Prediction of log P: ALOGPS Application in Medicinal Chemistry Education

ARTICLE *in* JOURNAL OF CHEMICAL EDUCATION · OCTOBER 2011

Impact Factor: 1.11 · DOI: 10.1021/ed100444h

4 AUTHORS, INCLUDING:

Jacek Kujawski
Poznan University of Medical Sciences
12 PUBLICATIONS **34** CITATIONS
SEE PROFILE

Marek Bernard
Poznan University of Medical Sciences
24 PUBLICATIONS **73** CITATIONS
SEE PROFILE

JOURNAL OF CHEMICAL EDUCATION

ARTICLE

pubs.acs.org/jchemeduc

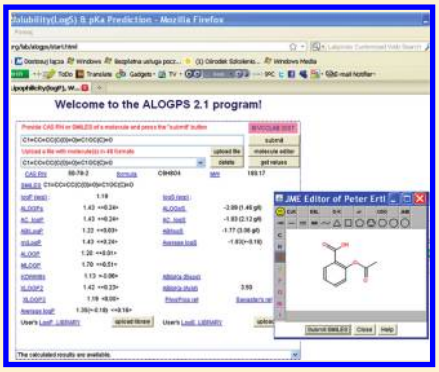
Prediction of log *P*: ALOGPS Application in Medicinal Chemistry Education

Jacek Kujawski,* Marek K. Bernard, Anna Janusz, and Weronika Kuźma

Department of Organic Chemistry, Faculty of Pharmacy, Poznan University of Medical Sciences, Grunwaldzka 6 str., 60-780 Poznań, Poland

Supporting Information

ABSTRACT: Molecular hydrophobicity (lipophilicity), usually quantified as log *P* where *P* is the partition coefficient, is an important molecular characteristic in medicinal chemistry and drug design. The log *P* coefficient is one of the principal parameters for the estimation of lipophilicity of chemical compounds and pharmacokinetic properties. The understanding of log *P* parameter in the undergraduate medicinal chemistry course seems to be a pitfall for students. This parameter has typically been measured using experimental methods, but recently, log *P* has been determined using computational methods. The number of publications about lipophilicity predictions has gradually increased over the last 10 years, but the number of programs available for an online prediction of this important parameter remains limited. An interesting tool for calculation of log *P* coefficients is presented: the Virtual Computational Chemistry Laboratory (VCCLAB) package. The package includes the ALOGPS 2.1 program suitable for log *P* calculations. This software is accessible online and may be easily mastered by the undergraduate medicinal chemistry student.



KEYWORDS: Second-Year Undergraduate, Chemoinformatics, Organic Chemistry, Internet/Web-Based Learning, Bioanalytical Chemistry, Drugs/Pharmaceuticals, Laboratory Computing/Interfacing

During transport to the receptor, a drug usually passes through lipid membranes. Thus, the relative drug distribution between aqueous and nonpolar media is of considerable interest. The molecular hydrophobicity (lipophilicity) is normally quantified as log *P* where *P* is the partition coefficient obtained by measuring the drug distribution between two immiscible solvents, usually 1-octanol and water because 1-octanol properties are similar to those of natural membranes.[1] The octanol/water coefficient, *P*, is the ratio of a neutral molecule concentration in 1-octanol to its concentration in water when the phases are at equilibrium. The obtained values are consistent for nonionizable compounds. For charged substances that have greater water solubility than can be predicted from the neutral structure, often the term log *D* is used to describe the lipophilicity. The distribution coefficient, *D*, is calculated for the partition of a drug between 1-octanol and aqueous buffer. Both the partition and distribution coefficients are measures of how hydrophilic (water loving) or hydrophobic (water fearing) a chemical substance is. The hydrophobic drugs with high partition coefficients are preferentially distributed to hydrophobic compartments such as lipid bilayers of cells, whereas hydrophilic drugs of low partition coefficients are preferentially localized in hydrophilic compartments such as blood serum. The optimal lipophilicity range along with low molar mass and low polar surface area is the driving force that leads to good absorption of chemicals in the intestine by passive diffusion. That is why the log *P* coefficient is one of the principal parameters that estimates lipophilicity of chemical compounds and, to a large degree, indicates the pharmacokinetic properties. It is also used as one of the standard properties identified by Lipinski in the "rule of 5" for drug-like molecules.[2] It can be measured using known experimental methods,[3−6] but recently, computational chemistry (in silico) methods have widely been applied. The first method of log *P* calculation was developed by Hansch, Fujita, and Iwana.[6] Despite the incredible growth of the Internet, the number of practical online applications in drug design remains limited, particularly for predictions of drug-like compounds. For example, the number of methodological publications about lipophilicity predictions has gradually increased over the last 10 years, but the number of programs available for online prediction of this important property includes few applications.[7] Methods for log *P* calculation can be divided roughly into two major classes: the substructure-based methods, and the whole-molecule approaches.[8]

