Prediction of the *n*-Octanol/Water Partition Coefficients in the SAMPL6 Blind Challenge from MST Continuum Solvation Calculations

William J. Zamora[1,2], Silvana Pinheiro[2], Kilian German[3], Clara Ràfols[3], Carles Curutchet[2] and F. Javier Luque[1]

[1] Department of Nutrition, Food Science and Gastronomy, Faculty of Pharmacy and Food Sciences, Institute of Biomedicine (IBUB) and Institute of Theoretical and Computational Chemistry (IQTCUB), Campus Torribera, University of Barcelona, 08921 Santa Coloma de Gramenet, Spain.

[2] Department of Pharmacy and Pharmaceutical Technology and Physical Chemistry and Institute of Theoretical and Computational Chemistry (IQTCUB), Faculty of Pharmacy and Food Sciences, University of Barcelona, Av. Joan XXIII 27-31, 08028 Barcelona, Spain.

[3] Department of Chemical Engineering and Analytical Chemistry and Institute of Biomedicine (IBUB), Universitat de Barcelona, Martí i Franquès 1-11, 08028 Barcelona, Spain

Abstract

The IEFPCM/MST continuum solvation model is used for the blind prediction of *n*-octanol/water partition of a set of 11 fragment-like small molecules within the SAMPL6 Part II Partition Coefficient Challenge. The partition coefficient of the neutral species (log *P*) was determined using an extended parametrization of the B3LYP/6-31G(d) version of the Miertus-Scrocco-Tomasi continuum solvation model in *n*-octanol. Comparison with the experimental data provided for partition coefficients yielded a root-mean square error (rmse) of 0.78 (log *P* units), which agrees with the accuracy reported for our method (rmse = 0.80) for nitrogen-containing heterocyclic compounds. Out of the 91 sets of log *P* values submitted by the participants, our submission is within those with an rmse < 1 and among the four best ranked physical methods. The largest errors involve three compounds: two with the largest positive deviations (SM13 and SM08), and one with the largest negative deviations (SM15). Here we report the potentiometric determination of the log *P* for SM13, leading to a value of 3.62 ± 0.02, which is in better agreement with most empirical predictions than the experimental value reported in SAMPL6. In addition, further inclusion of several conformations for SM08 significantly improved our results. Inclusion of these refinements led to an overall error of 0.51 (log *P* units), which supports the reliability of the IEFPCM/MST model for predicting the partitioning of neutral compounds.

Introduction

Lipophilicity is a crucial physicochemical property to understand the biological and pharmaceutical properties of drugs, as it reflects the differential solubility of solutes in aqueous and organic environments [1-4]. Generally, the lipophilicity is estimated from the partitioning of compounds between aqueous and *n*-octanol phases [5-7], which have been also used as reference systems in the development of a variety of computational approaches [8-12].

The logarithm of the partition coefficient of a neutral solute (log *P*) can be determined by combining the transfer free energy of the solute between water and *n*-octanol ($\Delta G^{o/w}$), which in turn is related to the difference in the solvation free energy upon transfer from the gas phase to the two solvents (ΔG_{sol}^{w} and ΔG_{sol}^{o}; Scheme 1). Therefore, a robust prediction of the partition coefficient for (bio)organic compounds depends on the accuracy of the calculated solvation free energies. In this context, quantum mechanical (QM) self-consistent reaction field models (SCRF) have demonstrated to be a powerful approach for the calculation of the solvation free energy, especially when one considers the tradeoff between chemical accuracy and computational cost [13-16]. Thus, after a careful parametrization of QM-SCRF methods, the solvation free energy of neutral compounds can generally be predicted with an uncertainty less than 1 kcal/mol [17, 18].

In this work we report the results obtained for the SAMPL6 Part II Partition Coefficient Challenge using the B3LYP/6-31G(d) version of the integral IEFPCM/MST solvation model [19], which relies on the Integral Equation Formalism of the Polarizable Continuum model [20, 21]. The reliability of the IEFPCM/MST model for predicting hydration and tautomerism free energy changes was already checked in the first two editions of the SAMPL blind test [22, 23], yielding

predictions of hydration free energies with a root-mean square error (rmse) of 2.3 kcal/mol compared to the experimental values. Recently a series of refinements have been introduced in order to improve the accuracy for predicting the solvation free energy of nitrogen-containing heterocyclic compounds [24]. In this context, the current edition of the SAMPL challenge is thus an excellent opportunity to reassess the performance of the MST method.

Scheme 1. Thermodynamic Cycle Used to Determine the Transfer Free Energy of a Compound (X) between Two Immiscible Solvents.

$$\log P = \frac{-\Delta G^{o/w}}{2.303RT} = \frac{\Delta G^w_{sol} - \Delta G^o_{sol}}{2.303RT}$$
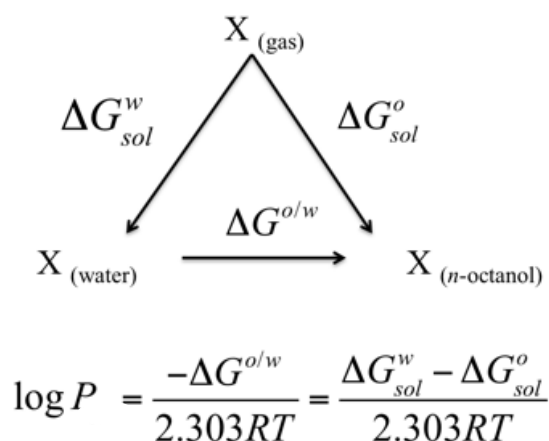
In the following a critical assessment of the results obtained from IEFPCM/MST continuum solvation calculations for the set of molecules proposed in the SAMPL6 challenge is made. The reader is addressed to ref. [25] for a comprehensive analysis of the performance of the different methods used in the blind challenge. The IEFPCM/MST results can be found under the identifier "kivfu" in the web page of the SAMPL6 challenge [26]. The overall performance of the MST model is discussed, with especial attention to the compounds that exhibit the largest deviations between experimental and calculated log P values. A critical review of the results estimated for

these compounds, including both the extension of the calculations to new chemical species and the potentiometric determination of the log P of compound SM13, leads to an improvement in the agreement between MST results and experimental data, as will be detailed below.

