Journal of Computer-Aided Molecular Design (2021) 35:813–818
https://doi.org/10.1007/s10822-021-00395-5

COSMO-RS predictions of logP in the SAMPL7 blind challenge

Judith Warnau[1] · **Karin Wichmann**[1] · **Jens Reinisch**[1]

Received: 24 February 2021 / Accepted: 4 June 2021 / Published online: 14 June 2021
© The Author(s), under exclusive licence to Springer Nature Switzerland AG 2021

Abstract
We applied the COSMO-RS method to predict the partition coefficient logP between water and 1-octanol for 22 small drug like molecules within the framework of the SAMPL7 blind challenge. We carefully collected a set of thermodynamically meaningful microstates, including tautomeric forms of the neutral species, and calculated the logP using the current COSMOtherm implementation on the most accurate level. With this approach, COSMO-RS was ranked as the 6st most accurate method (Measured by the mean absolute error (MAE) of 0.57) over all 17 ranked submissions. We achieved a root mean square deviation (RMSD) of 0.78. The largest deviations from experimental values are exhibited by five SAMPL molecules (SM), which seem to be shifted in most SAMPL7 contributions. In context with previous SAMPL challenges, COSMO-RS demonstrates a wide range of applicability and one of the best in class reliability and accuracy among the physical methods.

Keywords COSMO-RS · COSMOtherm · logP prediction · 1-Octanol–water partition coefficient

Introduction

The logP is defined as the logarithm of the partition coefficient between two phases, 1-octanol and water. It is used to measure the hydrophilicity/hydrophobicity of a chemical compound. logP is an important descriptor of substances for their bioavailability and pharmacological relevance. As a part of the Lipinski's rule of five [1, 2], the logP is a criterion which is used to estimate the drug likeness of a compound. Furthermore, the logP has an impact in environmental science, agrochemistry and toxicology.

It is also not always possible to determine the partition coefficient experimentally. A typical reason is the very small (or no) available amount of substance in early drug development or virtual screening stages. Other challenges where in-silico prediction are invaluable may include reaction intermediates, degradation states or metabolite intermediates as well as toxic or otherwise dangerous substances, e.g. when analyzing the environmental fate of warfare agents. There is a strong demand for accurate logP predictions in a number of industrial fields, where life-science is just the most prominent example.

There exist many different approaches, all aiming at an efficient and accurate calculation of logP, e.g. quantitative structure–property relationship (QSPR), linear-energy relationship (LFER), classical or quantum chemical physical methods, and empirical methods based on machine learning and artificial intelligent. The variety of methods is reflected in the last Statistical Assessment of Modeling of Proteins and Ligands (SAMPL) series of blind predictive challenges. Within the SAMPL competition, computational models are tested on their predictivity of properties related to challenges in the field of drug discovery. Initiated in 2008 [3], these blind challenges hosted from the Drug Design Data Resource initiative (D3R) [4] are aimed to advance computational techniques und give the opportunity to benchmark different methods against each other.

COSMO-RS performed very well in the last years SAMPL5 [5] and SAMPL6 [6] challenges. The task in SAMPL5 was to predict the logarithmic distribution coefficient logD of drug-like compounds between water and cyclohexane. The two phases represent two of the most extreme liquid phases occurring in nature regarding the hydrophilicity/hydrophobicity character. Inside a biological cell, they may reflect the different phases, cytosol and lipid membrane. The appropriate partition coefficient is of high relevance in order to predict the permeability through the biological membrane inside a living cell.