If a molecule contains basic or acidic groups, it becomes ionized and its distribution in octanol/water is pH dependent. At physiological pH, many basic or acidic drugs are ionized, and the partition coefficient is the distribution coefficient, *D*, which is generally accepted as the distribution between an aqueous buffer at pH 7.4 and 1-octanol. This distribution coefficient for monoprotic bases is defined as $\log D_{oct} = \log P_{oct} + \log [1/(1 + 10^{pKa - pH})]$. For

Published: October 06, 2011

Table 1. Examples of Some Web Resources in Chemoinformatics

No.	Name[a]	Provides	Link[b]
1	Corina	2D to 3D conversion of molecular structures	http://www.molecular-networks.com/software/corina/
2	Osiris	log *P*, solubility, toxicity, drug likeness	http://www.organicchemistry.org/prog/peo
3	VCCLAB	molecular descriptors, physicochemical properties	http://www.vcclab.org
4	Pre-ADMET	molecular descriptors and various ADME/T properties	http://preadme.bmdrc.org/preadme

[a] Information is adapted from ref 7. [b] All URLs accessed Sep 2011.

monoprotic acids, the equation has the same form, except that the exponent is written as "pH − pK_a". For polyprotic compounds, the equation becomes more complicated and is modified accordingly to incorporate correction terms for all of the ionized forms. Thus, log *D* prediction potentially accumulates errors due to the log *P* and pK_a predictions.[1] It is valuable to study log *P* to predict recognition and interactions between biological molecules because (i) log *P* is essentially an experimentally reproducible measurement; (ii) the partition experiments are inexpensive and can be performed relatively rapidly; and (iii) log *P* is directly related to the free energy of binding and solvation−desolvation effects.

DISCUSSION

The log *P* coefficient is a measure of lipophilicity, and many pharmaceutical and biological events are dependent on their lipophilic characteristics. Web resources are presented that can be applied by undergraduates for calculating log *P* coefficients for many organic derivatives. These tools can be helpful for rationalized drug design as the key step in medicinal chemistry education. From the examples of free computational Web resources in chemistry connected with this topic (Table 1), a user-friendly student package named Virtual Computational Chemistry Laboratory (VCCLAB), with emphasis on the ALOGPS 2.1 program,[9] is highlighted. The Web site of VCCLAB provides free online chemoinformatics tools, including the building and visualization of chemical structures, the calculation of molecular properties, and the analysis of relationships between chemical structure and properties. The ALOGPS 2.1 software provides an interactive online prediction of log *P*, water solubility, and pK_a of compounds for drug design and environmental chemistry studies. The software may be used in the classroom or laboratory and also in the dormitory, home, or local campus computer lab. Thus, it is a good supplementary reference for the undergraduate chemistry course and provides a new teaching method for the medicinal chemistry course.

The ALOGPS 2.1 program is built on the Associative Neural Network (ASNN) pattern, as a new challenge for the development of physicochemical data prediction methods.[10,11] In addition, in the parametrization of solubility prediction tools, some databases can be used, for example, PHYSPROP or AQUASOL.[12−15] To perform the log *P* calculations, the ALOGPS user can draw the molecule using the JME applet[16] or submit it in a format supported by freely available software such as OpenBabel.[17] A non-Java interface in the ALOGPS 2.1 package for structure submission is available as well. Moreover, the application calculates and compares lipophilicity and aqueous solubility of molecules using several methods including those described in references[16,18−21] (see Supporting Information), which definitely increases ALOGPS's practical viability. The prediction ability using ALOGPS 2.1 can also be increased in the library mode.[22]

HOW DOES IT WORK? AN EXAMPLE

To submit a compound into ALOGPS 2.1 program, the user needs to provide the compound SMILES code or enter its CAS registry number. These codes can be stored in databases, such as iResearch library[23] or ZINC.[24] To generate the SMILES code of a molecule, the student can use, for instance, the popular ACD/ChemSketch application.[25] For ionizable substances such as acetylsalicylic acid (aspirin, ASA, pK_a = 3.48), a well- known anti-inflammatory drug that can be easily synthesized by students,[26,27] the log *P* value is valid only at pH < 3; otherwise, the salt of acetylsalicylic acid becomes hydrophilic. Therefore, the log *D* parameter should be used for full characteristic of hydrophilic−hydrophobic properties of ASA. Because the problem of predicting log *D* is more complicated and it is computed from log *P* and pK_a, only the log *P* coefficient of ASA will be discussed here. After submission of the resulted SMILES code, the data from ALOGPS can be compared with that from other calculation methods, for example, ALOGPS = 1.43, MiLogP = 1.43, KOWWIN = 1.13, XLOGP2 = 1.42, XOLOGP3 = 1.19 (log *P* experimental = 1.19).[9] These differences are due to the calculation methods used in each of the different programs. The ALOGPS program, which is based on topological descriptors,[1,28−31] has been used in a series of studies on a variety of molecules, including libraries of drugs.[28,29] Although the program was developed to predict the partition coefficient log *P* for neutral compounds, it is user-friendly because it works in a completely automated fashion and does not require any user intervention or extended knowledge in computational chemistry.[28]

