Journal of Molecular Graphics and Modelling 60 (2015) 108–117

Contents lists available at ScienceDirect

Journal of Molecular Graphics and Modelling

journal homepage: www.elsevier.com/locate/JMGM

Computational prediction of octanol–water partition coefficient based on the extended solvent-contact model

Taeho Kim, Hwangseo Park*

Department of Bioscience and Biotechnology, Sejong University, 209 Neungdong-ro, Gwangjin-gu, Seoul 143-747, Republic of Korea

ARTICLE INFO

Article history:
Received 10 March 2015
Received in revised form 9 June 2015
Accepted 10 June 2015
Available online 16 June 2015

Keywords:
Partition coefficient
Solvation free energy
Solvent-contact model
Genetic algorithm

ABSTRACT

The logarithm of 1-octanol/water partition coefficient (LogP) is one of the most important molecular design parameters in drug discovery. Assuming that LogP can be calculated from the difference between the solvation free energy of a molecule in water and that in 1-octanol, we propose a method for predicting the molecular LogP values based on the extended solvent-contact model. To obtain the molecular solvation free energy data for the two solvents, a proper potential energy function was defined for each solvent with respect to atomic distributions and three kinds of atomic parameters. Total 205 atomic parameters were optimized with the standard genetic algorithm using the training set consisting of 139 organic molecules with varying shapes and functional groups. The LogP values estimated with the two optimized solvation free energy functions compared reasonably well with the experimental results with the associated squared correlation coefficient and root mean square error of 0.824 and 0.697, respectively. Besides the prediction accuracy, the present method has the merit in practical applications because molecular LogP values can be computed straightforwardly from the simple potential energy functions without the need to calculate various molecular descriptors. The methods for enhancing the accuracy of the present prediction model are also discussed.

© 2015 Elsevier Inc. All rights reserved.

1. Introduction

1-octanol/water partition coefficient (P) is defined as the ratio of the concentration of a neutral molecule in 1-octanol to that in water in a two-phase system at equilibrium. Since the first use of 1-octanol/water partitioning system by Collander [1], it has served as a fundamental physicochemical property related with cell permeability, metabolism, bioavailability, and toxicity of molecules. For example, the molecular lipophilicity can be quantified by the logarithm (to base 10) of P, which is one of the most important molecular design parameters [2]. Besides the role of the indicator for lipophilicity, molecular LogP values have also been useful for the estimation of the desolvation cost for binding of a ligand to the receptor protein [3]. LogP is indeed the most widely used molecular descriptor in contemporary drug discovery.

Actually the LogP value of a compound can be determined in a straightforward way by the shake-flask method or reverse phase high performance liquid chromatography. This has made it possible that the experimental LogP values for a large number of simple organic molecules are available in public chemical databases such as PubChem and ChEMBL. However, it has become difficult to cope with all the molecules in the experimental determination of LogP because of the rapid increase in the number of compounds due for example to the advent of combinatorial chemistry [4]. It is therefore, necessary to develop the fast and accurate theoretical methods for estimating the LogP values of organic molecules.

Accordingly, a variety of theoretical/computational methods for estimating the molecular LogP have been proposed and explored since the pioneering work of Hansch et al. [5]. The most popular methods for estimating LogP include the CLogP and ALogP methods in which the LogP value of a molecule is computed by the summation of all contributions made by the dissected fragments and the individual atoms, respectively [6,7]. To supplement the deficiencies of the fragment and atom-based methods, a whole-molecule approach using the topological indices was proposed as implemented in the MLogP method [8]. Various quantitative structure-property relationship (QSPR) models with high accuracy were also suggested with the classical [9–11] and some novel molecular descriptors such as the semi-empirical electrotopological index [12], intramolecular interactions between functional groups [13], and SMILES-based optimal descriptors [14]. Very recently, Daina et al. developed a rigorous method for predicting LogP by combining the generalized Born and solvent accessible surface area models [15], which was referred to as iLogP method.

* Corresponding author. Fax: +82 2 3408 4334.
E-mail address: hspark@sejong.ac.kr (H. Park).

http://dx.doi.org/10.1016/j.jmgm.2015.06.004
1093-3263/© 2015 Elsevier Inc. All rights reserved.

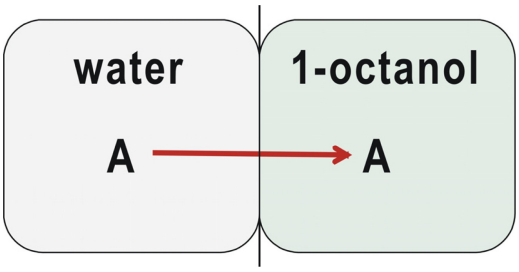

Fig. 1. Diffusion reaction of molecule A from water to 1-octanol solvent as the model system to calculate LogP.

Besides the high accuracy in estimating the molecular LogP values, iLogP method was also found to be adequate for coping with a large number of molecules in reasonably short time frame despite the requirement for computing various molecular descriptors. This indicated the usefulness of iLogP method in practical applications of drug discovery.

