COMPUTATIONAL METHODS IN SCIENCE AND TECHNOLOGY **18(2)**, 81-88 (2012)

The log *P* Parameter as a Molecular Descriptor in the Computer-aided Drug Design – an Overview

Jacek Kujawski*, Hanna Popielarska, Anna Myka, Beata Drabińska, Marek K. Bernard

Department of Organic Chemistry, Faculty of Pharmacy, Poznan University of Medical Sciences, Grunwaldzka 6 str., 60-780 Poznań
**e-mail: jacekkuj@ump.edu.pl*

(Received: 10 February 2012; accepted: 6 June 2012; published online: 27 August 2012)

Abstract: The computer-aided drug design is an important tool in modern medicinal chemistry. Molecular lipophilicity, usually quantified as log *P*, is an important molecular characteristic in medicinal chemistry and also in rationalized drug design. The log *P* coefficient is well-known as one of the principal parameters for the estimation of lipophilicity of chemical compounds and determines their pharmacokinetic properties. This parameter has been measured using known experimental methods, but recently huge progress in determination of log *P* using computational chemistry methods is observed. The number of methodological publications about lipophilicity predictions has gradually increased over the last ten years, but the number of programs available for an on-line prediction of this important parameter remains limited. This paper presents some of log *P* prediction methods and very popular programs connected to this topic. The prediction of log *P* is highly important for the pharmaceutical industry since it limits time-consuming experiments to measure log *P* required to optimize pharmacodynamic and pharmacokinetic properties of hits and leads. Development of the methods reviewed in this paper concerning log *P* prediction seems to be a significant tendency in the modern pharmaceutical industry.

Key words: chemoinformatics, computer-aided design, computer program, predictions and projections, pharmaceutical chemistry

I. INTRODUCTION

Bioavailability of a drug and its access to a therapeutic target are important considerations in a rational drug design. Before the drug can elicit an effect it usually has to pass through a series of barriers like biological membranes either by passive diffusion or carrier-mediated uptake. The affinity of a drug molecule for a target of interest and its ability to partition into a lipophilic environment at different pH values has to be quantified for a proper prediction of its ability to interact with the biological target and hence be efficacious. For many years the *n*-octanol/water partition coefficient has been used as a measure of *lipophilicity/hydrophobicity*, where *hydrophobicity* describes the ability for aggregation of organic compounds in water, while the *lipophilicity* is determined by intermolecular relationships between an organic substance and solvent [1]). This coefficient is usually quantified as log *P* and is an important molecular characteristics in medicinal chemistry and computer-aided drug design (CADD) as well [1]. Lipophilicity is a main physico-chemical determinant influencing the bioavailability, permeability and frequently the toxicity of drugs. Thus the octanol/water partition coefficient is the ratio of an unionized compound concentration in *n*-octanol to its concentration in water when the phases are at equilibrium. Hence these coefficients are a measure of differential solubility of the compound between these two solvents. Usually one of the chosen solvents is water, while the second one is hydrophobic, such as octanol. Both the partition and distribution coefficient are measures of how hydrophilic ("water loving") or hydrophobic ("water fearing") a chemical substance is. Partition coefficients are useful, e.g. in the estimation of drug distribution within the organism. The hydrophobic drugs with high partition coefficients are preferentially distributed to hydrophobic compartments such as lipid bilayers of cells, while hydrophilic drugs (low partition coefficients) are preferentially localized in hydrophilic compartments, such as blood serum. Optimal lipophilicity range along with low molecular weight and low polar surface area is the major driving force that leads to good absorption of chemicals, i.e. in the intestine by passive diffusion. That is why the log *P* coefficient is well-known as one of the principal parameters which estimates

lipophilicity (or solubility in lipids) of chemical compounds and, to a large degree, determines their pharmacokinetic properties. It is also used as one of the standard properties identified by Lipinski in the "rule of five" for 'drug-like' molecules. The Lipinski rule is actually the following four rules with cut-off numbers that are 5 or multiples of 5: (i) hydrogen bond donors (sum of hydroxyl and amine groups) less than 5, (ii) hydrogen bond acceptors (sum of nitrogen and oxygen atoms) less than 10, (iii) a molecular weight under 500 daltons, (iv) a log *P* coefficient of less than 5) [2]. It can be measured using experimental methods [3-5], but since Hansch and Fujita [6] invented the first method for log *P* calculation, a variety of computational procedures, including *in silico* methods, has been developed for the log *P* prediction. Despite the incredible growth of methodological publications about lipophilicity, the number available for on-line prediction of this important property is in the range of few applications [7]. The methods for log *P* calculation can be divided roughly into two major classes [8]:

a) the *substructure-based* methods, which have in common the observation that molecules are disconnected into atoms (*atom contribution methods*) or groups (*fragmental methods*); by summing the single-atom or fragmental contributions (supplemented by applying correction rules in the latter case) finally log *P* can be achieved;
b) the *whole molecule approaches*, which inspect the entire molecule using *molecular lipophilicity potentials (MLP), topological indices* or *molecular properties* to quantify log *P*.