To assess the efficacy of ALOGPS, the program was made available to students attending the organic chemistry course for the second-year pharmacy students. The details of the student activity and assessment are described. In the first graded exercise, the students were given several structures taken from the PHYSPROP database;[13] two compounds were obtained by our scientific group[32,33] in a multistep synthesis. Based on the compounds, students were supposed to answer the yes−no question about the given calculated log *P* values (see Table 2 in the Supporting Information). To do that, the SMILES notification of 13 substances was required and was generated using ACD/ChemSketch application[25] according to the procedure discussed above. The results of the first graded exercise were 73% of the students passed and 27% of the students failed (N = 24). In the second exercise, the students were asked about the influence of the R group (Figure 1) on the log *P* coefficient. The question was which of the resulting derivatives became more hydrophilic or lipophilic in comparison with the reference compound, R = H (see Table 3 in the Supporting Information). That exercise examined a possibility of modification of pyrazole ring as the part of strategy in the organic synthesis of more hydrophilic or lipophilic analogues. The results of the second exercise were 98% positive answers. That exercise and the ALOGPS package as an example of SAR (structure−activity relationship) in medicinal

dx.doi.org/10.1021/ed100444h |J. Chem. Educ. 2012, 89, 64–67

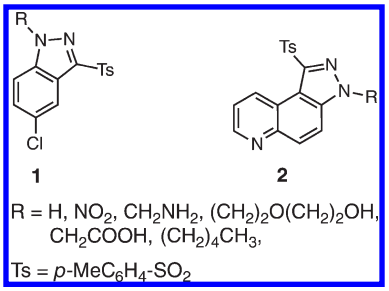


Figure 1. Structure of fused pyrazole derivatives used in the second exercise.

chemistry were then used to explain why some drugs containing electron-withdrawing substituents, such as the nitro group (e.g., nitro derivatives of benzodiazepines), were more lipophilic, able to diffuse over the blood–brain barrier, and to work longer in the human organism. It was also emphasized that oral drugs should have their lipophilicity between 1 and 4 on the log *D* scale to be absorbed by passive diffusion.[1] Because of this reasoning, the acetyl derivative of salicylic acid is used in oral treatment and is metabolized in the organism to the active salicylic acid.

PERSPECTIVES

Accurate prediction of log *P* is important for the pharmaceutical industry. Methods of log *P* prediction have attracted increasing interest during in the past decade. Given the benefits brought to bioinformatics by Web applications, it is attractive to encourage the development of these technologies in the chemoinformatics field. The Internet increases awareness about the existing software. The appearance of new protocols and standards for data sharing on the Web makes development of new applications easy and straightforward. The VCCLAB can be used as a prototype useful for developing such projects. The developed technology allows for integration of new third-party applications, which could be made available to the worldwide community. Fields such as chemistry and pharmacy benefit from having more chemists and pharmacists aware of Web software to reduce costly repetitions of work already done and to advance medicinal chemistry further and faster; VCCLAB, especially the ALOGPS 2.1 program, can be used as a springboard for research and development. With these tools becoming more accessible, students should be made aware of this information. Incorporating this program into instruction can enlarge the knowledge of undergraduate student on the chemoinformatics field and provide the medicinal chemistry instructor with interesting new material to incorporate across a range of classes. QSAR (quantitative structure–activity relationship) approaches such as ALOGPS, can improve prediction ability by self-learning on user-specific data and may also find significant application in the pharmaceutical industry in the near future.

ASSOCIATED CONTENT

Supporting Information

Materials used by students (handouts) and exercise examples. This material is available via the Internet at http://pubs.acs.org.

AUTHOR INFORMATION

Corresponding Author

*E-mail: jacekkuj@ump.edu.pl.

ACKNOWLEDGMENT

Support was partially provided by SBN Grants (Poznan University of Medical Sciences): 10/2009, 60/2010, as well as grant No. 502-143308417-08345 (Poznan University of Medical Sciences).

