SUPPLEMENTARY MATERIAL

Calculation of similarity

The matching of two nodes in the comparison between two molecules yields a weight given by the following equation:

$$sim(a, M(a)) = \sum_{k=1}^{11} W_k \min(a_k, M(a)_k)$$

where a is the node in the first molecule, M(a) its match, k runs over the indices of the features listed in Table 2, W_k, a_k, and $M(a)_k$ are the weight, and the value of the corresponding feature in the first and the second node, respectively.

The overall similarity is given as the sum of all similarities between matched nodes, normalized over the maximum of the self similarities of the compared molecules.

$$cw(K, L, M) = \sum_{a \in K} sim(a, M(a))$$

$$s(K, L, M) = \frac{cw(K, L, M)}{\max(cw(K, K, I), cw(L, L), I)}$$

where K and L are the ligands being compared, M is the mapping found by the algorithm, and I is the identity mapping.

Selected descriptors in the SVMR model

Table 1 Shown are the selected descriptors of the applied SVMR, their weights in the model and the selection order

Selection order	Dragon-ID	Description	Weight
1	ATS2p	Broto-Moreau autocorrelation of a topological structure - lag 2 / weighted by atomic polarizabilities	-4.865
2	L2s	2nd component size directional WHIM index / weighted by atomic electrotopological states	2.383
3	Mor22u	3D-MoRSE - signal 22 / unweighted	-1.794
4	nArOH	number of aromatic hydroxyl groups	-2.968
5	Ms	mean electrotopological state	2.975
6	R3u+	R maximal autocorrelation of lag 3 / unweighted	2.533
7	EEig11r	Eigenvalue 11 from edge adj. matrix weighted by resonance integrals	2.472
8	Mor17p	3D-MoRSE - signal 17 / weighted by atomic polarizabilities	1.182
9	MLOGP2	Squared Moriguchi octanol-water partition coeff. ($logP^2$)	-2.387
10	nCb-	number of substituted benzene C(sp2)	3.448
11	DISPe	d COMMA2 value / weighted by atomic Sanderson electronegativities	-0.941
12	RDF100m	Radial Distribution Function - 10.0 / weighted by	2.351

		atomic masses	
13	GGI5	topological charge index of order 5	-1.157
14	S-107	R2S / RS-SR atom-centred fragments	-1.161
15	H-049	H attached to C3(sp3) / C2(sp2) / C3(sp2) / C3(sp) atom-centred fragments	0.762
16	EEig10d	Eigenvalue 10 from edge adj. matrix weighted by dipole moments	1.548
17	Mor12m	3D-MoRSE - signal 12 / weighted by atomic masses	-0.668
18	Mor19p	3D-MoRSE - signal 19 / weighted by atomic polarizabilities	0.958
19	MATS5v	Moran autocorrelation - lag 5 / weighted by atomic van der Waals volumes	0.705
20	Mor03m	3D-MoRSE - signal 03 / weighted by atomic masses	1.558
21	C-015	=CH2 atom-centred fragments	-0.663
22	Mor21v	3D-MoRSE - signal 21 / weighted by atomic van der Waals volumes	0.888
23	BIC1	bond information content (neighborhood symmetry of 1-order)	0.966
24	SIC1	structural information content (neighborhood symmetry of 1-order)	-0.491
25	ISH	standardized information content on the leverage equality	-0.224
26	BELe6	lowest eigenvalue n. 6 of Burden matrix / weighted by atomic Sanderson electronegativities	-2.148
27	EEig06r	Eigenvalue 06 from edge adj. matrix weighted by resonance integrals	1.883
28	EEig09x	Eigenvalue 09 from edge adj. matrix weighted by edge degrees	1.442
29	T(O..F)	sum of topological distances between O..F	-0.557
30	G2	gravitational index G2 (bond-restricted)	-1.267
31	MAXDN	maximal electrotopological negative variation	-0.436
32	nRCOOR	number of esters (aliphatic)	-0.774
33	RDF050e	Radial Distribution Function - 5.0 / weighted by atomic Sanderson electronegativities	1.265
34	BLTD48	Verhaar model of Daphnia base-line toxicity from MLOGP (mmol/l)	2.663
35	C-031	X--CR—X atom-centred fragments	0.909
36	SPI	superpendentic index	-1.408
37	H5p	H autocorrelation of lag 5 / weighted by atomic polarizabilities	-0.842
38	RDF050u	Radial Distribution Function - 5.0 / unweighted	1.305
39	RTu	R total index / unweighted	-1.799