II. Log *P* PREDICTION – METHODS OF CALCULATIONS, EXAMPLES OF PROGRAMS AND TECHNIQUES

If a molecule contains basic or acid 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 indeed a distribution coefficient *D*, which is generally taken to be the distribution between an aqueous buffer at pH 7.4 and *n*-octanol. This distribution coefficient (in the form of its logarithm) for monoprotic bases is defined as: $\log D_{oct} = \log P_{oct} + \log [1/(1 + 10^{pKa - pH})]$. For 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 separate correction terms for each compound. Thus, log *D* prediction potentially accumulates errors due to the log *P* and pK_a predictions [1]. For this reason in our paper the log *P* coefficient prediction will be discussed. Why should we use log *P* to study and predict recognition and interactions between biological molecules? At least three reasonable answers could be possible: (i) log *P* is essentially an experimental reproducible measurement; (ii) partition experiments are cheap and relatively no time-consuming; and (iii) log *P* is directly related to the free energy of binding and solvation/desolvation effects.

Why is the *n*-octanol used in a log *P* estimation process? As mentioned above, the properties of *n*-octanol are regarded as being similar to those of lipid bilayer membranes. However, distribution of chemicals into *n*-octanol simulates, to certain extent, their ability to passively diffuse across biological membranes, but it is well known that lipophilicity is somewhat dependent on the hydrogen bonding ability of the analyte and solvent used. Unlike *n*-octanol, cyclohexane cannot form hydrogen bonds and for this the reason water-cyclohexane distribution might closer resemble the blood-brain barrier partitioning behavior [9].

As shown above, the first "by substituent" approach was proposed by Fujita and coworkers in 1964 [6]. Their technique is based on the following equation:

$$\pi = \log P_X - \log P_H, \quad (1)$$

where P_X represents the partition coefficient between *n*-octanol and water and P_H that of the parent compound. The log *P* parameter represents an additive-constitutive, free energy-related property, numerically equivalent to the sum of the parent log *P* compound representing the log *P* difference between a determinate substituent and a hydrogen atom which has been replaced [10]. As an example, the log *P* determination for the methyl group is given herein, namely:

$$\log P_{\text{methyl group}} = \log P_{\text{toulene}} - \log P_{\text{benzene}} \quad (2)$$

The following "by fragments" methods were supported by Rekker and Mannhold, who assumed that log *P* can be calculated as the sum of a fragment values plus certain correction factors. They determined the averaged contributions of simple fragments, using a large database of experimentally measured log *P* values [11, 12]. Rekker did not indicate which fragment could be considered a valid fragment. The log *P* of molecules can be calculated using the formula (Eq. (3)):

$$\log P = \Sigma anfn + \Sigma bmFm, \quad (3)$$

where *a* is the number of occurrences of fragment *f* of type *n,* while *b* is the number of occurrences of correction factor *F* of type *m*.

The well-known CLOGP method [1, 13] clearly represents an improvement of the Rekker approach and, in fact, can be expressed by the same equation. The CLOGP program breaks molecules into fragments and sums these constant fragment values and structure-dependent correction values taken from Hansch and Leo's database, to predict log *P* of several organic molecules. The program divides the target molecule into different fragments following a set of simple rules unalterable by users. The CLOGP represents the first stand-alone program developed by Pomona MedChem, following Rekker general formulation.

