LETTER

Findings of the Challenge To Predict Aqueous Solubility

Anton J. Hopfinger,*,†,‡ Emilio Xavier Esposito,‡ A. Llinàs,||,§ R. C. Glen,§ and J. M. Goodman§

College of Pharmacy, University of New Mexico, MSC09 5360, Albuquerque, New Mexico 87131-0001, The Chem21 Group, Inc., 1780 Wilson Drive, Lake Forest, Illinois 60045, and Unilever Centre for Molecular Science Informatics, Department of Chemistry, University of Cambridge, Lensfield Road, Cambridge CB2 1EW, U.K.

Received December 1, 2008

Recently some of us made intrinsic aqueous solubility measurements for 132 structurally diverse drugs and biologically significant molecules. We then issued, in conjunction with this Journal, an open prediction challenge in a paper where we reported the intrinsic solubilities for 100 of these compounds as a training set and listed the structures of the remaining 32 compounds which formed the prediction set.[1] This solubility challenge data were also made available online including a link on this Journal's Web site. The formal *Solubility Challenge* was held from approximately July 15, 2008 through September 15, 2008 although late and/or changed entries were accepted until the end of September. More than 100 entries to the *Solubility Challenge* have been received. In several cases multiple entries came from the same person or group. In addition, more than 5% of the prediction sheet entries were incomplete in that predictions were not reported for all 32 compounds of the prediction set. These incomplete entries were not included in the overall findings given here, but the submitted prediction sheets, like those of all other entries, were scored and returned to the contestants along with a copy of the findings of this challenge. Overall, 99 completed entries were scored and are reported here.

Before presenting some of the findings from this challenge, it is important to reiterate what was stated at the outset of the challenge. The goal of this Solubility Challenge *is not to identify a 'winner' but rather to advance our general understanding of how to better perform aqueous solubility estimations. The findings of this challenge also provide us a perspective on the current state of predictive capabilities.*

The prediction set of compounds and their measured intrinsic solubilities are given in Table 1. Some qualifiers had to be applied to the prediction set to fairly and unambiguously score the contestants' data. Two of the prediction set compounds exhibit polymorph solubility behavior, and the solubility of each polymorph has been measured. The lowest solubility polymorph value was used in scoring the predictions for these two compounds. There are four prediction set compounds that have "**too soluble to measure**" as their reported solubility measures. That is, no numerical value of S [logS] is given. In order to include and exclude these four compounds in computing the percent correct predictions, the following protocol was adopted to assess each contestant's measure of prediction performance, MPP.

Table 1. Prediction Set Compounds and Their Measured S Values with Error Limits

compound	measured solubility (μg/mL)	compound	measured solubility (μg/mL)
1 acebutolol	711 ± 140	17 indomethacin	410 ± 5
2 amoxicillin	3900 ± 200	18 ketoprofen	157 ± 4
3 bendroflumethiazide	21.2 ± 2.6	19 lidocaine	3130 ± 100
4 benzocaine	780 ± 150	20 marbofloxacin	too soluble to measure
5 benzthiazide	6.4 ± 0.1	21 meclofenamic acid	0.16 ± 0.01
6 2-chloromandelic acid	too soluble to measure	22 naphthoic acid	28.96 ± 0.01
7 clozapine	188.9 ± 0.7	23 1R-2R-pseudoephedrine	too soluble to measure
8 dibucaine	14 ± 2	24 probenecid	3.9 ± 0.1
9 diethylstilbestrol	10 ± 3	25 pyrimethamine	19.4 ± 2.7
10 diflunisal	four forms: 26 ± 8, 7.6 ± 0.7, 0.93 ± 0.05, **0.29 ± 0.02**[a]	26 salicylic acid	1620 ± 40
11 dipyridamole	3.46 ± 0.05	27 sulfamerazine	200 ± 6
12 1R-2S-ephedrine	too soluble to measure	28 sulfamethizole	450 ± 40
13 folic acid	2.5 ± 0.2	29 terfenadine	0.00856
14 furosemide	19.6 ± 2.0	30 thiabendazole	66 ± 4
15 hydrochlorothiazide	625 ± 11	31 tolbutamide	93 ± 5
16 imipramine	22 ± 1	32 trazodone	two forms: 460 ± 50, **127 ± 10**[a]

[a] The solubility measure used in scoring contestant entries.

a) If for any [up to all four] of these compounds, the predicted S [logS] value is larger than all of the 28 predictions for which actual measured solubility values are reported, the contestant was given credit for a correct answer. This is the "full set" measure.

b) These four compounds are neglected in computing the percent correct predictions for the "28 measured" and the "24 4-outlier" MPP.

