

Review

# Revisiting the Use of Quantum Chemical Calculations in $\text{LogP}_{\text{octanol-water}}$ Prediction

Dipankar Roy <sup>1,\*</sup>  and Chandan Patel <sup>2,\*</sup> <sup>1</sup> Department of Biological Sciences, University of Alberta, Edmonton, AB T6G 2E9, Canada<sup>2</sup> Department of Applied Sciences, COEP Technological University, Wellesely Road, Shivajinagar, Pune 411005, Maharashtra, India

\* Correspondence: droy1@ualberta.ca (D.R.); chandanz@gmail.com (C.P.)

**Abstract:** The partition coefficients of drug and drug-like molecules between an aqueous and organic phase are an important property for developing new therapeutics. The predictive power of computational methods is used extensively to predict partition coefficients of molecules. The application of quantum chemical calculations is used to develop methods to develop structure–activity relationship models for such prediction, either based on molecular fragment methods, or via direct calculation of solvation free energy in solvent continuum. The applicability, merits, and shortcomings of these developments are revisited here.

**Keywords:** octanol-water partition coefficients; quantum chemical calculations; ab initio methods; density functional theory; continuum solvation models; reference interaction site model

## 1. Introduction

The importance of the accurate prediction of physicochemical properties is quintessential, not only in drug development efforts but also toward applications in material science for molecules with desirable properties. The importance of quantitative structure activity/property relationships (QSAR/QSPR) is undoubtedly essential in chemistry and biology [1–4]. The linear free energy relationship (LFER) approach often provides a model system based on molecular topologies to relate with biological properties. It is, however, difficult to have a correct QSAR/QSPR for almost all of the cases, although a working relationship is of great demand in the pharmaceutical industry [5,6]. The essential components of such relationships depend on correct knowledge of solvation properties of a target molecule in multiple solvents of interest, with solvent polarities spreading across a broad scale. This is important for determining whether target molecules are adequate to get absorbed in an organ of interest. The important biophysical properties that help a drug to enter the body and distribute across various biological barriers consist of solubility, stability, and permeability [7,8]. The property of interest, lipophilicity, is the ability of a drug or drug-like molecule to dissolve in fat-like substances such as oils or non-polar solvents (toluene, cyclohexane, chloroform etc.). It is well recognized for the successful development of drug candidates, since the introduction by Hansch and Fujita's method to correlate biological activity with chemical structures [9]. Hydrophobic (water-fearing) drugs prefer to be in the hydrophobic organelles/compartments, such as the lipid bilayers of a cell, while the hydrophilic (water-loving) drugs will prefer to be dissolved in the blood serum (see Figure 1). In general,  $\text{logP} \leq 0$  means high hydrophilicity, and  $\text{logP} > 0$  means increased lipophilicity [10]. As per the well-known Lipinski's "rule of five" for drug-likeness, an orally active molecule has no more than one violation of the following criteria: (i) not more than five hydrogen bond donors (N/O atoms) with one or more hydrogen atoms, (ii) not more than ten hydrogen bond acceptors (N/O atoms), (iii) molecular weight under 500 Daltons, and (iv)  $\text{logP}$  coefficient of less than five [11]. The final criterion is based on the partitioning of molecules between organic and aqueous phase as in n-octanol/water mixture,



**Citation:** Roy, D.; Patel, C. Revisiting the Use of Quantum Chemical Calculations in  $\text{LogP}_{\text{octanol-water}}$  Prediction. *Molecules* **2023**, *28*, 801. <https://doi.org/10.3390/molecules28020801>

Academic Editor: Enrico Bodo

Received: 14 December 2022

Revised: 6 January 2023

Accepted: 10 January 2023

Published: 13 January 2023



**Copyright:** © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

traditionally expressed as the negative logarithm of partition coefficient ( $\log P_{\text{Organic-Water}}$  or simply  $\log P$ ) [12]. A study using Caco-2 cell monolayers has shown that the optimal permeability of drugs occurs at  $\text{Clog}P$  values between 4 and 5, while an increase in  $\text{Clog}P$  value from this range decreases the permeability of drugs [13].

From a physical chemistry perspective, for non-ionic and neutral forms of compounds,  $\log P$  is expressed as in Equation (1), where  $R$  is the molar gas constant and  $T$  is the temperature in K.

$$\log P = \frac{-\Delta G_{\text{transfer}}}{RT \ln 10},$$

where

$$\Delta G_{\text{transfer}} = \Delta G_{\text{solvation}} - \Delta G_{\text{hydration}} \quad (1)$$

Whereas, for ionic compounds, whose overall ionization in solution is pH dependent, often  $\log D$  is used instead of  $\log P$ . The  $\log D$  determines the ratio of the sum of the concentrations of all ionized and unionized forms of the compound in each of the two phases of interest (Equation (2)).