✉ Judith Warnau
judith.warnau@3ds.com

[1] Dassault Systemes Deutschland GmbH, Imbacher Weg 46, 51379 Leverkusen, Germany

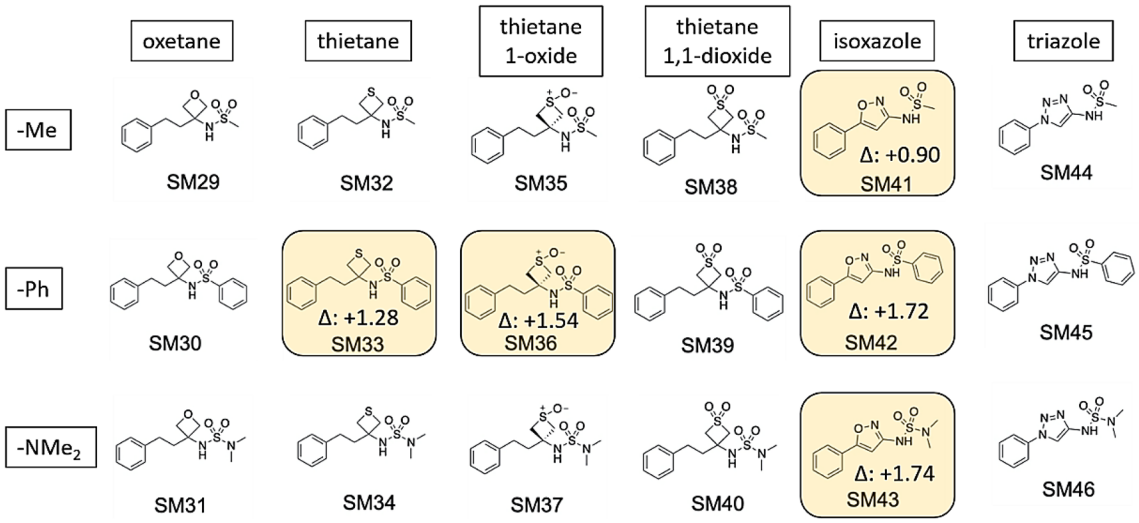

Fig. 1 Matrix of homologous series showing the structural relation of compounds SM29–SM46. 18 SM are ordered with respect to their characteristic functional groups. Bridging groups sorted in columns: oxetane, thietane, thietane 1-oxide, thietane 1,1-dioxide, isoxazole, and triazole group. Sulfone groups sorted in rows: methyl sulfone (-Me), phenyl sulfone (-Ph), and tertiary sulfonamide (-NMe_2) group. Orange boxes are highlighting the five outliers (SM33, SM36, SM41, SM42, and SM43). The largest deviations between predicted and experimentally determined logP values are indicated with Δ values

The two most recent SAMPL challenges, SAMPL6 and SAMPL7, focused on the prediction of the partition coefficient between water and 1-octanol phase. COSMO-RS achieved the most accurate logP prediction in the SAMPL6 challenge [6]. The pKa predictions in the SAMPL6 challenge with the lowest RMSE by Pracht et al. [7] used COSMO-RS computed free energies of solvation in combination with the conformational search with the MF-MD-GC//GFN-xTB workflow [8].

Method

The dataset of the SAMPL7 logP challenge consists of 22 small drug like molecules (SM) (Supplemental Fig. S1). All 22 SM share a similar chemistry. Common to all SM are the structural elements of a NH-acide group and an ethylbenzene. With a small deviation for SM41–SM46, where the ethyl bridge of this ethylbenzene group is replaced by a heteroaromatic ring, SM 29 to 46 build a consistent homologous series, therefore we constructed a matrix of homologues (Fig. 1) based on these 18 SM, comprising the following functional groups: oxetane, thietane, thietane 1-oxide, thietane 1,1-dioxide, isoxazole, and triazole. These structural elements appear in combination with three different sulfone groups: methyl sulfone, phenyl sulfone, and tertiary sulfonamide. The range of the experimental logP values reaches from 0.58 (SM41) to 2.96 (SM33) and shows only a part of the experimentally accessible [9] and pharmaceutical relevant logP values [1, 2].

The Conductor like Screening Model for Realistic Solvation (COSMO-RS) (Klamt1995 [10]) method has been described in previous publications for the SAMPL challenges 5 and 6 [5, 6]. For further and deeper explanations of the methodological foundation and definitions of the energy interaction terms, we refer to a review article [11].