Different from chemical group fragments, methods based on atomic contribution and/or surface area use atomic fragments and surface area data to predict hydrophobicity. The contribution of each atom to a molecule, in terms of hydrophobicity, can be evaluated by multiplying the corresponding atomic parameter by the degree of exposure to the surrounding solvent. The exposure degree is typically represented by the *solvent-accessible surface area* (SASA). The first promoters of this method were Broto with his colleagues, who developed a 222 descriptors set, made by combinations of up to four atoms with specific bonding pathways up to four in length, reaching precision of about 0.4 log units [14]. Later, the concept of SASA was used by Iwase [15] and Dunn [16] in principal component analysis to improve their log *P* estimations. Dunn computed an isotropic surface area, calculating the number of water molecules able to hydrate the polar portions of solute molecules. As an example, one water molecule was allowed for groups as nitro, aniline, ketones, and tertiary amines, while two waters are allowed for other amines, three for carboxyls, and five for amide groups. The use of SASA parameters has been extended and introduced in several log *P* calculation algorithms, like the HINT (**H**ydropathic **INT**eractions) program created by Abraham [17], which is able to directly calculate hydrophobic atomic constants for small molecules or to obtain them from a residue-based dictionary. This program was thus created with the purpose of rapid and proper estimation of biological interactions such as protein-protein, protein-DNA, and protein-igand. The HINT application can be defined as a natural and intuitive force field, helpful for calculation, using experimentally determined log *P* values, enthalpic, and entropic effects included in noncovalent interactions, like hydrogen bonding, Coulombic forces, acid-base and hydrophobic contacts. Hydrophobic and polar contacts are strictly related to solvent partitioning phenomena. In fact, the solubilization of a ligand in a mixed solvent system, like water and octanol, involves the same processes and atom–atom interactions as biomolecular interactions within or between proteins and ligands [18]. It was designed to consider and investigate hydrophobicity and hydropathic interactions in several biological areas. The HINT is able to: (i) calculate hydrophobic atomic constant for each atom in a small molecule or even in a macromolecule, and quantitatively score molecular interactions, (ii) create hydrophobic maps or fields for small molecules in protein environments, (iii) map the hydrophobic and polar nature of the surrounding receptor from the structure of small interacting molecules, providing a hydrophobic interaction template for a definition of secondary and tertiary protein structure, and (iv) suggest modes of inter-helix interactions in a trans-membrane ion channel [19]. All these features and capabilities make HINT a suitable tool, not only for the study of single and simple interactions, but also for a virtual screening of organic libraries and for *structure-based drug design*. Interactions between atom–atom couples are calculated using the following equation:

$$b_{ij} = a_i S_i aj S_j T_{ij} R_{ij} + r_{ij}, \tag{4}$$

where b_{ij} represents the interaction score between atoms i and j, a is the hydrophobic atomic constant, S is the SASA, T_{ij} is a logic function assuming –1 or +1 value, depending on the character of the interacting polar atoms, while R_{ij} and r_{ij} are functions of the distance between atoms i and j. The whole interaction between two molecules, like protein and ligand, or protein and DNA, can be represented as:

$$\Sigma\Sigma\, b_{ij} = \Sigma\Sigma\, a_i S_i a_j S_j T_{ij} R_{ij} + r_{ij}, \tag{5}$$

$b_{ij} > 0$ identifies favorable interactions, while $b_{ij} < 0$ the unfavorable ones.

Interactions can be divided into: polar-polar, hydrophobic-hydrophobic, and hydrophobic-polar. While hydrophobic-hydrophobic contacts are always positively scored, polar interactions depending on the charge of interacting groups can be favorable (acid-base) or unfavorable (acid-acid and base-base). Hydrophobic–polar contacts are constantly negatively scored by HINT, so they negatively contribute to the global binding energy. The HINT software, compared to CLOGP, allows us to reduce the information from bulk molecule solvent partitioning, to discrete interactions between biological molecules, i.e., ligand-protein, protein-protein, protein-DNA, and protein-ligand-water [19].

It is well known that scheduled, prior synthesis of many organic biologically active compound requires *rationalized drug design* techniques, including *in silico* applications for log *P* prediction, e.g. discussed above: HINT, CLOGP, and others programs like: KOWWIN [20], MiLog P [21], IA_logP [22], XLOGP [23], ALOGPS [24], AB/LogP [25], ACD Lab/ChemSketch [26], to name a few. In some

Table 1. Substructure-based methods – applications examples. Modified from [25]

Fragmental methods	Atom-based methods
Reductionist approaches	ALOGP
KLOGP	MOLCAD, TSAR, PrologP, Dragon
KOWWIN	ALOGP98
Constructionist approaches	Accelrys Discovery Studio
CLOGP	OsirisP
ACD/Log P (from ACDLab/ChemSketch)	*Atom types and correction factors*
Hierarchical clustering	XLOGP2, XLOGP3
AB/LogP	

Table 2. Property-based methods – applications examples. Modified from [25]

