The validation
of quantum chemical lipophilicity prediction
# of alcohols

Martin Michalík, Vladimír Lukeš

Department of Chemical Physics, Slovak University of Technology in Bratislava,
Radlinského 9, SK-812 37 Bratislava, Slovakia
martin.michalik@stuba.sk

Abstract: The validation of octanol-water partition coefficients ($\log P$) quantum chemical calculations is presented for 27 alkane alcohols. The chemical accuracy of predicted $\log P$ values was estimated for six DFT functionals (B3LYP, PBE0, M06-2X, ωB97X-D , B97-D3, M11) and three implicit solvent models. Triple-zeta basis set 6-311++G(d,p) was employed. The best linear correlation with the experimental $\log P$ values was achieved for the B3LYP and B97-D3 functionals combined with the SMD model. On the other hand, no linearity was found when IEF-PCM or C-PCM implicit models were employed.

Keywords: Lipophilicity prediction, Validation, DFT functionals, Alcohols

Introduction

The important biopharmaceutical properties that facilitate a drug entry into the body and across various cellular barriers include its solubility, stability, permeability, and first pass effect (Planey, 2013; Arnott, 2012). Moreover, the pharmacokinetic properties, e.g. clearance rate, biological half-life and volume of distribution, represent the next specific characteristics (Pieńko, 2016) responsible for the efficient drug delivery to the target sites. The lipophilicity, which refers to the ability of a compound to dissolve in fats, oils, lipids and non-polar solvents such as hexane or toluene, has long been recognized as an important factor for a drug successful passage through clinical development (Lombardo, 2000; Sârbu, 2012; Guillot, 2009; Komsta 2010). This quantity reflects the net result of all intermolecular forces involving a solute and the two phases between which it partitions (IUPAC, 1997). Generally, the lipophilicity is expressed experimentally as the octanol-water partition coefficient P ($\log P$) or as distribution coefficients ($\log D$). The $\log P$ describes the partition equilibrium of an un-ionized solute, while $\log D$ determines the ratio of the sum of the concentrations of all forms of the compound (pH-dependent mixture of ionized and unionized forms) in each of the two phases. The accurate and efficient measurement of the lipophilicity is an important requirement in drug design. In practice, the theoretical $\log P$ values evaluated using the quantitative structure–activity relationship (QSAR) models (Molinspiration, 2016; Lee, 1988; Estrada-Tejedor, 2013) are often used instead of expensive experimental $\log P$ measurements. However, these QSAR $\log P$ values used for screening virtual libraries can be also inaccurate (Liao, 2008; Mikulski, 2010). They can cause potentially promising compounds to be discarded and/or potentially flawed compounds to move forward (Khan, 2016).

The partition coefficients are tightly related to the electronic structure of the solvated molecules, therefore the methods of quantum chemistry can be also employed in this type of research. Especially during the last decade, computational methods for the precise description of electronic structure of single, non-interacting relatively large molecules have become common (Burke, 2013; Rappoport, 2008). On the other hand, the description of solvent-solute effects is not the trivial task and it brings a large number of unsolved problems. For example, the explicit inclusion of a solvent molecules surrounding the investigated solute leads to a large number of geometrical structures and consequently to a rapid increase of computational times. Plenty of computational time can be saved by using the implicit solvent approximation. This approach assumes that solvent molecules can be replaced by a homogeneously polarizable and thermally averaged isotropic medium (Mennucci, 2002). No explicit solvent molecules are present and no explicit solvent coordinates are given.

The polarizable continuum model (PCM) is a commonly used implicit method and it has seeded the birth of several variants (Tomasi, 2005; Cossi, 2003). This model is based on the Poisson-Boltzmann equation, which is an expansion of the original Poisson's equation. Two types of PCMs are well-known: the dielectric PCM (D-PCM), in which the continuum is polarizable, and the conductor-like PCM (C-PCM), in which the continuum is conductor-like such

as COnductor-like Screening MOdel (COSMO) (Klamt, 1993). The implicit COSMO solvation model uses scaled conductor boundary condition, which is a fast and robust approximation to the exact dielectric equations and reduces the outlying charge errors as compared to PCM. The approximations lead to a root mean square deviation in the order of 0.3 kJ·mol⁻¹ to the exact solutions (Klamt, 2015). Finally, Solvation Model based on Density (SMD) has also seen wide spread use (Marenich, 2009). This model starts from the generalized Born equation which is represented by the approximation of Poisson's equation suitable for arbitrary cavity shapes. It should be mentioned that in quantum chemistry, where charge distributions come from *ab initio* methods and Density Functional Theory (DFT), the implicit solvent models represent the solvent as a perturbation contribution to the solute Hamiltonian.