The present study is undertaken with the aim to establish a potential function from which molecular LogP values can be calculated in a straightforward way based on the difference in solvation free energies of a molecule with respect to water and 1-octanol solvents. For this purpose, P is assumed to be the equilibrium constant for the diffusion reaction of a molecule from water to 1-octanol solvent. Because the difference between the free energy of a molecule in water and that in 1-octanol should be calculated in this model prior to the estimation of LogP, we define the molecular solvation free energy functions for the two solvents based on the extended solvent-contact model. The present method is expected to be useful for coping with a large compound library in the early stage of drug discovery because molecular LogP values can be obtained directly using 3D atomic coordinates and the optimized atomic parameters without any additional calculation.

2. Theory and computational methods

2.1. Relation between partition coefficient and solvation free energy

Because P is defined as the ratio of molecular concentration in 1-octanol to that in water at equilibrium, it can be expressed as the equilibrium coefficient for the diffusion reaction of a molecule from water to 1-octanol solvent, which is illustrated in Fig. 1. Within this framework, LogP of a molecule can be expressed in the following form.

$$\mathrm{LogP} = \frac{\Delta G^0}{2.303RT} \tag{1}$$

Because ΔG^0 denotes the difference between the free energy of the solute molecule in 1-octanol and that in water, it can be approximated as the difference in solvation free energies of the solute with respect to the two solvents. LogP of the solute molecule at room temperature (298.15 K) is then expressed as follows in the unit of kcal/mol for ΔG.

$$\mathrm{LogP} = \frac{\Delta G_s^{\mathrm{wat}} - \Delta G_s^{\mathrm{oct}}}{1.364} \tag{2}$$

Here $\Delta G_s{}^{\mathrm{wat}}$ and $\Delta G_s{}^{\mathrm{oct}}$ denote solvation free energies of the solute molecule in water and 1-octanol, respectively.

To make it possible to calculate LogP using Eq. (2), we define the molecular solvation free energy functions in the two solvents based on the solvent-contact model [16,17]. As detailed in our previous papers on the extended solvent-contact model [18,19], the solvation free energy of a molecule can be obtained with the interatomic distances (r_{ij}'s) between solute atoms and the three atomic parameters.

$$\Delta G_s^{\mathrm{wat}} = \sum_i^{\mathrm{atoms}} S_i^{\mathrm{wat}} \left(O_{i,\max}^{\mathrm{wat}} - \sum_{j \neq i}^{\mathrm{atoms}} V_j e^{-\frac{r_{ij}^2}{2\sigma^2}} \right) \tag{3}$$

$$\Delta G_s^{\mathrm{oct,pure}} = \sum_i^{\mathrm{atoms}} S_i^{\mathrm{oct}} \left(O_{i,\max}^{\mathrm{oct}} - \sum_{j \neq i}^{\mathrm{atoms}} V_j e^{-\frac{r_{ij}^2}{2\sigma^2}} \right) \tag{4}$$

Here, we assume that the solute molecule can be stabilized in solution through the coordination between the intermolecular solvent-solute interactions and the intramolecular interactions between solute atoms. The gaussian-type envelope function is employed in the potential solvation free energy functions to reflect the effects of all surrounding atoms in the solute on the solvation of each atom in the distance-dependent manner. The key atomic parameters introduced in the solvation free energy function are the atomic solvation (S_i) energy per unit volume, the maximum atomic occupancy ($O_{i,\max}$), and the atomic fragmental volume (V_i). The negative and positive signs of S_i parameter indicate the stabilization and destabilization of the solute atom i, respectively, due to the combined effects of intermolecular interactions with solvent molecules and intramolecular interactions with the rest of solute atoms. The optimizations of these three atomic parameters for all possible atom types are required to calculate the solvation free energies of the solute with respect to water and 1-octanol, both of which are needed to estimate the LogP values. In the calculation of solvation free energies, we used the same V_i parameters for water and 1-octanol under the assumption that the change of the solvent would have little effect on the molecular volume of the solute molecule. On the other hand, S_i and $O_{i,\max}$ values were optimized separately in the two solvents using the corresponding experimental data for solvation free energies of the molecules in the data set.

Actually the 1-octanol phase is a mixture containing 96% of 1-octanol and 4% of water in the shake flask experiment. Therefore, we used the composition-weighted free energy function given by Eq. (5) to calculate the reference solvation free energy data for 1-octanol required for the estimation of LogP values.

$$\Delta G_s^{\mathrm{oct}} = 0.96\Delta G_s^{\mathrm{oct,pure}} + 0.04\Delta G_s^{\mathrm{wat}} \tag{5}$$

2.2. Preparation of training and test set

To obtain the complete forms of solvation free energy functions, we needed a reference data set with which all the atomic parameters could be optimized. Therefore, we constructed a chemical library containing 178 organic molecules for which experimental solvation free energy values were available for both solvents [20]. These 178 molecules were categorized into 39 subsets according to the structural similarity. More specifically, the structurally similar molecules with Tanimoto coefficient larger than 0.7 were collected into the same cluster. Structural similarities among the molecules were measured using the fingerprints of each molecule generated with the Daylight software as an ASCII string of 1's and 0's. For a subset containing n molecules, a single representative one was randomly selected with the probability $1/n$ as an element of the test set. The resultant training and test sets comprising 139 and 39 molecules, respectively, were utilized to build up and validate the LogP prediction method within the framework of solvent-contact model. The structures of the molecules selected as the elements of training and test set are shown in part in Figs. 2 and 3, respectively.

Download English Version:

https://daneshyari.com/en/article/442894

Download Persian Version:

https://daneshyari.com/article/442894

Daneshyari.com