Electronic Supplementary Material (ESI) for RSC Advances.

Supplementary 1

Table S1.1. Train and Test sets_ Seed 1

No.	Ligand Name	QSAR Set	Activity	No.	Ligand Name	QSAR Set	Activity
1	2.LHD	training	3.593	41	49.PHR	training	4.751
2	3.LHE	training	6.136	42	50.PHS	training	4.042
3	4.LHF	training	3.628	43	51.PHT	training	6.247
4	6.LHH	training	4.836	44	52.PHV	training	3.335
5	7.LHI	training	6.531	45	53.PHW	training	6.535
6	8.LHK	training	4.225	46	54.PHY	training	4.227
7	9.LHL	training	5.920	47	57.PWE	training	1.096
8	11.LHN	training	5.148	48	58.PWF	training	0.919
9	13.LHR	training	5.184	49	59.PWG	training	2.687
10	14.LHS	training	4.293	50	61.PWI	training	1.396
11	16.LHV	training	3.481	51	63.PWL	training	1.096
12	17.LHW	training	6.791	52	64.PWM	training	0.796
13	18.LHY	training	4.203	53	65.PWN	training	2.104
14	19.LWA	training	1.192	54	66.PWQ	training	1.202
15	20.LWD	training	1.717	55	67.PWR	training	2.705
16	21.LWE	training	1.717	56	68.PWS	training	1.096
17	22.LWF	training	1.414	57	70.PWV	training	1.008
18	24.LWH	training	3.212	58	71.PWW	training	2.899
19	25.LWI	training	1.111	59	72.PWY	training	1.114
20	26.LWK	training	1.899	60	73.RHA	training	5.205
21	27.LWL	training	0.606	61	74.RHD	training	3.304
22	28.LWM	training	1.394	62	75.RHE	training	5.096
23	29.LWN	training	1.313	63	76.RHF	training	3.300
24	30.LWQ	training	2.505	64	77.RHG	training	5.725
25	31.LWR	training	2.909	65	78.RHH	training	3.296
26	32.LWS	training	2.020	66	79.RHI	training	4.806
27	33.LWT	training	2.020	67	80.RHK	training	2.694
28	34.LWV	training	1.616	68	81.RHL	training	3.501
29	35.LWW	training	3.515	69	82.RHM	training	3.218
30	36.LWY	training	2.222	70	83.RHN	training	5.713
31	37.PHA	training	5.793	71	84.RHQ	training	3.108
32	38.PHD	training	4.622	72	85.RHR	training	4.302
33	39.PHE	training	6.152	73	86.RHS	training	3.386
34	40.PHF	training	3.916	74	87.RHT	training	5.987
35	42.PHH	training	6.051	75	88.RHV	training	3.206
36	43.PHI	training	4.916	76	89.RHW	training	5.878
37	44.PHK	training	3.426	77	91.RWA	training	1.212
38	45.PHL	training	5.311	78	92.RWD	training	0.909
39	46.PHM	training	3.714	79	93.RWE	training	1.091
40	48.PHQ	training	3.718	80	94.RWF	training	0.909

No.	Ligand Name	QSAR Set	Activity	No.	Ligand Name	QSAR Set	Activity
81	95.RWG	training	1.717	121	162.EYY	training	3.417
82	96.RWH	training	1.091	122	163.HYY	training	2.257
83	97.RWI	training	1.232	123	164.KYY	training	2.257
84	98.RWK	training	0.606	124	165.RYY	training	2.257
85	101.RWN	training	2.404	125	166.AYY	training	3.071
86	102.RWQ	training	0.606	126	169.FYY	training	1.911
87	103.RWR	training	2.384	127	172.GYY	training	5.071
88	105.RWT	training	3.818	128	174.QYY	training	5.071
89	106.RWV	training	0.606	129	175.MYY	training	1.991
90	108.RWY	training	0.808	130	176.SYY	training	3.070
91	110.EHH	training	0.905	131	177.TYY	training	3.070
92	114.IHH	training	2.020	132	178.CYY	training	0.470
93	115.FHH	training	1.803	133	179.YDY	training	3.047
94	117.YHH	training	1.803	134	180.YEY	training	3.047
95	118.GHH	training	1.089	135	181.YHY	training	9.886
96	119.NHH	training	1.089	136	182.YKY	training	9.886
97	120.QHH	training	1.089	137	185.YIY	training	3.607
98	121.MHH	training	2.015	138	186.YLY	training	3.607
99	122.SHH	training	1.320	139	187.YFY	training	2.233
100	123.THH	training	1.320	140	190.YGY	training	3.366
101	125.HDH	training	1.477	141	191.YNY	training	3.366
102	130.HAH	training	0.952	142	192.YQY	training	3.366
103	131.HIH	training	0.952	143	193.YMY	training	1.780
104	134.HWH	training	2.026	144	195.YTY	training	3.447
105	135.HYH	training	2.026	145	196.YCY	training	3.087
106	136.HGH	training	0.832	146	197.YYD	training	4.116
107	137.HNH	training	0.832	147	198.YYE	training	4.116
108	138.HQH	training	0.832	148	199.YYH	training	5.303
109	139.HMH	training	0.873	149	201.YYR	training	5.303
110	140.HSH	training	0.730	150	202.YYA	training	3.344
111	141.HTH	training	0.730	151	204.YYL	training	3.344
112	142.HCH	training	0.975	152	205.YYF	training	4.050
113	143.HHD	training	0.188	153	206.YYW	training	4.050
114	151.HHF	training	3.612	154	208.YYG	training	2.996
115	153.HHY	training	3.612	155	209.YYN	training	2.996
116	154.HHG	training	0.317	156	211.YYM	training	2.103
117	155.HHN	training	0.317	157	213.YYT	training	3.983
118	156.HHQ	training	0.317	158	214.YYC	training	0.637
119	160.HHC	training	0.128				
120	161.DYY	training	3.417				

No.	Ligand Name	QSAR Set	Activity
159	1.LHA	test	3.918
160	5.LHG	test	6.697
161	10.LHM	test	4.504
162	12.LHQ	test	4.136
163	15.LHT	test	5.584
164	23.LWG	test	1.313
165	41.PHG	test	5.197
166	47.PHN	test	6.061
167	55.PWA	test	1.396
168	56.PWD	test	1.096
169	60.PWH	test	1.184
170	62.PWK	test	0.407
171	69.PWT	test	2.598
172	90.RHY	test	3.378
173	99.RWL	test	3.212
174	100.RWM	test	0.727
175	104.RWS	test	0.808
176	107.RWW	test	2.707
177	109.DHH	test	0.905
178	113.AHH	test	2.020
179	116.WHH	test	1.803
180	124.CHH	test	0.937
181	126.HEH	test	1.477
182	132.HLH	test	0.952
183	133.HFH	test	2.026
184	144.HHE	test	0.188
185	152.HHW	test	3.612
186	167.IYY	test	3.071
187	168.LYY	test	3.071
188	170.WYY	test	1.911
189	173.NYY	test	5.071
190	183_YRY	test	9.886
191	184.YAY	test	3.607
192	188.YWY	test	2.233
193	194.YSY	test	3.447
194	200.YYK	test	5.303
195	203.YYI	test	3.344
196	210.YYQ	test	2.996
197	212.YYS	test	3.983

Table S1.2. Train and Test sets_ Seed 2

No	Ligand Name	QSAR Set	Activity	No	Ligand Name	QSAR Set	Activity
1	1.LHA	training	3.918	41	51.PHT	training	6.247
2	2.LHD	training	3.593	42	52.PHV	training	3.335
3	3.LHE	training	6.136	43	53.PHW	training	6.535
4	4.LHF	training	3.628	44	54.PHY	training	4.227
5	5.LHG	training	6.697	45	55.PWA	training	1.396
6	6.LHH	training	4.836	46	56.PWD	training	1.096
7	7.LHI	training	6.531	47	58.PWF	training	0.919
8	8.LHK	training	4.225	48	60.PWH	training	1.184
9	9.LHL	training	5.920	49	62.PWK	training	0.407
10	10.LHM	training	4.504	50	63.PWL	training	1.096
11	11.LHN	training	5.148	51	64.PWM	training	0.796
12	12.LHQ	training	4.136	52	65.PWN	training	2.104
13	13.LHR	training	5.184	53	66.PWQ	training	1.202
14	15.LHT	training	5.584	54	67.PWR	training	2.705
15	16.LHV	training	3.481	55	68.PWS	training	1.096
16	18.LHY	training	4.203	56	69.PWT	training	2.598
17	20.LWD	training	1.717	57	70.PWV	training	1.008
18	22.LWF	training	1.414	58	72.PWY	training	1.114
19	23.LWG	training	1.313	59	73.RHA	training	5.205
20	24.LWH	training	3.212	60	74.RHD	training	3.304
21	25.LWI	training	1.111	61	76.RHF	training	3.300
22	26.LWK	training	1.899	62	77.RHG	training	5.725
23	27.LWL	training	0.606	63	80.RHK	training	2.694
24	28.LWM	training	1.394	64	81.RHL	training	3.501
25	29.LWN	training	1.313	65	82.RHM	training	3.218
26	30.LWQ	training	2.505	66	83.RHN	training	5.713
27	31.LWR	training	2.909	67	85.RHR	training	4.302
28	32.LWS	training	2.020	68	86.RHS	training	3.386
29	33.LWT	training	2.020	69	87.RHT	training	5.987
30	34.LWV	training	1.616	70	88.RHV	training	3.206
31	36.LWY	training	2.222	71	89.RHW	training	5.878
32	38.PHD	training	4.622	72	90.RHY	training	3.378
33	40.PHF	training	3.916	73	91.RWA	training	1.212
34	41.PHG	training	5.197	74	92.RWD	training	0.909
35	42.PHH	training	6.051	75	93.RWE	training	1.091
36	44.PHK	training	3.426	76	94.RWF	training	0.909
37	47.PHN	training	6.061	77	95.RWG	training	1.717
38	48.PHQ	training	3.718	78	96.RWH	training	1.091
39	49.PHR	training	4.751	79	97.RWI	training	1.232
40	50.PHS	training	4.042	80	98.RWK	training	0.606

No	Ligand Name	QSAR Set	Activity	No	Ligand Name	QSAR Set	Activity
81	99.RWL	training	3.212	121	166.AYY	training	3.071
82	100.RWM	training	0.727	122	167.IYY	training	3.071
83	101.RWN	training	2.404	123	168.LYY	training	3.071
84	103.RWR	training	2.384	124	169.FYY	training	1.911
85	104.RWS	training	0.808	125	170.WYY	training	1.911
86	105.RWT	training	3.818	126	172.GYY	training	5.071
87	106.RWV	training	0.606	127	173.NYY	training	5.071
88	107.RWW	training	2.707	128	174.QYY	training	5.071
89	108.RWY	training	0.808	129	176.SYY	training	3.070
90	109.DHH	training	0.905	130	178.CYY	training	0.470
91	110.EHH	training	0.905	131	179.YDY	training	3.047
92	114.IHH	training	2.020	132	180.YEY	training	3.047
93	116.WHH	training	1.803	133	181.YHY	training	9.886
94	117.YHH	training	1.803	134	182.YKY	training	9.886
95	119.NHH	training	1.089	135	183_YRY	training	9.886
96	120.QHH	training	1.089	136	185.YIY	training	3.607
97	121.MHH	training	2.015	137	186.YLY	training	3.607
98	123.THH	training	1.320	138	187.YFY	training	2.233
99	124.CHH	training	0.937	139	188.YWY	training	2.233
100	125.HDH	training	1.477	140	190.YGY	training	3.366
101	126.HEH	training	1.477	141	191.YNY	training	3.366
102	131.HIH	training	0.952	142	192.YQY	training	3.366
103	132.HLH	training	0.952	143	193.YMY	training	1.780
104	133.HFH	training	2.026	144	194.YSY	training	3.447
105	135.HYH	training	2.026	145	195.YTY	training	3.447
106	136.HGH	training	0.832	146	196.YCY	training	3.087
107	137.HNH	training	0.832	147	198.YYE	training	4.116
108	140.HSH	training	0.730	148	199.YYH	training	5.303
109	141.HTH	training	0.730	149	200.YYK	training	5.303
110	142.HCH	training	0.975	150	201.YYR	training	5.303
111	144.HHE	training	0.188	151	203.YYI	training	3.344
112	151.HHF	training	3.612	152	204.YYL	training	3.344
113	152.HHW	training	3.612	153	205.YYF	training	4.050
114	153.HHY	training	3.612	154	206.YYW	training	4.050
115	154.HHG	training	0.317	155	208.YYG	training	2.996
116	156.HHQ	training	0.317	156	210.YYQ	training	2.996
117	160.HHC	training	0.128	157	211.YYM	training	2.103
118	161.DYY	training	3.417	158	212.YYS	training	3.983
119	163.HYY	training	2.257				
120	164.KYY	training	2.257				

No	Ligand Name	QSAR Set	Activity
159	14.LHS	test	4.293
160	17.LHW	test	6.791
161	19.LWA	test	1.192
162	21.LWE	test	1.717
163	35.LWW	test	3.515
164	37.PHA	test	5.793
165	39.PHE	test	6.152
166	43.PHI	test	4.916
167	45.PHL	test	5.311
168	46.PHM	test	3.714
169	57.PWE	test	1.096
170	59.PWG	test	2.687
171	61.PWI	test	1.396
172	71.PWW	test	2.899
173	75.RHE	test	5.096
174	78.RHH	test	3.296
175	79.RHI	test	4.806
176	84.RHQ	test	3.108
177	102.RWQ	test	0.606
178	113.AHH	test	2.020
179	115.FHH	test	1.803
180	118.GHH	test	1.089
181	122.SHH	test	1.320
182	130.HAH	test	0.952
183	134.HWH	test	2.026
184	138.HQH	test	0.832
185	139.HMH	test	0.873
186	143.HHD	test	0.188
187	155.HHN	test	0.317
188	162.EYY	test	3.417
189	165.RYY	test	2.257
190	175.MYY	test	1.991
191	177.TYY	test	3.070
192	184.YAY	test	3.607
193	197.YYD	test	4.116
194	202.YYA	test	3.344
195	209.YYN	test	2.996
196	213.YYT	test	3.983
197	214.YYC	test	0.637

