17302	Cook et al, 2009	2.6933661
Paclitaxel	t	Digoxin	4	Caco-2	-15.886762	Ekins et al, 2002	1.9879019
Pantoprazole 1 (cook)	t	Digoxin	4	Caco-2	-12.931607	Cook et al, 2009	1.8332452
Pantoprazole 2	t	calcein	2	K562-MDR	-12.931607	Richter et al, 2008	1.7323938
Paroxetine 1 (cook)	v	Digoxin	4	Caco-2	-12.471749	Cook et al, 2009	2.9943961
Paroxetine 2 (cook)	t	calcein	2	MDCK-MDR1	-12.471749	Cook et al, 2009	1.7715416
Procainamide	t	Digoxin	4	Caco-2	-10.075689	Ekins et al, 2002	0.9879019

Inhibitor	data split	Substrate	Substrate code	Cell System	Docking energy	References	Log K_i
Progesterone	t	Daunomycin	3	P388 lymphoma cells	-13.143506	Lan et al, 1996	0.6627578
Promethazine	t	Daunomycin	3	P388 lymphoma cells	-10.625359	Lan et al, 1996	0.5378191
Propafenone	t	Daunomycin	3	P388 lymphoma cells	-12.570536	Lan et al, 1996	-0.3565473
PSC-833 (Valspodar)	t	Digoxin	4	Caco-2	-20.943985	Wandel et al, 1999	-0.9586073
Pumafentrine	t	calcein	2	K562-MDR	-13.045033	Richter et al, 2008	0.1931246
Quinidine 1	t	Digoxin	4	MDCK II-MDR1	-14.042288	Cook et al, 2010	1.1903317
Quinidine 2	t	quinidine	8	MDCK II-MDR1	-14.042288	Lumen et al 2010	-1
Quinidine 3 (cook)	v	calcein	2	MDCK-MDR1	-14.042288	Cook et al, 2009; Richter et al 2008	0.8079154
Quinidine 4	t	Prazosin	7	MDCK II-MDR1	-14.042288	Rautio et al 2006	1.146128
Quinidine 5	v	Daunomycin	3	P388 lymphoma cells	-14.042288	Lan et al, 1996	0.3117539
Ranolazine 1 (cook)	v	Digoxin	4	Caco-2	-14.971668	Cook et al, 2009	1.6845922
Ranolazine 2 (cook)	t	calcein	2	MDCK-MDR1	-14.971668	Cook et al, 2009	1.5176906
Repaglinide (cook)	v	Digoxin	4	Caco-2	-13.316523	Cook et al, 2009	1.224845
Reserpine 1	v	calcein	2	MDR1 transfected LLC-PK1	-16.462879	Ekins et al, 2002	1.0863598
Reserpine 2	t	vinblastine	9	MDR1 transfected LLC-PK1	-16.462879	Ekins et al, 2002	-0.0132283
Reserpine 3	t	Digoxin	4	average of Caco-2 and MDCK-MDR1 data	-16.462879	Tang et al 2002	0.8088859
Reserpine 4	t	Daunomycin	3	P388 lymphoma cells	-16.462879	Lan et al, 1996	-0.853872
Ritonavir 1	t	Digoxin	4	Caco-2	-14.554341	Choo et al, 2000; Cook	0.7901279

