Supplementary File

Table S1. Correlation matrix of various pharmacophoric descriptors from R1 fragment training molecules.

Descriptors	H-D1-H	H-D2-HBA	HBA-D3-HBA	H-D4-H	H-D5-HBA	H-D6-HBA	HBA-D6-HBA	HBA-D6-HBD	H-D7-H	H-D8-HBA	H-D8-HBD	H-D9-HBA
H-D1-H	**1**	0.441	0.167	-0.354	0.611	0.240	0.091	0.354	0.645	0.560	0.645	0.804
H-D2-HBA	0.441	**1**	-0.220	0.624	0.735	0.635	0.000	0.000	0.683	0.741	0.683	0.665
HBA-D3-HBA	0.167	-0.220	**1**	-0.471	0.361	-0.520	0.867	0.884	0.258	0.420	0.258	0.251
H-D4-H	-0.354	0.624	-0.471	**1**	0.196	0.510	-0.258	-0.417	0.091	0.149	0.091	-0.036
H-D5-HBA	0.611	0.735	0.361	0.196	**1**	0.053	0.304	0.629	0.947	0.887	0.947	0.921
H-D6-HBA	0.240	0.635	-0.520	0.510	0.053	**1**	-0.132	-0.510	-0.062	0.135	-0.062	0.024
HBA-D6-HBA	0.091	0.000	0.867	-0.258	0.304	-0.132	**1**	0.710	0.141	0.499	0.141	0.138
HBA-D6-HBD	0.354	0.000	0.884	-0.417	0.629	-0.510	0.710	**1**	0.548	0.545	0.548	0.533
H-D7-H	0.645	0.683	0.258	0.091	0.947	-0.062	0.141	0.548	**1**	0.868	1.000	0.973
H-D8-HBA	0.560	0.741	0.420	0.149	0.887	0.135	0.499	0.545	0.868	**1**	0.868	0.844
H-D8-HBD	0.645	0.683	0.258	0.091	0.947	-0.062	0.141	0.548	1.000	0.868	**1**	0.973
H-D9-HBA	0.804	0.665	0.251	-0.036	0.921	0.024	0.138	0.533	0.973	0.844	0.973	**1**

Table S2. Correlation matrix of various pharmacophoric descriptors from R2 fragment training molecules.

Descriptors	H-D1-H	H-D2-HBA	H-D2-HBD	H-D3-HBA	H-D3-HBD	HBA-D3-HBD	H-D4-HBA	HBA-D4-HBA	HBA-D4-HBD	H-D5-HBA	H-D5-HBD	HBA-D5-HBA	H-D6-HBA	H-D7-H	H-D7-HBA
H-D1-H	**1**	0.731	-0.113	0.679	0.077	0.113	0.077	0.113	-0.135	0.132	0.615	0.113	0.881	0.881	0.881
H-D2-HBA	0.731	**1**	-0.113	0.283	-0.192	0.510	-0.192	0.510	0.336	0.439	0.885	0.510	0.881	0.881	0.881
H-D2-HBD	-0.113	-0.113	**1**	0.417	0.906	-0.417	0.906	-0.417	-0.198	0.194	-0.283	-0.417	-0.354	-0.354	-0.354
H-D3-HBA	0.679	0.283	0.417	**1**	0.510	-0.417	0.510	-0.417	-0.545	-0.258	0.113	-0.417	0.471	0.471	0.471
H-D3-HBD	0.077	-0.192	0.906	0.510	**1**	-0.510	1.000	-0.510	-0.336	0.175	-0.346	-0.510	-0.320	-0.320	-0.320
HBA-D3-HBD	0.113	0.510	-0.417	-0.417	-0.510	**1**	-0.510	1.000	0.891	0.710	0.679	1.000	0.354	0.354	0.354
H-D4-HBA	0.077	-0.192	0.906	0.510	1.000	-0.510	**1**	-0.510	-0.336	0.175	-0.346	-0.510	-0.320	-0.320	-0.320
HBA-D4-HBA	0.113	0.510	-0.417	-0.417	-0.510	1.000	-0.510	**1**	0.891	0.710	0.679	1.000	0.354	0.354	0.354
HBA-D4-HBD	-0.135	0.336	-0.198	-0.545	-0.336	0.891	-0.336	0.891	**1**	0.767	0.605	0.891	0.070	0.070	0.070