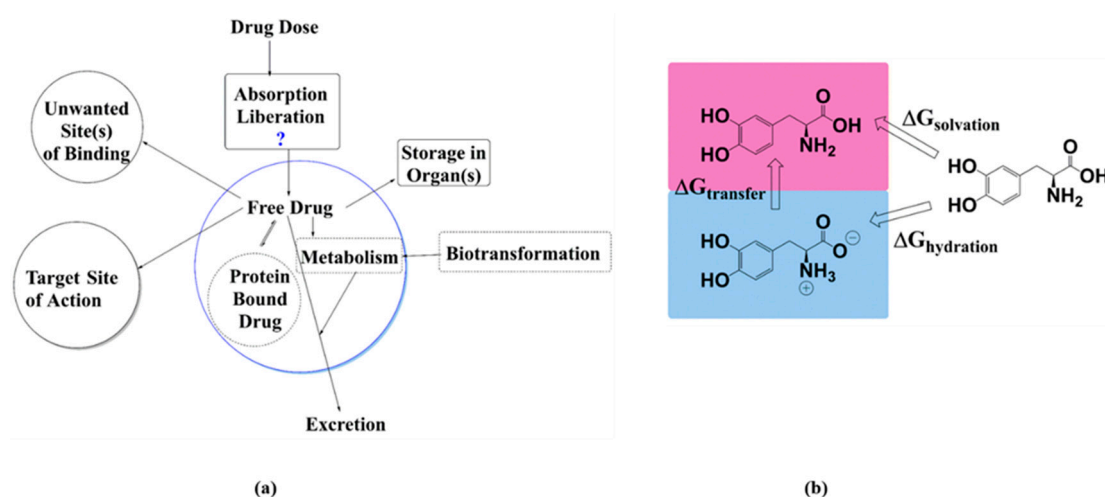
$$\log D_{\text{Organic-Water}} = \log \left( \frac{[\text{Solute}]_{\text{neutral,Organic}}}{[\text{Solute}]_{\text{neutral,Water}} + [\text{Solute}]_{\text{ionic,Water}}} \right), \quad (2)$$

For molecules whose aqueous ionization constants are known,  $\log D$  is related to  $\log P$  via Equation (3). The significance of  $\log D$  values becomes clear if one looks into the vast range of physiological pH values in the human body, e.g., stomach: ~2; plasma: 7.4; small intestine: ~5–6.8. [14].

$$\log D_{\text{Organic-Water}} = \log P_{\text{Organic-Water}} - \log \left( 1 + 10^{(pH - pK_a) \Delta i} \right) \quad (3)$$

where  $\Delta i = +1$  for acids and  $\Delta i = -1$  for bases [15].

Evidently, the solvation free energy terms are guiding forces here to determine the ability of a molecule to pass through different biological barriers before reaching the target receptor(s). While the distribution of a compound between n-octanol and water mixture (expressed as  $\log P_{\text{OW}}$ ) was traditionally used as a measure of the lipophilicity and/or hydrophobicity of compounds, strictly speaking the hydrophobicity can be defined as the aggregation propensity of substances in aqueous medium, and n-octanol very poorly represents lipid bilayers. For passive diffusion of drug molecules, which is considered as one of the major pathways of drug absorption, the partition coefficient between aqueous phase and organic phase is directly related to membrane permeability (see Figure 1a for possible fates of a drug inside the human body after ingestion). To gain knowledge of solvation properties, experimental studies are the most reliable; however, for a large molecular dataset, it is a tremendous effort to explore these properties in multiple solvents. This is often not so useful from an industrial point of view, as many compounds fail to clear the initial screening processes based on solvation properties; rendering initial molecular screening expensive. The other and most efficient method is the application of computational methods to explore molecular solvation properties. The use of computational and theoretical tools in understanding molecular solvation is a well-researched area, and is continuously evolving based on need [16]. In the context of  $\log P_{\text{OW}}$  calculation/prediction using theoretical and computational techniques, empirical, semi-empirical, and quantum chemical calculations-based theories have been applied [17–19]. The QSAR computed values of  $\log P$  are often used instead of expensive experimental studies. This is a very effective technique for high throughput virtual screening (HTVS), although they can generate significant inaccuracies, and potential compounds can be rejected, or false positive results may skew efforts [20–22].



**Figure 1.** (a) Possible fates of a drug molecule inside human body; (b) the physical process of distribution of a drug molecule between two phases.

## 2. Methods for logP Empirical Predictions

The logP prediction methods are broadly classified into two categories, viz. the sub-structure (or fragment) based methods, and the whole molecule approach. The fragment-based approach is guided by the principle of averaged contributions of simple fragments corrected against large experimental datasets [23,24]. An excellent example of an improved version of this methodology is the clogP program module [25]. The whole molecule approach develops upon molecular lipophilicity potential (MLP), topological indices, and/or molecular quantities to predict logP. While these two methods are complementary to each other as they try to circumvent the shortcomings of each other, they have their own limitations and efficiency considerations [26,27]. For a comparison of *logP* methods on an extensive set of molecules, see references [5,28]. With phenomenal improvements in computer hardware, and with the advent of logic-high programming languages, the scope of implementing and validating QSAR/QSPR models has become a part of routine development [29]. The earlier studies of linear regression analysis were eventually refined using neural networks and genetic algorithms-based computing schemes for logP prediction/validation. The latest in this field is the application of machine learning protocols (random Forest, support vector machine) for building non-linear QSAR/QSPR models and principle component analysis [30–33].

The solvation properties are a direct outcome of the electronic structure of a molecule, and therefore, electronic structure calculations seem to be an excellent starting point for calculation of logP. The use of quantum chemical methods is applied for two very specific regimes of methods: first, for the development of descriptors to assist in computation of logP using the whole molecule-based approaches; and second, direct calculation of solvation free energy using different continuum solvation models. The following discussion is divided into two sections, viz. applications of quantum mechanics based molecular descriptors for logP prediction, and direct prediction of logP via free energy of solvation calculations.