Although the use of above mentioned implicit models for obtaining various properties of solvated molecules is straightforward, the reliability of the results still needs to be checked with respect to the experimental data. Therefore we decided to perform the validation study of 27 model primary, secondary and tertiary alcohols. The main goals of this work can be defined as follows:

a) to evaluate theoretically $\log P$ values for six distinct DFT functionals together with three different implicit solvent models;

b) to correlate these quantities with the available experimental data;

c) to estimate chemical accuracy of tested approaches.

Finally, the quantum chemically evaluated $\log P$ values will be compared with the QSAR models.

Computational details

The quantum chemical calculations were performed using Gaussian 09 program package (Frisch et al., 2009). The optimal geometries of the studied molecules were calculated in *n*-octanol and water by DFT method with six functionals: B3LYP (Becke's three parameter Lee–Yang–Parr) (Lee et al., 1988; Becke, 1988), PBE0 (Perdew, Burke and Ernzerhof) (Perdew et al., 1996), M06-2X (Zhao et al., 2008), ωB97X-D (Chai et al., 2008), B97-D3 (Grimme et al., 2011), M11 (Peverati, 2011). The energy cut-off was of 10^{-5} kJ mol⁻¹ and final RMS energy gradient was under 0.01 kJ mol⁻¹ Å⁻¹. Within all calculations, the 6-311++G(d,p) basis set was employed for all atoms (Hariharan et al., 1973; Rassolov et al., 1998). This basis set can provide reliable molecular geometries and reaction enthalpies and can be considered sufficiently large (Poliak et al.,

2013; Michalík et al., 2014; Škorňa et al., 2014). The influence of the solvents was approximated by three implicit continuum models, i.e. SMD, IEF-PCM and C-PCM. The optimized structures were confirmed to be real minima by vibration analysis (no imaginary frequencies). The theoretical logarithm of partition coefficient (DFT-$\log P$) for the water/n-octanol mixture was calculated according Eq. 1

$$\text{DFT-}\log P = \frac{\Delta G_{\text{water}} - \Delta G_{\text{n-octanol}}}{2.303\,RT} \qquad (1)$$

where ΔG values are the Gibbs energies of the solvated molecules in the relevant solvent (Garrido, 2012). Visualization of obtained theoretical results was done by Molekel program package (Varetto, 2009).

Results and Discussion

From the geometrical point of view, the studied compounds consist of methyl, methylene bridge, methine and hydroxyl moieties connected by single bonds. This may lead to a large number of conformations. In many cases, the energy differences between the individual conformations are minimal and they can be uniformly thermally populated in real samples. Next, the free rotation of terminal alkyl moieties or hydroxyl groups is possible at the room temperature as the energy barriers for the rotation around single C—C and C—O bonds are very low. For the sake of simplicity, we decided to use in our calculations the consistent definition of the conformations for the individual series of studied alcohols. For example, the geometry optimisation of the primary and selected tertiary alcohols started from symmetric structures. The point group symmetry was fixed to C_s. Next, the starting geometries for the remaining non-symmetric molecules were derived from the addition of hydroxyl group to the symmetric alkyl chain. The notation and gas-phase geometries of all studied molecules are presented in Fig. 1.