Lastly, the four "**too soluble to measure**" compounds were not used in computing the square of the correlation coefficient, R^2, of predicted versus measured solubility values. Thus, only compounds having measured numerical values for S [logS] have been employed in computing the R^2.

The distributions of the 28 measured prediction set compounds in S and logS space are shown in Figure 1. The pair of red lines in each plot defines the respective envelopes within which S and logS predictions are considered correct. The purpose for including these two plots is to illustrate the differences in how the prediction set compounds locate in their respective S and logS spaces. These differences may

* Corresponding author fax: 847-234-3992; e-mail: hopfingr@unm.edu.
† University of New Mexico.
‡ The Chem21 Group, Inc.
|| Current address: Physical Chemistry & DMPK Lead Generation, Astra Zeneca R&D, Mölndal, SE-431 83 Mölndal, Sweden.
§ University of Cambridge.

10.1021/ci800436c CCC: $40.75 © 2009 American Chemical Society
Published on Web 12/31/2008

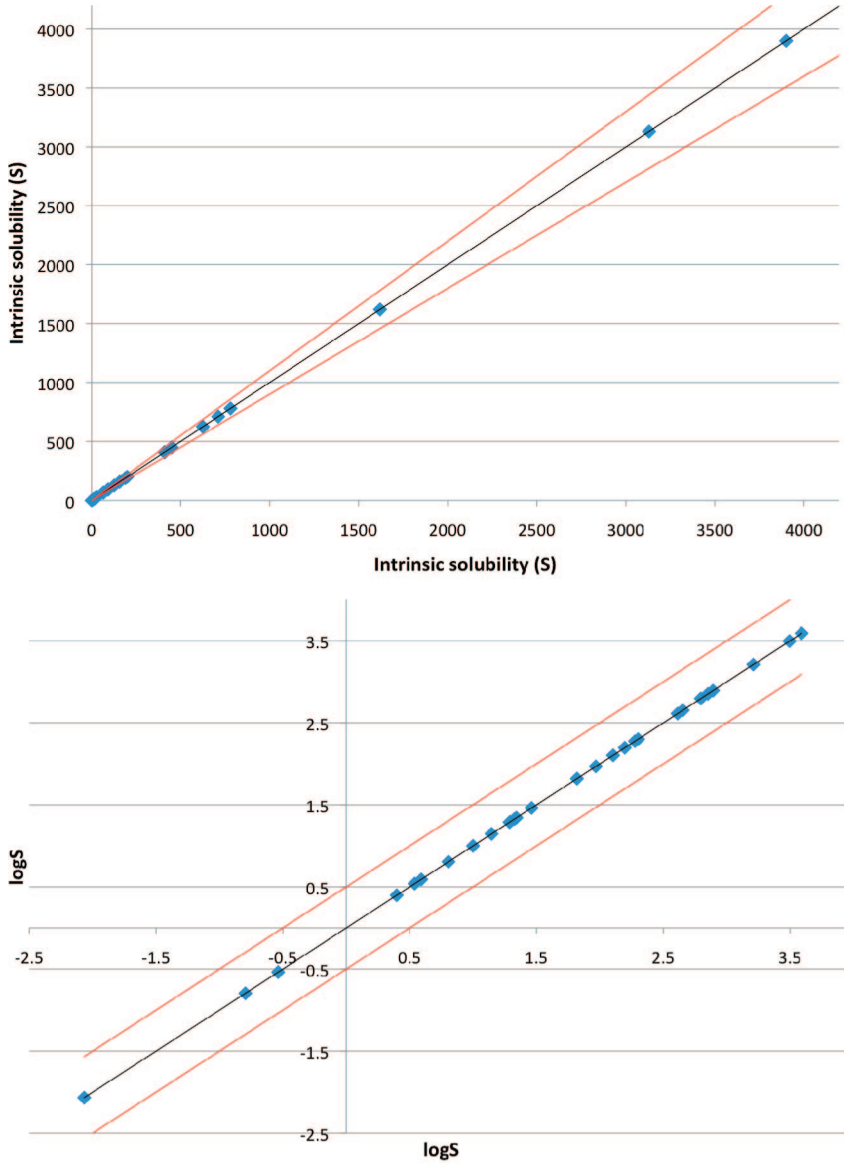

Figure 1. The S and logS distribution plots for the 28 compounds of the prediction set for which numerical measures are available. The pair of red lines of each plot defines the envelopes within which the predictions are considered correct.

influence *both* the estimation of the solubility measures with respect to methodology and the MPP values reported here. The measures of logS are more evenly distributed over their range than are the S values, and, overall, correct predictions of direct intrinsic solubilities, S, are seen as more difficult to achieve than logS.