Calculation methods	Examples of programs
Empirical approaches	
Molecular size and H-bond strenght	SLIPPER
Methods that used 3D-structures of molecules	
Quantum mechanical semi-empirical calculations	
Calculated quantum mechanical parameters	BLOGP
Molecular dynamics calculations	
Molecular lipophilicity potential (MLP)	HINT
Methods based on topological descriptors	
Electrotopological-state descriptors	
Associative Neutral Network (ANN) model using 75 E-state indices	ALOGPS

of them [25] the *substructure-based methods* (Table 1) are used, in others – *property-based approaches* are imple mented (Table 2). The MiLogP application is based on group contributions and was developed using 35 small basic fragments as well as 185 larger fragments characterizing intramolecular H-bonding contribution and charge interactions [25]. The KOWWIN application uses the *atom/fragment contribution method* (AFC) [25]. Its extension, the *Experimental Value Adjusted* (EVA) approach, starts from the measured log *P* for a structural analogue of the query compound. The analogue is modified by subtracting and adding fragments and factors to build the query compound. The 155 atoms/fragments are used in this software. In general, each nonhydrogen atom in a structure is the fragment core and the exact fragment is determined by the type of atoms connected to the core. The estimate is a sum of the experimental log *P* and the value of fragment/factor modifications. Such estimations are more accurate than using the original method [1, 25]. The XLOGP algorithm is essentially an atom-additive model which is supplemented by a small number of the correction factors [1, 25]. The XLOGP2 program uses a total of 90 atom types to classify atoms in neutral organic compounds. Atoms classifications are made according to: element, hybridization state, solvent accessibility, nature of the neighboring atoms, and adjacency to π-systems [1, 25]. The XLOGP3 program adopts an optimized classification scheme of 87 atom/group types as well as two correction factors accounting for internal H-bonds and amino acids. Fragmental methods (used in the KOWWIN program) are *reductionistic* approaches – fragment and correction factor coefficients are derived by multiple regression of experimental data. They are identified and evaluated concurrently [1]. The CLOGP and ACD/logP packages [1, 25] are *constructionistic* approaches – the basic fragment values are derived from measured log *P* data of

simple molecules, then the remaining fragment set is constructed. These methods systematically interpret and generalize all the possible increments [1]. The AB/logP program combines the advantages of both approaches described above by using hierarchical cluster analysis [1, 25].

Among the most-used and recently popular tools for log *P* prediction as a part of computational methods are: the VCCLAB platform and the included ALOGPS2.1 package [1, 24, 25, 27], and the Novartis Intranet for *in silico* profiling as well [35]. The VCCLAB includes three main parts: Applet Clients, Super Server and Calculation Servers [28]. Applet Clients represents a front-end part of the site and allows the users to provide data, specify parameters, execute tasks and collect calculated results. The Calculation Servers execute the tasks submitted by the clients. The Super Server provides a link between the Applet Clients and Calculation Servers. Basically the ALOGPS 2.1 package is built on the *Associative Neural Network* (ASNN) pattern, as a new challenge for development of physicochemical data prediction methods [29]. This method is very important in chemoinformatics and is introduced as a combination of *k*-nearest neighbour and artificial neural network methods [30]. The ALOGPS was developed from a set of 12908 molecules from the Physical Properties Database (PHYSPROP) using 75 E-state indices [24, 31, 32], and it is also very well correlated with free online available, easy to install software for drawing chemical molecules (ACD/ChemSketch^{TM}, ISIS Draw^{TM}). The PHYSPROP contains chemical structures, names, and physical properties for over 41 000 chemicals. Physical properties are based on a wide variety of sources. This database shows at the same time the experimental, extrapolated, and estimated values for melting point, boiling point, water solubility, octanol-water partition coefficient, vapor pressure, pK_a, Henry's Law Constant, and OH rate constant under standard conditions [31]. In the parameterization of solubility prediction tools, other databases can be used as well, e.g. AQUASOL that contains almost 20 000 solubility records for almost 6000 substances [33]. In order to perform the log *P* calculations the ALOGPS user can draw the molecule using the JME applet of Peter Ertl or submit it in a format supported by OpenBabel [34]. The OpenBabel is a free software available for Windows, Unix, and Mac OS, and is mainly used for converting chemical file formats. It can be easily applied by undergraduate students in the classroom or dormitory and it has a close relationship with chemoinformatics and molecular modeling. A non-Java interface in the ALOGPS 2.1 package for structure submission is available as well. Moreover, the user should know that this application calculates and compares lipophilicity and aqueous solubility of molecules using several other programs including CLOGP, KOWWIN, MiLogP, IA_logP and XLOGP, which definitely increases ALOGPS's practical viability. Moreover, the Novartis Intranet for *in silico* profiling offers properties that can be calculated, like: log *P* coefficient, molar refractivity, etc. Its batch version offers estimation of molecular properties for all virtual libraries as well [35].