## 3. Applications of Quantum Mechanics Based Molecular Descriptors for logP Prediction

Since the first demonstration by Hanch and Fujita, that logP can be calculated using a linear correlation method, several similar methods were reported for various purposes, mostly for QSAR/QSPR analysis (see Figure 2 for information on different types of descriptors used in logP prediction modules). The Hansch–Fujita method predicts logP of a compound with substitution of H with X, using a relation with the unsubstituted form as  $\pi = \log P_X - \log P_H$  where the parameter  $\pi$  is obtained by regression analysis against

standard databases. This simple looking equation leads to a plethora of models involving molecular properties, and more significantly intrinsic electronic properties, which can be partitioned over atoms present in a molecule. The application of empirical and quantum chemical (QC) methods to derive such descriptors have become very popular. The QC methods were more accurate and well suited for calculating atomic partial-charges, molecular orbital (MO) energies, dipole moments, and many other properties. In earlier days, the demand for heavy computational resources to generate molecular wave function had restricted the use of the QM descriptors; however, semi-empirical calculations became quite popular, which were developed within the mathematical framework of molecular orbital theory, but with more generalizations and approximations (SCF-MO). There are several semi-empirical methods available which have been benchmarked against experimental atomic and molecular data to be used for generation of molecular descriptors [34]. The first descriptor to be used for logP prediction were the atomic charge densities. The method is based on building a linear relation for the set of 61 compounds covering hydrocarbons, alcohols, ethers, carbonyls, amines, nitriles, and amides [35]. The squared atomic charges, calculated using semi-empirical MNDO/3 Hamiltonian method, are used to develop a correlation equation as

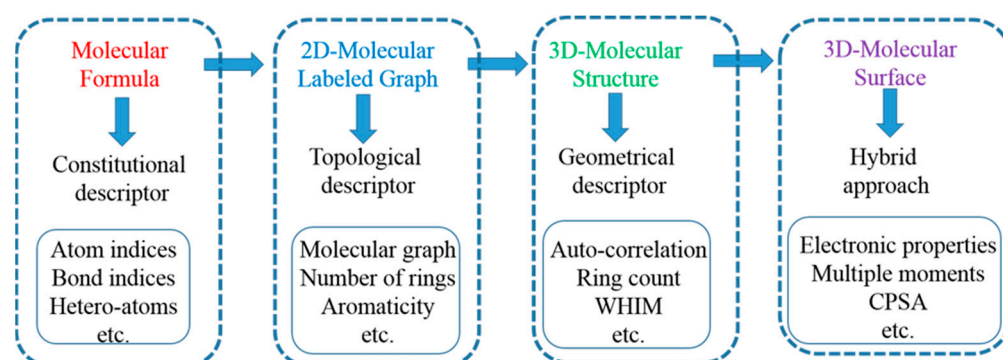
$$\log P = 0.344 + 0.2078n_H + 0.093n_C - 2.119n_N - 1.937n_O - 1.389\sum q_C^2 - 17.28\sum q_N^2 + 0.7316\sum q_O^2 + 2.844NA + 0.910NT + 1.709NM,$$

resulting in a statistical correlation of  $n = 61$ ,  $r^2 = 0.985$ ,  $s = 0.15$ .

The parameters used in this relation are  $n_H$ ,  $n_C$ ,  $n_N$ , and  $n_O$ , and represent the numbers of hydrogen, carbon, nitrogen, and oxygen atoms, respectively;  $q_i$  is the atomic charge on the corresponding atoms; NA, NT, and NM are numbers of carboxy, cyano, and amide groups, respectively. Interestingly, this approach yielded better results than fragment-based methods, despite being simplistic. A subsequent development was made involving curvature of molecular surface, molecular surface area, and dipole moment. This new model predicted logP with very good correlations for 118 molecules [36,37]. For a series of substituted phenols, the logP prediction scheme was reported using HOMO and LUMO energy gaps, molecular volume, total surface area, molar mass, refractivity, melting point, charge on oxygen atom, and polarizability of the molecule, with excellent success [38]. For a larger dataset of 592 molecules, an excellent QSAR model was developed using only polarizability and partial atomic charges on nitrogen and oxygen. This three-parameter based model is simple and covers a vast molecular space [39]. The van der Waal volume was incorporated in a more recent model computed using density functional theory (DFT) based B3LYP, functional with a double- $\xi$  quality basis set. The authors have used multivariate analysis revealing differences between the chemical classes in terms of their electronic properties, and importance of the frontier MOs in logP prediction [40]. A theoretical development of heuristic molecular lipophilicity potential, using quantum chemically calculated electrostatic potential (ESP) as a unified solvent-philicity potential was reported by Du et al. [41]. The resultant linear regression analysis has the form of:

$$\log P = \frac{a \times L_M + b \times H_M}{RT \ln 10} + c$$

where molecular lipophilicity ( $L_M$ ) and hydrophilicity ( $H_M$ ) indices were calculated using quantum chemical calculation (semi-empirical, ab initio Hartree-Fock, and DFT-B3LYP levels). For an overview of the lipophilicity predictions, based on QSAR models in drug design, please see reference [41]. There are numerous reports on development of QSAR models and logP predictions based on QM-methods, although the basic number and types of molecular descriptors that were employed are limited in number as to make the model easier to compute and accessible. The use of QM-methods has often overcome deficiencies of force-field based computation schemes.