In the following paragraph we will explain the concept of a COSMO surface and σ-profile, which is needed to understand Fig. 3. The COSMO surfaces consisting of charge elements σ are determined by quantum chemical density functional theory calculations within a virtual conductor implemented in an implicit solvation model (COSMO [10]). The COSMO reference state is a virtual conductor, meaning it is assumed all charges of the molecule are perfectly screened. [10]). σ is the charge which is needed to compensate the charge of the molecule at that place. σ is noted in the perspective of the conductor environment. A histogram of all σ of a molecule ordered by their charge is called a σ-profile. For a construction of a COSMO surface, all σ surface segments are mapped on the appropriate geometrical position on the molecular surface. The color represents the charge of a σ surface segment. The charge of a σ segment ranging from -0.03 to +0.03 e/A^2 is mapped to the color spectrum spanning from blue to red. That means, a negative-blue σ charge is compensating a positive charge of the molecule at that position, and a positive-red σ charge is compensating a negative charge of the molecule at that position.

This year, the task for the physical property logP challenge was to submit transfer energies between water and wet 1-octanol of all 22 SM [12]. Equation 1 gives the

relation of the transfer energy to the logarithm of the partition coefficient P. We calculated the transfer energies with the COSMO-RS method based on the difference of the chemical potential μ at infinite dilution in both phases, modelling the wet 1-octanol phase with a mole fraction of 0.2705 of water and the second phase as pure water. By adding the quotient of the molar volumes (V_2/V_1) to we converted the mole fraction based reference state typically used in COSMO-RS into a molar concentration-based framework (Eq. 2).[1]

$$\log_{10}(P) = \frac{\Delta G_{Transfer}}{RT\ \ln(10)} \tag{1}$$

$$\log_{10}(P) = \log_{10}\left(\frac{c_1}{c_2}\right) = \frac{\left(\mu_2^x - \mu_1^x\right)}{RT\ln(10)} + \log_{10}\left(\frac{V_2}{V_1}\right) \tag{2}$$

We started our workflow of preparing the SAMPL molecules with a systematic tautomer search by using COSMOquick [13–15]. The search algorithm for tautomers and conformers consists of several steps. First, we used COSMOquick to generate possible tautomeric states. Irrelevant states where discarded due to an internal energy threshold of 25 kcal/mol implemented in COSMOquick. By a subsequent manual check, we sorted out destroyed molecules and failed calculations. Second, we conducted a conformational search of every microstate with COSMOconf using up to 150 conformers. The conformational search based on Balloon [16] and RDkit [17, 18]. Different start conformations are generated in a random manner by using a distance geometry approach. In subsequent steps, the generated conformers are filtered by an identity check to delete redundant conformers and by an alignment und clustering step. A diversity clustering approach limits the amount to 150 conformers. The applied COSMOconf procedure performs a geometry optimization using the BP86 functional [19, 20] with a TZVP basis set [21] and the COSMO [22] solvation scheme for each conformer. A key step in the COSMOconf workflow is to reduce the number of conformers without changing the quality of the conformer set. Conformers are deleted in a systematic way, under the precondition that the Gibbs free energy does not change more than 0.2 kcal/mol. Subsequently, a single point energy calculation was done for the surviving conformers, using the BP86 functional with a def2-TZVPD [23] basis set and the FINE COSMO cavity [22]. All density functional theory calculations were carried out with the TURBOMOLE 7.5 [24] program package. The final set of microstates was generated by merging the conformer sets of all tautomeric forms of each individual SAMPL molecule and contains only those conformers and microstates that are relevant in liquid solutions.