Generally, the measurement of experimental log *P* values has been accomplished using reversed-phase high-pressure liquid chromatography [36] or reversed-phase thin-layer chromatography techniques [37]. Moreover, the prediction of log *P* is often developed using complicated, characterized *substructure-based methods* and *whole-molecule approaches*, but one of the possible reasons that these methods have not been successful for complex structures is due to the complexity of quantum mechanical methods and the fact that the solute-solvent interactions are described by the Boltzmann's distribution [36]. The observed chemical shifts in nuclear magnetic resonance (NMR) spectroscopy are related to the electrostatics at the nucleus, which are influenced by solute-solvent interactions. The different solvation effects on a molecule by either water or methanol considerably influences the NMR chemical shifts values which can be observed in aqueous and organic solvent to correlate log *P*. Therefore, a significant path for the log *P* calculation, based on molar volume, hydrogen bonds, and differences in calculated ^{13}C NMR chemical shifts of 162 compounds was discussed [36]. The authors used the ACD/Labs 8.0 CNMR and HNMR software for ^{1}H NMR and ^{13}C NMR of compounds studied in the paper. The results show that the sum of differences in predicted carbon chemical shifts had an important contribution to the log *P* model. Such phenomena were not observed for predicted proton chemical shifts. Thus, a larger interaction with the solvent is connected to protons, and electrostatic effects of the solvent have a greater influence on proton chemical shifts than carbon chemical shifts. It was concluded that a sum of the NMR chemical shifts differences between analyzed substance and methanol, water or octanol represents simple, computer based, *whole-molecule approach* to the log *P* model and, in addition, can be applied to an even larger and diverse set of compounds [36].

Taking into account the relationship between log *P* estimation and NMR measurement a novel and efficient protocol for this purpose, alternative for other procedures, was proposed [40]. In the paper, analyzed pharmaceutical agents were dissolved to near saturation in water and 1-octanol with the usual shake-flask equilibration, and

log *P* was calculated from the concentrations determined in the two phases by NMR, as almost every pharmaceutical agent contains at least one proton signal for observation. Moreover, the analyte was observed in 1D proton spectra either after WET-based (Water suppression Enhanced through T1 effects) solvent (water or octanol) suppression, if their concentrations were low (<10 mM), or with the solvent, if the concentrations were sufficiently high. In this experiment, only deuterium-free solvents were utilized. It was assumed that under ideal conditions, the NMR peak integration A is proportional to the associated proton concentration c and the sine of the pulse excitation angle Θ (no larger than 90°; subscripts s and a represent the solvent and analyte given in the Eq. (6)) [40]. The lipophilic pharmaceutical agent can be observed without any solvent suppression in 1-octanol. The below equation provides the determination of the solute, based on comparisons of peak integrations. On the other hand, the solute concentration in water tends to be much lower. Thus, the water-like solvent might be observed with a very small excitation angle and solute observed in a separate but almost identical acquisition with the solvent suppression and larger excitation angle as the sole changes

$$A_a/A_s=c_a\sin(\Theta_a)/c_s\sin(\Theta_s) \quad (6)$$

This NMR method is superior in its direct detection of analyte, simple data interpretation, and minimal sample manipulation. On the other hand, in such techniques the NMR sensitivity should be considered, which may require hours of acquisition time if the sample concentration falls under 1 μM. Much more important is the lack of clear analyte signal due to interference with the solvent (water or 1-octanol). Nevertheless, both situations occur rarely in drug discovery: i) a drug candidate typically needs at least a modest level of aqueous solubility for good bioavailability, ii) most 'drug-like' compounds have proton signals that can be readily identified and are distinct from the solvent. Therefore, they can be quantified with high accuracy and confidence. Basing on these facts, the authors demonstrated that NMR is a robust tool to determine the log *P* for pharmaceutical agents, without the use of deuterated solvents and internal standards [40].

