J Mol Model (2009) 15:1509–1515
DOI 10.1007/s00894-009-0512-3

ORIGINAL PAPER

DFT-based QSAR study of alkanols and alkanthiols using the conductor-like polarizable continuum model (CPCM)

Khaled Azizi · Mohammad Ali Safarpour · Maryam Keykhaee · Ahmad Reza Mehdipour

Received: 24 November 2008 / Accepted: 6 March 2009 / Published online: 22 May 2009

Abstract The usefulness of the CPCM method, calculated at the level of the DFT theory using 6-311++G** basis set for QSAR study of anesthetic activity of alkanol(thiol)s was examined. Three classes of molecular descriptors including AIM, chemical and quantum chemical were used to model the relationships between the anesthetics activity and structural characteristics. Multiple linear regressions were performed to model the relationships between molecular descriptors and biological activity of these molecules using stepwise method and as variable selection tool. A multi-parametric equation containing four descriptors with good statistical qualities was obtained by multiple linear regression (MLR) using stepwise method.

Keywords QSAR · Alkanethiol · Alkanol · Anesthetic · CPCM · DFT

K. Azizi · M. Keykhaee
Faculty of Science,
University of Kurdistan Department of Chemistry,
University of Kurdistan,
Sanandaj, Iran

M. A. Safarpour (✉)
Department of Physical Chemistry, Faculty of Chemistry,
Iran University of Science and Technology (IUST),
Narmak, Tehran, Iran
e-mail: msafarpour@iust.ac.ir

A. R. Mehdipour
Medicinal and Natural Product Chemistry Research Center,
Shiraz University of Medicinal Sciences,
Shiraz, Iran

Abbreviations

AIM	Atoms in molecules
CPCM	Conductor-like polarizable continuum model
DFT	Density functional theory
MLR	Multiple linear regression
QSAR	Quantitative structure–activity relationship

Introduction

Many chemical and biological reactions occur in water where the polar and ionic processes are much more favorable than in the gas phase. Many efforts have been devoted to the development of methods to compute reaction barriers and energetic occurring in condensed phases with experimental accuracy. Effective explicit water models become available for the description of chemical systems in liquid solution. However, with high-level quantum mechanics, only a limited number of solvent molecules can be included explicitly due to the high cost of the calculations. The goal of this work is to determine which theoretical procedure provides the most quantitative estimate of aqueous solvation effects, so that the rates of chemical and biological reactions in water can be computed accurately. One of the most successful solvation models is the conductor-like polarizable continuum model CPCM [1–3].

Recently, computational chemistry has developed into an important contributor to rational drug design. Quantitative structure–activity/property relationship (QSAR/QSPR) models, mathematical equations relating chemical structure to their biological activity, give information that is useful for drug design and medicinal chemistry [4–6]. The derived relationships between molecular descriptors and activity are used to estimate the property of other molecules and/or

finding the parameters affecting the biological activity. The QSAR approach attempts to find consistent relationships between the variations in the values of molecular properties (molecular descriptors) and the biological activity. A major step in constructing the QSAR models is to find one or more molecular descriptors that represent variation in the structural property of the molecules by a number. Nowadays a wide variety of descriptors have been used in QSAR analysis [7–9]. A major step in constructing the QSAR models is finding one or more molecular descriptors that represent variation in the structural property of the molecules by a number. Recent progress in computational hardware and the development of efficient algorithms have assisted the routine development of molecular quantum chemical calculations. Quantum chemical descriptors offer an attractive alternative to traditional QSAR molecular descriptors by expressing a more accurate and detailed description of the electronic and geometric molecular properties and the interaction between them [10]. Recently, Karelson *et al*. reported a comprehensive review on these types of descriptors [11]. Also, Thanikaivelan *et al.* defined some new quantum chemical descriptors, including hardness, softness, electronegativity and electrophilicity, and used them for a QSAR study of alkanes [12]. Very recently, we have successfully applied the *ab initio* theory to derive quantum chemical descriptors for the QSAR studies of some drugs [4–6]. Semiempirical molecular orbital (MO) calculations have been used to obtain electronic descriptors for many years. However, the latest development of the computer technology and software of electronic structure theory allows calculating quantum chemical descriptors at first-principles levels, such as DFT, with higher accuracy including some effective consideration of electron correlation effects. One of the major reasons for the acceleration of the use of electronic structure theory in predicting molecular properties for larger molecules has been the development of density functional theory (DFT). The combination of relatively low computational cost with reasonable accuracy has led to the successful application of the DFT method to the prediction of a broad range of properties of molecules. So, DFT has emerged as a practical and versatile tool to obtain accurate information on molecular systems of chemical interest [13–15]. The performance of the DFT theory in the description of structural, energetic, and magnetic molecular properties has been quite substantially confirmed in recent times [16–18]. The DFT method puts the spotlight on ρ rather than on the wave function, thereby inviting an interpretative electron density theory such as AIM. The theory of atoms in molecules is unique in the sense that it provides a rigorous link between intuitive chemical concepts and quantum mechanics through analysis of the electron density $\rho(r)$(r is a space coordinate) [19]. So, for our purposes, we aimed to use AIM theory as a representational front-end to quantum mechanics that is suitable for QSAR modeling.