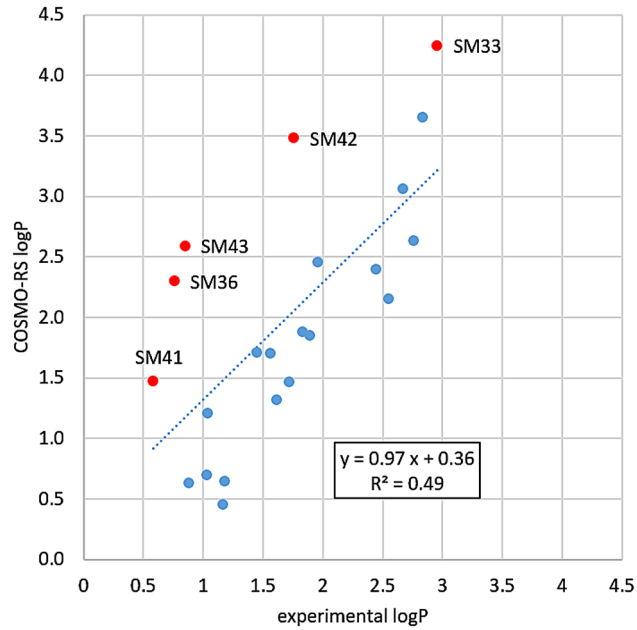

Fig. 2 Correlation plot of experimental measured logP [28] and logP predicted with COSMO-RS. The outliers SM33, SM36, SM41, SM42, and SM43 are indicated with red dots. The linear regression between COSMO-RS predicted and experimental measured logP values gives a regression line (blue dotted line) with a functional equation y = 0.97 x + 0.36. Including the outliers, the squared Pearson's correlation coefficient R^2 is 0.49

To predict the transfer energies between water and wet 1-octanol, we calculated the free energy difference in the respective phases for each molecule set. Within the COSMO-RS workflow, an ensemble of microstates is used. The microstates are weighted automatically by COSMOtherm according to their total free energy in the respective liquid phase, i.e. different weights are used in water and wet 1-octanol. For the free energy calculations we used the COSMOtherm 2020 [10, 25–27] implementation of COSMO-RS in combination with the BP-TZVPD-FINE 20 parameterization. The applied COSMO-RS version in the parametrization BP-TZVPD-FINE 20 shows a general root mean square deviation (RMSD) of 0.6 kcal/mol in our internal data sets and with regard to experimental 1-octanol/water partitioning data. The COSMO-RS method by itself has no statistical error and a numerical error of smaller 0.01 kcal/mol. The overall workflow including the conformational search has a statistical noise smaller than 0.1 kcal/mol on average.

[1] Originally, we submitted transfer free energies in the mol-fraction based reference state (Eq. 2 without the additional term of molar volumes). After realizing that the concentration-based framework was asked, we got the opportunity to resubmit our predicted transfer free energies in the relevant reference state by adding the logarithm of the quotient of the molar volumes (like in Eq. 2).

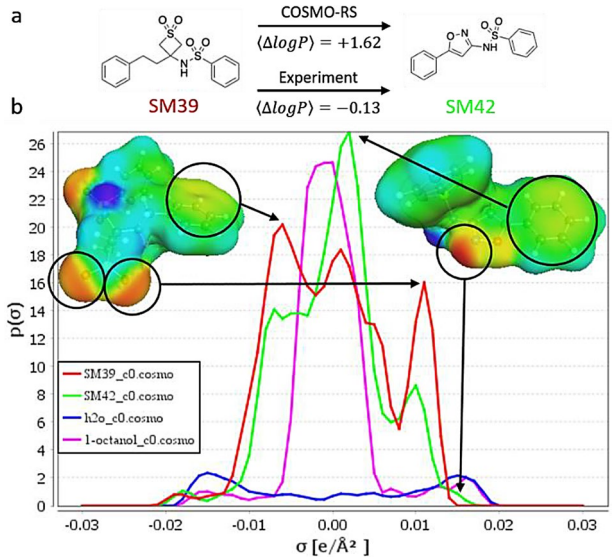

Fig. 3 Comparison between SM39 and SM42 as an example of a transfer of the functional group from thietane 1,1-dioxide to isoxazole. **a** 2D structural representation of SM39 (red) and SM42 (green). The average shift in logP going from the thietane 1,1-dioxide group (SM38, SM39, and SM40) to the isoxazole group (SM41, SM42, and SM43) are given for experimental values (− 0.13) and COSMO-RS prediction (+ 1.62). **b** σ-profiles. p(σ) of SM39 (red), SM42 (green), water (blue), and 1-octanol (magenta) are represented in the units elementary charge *e* divided by $Å^2$. The COSMO surfaces, calculated in a conductor environment, are represented as insets for SM39 and SM42 on the left and right side, respectively. Black circles emphasize key sigma surface elements and the black arrow point to their peaks in the corresponding σ-profile. The green COSMO surface at the phenyl next to the isoxazole group of SM42 indicates the more hydrophobic character of an enlarged mesomeric system. These neutral σ-surface segments of isoxazole SM are dominating the increase in solubility in 1-octanol and therefor the increase of logP