Inhibitor	data split	Substrate	Substrate code	Cell System	Docking energy	References	Log K_i
						et al, 2010	
Ritonavir 2 (cook)	t	calcein	2	MDCK-MDR1	-14.554341	Cook et al, 2009	1.5425142
Roxithromycin	v	Digoxin	4	Caco-2	-15.752835	Eberl et al, 2007	1.1754226
Saquinavir 1	t	Digoxin	4	Caco-2	-15.479135	Choo et al, 2000; Cook et al, 2010	0.8809862
Saquinavir 2 (cook)	t	calcein	2	MDCK-MDR1	-15.479135	Cook et al, 2009	1.6489695
Sitagliptin (cook)	v	Digoxin	4	Caco-2	-13.506649	Cook et al, 2009	2.4715174
S-Mephenytoin	t	Digoxin	4	Caco-2	-9.6019487	Ekins et al, 2002	1.9879019
Sparfloxacin (cook)	t	Digoxin	4	Caco-2	-12.984799	Cook et al, 2009	2.4715174
Spironolactone	t	Daunomycin	3	P388 lymphoma cells	-12.21499	Lan et al, 1996	0.6170003
Sufentanil	v	Digoxin	4	Caco-2	-11.046084	Ekins et al, 2002	0.6111512
Talinolol 1 (cook)	t	Digoxin	4	Caco-2	-13.549482	Cook et al, 2009	2.4627435
Talinolol 2 (cook)	t	calcein	2	MDCK-MDR1	-13.549482	Cook et al, 2009	1.667453
Tamoxifen 1	v	Digoxin	4	Caco-2	-11.870214	Ekins et al, 2002	1.7282646
Tamoxifen 2	t	Daunomycin	3	P388 lymphoma cells	-11.870214	Lan et al, 1996	0.1430148
Telithromycin	t	Digoxin	4	Caco-2	-15.286554	Ebrel et al, 2007	0.2431744
Telmisartan 1 (cook)	t	Digoxin	4	Caco-2	-14.009093	Cook et al, 2009	0.6933661
Telmisartan 2 (cook)	t	calcein	2	MDCK-MDR1	-14.009093	Cook et al, 2009	0.764363
Terfenadine 1	t	Digoxin	4	Caco-2	-16.656696	Ekins et al, 2002	0.9879019
Terfenadine 2	t	Daunomycin	3	P388 lymphoma cells	-16.656696	Lan et al, 1996	-0.2006595
testosterone	t	calcein	2	K562-MDR	-13.23076	Richter et al, 2008	1.4502491
Thioridazine	v	calcein	2	MDCK II-MDR1	-10.468631	Matsson P. et al, 2009	1.6118198
Tiapamil	v	vinblastine	9	Caco-2	-13.103317	Neuhoff et al, 2000	0.98428
Tolafentri	v	calcein	2	K562-	-13.423183	Richter et al,	0.6748

Inhibitor	data split	Substrate	Substrate code	Cell System	Docking energy	References	Log K_i
ne				MDR		2008	611
Tolbutamide	t	Digoxin	4	Caco-2	-10.450907	Ekins et al, 2002	1.9879019
Trifluoperazine	v	Daunomycin	3	P388 lymphoma cells	-11.762858	Lan et al, 1996	0.5797836
Troglitazone 1 (cook)	t	Digoxin	4	Caco-2	-15.617843	Cook et al, 2009	1.4857578
Troglitazone 2 (cook)	t	calcein	2	MDCK-MDR1	-15.617843	Cook et al, 2009	1.2649653
Troleandomycin 1	t	calcein	2	Caco-2	-18.369055	Ekins et al, 2002	2.6842168
Troleandomycin 2	t	vinblastine	9	MDR1 transfected LLC-PK1	-16.349228	Ekins et al, 2002	1.9427024
Verapamil 1	t	Daunomycin	3	P388 lymphoma cells	-14.540354	Lan et al, 1996	-0.1611509
Verapamil 2	t	vinblastine	9	Caco-2	-14.540354	Neuhoff et al, 2000	0.0746372
Verapamil 3	t	Digoxin	4	Caco-2	-14.540354	Pauli-Magnus et al, 2000; Tang et al, 2002	0.9081235
Verapamil 4 (cook)	t	calcein	2	MDCK-MDR1	-14.540354	Cook et al, 2009; Richter et al, 2008	1.1814505
Verapamil 5 (efflux)	v	Irinotecan (cpt -11)	6	MDCK II-MDR1	-14.540354	Luo et al, 2002	2.2829343
Verapamil 6	t	Prazosin	7	MDCK II-MDR1	-14.540354	Rautio et al 2006	0.071882
Verapamil 7	t	fexofenadine	5	Caco-2	-14.540354	Petri et al, 2004	0.8983137
Verapamil 8	t	vincristine	10	K562-MDR	-14.540354	Richter et al, 2008	-0.7478186
Vinblastine 1	t	Digoxin	4	Average of Caco-2 and MDCK-MDR1 data	-12.455927	Tang et al 2002	1.7233059
Vinblastine 2	t	Prazosin	7	MDCK II-MDR1	-12.455927	Rautio et al 2006	1.3404441
Vincristine	t	Digoxin	4	Average of Caco-2	-15.685461	Tang et al 2002	2.1524412