**Figure 2.** Hierarchical developmental tree of QSAR descriptors for applications in molecular design.

#### 4. Direct Prediction of logP via Free Energy of Solvation Calculations

The accuracy of implicit solvation methods for solvation free energy calculations, and the role of molecular conformations are two major limitations that handicap applications of direct logP estimations using electronic structure calculations. The positive sides of implicit solvation models are (i) reduced computation time in comparison to an explicit all atom calculation, (ii) inherent parameterization to reproduce experimental macroscopic properties of small molecules and ions, and (iii) for an arbitrary geometry of solute “... continuum models automatically give configurationally sampled solvent effect” [42]. One should note that in the case of flexible molecules, the single conformation implicit solvation free energy may not correctly represent conformational ensembles in different phases. Amongst implicit quantum chemical solvation models, the most prominent ones are (i) the polarization continuum models based on the extension of the ‘Integral Equation Formalism’ (IEF), (ii) Minnesota solvation models developed by Truhlar, Cramer, and coworkers traditionally designated as SMD, and (iii) COnductor like Screening MOdel for Real Solvents (COSMO-RS). The description of the molecular solvation theory and other derivatives of it are beyond the scope of the present review. In short, all these methods depend on generating a molecular shaped cavity to place the solute molecule inside a solvent continuum. The deviations in calculating molecular electronic properties as well as solvation energy are mostly arising from the choice of atomic radii sets, which varies significantly between different methods, and calculation of electrostatic field. The electrostatic calculation based on the Density Functional Theory (DFT) grid with a cavity radius of  $\sim 1.2 \times$  van der Waals (vdw) radii is a reliable starting point for molecules covering diverse molecular space.

The accurate prediction of hydration free energy by Minnesota solvation model (SMn series;  $n = 5, 6, 8, 12$  etc.) was an encouraging step towards using an implicit solvation model for logP prediction, as well as using hydration free energy as a descriptor in drug design [43,44]. In a study by Kolar and coworkers on the accuracy of implicit solvation methods in calculating  $\log P_{OW}$  of 20 neutral drug molecules, the best performers were SMD and COSMO-RS methods, whereas molecular mechanics (MM) based methods performed poorly [45]. The authors also found significant deviations between computed and experimental logP for flexible molecules, when a single rigid conformation is used for computation. In a recent work, the applicability of the Onsager equation for realistic systems has been questioned [46]. Indeed, the computational complexity in calculating  $\Delta G_{solvation}$  is often the limiting step for correct prediction of the partition coefficients. The COSMO-RS method can use a priori prediction of solubilities of a large class of compounds. The first report of training COSMO-RS on a 150 drug-like compound database and testing on a pesticides database containing 107 entries provided rms deviation of 0.66–0.61 log units [47]. One of the best applications of COSMO-RS in predicting cyclohexane–water distribution coefficients in SAMPL5 dataset was reported by Klamt and coworkers [48]. Ouimet and Paluch made a reference database from Drugbank for a large number of fragments based on 4-amino quinazoline structure with experimental logP values in order to develop a calibration curve for logP prediction of select SAMPL6 compounds, using the Minnesota solvation

models [49]. These authors found building a training set did not improve prediction for the SAMPL6 set. Their best results were from untrained SM8 solvation level calculations at the M06-2X/6-31G(d) level. Kundi and Ho calculated logP of two sets of compounds, viz. 34 organic compounds and 55 fluorinated alkanols and carbohydrates using implicit and explicit solvents [50]. These authors found that implicit solvent models performed better than the explicit solvent models with a mean absolute deviation of 0.6 log units using the COSMO-RS continuum model. The best results in this study are obtained from empirical fragment-based predictions (ALOGP and miLogP) with a mean deviation in the range 0.2–0.4 log units. Fluorinated compounds receive a lot of attention in drug development owing to their use as hydrophobicity modulating moiety (e.g.,  $-\text{CF}_3$ ,  $p\text{-F}_3\text{C-C}_6\text{H}_4\text{-}$ ). Very recently, Patel and Roy reported a systematic analysis of the effect of atomic radii on logP calculations of 56 fluorinated drug molecules using DFT functionals [51]. The standard SMD model is found to outperform the conductor like the polarizable continuum model, while used in combination with various density functionals. These authors reported no significant basis set size issues in logP calculations. The use of the dispersion corrected density functionals did not improve logP predictions in comparison to those computed by non-dispersion corrected ones, as reported by Patel and Roy [51].

A direct logP<sub>OW</sub> calculation for a large set of compounds using semi-empirical quantum chemical methods with dielectric continuum models found significant deviations between computed and experimental data [52]. An improved prediction was obtained by using multiple linear regression analysis based on structural and constitutional features of the molecules in the same database. A back propagating neural network model was proposed by Briendl et al. for logP prediction for 104 organic molecules based on AM1 and PM3 semi-empirical MO calculations [53]. Fizer et al. reported a detailed analysis of the performance of a combination of different electrostatic charge generation schemes with semiempirical, Hartree-Fock, and DFT level computations for hydrophilic/lipophilic index and logP calculations of 50 mono-charged organic ions [54]. For this set of ions, the authors obtained the best results with the semi-empirical AM1 Ford–Wang parametric electrostatic potential charges. The performance of the natural population with DFT also resulted in optimal performances.