In this paper, we discuss the results of our work on the quantitative structure activity relation study of alkanol (thiol) derivatives reported in the literature as potent and anesthetics agents. It is proposed that the mechanism of action of the alkanol derivatives involves electronic interactions with receptors and therefore, we aimed to study the effect of different electronic properties of alkanol derivatives on their biological activity. Therefore we applied the DFT theory to derive quantum chemical descriptors for the QSAR study of the 24 alkanol derivatives. AIM theory was also used to calculate electronic descriptors.

Methods

Activity data

The biological data used in this study are the anesthetic activity in mice (MAC) of the set of 18 alkanol derivatives [20–22]. The biological activity data (MAC) were converted to logarithmic scale (pMAC) and then used for subsequent QSAR analysis as dependent variables. The basic structure of these compounds is shown in Fig. 1. In Table 1 the biological activity values of the alkanol and alkanthiol derivatives used in this study are presented.

Quantum chemical calculations

The molecular structures of all the alkanol derivatives were built with Hyperchem (Version 7, Hyper Cube Inc). Solvent-phase full geometry optimization for the investigated molecules was carried out with the Gaussian 98 series of programs [23].

Dielectric continuum theories are now widely used to describe hydration in conjunction with quantum mechanical calculations due to the relatively low cost of the calculation. CPCM is one of many successful solvation models. In their approaches, the solute interacts with the solvent represented by a dielectric continuum model. The solute molecule is embedded into a cavity surrounded by a dielectric continuum of permittivity. The accuracy of continuum solvation models depends on several factors; the most important one is the use of proper boundary conditions on the surface of the cavity containing the solute. CPCM define the cavities as envelopes of spheres centered on atoms or atomic groups: a number of cavity models have been suggested. Inside the cavity the dielectric constant is the same as in vacuum, outside it takes the value of the desired solvent. Once the cavity has been defined, the surface is smoothly mapped by small regions, called tesserae. Each tessera is characterized by the position of its center, its area, and the