Methods

The IEF-MST model

Since the IEFPCM/MST model has been widely described in the literature, the reader is addressed to Refs. [14, 27] for details about the formalism of this continuum solvation model. Here, we limit ourselves to provide a succint description of the MST method. The IEFPCM/MST model computes the ΔG_{sol} from the addition of electrostatic (ΔG_{ele}), cavitation (ΔG_{cav}) and van der Waals (ΔG_{vW}) components, which are calculated using a double molecular-shaped cavity for the solute embedded in the polarizable continuum medium.

The ΔG_{ele} term measures the work needed to build up the solute charge distribution in the solvent. To this end, a solvent-excluded surface is obtained by scaling the atomic radii by a solvent-dependent factor (λ), which varies from 1.25 for the solvation in water to 1.80 for the solvation in carbon tetrachloride. The value of this scaling factor for solvation in n-octanol was adjusted to be 1.50 [28]. In contrast, no scaling is used for the non-electrostatic terms (ΔG_{cav}, ΔG_{vW}), which are evaluated using a van der Waals surface.

The ΔG_{cav} contribution is determined following Claverie-Pierotti's scaled particle theory (Eq. 1) [29, 30].

$$\Delta G_{cav} = \sum_{i=1}^{N} w_i \ \Delta G_{cav,i}^{P} \tag{1}$$

where N is the number of atoms, w_i is a weighting factor determined from the ratio between the solvent-exposed surface of atom i (S_i) and the total surface of such an atom, and $\Delta G_{cav,i}^{P}$ is the cavitation free energy of the sphere associated to atom i.

Finally, the ΔG_{vW} term is computed using a linear relationship to the solvent-exposed surface of each atom (Eq. 2).

$$\Delta G_{vW} = \sum_{i=1}^{N} \xi_i \ S_i \tag{2}$$

where ξ_i denotes the atomic surface tension of atom i, which is determined by fitting to the experimental free energies of solvation in the parametrization procedure.

Dataset

The SAMPL6 dataset consists of 11 fragment-like small molecules (see Fig. 1) endowed with kinase inhibitory activity. The dataset includes nine basic compounds, an acidic molecule, and a zwitterionic one, with experimental log P values in the range of 1.95–4.09 [31].
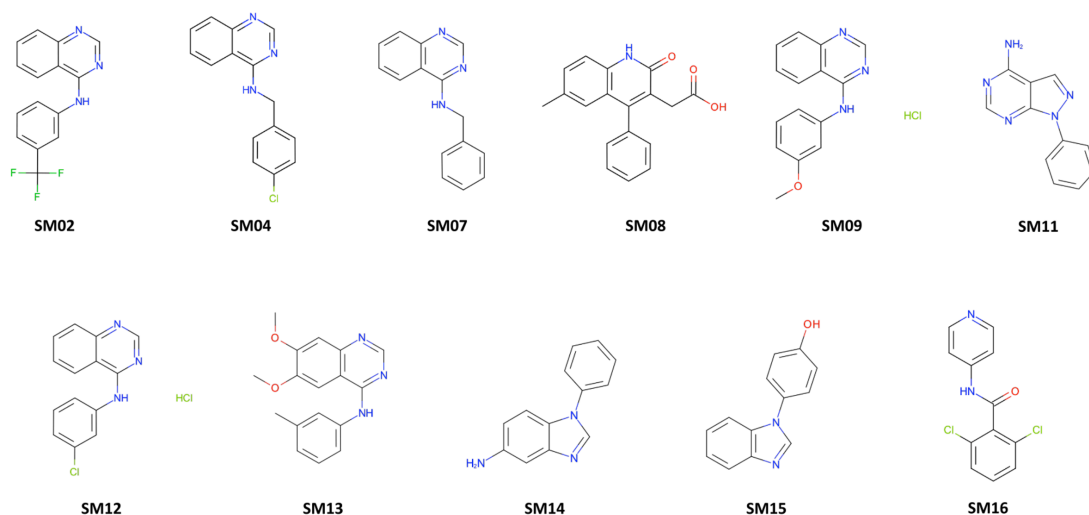

Figure 1. Dataset of 11 fragment-like small molecules in the SAMPL6 log P challenge.

Computational details

The structures of the 11 molecules were generated from the SMILES codes given in the SAMPL6 log P webpage [32] using the online SMILES translator and structure file generator of the National Cancer Institute [33]. Starting from these structures, the molecular geometries of the compounds were fully optimized at the B3LYP/6-31G(d) level of theory, without exploration of additional conformations. The solvation effect of water and *n*-octanol on the geometrical parameters of solutes was taken into account in geometry optimizations, which were performed using the IEFPCM version of the MST model. Single-point calculations in the gas phase and in solution were performed for the optimized geometries of the compounds to estimate the ΔG_{sol} in the two solvents. All calculations were performed using Gaussian 09 [34].