Thus it is clear that NMR spectra prediction might have a strong influence on series of QSAR/QSPR models, which correlate molecular structures with a measured activity or property including so called molecular descriptors [39, 41]. From this point of view a very interesting study was proposed, in which simulated ^{1}H NMR and ^{13}C NMR spectra were compared with theoretically calculated descriptors, including partition coefficient. One-dimensional NMR spectra have been simulated with ACDs ^{1}H Predictor and ^{13}C Predictor 7.0. Theoretical molecular descriptors were calculated using Dragon 5 package. Moreover, to create mathematical models the partial least square (PLS) methods were utilized, which relate the molecular descriptor set with the activity. Additionally, the root mean square error (RMSE) of leave-on-out cross validation (LOOCV) was used to pick the number of latent variables for the model. Finally, it was concluded that predictive powers of the PLS model for the analyzed data sets indicate that proton NMR prediction is not suitable for building QSPR models. Carbon NMR-based approaches, however, were acceptable, but the Dragon software performed better results. Thus the prediction of proton NMR spectra in order to achieve good correlation with log *P* parameter should probably not be used at all [41].

III. CONCLUSIONS

The chemical compounds synthesized in a framework of academic research are considered as a valuable source for the discovery of new leads [1]. Therefore, medicinal chemists and pharmacists should apply rational approaches at the earliest stages of its studies, ideally during the synthesis. The computer-aided methods are widely used for this purpose, providing estimation of octanol/water coefficient. In addition to numerous commercially-available computer programs, there are computational log *P* prediction tools freely available via the Internet [1, 7, 24, 30]. However, in the majority of publications related to the medicinal chemistry, there is no clear argumentation why this or that particular computational tool has been selected to obtain the computer-aided estimations [1]. Usually, researchers who are not specialized in (Q)SAR/(Q)SPR analysis do not examine if a specific tool is applicable to molecules from the particular chemical series. Also, the validation of particular methods has mostly been provided by these authors. Despite the importance of this problem [37], the independent comparative analysis of prediction accuracy for different tools has not been performed in many cases. By comparison of the prediction results for the same properties provided by different tools (lipophilicity, solubility and drug-likeness), it was shown that these data may vary significantly, and in general there are no objective criteria for selecting ''the best'' method. Moreover, for optimization of the log *P* prediction algorithms, some NMR techniques might be applied, especially ^{13}C NMR with the use of free-deuterium solvents or predictive software, but even then some errors may occur. Thus, our conclusion is that the validation of computational approaches can be performed only by comparison of predicted properties with the results of experimental studies [27, 38, 39, 41].

Acknowledgements

This study was supported by PCSS (Poznan Supercomputing and Networking Center) Grant No. 92/2011, as well as SBN Grant No. 96/2011 (Poznan University of Medical Sciences).

References

[1] J. Kujawski, M.K. Bernard, A. Janusz, W. Kuźma, *Prediction of log P – ALOGPS Application in Medicinal Chemistry Education*. J. Chem. Educ. 89, 64-67 (2012).

[2] C.A. Lipinski, F.M. Lombardo, B.W. Dominy, P.J. Feeney, *Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings*. Adv. Drug Delivery Rev. 46, 3-26 (2001).

[3] F.M. Lombardo, M.Y. Shalaeva, K.A. Tupper, F. Gao, *ElogDoct: A Tool for Lipophilicity Determination in Drug Discovery. 2. Basic and Neutral Compounds*. J. Med. Chem. 44, 2490-2497 (2001).

[4] S.M. Ulmeanu, H. Jensen, G. Bouchard, P.A. Carrupt, H.H. Girault, *Water-Oil Partition Profiling of Ionized Drug Molecules using Cyclic Voltammetry and a 96-Well Microfilter Plate System*. Pharm. Res. 20, 1317-1322 (2003).

[5] A.M. Bond, F. Marken, *Mechanistic aspects of the electron and ion transport processes across the electrode solid solvent (electrolyte) interface of microcrystalline decamethylferrocene attached mechanically to a graphite electrode*. J. Electroanalyt. Chem. 372, 125-135 (1994).

[6] T. Fujita, J. Iwana, C.J. Hansch, *A New Substituent Constant, π, Derived from Partition Coefficients*. J. Am. Chem. Soc. 86, 5175-5180 (1964).

[7] I.V. Tetko, *Computing chemistry on the web*. Drug Discovery Today 10, 1497-1500 (2005).

[8] R. Mannhold, van H. Waterbeemd, *Substructure and whole molecule approaches for calculating log P*. J. Comp-Aided Mol. Des. 15, 337-354 (2001).