Inhibitor	data split	Substrate	Substrate code	Cell System	Docking energy	References	Log K_i
				and MDCK-MDR1 data			

Supplementary Table II. A brief description of the most important molecular descriptors selected and used by the models.

Descriptor	Model	Description
a_nC	BT	Number of carbon atoms.
ASA_H	RF	Calculated water accessible surface area of all hydrophobic atoms.
apol	BT	Sum of the atomic polarisabilities.
b-ar	RT, RF	Number of aromatic bonds
b_double	CHAID	Number of double bonds.
chi0_C	BT	Carbon connectivity index (order 0). This is calculated as the sum of 1/sqrt(d_i) over all carbon atoms i with $d_i > 0$.
chi0v_C	BT	Carbon valence connectivity index (order 0). This is calculated as the sum of 1/sqrt(v_i) over all carbon atoms i with $v_i > 0$.
chi1v_C	BT, MARS	Carbon valence connectivity index (order 1). This is calculated as the sum of 1/sqrt over all bonds between carbon atoms i and j where $i < j$
Docking energy (MOE)	I-tree	Docking score (kcal/mol) for enzyme-ligand docking of the compounds into the active site of P-glycoprotein (Aller et al, 2009) calculated using MOE software.
fiA	RF	The fractions of compounds that is ionised at pH 7.4 as acid.
LogD(10)	RF, MARS	Logarithm of distribution coefficient D of a compound between octanol and buffer layers at pH value 10.
Num Rings 5	RT	Number of rings 5.
Opr_leadlike	MARS	This is one if and only if there are fewer than two violations from Oprea's lead like rules, otherwise zero.
Opr_nrot	CHAID	The number of rotatable bonds from Tudor Oprea.
PEOE_VSA-6	CHAID	Van der Waals surface area of atoms with atomic charge in the range [0.00, 0.05).
S-fU	CHAID	Fractions of compounds unionised for substrates.
SlogP	RT, BT, CHAID, RF, MARS	The octanol/water partition coefficient.
SlogP_VSA5	RT	Atoms with logP(o/w) contribution of (0.15,0.20].
Substrate	BT	The substrate type (categorical variable)

Descriptor	Model	Description
VDistEq	RF	Adjacency and Distance Matrix descriptor: VdistEq is the sum of $\log_2 m - p_i \log_2 p_i / m$ where m is the sum of the distance matrix entries, i represents all the atoms and p_i is the number of distance matrix entries equal to i.
vsa_pol	RT	Polar van der Waals surface area.
vsurf_CW1	I-tree	Capacity factor is the ratio of the hydrophilic surface over the total molecular surface, calculated at eight different energy levels (from -0.2 to -6.0 kcal/mol).
vsurf_CW4	RF	Capacity factor is the ratio of the hydrophilic surface over the total molecular surface, calculated at eight different energy levels (from -0.2 to -6.0 kcal/mol).
vsurf_CW7	RF	Capacity factor is the ratio of the hydrophilic surface over the total molecular surface, calculated at eight different energy levels (from -0.2 to -6.0 kcal/mol).
vsurf_D1	BT	Hydrophobic volume at -0.2 kcal/mol.
vsurf_D2	BT	Hydrophobic volume at -0.4 kcal/mol.
vsurf_D6	RF	Hydrophobic volume at -1.2 kcal/mol.
vsurf_D8	I-tree	Hydrophobic volume at -1.6 kcal/mol.
vsurf_HL1	RF	Hydrophilic-Lipophilic balance; it is the ratio between the hydrophilic regions measured at -3 and -4 kcal/mol and the hydrophobic regions measured at -0.6 and -0.8 kcal/mol. The balance describes which effect dominates in the molecule, or if they are roughly equally balanced.
vsurf_Wp1	BT	Polar (van der Waals) volume (8 descriptors).
vsurf_Wp2	BT	Polar (van der Waals) volume (8 descriptors).