Table S1.3. Train and Test sets_ Seed 3

No.	Ligand Name	QSAR Set	Activity	No.	Ligand Name	QSAR Set	Activity
1	2.LHD	training	3.593	41	51.PHT	training	6.247
2	3.LHE	training	6.136	42	54.PHY	training	4.227
3	4.LHF	training	3.628	43	56.PWD	training	1.096
4	5.LHG	training	6.697	44	57.PWE	training	1.096
5	7.LHI	training	6.531	45	58.PWF	training	0.919
6	9.LHL	training	5.92	46	59.PWG	training	2.687
7	10.LHM	training	4.504	47	60.PWH	training	1.184
8	11.LHN	training	5.148	48	61.PWI	training	1.396
9	12.LHQ	training	4.136	49	62.PWK	training	0.407
10	13.LHR	training	5.184	50	63.PWL	training	1.096
11	14.LHS	training	4.293	51	65.PWN	training	2.104
12	17.LHW	training	6.791	52	66.PWQ	training	1.202
13	18.LHY	training	4.203	53	67.PWR	training	2.705
14	21.LWE	training	1.717	54	69.PWT	training	2.598
15	22.LWF	training	1.414	55	70.PWV	training	1.008
16	24.LWH	training	3.212	56	71.PWW	training	2.899
17	25.LWI	training	1.111	57	72.PWY	training	1.114
18	26.LWK	training	1.899	58	74.RHD	training	3.304
19	27.LWL	training	0.606	59	75.RHE	training	5.096
20	28.LWM	training	1.394	60	76.RHF	training	3.3
21	29.LWN	training	1.313	61	77.RHG	training	5.725
22	30.LWQ	training	2.505	62	78.RHH	training	3.296
23	31.LWR	training	2.909	63	79.RHI	training	4.806
24	32.LWS	training	2.02	64	81.RHL	training	3.501
25	33.LWT	training	2.02	65	82.RHM	training	3.218
26	34.LWV	training	1.616	66	83.RHN	training	5.713
27	35.LWW	training	3.515	67	84.RHQ	training	3.108
28	36.LWY	training	2.222	68	85.RHR	training	4.302
29	37.PHA	training	5.793	69	86.RHS	training	3.386
30	38.PHD	training	4.622	70	87.RHT	training	5.987
31	40.PHF	training	3.916	71	88.RHV	training	3.206
32	41.PHG	training	5.197	72	90.RHY	training	3.378
33	42.PHH	training	6.051	73	91.RWA	training	1.212
34	44.PHK	training	3.426	74	92.RWD	training	0.909
35	45.PHL	training	5.311	75	94.RWF	training	0.909
36	46.PHM	training	3.714	76	95.RWG	training	1.717
37	47.PHN	training	6.061	77	96.RWH	training	1.091
38	48.PHQ	training	3.718	78	97.RWI	training	1.232
39	49.PHR	training	4.751	79	98.RWK	training	0.606
40	50.PHS	training	4.042	80	100.RWM	training	0.727

81	101.RWN	training	2.404	121	166.AYY	training	3.071
82	103.RWR	training	2.384	122	168.LYY	training	3.071
83	104.RWS	training	0.808	123	169.FYY	training	1.911
84	106.RWV	training	0.606	124	170.WYY	training	1.911
85	108.RWY	training	0.808	125	172.GYY	training	5.071
86	109.DHH	training	0.905	126	173.NYY	training	5.071
87	113.AHH	training	2.02	127	174.QYY	training	5.071
88	114.IHH	training	2.02	128	176.SYY	training	3.07
89	115.FHH	training	1.803	129	177.TYY	training	3.07
90	117.YHH	training	1.803	130	179.YDY	training	3.047
91	118.GHH	training	1.089	131	181.YHY	training	9.886
92	119.NHH	training	1.089	132	182.YKY	training	9.886
93	120.QHH	training	1.089	133	183_YRY	training	9.886
94	121.MHH	training	2.015	134	184.YAY	training	3.607
95	122.SHH	training	1.32	135	185.YIY	training	3.607
96	123.THH	training	1.32	136	187.YFY	training	2.233
97	124.CHH	training	0.937	137	188.YWY	training	2.233
98	125.HDH	training	1.477	138	190.YGY	training	3.366
99	126.HEH	training	1.477	139	192.YQY	training	3.366
100	130.HAH	training	0.952	140	193.YMY	training	1.78
101	131.HIH	training	0.952	141	194.YSY	training	3.447
102	132.HLH	training	0.952	142	195.YTY	training	3.447
103	134.HWH	training	2.026	143	196.YCY	training	3.087
104	135.HYH	training	2.026	144	197.YYD	training	4.116
105	137.HNH	training	0.832	145	198.YYE	training	4.116
106	138.HQH	training	0.832	146	199.YYH	training	5.303
107	139.HMH	training	0.873	147	200.YYK	training	5.303
108	140.HSH	training	0.73	148	201.YYR	training	5.303
109	141.HTH	training	0.73	149	202.YYA	training	3.344
110	144.HHE	training	0.188	150	203.YYI	training	3.344
111	151.HHF	training	3.612	151	204.YYL	training	3.344
112	152.HHW	training	3.612	152	205.YYF	training	4.05
113	153.HHY	training	3.612	153	208.YYG	training	2.996
114	154.HHG	training	0.317	154	209.YYN	training	2.996
115	155.HHN	training	0.317	155	210.YYQ	training	2.996
116	156.HHQ	training	0.317	156	212.YYS	training	3.983
117	160.HHC	training	0.128	157	213.YYT	training	3.983
118	162.EYY	training	3.417	158	214.YYC	training	0.637
119	163.HYY	training	2.257				
120	165.RYY	training	2.257				

159	1.LHA	test	3.918
160	6.LHH	test	4.836
161	8.LHK	test	4.225
162	15.LHT	test	5.584
163	16.LHV	test	3.481
164	19.LWA	test	1.192
165	20.LWD	test	1.717
166	23.LWG	test	1.313
167	39.PHE	test	6.152
168	43.PHI	test	4.916
169	52.PHV	test	3.335
170	53.PHW	test	6.535
171	55.PWA	test	1.396
172	64.PWM	test	0.796
173	68.PWS	test	1.096
174	73.RHA	test	5.205
175	80.RHK	test	2.694
176	89.RHW	test	5.878
177	93.RWE	test	1.091
178	99.RWL	test	3.212
179	102.RWQ	test	0.606
180	105.RWT	test	3.818
181	107.RWW	test	2.707
182	110.EHH	test	0.905
183	116.WHH	test	1.803
184	133.HFH	test	2.026
185	136.HGH	test	0.832
186	142.HCH	test	0.975
187	143.HHD	test	0.188
188	161.DYY	test	3.417
189	164.KYY	test	2.257
190	167.IYY	test	3.071
191	175.MYY	test	1.991
192	178.CYY	test	0.47
193	180.YEY	test	3.047
194	186.YLY	test	3.607
195	191.YNY	test	3.366
196	206.YYW	test	4.05
197	211.YYM	test	2.103

Table S1.4. Train and Test sets_ Seed 4

No.	Ligand Name	QSAR Set	Activity	No.	Ligand Name	QSAR Set	Activity
1	1.LHA	training	3.918	41	53.PHW	training	6.535
2	2.LHD	training	3.593	42	54.PHY	training	4.227
3	3.LHE	training	6.136	43	55.PWA	training	1.396
4	4.LHF	training	3.628	44	56.PWD	training	1.096
5	5.LHG	training	6.697	45	58.PWF	training	0.919
6	7.LHI	training	6.531	46	59.PWG	training	2.687
7	8.LHK	training	4.225	47	60.PWH	training	1.184
8	9.LHL	training	5.92	48	61.PWI	training	1.396
9	10.LHM	training	4.504	49	62.PWK	training	0.407
10	11.LHN	training	5.148	50	63.PWL	training	1.096
11	14.LHS	training	4.293	51	64.PWM	training	0.796
12	16.LHV	training	3.481	52	65.PWN	training	2.104
13	18.LHY	training	4.203	53	66.PWQ	training	1.202
14	19.LWA	training	1.192	54	67.PWR	training	2.705
15	21.LWE	training	1.717	55	68.PWS	training	1.096
16	24.LWH	training	3.212	56	70.PWV	training	1.008
17	25.LWI	training	1.111	57	71.PWW	training	2.899
18	26.LWK	training	1.899	58	72.PWY	training	1.114
19	28.LWM	training	1.394	59	73.RHA	training	5.205
20	29.LWN	training	1.313	60	74.RHD	training	3.304
21	30.LWQ	training	2.505	61	75.RHE	training	5.096
22	31.LWR	training	2.909	62	76.RHF	training	3.3
23	32.LWS	training	2.02	63	77.RHG	training	5.725
24	34.LWV	training	1.616	64	79.RHI	training	4.806
25	35.LWW	training	3.515	65	80.RHK	training	2.694
26	36.LWY	training	2.222	66	82.RHM	training	3.218
27	37.PHA	training	5.793	67	83.RHN	training	5.713
28	38.PHD	training	4.622	68	84.RHQ	training	3.108
29	39.PHE	training	6.152	69	86.RHS	training	3.386
30	40.PHF	training	3.916	70	87.RHT	training	5.987
31	41.PHG	training	5.197	71	88.RHV	training	3.206
32	43.PHI	training	4.916	72	89.RHW	training	5.878
33	44.PHK	training	3.426	73	91.RWA	training	1.212
34	45.PHL	training	5.311	74	93.RWE	training	1.091
35	46.PHM	training	3.714	75	94.RWF	training	0.909
36	47.PHN	training	6.061	76	95.RWG	training	1.717
37	48.PHQ	training	3.718	77	96.RWH	training	1.091
38	49.PHR	training	4.751	78	97.RWI	training	1.232
39	50.PHS	training	4.042	79	98.RWK	training	0.606
40	52.PHV	training	3.335	80	99.RWL	training	3.212

81	100.RWM	training	0.727	121	166.AYY	training	3.071
82	102.RWQ	training	0.606	122	167.IYY	training	3.071
83	103.RWR	training	2.384	123	168.LYY	training	3.071
84	104.RWS	training	0.808	124	169.FYY	training	1.911
85	107.RWW	training	2.707	125	170.WYY	training	1.911
86	109.DHH	training	0.905	126	172.GYY	training	5.071
87	110.EHH	training	0.905	127	173.NYY	training	5.071
88	113.AHH	training	2.02	128	174.QYY	training	5.071
89	114.IHH	training	2.02	129	175.MYY	training	1.991
90	115.FHH	training	1.803	130	176.SYY	training	3.07
91	116.WHH	training	1.803	131	177.TYY	training	3.07
92	118.GHH	training	1.089	132	178.CYY	training	0.47
93	120.QHH	training	1.089	133	179.YDY	training	3.047
94	121.MHH	training	2.015	134	181.YHY	training	9.886
95	123.THH	training	1.32	135	182.YKY	training	9.886
96	124.CHH	training	0.937	136	183_YRY	training	9.886
97	125.HDH	training	1.477	137	184.YAY	training	3.607
98	126.HEH	training	1.477	138	185.YIY	training	3.607
99	130.HAH	training	0.952	139	186.YLY	training	3.607
100	131.HIH	training	0.952	140	187.YFY	training	2.233
101	132.HLH	training	0.952	141	190.YGY	training	3.366
102	134.HWH	training	2.026	142	191.YNY	training	3.366
103	135.HYH	training	2.026	143	192.YQY	training	3.366
104	136.HGH	training	0.832	144	193.YMY	training	1.78
105	137.HNH	training	0.832	145	194.YSY	training	3.447
106	139.HMH	training	0.873	146	197.YYD	training	4.116
107	140.HSH	training	0.73	147	198.YYE	training	4.116
108	141.HTH	training	0.73	148	199.YYH	training	5.303
109	143.HHD	training	0.188	149	201.YYR	training	5.303
110	151.HHF	training	3.612	150	202.YYA	training	3.344
111	153.HHY	training	3.612	151	204.YYL	training	3.344
112	154.HHG	training	0.317	152	205.YYF	training	4.05
113	155.HHN	training	0.317	153	208.YYG	training	2.996
114	156.HHQ	training	0.317	154	210.YYQ	training	2.996
115	160.HHC	training	0.128	155	211.YYM	training	2.103
116	161.DYY	training	3.417	156	212.YYS	training	3.983
117	162.EYY	training	3.417	157	213.YYT	training	3.983
118	163.HYY	training	2.257	158	214.YYC	training	0.637
119	164.KYY	training	2.257				
120	165.RYY	training	2.257				

159	6.LHH	test	4.836
160	12.LHQ	test	4.136
161	13.LHR	test	5.184
162	15.LHT	test	5.584
163	17.LHW	test	6.791
164	20.LWD	test	1.717
165	22.LWF	test	1.414
166	23.LWG	test	1.313
167	27.LWL	test	0.606
168	33.LWT	test	2.02
169	42.PHH	test	6.051
170	51.PHT	test	6.247
171	57.PWE	test	1.096
172	69.PWT	test	2.598
173	78.RHH	test	3.296
174	81.RHL	test	3.501
175	85.RHR	test	4.302
176	90.RHY	test	3.378
177	92.RWD	test	0.909
178	101.RWN	test	2.404
179	105.RWT	test	3.818
180	106.RWV	test	0.606
181	108.RWY	test	0.808
182	117.YHH	test	1.803
183	119.NHH	test	1.089
184	122.SHH	test	1.32
185	133.HFH	test	2.026
186	138.HQH	test	0.832
187	142.HCH	test	0.975
188	144.HHE	test	0.188
189	152.HHW	test	3.612
190	180.YEY	test	3.047
191	188.YWY	test	2.233
192	195.YTY	test	3.447
193	196.YCY	test	3.087
194	200.YYK	test	5.303
195	203.YYI	test	3.344
196	206.YYW	test	4.05
197	209.YYN	test	2.996