H-D5-HBA	0.132	0.439	0.194	-0.258	0.175	0.710	0.175	0.710	0.767	**1**	0.482	0.710	0.091	0.091	0.091
H-D5-HBD	0.615	0.885	-0.283	0.113	-0.346	0.679	-0.346	0.679	0.605	0.482	**1**	0.679	0.801	0.801	0.801
HBA-D5-HBA	0.113	0.510	-0.417	-0.417	-0.510	1.000	-0.510	1.000	0.891	0.710	0.679	**1**	0.354	0.354	0.354
H-D6-HBA	0.881	0.881	-0.354	0.471	-0.320	0.354	-0.320	0.354	0.070	0.091	0.801	0.354	**1**	1.000	1.000
H-D7-H	0.881	0.881	-0.354	0.471	-0.320	0.354	-0.320	0.354	0.070	0.091	0.801	0.354	1.000	**1**	1.000
H-D7-HBA	0.881	0.881	-0.354	0.471	-0.320	0.354	-0.320	0.354	0.070	0.091	0.801	0.354	1.000	1.000	**1**

Table S3. Correlation between R1 fragment pharmacophoric descriptors and its PCA factors.

Descriptors	F1	F2	F3	F4	F5	F6
H-D1-H	0.708	-0.072	-0.466	0.512	-0.080	0.087
H-D2-HBA	0.672	-0.698	0.231	0.058	0.014	-0.067
HBA-D3-HBA	0.447	0.831	0.288	0.056	-0.036	0.150
H-D4-H	0.009	-0.753	0.525	-0.367	-0.116	0.099
H-D5-HBA	0.969	-0.112	0.069	-0.127	-0.163	0.014
H-D6-HBA	0.011	-0.763	0.268	0.581	-0.084	-0.023
HBA-D6-HBA	0.393	0.612	0.612	0.297	0.078	-0.046
HBA-D6-HBD	0.675	0.684	0.097	-0.073	-0.220	-0.117
H-D7-H	0.954	-0.128	-0.160	-0.211	0.051	-0.016

H-D8-HBA	0.936	-0.075	0.238	0.042	0.245	0.017
H-D8-HBD	0.954	-0.128	-0.160	-0.211	0.051	-0.016
H-D9-HBA	0.956	-0.121	-0.265	-0.010	0.015	0.014

Table S4. Correlation between R2 fragment pharmacophoric descriptors and its PCA factors.

Descriptors	F1	F2	F3	F4	F5	F6
H-D1-H	0.534	0.790	0.059	-0.254	-0.120	0.092
H-D2-HBA	0.822	0.429	0.196	0.244	-0.146	-0.149
H-D2-HBD	-0.536	0.279	0.734	0.257	0.121	-0.123
H-D3-HBA	-0.084	0.928	0.131	-0.082	0.327	0.022
H-D3-HBD	-0.582	0.400	0.700	-0.088	-0.034	0.042
HBA-D3-HBD	0.843	-0.454	0.229	-0.111	0.117	-0.068
H-D4-HBA	-0.582	0.400	0.700	-0.088	-0.034	0.042
HBA-D4-HBA	0.843	-0.454	0.229	-0.111	0.117	-0.068
HBA-D4-HBD	0.641	-0.606	0.406	0.120	0.075	0.191
H-D5-HBA	0.484	-0.284	0.793	-0.117	-0.204	-0.041
H-D5-HBD	0.912	0.194	0.167	0.221	0.003	0.231
HBA-D5-HBA	0.843	-0.454	0.229	-0.111	0.117	-0.068
H-D6-HBA	0.783	0.603	-0.148	0.009	-0.004	-0.035
H-D7-H	0.783	0.603	-0.148	0.009	-0.004	-0.035
H-D7-HBA	0.783	0.603	-0.148	0.009	-0.004	-0.035

Table S5. Factor scores of R1 fragment training, test and external test set molecules.