The obtained gas-phase geometries were used as the starting geometries for the next geometry optimisation within the tested implicit solvent models. Then, based on the Gibbs energies, the logarithms of partition coefficients were evaluated (see Eq. 1). These theoretical results are presented in Tab. 1. The only negative $\log P$ value within SMD model was obtained for the smallest methanol molecule. The calculated lipophilicity values ranged from –0.35 to 3.87. The maximal positive value predicted B97-D3 functional for the largest primary alcohol, i.e. octan-4-ol. Interestingly, the $\log P$ values of 2-butanol and *tert*-butanol are comparable. On the other hand, isobutanol has by about 15 % larger
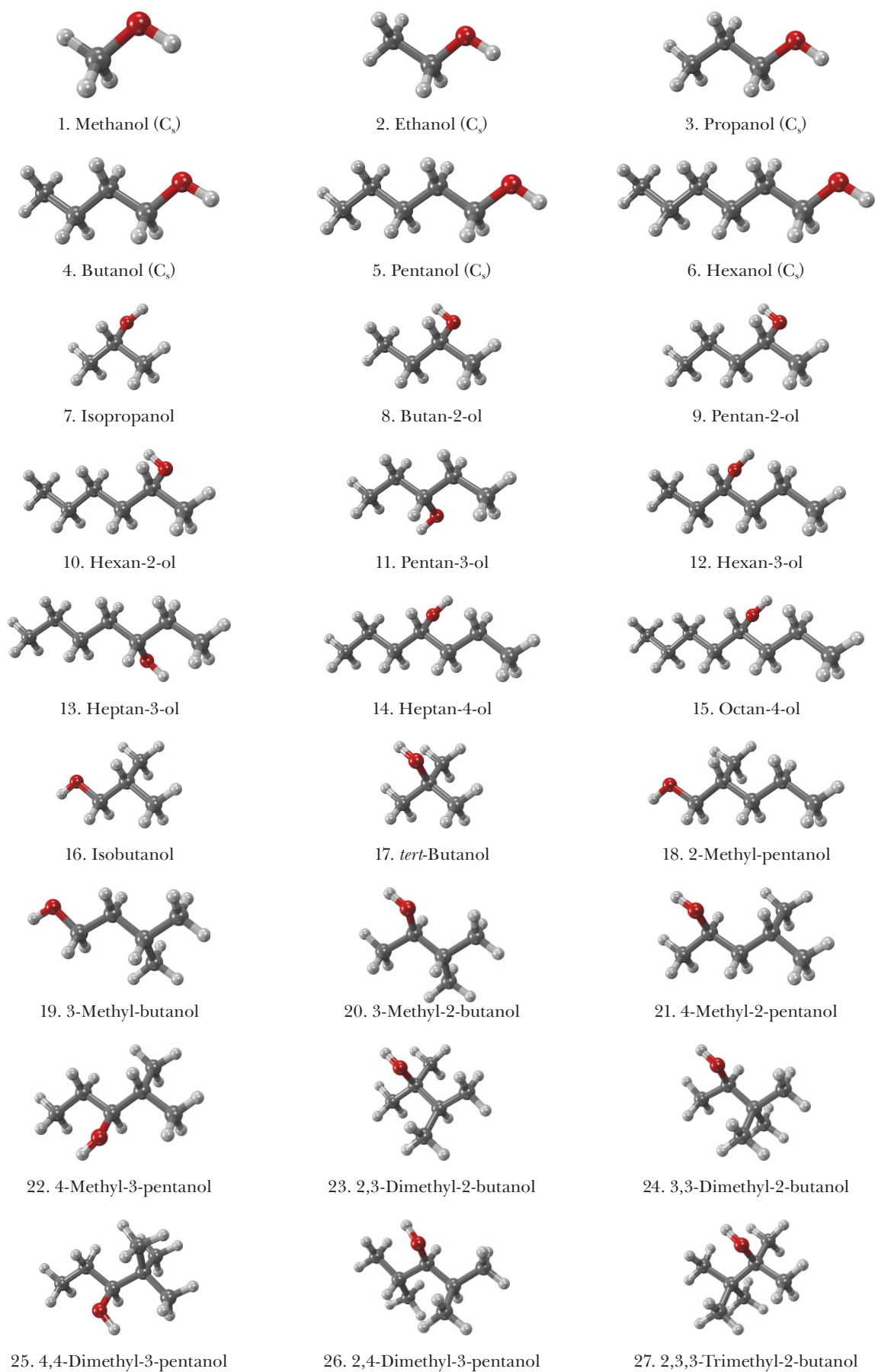
Fig. 1. The notation and optimal gas-phase B3LYP geometries of the studied alcohols. The point group symmetry C$_s$ is indicated in parentheses.

Tab. 1. The Log$P$ lipophilicity values of the studied alcohols calculated using different DFT functionals with SMD solvation model, the QSAR predicted MI-Log$P$ and the experimental Exp-Log$P$ (Khadikar et al. 2005) values.