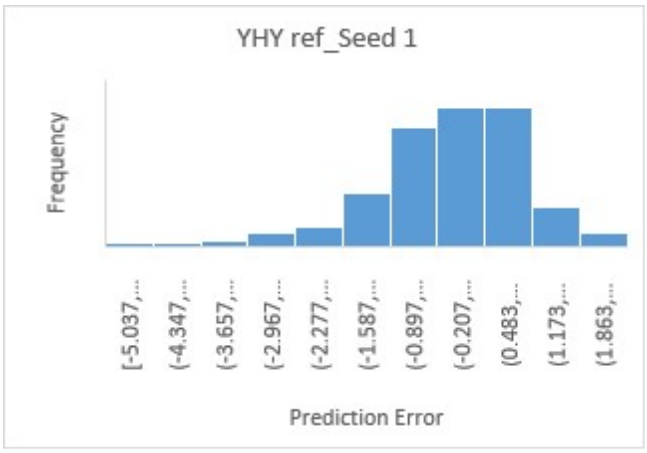
Table S1.5. Train and Test sets_ Seed 5

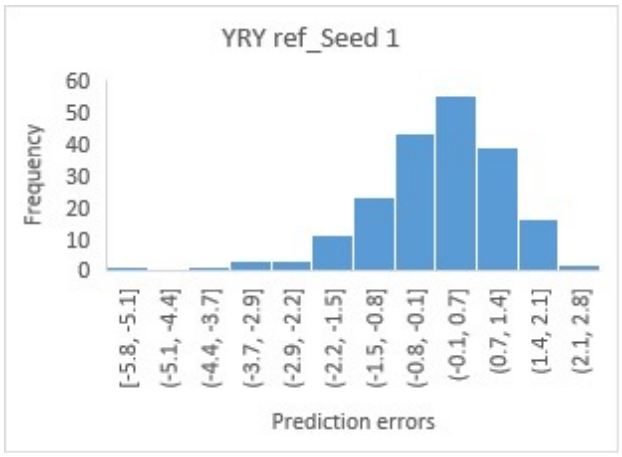
No.	Ligand Name	QSAR Set	Activity	No.	Ligand Name	QSAR Set	Activity
1	1.LHA	training	3.918	41	51.PHT	training	6.247
2	2.LHD	training	3.593	42	52.PHV	training	3.335
3	3.LHE	training	6.136	43	53.PHW	training	6.535
4	4.LHF	training	3.628	44	54.PHY	training	4.227
5	7.LHI	training	6.531	45	55.PWA	training	1.396
6	8.LHK	training	4.225	46	56.PWD	training	1.096
7	9.LHL	training	5.92	47	57.PWE	training	1.096
8	10.LHM	training	4.504	48	58.PWF	training	0.919
9	12.LHQ	training	4.136	49	59.PWG	training	2.687
10	13.LHR	training	5.184	50	60.PWH	training	1.184
11	14.LHS	training	4.293	51	61.PWI	training	1.396
12	16.LHV	training	3.481	52	63.PWL	training	1.096
13	19.LWA	training	1.192	53	64.PWM	training	0.796
14	21.LWE	training	1.717	54	65.PWN	training	2.104
15	22.LWF	training	1.414	55	66.PWQ	training	1.202
16	23.LWG	training	1.313	56	67.PWR	training	2.705
17	24.LWH	training	3.212	57	69.PWT	training	2.598
18	25.LWI	training	1.111	58	70.PWV	training	1.008
19	26.LWK	training	1.899	59	72.PWY	training	1.114
20	27.LWL	training	0.606	60	74.RHD	training	3.304
21	28.LWM	training	1.394	61	75.RHE	training	5.096
22	31.LWR	training	2.909	62	76.RHF	training	3.3
23	32.LWS	training	2.02	63	77.RHG	training	5.725
24	33.LWT	training	2.02	64	78.RHH	training	3.296
25	34.LWV	training	1.616	65	79.RHI	training	4.806
26	35.LWW	training	3.515	66	80.RHK	training	2.694
27	36.LWY	training	2.222	67	81.RHL	training	3.501
28	37.PHA	training	5.793	68	83.RHN	training	5.713
29	39.PHE	training	6.152	69	85.RHR	training	4.302
30	40.PHF	training	3.916	70	86.RHS	training	3.386
31	41.PHG	training	5.197	71	88.RHV	training	3.206
32	42.PHH	training	6.051	72	89.RHW	training	5.878
33	43.PHI	training	4.916	73	91.RWA	training	1.212
34	44.PHK	training	3.426	74	92.RWD	training	0.909
35	45.PHL	training	5.311	75	93.RWE	training	1.091
36	46.PHM	training	3.714	76	94.RWF	training	0.909
37	47.PHN	training	6.061	77	95.RWG	training	1.717
38	48.PHQ	training	3.718	78	97.RWI	training	1.232
39	49.PHR	training	4.751	79	98.RWK	training	0.606
40	50.PHS	training	4.042	80	99.RWL	training	3.212

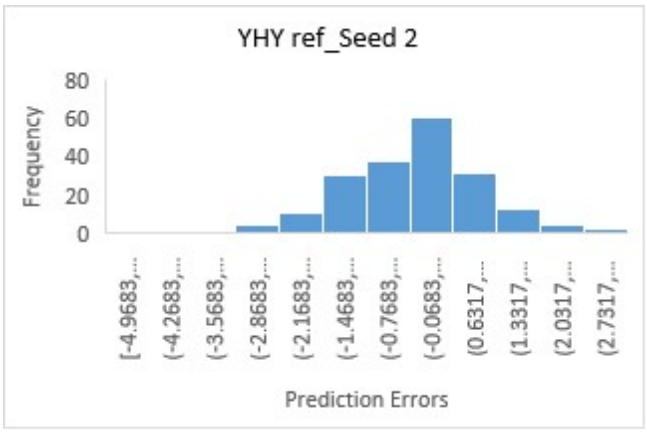
No.	Ligand Name	QSAR Set	Activity	No.	Ligand Name	QSAR Set	Activity
81	100.RWM	training	0.727	121	163.HYY	training	2.257
82	101.RWN	training	2.404	122	165.RYY	training	2.257
83	102.RWQ	training	0.606	123	166.AYY	training	3.071
84	103.RWR	training	2.384	124	167.IYY	training	3.071
85	104.RWS	training	0.808	125	168.LYY	training	3.071
86	105.RWT	training	3.818	126	169.FYY	training	1.911
87	106.RWV	training	0.606	127	170.WYY	training	1.911
88	107.RWW	training	2.707	128	172.GYY	training	5.071
89	108.RWY	training	0.808	129	173.NYY	training	5.071
90	110.EHH	training	0.905	130	174.QYY	training	5.071
91	114.IHH	training	2.02	131	176.SYY	training	3.07
92	115.FHH	training	1.803	132	178.CYY	training	0.47
93	116.WHH	training	1.803	133	180.YEY	training	3.047
94	117.YHH	training	1.803	134	181.YHY	training	9.886
95	119.NHH	training	1.089	135	182.YKY	training	9.886
96	120.QHH	training	1.089	136	183_YRY	training	9.886
97	121.MHH	training	2.015	137	184.YAY	training	3.607
98	122.SHH	training	1.32	138	186.YLY	training	3.607
99	124.CHH	training	0.937	139	187.YFY	training	2.233
100	125.HDH	training	1.477	140	188.YWY	training	2.233
101	126.HEH	training	1.477	141	190.YGY	training	3.366
102	131.HIH	training	0.952	142	192.YQY	training	3.366
103	132.HLH	training	0.952	143	195.YTY	training	3.447
104	134.HWH	training	2.026	144	196.YCY	training	3.087
105	135.HYH	training	2.026	145	198.YYE	training	4.116
106	136.HGH	training	0.832	146	199.YYH	training	5.303
107	137.HNH	training	0.832	147	200.YYK	training	5.303
108	139.HMH	training	0.873	148	201.YYR	training	5.303
109	140.HSH	training	0.73	149	202.YYA	training	3.344
110	141.HTH	training	0.73	150	203.YYI	training	3.344
111	142.HCH	training	0.975	151	204.YYL	training	3.344
112	143.HHD	training	0.188	152	205.YYF	training	4.05
113	144.HHE	training	0.188	153	206.YYW	training	4.05
114	152.HHW	training	3.612	154	208.YYG	training	2.996
115	153.HHY	training	3.612	155	209.YYN	training	2.996
116	154.HHG	training	0.317	156	210.YYQ	training	2.996
117	156.HHQ	training	0.317	157	211.YYM	training	2.103
118	160.HHC	training	0.128	158	213.YYT	training	3.983
119	161.DYY	training	3.417				
120	162.EYY	training	3.417				

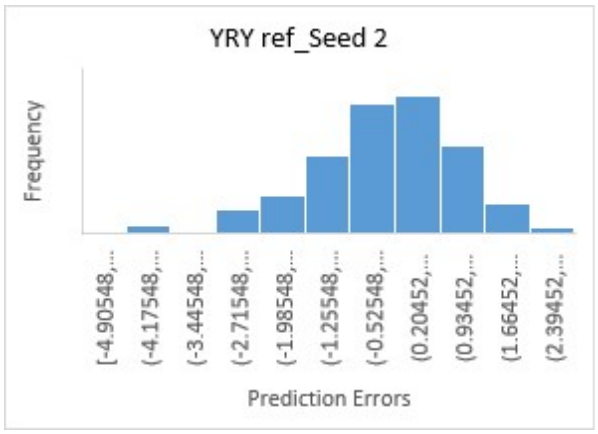
No.	Ligand Name	QSAR Set	Activity
159	5.LHG	test	6.697
160	6.LHH	test	4.836
161	11.LHN	test	5.148
162	15.LHT	test	5.584
163	17.LHW	test	6.791
164	18.LHY	test	4.203
165	20.LWD	test	1.717
166	29.LWN	test	1.313
167	30.LWQ	test	2.505
168	38.PHD	test	4.622
169	62.PWK	test	0.407
170	68.PWS	test	1.096
171	71.PWW	test	2.899
172	73.RHA	test	5.205
173	82.RHM	test	3.218
174	84.RHQ	test	3.108
175	87.RHT	test	5.987
176	90.RHY	test	3.378
177	96.RWH	test	1.091
178	109.DHH	test	0.905
179	113.AHH	test	2.02
180	118.GHH	test	1.089
181	123.THH	test	1.32
182	130.HAH	test	0.952
183	133.HFH	test	2.026
184	138.HQH	test	0.832
185	151.HHF	test	3.612
186	155.HHN	test	0.317
187	164.KYY	test	2.257
188	175.MYY	test	1.991
189	177.TYY	test	3.07
190	179.YDY	test	3.047
191	185.YIY	test	3.607
192	191.YNY	test	3.366
193	193.YMY	test	1.78
194	194.YSY	test	3.447
195	197.YYD	test	4.116
196	212.YYS	test	3.983
197	214.YYC	test	0.637

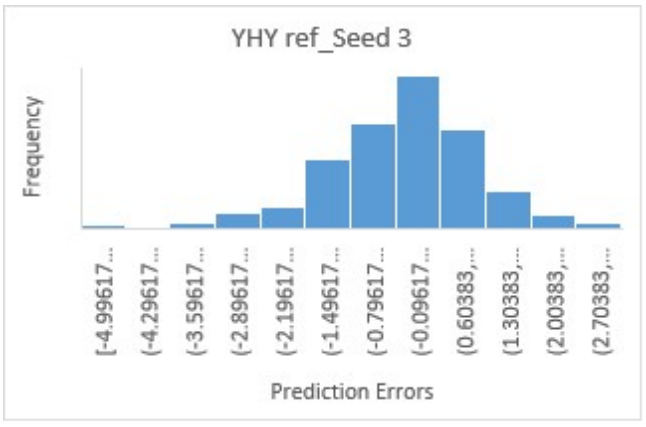
Figure S1.1. Distribution of Prediction Errors for PLS number of 3