The MPP of the 99 individual entries to the *Solubility Challenge* are given in Table 2. The MPP are first classified with respect to the actual intrinsic solubility, S, and with respect to logS. Next, the MPP are divided with respect to the full 32 compound prediction set, the 28 measured compounds in the prediction set, and also with respect to removal of the four largest predicted outliers [∼15% of the 28 compound predictions set] to yield a pruned 24 compound prediction set. Finally, the percent of correct predictions, using a ±10% error constraint for S and a ±0.5 log error limit for logS, are reported as relative MPP. The correlation coefficient, R^2, of the linear regression fits of the predicted versus measured S, and logS, values have been used as absolute MPP.

A histogram plot of the prediction set compounds ranked with respect to their measured solubility versus the percent of correct ±0.5logS predictions by the contestants has also been constructed and is shown in Figure 2. The X-axis lists the 32 test set compounds ordered from left [least soluble] to right [most soluble] with respect to solubility. The left Y-axis measures the percent of correct predictions of each compound, and the right Y-axis refers to logS. The red line with squares running across the plot contains the measured logS values of the prediction set.

Major findings from the *Solubility Challenge* include the following:

1. The ranges in the percent correction predictions for ±10% of the measured value of S are as follows: i) full set of 32 compounds [0.0% to 21.9%], ii) measured set of 28 compounds [0.0% to 17.9%], and iii) 4-outlier set of 24 compounds [0.0% to 20.8%].

2. The ranges in the predicted versus measured R^2 for S are as follows: i) measured set of 28 compounds [0.000 to 0.642] and ii) 4-outlier set of 24 compounds [0.002 to 0.987].