No. Molecule	B3LYP	M06-2X	PBE0	ωB97X-D	B97-D3	M11	MI-Log$P$	Exp-Log$P$
1. Methanol	–0.22	–0.29	–0.23	–0.26	–0.35	–0.35	–0.32	–0.764
2. Ethanol	0.51	0.44	0.50	0.46	0.56	0.37	0.06	–0.235
3. Propanol	1.10	1.01	1.07	1.04	1.15	0.95	0.56	0.294
4. Butanol	1.65	1.57	1.63	1.59	1.71	1.51	1.12	0.823
5. Pentanol	2.19	2.12	2.18	2.14	2.26	2.06	1.62	1.352
6. Hexanol	2.76	2.69	2.75	2.72	2.86	2.64	2.13	1.881
7. Propan-2-ol	0.78	0.80	0.85	0.79	0.90	0.72	0.42	0.154
8. Butan-2-ol	1.35	1.27	1.39	1.42	1.39	1.27	0.92	0.603
9. Pentan-2-ol	1.81	1.72	1.75	1.79	1.82	1.62	1.48	1.132
10. Hexan-2-ol	2.72	2.34	2.55	2.54	2.63	2.46	1.99	1.661
11. Pentan-3-ol	1.99	2.00	1.89	1.95	1.90	1.98	1.43	1.132
12. Hexan-3-ol	2.59	2.55	2.46	2.57	2.54	2.54	1.98	1.661
13. Heptan-3-ol	2.97	3.02	2.67	3.07	2.91	3.12	2.49	2.190
14. Heptan-4-ol	3.19	3.23	3.23	3.23	3.25	3.43	2.54	2.190
15. Octan-4-ol	3.79	3.81	3.75	3.81	3.87	3.79	3.05	2.680
16. Isobutanol	1.55	1.57	1.57	1.53	1.59	1.51	0.80	0.805
17. *tert*-Butanol	1.30	1.25	1.28	1.26	1.34	1.16	0.87	0.532
18. 2-Methyl-pentanol	2.64	2.42	2.57	2.60	2.70	2.34	1.99	1.693
19. 3-Methyl-butanol	1.86	1.79	1.85	1.89	2.01	1.66	1.33	1.280
20. 3-Methyl-2-butanol	1.88	1.81	1.86	1.75	1.93	1.69	1.17	1.280
21. 4-Methyl-2-pentanol	2.41	2.28	2.28	2.29	2.41	2.22	1.70	1.687
22. 4-Methyl-3-pentanol	2.39	2.05	2.26	2.36	2.37	1.99	1.67	1.687
23. 2,3-Dimethyl-2-butanol	2.51	2.03	2.39	2.28	2.52	1.73	1.61	1.529
24. 3,3-Dimethyl-2-butanol	2.17	1.95	2.01	2.12	2.05	1.87	1.75	1.480
25. 4,4-Dimethyl-3-pentanol	2.86	2.71	2.74	2.63	2.86	2.65	2.25	2.154
26. 2,4-Dimethyl-3-pentanol	2.85	2.74	2.69	2.65	2.96	2.62	1.91	2.148
27. 2,3,3-Trimethyl-2-butanol	2.47	2.56	2.42	2.39	2.42	2.21	2.19	1.996

Tab. 2. The line parameters and regression coefficients for the dependence of Exp-Log$P$ on the DFT(SMD)-Log$P$ values.

DFT Functional	Slope	Intercept	R
B3LYP	0.905(30)	–0.582(67)	0.986
M06-2X	0.906(35)	–0.497(76)	0.982
PBE0	0.940(37)	–0.594(81)	0.981
ωB97X-D	0.908(36)	–0.539(78)	0.981
B97-D3	0.894(33)	–0.576(76)	0.983
M11	0.863(48)	–0.35(10)	0.964

log$P$ value. These predicted trends very well agree with the experimental Exp-log$P$ values published by (Khadikar et al. 2005) as well as with the lipophilicity parameters predicted by QSAR model Mol-Inspiration. Next, regression of Exp-Log$P$ on the MI-Log$P$ dependence provided equation:

$$\text{Exp-Log}P = 1.015(43) \times \text{MI-log}P - 0.234(72) \qquad (2)$$

with the regression coefficient of 0.979. The correlations between the experimental Exp-log$P$ and DFT-log$P$ values lead to linear dependences with the line parameters collected in Tab. 2. The mutual comparison of intercepts show that DFT(SMD) results are higher and uniformly shifted with respect to the experimental results. The best linearity was found for hybrid B3LYP (Fig. 2) and pure B97-D3 DFT functional. On the other hand M11 performed worst with regression coefficient R of 0.964 being the only functional that did not surpass QSAR MI-Log$P$ approach in terms of regression coefficients.