For one of our outliers, SM08, a more comprehensive conformational analysis was performed using FRee Online druG conformation generation (Frog, version 2) [35], including both keto and enol tautomeric forms (see Fig. S1).

For all conformers, the geometries were fully optimized at the B3LYP/6-31G(d) level, and the resulting minima were verified by vibrational frequency analysis, which gave positive frequencies in all cases. Then, the relative energies of the whole set of conformational species of both tautomers were refined from single-point computations performed at the MP2/aug-cc-pVDZ and MP2/aug-cc-pVTZ levels of theory, which were used to estimate the energy difference by extrapolation to the complete basis set (CBS) limit. Since the spin-component-scaled version of MP2 (SCS-MP2) provides a significant improvement in ground state energies by scaling parallel and antiparallel-spin pair correlation energies [36], the CBS energy was determined upon extrapolation of the SCS-MP2 correlation energies computed using

Dunning's aug-cc-pVDZ and aug-cc-pVTZ basis sets according to the formalisms proposed by Halkier [37] and Truhlar [38]. Finally, higher-order electron correlation effects were estimated from the difference between CCSD and SCS-MP2 energies calculated with the 6-31G(d) basis set.

The best gas phase estimate of the free energy difference was derived by combining the SCS-MP2/CBS results with the CCSD higher-order electron correlation correction, and the thermal contribution obtained in the B3LYP vibrational analysis (SCS-MP2/CBS+[CCSD-MP2/6-31G(d)]+G_{corr}; Eq 3).

$$G_{gas} = E_{SCS-MP2/CBS} + \Delta E_{CCSD} + G_{corr} \tag{3}$$

The gas-phase free energies were then combined with solvation free energies in both water and *n*-octanol (at 298 K) computed using the B3LYP/6-31G(d) version of the IEFPCM/MST model. The partition coefficient was determined using a Boltzmann's weighting scheme to the relative stabilities of the conformational species determined for the tautomers of this compound in the two solvents.

Determination of partition coefficient (log P) for SM13 by potentiometry

The compound 6,7-dimethoxy-N-(3-methylphenyl)quinazolin-4-amine (SM13) was purchased from MolPort with a purity grade of ~ 97.5% [31]. The log *P* values were obtained from the difference between the aqueous pK_a of the species and the apparent pK_a determined from dual-phase titrations (*n*-octanol/KCl 0.15M) using PCA200 from Sirius Instruments (UK). The experimental aqueous pK_a (5.77 ± 0.01) was taken from the value reported in the SAMPL6 part 1 [39].

Typically, ~3 mg of the samples were dissolved in the appropriate volume ratio of n-octanol (saturated with 0.15 M KCl aqueous solution) and 0.15 M KCl aqueous solution (saturated with n-octanol), followed by a preacidification of the sample with HCl 0.5 M and subsequent titration with KOH 0.5 M. Due to the low solubility of the sample, it was necessary to increase the temperature to approximately 30 ºC for obtaining a totally homogeneous sample. Several phase ratios, R, (between 0.03 and 0.4 v_o/v_w) were selected according the expected solubility and hydrophobicity of the sample [40, 41]. The log P values were estimated and refined by a weighted non-linear least-squares fit, where the aqueous pK_a values were used as unrefined contributions (see Fig. S4).

Results and Discussion

The partition coefficients determined from IEFPCM/MST B3LYP/6-31G(d) calculations in water and n-octanol are given in Table 1. On average, the IEFPCM/MST results deviate from the experimental values by a root-mean square error (rmse) of 0.78 log units, which places the IEFPCM/MST results among the most accurate values provided by physical methods, keeping in mind the small differences observed between methods with rmse < 1 (see Fig. S2). Thus, the rmse of the other best ranked QM-based solvation models, the *Cosmotherm* version of COSMO-RS [16, 42] and the Minnesota's solvation models SMD [43], SM8 [44] and SM12 [45], are in a narrow range comprised between 0.38 and 0.65 log units, whereas the use of 3D integral equation theory with a cluster embedding approach [46] yields a rmse of 0.47. The expected accuracy of the IEFPCM/MST for predicting log P values of

nitrogen-containing heterocyclic compounds was estimated to be 0.80 log units, after refinement of the atomic surface tension for nitrogen atoms (ξ_N and ξ_{NH}) in heterocyclic aromatic compounds in the calculation of the van der Waals component of solvation free energy in n-octanol [24]. Thus, present results support the suitability of the latest refinements introduced in the IEFPCM/MST model.

Table 1. Calculated (submission ID "kivfu") and experimental n-octanol/water partition coefficient (log P) determined for the set of compounds included in the SAMPL6 dataset.

Compound	Calculated	Experimental[a]	Δlog P (calc - exptl)
SM02	4.18	4.09 ± 0.03	0.09
SM04	3.45	3.98 ± 0.03	-0.53
SM07	2.78	3.21 ± 0.04	-0.43
SM08	4.69	3.10 ± 0.03	**1.59**
SM09	3.21	3.03 ± 0.07	0.18
SM11	1.71	2.10 ± 0.04	-0.39
SM12	3.70	3.83 ± 0.03	-0.13
SM13	4.28	2.92 ± 0.04	**1.36**
SM14	1.77	1.95 ± 0.03	-0.18
SM15	1.81	3.07 ± 0.03	**-1.26**
SM16	2.67	2.62 ± 0.01	0.05
mse[b]	-0.03		
mue[b]	0.56		
rsme[b]	0.78		

[a]Ref. [31] [b]Mean signed error (mse), mean unsigned error (mue), and root-mean square error (rmse) calculated relative to the experimental values are given in log P units.

