Supporting Information for "TMACC: Interpretable Correlation Descriptors for Quantitative Structure-Activity Relationships", James L. Melville and Jonathan D. Hirst.

LogS rules.

The training set supplied by Hou et al.¹ in their supporting information does not lead to the same atom type frequencies given in Table 1 of the main paper. Inclusion of the Yalkowsky test set of 21 molecules gives closer results, with only 7 atom types incorrectly assigned. Inspection of the SMARTS rules reveals that some atoms that are matches for pattern 36, are also matches for rule 39. By modifying rule 39, we were able to generate the correct frequency of atom types.

Old rule 39: [c;H0](~[!#6])(~c)~[a;!c]

New rule 39: [c;H0](~[!#6;!F;!Cl;!Br;!I])(~c)~[a;!c]

The numerical contribution of each rule was not modified.

1. Hou, T. J.; Xia, K.; Zhang, W.; Xu, X. J. ADME evaluation in drug discovery. 4. Prediction of aqueous solubility based on atom contribution approach. *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 266-275.

Training/Test set statistics for TMACC and HQSAR

dataset	TMACC		HQSAR		Ref 2
	q^2	R^2_{pred}	q^2	R^2_{pred}	R^2_{pred}
ace	0.73	0.56	0.67	0.45	0.30
ache	0.48	0.47	0.49	0.57	0.37
bzr	0.40	-0.22	0.47	-0.02	0.17
cox2	0.53	0.15	0.50	0.09	0.27
dhfr	0.54	0.53	0.69	0.71	0.63
gpb	0.49	0.61	0.62	0.54	0.58
therm	0.26	0.19	0.66	0.46	0.53
thr	0.65	0.48	0.68	0.19	-0.25

To enable comparison with those in Reference 2, the q^2 values use the standard definition of TSS. In Reference 2, the authors based their conclusion on R^2_{pred} results after removing outliers from the test set; we reproduce the figures obtained before any outlier removal, to preclude any subjectivity in defining outliers.

2. Sutherland, J. J.; O'Brien, L. A.; Weaver, D. F. A Comparison of Methods for Modeling Quantitative Structure-Activity Relationships. *J. Med. Chem.* **2004**, *47*, 5541-5554.