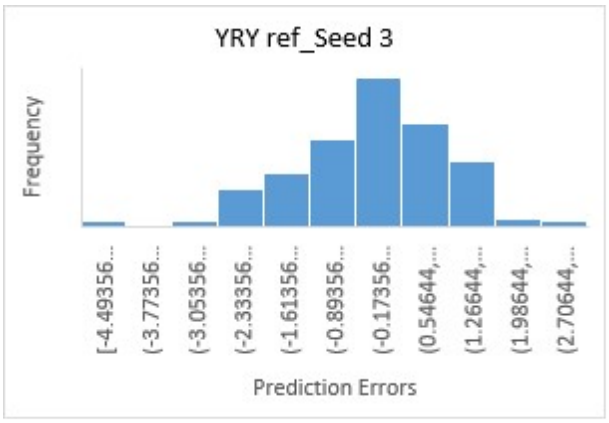


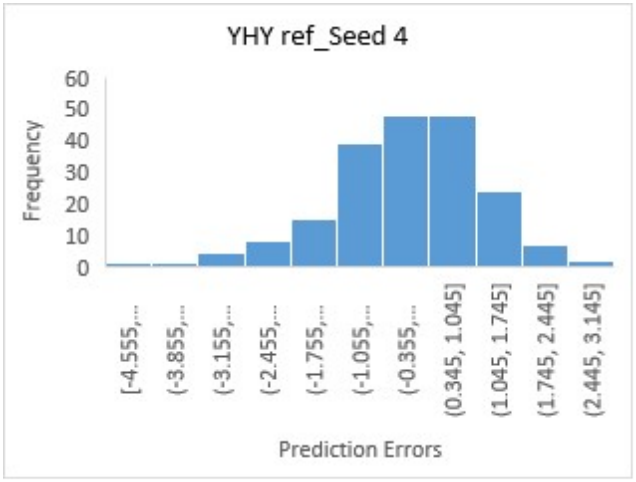


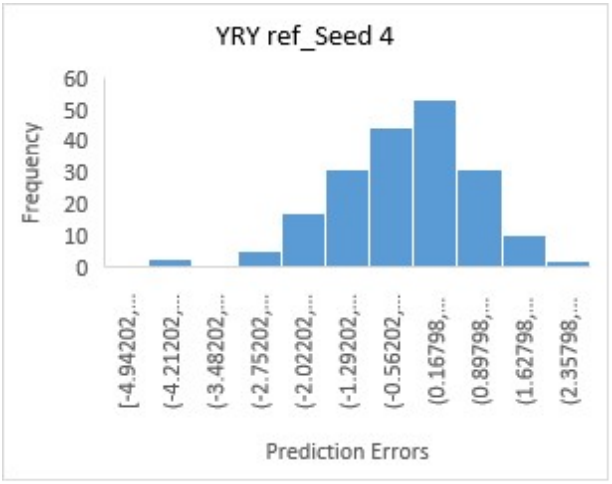


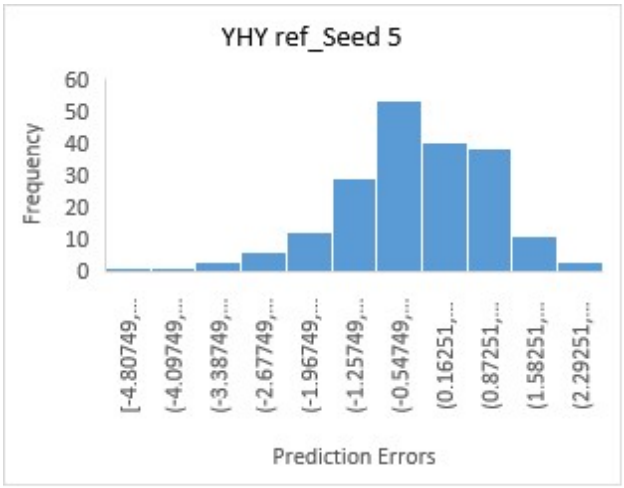


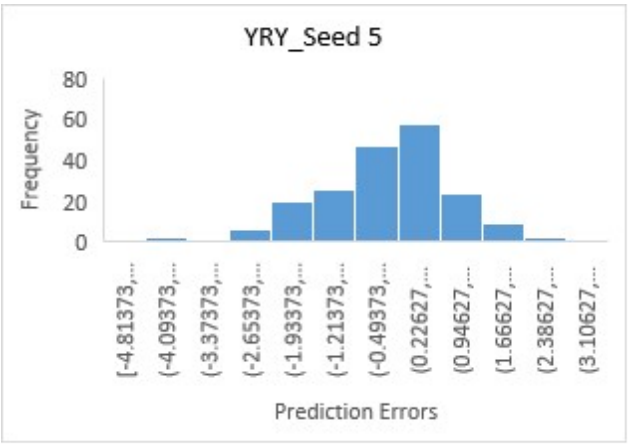





YHY ref_Seed 4
Frequency
60
50
40
30
20
10
0
[-4.555,...
(-3.855,...
(-3.155,...
(-2.455,...
(-1.755,...
(-1.055,...
(-0.355,...
(0.345, 1.045]
(1.045, 1.745]
(1.745, 2.445]
(2.445, 3.145]
Prediction Errors


YRY ref_Seed 4
Frequency
60
50
40
30
20
10
0
[-4.94202,...
(-4.21202,...
(-3.48202,...
(-2.75202,...
(-2.02202,...
(-1.29202,...
(-0.56202,...
(0.16798,...
(0.89798,...
(1.62798,...
(2.35798,...
Prediction Errors


YHY ref_Seed 5
Frequency
60
50
40
30
20
10
0
[-4.80749,...
(-4.09749,...
(-3.38749,...
(-2.67749,...
(-1.96749,...
(-1.25749,...
(-0.54749,...
(0.16251,...
(0.87251,...
(1.58251,...
(2.29251,...
Prediction Errors


YRY_Seed 5
Frequency
80
60
40
20
0
[-4.81373,...
(-4.09373,...
(-3.37373,...
(-2.65373,...
(-1.93373,...
(-1.21373,...
(-0.49373,...
(0.22627,...
(0.94627,...
(1.66627,...
(2.38627,...
(3.10627,...
Prediction Errors

Supplementary 2. Statistical parameters used in OPLS, Tripos and ML based CoMSIA models

1. Coefficient of multiple determination for the training set (R^2)

$$R^2 = \frac{MSS}{TSS} = 1 - \frac{RSS}{TSS}$$

$$\text{MSS} = \sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2$$

$$\text{TSS} = \sum_{i=1}^{n} (y_i - \bar{y})^2$$

$$\text{RSS} = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$$

y_i: observed response for the i^{th} compound

$\bar{y}$: the mean of responses in the training set

$\hat{y}_i$: response of i^{th} compound calculated using the regression model

n: total number of compounds in the training set

2. Root mean square error in fitting/ root mean square error for training set (RMSE)

$$RMSE = \sqrt{\frac{RSS}{n}} = \sqrt{\frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{n}}$$

3. The standard error of prediction (SEP)

$$SEP = \sqrt{\frac{RSS}{n-m}} = \sqrt{\frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{n-m}}$$

n: total number of compounds in the training set

m: the number of PLS factors

4. F-value (F)

$$F = \frac{MSS}{RSS} \times \frac{n-m-1}{m}$$

n: total number of compounds in the training set

m: the number of PLS factors

The F-statistic represents the ratio of model variance distributed across m degrees of freedom to the activity variance distributed across n–m–1 degrees of freedom. Higher F-values indicate a stronger statistical significance in the regression analysis.

5. Cross-validated R^2_{CV} value, computed from predictions obtained by a leave-M-out approach (if M = 1, $R^2_{CV} = R^2_{LOO}$)

$$R^2_{CV} = 1 - \frac{PRESS_{CV}}{TSS} = \frac{\sum_{i=1}^{n} (y_i - \hat{y}_{i/i})^2}{TSS}$$

$PRESS_{CV}$ cross-validated residual sum of squares

y_i: observed response for the i[th] compound

$\hat{y}_{i/i}$: response of the i[th] object estimated by using a model obtained without using M compounds (the i[th] object is one of the M compounds)

6. Variance explained in external prediction using the mean of the training set (Q^2 also known as Q^2_{F1})

$$Q^2 = 1 - \frac{PRESS_{test}}{TSS_{(\bar{y}_{train})}}$$

$$TSS_{(\bar{y}_{train})} = \sum_{i=1}^{t} (y_i - \bar{y}_{train})^2$$

$$PRESS_{test} = \sum_{i=1}^{t} (y_i - \hat{y}_i)^2$$

$\bar{y}_{train}$: the mean of the training-set responses

y_i: observed response for the i[th] compound in the test set

$\hat{y}_i$: predicted response for the i[th] compound in the test set

t: total number of compounds in the test set

7. Root mean square error in external prediction (RMSE_test)

$$RMSE_{test} = \sqrt{\frac{\sum_{i=1}^{t} (y_i - \hat{y}_i)^2}{t}}$$

8. Variance explained in external prediction using the mean of the test set (R^2_test also known as Q^2_{F2})

$$R^2_{test} = 1 - \frac{PRESS_{test}}{TSS_{(\bar{y}_{test})}}$$

$$TSS_{(\bar{y}_{test})} = \sum_{i=1}^{t} (y_i - \bar{y}_{test})^2$$

$$\text{PRESS}_{\text{test}} = \sum_{i=1}^{t} (y_i - \hat{y}_i)^2$$

$\bar{y}_{test}$: the mean of the test-set responses

y_i: observed response for the i[th] compound in the test set

$\hat{y}_i$: predicted response for the i[th] compound in the test set

t: total number of compounds in the test set

9. R^2_{Scramble}: Average value of R^2 from a series of models built using scrambled activities, measures the degree to which the molecular fields can fit random data. A low value means that the model cannot fit random data, but a high value merely means that the variable set is fairly complete and can fit anything.

Supplementary 3

Table S3.1. Five CoMSIA interaction fields from FTC dataset with YRY reference

The Table S3.1 can be found at:
https://docs.google.com/spreadsheets/d/1UYyIjOepMVyfRX5VJnK75_VJ3KzvX2Vo/edit?usp=sharing&ouid=106421106263255583273&rtpof=true&sd=true

Supplementary 4

Table S4.1. The statistics of OPLS models with **reference YHY** and random seed 1

# Factors	SD	R^2	R^2_{cv}	$R^2\_{Scramble}$	Stability	F	RMSE_test	Q^2	Pearson-r
1	1.542	0.314	0.272	0.056	0.997	71.500	1.630	0.335	0.588
2	1.184	0.598	0.528	0.086	0.992	115.300	1.230	0.619	0.800
3	1.136	0.633	0.555	0.123	0.992	88.400	1.220	0.628	0.804
4	1.091	0.663	0.542	0.166	0.978	75.400	1.220	0.626	0.796
5	1.059	0.685	0.526	0.188	0.959	66.100	1.200	0.638	0.803
6	1.035	0.701	0.523	0.214	0.942	58.900	1.170	0.657	0.814
7	1.014	0.715	0.518	0.235	0.924	53.700	1.180	0.649	0.809
8	1.000	0.724	0.517	0.252	0.911	49.000	1.130	0.680	0.827
9	0.975	0.740	0.530	0.271	0.897	46.800	1.090	0.702	0.840
10	0.965	0.747	0.529	0.287	0.887	43.400	1.100	0.698	0.838
11	0.948	0.757	0.520	0.303	0.873	41.400	1.130	0.682	0.832
12	0.944	0.761	0.513	0.316	0.869	38.500	1.140	0.672	0.827
13	0.937	0.766	0.505	0.323	0.863	36.300	1.170	0.654	0.818
14	0.928	0.772	0.495	0.333	0.861	34.600	1.200	0.637	0.811
15	0.918	0.779	0.495	0.344	0.859	33.300	1.220	0.625	0.809
16	0.901	0.789	0.486	0.353	0.854	32.900	1.220	0.625	0.816
17	0.878	0.800	0.443	0.365	0.810	33.000	1.260	0.604	0.817
18	0.857	0.811	0.440	0.373	0.782	33.100	1.300	0.578	0.804
19	0.846	0.817	0.438	0.382	0.761	32.500	1.310	0.573	0.800
20	0.835	0.824	0.433	0.389	0.743	31.900	1.300	0.574	0.801

Table S4.2. The statistics of OPLS models with reference YHY and random seed 2

# Factors	SD	R^2	R^2_{cv}	$R^2\_{Scramble}$	Stability	F	RMSE_test	Q^2	Pearson-r
1	1.604	0.309	0.274	0.041	0.998	69.700	1.630	0.105	0.379
2	1.248	0.584	0.516	0.085	0.993	108.800	1.400	0.334	0.617
3	1.139	0.656	0.571	0.128	0.989	97.800	1.270	0.449	0.695
4	1.054	0.707	0.592	0.172	0.973	92.500	1.300	0.429	0.688
5	1.007	0.735	0.600	0.211	0.960	84.100	1.240	0.481	0.716

6	0.970	0.755	0.614	0.242	0.949	77.600	1.200	0.513	0.731
7	0.956	0.764	0.619	0.263	0.949	69.300	1.150	0.551	0.752
8	0.951	0.768	0.621	0.284	0.949	61.700	1.160	0.541	0.747
9	0.936	0.777	0.604	0.297	0.940	57.300	1.130	0.567	0.763
10	0.929	0.782	0.595	0.309	0.931	52.600	1.110	0.584	0.771
11	0.914	0.790	0.558	0.326	0.903	49.900	1.160	0.545	0.753
12	0.903	0.796	0.538	0.339	0.875	47.300	1.150	0.549	0.761
13	0.896	0.801	0.528	0.351	0.861	44.500	1.170	0.537	0.755
14	0.889	0.806	0.519	0.361	0.850	42.300	1.150	0.550	0.762
15	0.883	0.810	0.514	0.371	0.849	40.200	1.150	0.555	0.767
16	0.876	0.814	0.508	0.380	0.846	38.500	1.120	0.576	0.778
17	0.874	0.816	0.509	0.390	0.847	36.500	1.110	0.579	0.777
18	0.870	0.819	0.499	0.398	0.844	34.900	1.130	0.564	0.767
19	0.859	0.825	0.471	0.407	0.824	34.200	1.150	0.553	0.761
20	0.851	0.829	0.455	0.416	0.806	33.200	1.160	0.542	0.761

Table S4.3. The statistics of OPLS models with reference YHY and random seed 3.

# Factors	SD	R^2	R^2_{cv}	$R^2\_{Scramble}$	Stability	F	RMSE_test	Q^2	Pearson-r
1	1.591	0.322	0.280	0.045	0.997	74.100	1.510	0.215	0.494
2	1.228	0.599	0.527	0.083	0.992	115.700	1.350	0.372	0.667
3	1.157	0.646	0.565	0.122	0.990	93.700	1.230	0.482	0.735
4	1.090	0.688	0.580	0.159	0.978	84.400	1.160	0.538	0.766
5	1.053	0.711	0.587	0.187	0.967	74.600	1.180	0.519	0.769
6	1.024	0.728	0.597	0.212	0.959	67.400	1.130	0.562	0.781
7	1.010	0.738	0.604	0.229	0.956	60.200	1.120	0.574	0.791
8	1.002	0.743	0.604	0.247	0.956	53.900	1.150	0.547	0.778
9	0.993	0.750	0.603	0.263	0.956	49.200	1.120	0.567	0.786
10	0.987	0.754	0.595	0.277	0.951	45.100	1.120	0.568	0.787
11	0.973	0.763	0.563	0.291	0.931	42.600	1.130	0.565	0.795
12	0.970	0.766	0.553	0.302	0.923	39.500	1.140	0.554	0.786
13	0.962	0.771	0.536	0.313	0.910	37.400	1.180	0.527	0.778
14	0.947	0.780	0.513	0.330	0.886	36.200	1.200	0.508	0.770
15	0.931	0.789	0.489	0.340	0.859	35.300	1.250	0.462	0.749
16	0.917	0.796	0.484	0.351	0.839	34.400	1.260	0.456	0.764
17	0.902	0.804	0.478	0.364	0.817	33.800	1.240	0.470	0.771
18	0.889	0.811	0.499	0.376	0.826	33.200	1.230	0.478	0.769
19	0.878	0.817	0.495	0.389	0.823	32.500	1.290	0.432	0.763
20	0.873	0.821	0.505	0.400	0.831	31.400	1.350	0.377	0.748

Table S4.4. The statistics of OPLS models with reference YHY and random seed 4.