The largest errors in the IEFPCM/MST results involve three compounds: the overestimation of the log P of compounds SM08 and SM13, and the underestimation of the log P of SM15 (see Fig. 2 and Fig. S3). Upon exclusion of these compounds, the statistical error is significantly reduced (rmse of 0.30 log units),and the correlation between calculated and experimental values improves from 0.64 to 0.95. This encouraged us to examine the potential factors that may underlie the deviation of the predicted values for these compounds.

Figure 2. Comparison between experimental and IEFPCM/MST n-octanol/water log P for the SAMPL6 Dataset. Red points represent the compounds with the largest errors in the original submission. Statistical analyses are shown for (top left) all compounds and (bottom right) after exclusion of SM08, SM13 and SM15.
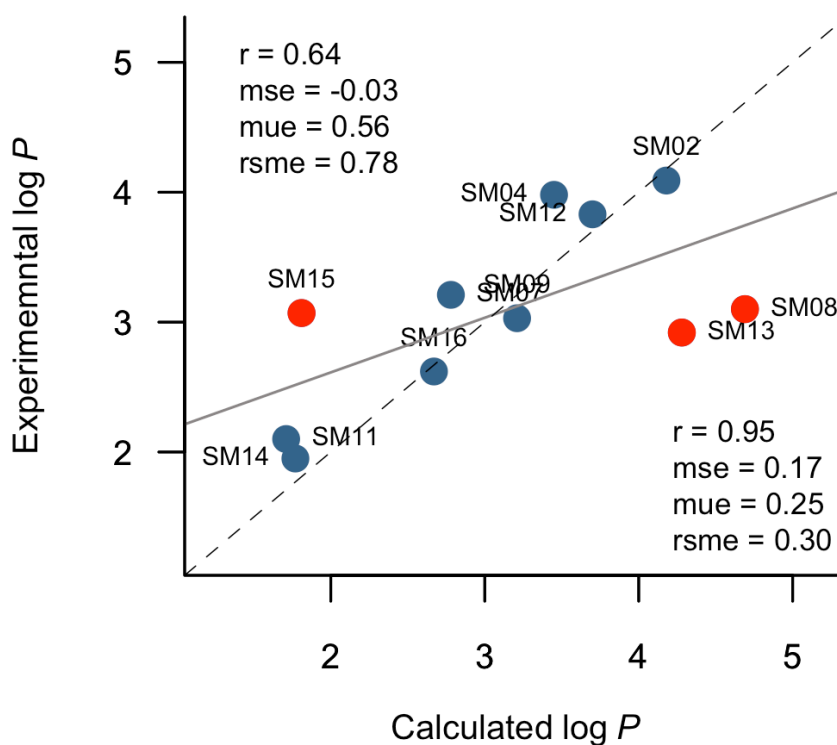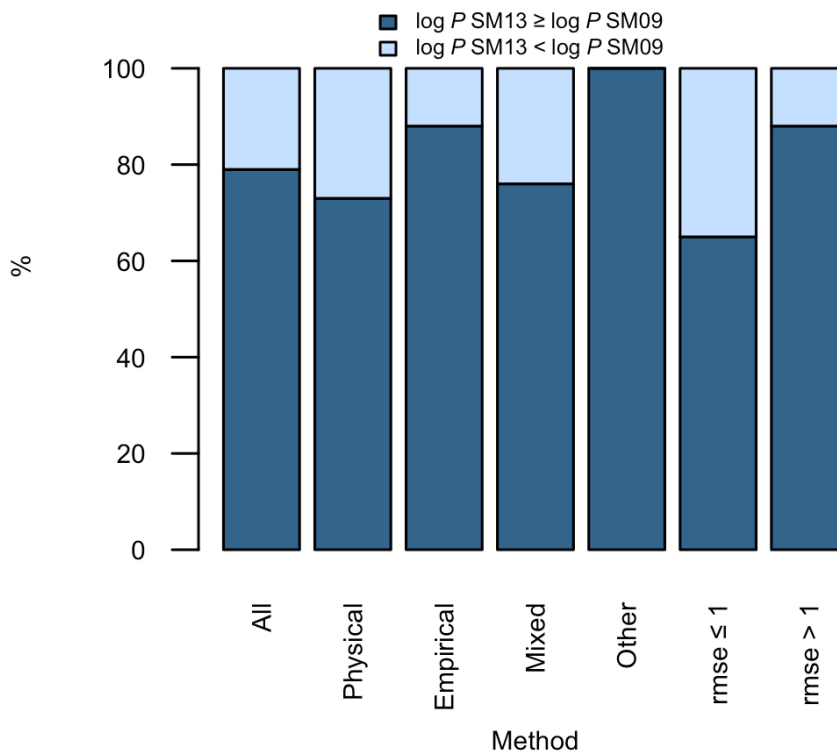
Compound SM13

Both SM13 and SM09 share the 4-aminoquinazoline scaffold (Fig. 1), but the Δlog P error increases from 0.18 for SM09 to 1.36 for SM13 (Table 1). Compared to SM09, the additional methoxy and methyl substituents present in SM13 suggest that this compound would be more lipophilic than SM09, a trend not reflected in the experimental log P values of 2.92 and 3.03 for SM13 and SM09, respectively. In fact, comparison of the log P values calculated from the different computational approaches that participated in the challenge reveals that the log P of SM13 is in most cases predicted to be larger compared to the log P of SM09 (Fig. 3). Furthermore, a survey of the main outliers in the best-ranked physical methods and in the most accurate submissions in other categories points out that SM13 is an outlier in the most succesful methods (Table 2).