REFERENCES

(1) *Molecular Drug Properties: Measurement and Prediction*; Mannhold, R., Ed.; Wiley-VCH Verlag: Weinheim, 2008.
(2) Lipinski, C. A.; Lombardo, F.; Dominy, B. W.; Feeney, P. J. *Adv. Drug Delivery Rev.* **2001**, *46*, 3–26.
(3) Lombardo, F,M.; Shalayeva, M. Y.; Tupper, K. A.; Gao, F. *J. Med. Chem.* **2001**, *44*, 2490–2497.
(4) Ulmeanu, S. M.; Jensen, H.; Bouchard, G.; Carrupt, P. A.; Girault, H. H. *Pharm. Res.* **2003**, *20*, 1317–1322.
(5) Bond, A. M.; Marken, F. *J. Electroanal. Chem.* **1994**, *372*, 125–135.
(6) Fujita, T.; Iwana, J.; Hansch, C. J. *J. Am. Chem. Soc.* **1964**, *86*, 5175–5180.
(7) Tetko, I. V. *Drug Discovery Today* **2005**, *10*, 1497–1500.
(8) Mannhold, R.; Waterbeemd, van H. *J. Comput.-Aided Mol. Des.* **2001**, *15*, 337–354.
(9) Virtual Computational Chemistry Laboratory (VCCLAB), http://www.vcclab.org/ (accessed Sep 2011).
(10) Tetko, I. V.; Gasteiger, J.; Todeschini, R.; Mauri, A.; Livingstone, D.; Ertl, P.; Palyulin., V. A.; Radchenko, E. V.; Zefirov, N. S.; Makarenko, A. S.; Tanchuk, V. Y.; Prokopenko, V. V. *J. Comput.-Aided Mol. Des.* **2005**, *19*, 453–463.
(11) Tetko, I. V. Associative Neural Network, CogPrints archive code: cog00001441, http://www.vcclab.org/lab/alogps/library.html (accessed Sep 2011).
(12) Tetko, I. V. *J. Chem. Inf. Comput. Sci.* **2002**, *42*, 717–728.
(13) Syracuse Research Corporation. Physical/Chemical Property Database (PHYSPROP); SRC, Environmental Science Center: Syracuse, NY.
(14) Tetko, I. V.; Tanchuk., V. Y.; Villa, A. E. *J. Chem. Inf. Comput. Sci.* **2001**, *41*, 1407–1421.
(15) Stahura, F. L.; Godden, J. W.; Bajorath, J. *J. Chem. Inf. Comput. Sci.* **2002**, *42*, 550–558.
(16) Molinspiration. http://www.molinspiration.com (accessed Sep 2011).
(17) Open Babel as the Open Source Chemistry Toolbox. http://openbabel.sourceforge.net (accessed Sep 2011).
(18) BioByte. http://www.biobyte.com/bb/prod/clogp40.html (accessed Sep 2011).
(19) SRC KOWWIN. http://www.syrres.com/search.aspx?searchtext=KOWWIN &folderid=0&searchfor=all&orderby=id&orderdirection=ascending (accessed Sep 2011).
(20) Virtual logP. http://nova.colombo58.unimi.it/vlogp.htm http://www.logp.com (accessed Sep 2011).
(21) Wang, R. *Perspect. Drug Discovery Des.* **2000**, *19*, 47–66.
(22) The Library Mode of the platform VCCLAB. http://www.vcclab.org/lab/alogps/library.html (accessed Sep 2011).
(23) ChemNavigator. http://www.chemnavigator.com (accessed Sep 2011).
(24) Irvin, J. J.; Shoichet, B. K. *J. Chem. Inf. Model.* **2005**, *45*, 177–182.
(25) Advanced Chemical Development (ACD) as ACD/log P freeware, www.acdlabs.com/download/ (accessed Sep 2011).
(26) Montes, I.; Sanabria, D.; García, M.; Castro, J.; Fajardo, J. *J. Chem. Educ.* **2006**, *83*, 628–631.
(27) Olmsted, J. A., III. *J. Chem. Educ.* **1998**, *75*, 1261–1263.
(28) Tetko, I. V.; Poda, I. *J. Med. Chem.* **2004**, *47*, 5601–5604.
(29) Mannhold, R.; Poda, G. I.; Ostermann, C.; Tetko, I. V. *J. Pharm. Sci.* **2009**, *98*, 861–893.
(30) *Topological Indices and Related Descriptors in QSAR and QSPR*; Devillers, J; Balaban, A. T., Eds.; Gordon and Breach Scientific Publishers: Amsterdam, 1999.

(31) *Molecular Descriptors for Chemoinformatics*, 2nd ed.; Todeschini, R., Consonni, V., Eds.; Wiley-VCH: Weinheim, 2009.
(32) Bernard, M. K.; Kujawski, J.; Janusz, A.; Kuźma, W.; Totoń, E.; Wierzchowski, M.; Rybczyńska, M. Synthesis of some fused aromatic compounds with potential antitumor activity. In 5th German-Polish Symposium "New challenges for pharmaceutical sciences", May 15−16, 2009; Book of Abstracts, pp 29.
(33) Mąkosza, M.; Wojciechowski, K. *Chem. Rev.* **2004**, *104*, 2631–2666.