Electronic Supplementary Material (ESI) for Molecular BioSystems.

Table S.1 The explanation of the BOND and NOSP attributes.

=	#	@	Comments
Calculation of the BOND index			
0	0	0	There are no double, triple, or stereochemical bonds.
0	0	1	The molecule contains only stereochemical bonds.
0	1	0	The molecule contains only triple bonds.
0	1	1	The molecule contains triple and stereochemical bonds.
1	0	0	The molecule contains only double bonds.
1	0	1	The molecule contains double bonds and stereochemical bonds.
1	1	0	The molecule contains double and triple bonds.
1	1	1	The molecule contains double, triple, and stereochemical bonds.

N	O	S	P	Comments
Calculation of the NOSP index				
0	0	0	0	Nitrogen, oxygen, sulfur, and phosphorus are absent.
0	0	0	1	The molecule contains only phosphorus.
0	0	1	0	The molecule contains only sulfur.
0	0	1	1	The molecule contains sulfur and phosphorus.
0	1	0	0	The molecule contains only oxygen.
0	1	0	1	The molecule contains oxygen and phosphorus.
0	1	1	0	The molecule contains oxygen and sulfur.
0	1	1	1	The molecule contains oxygen, sulfur, and phosphorus.
1	0	0	0	The molecule contains only nitrogen.
1	0	0	1	The molecule contains nitrogen and phosphorus.
1	0	1	0	The molecule contains nitrogen and sulfur.
1	0	1	1	The molecule contains nitrogen, sulfur, and phosphorus.
1	1	0	0	The molecule contains nitrogen and oxygen.
1	1	0	1	The molecule contains nitrogen, oxygen and phosphorus.
1	1	1	0	The molecule contains nitrogen, oxygen, and sulfur.
1	1	1	1	The molecule contains nitrogen, oxygen, sulfur, and phosphorus.

Table S.2 The methods and results in the SMILES-based models.

Method	Split1	Training set n=41				Test set n=14			
		R^{2a}	Q^{2b}	S^c	F^d	R^2	Q^2	S	F
S_k,SS_k,	Run1(0, 7)	0.9320	0.9243	0.163	535	0.8262	0.7582	0.289	57
SSS_k,	Run2(2, 8)	0.9345	0.9272	0.160	556	0.8138	0.7488	0.295	52
BOND,	Run3(0, 6)	0.9116	0.9017	0.186	402	0.8413	0.7698	0.282	64
NOSP	Average	0.9274	0.9177	0.170	498	0.8271	0.7589	0.289	58
Method	Split2	Training set n=41				Test set n=14			
		R^2	Q^2	S	F	R^2	Q^2	S	F
S_k,SS_k,	Run1(4, 9)	0.9178	0.9094	0.180	436	0.7539	0.6811	0.353	37
SSS_k,	Run2(5, 9)	0.9136	0.9045	0.184	412	0.7141	0.6253	0.380	30
BOND,	Run3(4, 8)	0.9134	0.9041	0.185	411	0.7358	0.6528	0.355	33
NOSP	Average	0.9149	0.9060	0.183	420	0.7346	0.6530	0.363	33
Method	Split3	Training set n=42				Test set n=13			
		R^2	Q^2	S	F	R^2	Q^2	S	F
S_k,SS_k, BOND, NOSP	Run1(5, 17)	0.8696	0.8559	0.233	267	0.8468	0.7969	0.292	61
	Run2(4, 19)	0.8787	0.8659	0.225	290	0.7831	0.7194	0.368	40
	Run3(4, 19)	0.8705	0.8568	0.232	269	0.7942	0.7401	0.322	42
	Average	0.8729	0.8595	0.230	275	0.8080	0.7521	0.327	48
Method	Split4	Training set n=42				Test set n=13			
		R^2	Q^2	S	F	R^2	Q^2	S	F
S_k, BOND, NOSP	Run1(4, 9)	0.8366	0.8213	0.261	205	0.7327	0.6297	0.348	30
	Run2(2, 8)	0.8139	0.7981	0.279	175	0.7474	0.6506	0.343	33
	Run3(2, 8)	0.8275	0.8121	0.268	192	0.7380	0.6400	0.347	31
	Average	0.8260	0.8105	0.269	191	0.7394	0.6401	0.346	31

[a] R^2 is correlation coefficient;

[b] Q^2 is cross-validated correlation coefficient;

[c] S is standard error of estimation;

[d] F is Fischer F-ratio.