Amongst physics-based methods, the three-dimensional reference interaction site model (3D-RISM) theory has also been applied successfully toward prediction of partition coefficients. The theory uses first principle statistical mechanics to describe properties of molecular solvents using modest computational resources. The theory has been extended significantly to apply in quantum-chemical calculation using the RISM-SCF method [55,56]. The RISM-SCF method is used for the calculations of molecular aggregates, solvation free energy, and hydration of organometallic systems [57–61]. Successful applications of the 3D-RISM theory with Kovalenko–Hirata (KH) closure relation in predicting cyclohexane–water partition coefficients for SAMPL5 and logP<sub>octanol-water</sub> were reported by Kovalenko and co-workers [62,63]. It is obvious at this point that the need for good computational/theoretical solvation models is essential at present to address solvation modeling. The physics-based methods, relying on quantum chemical description of solvation, are far away from direct applications in logP calculation based on pure hydration free energy. Methods such as 3D-RISM(KH) and its variants are a suitable replacement; although, they are often plagued with issues related to calculation convergence, system size, and parallelization.

## 5. Conclusions

It is important to address the suitability of a choice of logP calculation/prediction based on a case specific manner. For similar molecular structures to be used for high throughput virtual screening, the empirical logP predictions based on atom/fragment addition approaches may often suffice. Thus, QSAR/QSPR approaches are adequate here. However, this is clearly not the case for systems with novel and/or exotic fragments, which is often the case for development of new pharmaceuticals. The descriptor-based

methods should be developed with special care so as to have a minimum of five times more compounds in the training database, than the number of molecular descriptors used. The whole molecule-based methods may work very well for such a scenario. It is worth, for such cases, to treat the molecular descriptor in the most accurate way possible. For instance, an error in calculating atomic charges and resultant multiple moments would result in significant deviation of molecular dipole moment from the actual value, and would interfere with all subsequent calculations based on these descriptors. The semi empirical AM1 method-based charges are often a good choice to start with. Further desired improvements should be guided by necessity and available computational resources. The quantum chemical calculations are an excellent choice for specific solvation property calculations and should be used judiciously. For small molecular sizes, electronic structure calculations with explicit solvation methods are the best; moreover, for a relatively larger system size, a hybrid approach is recommended. Deploying a hybrid approach is non-trivial and requires understanding of the amount of explicit solvent molecules to be added per molecule of substrate to represent the solvation shell, before immersing the entire supramolecular structure in a dielectric continuum model. The molecular solvation theory-based methods, such as 3D-RISM, should be properly standardized for their use in different solvents before setting off to use them for large molecular databases [64]. Of course, if experimental results are available for a reasonable number of systems, then multiple methods/models should be evaluated for correctness before finalizing the method of choice.

**Author Contributions:** D.R. and C.P. contributed equally to this work. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Not applicable.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Dearden, J.C. Prediction of physicochemical properties. *Methods Mol Biol.* **2012**, *929*, 93–138.
2. Wildman, S.A.; Crippen, G.M. Prediction of physicochemical parameters by atomic contributions. *J. Chem. Inf. Comput. Sci.* **1999**, *39*, 868–873. [[CrossRef](#)]
3. Basak, S.C.; Harriss, D.K.; Magnuson, V.R. Comparative study of lipophilicity versus topological molecular descriptors in biological correlations. *J. Pharm. Sci.* **1984**, *73*, 429–437. [[CrossRef](#)] [[PubMed](#)]
4. Zhao, Y.H.; Le, J.; Abraham, M.H.; Hersey, A.; Eddershaw, P.J.; Luscombe, C.N.; Butina, D.; Beck, G.; Sherborne, B.; Cooper, I.; et al. Evaluation of human intestinal absorption data and subsequent derivation of a quantitative structure-activity relationship (QSAR) with the Abraham descriptors. *J. Pharm. Sci.* **2001**, *90*, 749–784. [[CrossRef](#)] [[PubMed](#)]
5. Guha, R.; Willighagen, E. A Survey of Quantitative Descriptions of Molecular Structure. *Curr. Top. Med. Chem.* **2012**, *12*, 1946–1956. [[CrossRef](#)]
6. Spjuth, O.; Willighagen, E.L.; Guha, R.; Eklund, M.; Wikberg, J.E.E. Towards interoperable and reproducible QSAR analyses: Exchange of datasets. *J. Cheminformatics* **2010**, *2*, 5. [[CrossRef](#)]
7. Lombardo, F.; Shalaeva, M.Y.; Tupper, K.A.; Gao, F.; Abraham, M.H. ElogP<sub>oct</sub>: A Tool for Lipophilicity Determination in Drug Discovery. *J. Med. Chem.* **2000**, *43*, 2922–2928. [[CrossRef](#)]
8. Guillot, A.; Henchoz, Y.; Moccand, C.; Guillaume, D.; Veuthey, J.L.; Carrupt, P.A.; Martel, S. Lipophilicity Determination of Highly Lipophilic Compounds by Liquid Chromatography. *Chem. Biodivers.* **2009**, *6*, 1828–1836. [[CrossRef](#)]
9. Fujita, T.; Iwasa, J.; Hansch, C. A New Substituent Constant,  $\pi$ , Derived from Partition Coefficients. *J. Am. Chem. Soc.* **1964**, *86*, 5175–5180. [[CrossRef](#)]
10. Sawant, P.D.; Luu, D.; Ye, R.; Buchta, R. Drug release from hydroethanolic gels. Effect of drug's lipophilicity (logP), polymer-drug interactions and solvent lipophilicity. *Int. J. Pharm.* **2010**, *396*, 45–52. [[CrossRef](#)]
11. Lipinski, C.A. Lead- and drug-like compounds: The rule-of-five revolution. *Drug Discov. Today Technol.* **2004**, *1*, 337–341. [[CrossRef](#)]
12. Sangster, J. *Octanol-Water Partition Coefficients: Fundamentals and Physical Chemistry*; Wiley Series in Solution Chemistry; Wiley: New York, NY, USA, 1997; Volume 2.