# Factors	SD	R^2	R^2_{cv}	$R^2\_{Scramble}$	Stability	F	RMSE_test	Q^2	Pearson-r

1	1.448	0.434	0.376	0.061	0.995	119.500	1.450	0.308	0.582
2	1.202	0.612	0.534	0.108	0.991	122.400	1.280	0.460	0.695
3	1.169	0.636	0.545	0.154	0.988	89.700	1.190	0.533	0.740
4	1.100	0.680	0.549	0.187	0.970	81.200	1.180	0.543	0.754
5	1.034	0.719	0.545	0.220	0.941	77.600	1.230	0.504	0.751
6	0.983	0.747	0.563	0.249	0.923	74.500	1.200	0.525	0.762
7	0.968	0.757	0.574	0.270	0.920	66.700	1.180	0.545	0.768
8	0.957	0.764	0.586	0.287	0.925	60.300	1.170	0.552	0.766
9	0.931	0.778	0.600	0.303	0.925	57.600	1.210	0.519	0.754
10	0.918	0.786	0.576	0.320	0.908	53.900	1.240	0.497	0.746
11	0.908	0.792	0.554	0.331	0.896	50.400	1.260	0.478	0.745
12	0.901	0.796	0.548	0.345	0.890	47.200	1.290	0.453	0.733
13	0.895	0.801	0.548	0.358	0.886	44.400	1.330	0.417	0.715
14	0.888	0.805	0.535	0.369	0.873	42.200	1.380	0.374	0.685
15	0.878	0.811	0.522	0.381	0.862	40.500	1.400	0.358	0.685
16	0.870	0.815	0.523	0.390	0.851	38.900	1.380	0.373	0.696
17	0.863	0.820	0.502	0.399	0.836	37.400	1.370	0.382	0.707
18	0.852	0.826	0.482	0.409	0.818	36.600	1.330	0.422	0.726
19	0.837	0.833	0.471	0.420	0.804	36.100	1.290	0.450	0.748
20	0.828	0.837	0.442	0.434	0.785	35.300	1.310	0.439	0.746

Table S4.5. The statistics of OPLS models with reference YHY and random seed 5.

# Factors	SD	R^2	R^2_{cv}	$R^2\_{Scramble}$	Stability	F	RMSE_test	Q^2	Pearson-r
1	1.529	0.367	0.321	0.059	0.997	90.500	1.610	0.157	0.445
2	1.251	0.579	0.510	0.096	0.993	106.800	1.510	0.255	0.545
3	1.154	0.644	0.558	0.136	0.990	93.000	1.310	0.442	0.676
4	1.076	0.693	0.574	0.168	0.976	86.200	1.230	0.506	0.728
5	1.051	0.709	0.588	0.195	0.975	74.000	1.220	0.514	0.734
6	0.994	0.741	0.605	0.223	0.955	72.100	1.230	0.512	0.739
7	0.965	0.758	0.617	0.248	0.949	66.900	1.210	0.526	0.740
8	0.957	0.763	0.623	0.266	0.950	60.100	1.230	0.508	0.730
9	0.944	0.771	0.630	0.283	0.954	55.500	1.240	0.504	0.727
10	0.928	0.781	0.625	0.301	0.952	52.300	1.250	0.493	0.715
11	0.912	0.789	0.608	0.320	0.944	49.700	1.260	0.481	0.715
12	0.908	0.793	0.595	0.332	0.937	46.200	1.280	0.471	0.712
13	0.904	0.796	0.582	0.349	0.928	43.100	1.290	0.459	0.707
14	0.894	0.802	0.559	0.364	0.910	41.300	1.340	0.418	0.686
15	0.889	0.805	0.556	0.375	0.905	39.100	1.340	0.418	0.684
16	0.883	0.809	0.550	0.389	0.901	37.400	1.300	0.450	0.700
17	0.870	0.816	0.534	0.399	0.881	36.500	1.320	0.436	0.698
18	0.848	0.827	0.490	0.411	0.830	36.800	1.280	0.470	0.713
19	0.839	0.831	0.496	0.424	0.832	35.800	1.270	0.474	0.716

20	0.831	0.836	0.483	0.433	0.825	34.900	1.280	0.469	0.715

Table S4.6. The statistics of OPLS models with **reference YRY** and random seed 1

# Factors	SD	R^2	R^2cv	R^2_Scramble	Stability	F	RMSE_test	Q^2	Pearson-r
1	1.572	0.287	0.243	0.044	0.997	62.800	1.710	0.270	0.531
2	1.204	0.585	0.498	0.083	0.987	109.100	1.470	0.459	0.687
3	1.122	0.641	0.543	0.126	0.985	91.800	1.470	0.458	0.689
4	1.089	0.665	0.545	0.160	0.972	75.800	1.490	0.441	0.675
5	0.995	0.722	0.545	0.184	0.921	78.800	1.610	0.348	0.612
6	0.962	0.741	0.552	0.212	0.895	72.100	1.650	0.316	0.591
7	0.934	0.758	0.568	0.230	0.890	67.100	1.690	0.281	0.563
8	0.906	0.774	0.589	0.248	0.872	63.800	1.680	0.292	0.578
9	0.897	0.780	0.582	0.263	0.874	58.200	1.690	0.282	0.573
10	0.882	0.788	0.574	0.275	0.858	54.800	1.710	0.266	0.561
11	0.878	0.792	0.573	0.287	0.855	50.400	1.710	0.264	0.559
12	0.862	0.801	0.568	0.301	0.857	48.600	1.700	0.277	0.563
13	0.847	0.809	0.568	0.315	0.868	46.900	1.710	0.265	0.563
14	0.844	0.812	0.572	0.326	0.866	44.000	1.710	0.265	0.567
15	0.840	0.815	0.567	0.334	0.863	41.700	1.720	0.261	0.568
16	0.835	0.818	0.570	0.342	0.859	39.700	1.700	0.277	0.582
17	0.831	0.821	0.561	0.349	0.841	37.800	1.670	0.302	0.595
18	0.829	0.823	0.556	0.355	0.839	36.000	1.670	0.299	0.593
19	0.829	0.824	0.554	0.362	0.834	34.100	1.660	0.312	0.599
20	0.828	0.826	0.530	0.369	0.819	32.600	1.660	0.313	0.601

Table S4.7. The statistics of OPLS models with **reference YRY** and random seed 2

# Factors	SD	R^2	R^2cv	R^2_Scramble	Stability	F	RMSE_test	Q^2	Pearson-r
1	1.609	0.305	0.266	0.045	0.998	68.400	1.620	0.115	0.389
2	1.244	0.587	0.498	0.082	0.989	110.000	1.290	0.433	0.685
3	1.216	0.608	0.508	0.110	0.987	79.700	1.230	0.491	0.719
4	1.147	0.654	0.506	0.152	0.967	72.200	1.270	0.457	0.697
5	1.104	0.681	0.492	0.181	0.937	65.000	1.290	0.434	0.675
6	1.064	0.706	0.497	0.213	0.913	60.400	1.240	0.481	0.701
7	1.044	0.719	0.501	0.233	0.899	54.700	1.180	0.528	0.731
8	1.014	0.736	0.474	0.253	0.867	52.000	1.150	0.549	0.749
9	0.993	0.749	0.458	0.267	0.839	49.000	1.170	0.535	0.741
10	0.967	0.763	0.419	0.277	0.776	47.400	1.100	0.589	0.772
11	0.946	0.775	0.353	0.287	0.702	45.700	1.110	0.584	0.769
12	0.930	0.784	0.336	0.300	0.656	43.900	1.120	0.573	0.761
13	0.913	0.794	0.336	0.311	0.627	42.600	1.100	0.592	0.773
14	0.893	0.804	0.334	0.320	0.626	41.800	1.100	0.591	0.776
15	0.889	0.807	0.320	0.329	0.611	39.600	1.090	0.594	0.777

16	0.886	0.810	0.326	0.339	0.592	37.500	1.070	0.611	0.786
17	0.883	0.812	0.342	0.346	0.588	35.600	1.080	0.606	0.783
18	0.881	0.815	0.340	0.354	0.588	33.900	1.080	0.607	0.784
19	0.879	0.816	0.347	0.366	0.585	32.300	1.080	0.608	0.784
20	0.877	0.819	0.370	0.379	0.609	31.000	1.090	0.595	0.776

Table S4.8. The statistics of OPLS models with **reference YRY** and random seed 3

# Factors	SD	R^2	R^2cv	R^2_Scramble	Stability	F	RMSE_test	Q^2	Pearson-r
1	1.624	0.293	0.250	0.050	0.997	64.700	1.510	0.217	0.487
2	1.251	0.584	0.497	0.090	0.989	108.600	1.360	0.362	0.651
3	1.185	0.628	0.540	0.127	0.988	86.800	1.310	0.407	0.699
4	1.136	0.661	0.544	0.175	0.977	74.500	1.240	0.471	0.734
5	1.069	0.702	0.534	0.202	0.943	71.500	1.310	0.415	0.708
6	1.027	0.727	0.531	0.229	0.915	66.900	1.270	0.445	0.715
7	1.003	0.741	0.543	0.248	0.906	61.300	1.220	0.490	0.740
8	0.976	0.756	0.540	0.266	0.882	57.800	1.270	0.448	0.731
9	0.947	0.772	0.517	0.281	0.848	55.700	1.300	0.423	0.707
10	0.915	0.789	0.486	0.297	0.794	54.900	1.270	0.446	0.725
11	0.911	0.792	0.466	0.308	0.776	50.500	1.290	0.434	0.719
12	0.887	0.804	0.437	0.319	0.736	49.600	1.290	0.432	0.728
13	0.869	0.813	0.429	0.329	0.718	48.200	1.330	0.397	0.716
14	0.862	0.818	0.441	0.337	0.716	45.800	1.310	0.409	0.721
15	0.851	0.823	0.450	0.346	0.689	44.100	1.280	0.438	0.730
16	0.846	0.827	0.443	0.357	0.684	42.100	1.260	0.459	0.745
17	0.844	0.829	0.448	0.363	0.676	39.900	1.230	0.481	0.749
18	0.839	0.832	0.465	0.374	0.702	38.200	1.230	0.482	0.748
19	0.835	0.835	0.465	0.384	0.710	36.700	1.210	0.499	0.756
20	0.830	0.838	0.485	0.395	0.731	35.400	1.190	0.511	0.761

Table S4.9. The statistics of OPLS models with **reference YRY** and random seed 4

# Factors	SD	R^2	R^2cv	R^2_Scramble	Stability	F	RMSE_test	Q^2	Pearson-r
1	1.568	0.336	0.284	0.053	0.996	79.100	1.440	0.318	0.568
2	1.273	0.565	0.475	0.096	0.990	100.800	1.170	0.547	0.742
3	1.236	0.593	0.485	0.135	0.987	74.800	1.130	0.579	0.764
4	1.151	0.649	0.489	0.182	0.964	70.800	1.180	0.540	0.739
5	1.100	0.682	0.491	0.211	0.943	65.200	1.180	0.542	0.741
6	1.054	0.710	0.505	0.236	0.922	61.600	1.260	0.480	0.705
7	1.030	0.725	0.511	0.256	0.910	56.400	1.290	0.456	0.695
8	1.006	0.739	0.508	0.275	0.899	52.700	1.240	0.492	0.717
9	0.987	0.751	0.512	0.289	0.887	49.500	1.200	0.524	0.738
10	0.961	0.765	0.494	0.301	0.855	47.900	1.250	0.488	0.723

11	0.939	0.777	0.468	0.315	0.814	46.300	1.300	0.447	0.706
12	0.917	0.789	0.443	0.326	0.772	45.200	1.330	0.418	0.692
13	0.902	0.797	0.424	0.339	0.743	43.600	1.320	0.429	0.690
14	0.882	0.808	0.418	0.351	0.722	42.900	1.330	0.417	0.690
15	0.871	0.814	0.410	0.362	0.700	41.300	1.340	0.407	0.684
16	0.866	0.817	0.398	0.371	0.676	39.400	1.320	0.425	0.700
17	0.859	0.821	0.397	0.380	0.661	37.800	1.310	0.438	0.703
18	0.858	0.823	0.396	0.388	0.641	35.800	1.270	0.467	0.713
19	0.857	0.825	0.395	0.397	0.643	34.100	1.230	0.504	0.732
20	0.857	0.826	0.400	0.405	0.650	32.500	1.200	0.525	0.742

Table S4.10. The statistics of OPLS models with **reference YRY** and random seed 5

# Factors	SD	R^2	R^2cv	R^2_Scramble	Stability	F	RMSE_test	Q^2	Pearson-r
1	1.516	0.378	0.326	0.053	0.996	94.700	1.580	0.192	0.478
2	1.279	0.560	0.491	0.095	0.993	98.800	1.480	0.293	0.569
3	1.188	0.623	0.537	0.124	0.990	84.700	1.270	0.473	0.696
4	1.083	0.689	0.509	0.159	0.913	84.600	1.310	0.445	0.672
5	1.053	0.708	0.541	0.185	0.943	73.700	1.350	0.412	0.652
6	1.018	0.729	0.551	0.204	0.931	67.500	1.320	0.438	0.672
7	0.998	0.741	0.568	0.224	0.928	61.200	1.310	0.440	0.674
8	0.984	0.750	0.571	0.238	0.924	55.800	1.340	0.420	0.659
9	0.954	0.766	0.547	0.252	0.901	54.000	1.330	0.424	0.659
10	0.937	0.776	0.542	0.264	0.885	51.000	1.360	0.398	0.647
11	0.916	0.788	0.513	0.277	0.854	49.200	1.370	0.393	0.649
12	0.886	0.803	0.462	0.292	0.783	49.200	1.390	0.374	0.641
13	0.870	0.811	0.464	0.306	0.757	47.500	1.360	0.403	0.661
14	0.856	0.818	0.471	0.318	0.738	46.000	1.360	0.398	0.657
15	0.850	0.822	0.457	0.330	0.724	43.800	1.350	0.406	0.659
16	0.846	0.825	0.473	0.343	0.724	41.500	1.310	0.440	0.680
17	0.844	0.827	0.478	0.354	0.724	39.400	1.300	0.447	0.682
18	0.839	0.830	0.490	0.365	0.727	37.800	1.290	0.457	0.688
19	0.833	0.834	0.498	0.376	0.737	36.500	1.280	0.468	0.695
20	0.831	0.836	0.503	0.386	0.731	34.900	1.280	0.471	0.694

Supplementary 5. Statistics of Tripos-based models with different random seeds.