Table S.3 The verification for the test set in the SMILES-based models.

Split	Run	R_0^2	R'^2_0	$\frac{(R^2-R_0^2)}{R^2}$	$\frac{(R^2-R'^2_0)}{R^2}$	R_m^2	R'^2_m	$\overline{R_m}$	ΔR_m	k	k'
Split1	Run1	0.8101	0.6908	0.0195	0.1638	0.7214	0.5222	0.6218	0.1991	1.0015	0.9957
	Run2	0.8045	0.7002	0.0114	0.1396	0.7355	0.5395	0.6375	0.1959	1.0069	0.9903
	Run3	0.8211	0.7036	0.0240	0.1637	0.7218	0.5291	0.6255	0.1927	0.9944	1.0030
Split2	Run1	0.7440	0.7272	0.0131	0.0354	0.6790	0.6308	0.6549	0.0482	0.9792	1.0174
	Run2	0.6880	0.6928	0.0365	0.0299	0.5988	0.6098	0.6043	0.0110	0.9830	1.0127
	Run3	0.7275	0.6983	0.0113	0.0511	0.6687	0.5932	0.6310	0.0755	0.9842	1.0121
Split3	Run1	0.7626	0.8377	0.0994	0.0107	0.6011	0.7662	0.6836	0.1651	1.0063	0.9909
	Run2	0.7435	0.7818	0.0505	0.0017	0.6273	0.7547	0.6910	0.1274	1.0407	0.9579
	Run3	0.7290	0.7937	0.0821	0.0005	0.5914	0.7780	0.6847	0.1866	1.0163	0.9807
Split4	Run1	0.6603	0.7318	0.0988	0.0013	0.5356	0.7100	0.6228	0.1744	1.0032	0.9928
	Run2	0.6802	0.7468	0.0899	0.0008	0.5537	0.7296	0.6417	0.1760	1.0123	0.9841
	Run3	0.6676	0.7372	0.0954	0.0011	0.5421	0.7169	0.6295	0.1748	1.0087	0.9874

Table S.4 The parameters of the CoMSIA models after changing the combination of fields.

Split	Parameters	CoMSIA (SEA)	CoMSIA (SEH)	CoMSIA (SED)	CoMSIA (SEAH)	CoMSIA (SEAD)	CoMSIA (SEADH)[f]
Split1	Q^2[a]	0.651	0.488	0.623	0.565	0.591	0.521
	ONC[b]	7	5	7	6	8	7
	R^2[c]	0.937	0.923	0.939	0.935	0.941	0.940
	S[d]	0.173	0.185	0.170	0.173	0.171	0.169
	F[e]	90	84	73	82	63	73
Split2	Q^2	0.754	0.636	0.700	0.710	0.694	0.659
	ONC	7	8	9	9	7	9
	R^2	0.980	0.989	0.987	0.989	0.976	0.987
	S	0.098	0.074	0.082	0.074	0.107	0.080
	F	229	351	259	320	191	267
Split3	Q^2	0.627	0.649	0.573	0.641	0.577	0.591
	ONC	8	9	9	6	9	10
	R^2	0.935	0.963	0.944	0.930	0.940	0.959
	S	0.184	0.141	0.173	0.185	0.179	0.150
	F	59	92	60	77	56	73

[a] Cross-validated correlation coefficient after the leave-one-out procedure;

[b] Optimum number of components;

[c] Non-cross-validated correlation coefficient;

[d] Standard error of estimate;

[e] F is Fischer F-ratio;

[f] S, steric; E, electrostatic; H, hydrophobic field; D, hydrogen bond donor; A, hydrogen bond acceptor;