In the case of the IEF-PCM and C-PCM implicit models, all calculated log$P$ values were found in relatively small interval from –0.39 to –0.50 for IEF-PCM and from –0.33 to –0.26 for C-PCM model.
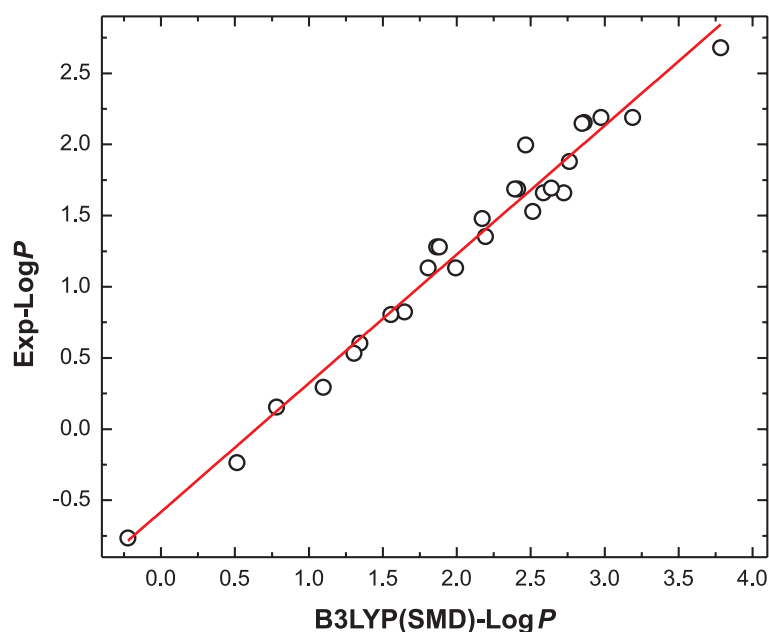
Fig. 2. The dependence of experimental Exp-Log*P* data (Khadikar et al. 2005) on theoretical B3LYP-Log*P* values employing SMD solvation model.

The direct comparison with the experimental or QSAR data reveals no linear correlation.

Conclusion

The validation of quantum chemical octanol-water partition coefficient (log*P*) calculations is presented for 27 primary, secondary and tertiary alkane alcohols. The chemical accuracy of predicted log*P* values was estimated for hybrid, pure and long-range DFT functionals represented by B3LYP, PBE0, M06-2X, B97-D3 and M11, ωB97X-D. The solvent effect was estimated using SMD, IEF-PCM and C-PCM implicit solvent models. The best correlation with the experiment was achieved for SMD model combined with the B3LYP and B97-D3 functionals. Except M11, all functionals performed better that Molinspiration approach in lipophilicity prediction. However, the quality of obtained theoretical results was much more sensitive to the implicit solvation model selection than to the DFT functional. We can conclude that the C-PCM and IEF-PCM implicit models seem unsuitable for the quantum chemical log*P* prediction probably due to the non-adequate description of solvent continuum.

References

Arnott JA, Planey SL (2012) Expert Opin. Drug Dis. 7: 863−875.

Becke AD (1988) Phys. Rev. A. 38: 3098.

Burke K, Wagner LO (2013) Int. J. Quantum Chem. 113: 96−101.

Chai JD, Head-Gordon M (2008) J. Chem. Phys. 128: 084106.

Cossi M, Rega N, Scalmani G, Barone V (2003) J. Comp. Chem. 24: 669−681.

Estrada-Tejedor R, Sabaté N, Broto F, Nonell S (2013) Afinidad 70: 564.

Francl MM, Pietro WJ, Hehre WJ, Binkley JS, Gordon MS, DeFrees DJ, Pople JA (1982) J. Chem. Phys. 77: 3654−3665.

Frisch MJ, Trucks GW, Schlegel HB, G. E. Scuseria GE, Robb MA, Cheeseman JR, Scalmani G, Barone V, Mennucci B, Petersson GA, Nakatsuji H, Caricato M, Li X, Hratchian HP, Izmaylov AF, Bloino J, Zheng G, Sonnenberg JL, Hada M, Ehara M, Toyota K, Fukuda R, Hasegawa J, Ishida M, Nakajima T, Honda Y, Kitao O, Nakai H, Vreven T, Montgomery JA Jr., Peralta JE, Ogliaro F, Bearpark M, Heyd JJ, Brothers E, Kudin KN, Staroverov VN, Keith T, Kobayashi R, Normand J, Raghavachari K, Rendell A, Burant JC, Iyengar SS,Tomasi J, Cossi M, Rega N, Millam JM, Klene M, Knox JE, Cross JB, Bakken V, Adamo C, Jaramillo J, Gomperts R, Stratmann RE, Yazyev O, Austin AJ, Cammi R, Pomelli C, Ochterski JW, Martin

RL, Morokuma K, Zakrzewski VG, Voth GA, Salvador P, Dannenberg JJ, Dapprich S, Daniels AD, Farkas O, Foresman JB, Ortiz JV, Cioslowski J and Fox DJ (2009) Gaussian 09, Revision D.01, Gaussian, Inc. Wallingford CT.