13. Yang, Y.; Engkvist, O.; Llinas, A.; Chen, H. Beyond Size, Ionization State, and Lipophilicity: Influence of Molecular Topology on Absorption, Distribution, Metabolism, Excretion, and Toxicity for Druglike Compounds. *J. Med. Chem.* **2012**, *55*, 3667–3677. [[CrossRef](#)]
14. Kerns, E.H.; Di, L. Pharmaceutical Profiling in Drug Discovery. *Drug Discov. Today* **2003**, *8*, 316–323. [[CrossRef](#)]
15. Aligas, I.; Gobbi, A.; Lee, M.-L.; Sellers, B.D. Comparison of logP and logD correction models trained with public and proprietary data sets. *J. Comput.-Aided Mol. Des.* **2022**, *36*, 253–262. [[CrossRef](#)] [[PubMed](#)]
16. Skyner, R.E.; McDonagh, J.L.; Groom, C.R.; van Mourik, T.; Mitchell, J.B.O. A review of methods for the calculation of solution free energies and the modelling of systems in solution. *Phys. Chem. Chem. Phys.* **2015**, *17*, 6174–6191. [[CrossRef](#)] [[PubMed](#)]
17. da Silva, C.O.; da Silva, E.C.; Nascimento, M.A.C. Ab Initio Calculations of Absolute pKa Values in Aqueous Solution I. Carboxylic Acids. *J. Phys. Chem. B* **1999**, *103*, 11194–11199. [[CrossRef](#)]
18. Citra, M.J. Estimating the pKa of Phenols, Carboxylic Acids and Alcohols from Semiempirical Quantum Chemical Methods. *Chemosphere* **1999**, *38*, 191–206. [[CrossRef](#)] [[PubMed](#)]
19. Duprat, A.F.; Huynh, T.; Dreyfus, G. Toward a Principled Methodology for Neural Network Design and Performance Evaluation in QSAR. Application to the Prediction of logP. *J. Chem. Inf. Comput. Sci.* **1998**, *38*, 586–594. [[CrossRef](#)] [[PubMed](#)]
20. Liao, S.Y.; Chen, J.C.; Qian, L.; Shen, Y.; Zheng, K.C. QSAR, action mechanism and molecular design of flavone and isoflavone derivatives with cytotoxicity against HeLa. *Eur. J. Med. Chem.* **2008**, *43*, 2159–2170. [[CrossRef](#)]
21. Khan, D.A.U. Descriptors and their selection methods in QSAR analysis: Paradigm for drug design. *Drug Discov. Today* **2016**, *21*, 1291–1302.
22. Mikulski, D.; Górnica, R.; Molski, M. A theoretical study of the structure–radical scavenging activity of trans-resveratrol analogues and cis-resveratrol in gas phase and water environment. *Eur. J. Med. Chem.* **2010**, *45*, 1015–1027. [[CrossRef](#)] [[PubMed](#)]
23. Leo, A.J.; Jow, P.Y.C.; Silipo, C.; Hansch, C. Calculation of hydrophobic constant (log P) from  $\pi$  and  $f$  constants. *J. Med. Chem.* **1975**, *18*, 865–868. [[CrossRef](#)] [[PubMed](#)]
24. Hansch, C.; Leo, A.J. *Substituent Constants for Correlation Analysis in Chemistry and Biology*; Wiley: New York, NY, USA, 1979.
25. Leo, A.J. Calculating log P<sub>oct</sub> from structures. *Chem. Rev.* **1993**, *93*, 1281–1306. [[CrossRef](#)]
26. Bodor, N.; Buchwald, P. Molecular Size Based Approach To Estimate Partition Properties for Organic Solutes. *J. Phys. Chem. B* **1997**, *101*, 3404–3412. [[CrossRef](#)]
27. Mannhold, R.; Dross, K. Calculation Procedures for Molecular Lipophilicity: A Comparative Study. *Quant. Struct.-Act. Relat.* **1996**, *15*, 403–409. [[CrossRef](#)]