On the basis of these considerations, we performed a potentiometric determination to determine the log P of compound SM13, obtaining a value of 3.62 ± 0.02 (see Fig. S4), which seems in better agreement with chemical intuition. After insightful discussion with the organizers of the challenge, we think that the main difference between the two experimental measurements may arise from the sensitivity of the profiles of volumetric ratios between partition solvents, expressed as log R, to both the partition of neutral and ionic species of the compound, and the low solubility of SM13. In the original measurement [31], the log R range was comprised between -1.20 and -0.18 but with a maximum analysis vial volume of 3 mL. In our case, we have used a similar log R range (-1.48 to -0.40). However, the volumes of partition solvents were larger than those used in the original measurement, as our volume limit was 25 mL. We considered that these experimental conditions allowed both a better solubility of SM13 and partition of ionic species. According to the experimental data

given for SM13 by the organizers of SAMPL6 [31], the partition of the ionic species was negligible in most replicas (-4.35, -7.57 and 0.09). On the other hand, for the structurally similar compound SM09, two replicas show a similar partition of the ionic species (1.05 and 0.91). The discrepancies in the partition of the ionic species may also affect the experimental determination of the log P for the neutral species of these compounds [47].

On the other hand, we do not think that temperature (25 $^{\circ}$C in the experimental values of the challenge, and 30 $^{\circ}$C in our measurement) might justify the difference between the two values, since according to previous studies one should expect an increase in hydrophobicity with a decrease in temperature [48, 49].

Figure 3. Representation of the relative log *P* prediction between SM13 and SM09 for all 91 submissions and classified by category and performance in the challenge.

Compound SM08

The error in the estimated log P of SM08 was partly due to a misassignment of the atomic surface tension for the nitrogen atom (ξ_{NH} = −0.295 kcal mol^{-1} Å^{-2} instead of ξ_{NH} = −0.234 kcal mol^{-1} Å^{-2}) in the original submission. Correction of this mistake would have led to a log P of 4.20. However, this value still deviates significantly from the experimental one (log P = 3.10), which could originate from the use a single conformational species of the keto tautomer (denoted SM08-T1 hereafter). This led us to carry out a more detailed analysis considering the involvement of alternative conformational and/or tautomeric species.

Table 2. Ranking of SM13 and SM15 with respect to the main ourliers in the best-ranked physical methods and the most accurate submissions from other categories.
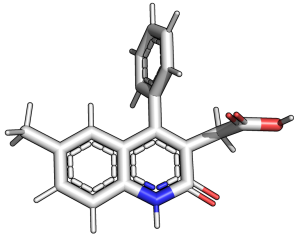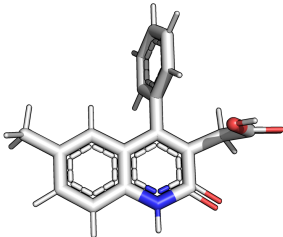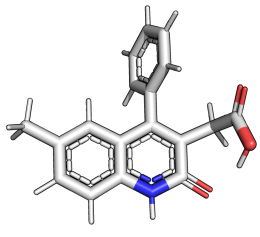
Sub. ID	Method	Category	Ranking (by rmse)		Ranking relative to the main outlier	
			Global	Category	SM13	SM15
hmz0n	COSMO-RS	Physical	1	1	1	2
j8nwc	EC_RISM	Physical	5	2	3	1
dqxk4	SMD	Physical	8	3	9	2
kivfu	IEFPCM/MST	Physical	25	4	2	3
gmoq5	Global XGBoost-Based QSPR	Empirical	2	1	3	4
3vqbi	Cosmoquick TZVP18+ML	Mixed	3	1	3	1
7dhtp	LogP-prediction	Other	14	1	1	2
Average					2[a]	2

[a] dqxk4 submission was not employed because SM13 was not an outlier in this case.

Three main conformational species were found for SM08-T1, which are shown in Table 3 together with the estimated population in the gas phase, *n*-octanol and water (populations in the gas phase are presented with the aim to demonstrate the relevant effect played by the solvent on the conformational preferences). The results point out the decrease of the intramolecularly hydrogen-bonded conformation (**III**) upon transition from the gas phase to *n*-octanol to water. Furthermore, the results show the modulation of the relative weigh for the conformations with the solvent-exposed carboxylic acid.

Table 3. Relative populations (%) of the major conformational species for the keto tautomer of compound SM08 estimated from SCS-MP2/CBS + ΔCCSD calculations. Only one of the symmetry-related conformations is shown.

| Solvent	**I**	**II**	**III**
Gas phase	50	8	42
n-octanol	42	20	38
water	62	32	6

Taking into account the weigh of the distinct conformers improves the prediction of the log *P* of SM08-T1, which is estimated to be 4.01 when the population distribution is determined at the DFT level. Moreover, weighting such conformations according to the SCS-MP2 energies corrected for CCSD high-order electron correlation effects