Result

COSMO-RS predicts logP values of small drug like molecules with an overall good accuracy. A correlation plot (Fig. 2) suggests, except for a few outliers, an overall good agreement of COSMO-RS predicted and experimentally measured logP values. The Kendall's tau correlation coefficient, which is especially suited for ranking purposes, is 0.53, which is the highest value among all physical-based method submissions. We achieved an RMSD of 0.78 log units for our prediction of the 22 SM. Compared to the best-ranked method, TFE MLR with an RMSD of 0.58, we are just 0.2 log units off.

We identified five outliers in our logP prediction deviating more than 0.90 log units from the experimental data [28] (see also Supplemental Table S1): SM33 (+ 1.28), SM36 (+ 1.54), SM41 (+ 0.90), SM42 (+ 1.72), SM43 (+ 1.74). We discuss the consequences of the outliers to the performance later in the discussion section.

The isoxazole series (SM41, SM42, and SM43) deserve a special role in the SAMPL challenge. We predict a contradicting trend of logP shift compared to experimental values. Transitioning from the thietane 1,1-dioxide group (SM38, SM39, and SM40) to the isoxazole group (SM41, SM42, and SM43), the average shift in experimental logP values is -0.52 whereas for COSMO-RS prediction it is + 1.04 log units (Supplemental Table S2). The isoxazoles appear more hydrophobic based on our calculations than based on experimental measurements. The COSMO surfaces and σ-profiles illustrates the more hydrophobic character of the SAMPL molecules of the isoxazole group compared to the SAMPL molecules of the thietane 1,1-dioxide group (Fig. 3), which is due to the higher amount of neutral surface area.

Discussion

The COSMO-RS logP predictions showed an overall very good agreement with experimental values in the last SAMPL blind challenges. In the last years, we achieved the highest accurate logP prediction in SAMPL5 (2016) and SAMPL6 (2019). COSMO-RS demonstrated to predict reliable logP values of high quality [5, 6] independent of the solvent and chemical diversity. In SAMPL5 COSMO-RS predicted most accurately the partition coefficient between hexane and water [5]. Last year, we achieved a very accurate result for logP (1-octanol–water) of SAMPL molecules with multiple functional groups of chemical diversity, for example also including halogens. This year, empirical machine learning based submissions achieved a higher accuracy. COSMO-RS was, however, the 2nd most accurate physical method and delivered the highest ranking correlation coefficient in this class.

In SAMPL7, COSMO-RS quite accurately predicted logP values (RMSD of 0.78), apart from five outliers (SM33 (+ 1.28), SM36 (+ 1.54), SM41 (+ 0.90), SM42 (+ 1.72), and SM43 (+ 1.74)). To deepen our understanding about the nature of the large deviation in these cases, we compared the experimental logP values between the homologous series (Fig. 1). We grouped the SAMPL molecules depending on their characteristic functional group and ordered 18 of the 22 SM in a regular matrix (Fig. 1), revealing structural relationships. We checked for the consistency of the experimental logP values by calculating average shifts in logP between the functional groups ordered in rows and columns of the homology matrix (Supplemental Fig. S2). By this analysis, the experimental logP values show consistent trends in logP shifts between the homologous series, except for SM33, SM36 and the isoxazole group (SM41, SM42, and SM43) (Supplemental Fig. S2). Comparing the shifts in logP between the first (-Me: methyl sulfone group) and the second row (-Ph: phenyl sulfone group) reveals a notably small shift of the transition going from SM32 to SM33 and from

SM35 to SM36 (Supplemental Fig. S2A). Surprisingly, the shift of logP is even negative for the transition from SM35 to SM36, meaning that a substitution of a methyl group by a phenyl group leads to increase of hydrophilicity. The shifts in logP from SM33 to SM36 and from SM36 to SM39 are deviating strongly from the average shift in logP between thietane to thietane 1-oxide and thietane 1-oxideto thietane 1,1-dioxide group. Due to this large deviation in the logP shifts from common trends, caution has to be taken by comparing predicted logP values of SM33 and SM36 with the experimental logP values.