28. Mannhold, R.; Poda, G.I.; Osterman, C.; Tetko, I.V. Calculation of molecular lipophilicity: State-of-the-art and comparison of log P methods on more than 96,000 compounds. *J. Pharma. Sci.* **2009**, *98*, 861–893. [[CrossRef](#)] [[PubMed](#)]
29. Baldi, P. Deep learning in biomedical data science. *Ann. Rev. Biomed. Data Sci.* **2018**, *1*, 181–205. [[CrossRef](#)]
30. Boobier, S.; Osbourne, A.; Mitchell, J.O.B. Can human experts predict solubility better than computers? *J. Cheminformatics* **2017**, *9*, 63. [[CrossRef](#)]
31. Lusci, A.; Pollastri, G.; Baldi, P. Deep Architectures and Deep Learning in Chemoinformatics: The Prediction of Aqueous Solubility for Drug-Like Molecules. *J. Chem. Inf. Model.* **2013**, *53*, 1563–1575. [[CrossRef](#)]
32. Panteleev, J.; Gao, H.; Jia, L. Recent applications of machine learning in medicinal chemistry. *Bioorg. Med. Chem. Lett.* **2018**, *28*, 2807–2815. [[CrossRef](#)]
33. Fuchs, J.-A.; Grisoni, F.; Kossenjans, M.; Hiss, J.A.; Schneider, G. Lipophilicity prediction of peptides and peptide derivatives by consensus machine learning. *Med. Chem. Commun.* **2018**, *9*, 1538–1546. [[CrossRef](#)]
34. Cartier, A.; Rivail, J.-L. Electronic descriptors in quantitative structure—Activity relationships. *Chemom. Intell. Lab. Sys.* **1987**, *1*, 335–347. [[CrossRef](#)]
35. Klopman, G.; Iroff, L.D. Calculation of partition coefficients by the charge density method. *J. Comput. Chem.* **1981**, *2*, 157–160. [[CrossRef](#)]
36. Bodor, N.; Gabanyi, Z.; Wong, C.-K. A new method for the estimation of partition coefficient. *J. Am. Chem. Soc.* **1989**, *111*, 3783–3786. [[CrossRef](#)]
37. Bodor, N.; Huang, M.-J. Predicting partition coefficients for isomeric diastereoisomers of some tripeptide analogs. *J. Comput. Chem.* **1991**, *12*, 1182–1186. [[CrossRef](#)]
38. Makovskaya, V.; Dean, J.R.; Tomlinson, W.R.; Comber, M. Octanol-water partition coefficients of substituted phenols and their correlation with molecular descriptors. *Anal. Chim. Acta* **1995**, *315*, 193–200. [[CrossRef](#)]
39. Larsson, M.; Mishra, B.K.; Tysklind, M.; Linusson, A.; Anderson, P.L. On the use of electronic descriptors for QSAR modelling of PCDDs, PCDFs and dioxin-like PCBs. *SAR QSAR Environ. Res.* **2013**, *24*, 461–479. [[CrossRef](#)] [[PubMed](#)]
40. Du, Q.; Liu, P.-J.; Mezey, P.G. Theoretical Derivation of Heuristic Molecular Lipophilicity Potential: A Quantum Chemical Description for Molecular Solvation. *J. Chem. Inf. Model.* **2005**, *45*, 347–353. [[CrossRef](#)]
41. Ginex, T.; Vazquez, J.; Gilbert, E.; Herrero, E.; Luque, L.J. Lipophilicity in drug design: An overview of lipophilicity descriptors in 3D-QSAR studies. *Future Med. Chem.* **2019**, *11*, 1177–1193. [[CrossRef](#)]
42. Mennucci, B. Continuum Solvation Models: What Else Can We Learn from Them? *J. Phys. Chem. Lett.* **2010**, *1*, 1666–1674. [[CrossRef](#)]
43. Marenich, A.V.; Cramer, C.J.; Truhlar, D.G. Performance of SM6, SM8, and SMD on the SAMPL1 Test Set for the Prediction of Small-Molecule Solvation Free Energies. *J. Phys. Chem. B* **2009**, *113*, 4536–4543. [[CrossRef](#)]