yields an estimated log P of 3.48, which almost matches the experimental value (log P = 3.10). Finally, we also determined the partition coefficient of the enol tautomer (SM08-T2; Fig. S1), leading to a log P of 3.30 (detailed data not shown). However, the contribution of the enol tautomer is expected to be negligible according to the relative stability in water and *n*-octanol (Table S1).
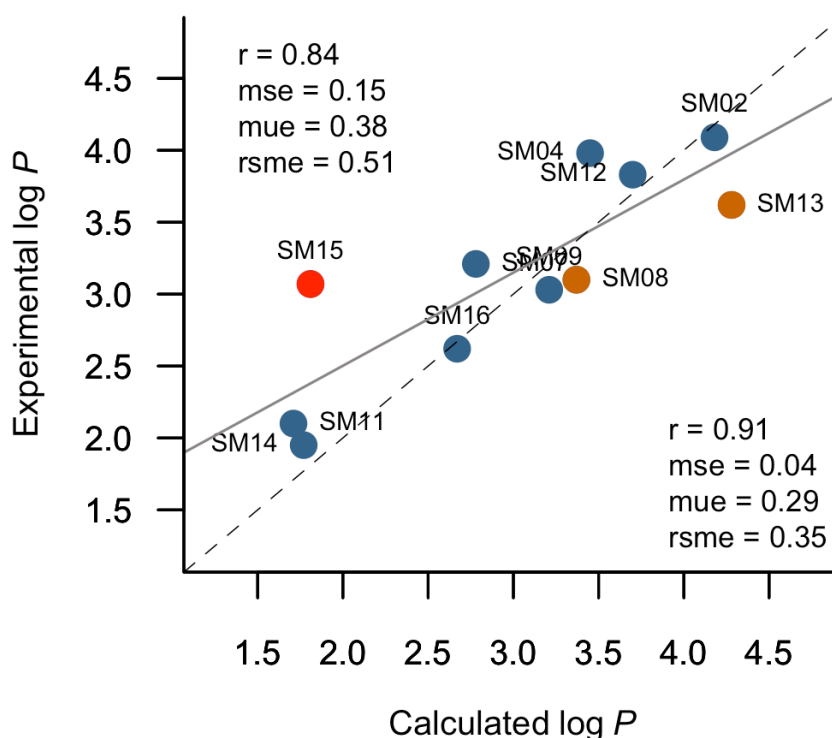
Compound SM15

The IEFPCM/MST calculations lead to remarkable different errors in the predicted log P of the benzimidazole-containing compounds SM14 and SM15, as the Δlog P error is much higher in the latter (-0.18 and -1.26 for SM14 and SM15, repectively). In fact, all the submissions predict that SM15 is less lipophilic compared to the experimental value (log P = 3.07), as noted in Fig. 4. In addition, Table 2 also points out that SM15 can be viewed as an outlier in the best-ranked submissions. The reasons that may explain this behaviour are unclear, but the uncertainty might be attributed to the population of cationic and anionic species during the potentiometric determination of the log P for a compound with distinct titratable sites. In this case, according to previous studies [50], the use of a shake-flask technique would be valuable for comparison purposes with the potentiometric method, because these compounds are prone to exhibit a larger difference between the log P values obtained by using these experimental techniques.

Figure 4. Representation of the average log *P* prediction for SM15 (dark blue circle) by all 91 submissions and classified by category in the challenge. The mean signed error (mse) is shown as light blue circles. The experimental value is shown as the dashed line.
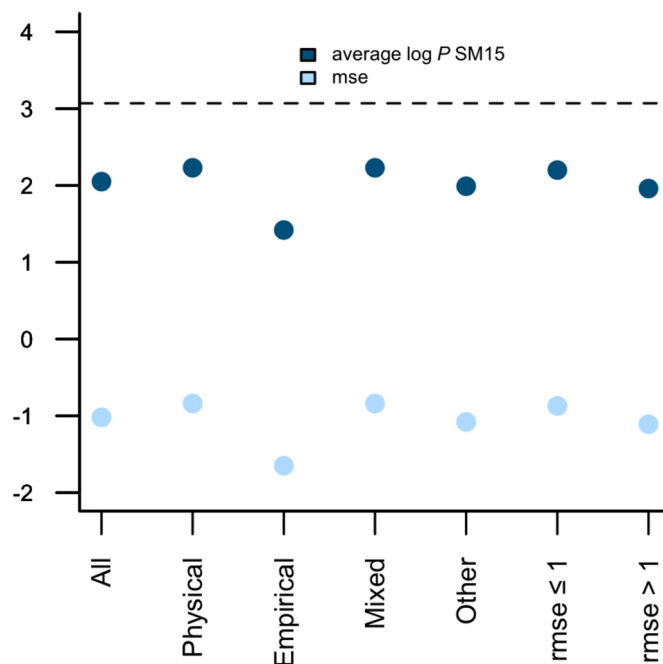
On the basis of the preceding considerations, if one keeps in mind the new experimental determination of log *P* reported here for SM13 and the recalculated value that accounts for the conformational flexibility of SM08, a significant improvement is found in the ability of the IEFPCM/MST model for predicting the experimental data of the compounds in the SAMPL6 challenge, as the rmse is reduced to 0.51 log units (0.35 upon exclusion of SM15; Fig. 5). Overall, the results support the suitability of the IEFPCM/MST model, as well as other QM-based continuum solvation models, for predicting the partitioning of neutral compounds with excellent accuracy.

Figure 5. Comparison between experimental and IEFPCM/MST *n*-octanol/ water log *P* for the SAMPL6 Dataset. Recalculated values for SM08 and our experimental log *P* for SM13 are represented in orange points, whereas SM15, the new main outlier of our approach, is shown as a red point. Statistical analyses are shown for all compounds (top left) and for 10 compounds excluding SM15 (bottom right).