To analyze if the outliers are only present in COSMO-RS we compare how all other methods perform on these molecules. We calculated the **A**verage prediction of the **B**est **5** submissions (AB5)[2] (Supplemental Equation S1), which are yielding a smaller RMSD than COSMO-RS, as well as the **A**verage prediction **W**eighted by their **I**nverse overall RMSD of all 17 ranked submissions (AWI17)[4] (Supplemental Table S1 and Supplemental Equation S2). AB5 and AWI17 show the same trend as our predictions, an increase of average logP shift from the thietane 1,1-dioxide group to the isoxazole group by 0.74 (AB5) and 0.96 (AWI17) log units, respectively (Supplemental Table S2). The largest deviation of AB5 and AWI17 to experimental logP values occurs at SM33 (−0.68, −0.71), SM36 (−1.72, −1.53), SM41 (−0.98, −1.00), SM42 (−1.11, −1.40), and SM43 (−1.04, −1.24) (Supplemental Table S1). The AB5 and AWI17 analysis reveals that all submissions show the same deviations from the experimental logP values (Supplemental Table S1, Supplemental Fig. S3). Therefore, we conclude that SM33, SM36, SM41, SM42, and SM43 are not only outliers in COSMO-RS, also for various other methods [29].

The large deviation between predicted and experimental determined logP values of isoxazoles remain to be answered. By additional single point calculations with different QM methods, we can so far exclude that the tautomer relation and therefore the composition of our ensemble is independent of a specific choice of QM method (see also Supplemental Fig. S6 and Table S3). To understand the nature of this deviation we would have to apply further investigations to systematically evaluate the major source of the deviation. In order to conduct a broad study with many isoxazole derivatives, reliable experimental logP measurements would be crucial.

We assessed the accuracy of COSMO-RS method without the outliers (SM33, SM36, SM41, SM42, and SM43). The performance of COSMO-RS based on the remaining 17 SM without the five outliers would increase significantly. The RMSD would decreases to 0.38 log units. The correlation coefficient R^2 would increase to 0.76 (Supplemental Fig. S4). The Kendall's tau correlation coefficient would rise to 0.78. These statistics without the outliers is in good agreement with previous performance evaluation of COSMO-RS [5, 6]. We also analyzed the RMSD of all 17 ranked submissions without the five outliers (Supplemental Fig. S5).[3] All methods would perform better. A consequence of a statistical analysis without outliers is that the top eight submissions would lie close together in the RMSD range of 0.39–0.66 (Supplemental Fig. S5). In this regime, some empirical methods would swap their places in ranking. This outcome underlines the robustness and reliability of physical methods like COSMO-RS.

Conclusion

COSMO-RS performed quite well in the logP prediction challenge SAMPL7. We achieved an accuracy of 0.78 log units, which is only 0.2 log units above the best submission.[4] With that result our submission was ranked to the 6th place according to accuracy (MAE) and the 3rd place according to the ranked correlation coefficient Kendall's tau ($\tau = 0.53$). Among the physical methods, COSMO-RS performed 2nd best. In comparison with SAMPL5 and SAMPL6, the performance and ranking of COSMO-RS submission dropped significantly, indicating the impressive improvement of machine learning techniques and their increasing applicability for physical properties predictions. Based on these outcomes, we aim to improve the COSMO-RS theory with machine learning techniques in the next years.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s10822-021-00395-5.