YHY REFERENCE_SEED 1

Determine the optimal number of factors

Parameters: CoMFA standard scaling, Leave-one-out, SAMPLS, 20 components

```
Option Name       Description                              Value
----------- ----------- -----
BOOTSTRAPPING     Number of bootstrap runs to make         0
CENTERING         Force intercept thru 0.0?                NO
COMPONENTS        Number of components to use              20
CROSSVALIDATION   Number of crossvalidation groups         0
EPSILON           Convergence criterion                    1.e-4
ITERATION         Maximum number of iterations             100
SCALING_METHOD    Pre-analysis scaling          (COMFA_STD for CoMF COMFA_STD
XLIST Print info about xval groups?                  NO
SAMPLS has finished running. This is NOT a full PLS analysis.
Standard Error of Prediction for 20 components:
1.426 1.403 1.417 1.434 1.433 1.431 1.441 1.444 1.443 1.431 1.432 1.441 1.450
1.484 1.497 1.512 1.530 1.536 1.565 1.589
Crossvalidated R2 for 20 components:
0.413 0.436 0.428 0.418 0.422 0.428 0.424 0.425 0.430 0.443 0.446 0.443 0.440
0.417 0.411 0.404 0.394 0.393 0.374 0.360
-- optimum is 0.446 at 11 components
##################################################
```

Determine the coefficient of determination (without cross-validation) R^2

Parameters: CoMFA standard scaling, No cross validation, Filtering: 0 kcal, 11 components

```
Standard Error of Estimate               0.950
R squared                                0.756
F values          (n1=11, n2=146)        41.227
Prob.of R2=0      (n1=11, n2=146)        0.000
Relative Contributions
#                            Norm.Coeff. Fraction
- ----------- ---------
1 COMSIA_DONOR (1100 vars)          2.180 0.275
2 COMSIA_ACCEPTOR (1100 vars)       0.831 0.105
3 COMSIA_HYDROPHOBIC (1100 vars)    2.654 0.334
4 COMSIA_STERIC (1100 vars)         0.519 0.065
5 COMSIA_ELECTROSTATIC (1100 vars)  1.754 0.221
Summary output
Standard Error of Estimate          0.950
R squared                           0.756
F values    (n1=11, n2=146)         41.227
Prob.of R2 = 0 (n1=11, n2=146)      0.000
```

Determine the coefficient of determination due to bootstrapping activity R^2_{bstr}

Parameters: CoMFA standard scaling, Bootstrap 100 runs, Filtering: 0 kcal, 11 components

```
Means
#                            Norm.Coeff. Fraction
- ----------- ---------
1 COMSIA_DONOR (1100 vars)          1.911 0.264
2 COMSIA_ACCEPTOR (1100 vars)       0.779 0.107
3 COMSIA_HYDROPHOBIC (1100 vars)    2.397 0.331
4 COMSIA_STERIC (1100 vars)         0.534 0.074
5 COMSIA_ELECTROSTATIC (1100 vars)  1.627 0.225
```

```
Standard Deviations
#                              Norm.Coeff. Fraction
- ----------- ---------
1 COMSIA_DONOR (1100 vars)          1.535 0.297
2 COMSIA_ACCEPTOR (1100 vars)       0.623 0.121
3 COMSIA_HYDROPHOBIC (1100 vars)    1.479 0.286
4 COMSIA_STERIC (1100 vars)         0.368 0.071
5 COMSIA_ELECTROSTATIC (1100 vars)  1.159 0.224
Summary output
Standard Error of Estimate
#                 Mean StdDev
- --------- ---------
6 r_m_Activity    0.781 0.406
R squared
#                 Mean StdDev
- ------ ------
6 r_m_Activity    0.826 0.052
```

YHY REFERENCE_SEED 2

Determine the optimal number of factors
Parameters: CoMFA standard scaling, Leave-one-out, SAMPLS, 20 components

```
Option Name       Description                                  Value
----------- ----------- -----
BOOTSTRAPPING     Number of bootstrap runs to make          0
CENTERING         Force intercept thru 0.0?                 NO
COMPONENTS        Number of components to use               20
CROSSVALIDATION   Number of crossvalidation groups          0
EPSILON           Convergence criterion                     1.e-4
ITERATION         Maximum number of iterations              100
SCALING_METHOD    Pre-analysis scaling          (COMFA_STD for CoMF COMFA_STD
XLIST Print info about xval groups? NO
SAMPLS has finished running. This is NOT a full PLS analysis.
Standard Error of Prediction for 20 components:
1.506 1.442 1.461 1.476 1.491 1.478 1.479 1.464 1.452 1.458 1.462 1.486 1.500
1.527 1.523 1.540 1.538 1.545 1.559 1.582
Crossvalidated R2 for 20 components:
0.391 0.445 0.434 0.426 0.418 0.432 0.435 0.450 0.463 0.462 0.462 0.448 0.442
0.426 0.433 0.425 0.430 0.429 0.423 0.410
-- optimum is 0.463 at 9 components
################################################
```

Determine the coefficient of determination (without cross validation) R^2
Parameters: CoMFA standard scaling, No cross validation, Filtering: 0 kcal, 9 components

```
Standard Error of Estimate          0.969
R squared                           0.761
F values      (n1= 9, n2=148)       52.305
Prob.of R^2 = 0 (n1= 9, n2=148)     0.000
Relative Contributions
                                    # Norm.Coeff. Fraction
- ----------- ---------
1 COMSIA_DONOR (1100 vars)                1.797 0.281
2 COMSIA_ACCEPTOR (1100 vars)             0.743 0.116
3 COMSIA_HYDROPHOBIC (1100 vars)          2.147 0.336
4 COMSIA_STERIC (1100 vars)               0.420 0.066
```

```
5 COMSIA_ELECTROSTATIC (1100 vars)        1.282 0.201
Summary output
Standard Error of Estimate                0.969
R squared                                 0.761
F values (n1= 9, n2=148)                  52.305
Prob.of R2 = 0 (n1= 9, n2=148)            0.000
```

Determine the coefficient of determination due to bootstrapping activity R^2_{bstr}

Parameters: CoMFA standard scaling, Bootstrap 10 runs, Filtering: 0 kcal, 9 components

```
Means
                                    # Norm.Coeff. Fraction
- ----------- ---------
1 COMSIA_DONOR (1100 vars)                1.658 0.273
2 COMSIA_ACCEPTOR (1100 vars)             0.692 0.114
3 COMSIA_HYDROPHOBIC (1100 vars)          2.060 0.340
4 COMSIA_STERIC (1100 vars)               0.411 0.068
5 COMSIA_ELECTROSTATIC (1100 vars)        1.245 0.205
Standard Deviations
# Norm.Coeff. Fraction
- ----------- ---------
1 COMSIA_DONOR (1100 vars) 1.144 0.267
2 COMSIA_ACCEPTOR (1100 vars) 0.519 0.121
3 COMSIA_HYDROPHOBIC (1100 vars) 1.277 0.298
4 COMSIA_STERIC (1100 vars) 0.305 0.071
5 COMSIA_ELECTROSTATIC (1100 vars) 1.033 0.241
Summary output
Standard Error of Estimate
# Mean StdDev
- --------- ---------
6 r_m_Activity 0.789 0.369
R squared
                    # Mean StdDev
- ------ ------
6 r_m_Activity 0.840 0.033
```

YHY REFERENCE_SEED 3

Determine the optimal number of factors

Parameters: CoMFA standard scaling, Leave-one-out, SAMPLS, 20 components

```
Option Name Description Value
----------- ----------- -----
BOOTSTRAPPING Number of bootstrap runs to make        0
CENTERING Force intercept thru 0.0?                   NO
COMPONENTS Number of components to use                20
CROSSVALIDATION Number of crossvalidation groups      0
EPSILON Convergence criterion                         1.e-4
ITERATION Maximum number of iterations                100
SCALING_METHOD Pre-analysis scaling (COMFA_STD for CoMF COMFA_STD
XLIST Print info about xval groups?                   NO
SAMPLS has finished running. This is NOT a full PLS analysis.
Standard Error of Prediction for 20 components:
1.497 1.454 1.487 1.506 1.503 1.491 1.468 1.437 1.431 1.455 1.458 1.472 1.488
1.513 1.529 1.518 1.550 1.540 1.559 1.574
Crossvalidated R2 for 20 components:
0.400 0.437 0.415 0.404 0.410 0.423 0.445 0.471 0.480 0.465 0.467 0.461 0.452
0.438 0.430 0.442 0.423 0.434 0.424 0.417
```

```
-- optimum is 0.48 at 9 components
```

Determine the coefficient of determination (without cross validation) R^2

Parameters: CoMFA standard scaling, No cross validation, Filtering: 0 kcal, 9 components

```
Standard Error of Estimate 0.987
R squared                          0.753
F values          (n1= 9, n2=148)   50.029
Prob.of R2=0      (n1= 9, n2=148)   0.000
Relative Contributions
#                            Norm.Coeff. Fraction
- ----------- ---------
1 COMSIA_DONOR (1100 vars)          1.937 0.288
2 COMSIA_ACCEPTOR (1100 vars)       0.690 0.102
3 COMSIA_HYDROPHOBIC (1100 vars)    2.322 0.345
4 COMSIA_STERIC (1100 vars)         0.452 0.067
5 COMSIA_ELECTROSTATIC (1100 vars)  1.328 0.197
Summary output
Standard Error of Estimate          0.987
R squared                           0.753
F values (n1= 9, n2=148)            50.029
Prob.of R2=0 (n1= 9, n2=148)  0.000
```

Determine the coefficient of determination due to bootstrapping activity R^2_{bstr}

Parameters: CoMFA standard scaling, Bootstrap 10 runs, Filtering: 0 kcal, 9 components

```
Means
# Norm.Coeff. Fraction
- ----------- ---------
1 COMSIA_DONOR (1100 vars)                1.628 0.293
2 COMSIA_ACCEPTOR (1100 vars)             0.565 0.102
3 COMSIA_HYDROPHOBIC (1100 vars)          1.942 0.349
4 COMSIA_STERIC (1100 vars)               0.413 0.074
5 COMSIA_ELECTROSTATIC (1100 vars)        1.013 0.182
Standard Deviations
# Norm.Coeff. Fraction
- ----------- ---------
1 COMSIA_DONOR (1100 vars)                1.136 0.265
2 COMSIA_ACCEPTOR (1100 vars)             0.548 0.128
3 COMSIA_HYDROPHOBIC (1100 vars)          1.260 0.294
4 COMSIA_STERIC (1100 vars)               0.325 0.076
5 COMSIA_ELECTROSTATIC (1100 vars)        1.017 0.237
Summary output
Standard Error of Estimate
# Mean StdDev
- --------- ---------
6 r_m_Activity 0.855 0.431
R squared
# Mean StdDev
- ------ ------
6 r_m_Activity 0.809 0.038
```

YHY REFERENCE_SEED 4

Determine the optimal number of factors

Parameters: CoMFA standard scaling, Leave-one-out, SAMPLS, 20 components

```
Option Name Description Value
```

```
---------- ---------- -----
BOOTSTRAPPING Number of bootstrap runs to make              0
CENTERING Force intercept thru 0.0?                         NO
COMPONENTS Number of components to use                      20
CROSSVALIDATION Number of crossvalidation groups            0
EPSILON Convergence criterion                               1.e-4
ITERATION Maximum number of iterations                      100
SCALING_METHOD Pre-analysis scaling (COMFA_STD for CoMF COMFA_STD
XLIST Print info about xval groups?                         NO
SAMPLS has finished running. This is NOT a full PLS analysis.
Standard Error of Prediction for 20 components:
1.500 1.476 1.522 1.545 1.530 1.516 1.499 1.465 1.442 1.428 1.430 1.446 1.464
1.480 1.498 1.507 1.520 1.523 1.556 1.559
Crossvalidated R2 for 20 components:
0.392 0.416 0.382 0.368 0.385 0.400 0.417 0.447 0.467 0.481 0.483 0.475 0.466
0.458 0.449 0.446 0.441 0.442 0.422 0.424
-- optimum is 0.483 at 11 components
```