Conclusions

The SAMPL6 Part II partition coefficient challenge has allowed us to examine the refinements made in the IEFPCM/MST continuum solvation model for the treatment of nitrogen-containing heterocyclic compounds in *n*-octanol. The results obtained with our submision ID "kivfu" support the suitability of the IEFPCM/MST model for predicting the *n*-octanol/water partitioning of neutral compounds. However, the

analysis of the compounds with the largest uncertainties reveals the need to maintain a close cross talk between theoretical predictions and experimental measurements. Thus, the experimental evaluation of the partition coefficient for SM13 yields a value that is in better agreement with chemical intuition, as it reflects an increased lipophilicity compared with the structurally-related compound SM09, but also with the general behaviour predicted from theoretical calculations. In turn, the experimental and predicted log P values for SM08 can be reconciled after accounting for the solvent effect on the relative stability of the different conformational species. Overall, the availability of blind high-quality datasets, such as the set of compounds included in the SAMPL6 Part II log P challenge, is a powerful tool not only to calibrate the strengths and weaknesses of theoretical methods, but also to implement refinements to improve the chemical accuracy of predictions and to gain deeper insight into the physicochemical factors that regulate the partitioning of (bio)organic compounds.

Acknowledgements

Supplementary Material

The overall ranking of submissions, the analysis of errors for the compounds in the

SAMPL6 challenge, and data about the experimental determination of the log P for

compound SM13 are provided.

References

1.   Testa B, Carrupt PA, Gaillard P, Tsai RS (1996) Intramolecular interactions encoded in lipophilicity: Their nature and significance. In: Pliska V, Testa B, van de Waterbeemd H (eds), Lipophilicity in Drug Action and Toxicology, VCH, Weinheim, pp 49-71

2.   van de Waterbeemd H, Lennernäs H, Arturssin P (eds). (2003) Drug Bioavailability: Estimation of solubility, permeability, absorption and bioavailability, Wiley-VCH, Weinheim

3.   Caron G, Ermondi G, Scherer RA (2007) Lipophilicity, polarity, and hydrophobicity. In: Taylor JB, Triggle DJ (eds), Comprehensive Medicinal Chemistry, Elsevier Science, Oxford, pp 425-452.

4.   Muñoz-Muriedas J (2012) Bioavailability prediction at early drug discovery stages: In vitro assays and simple physico-chemical rules. In: Luque, FJ, Barril X (eds), Physico-chemical and computational approaches to drug discovery. Royal Society of Chemistry, Cambridge, pp 104-127.

5.   Leo A, Hansch C, Elkins D (1971) Partition coefficients and their uses. Chem Rev 71:525–616

6.   Sangster J (1989) Octanol-water partition coefficients of simple organic compounds. J Phys Chem Ref Data 18:1111–1227

7.   Leo AJ (1993) Calculating log Poct from structures. Chem Rev 93:1281–1306

8.   Mannhold R, Dross K (1996) Calculation procedures for molecular lipophilicity: A comparative study. Quant Struct Relationships 15:403–409

9.   Ghose AK, Viswanadhan VN, Wendoloski JJ (1998) Prediction of hydrophobic (lipophilic) properties of small organic molecules using fragmental methods: An analysis of ALOGP and CLOGP methods. J Phys Chem A 102:3762–

3772

10.   Mannhold R, Van De Waterbeemd H (2001) Substructure and whole molecule approaches for calculating log P. J Comput Aided Mol Des 15:337–354

11.   Mannhold, R.; Poda, G. I.; Ostermann, C.; Tetko I V (2009) Calculation of molecular lipophilicity: State-of-the-art and comparison of LogP methods on more than 96,000 compounds. J Pharm Sci 98:861–893

12.   Ginex T, Vazquez J, Gibert E, Herrero E, Luque FJ (2019) Lipophilicity in drug design: An overview of lipophilicity descriptors on 3D-QSAR studies. Fut Med Chem 11:1177–1193.

13.   Cramer CJ, Truhlar DG (1999) Implicit solvation models: Equilibria, structure, spectra, and Dynamics. Chem Rev 99:2161–2200.

14.   Luque FJ, Curutchet C, Muñoz-Muriedas J, Bidon-Chanal A, Soteras I, Morreale A, Gelpí JL, Orozco M (2003) Continuum solvation models: Dissecting the free energy of solvation. Phys Chem Chem Phys 5:3827–383

15.   Tomasi J, Mennucci B, Cammi R (2005) Quantum mechanical continuum solvation models. Chem Rev 105:2999–3093

16.   Klamt A (2018) The COSMO and COSMO-RS solvation models. WIREs Comput Mol Sci 1:1–11

17.   Cramer CJ, Truhlar DG (2008) A universal approach to solvation modeling. Acc Chem Res 41:760–768

18.   Klamt A, Mennucci B, Tomasi J, Barone V, Curutchet C, Orozco M, Luque FJ (2009) On the performance of continuum solvation methods. A comment on "Universal approaches to solvation modeling". Acc Chem Res 42:489–492

19.   Soteras I, Curutchet C, Bidon-Chanal A, Orozco M, Luque FJ (2005) Extension of the MST model to the IEF formalism: HF and B3LYP parametrizations.

J Mol Struct THEOCHEM 727:29–40

20.   Cancès E, Mennucci B, Tomasi J (1997) A new integral equation formalism for the polarizable continuum model: Theoretical background and applications to isotropic and anisotropic dielectrics. J Chem Phys 107:3032

21.   Mennucci B, Cancès E, Tomasi J (1997) Evaluation of solvent effects in isotropic and anisotropic dielectrics and in ionic solutions with a unified integral equation method: Theoretical bases, computational implementation, and numerical applications. J Phys Chem B 101:10506–10517

22.   Soteras I, Forti, F, Orozco M, Luque FJ (2009) Performance of the IEF-MST solvation continuum model in a blind test prediction of hydration free energies. J Phys Chem B 27:9330–9334.