Table 2. Measures of Prediction Performance for All 99 Fully Completed Entries

	S					logS				
	full 32[a]	28 measured[b]	28	24 4-outliers[c]	24	full 32	28 measured	28	24 4-outliers	24
entry code	±10% percent correct	±10% percent correct	R^2	±10% percent correct	R^2	±0.5logS percent correct	±0.5logS percent correct	R^2	±0.5logS percent correct	R^2
6Y8B4	18.8	7.1	0.011	8.3	0.032	46.9	39.3	0.313	45.8	0.581
A69EW	12.5	3.6	0.580	4.2	0.721	46.9	42.9	0.583	50.0	0.671
ADIBE	21.9	17.9	0.005	20.8	0.134	34.4	32.1	0.174	37.5	0.470
AEJAK	9.4	7.1	0.502	8.3	0.846	46.9	50.0	0.562	58.3	0.797
AM108	9.4	0.0	0.188	0.0	0.572	46.9	42.9	0.620	50.0	0.793
ASTRA	0.0	0.0	0.020	0.0	0.322	28.1	32.1	0.298	37.5	0.630
AULT1	6.3	3.6	0.043	4.2	0.732	40.6	42.9	0.357	50.0	0.659
AULT2	15.6	7.1	0.009	8.3	0.647	53.1	50.0	0.361	58.3	0.669
AULT3	9.4	7.1	0.023	8.3	0.727	50.0	53.6	0.366	62.5	0.663
AZECA	6.3	0.0	0.221	0.0	0.109	56.3	57.1	0.291	66.7	0.565
AZECB	18.8	10.7	0.273	12.5	0.279	53.1	50.0	0.290	58.3	0.499
AZECC	15.6	7.1	0.389	8.3	0.603	53.1	50.0	0.548	58.3	0.605
AZECd	15.6	10.7	0.096	12.5	0.058	40.6	39.3	0.221	45.8	0.592
BEREL	6.3	3.6	0.046	4.2	0.786	43.8	46.4	0.305	54.2	0.648
BHNMY	3.1	0.0	0.113	0.0	0.019	21.9	21.4	0.144	25.0	0.459
CAODS	15.6	10.7	0.070	12.5	0.880	50.0	50.0	0.234	58.3	0.611
CRSCF	9.4	3.6	0.360	4.2	0.699	34.4	32.1	0.509	37.5	0.777
CRSFU	6.3	0.0	0.389	0.0	0.011	31.3	28.6	0.465	33.3	0.622
DBOPK	6.3	0.0	0.089	0.0	0.527	34.4	32.1	0.444	37.5	0.799
DCG82	9.4	3.6	0.194	4.2	0.142	28.1	25.0	0.357	29.2	0.692
DDDC1	12.5	10.7	0.022	12.5	0.720	46.9	50.0	0.365	58.3	0.723
FRJTU	6.3	0.0	0.042	0.0	0.582	28.1	25.0	0.504	29.2	0.780
GPAPO	6.3	3.6	0.016	4.2	0.247	37.5	39.3	0.366	45.8	0.603
GRAP1	3.1	0.0	0.018	0.0	0.283	34.4	35.7	0.372	41.7	0.609
HELIX	3.1	0.0	0.022	0.0	0.851	28.1	28.6	0.452	33.3	0.590
HIRST	6.3	0.0	0.625	0.0	0.941	37.5	35.7	0.269	41.7	0.621
I-GPS	12.5	7.1	0.049	8.3	0.472	37.5	35.7	0.326	41.7	0.731
I-LIB	12.5	7.1	0.049	8.3	0.467	37.5	35.7	0.320	41.7	0.717
I-OGP	9.4	3.6	0.523	4.2	0.488	43.8	42.9	0.360	50.0	0.617
I-OGS	6.3	3.6	0.190	4.2	0.881	34.4	35.7	0.439	41.7	0.686
ICAM1	6.3	3.6	0.033	4.2	0.074	28.1	28.6	0.066	33.3	0.191
ICAM2	9.4	0.0	0.033	0.0	0.336	46.9	42.9	0.243	50.0	0.575
ICAM3	9.4	0.0	0.030	0.0	0.422	31.3	25.0	0.057	29.2	0.040
ICAM4	12.5	3.6	0.020	4.2	0.002	34.4	28.6	0.170	33.3	0.434
