



Benchmark of different charges for prediction of the partitioning coefficient through the hydrophilic/lipophilic index

Oksana Fizer^{1,2} · Maksym Fizer¹ · Vasyly Sidey³ · Yaroslav Studenyak¹ · Ruslan Mariychuk²

Received: 8 March 2018 / Accepted: 22 May 2018
© Springer-Verlag GmbH Germany, part of Springer Nature 2018

Abstract

A few different theoretical methods for assigning the partial atomic charges were benchmarked for calculation of the hydrophilic/lipophilic index (HLI). The coefficients were selected to produce the best correlation of the HLI values with the experimental octanol-water partition. Different parameters were checked in calculations of partial charges to get the best performance of the HLI values obtained. Thus, four partitioning schemes (Coulson, Mulliken, Merz-Kollman, Ford-Wang) were benchmarked for calculations of atomic charges with six semiempirical methods (AM1, PM3, RM1, PM6, PM6-D3H4, PM7). Moreover, five distinct types of partial atomic charges (Mulliken, Hirshfeld, Löwdin, CHELPG, NPA), obtained at the Hartree-Fock and DFT levels of theory with three basis sets, were tested for their ability to produce the HLI values with the best correlation to experimental logP coefficients of 50 mono-charged organic anions. In the case of the semiempirical methods, the best correlation between the HLI and logP values (the correlation coefficient $r = 0.9216$) was obtained with the AM1 Ford-Wang parametric electrostatic potential charges. The Mulliken and Coulson charges calculated with the PM7 method can be used as an alternative to AM1, with the r values of 0.9107 and 0.8984, respectively. In the case of the DFT, the PBE/def2-TZVP natural population analysis charges produce the best correlation ($r = 0.9220$). Nevertheless, in spite of a marginally lower performance ($r = 0.9159$), the NPA charges computed at the PBE/def2-SVP level are more robust and can be regarded as the optimum choice for calculating the HLI values.

Keywords logP · Partitioning coefficient · Partial charge · Semiempirical · DFT

Introduction

One of the most widely used descriptors of molecule's lipophilicity is logP, which is the logarithm of the partition coefficient between octanol and water [1]. Being the part of the

famous "Lipinski's rule of five" [2], this descriptor is an extremely important parameter in the drug discovery process. logP of a compound can be measured directly by partitioning between octanol and water, or it can be measured indirectly by chromatography, potentiometry or some other procedures [3–6]. However, depending upon research needs, the experimental measurements of logP are often replaced with the theoretical estimations, especially if these measurements are difficult or impossible at all, as in the cases of labile and unstable compounds. Furthermore, in the case of surfactant-like molecules, the difficulties are primarily associated with the formation of emulsions during the shaking of compounds in two-solvent systems. The above problems could explain why the computational approaches for determining or predicting the partition coefficients are still in the scientific focus [7–11]. Nowadays a huge number of the experimentally determined logP values of organic structures are reported in the literature, but this is still relatively small in comparison with the number of the compounds for which the logP values are of interest. The exhaustive list of the known and popular theoretical

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s00894-018-3692-x>) contains supplementary material, which is available to authorized users.

✉ Maksym Fizer
max.fizer@uzhnu.edu.ua

¹ Faculty of Chemistry, Uzhhorod National University, Pidhima Str. 46, Uzhhorod 88000, Ukraine

² Faculty of Humanity and Natural Sciences, University of Prešov in Prešov, 17th November 1, Prešov 08116, Slovak Republic

³ Research Institute for Physics and Chemistry of Solid State, Uzhhorod National University, Pidhima Str. 46, Uzhhorod 88000, Ukraine