Molecules	F1	F2	F3	F4
Training molecules				
R1a	-1.079	-2.402	1.492	1.025
R1b	-1.325	1.923	-0.261	-0.104
R1c	3.226	-0.409	0.716	-1.867
R1d	4.357	-0.318	-1.285	1.218
R1e	-0.379	3.065	1.209	0.509
R1f	-2.735	0.023	-1.710	-0.390
R1g	-2.065	-1.882	-0.161	-0.392
Test molecules				
Comp6-R1	5.313	-0.225	-1.601	0.499
Comp9-R1	-1.979	-0.754	-0.672	-0.098
Comp10-R1	-1.979	-0.754	-0.672	-0.098
Comp13-R1	-1.979	-0.754	-0.672	-0.098
External test molecules				
CP38	-0.083	2.812	0.836	0.938
CP40	-1.979	-0.754	-0.672	-0.098

CP51	-1.979	-0.754	-0.672	-0.098
CP96	-1.168	1.816	0.130	-0.378
CP110	-0.717	3.051	0.268	0.387

Table S6. Factor scores of R2 fragment training, test and external test set molecules.

Molecules	F1	F2	F3	F4
Training molecules				
R2a	-2.983	1.372	-0.340	0.590
R2b	-3.022	2.214	1.664	-0.765
R2c	5.210	3.051	-0.587	0.013
R2d	0.511	-2.242	-0.684	-0.849
R2e	-1.933	-0.341	-2.607	0.394
R2f	1.255	-2.465	-0.241	-0.066
R2g	0.962	-1.590	2.795	0.684
Test molecules				
Comp6-R2	-2.774	1.680	0.171	0.567
Comp9-R2	5.099	1.940	-0.651	0.001
Comp10-R2	0.262	-2.454	-0.063	0.691

Comp13-R2	-2.350	-0.253	-2.598	-1.152
External test molecules				
CP38	-2.602	0.861	-0.408	-0.235
CP40	-2.080	2.164	0.073	2.243
CP51	-1.053	1.103	0.336	1.882
CP96	-1.053	1.103	0.336	1.882
CP110	-2.600	1.753	1.773	-1.694

Table S7. Z score and percentiles for R1 and R2 fragment training, test and external test set molecules.

Molecules	Z score on Scaled Biological Activity	Z score on Predicted Biological Activity	Molecules	Z score on Scaled Biological Activity	Z score on Predicted Biological Activity
Training molecules					
R1a	-0.833437498	-0.619089063	R2a	-1.309531488	-1.137403081
R1b	-1.01504779	-0.59720668	R2b	-0.977550687	-1.390117681
R1c	-0.139282602	-0.130382527	R2c	-0.04453275	-0.055714637
R1d	2.043067744	2.127605784	R2d	-0.341073335	-0.139224133
R1e	-0.139282602	-0.723030378	R2e	0.425592567	0.531912391
R1f	0.21317589	0.280823905	R2f	1.626220301	1.281311741

R1g	-0.129193142	-0.338721041	R2g	0.620875391	0.909235401
Test molecules					
Comp6-R1	1.5	2.572	Comp6-R2	-0.844308462	-0.707
Comp9-R1	-0.5	-0.110	Comp9-R2	0.525614436	0.661
Comp10-R1	-0.5	-0.110	Comp10-R2	1.14931917	2.350
Comp13-R1	-0.5	-0.110	Comp13-R2	-0.830625144	0.144
External test molecules					
CP38-R1	-1.11402	-0.680	CP38-R2	-1.11402	-0.707
CP40-R1	0.702928	-0.110	CP40-R2	0.702928	-0.540
CP51-R1	0.721263	-0.157	CP51-R2	0.721263	-0.205
CP96-R1	0.764561	-0.641	CP96-R2	0.764561	-0.205
CP110-R1	-1.0772	-0.664	CP110-R2	-1.0772	-0.801