ICAM5	3.1	0.0	0.397	0.0	0.904	37.5	39.3	0.080	45.8	0.210
JDSYN	6.3	0.0	0.397	0.0	0.312	40.6	39.3	0.503	45.8	0.763
JDW01	12.5	3.6	0.164	4.2	0.776	53.1	50.0	0.362	58.3	0.655
JWANG	6.3	3.6	0.008	4.2	0.467	46.9	50.0	0.193	58.3	0.651
KATYA	6.3	3.6	0.000	4.2	0.365	40.6	42.9	0.274	50.0	0.466
KBSZT	6.3	3.6	0.032	4.2	0.013	28.1	28.6	0.153	33.3	0.475
KHQGF	9.4	3.6	0.002	4.2	0.423	34.4	32.1	0.409	37.5	0.771
KMSDA	3.1	0.0	0.121	0.0	0.448	43.8	46.4	0.412	54.2	0.688
LFMSC	6.3	3.6	0.041	4.2	0.054	37.5	39.3	0.040	45.8	0.439
LGGAV	12.5	7.1	0.568	8.3	0.616	53.1	53.6	0.535	62.5	0.759
LJMU1	12.5	7.1	0.002	8.3	0.322	40.6	39.3	0.181	45.8	0.516
LJMU2	9.4	0.0	0.068	0.0	0.649	56.3	53.6	0.526	62.5	0.762
LJMU3	9.4	3.6	0.029	4.2	0.025	18.8	14.3	0.023	16.7	0.019
MAHEN	9.4	3.6	0.005	4.2	0.722	37.5	35.7	0.143	41.7	0.383
MJEMK	18.8	10.7	0.152	12.5	0.122	56.3	53.6	0.299	62.5	0.445
MKETT	3.1	0.0	0.074	0.0	0.802	28.1	28.6	0.411	33.3	0.598
MKMJE	9.4	0.0	0.309	0.0	0.455	62.5	60.7	0.471	70.8	0.835
MODE9	9.4	3.6	0.455	4.2	0.530	43.8	42.9	0.154	50.0	0.454
MOEXY	6.3	0.0	0.142	0.0	0.455	46.9	46.4	0.316	54.2	0.614
MRENS	6.3	3.6	0.003	4.2	0.439	37.5	39.3	0.317	45.8	0.740
MRSOD	12.5	10.7	0.003	12.5	0.022	37.5	39.3	0.168	45.8	0.315
MSUHB	6.3	0.0	0.631	0.0	0.575	43.8	42.9	0.608	50.0	0.811
NIRAZ	9.4	7.1	0.042	8.3	0.031	28.1	28.6	0.394	33.3	0.688
NLDTH	12.5	3.6	0.347	4.2	0.459	46.9	42.9	0.403	50.0	0.625
NORMA	6.3	0.0	0.007	0.0	0.010	15.6	10.7	0.376	12.5	0.531
NORMB	12.5	7.1	0.025	8.3	0.784	46.9	46.4	0.373	54.2	0.735
NSLIC	9.4	3.6	0.490	4.2	0.684	50.0	50.0	0.541	58.3	0.709
OLASM	9.4	3.6	0.552	4.2	0.648	46.9	46.4	0.558	54.2	0.766
P4G4W	12.5	10.7	0.072	12.5	0.058	43.8	46.4	0.439	54.2	0.579
PFRGB	9.4	0.0	0.089	0.0	0.673	50.0	46.4	0.312	54.2	0.548
PLIND	6.3	0.0	0.095	0.0	0.275	37.5	35.7	0.109	41.7	0.342
PLJTU	6.3	3.6	0.376	4.2	0.873	37.5	39.3	0.433	45.8	0.500
PVUHD	6.3	0.0	0.105	0.0	0.532	31.3	28.6	0.478	33.3	0.819
ROBOT	6.3	7.1	0.026	8.3	0.739	31.3	35.7	0.093	41.7	0.192
RONLY	9.4	3.6	0.008	4.2	0.400	37.5	35.7	0.250	41.7	0.432
RTRMT	3.1	0.0	0.014	0.0	0.000	15.6	14.3	0.018	16.7	0.002
SAVLF	12.5	3.6	0.125	4.2	0.792	46.9	42.9	0.352	50.0	0.691
SFBDD	3.1	0.0	0.009	0.0	0.290	31.3	32.1	0.283	37.5	0.651
SHARK	9.4	3.6	0.572	4.2	0.722	46.9	46.4	0.591	54.2	0.707
SHYMW	9.4	3.6	0.319	4.2	0.914	46.9	46.4	0.335	54.2	0.546