[9] R.C. Young, R.C. Mitchell, T.H. Brown, C.R. Ganellin, R. Griffiths, M. Jones, K.K. Rana, D. Saunders, I.R. Smith, N.E. Sore*, Development of a new physicochemical model for brain penetration and its application to the design of centrally acting H_2 receptor histamine antagonists*. J. Med. Chem. 31, 656-671 (1988).

[10] A.J. Leo, *Calculating log Poct from structures*. Chem. Rev. 93, 1281-1306 (1993).

[11] R. Rekker, *The Hydrophobic Fragmental Constant*. Elsevier, Amsterdam 1997.

[12] R. Rekker, R. Mannhold *Calculation of Drug Lipophilicity*. VCH, Weinheim 1992.

[13] http://www.biobyte.com/bb/prod/clogp40.html (accessed: Jan 2012).

[14] P. Broto, G. Moreau, C. Vandycke, *Molecular structures: perception, autocorrelation descriptor and SAR studies*. Eur. J. Med. Chem. 19, 71-78 (1984).

[15] K. Iwase, K. Komatsu, S. Hirono, S. Nakagawa, I. Moriguchi, *Estimation of Hydrophobicity Based on The Solvent-Accessible Surface Area of Molecules*. Chem. Pharm. Bull. 33, 2114-2121 (1985).

[16] W.J. Dunn III, *The role of solvent-accessible surface area in determining partition coefficients*. J. Med. Chem. 30, 1121-1126 (1987).

[17] D.J. Abraham, A.J. Leo, *Extension of the fragment method to calculate amino acid zwitterion and side chain partition coefficients*. Proteins 2, 130-52 (1987).

[18] G.E. Kellogg, D.J. Abraham, *Hydrophobicity: is $LogP_{o/w}$ more than the sum of its parts?* Eur. J. Med. Chem. 35, 651-61 (2000).

[19] G.E. Kellogg, G.S. Joshi, D.J. Abraham, *New tools for modeling and understanding hydrophobicity and hydrophobic interactions*. Med. Chem. Res. 1, 444-53 (1992).

[20] SRC KOWWIN, http://www.syrres.com/search.aspx?searchtext=KOWWIN&folderid=0&searchfor=all&orderby=id&orderdirection=ascending (accessed: Jan 2012).

[21] Molinsipartion, http://www.molinspiration.com (accessed: Jan 2012).

[22] Virtual log *P*, http://nova.colombo58.unimi.it/vlogp.htm (accessed: Jan 2012).

[23] R. Wang, *Calculating partition coefficient by atom-additive method*. Persp. Drug Discov. Des. 19, 47-66 (2000).

[24] Virtual Computational Chemistry Laboratory (VCCLAB), http://www.vcclab.org/ (accessed: Jan 2012).

[25] R. Mannhold, G.I. Poda, C. Ostermann, I.V. Tetko, *Calculation of molecular lipophilicity: State-of-the-art and comparison of log P methods on more than 96,000 compounds*. J. Pharm. Sci . 98, 861-93 (2009).

[26] Advanced Chemical Development (ACD) as ACD/log *P* freeware, www.acdlabs.com/download/ (accessed: Jan 2012).

[27] A. Geronikaki, D. Druzhilovsky, A. Zakharov, V. Poroikov, *Computer-aided prediction for medicinal chemistry via the Internet*. SAR and QSAR Eniromen. Res. 19, 1-2, 27-38 (2008).

[28] I.V. Tetko, J. Gasteiger, R. Todeschini, A. Mauri, D. Livingstone, P. Ertl, V.A. Palyulin, E.V. Radchenko, N.S. Zefirov, A.S. Makarenko, V.Y. Tanchuk, V.V. Prokopenko, *Virtual Computational Chemistry Laboratory – Design and Description*. J. Com.p-Aided Mol. Des. 19, 453-63 (2005).

[29] I.V. Tetko, *Associative Neural Network, CogPrints archive code: cog00001441*, http://www.vcclab.org/lab/alogps/library.html (accessed: Jan 2012).

[30] I.V. Tetko, *Introduction to associative neural networks*. J. Chem. Inf. Comp. Sci. 42, 717-28 (2002).

[31] Syracuse Research Corporation. Physical/Chemical Property Database (PHYSPROP); SRC, Environmental Science Center: Syracuse, NY.

[32] I.V. Tetko, V.Y. Tanchuk, A.E. Villa, *Prediction of n-Octanol/Water Partition Coefficients from PHYSPROP Database Using Artificial Neural Networks and E-State Indices*. J. Chem. Inf. Comp. Sci. 41, 1407-1421 (2001).