Acknowledgements The authors thank David Mobley and the organizers for setting up the SAMPL7 challenge and fruitful discussions. We appreciate the National Institutes of Health for its support of the SAMPL project via R01GM124270 to David L. Mobley (UC Irvine). We thank the Ballatore lab at UCSD for carrying out the experimental logP measurements. We thank Andreas Klamt for helpful discussions. We thank Michael Diedenhofen, Johannes Schwöbel and Frank Eckert for conducting a big part of the tautomer and conformer workflow with COSMOconf and COSMOquick as well as for helping us in refining the data and supporting discussions. Judith Warnau thanks Felix Hanke for revising the manuscript.

Author contributions All authors contributed to developing the research design. Judith Warnau performed material preparation, data collection and analysis. All authors contributed to data analysis decisions and interpretation of the results. Judith Warnau wrote the first draft of the manuscript and all authors revised and commented on

[2] The AB5 and AWI17 analysis was based on the SAMPL7 analysis of Version 0.7 (https://github.com/samplchallenges/SAMPL7/releases/tag/0.7).

[3] This analysis was based on the SAMPL7 analysis of Version 0.7 (https://github.com/samplchallenges/SAMPL7/releases/tag/0.7).

[4] In SAMPL7, TFE MLR was the most accurate method with an RMSD of 0.58 log units.

previous versions of the manuscript. All authors read and approved the final manuscript.

Funding Financial interests: All author are employed and have received research support from Company Dassault Systemes Deutschland GmbH. Apart from that, no funding was received for conducting this study.

Availability of data and materials The datasets generated during and analyzed during the current study are available from the corresponding author on reasonable request.

Code availability The code of our software is commercially available.

Declarations

Conflict of interest All authors are employed by Dassault Systemes Deutschland GmbH. Dassault Systemes commercially distributes the software (COSMOtherm, COSMOconf, COSMOquick and Turbomole) which was used to conduct the calculations for this manuscript. Apart from that, there are no conflicts of interests or competing interests.