Table 2. Continued

	S					logS				
	full 32[a]	28 measured[b]	28	24 4-outliers[c]	24	full 32	28 measured	28	24 4-outliers	24
SIMPL	3.1	0.0	0.075	0.0	0.494	31.3	32.1	0.261	37.5	0.235
SJEPC	21.9	14.3	0.122	16.7	0.722	62.5	60.7	0.434	70.8	0.765
SJEPN	18.8	10.7	0.091	12.5	0.931	56.3	53.6	0.437	62.5	0.768
SJEPP	9.4	3.6	0.084	4.2	0.121	43.8	42.9	0.303	50.0	0.457
SJNUC	18.8	10.7	0.161	12.5	0.675	56.3	53.6	0.365	62.5	0.766
SJNUN	9.4	0.0	0.133	0.0	0.961	50.0	46.4	0.368	54.2	0.769
SJNUP	6.3	0.0	0.071	0.0	0.747	43.8	42.9	0.278	50.0	0.451
SOMAD	3.1	0.0	0.002	0.0	0.224	25.0	25.0	0.143	29.2	0.532
STAND	0.0	0.0	0.237	0.0	0.678	31.3	35.7	0.301	41.7	0.593
SUNAL	15.6	3.6	0.228	4.2	0.765	56.3	50.0	0.517	58.3	0.711
SUNHM	21.9	10.7	0.029	12.5	0.483	50.0	42.9	0.427	50.0	0.683
SUNNC	12.5	0.0	0.019	0.0	0.826	50.0	42.9	0.433	50.0	0.687
SYPJA	9.4	0.0	0.000	0.0	0.429	34.4	28.6	0.408	33.3	0.648
SYPJS	6.3	0.0	0.000	0.0	0.007	31.3	28.6	0.290	33.3	0.502
TIWA2	6.3	0.0	0.004	0.0	0.130	28.1	25.0	0.269	29.2	0.587
TOLIK	6.3	3.6	0.007	4.2	0.484	46.9	50.0	0.278	58.3	0.480
TROP_A	9.4	0.0	0.119	0.0	0.800	43.8	39.3	0.434	45.8	0.503
TROP_B	15.6	7.1	0.129	8.3	0.443	50.0	46.4	0.480	54.2	0.644
TROP_C	6.3	0.0	0.144	0.0	0.604	43.8	42.9	0.506	50.0	0.647
TSTFY	9.4	3.6	0.081	4.2	0.673	43.8	42.9	0.479	50.0	0.798
UIQBB	9.4	0.0	0.642	0.0	0.907	53.1	50.0	0.650	58.3	0.803
UTITE	9.4	3.6	0.082	4.2	0.371	43.8	42.9	0.254	50.0	0.559
UX100	12.5	7.1	0.123	8.3	0.778	34.4	32.1	0.346	37.5	0.640
Y1008	15.6	7.1	0.389	8.3	0.603	53.1	50.0	0.548	58.3	0.605
YINAN	9.4	3.6	0.533	4.2	0.987	53.1	53.6	0.552	62.5	0.750

[a] All 32 compounds of the prediction set used. [b] The four compounds being 'too soluble to measure' were NOT considered leading to 8 compounds being used. [c] In addition to not considering the four compounds 'too soluble to measure', the four largest outliers of each individual prediction set were also NOT considered thus leading to 24 compounds being used.