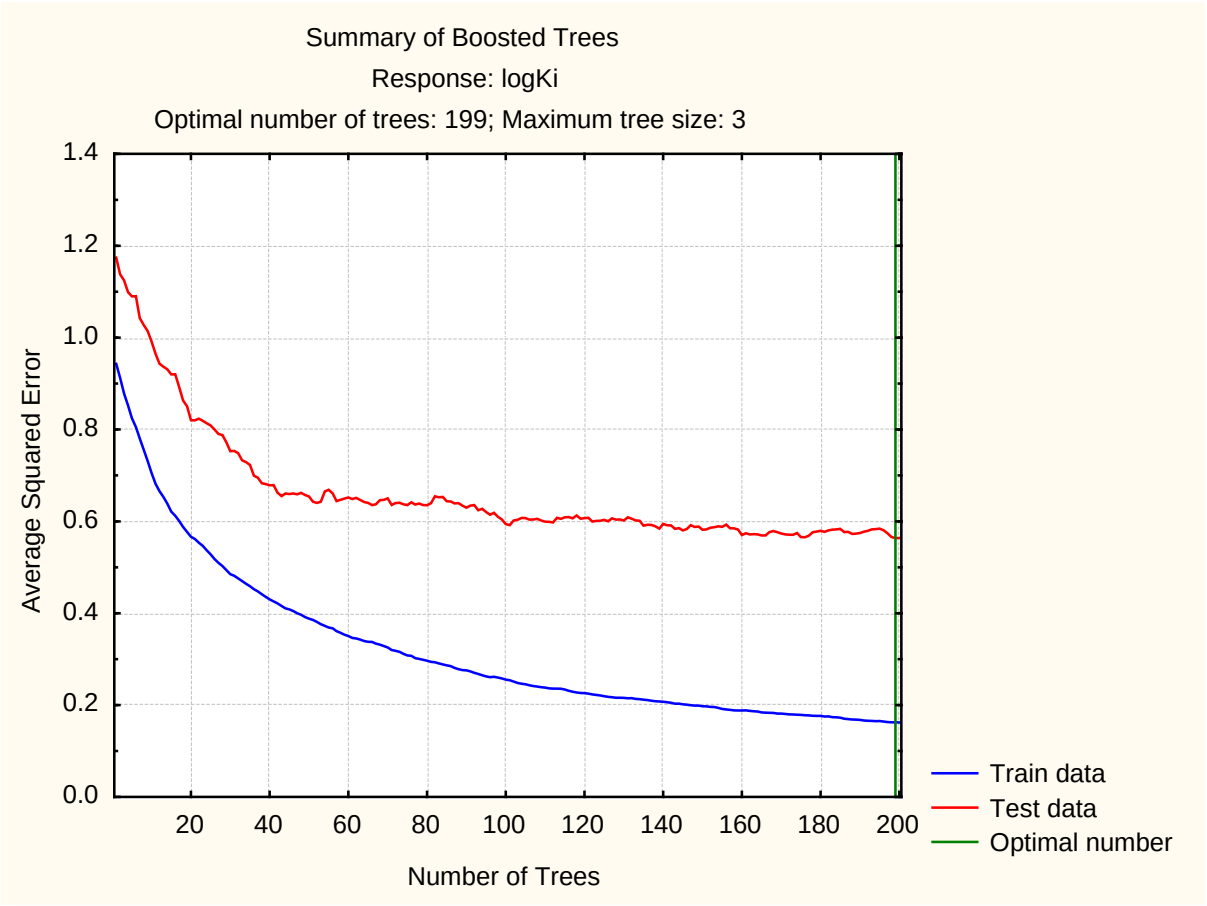

Figure S1. Effect of number of trees in boosted trees model on the average error for training and test sets