Garrido NM, Economou IG, Queimada AJ, Jorge M, Macedo EA (2012) AIChE J., 58: 1929—1938.

Grimme S, Ehrlich S, Goerigk L (2011) J. Comp. Chem. 32: 1456—1465.

Guillot A, Henchoz Y, Moccand C, Guillarme D, Veuthey JL, Carrupt PA, Martel S (2009) Chem. Biodivers. 6: 1828—1836.

Hariharan PC, Pople JA (1973) Theor. Chim. Acta. 28: 213—222.

IUPAC, Compendium of Chemical Terminology, 2nd ed., Blackwell Scientific Publications, Oxford, 1997.

Khadikar PV, Sharma V, Varma RG (2005) Bioorg. Med. Chem. Lett. 15: 421—425.

Khan AU (2016) Drug Discov. Today 21: 1291—1302.

Klamt A, Moya C, Palomar J (2015) J. Chem. Theory Comput. 11: 4220—4225.

Klamt A, Schüürmann GJGJ (1993) J. Chem. Soc. Perk. T. 2. 5: 799—805.

Komsta Ł, Skibiński R, Berecka A, Gumieniczek A, Radkiewicz B, Radoń M (2010) J. Pharmaceut. Biomed. 53: 911—918.

Lee C, Yang W, Parr RG (1988) Phys. Rev B. 37: 785.

Liao SY, Chen JC, Qian L, Shen Y, Zheng KC (2008) Eur. J. Med. Chem. 43: 2159—2170.

Lombardo F, Shalaeva MY, Tupper KA, Gao F, Abraham MH (2000) J. Med. Chem. 43: 2922—2928.

Marenich AV, Cramer CJ, Truhlar DG (2009) J. Phys. Chem. B 113: 6378—6396.

Mennucci B, Tomasi J, Cammi R, Cheeseman JR, Frisch MJ, Devlin FJ, Stephens PJ (2002) J. Phys. Chem. A 106: 6102—6113.

Michalík M, Vagánek A, Poliak P (2014) Acta Chim. Slovaca 7: 123—128.

Mikulski D, Górniak R, Molski M (2010) Eur. J. Med. Chem. 45: 1015—1027.

Molinspiration. CHEMINFORMATICS. Bratislava, Slovak Republic. 2016.

Perdew JP, Burke K, Ernzerhof M (1996) Phys. Rev. Lett. 77: 3865.

Peverati R, Truhlar DG (2011) J. Phys. Chem. Lett. 2: 2810—2817.

Pieńko T, Grudzień M, Taciak PP, Mazurek AP (2016) J. Mol. Graph. Model. 63: 15—21.

Planey SL, Kumar R (2013) J. App. Pharmacokinet. Biop. 1: 31—36.

Poliak P, Vagánek A (2013) Acta Chim. Slovaca 6: 64—72.

Rappoport D, Crawford NR, Furche F, Burke K (2008) In: Edward IS (Ed) Computational Inorganic and Bioinorganic Chemistry, Vol 2. (pp. 159—173) Wiley, Chichester.

Rassolov VA, Pople JA, Ratner MA, Windus TL (1998) J. Chem. Phys. 109: 1223—1229.

Sârbu C, Naşcu-Briciu RD, Casoni D, Kot-Wasik A, Wasik A, Namieśnik J (2012) J. Chromatogr. A. 1266: 53—60.

Škorňa P, Lengyel J, Rimarčík J, Klein E (2014) Comput. Theor. Chem. 1038: 26—32.

Tomasi J, Mennucci B, Cammi R (2005) Chem. Rev. 105: 2999—3094.

Varetto U (2009) Molekel 5.4.0.8; Swiss National Supercomputing Centre: Lugano (Switzerland).

Zhao Y, Truhlar DG (2008) Theor. Chem. Acc. 120: 215—241.