Determine the coefficient of determination (without cross validation) R^2

Parameters: CoMFA standard scaling, No cross validation, Filtering: 0 kcal, 11 components

```
Standard Error of Estimate          0.910
R squared                           0.791
F values (n1=11, n2=146)            50.186
Prob.of R2=0 (n1=11, n2=146)        0.000
Relative Contributions
#                            Norm.Coeff. Fraction
- ---------- ---------
1 COMSIA_DONOR (1100 vars)          2.400 0.267
2 COMSIA_ACCEPTOR (1100 vars)       0.918 0.102
3 COMSIA_HYDROPHOBIC (1100 vars)    3.078 0.342
4 COMSIA_STERIC (1100 vars)         0.661 0.073
5 COMSIA_ELECTROSTATIC (1100 vars)  1.939 0.216
Summary output
Standard Error of Estimate          0.910
R squared                           0.791
F values (n1=11, n2=146)            50.186
Prob.of R2=0 (n1=11, n2=146)        0.000
```

Determine the coefficient of determination due to bootstrapping activity R^2_{bstr}

Parameters: CoMFA standard scaling, Bootstrap 10 runs, Filtering: 0 kcal, 11 components

```
Means
# Norm.Coeff. Fraction
- ---------- ---------
1 COMSIA_DONOR (1100 vars)              1.856 0.267
2 COMSIA_ACCEPTOR (1100 vars)           0.704 0.101
3 COMSIA_HYDROPHOBIC (1100 vars)        2.364 0.340
4 COMSIA_STERIC (1100 vars)             0.573 0.082
5 COMSIA_ELECTROSTATIC (1100 vars)      1.463 0.210
Standard Deviations
# Norm.Coeff. Fraction
- ---------- ---------
1 COMSIA_DONOR (1100 vars)              1.372 0.288
2 COMSIA_ACCEPTOR (1100 vars)           0.585 0.123
3 COMSIA_HYDROPHOBIC (1100 vars)        1.259 0.265
4 COMSIA_STERIC (1100 vars)             0.356 0.075
5 COMSIA_ELECTROSTATIC (1100 vars)      1.187 0.249
Summary output
Standard Error of Estimate
```

```
# Mean StdDev
- --------- ---------
6 r_m_Activity 0.706 0.336
R squared
# Mean StdDev
- ------ ------
6 r_m_Activity 0.873 0.034
```

YHY REFERENCE_SEED 5

Determine the optimal number of factors

Parameters: CoMFA standard scaling, Leave-one-out, SAMPLS, 20 components

```
Option Name Description Value
---------- ---------- -----
BOOTSTRAPPING Number of bootstrap runs to make        0
CENTERING Force intercept thru 0.0?                   NO
COMPONENTS Number of components to use                20
CROSSVALIDATION Number of crossvalidation groups      0
EPSILON Convergence criterion                         1.e-4
ITERATION Maximum number of iterations                100
SCALING_METHOD Pre-analysis scaling (COMFA_STD for CoMF COMFA_STD
XLIST Print info about xval groups?                   NO
SAMPLS has finished running. This is NOT a full PLS analysis.
Standard Error of Prediction for 20 components:
1.485 1.431 1.479 1.497 1.503 1.470 1.464 1.430 1.424 1.422 1.431 1.452 1.471
1.485 1.491 1.497 1.496 1.523 1.549 1.546
Crossvalidated R2 for 20 components:
0.403 0.450 0.416 0.405 0.405 0.434 0.442 0.471 0.479 0.484 0.481 0.470 0.460
0.453 0.453 0.452 0.456 0.441 0.425 0.432
-- optimum is 0.484 at 10 components
```

Determine the coefficient of determination (without cross validation) R^2

Parameters: CoMFA standard scaling, No cross validation, Filtering: 0 kcal, 9 components

```
Standard Error of Estimate                0.959
R squared                                 0.766
F values (n1=10, n2=147)                  47.987
Prob.of R2=0 (n1=10, n2=147)              0.000
Relative Contributions
#                                  Norm.Coeff. Fraction
- ----------- ---------
1 COMSIA_DONOR (1100 vars)                1.947 0.263
2 COMSIA_ACCEPTOR (1100 vars)             0.866 0.117
3 COMSIA_HYDROPHOBIC (1100 vars)          2.519 0.340
4 COMSIA_STERIC (1100 vars)               0.518 0.070
5 COMSIA_ELECTROSTATIC (1100 vars)        1.553 0.210
Summary output
Standard Error of Estimate 0.959
R squared 0.766
F values (n1=10, n2=147) 47.987
Prob.of R2=0 (n1=10, n2=147) 0.000
```

Determine the coefficient of determination due to bootstrapping activity R^2_{bstr}

Parameters: CoMFA standard scaling, Bootstrap 10 runs, Filtering: 0 kcal, 9 components

```
Means
# Norm.Coeff. Fraction
- ----------- ---------
```

```
1 COMSIA_DONOR (1100 vars)                1.541 0.265
2 COMSIA_ACCEPTOR (1100 vars)             0.670 0.115
3 COMSIA_HYDROPHOBIC (1100 vars)          1.955 0.337
4 COMSIA_STERIC (1100 vars)               0.408 0.070
5 COMSIA_ELECTROSTATIC (1100 vars)        1.234 0.212
Standard Deviations
# Norm.Coeff. Fraction
- ----------- ---------
1 COMSIA_DONOR (1100 vars)                1.275 0.268
2 COMSIA_ACCEPTOR (1100 vars)             0.597 0.125
3 COMSIA_HYDROPHOBIC (1100 vars)          1.343 0.282
4 COMSIA_STERIC (1100 vars)               0.339 0.071
5 COMSIA_ELECTROSTATIC (1100 vars)        1.210 0.254
Summary output
Standard Error of Estimate
# Mean StdDev
- --------- ---------
6 r_m_Activity 0.850 0.344
R squared
# Mean StdDev
- ------ ------
6 r_m_Activity 0.820 0.031
```

YRY REFERENCE_SEED 1

Determine the optimal number of factors

Parameters: CoMFA standard scaling, Leave-one-out, SAMPLS, 20 components

```
Option Name    Description                              Value
-----------  -----------  -----
BOOTSTRAPPING Number of bootstrap runs to make             0
CENTERING Force intercept thru 0.0?                        NO
COMPONENTS Number of components to use 20
CROSSVALIDATION Number of crossvalidation groups           0
EPSILON Convergence criterion                              1.e-4
ITERATION Maximum number of iterations                     100
SCALING_METHOD Pre-analysis scaling (COMFA_STD for COMFA_STD
XLIST Print info about xval groups?                        NO
SAMPLS has finished running. This is NOT a full PLS analysis.
XLIST Print info about xval groups? NO
SAMPLS has finished running. This is NOT a full PLS analysis.
Standard Error of Prediction for 20 components:
1.494 1.408 1.478 1.398 1.420 1.378 1.363 1.365 1.353 1.341 1.337 1.325 1.335
1.333 1.357 1.343 1.350 1.350 1.366 1.383
Crossvalidated R2 for 20 components:
0.356 0.432 0.378 0.447 0.433 0.470 0.484 0.486 0.499 0.511 0.518 0.529 0.525
0.530 0.516 0.530 0.528 0.531 0.524 0.515
-- optimum is 0.531 at 18 components
```

Determine the coefficient of determination (without cross-validation) R^2

Parameters: CoMFA standard scaling, No cross validation, Filtering: 0 kcal, 18 components

```
Relative Contributions
                              # Norm.Coeff. Fraction
- ----------- ---------
1 COMSIA_DONOR (1296 vars)          0.629 0.217
```

```
2 COMSIA_ACCEPTOR (1296 vars)       0.385 0.133
3 COMSIA_HYDROPHOBIC (1296 vars)    1.042 0.359
4 COMSIA_STERIC (1296 vars)         0.204 0.070
5 COMSIA_ELECTROSTATIC (1296 vars)  0.644 0.222
Summary output
Standard Error of Estimate          1.116
R squared                           0.636
F values (n1= 3, n2=153)            89.301
Prob.of R2=0(n1= 3, n2=153)         0.000
```

Determine the coefficient of determination due to bootstrapping activity R^2_{bstr}

Parameters: CoMFA standard scaling, Bootstrap 100 runs, Filtering: 0 kcal, 18 components

```
Means
#                           Norm.Coeff. Fraction
- ----------- ---------
1 COMSIA_DONOR (1296 vars)          0.597 0.211
2 COMSIA_ACCEPTOR (1296 vars)       0.375 0.133
3 COMSIA_HYDROPHOBIC (1296 vars)    1.013 0.358
4 COMSIA_STERIC (1296 vars)         0.201 0.071
5 COMSIA_ELECTROSTATIC (1296 vars)  0.641 0.227
Standard Deviations
#                           Norm.Coeff. Fraction
- ----------- ---------
1 COMSIA_DONOR (1296 vars)          0.358 0.269
2 COMSIA_ACCEPTOR (1296 vars)       0.191 0.144
3 COMSIA_HYDROPHOBIC (1296 vars)    0.366 0.275
4 COMSIA_STERIC (1296 vars)         0.095 0.071
5 COMSIA_ELECTROSTATIC (1296 vars)  0.319 0.240
Summary output
Standard Error of Estimate
#                           Mean StdDev
- --------- ---------
1 R_USER_ACTIVITY           1.029 0.372
R squared
#                           Mean StdDev
- ------ ------
1 R_USER_ACTIVITY           0.684 0.051
```

YRY REFERENCE_SEED 2

Determine the optimal number of factors

Parameters: CoMFA standard scaling, Leave-one-out, SAMPLS, 15 components

```
Option Name Description Value
----------- ----------- -----
BOOTSTRAPPING Number of bootstrap runs to make              0
CENTERING Force intercept thru 0.0?                         NO
COMPONENTS Number of components to use                      15
CROSSVALIDATION Number of crossvalidation groups 0
EPSILON Convergence criterion 1.e-4
ITERATION Maximum number of iterations 100
SCALING_METHOD Pre-analysis scaling (COMFA_STD for CoMF COMFA_STD
XLIST Print info about xval groups? NO
SAMPLS has finished running. This is NOT a full PLS analysis.
Standard Error of Prediction for 15 components:
```

```
1.556 1.473 1.479 1.503 1.519 1.520 1.532 1.553 1.557 1.579 1.602 1.610 1.613
1.629 1.625
Crossvalidated R2 for 15 components:
0.350 0.421 0.420 0.405 0.396 0.399 0.394 0.381 0.382 0.369 0.354 0.353 0.355
0.347 0.354
-- optimum is 0.421 at 2 components
```

Determine the coefficient of determination (without cross-validation) R^2

Parameters: CoMFA standard scaling, No cross validation, Filtering: 0 kcal, 2 components

```
Standard Error of Estimate                1.325
R squared                                 0.531
F values ( n1= 2, n2=155 )                87.797
Prob.of R2=0 ( n1= 2, n2=155 )            0.000
Relative Contributions
#                                    Norm.Coeff. Fraction
- ----------- ---------
1 COMSIA_DONOR (1152 vars)                0.440 0.196
2 COMSIA_ACCEPTOR (1152 vars)             0.333 0.148
3 COMSIA_HYDROPHOBIC (1152 vars)          0.744 0.330
4 COMSIA_STERIC (1152 vars)               0.146 0.065
5 COMSIA_ELECTROSTATIC (1152 vars)        0.589 0.262
Summary output
Standard Error of Estimate                1.325
R squared                                 0.531
F values (n1= 2, n2=155)                  87.797
Prob.of R2=0 ( n1= 2, n2=155 )            0.000
```

Determine the coefficient of determination due to bootstrapping activity R^2_{bstr}

Parameters: CoMFA standard scaling, Bootstrap 100 runs, Filtering: 0 kcal, 2 components

```
Means
#                                    Norm.Coeff. Fraction
- ----------- ---------
1 COMSIA_DONOR (1152 vars)                0.441 0.197
2 COMSIA_ACCEPTOR (1152 vars)             0.324 0.145
3 COMSIA_HYDROPHOBIC (1152 vars)          0.749 0.334
4 COMSIA_STERIC (1152 vars)               0.146 0.065
5 COMSIA_ELECTROSTATIC (1152 vars)        0.580 0.259
Standard Deviations
#                                    Norm.Coeff. Fraction
- ----------- ---------
1 COMSIA_DONOR (1152 vars)                0.268 0.260
2 COMSIA_ACCEPTOR (1152 vars)             0.154 0.149
3 COMSIA_HYDROPHOBIC (1152 vars)          0.279 0.270
4 COMSIA_STERIC (1152 vars)               0.078 0.075
5 COMSIA_ELECTROSTATIC (1152 vars)        0.253 0.245
Summary output
Standard Error of Estimate
# Mean StdDev
- --------- ---------
6 r_m_Activity 1.250 0.492
R squared
# Mean StdDev
- ------ ------
6 r_m_Activity 0.583 0.041
```

YRY REFERENCE_SEED 3

Determine the optimal number of factors

Parameters: CoMFA standard scaling, Leave-one-out, SAMPLS, 20 components

```
Option Name Description Value
---------- ---------- -----
BOOTSTRAPPING Number of bootstrap runs to make        0
CENTERING Force intercept thru 0.0?                   NO
COMPONENTS Number of components to use                20
CROSSVALIDATION Number of crossvalidation groups      0
EPSILON Convergence criterion                         1.e-4
ITERATION Maximum number of iterations                100
SCALING_METHOD Pre-analysis scaling (COMFA_STD for CoMF COMFA_STD
XLIST Print info about xval groups? NO
SAMPLS has finished running. This is NOT a full PLS analysis.
Standard Error of Prediction for 20 components:
1.556 1.475 1.473 1.491 1.490 1.476 1.481 1.498 1.486 1.507 1.514 1.502 1.519
1.530 1.491 1.493 1.493 1.507 1.512 1.516
Crossvalidated R2 for 20 components:
0.352 0.421 0.426 0.416 0.421 0.435 0.435 0.426 0.439 0.427 0.425 0.439 0.429
0.425 0.458 0.461 0.464 0.458 0.458 0.459
-- optimum is 0.464 at 17 components
```