[33] F.L. Stahura, J.W. Godden, J. Bajorath, *Differential Shannon Entropy Analysis Identifies Molecular Property Descriptors that Predict Aqueous Solubility of Synthetic Compounds with High Accuracy in Binary QSAR Calculations*. J. Chem. Inf. Comp. Sci. 42, 550-558 (2002).

[34] OpenBabel, http://openbabel.sourceforge.net (accessed: Jan 2012).

[35] P. Ertl, P. Selzer, J. Műhlbacher, *Web-based cheminformatics tools deployed via corporate Intranets*. Drug Discovery Today: Biosilico 2, 5, 201-207 (2004).

[36] L.K. Schnackenberg, R.D. Beger, *Whole-Molecule Calculation of Log P Based on Molar Volume, Hydrogen Bonds, and Simulated ^{13}C NMR Spectra*. J. Chem. Inf. Model. 45, 360-365 (2005).

[37] M. Stefaniak, A. Niestrój, J. Klupsch, J. Śliwok, A. Pyka. *Use of RP-TLC to determine the log P Values of Isomers of Organic Compounds*. Chromatographia 62, 87-89 (2005).

[38] I.V. Tetko, *Can we estimate the accuracy of ADME-Tox predictions?* Drug Discovery Today 11, 700-706 (2006).
[39] H. Sun, *A Universal Molecular Descriptor System for Prediction of LogP, LogS, LogBB, and Absorption*. J. Chem. Inf. Comput. Sci. 44, 748-757 (2004).
[40] H. Mo, K.M. Balko, D.A. Colby, *A practical deuterium-free NMR method for the rapid determination of 1-octanol/water partition coefficient of pharmaceutical agents.* Bioorg. Med. Chem. Lett. 20, 6712-6715 (2010).
[41] E.L. Willighagen, H.M.G.W. Denissen, R. Wehrens, L.M.C. Buydens. *On the Use of ^{1}H and ^{13}C 1D NMR Spectra as QSPR Descriptors.* J. Chem. Inf. Model. 46, 487-494 (2006).

JACEK KUJAWSKI, M.Sc., was born in Świebodzin, Poland in 1978. He studied pharmacy at Poznan University of Medical Sciences in Poznań, where he received his M.Sc. in Pharmacy. Since 2007, he has been working as assistant, toward earning his Ph.D. in the Department of Organic Chemistry at Poznan University of Medical Sciences under the supervision of Assoc. Prof. Bernard. He is the supervisor of a student scientific group (*SKN Chemii Organicznej*) in the local Department.

HANNA POPIELARSKA was born in Leszno, Poland in 1987. She is a a student of pharmacy (5th year) at the Faculty of Pharmacy at Poznan University of Medical Sciences in Poznań, Poland. She is also a member of a student scientific group in the local Department of Organic Chemistry under the supervision of Jacek Kujawski, M.Sc. Her area of interest includes organic synthesis of heterocycles with potential biological activity and *in silico* designing.

ANNA MYKA was born in Mogilno, Poland in 1989. She has been studying pharmacy at Poznan University of Medical Sciences since 2008. She is the main member of a scientific group in the Department of Organic Chemistry. This group has been working on synthesis of heterocyclic derivatives with potential cytostatic activity under the supervision of Jacek Kujawski, M.Sc.

BEATA DRABIŃSKA was born in Bydgoszcz, Poland in 1990. She has been studying pharmacy at Poznan University of Medical Sciences since 2009. She is the member of a scientific group working under the supervision of Jacek Kujawski, M.Sc. at the Department of Organic Chemistry.

MAREK K. BERNARD, Assoc. Prof., Ph.D., D. Sc., graduated from the Poznan University of Medical Sciences in Poznań where he earned his Ph. D. degree under the supervision of professor U. Wrzeciono in 1987, and then D. Sc. degree (2002) in organic chemistry. He was twice a research fellow at the Institute of Organic Chemistry PAS in the professor M. Mąkosza group (1985, 1988) and held post-doctoral position at the University of Florida (Gainesville) in the professor A.R. Katritzky group (1989-1991). In years 2002-2008 he was a coordinator of the English program in pharmacy at the Poznan University of Medical Sciences. His main scientific interest include: nucleophilic substitution in azoles and azines, organic synthesis in solid phase, weak hydrogen bonding and molecular recognition.