23.   Soteras I, Orozco M, Luque FJ (2010) Performance of the IEF-MST solvation continuum model in the SAMPL2 blind test prediction of hydration and tautomerization free energies. J Comput Aided Mol Des 24:281–291

24.   Zamora WJ, Curutchet C, Campanera JM, Luque FJ (2017) Prediction of pH-dependent hydrophobic profiles of small molecules from Miertus-Scrocco-Tomasi continuum solvation calculations. J Phys Chem B 121:9868–9880

25. Işık M, Bergazin TD, Fox T, Rizzi A, Chodera JD, Mobley DL (2020) Assessing the accuracy of octanol-water partition coefficients in the SAMPL6 Part II log P challenge. J Comput Aided Mol Des (to be published in the special issue)

26.

        https://github.com/samplchallenges/SAMPL6/tree/master/physical_properties/logP/predictions/submission_files

27.   Luque FJ, Barril X, Orozco M (1999) Fractional description of free energies of solvation. J Comput Aided Mol Des 13:139–152

28. Curutchet C, Orozco M, Luque FJ (2001) Solvation in octanol: Parametrization of the continuum MST model. J Comput Chem 22:1180–1193

29. Claverie P, Daudey JP, Langlet J, Pullman B, Piazzola D, Huron MJ (1978) Studies of solvent effects. 1. Discrete, continuum, and discrete-continuum models and their comparison for some simple cases: NH_4^+, CH_3OH, and substituted NH_4^+. J Phys Chem 82:405–418

30. Pierotti RA (1976) A scaled particle theory of aqueous and nonaqueous aolutions. Chem Rev 76:717–726

31. Işık M, Levorse D, Mobley DL, Rhodes T, Chodera JD (2019) Octanol-water partition coefficient measurements for the SAMPL6 blind prediction challenge. bioRxiv. https://doi.org/10.1101/757393

32.

https://github.com/samplchallenges/SAMPL6/blob/master/physical_properties /logP/molecule_ID_and_SMILES.csv

33. https://cactus.nci.nih.gov/translate

34. Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; et al. (2009) Gaussian 09, revision D.01; Gaussian, Inc.: Wallingford CT

35. Miteva MA, Guyon F, Tufféry P (2010) Frog2: Efficient 3D conformation ensemble generator for small compounds. Nuc Acids Res 38:622–627

36. Grimme S (2003) Improved second-order Møller-Plesset perturbation theory by separate scaling of parallel- and antiparallel-spin pair correlation energies. J Chem Phys 118:9095–9102

37. Halkier A, Helgaker T, Jørgensen P, Klopper W, Koch H, Olsen J, Wilson AK (1998) Basis-set convergence in correlated calculations on Ne, N_2, and H_2O. Chem

Phys Lett 286:243–252.

38. Truhlar DG (1998) Basis-set extrapolation. Chem Phys Lett 294:45–48.

39. Işık M, Levorse D, Rustenburg AS, Ndukwe IE, Wang H, Wang X, Reibarkh M, Martin GE, Makarov AA, Mobley DL, Rhodes T, Chodera JD. pKa measurements for the SAMPL6 prediction challenge for a set of kinase inhibitor-like fragments. J Comput Aided Mol Des 32:1117−1138.

40. Ràfols C, Bosch E, Ruiz R, Box KJ, Reis M, Ventura C, Santos S, Araújo ME, Martins F (2012) Acidity and hydrophobicity of several new potential antitubercular drugs: Isoniazid and benzimidazole derivatives. J Chem Eng Data 57:330–338

41. Kramer, S. D.; Gautier, J. C.; Saudemon, P (1998) Considerations on the potentiometric log P determination. J Pharm Res 15:1310−1313

42. Eckert F, Klamt A (2002) Fast solvent screening via quantum chemistry: COSMO-RS approach. AIChE J 48:369–385

43. Marenich A V, Cramer CJ, Truhlar DG (2009) Universal solvation model based on solute electron density and on a continuum model of the solvent defined by the bulk dielectric constant and atomic surface tensions. J Phys Chem B 113:6378–6396

44. Marenich A V, Cramer CJ, Truhlar DG (2007) Self-consistent reaction field model for aqueous and nonaqueous solutions based on accurate polarized partiacl charges. J Chem Theory Comput 3:2011–2033.

45. Marenich A V, Cramer CJ, Truhlar DG (2013) Generalized Born solvation model SM12. J Chem Theory Comput 9:609–620

46. Kloss T, Heil J, Kast SM (2008) Quantum chemistry in solution by combining 3D integral equation theory with a cluster embedding approach. J Phys Chem B 112:4337–4343.

47.   Barzanti C, Evans R, Fouquet J, Gouzin L, Howarth NM, Kean G, Levet E, Wang D, Wayemberg E, Yeboah AA, Kraft A (2007) Potentiometric determination of octanol-water and liposome-water partition coefficients (log P) of ionizable organic compounds. Tetrahedron Lett 48:3337–3341.

48.   Bahadur NP, Shiu WY, Boocock DGB, Mackay D (1997) Temperature dependence of octanol-water partition coefficient for selected chlorobenzenes. J Chem Eng Data 42:685–688.

49.   Borrirukwisitsak S, Keenan HE, Gauchotte-Lindsay C (2012) Effects of salinity, pH and temperature on the octanol-water partition coefficient of bisphenol A. Int J Environ Sci Dev 3:460–464

50.   Port A, Bordas M, Enrech R, Pascual R, Rosés M, Ràfols C, Subirats X, Bosch E (2018) Critical comparison of shake-flask, potentiometric and chromatographic methods for lipophilicity evaluation (log Po/w) of neutral, acidic, basic, amphoteric, and zwitterionic drugs. Eur J Pharm Sci 122:331–340