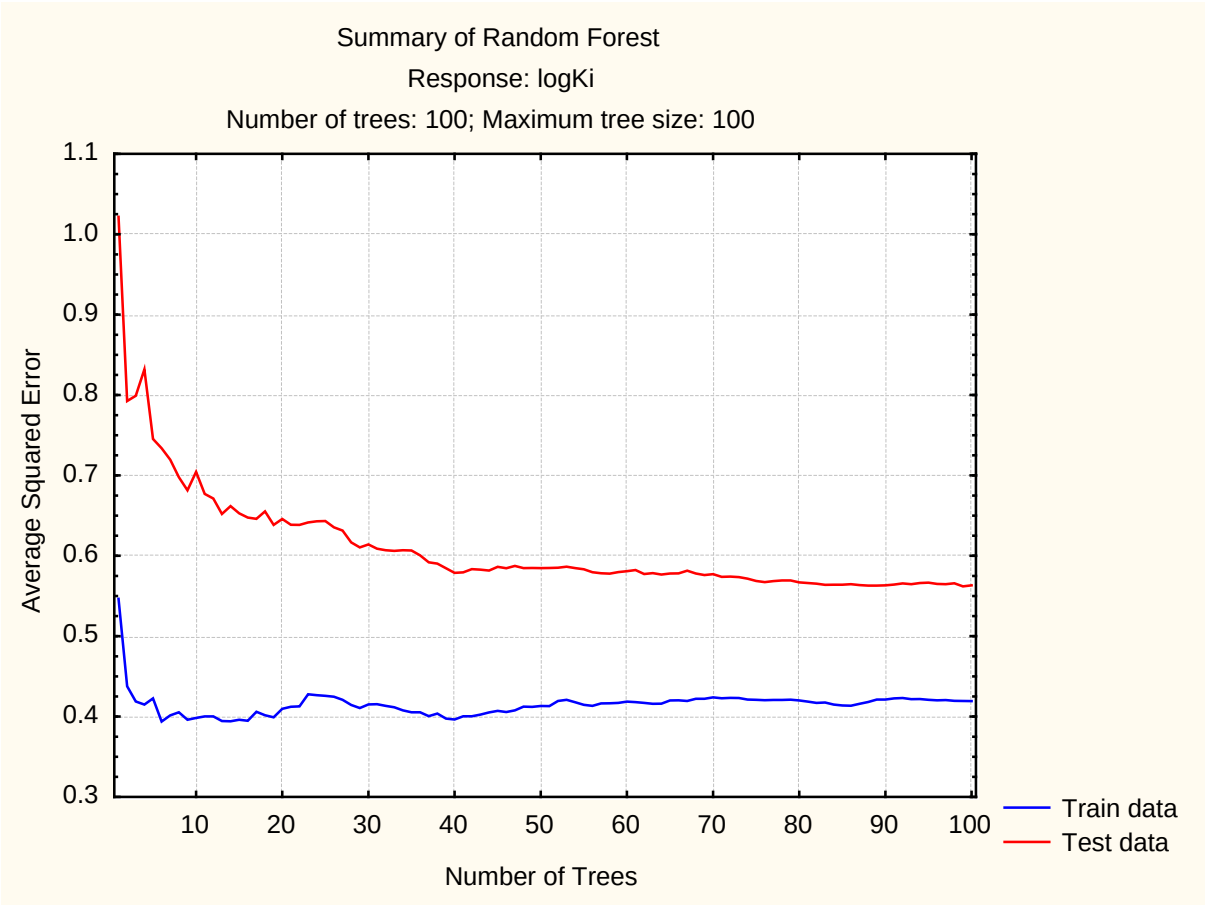

Figure S2. Effect of number of trees in random forest model on the average error for training and test sets