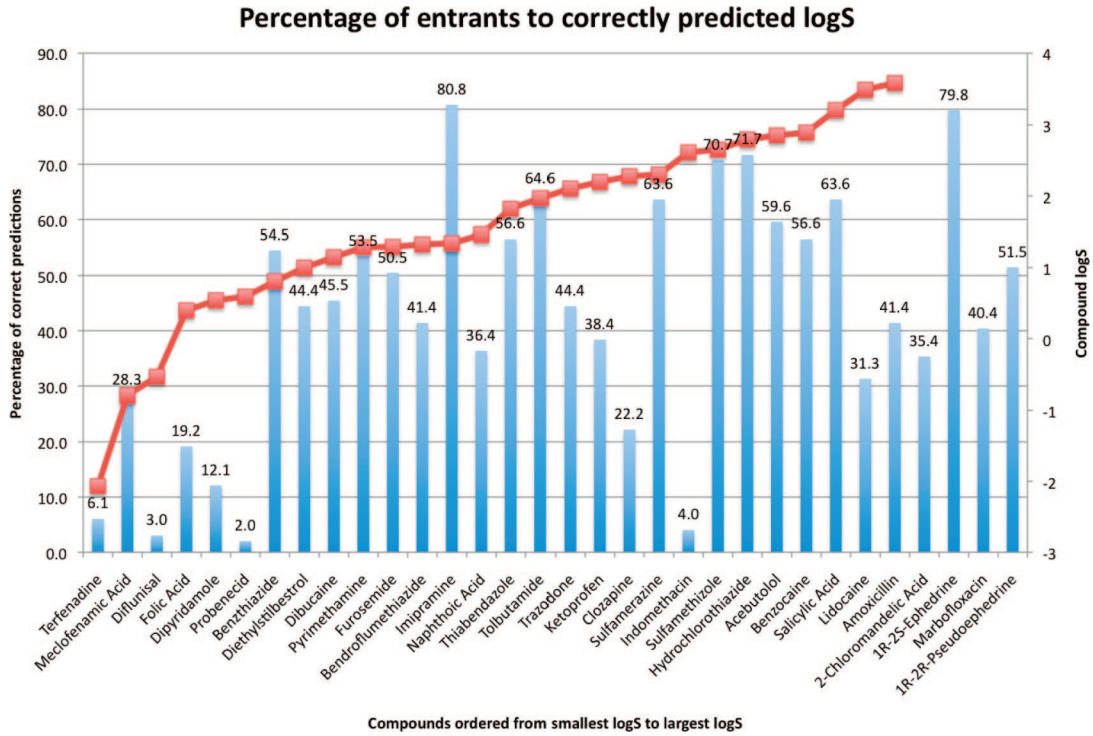

Figure 2. The histogram plot of the prediction set compounds ranked with respect to their measured solubilities, given by the red curve, versus the percent of correct predictions, where correct is defined as being less than ±0.5logS from the measured value. LogS values are not given for the final four compounds on the right as they are too soluble to measure. The definition for a correct prediction of a too soluble to measure compound is given in the text.

3. The ranges in the percent correction predictions for ±0.5logS of the measured value of logS are as follows: i) full set of 32 compounds [15.6% to 62.5%], ii) measured set of 28 compounds [10.7% to 60.7%], and iii) 4-outlier set of 24 compounds [12.5% to 70.8%].

4. The ranges in the predicted versus measured R^2 for logS are as follows: i) measured set of 28 compounds [0.018 to 0.650] and ii) 4-outlier set of 24 compounds [0.002 to 0.835].

5. No contestant suggested that any compound might be a polymorph. Correspondingly, no contestant made any prediction of solubility as a function of polymorphic state.

6. The majority of the compounds in the test set have solubilities in the range of logS = [0.5 to 3]. Within this

range, the likelihood of an accurate prediction is higher than for both lower and higher ranges in solubility.

7. Prediction accuracy depends on chemical structure as can be inferred from comparison of imipramine [S = 22.1] which is correctly predicted by 81% of the contestants to naphthoic acid [S = 28.9] which is correctly predicted by less than 37% of the contestants.

8. The two least well predicted compounds having measured logS values in the range of [0.5 to 3] are probenecid [2% correct predictions] and indomethacin [4% correct predictions].

Challenge contestants employed the entire spectrum of current approaches to predict aqueous solubility and other physicochemical end points, both with respect to statistical analyses and computational chemistry. The challenge data cover a wide range in measured solubility and structural diversity. Hence, the findings of the *Solubility Challenge* given above provide us an overall perspective as to the current ability to estimate aqueous solubility. It would be nice to have been able to say that an analysis of the set of contestant entries identifies definitive methods that are best to use in predicting aqueous solubility. However, this was not possible; rather, a variety of methods and combinations of methods all perform about equally well.

Finally, open and objective challenges to predict important physicochemical and biological properties may increasingly become an important way to evaluate the computational chemistry capabilities and measure progress in the field. However, in carrying out this solubility challenge it became clear to us that this is a time-consuming activity that is dependent upon the data used and the manner in which results are scored. Consequently, going forward considerable thought needs to be given regarding who should do challenges, how experimental data are selected and measured, how scoring and evaluation are carried out, and what specific properties should be explored.

REFERENCES AND NOTES

(1) Llinàs, A.; Glen, R. C.; Goodman, J. M. Solubility Challenge: Can You Predict Solubilities of 32 Molecules Using a Database of 100 Measurements? *J. Chem. Inf. Model.* **2008**, *48*, 1289–1303.

CI800436C