Determine the coefficient of determination (without cross-validation) R^2

Parameters: CoMFA standard scaling, No cross validation, Filtering: 0 kcal, 17 components

```
Standard Error of Estimate 0.825
R squared 0.837
F values ( n1=17, n2=140 ) 42.148
Prob.of R2=0 ( n1=17, n2=140 ) 0.000
Relative Contributions
# Norm.Coeff. Fraction
- ----------- ---------
1 COMSIA_DONOR (1296 vars) 2.843 0.261
2 COMSIA_ACCEPTOR (1296 vars) 1.493 0.137
3 COMSIA_HYDROPHOBIC (1296 vars) 3.734 0.342
4 COMSIA_STERIC (1296 vars) 0.887 0.081
5 COMSIA_ELECTROSTATIC (1296 vars) 1.947 0.179
Summary output
Standard Error of Estimate 0.825
R squared 0.837
F values ( n1=17, n2=140 ) 42.148
Prob.of R2=0 ( n1=17, n2=140 ) 0.000
```

Determine the coefficient of determination due to bootstrapping activity R^2_{bstr}

Parameters: CoMFA standard scaling, Bootstrap 100 runs, Filtering: 0 kcal, 17 components

```
Means
#                                      Norm.Coeff. Fraction
- ----------- ---------
1 COMSIA_DONOR (1296 vars)                   2.253 0.246
2 COMSIA_ACCEPTOR (1296 vars)                1.174 0.128
3 COMSIA_HYDROPHOBIC (1296 vars)             3.265 0.356
4 COMSIA_STERIC (1296 vars)                  0.834 0.091
5 COMSIA_ELECTROSTATIC (1296 vars)           1.646 0.179
Standard Deviations
#                                      Norm.Coeff. Fraction
- ----------- ---------
1 COMSIA_DONOR (1296 vars)                   2.258 0.303
2 COMSIA_ACCEPTOR (1296 vars)                0.992 0.133
```

```
3 COMSIA_HYDROPHOBIC (1296 vars)                1.967 0.264
4 COMSIA_STERIC (1296 vars)                     0.638 0.086
5 COMSIA_ELECTROSTATIC (1296 vars)              1.601 0.215
Summary output
Standard Error of Estimate
# Mean StdDev
- --------- ---------
6 r_m_Activity 0.700 0.282
R squared
# Mean StdDev
- ------ ------
6 r_m_Activity 0.879 0.026
```

YRY REFERENCE_SEED 4

Determine the optimal number of factors

Parameters: CoMFA standard scaling, Leave-one-out, SAMPLS, 20 components

```
Option Name Description Value
----------- ----------- -----
BOOTSTRAPPING Number of bootstrap runs to make 1            00
CENTERING Force intercept thru 0.0?                         NO
COMPONENTS Number of components to use                      20
CROSSVALIDATION Number of crossvalidation groups            0
EPSILON Convergence criterion                               1.e-4
ITERATION Maximum number of iterations                      100
SCALING_METHOD Pre-analysis scaling (COMFA_STD for CoMF COMFA_STD
XLIST Print info about xval groups?                         NO
SAMPLS has finished running. This is NOT a full PLS analysis.
Standard Error of Prediction for 20 components:
1.567 1.506 1.492 1.513 1.498 1.486 1.488 1.498 1.503 1.499 1.538 1.536 1.528
1.554 1.543 1.541 1.534 1.539 1.565 1.573
Crossvalidated R2 for 20 components:
0.337 0.392 0.406 0.394 0.410 0.423 0.426 0.422 0.422 0.429 0.403 0.408 0.418
0.402 0.415 0.420 0.430 0.431 0.415 0.414
-- optimum is 0.431 at 18 components
```

Determine the coefficient of determination (without cross-validation) R^2

Parameters: CoMFA standard scaling, No cross validation, Filtering: 0 kcal, 18 components

```
Standard Error of Estimate 0.847
R squared 0.827
F values ( n1=18, n2=139 ) 36.988
Prob.of R2=0 (n1=18, n2=139 ) 0.000
Relative Contributions
# Norm.Coeff. Fraction
- ----------- ---------
1 COMSIA_DONOR (1152 vars) 2.768 0.259
2 COMSIA_ACCEPTOR (1152 vars) 1.495 0.140
3 COMSIA_HYDROPHOBIC (1152 vars) 3.633 0.340
4 COMSIA_STERIC (1152 vars) 0.892 0.084
5 COMSIA_ELECTROSTATIC (1152 vars) 1.887 0.177
Summary output
Standard Error of Estimate 0.847
R squared 0.827
F values (n1=18, n2=139 ) 36.988
Prob.of R2=0 (n1=18, n2=139 ) 0.000
```

Determine the coefficient of determination due to bootstrapping activity R^2_{bstr}

Parameters: CoMFA standard scaling, Bootstrap 10 runs, Filtering: 0 kcal, 18 components

```
Means
#                                  Norm.Coeff. Fraction
- ----------- ---------
1 COMSIA_DONOR (1152 vars)          2.265 0.243
2 COMSIA_ACCEPTOR (1152 vars)       1.230 0.132
3 COMSIA_HYDROPHOBIC (1152 vars)    3.398 0.365
4 COMSIA_STERIC (1152 vars)         0.813 0.087
5 COMSIA_ELECTROSTATIC (1152 vars)  1.614 0.173
Standard Deviations
#                             Norm.Coeff. Fraction
- ----------- ---------
1 COMSIA_DONOR (1152 vars)          2.323 0.311
2 COMSIA_ACCEPTOR (1152 vars)       0.974 0.130
3 COMSIA_HYDROPHOBIC (1152 vars)    1.993 0.267
4 COMSIA_STERIC (1152 vars)         0.661 0.089
5 COMSIA_ELECTROSTATIC (1152 vars)  1.517 0.203
Summary output
Standard Error of Estimate
# Mean StdDev
- --------- ---------
6 r_m_Activity 0.697 0.259
R squared
# Mean StdDev
- ------ ------
6 r_m_Activity 0.881 0.022
```

YRY REFERENCE_SEED 5

Determine the optimal number of factors

Parameters: CoMFA standard scaling, Leave-one-out, SAMPLS, 20 components

```
Option Name Description Value
----------- ----------- -----
BOOTSTRAPPING Number of bootstrap runs to make              0
CENTERING Force intercept thru 0.0?                         NO
COMPONENTS Number of components to use                      20
CROSSVALIDATION Number of crossvalidation groups            0
EPSILON Convergence criterion                               1.e-4
ITERATION Maximum number of iterations                      100
SCALING_METHOD Pre-analysis scaling (COMFA_STD for CoMF COMFA_STD
XLIST Print info about xval groups?                         NO
SAMPLS has finished running. This is NOT a full PLS analysis.
Standard Error of Prediction for 20 components:
1.539 1.464 1.461 1.476 1.482 1.461 1.467 1.464 1.461 1.465 1.501 1.484 1.506
1.493 1.497 1.486 1.492 1.490 1.502 1.518
Crossvalidated R2 for 20 components:
0.359 0.424 0.430 0.421 0.421 0.441 0.440 0.446 0.452 0.453 0.429 0.446 0.433
0.447 0.448 0.460 0.459 0.465 0.460 0.453
-- optimum is 0.465 at 18 components
```

Determine the coefficient of determination (without cross-validation) R^2

Parameters: CoMFA standard scaling, No cross validation, Filtering: 0 kcal, 18 components

Standard Error of Estimate 0.836
R squared 0.831
F values (n1=18, n2=139) 38.077
Prob.of R2=0 (n1=18, n2=139) 0.000

Relative Contributions

#		Norm.Coeff.	Fraction
1	COMSIA_DONOR (1296 vars)	2.681	0.255
2	COMSIA_ACCEPTOR (1296 vars)	1.450	0.138
3	COMSIA_HYDROPHOBIC (1296 vars)	3.581	0.340
4	COMSIA_STERIC (1296 vars)	0.888	0.084
5	COMSIA_ELECTROSTATIC (1296 vars)	1.934	0.184

Determine the coefficient of determination due to bootstrapping activity R^2_{bstr}

Parameters: CoMFA standard scaling, Bootstrap 10 runs, Filtering: 0 kcal, 18 components

Means

#		Norm.Coeff.	Fraction
1	COMSIA_DONOR (1296 vars)	2.045	0.236
2	COMSIA_ACCEPTOR (1296 vars)	1.109	0.128
3	COMSIA_HYDROPHOBIC (1296 vars)	3.120	0.361
4	COMSIA_STERIC (1296 vars)	0.812	0.094
5	COMSIA_ELECTROSTATIC (1296 vars)	1.566	0.181

Standard Deviations

#		Norm.Coeff.	Fraction
1	COMSIA_DONOR (1296 vars)	2.246	0.304
2	COMSIA_ACCEPTOR (1296 vars)	0.933	0.126
3	COMSIA_HYDROPHOBIC (1296 vars)	1.954	0.265
4	COMSIA_STERIC (1296 vars)	0.637	0.086
5	COMSIA_ELECTROSTATIC (1296 vars)	1.613	0.219

Summary output

Standard Error of Estimate

#	Mean	StdDev
6	r_m_Activity	0.713	0.233

R squared

#	Mean	StdDev
6	r_m_Activity	0.865	0.020

Supplementary 6. Observed vs. predicted FTC values derived from OPLS, Tripos-based models and GBR model.

OPLS-based models				Tripos-based models					
YHY ref_seed 1		YRY ref_seed 1		YHY ref_seed 1		YRY ref_seed 1		RFE-GBR	
Observed	Predicted	Observed	Predicted	Observed	Predicted	Observed	Predicted	Observed	Predicted
3.918	4.568	3.918	4.201	3.918	4.720	3.918	4.475	1.313	2.025
6.697	4.482	6.697	4.143	6.697	5.015	6.697	5.010	1.320	2.884
4.504	4.537	4.504	4.079	4.504	3.675	4.504	3.498	2.233	2.111
4.136	4.095	4.136	4.044	4.136	3.979	4.136	3.963	1.091	1.434
5.584	5.276	5.584	4.833	5.584	5.773	5.584	5.448	2.505	1.802
1.313	1.670	5.197	4.441	1.313	2.342	5.197	5.179	5.197	4.814
5.197	4.468	6.061	4.699	5.197	5.018	6.061	4.872	1.192	1.971
6.061	4.826	1.396	1.474	6.061	4.298	1.396	1.621	9.886	7.256
1.396	1.658	1.096	1.689	1.396	2.083	1.096	0.995	3.378	3.994
1.096	1.421	2.598	2.063	1.096	1.245	2.598	2.053	3.417	2.478
1.184	0.923	3.378	4.211	1.184	1.697	3.378	4.449	4.136	4.408
0.407	1.505	0.904	2.198	0.407	1.175	0.904	2.764	0.808	1.910
2.598	2.431	1.803	3.922	2.598	3.065	1.803	3.462	2.026	0.984
3.378	4.640	0.937	2.488	3.378	4.418	0.937	-0.629	2.020	1.343
3.212	1.303	5.303	4.104	3.212	1.373	5.303	4.479	0.905	2.129
0.727	1.305	3.344	3.626	0.727	0.655	3.344	3.849	5.311	4.244
0.808	1.117	3.983	3.760	0.808	0.687	3.983	4.012	4.227	4.185
2.707	2.621	1.313	1.429	2.707	3.039	1.313	2.284	6.791	5.601
0.905	2.078	2.02	1.818	0.905	2.499	2.02	2.019	4.302	4.370
2.020	2.065	3.071	3.971	2.02	2.423	3.071	5.591	3.607	2.395
1.803	2.506	3.071	3.590	1.803	3.301	3.071	3.563	1.394	1.733
0.937	2.100	5.071	1.782	0.937	0.881	5.071	1.794	0.188	1.260
1.477	0.554	3.212	1.253	1.477	0.610	3.212	1.307	1.096	1.601
0.952	0.503	0.727	0.896	0.952	0.757	0.727	0.255	0.727	0.670
2.026	0.228	0.808	1.374	2.026	-0.304	0.808	1.431	1.089	1.766
0.188	2.368	2.707	2.787	0.188	2.037	2.707	2.757	6.152	4.637
3.612	3.524	3.607	3.548	3.612	5.030	3.607	3.399	3.047	3.191
3.071	3.306	1.184	1.576	3.071	3.680	1.184	2.236	3.718	3.911
3.071	4.161	2.233	1.841	3.071	4.873	2.233	1.175	1.803	1.986
1.911	2.974	3.447	3.226	1.911	3.945	3.447	3.395	3.366	3.782
5.071	2.343	9.886	4.061	5.071	2.016	9.886	3.374	6.697	4.366
9.886	5.737	1.911	2.929	9.886	5.839	1.911	3.518	3.607	2.487
3.607	3.394	0.188	2.276	3.607	3.473	0.188	1.912	2.996	3.527
2.233	1.880	3.612	2.246	2.233	1.565	3.612	0.932	2.384	1.835
3.447	3.613	2.996	3.359	3.447	3.542	2.996	4.350	4.116	3.734
5.303	3.530	0.407	2.135	5.303	3.823	0.407	1.449	1.320	3.469
3.344	3.433	1.477	0.711	3.344	4.011	1.477	0.980	3.344	3.603
2.996	3.095	0.952	0.722	2.996	3.088	0.952	1.325	2.996	3.339
3.983	3.314	2.026	0.746	3.983	2.273	2.026	0.251	0.606	1.732

Supplementary 7.

Table S7.1. Five CoMSIA interaction fields from 13 Tryptophyllin peptides.

Table S7.1 can be found at: https://docs.google.com/spreadsheets/d/1YOqhC-kpOeS_yjLj2ngXLlzDuYDIQl4S/edit?usp=sharing&ouid=106421106263255583273&rtpof=true&sd=true

Script S1- 6.

Script S1-6 can be found at https://drive.google.com/drive/folders/1vdrckX9X1GjO3MTsCyV7At1yyc0PGaBG?usp=sharing