44. Zafar, A.; Reynisson, J. Hydration Free Energy as a Molecular Descriptor in Drug Design: A Feasibility Study. *Mol. Inform.* **2016**, *35*, 207–214. [[CrossRef](#)] [[PubMed](#)]
45. Kolář, M.; Fanfrlík, J.; Lepšík, M.; Forti, F.; Luque, F.J.; Hobza, P. Assessing the Accuracy and Performance of Implicit Solvent Models for Drug Molecules: Conformational Ensemble Approaches. *J. Phys. Chem. B* **2013**, *117*, 5950–5962. [[CrossRef](#)] [[PubMed](#)]
46. Zhang, J.; Zhang, H.; Wu, T.; Wang, Q.; van der Spoel, D. Comparison of Implicit and Explicit Solvent Models for the Calculation of Solvation Free Energy in Organic Solvents. *J. Chem. Theor. Comput.* **2017**, *13*, 1034–1043. [[CrossRef](#)]
47. Klamt, A.; Eckert, F.; Hornig, M.; Beck, M.E.; Bürger, T. Prediction of aqueous solubility of drugs and pesticides with COSMO-RS. *J. Comput. Chem.* **2002**, *23*, 275–281. [[CrossRef](#)] [[PubMed](#)]
48. Klamt, A.; Eckert, F.; Reinisch, J.; Wichmann, K. Prediction of cyclohexane-water distribution coefficients with COSMO-RS on the SAMPL5 data set. *J. Comput. Aided Mol. Des.* **2016**, *30*, 959–967. [[CrossRef](#)]
49. Ouimet, J.A.; Paluch, A.S. Predicting octanol/water partition coefficients for the SAMPL6 challenge using the SM12, SM8, and SMD solvation models. *J. Comput.-Aided Mol. Des.* **2020**, *34*, 575–588. [[CrossRef](#)]
50. Kundi, V.; Ho, J. Predicting Octanol-Water Partition Coefficients: Are Quantum Mechanical Implicit Solvent Models Better than Empirical Fragment-Based Methods? *J. Phys. Chem. B* **2019**, *123*, 6810–6822. [[CrossRef](#)]
51. Patel, C.; Roy, D. Octanol-Water Partition Coefficients of Fluorinated Drug Molecules with Continuum Solvation Models. *J. Phys. Chem. A* **2022**, *126*, 4185–4190. [[CrossRef](#)]
52. Takahashi, T.; Matsui, T.; Hengphasatporn, K.; Shigeta, Y. A Practical Prediction of LogPo/w through Semiempirical Electronic Structure Calculations with Dielectric Continuum Model. *Bul. Chem. Soc. Jpn.* **2021**, *94*, 1807–1814. [[CrossRef](#)]
53. Breindl, A.; Beck, B.; Clark, T.; Glen, R.C. Prediction of the n-Octanol/Water Partition Coefficient, logP, Using a Combination of Semiempirical MO-Calculations and a Neural Network. *Mol. Model. Ann.* **1997**, *3*, 142–155. [[CrossRef](#)]
54. Fizer, O.; Fizer, M.; Sidey, V.; Studenyak, Y.; Mariychuk, R. Benchmark of different charges for prediction of the partitioning coefficient through the hydrophilic/lipophilic index. *J. Mol. Model.* **2018**, *24*, 141. [[CrossRef](#)] [[PubMed](#)]
55. Gusarov, S.; Ziegler, T.; Kovalenko, A. Self-Consistent Combination of the Three-Dimensional RISM Theory of Molecular Solvation with Analytical Gradients and the Amsterdam Density Functional Package. *J. Phys. Chem. A* **2006**, *110*, 6083–6090. [[CrossRef](#)] [[PubMed](#)]
56. Casanova, D.; Gusarov, S.; Kovalenko, A.; Zeigler, T. Evaluation of the SCF Combination of KS-DFT and 3D-RISM-KH; Solvation Effect on Conformational Equilibria, Tautomerization Energies, and Activation Barriers. *J. Chem. Theor. Comput.* **2007**, *3*, 458–476. [[CrossRef](#)] [[PubMed](#)]
57. Hlushak, S.; Stoyanov, S.R.; Kovalenko, A. A 3D-RISM-KH Molecular Theory of Solvation Study of the Effective Stacking Interactions of Kaolinite Nanoparticles in Aqueous Electrolyte Solution Containing Additives. *J. Phys. Chem. C* **2016**, *120*, 21344–21357. [[CrossRef](#)]
58. Yokogawa, D.; Sato, H.; Imai, T.; Sakaki, S. A highly parallelizable integral equation theory for three dimensional solvent distribution function: Application to biomolecules. *J. Chem. Phys.* **2009**, *130*, 06411. [[CrossRef](#)]
59. Ano, S.; Hosoya, T.; Sakaki, S. A 3D-RISM-SCF method with dual solvent boxes for a highly polarized system: Application to 1,6-anhydrosugar formation reaction of phenyl  $\alpha$ - and  $\beta$ -D-glucosides under basic conditions. *Phys. Chem. Chem. Phys.* **2013**, *15*, 6368–6381. [[CrossRef](#)]
60. Ano, S.; Nakagaki, M.; Sakaki, S. Theoretical study of one-electron-oxidized salen complexes of group 7 (Mn(III), Tc(III), and Re(III)) and group 10 metals (Ni(II), Pd(II), and Pt(II)) with the 3D-RISM-GMC-QDPT method: Localized vs. delocalized ground and excited states in solution. *Phys. Chem. Chem. Phys.* **2017**, *19*, 16831–16849. [[CrossRef](#)]
61. Ano, S.; Mori, T.; Sakaki, S. 3D-RISM-MP2 Approach to Hydration Structure of Pt(II) and Pd(II) Complexes: Unusual H-Ahead Mode vs. Usual O-Ahead One. *J. Chem. Theory. Comput.* **2016**, *12*, 1189–1206. [[CrossRef](#)]
62. Luchko, T.; Blinov, N.; Limon, G.C.; Joyce, K.P.; Kovalenko, A. SAMPL5: 3D-RISM partition coefficient calculations with partial molar volume corrections and solute conformational sampling. *J. Comput. Aided Mol. Des.* **2016**, *30*, 1115. [[CrossRef](#)]
63. Huang, W.J.; Blinov, N.; Kovalenko, A. Octanol-Water Partition Coefficient from 3D-RISM-KH Molecular Theory of Solvation with Partial Molar Volume Correction. *J. Phys. Chem. B* **2015**, *119*, 5588–5597. [[CrossRef](#)] [[PubMed](#)]
64. Roy, D.; Blinov, N.; Kovalenko, A. Predicting Accurate Solvation Free Energy in n-Octanol Using 3D-RISM-KH Molecular Theory of Solvation: Making Right Choices. *J. Phys. Chem. B* **2017**, *121*, 92688–9273. [[CrossRef](#)] [[PubMed](#)]

**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.