References

1. Lipinski CA, Lombardo F, Dominy BW, Feeney PJ (2001) Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. Adv Drug Deliv Rev 46:3–26. https://doi.org/10.1016/s0169-409x(00)00129-0
2. Lipinski CA (2004) Lead- and drug-like compounds: the rule-of-five revolution. Drug Discov Today 1:337–341. https://doi.org/10.1016/j.ddtec.2004.11.007
3. Nicholls A, Mobley DL, Guthrie JP et al (2008) Predicting small-molecule solvation free energies: an informal blind test for computational chemistry. J Med Chem 51:769–779. https://doi.org/10.1021/jm070549+
4. (2019) Drug design data resource. In: Drug Des. Data Resour. https://drugdesigndata.org. Accessed 1 Feb 2019
5. Klamt A, Eckert F, Reinisch J, Wichmann K (2016) Prediction of cyclohexane-water distribution coefficients with COSMO-RS on the SAMPL5 data set. J Comput Aided Mol Des 30:959–967. https://doi.org/10.1007/s10822-016-9927-y
6. Loschen C, Reinisch J, Klamt A (2020) COSMO-RS based predictions for the SAMPL6 logP challenge. J Comput Aided Mol Des 34:385–392. https://doi.org/10.1007/s10822-019-00259-z
7. Pracht P, Wilcken R, Udvarhelyi A et al (2018) High accuracy quantum-chemistry-based calculation and blind prediction of macroscopic pKa values in the context of the SAMPL6 challenge. J Comput Aided Mol Des 32:1139–1149. https://doi.org/10.1007/s10822-018-0145-7
8. Grimme S, Bannwarth C, Dohm S et al (2017) Fully automated quantum-chemistry-based computation of spin-spin-coupled nuclear magnetic resonance spectra. Angew Chem Int Ed. https://doi.org/10.1002/anie.201708266
9. Sangster J (1989) Octanol-water partition coefficients of simple organic compounds. J Phys Chem Ref Data 18:1111–1229. https://doi.org/10.1063/1.555833
10. Klamt A (1995) Conductor-like screening model for real solvents: a new approach to the quantitative calculation of solvation phenomena. J Phys Chem 99:2224–2235. https://doi.org/10.1021/j100007a062
11. Klamt A (2018) The COSMO and COSMO-RS solvation models: COSMO and COSMO-RS. Wiley Interdiscip Rev Comput Mol Sci 8:e1338. https://doi.org/10.1002/wcms.1338
12. Bannan CC, Calabró G, Kyu DY, Mobley DL (2016) Calculating partition coefficients of small molecules in octanol/water and cyclohexane/water. J Chem Theory Comput 12:4015–4024. https://doi.org/10.1021/acs.jctc.6b00449
13. *COSMOquick*, Version 1.8; Dassault Systemes, 2020. https://www.3ds.com/
14. Hornig M, Klamt A (2005) COSMOfrag: a novel tool for high-throughput ADME property prediction and similarity screening based on quantum chemistry. J Chem Inf Model 45:1169–1177. https://doi.org/10.1021/ci0501948
15. Loschen C, Klamt A (2012) COSMOquick: a novel interface for fast σ-profile composition and its application to COSMO-RS solvent screening using multiple reference solvents. Ind Eng Chem Res 51:14303–14308. https://doi.org/10.1021/ie3023675
16. Vainio MJ, Johnson MS (2007) Generating conformer ensembles using a multiobjective genetic algorithm. J Chem Inf Model 47:2462–2474. https://doi.org/10.1021/ci6005646
17. RDKit: Open-Source Cheminformatics Software. https://www.rdkit.org/ (Release_2016.09). Accessed 4 Nov 2016
18. Ebejer J-P, Morris GM, Deane CM (2012) Freely available conformer generation methods: how good are they? J Chem Inf Model 52:1146–1158. https://doi.org/10.1021/ci2004658
19. Becke AD (1988) Density-functional exchange-energy approximation with correct asymptotic behavior. Phys Rev A 38:3098–3100. https://doi.org/10.1103/PhysRevA.38.3098
20. Perdew JP (1986) Density-functional approximation for the correlation energy of the inhomogeneous electron gas. Phys Rev B 33:8822–8824. https://doi.org/10.1103/PhysRevB.33.8822
21. Schäfer A, Huber C, Ahlrichs R (1994) Fully optimized contracted Gaussian basis sets of triple zeta valence quality for atoms Li to Kr. J Chem Phys 100:5829. https://doi.org/10.1063/1.467146
22. Klamt A, Schüürmann G (1993) COSMO: a new approach to dielectric screening in solvents with explicit expressions for the screening energy and its gradient. J Chem Soc Perkin Trans 2(1993):799–805. https://doi.org/10.1039/P29930000799
23. Rappoport D, Furche F (2010) Property-optimized Gaussian basis sets for molecular response calculations. J Chem Phys 133:134105. https://doi.org/10.1063/1.3484283
24. TURBOMOLE V7.4 2019, a development of University of Karlsruhe and Forschungszentrum Karlsruhe GmbH, 1989-2007, TURBOMOLE GmbH,since 2007; available from https://www.turbomole.org.
25. Klamt A, Jonas V, Bürger T, Lohrenz JC (1998) Refinement and parametrization of COSMO-RS. J Phys Chem A 102:5074–5085. https://doi.org/10.1021/jp980017s
26. Eckert F, Klamt A (2002) Fast solvent screening via quantum chemistry: COSMO-RS approach. AIChE J 48:369–385. https://doi.org/10.1002/aic.690480220
27. BIOVIA COSMOtherm, Release 2020; Dassault Systèmes. https://www.3ds.com/
28. Mobley DL, Bergazin D, et al (2020) samplchallenges/SAMPL7: version 0.6: Add logP challenge inputs/experimental results. Zenodo. https://doi.org/10.5281/zenodo.4086044
29. Bergazin TD, Tielker N, Zhang Y, Mao J, Gunner MR, Francisco K, Ballatore C, Kast SM, Mobley DL. Evaluation of log P, pKa, and log D predictions from the SAMPL7 blind challenge

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.