40	nPyrazines	number of pyrazines	-0.132
41	G3e	component symmetry directional WHIM index / weighted by atomic Sanderson electronegativities	-1.483
42	Lop	Lopping centric index	0.613
43	nArNR2	number of tertiary amines (aromatic)	0.827
44	piPC08	molecular multiple path count of order 08	1.22
45	C-037	Ar-CH=X atom-centred fragments	0.674
46	G3v	3st component symmetry directional WHIM index / weighted by atomic van der Waals volumes	1.125
47	piPC07	molecular multiple path count of order 07	-1.328
48	BLI	Kier benzene-likeliness index	-0.285
49	EEig14r	Eigenvalue 14 from edge adj. matrix weighted by resonance integrals	0.724
50	X2	connectivity index chi-2	-1.211
51	HATS6u	leverage-weighted autocorrelation of lag 6 / unweighted	0.552
52	Depressant-50	Ghose-Viswanadhan-Wendoloski antidepressant at 50% (drug-like index)	-0.614
53	D/Dr08	distance/detour ring index of order 8	0.297
54	ATS6p	Broto-Moreau autocorrelation of a topological structure - lag 6 / weighted by atomic polarizabilities	-1.059
55	HATS8p	leverage-weighted autocorrelation of lag 8 / weighted by atomic polarizabilities	0.69
56	R7u	R autocorrelation of lag 7 / unweighted	-0.861
57	nPyrazoles	number of Pyrazoles	0.441
58	RTe+	R maximal index / weighted by atomic Sanderson electronegativities	-0.188
59	R6e+	R maximal autocorrelation of lag 6 / weighted by atomic Sanderson electronegativities	-0.279
60	ATS7m	Broto-Moreau autocorrelation of a topological structure - lag 7 / weighted by atomic masses	-0.373
61	C-022	#CR / R=C=R atom-centred fragments	0.224
62	H7v	H autocorrelation of lag 7 / weighted by atomic van der Waals volumes	0.274
63	ATS8v	Broto-Moreau autocorrelation of a topological structure - lag 8 / weighted by atomic van der Waals volumes	-0.252
64	X1Av	average valence connectivity index chi-1	-0.256
65	IAC	total information index of atomic composition	-0.475
66	IDET	total information content on the distance equality	0.835
67	ATS7p	Broto-Moreau autocorrelation of a topological structure - lag 7 / weighted by atomic polarizabilities	0.301
68	nArCOOR	number of esters (aromatic)	-0.069

Remaining plots of the double cross validation

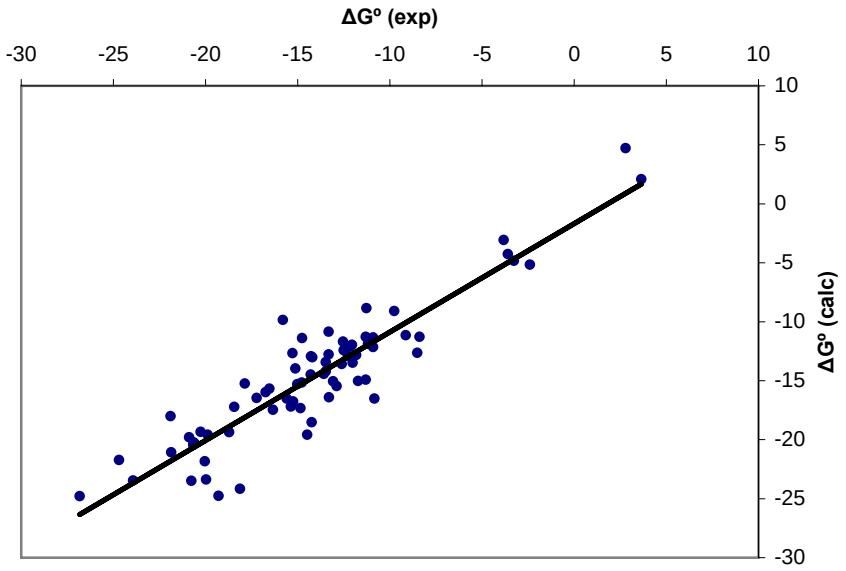

Figure 1 Prediction of ΔG° for subset S2 by means of the regression model for validation set V2. (r^2 = 0.84, RMSE = 2.32).

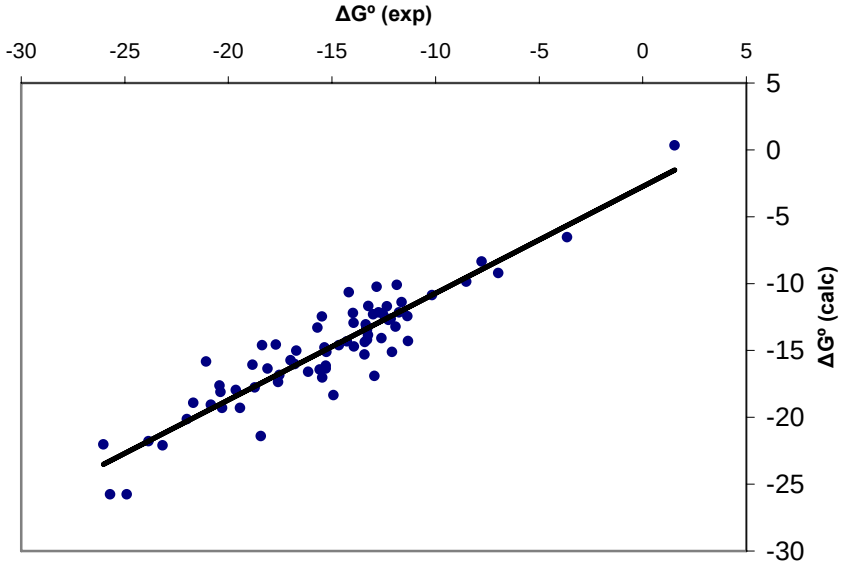

Figure 2 Prediction of ΔG° for subset S3 by means of the regression model for validation set V1 (r^2 = 0.85, RMSE = 1.88).