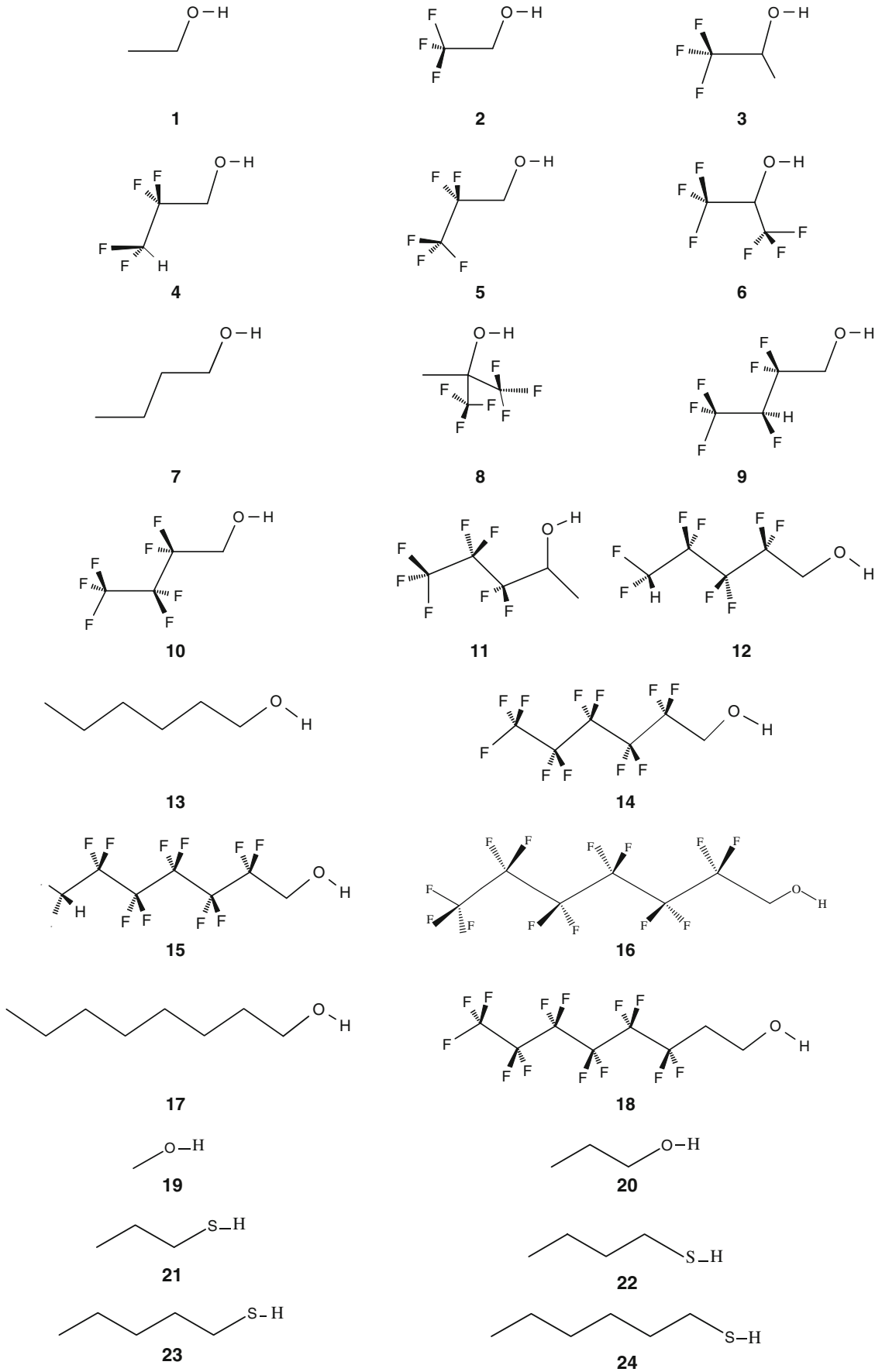

Fig. 1 Structure of the alkanol derivatives used in this study

Table 1 The experimental activity of the alkanol(thiol)s used in this study and their predicted values by MLR

Name	-Log MAC		Name	-Log MAC	
	Experimental	predicted		Experimental	predicted
1	3.005	2.912	13	4.669	4.826
2	3.155	3.531	14	3.344	3.387
3	3.278	3.207	15	4.105	**5.376**[a]
4	4.243	4.185	16	3.374	3.275
5	3.276	3.307	17	5.932	5.799
6	4.355	4.076	18	3.650	3.759
7	3.876	3.933	19	2.699	2.614
8	3.638	3.587	20	3.493	3.402
9	4.352	4.338	21	1.991	2.022
10	3.398	3.344	22	2.382	2.512
11	3.303	3.405	23	2.759	2.813
12	4.699	4.849	24	3.289	3.078

[a] Outlier compound in final model

electrostatic vector normal to the surface passing through its center. Recently, the CPCM method has been improved and extended in so that the cavity can be selected in a number of different ways [1].

The structures were optimized with DFT method at the hybrid functional B3LYP (Becke's three-parameter [24] functional employing the Lee, Yang, and Parr correlation functional) [25] and the medium-size basis set 6-311++G(d, p) level. No molecular symmetry constraint was applied; rather full optimization of all bond lengths and angles was carried out. Local charge (LC) calculated according to Mulliken population analysis (MPA) [26], natural population analysis (NPA) [27] and electrostatic potential (EP) [28] at each atom, highest occupied molecular orbital (HOMO) and lowest unoccupied molecular orbital (LUMO) energies, difference between LUMO and HOMO orbital energies, molecular dipole moment (MDP) and molecular polarizability (MP), total free energy in solution with all non electrostatic terms calculated by Gaussian 98. The molecular modeling system Hyperchem software was further employed to calculate the following parameters from the energy minimized structures: molecular surface area, molar refractivity and n-octanol/water partition (log P). Quantum chemical indices of hardness (η), softness (S), electronegativity (χ) and electrophilicity (ω) were calculated according to the method proposed by Thanikaivelan *et al.* [12]. In addition, the topological analysis of electronic density distribution in the theory of atoms in molecules (AIM) was used to compute some other molecular properties. A brief description of the descriptors used in this study is represented in Table 2.

Data processing and modeling

The multiple linear regression (MLR) analysis was employed to derive the QSAR models for different alkanol(thiol)s derivatives (Table 3). MLR analysis and correlation analysis were carried out by the statistics software SPSS 13.0 version. Before any MLR analysis, the correlation between the selected descriptors was examined (Table 4) and collinear descriptors ($r>0.90$) were determined. Among these descriptors one of them, which had higher correlation with the dependent variable, was retained and the others were removed from the descriptor

Table 2 The calculated descriptors used in this study

Descriptor type	Molecular descriptors
Quantum chemical descriptors	Highest occupied molecular orbital energy (E_{HOMO}), lowest unoccupied molecular orbital energy (E_{LUMO}), the local charges at each atom of the unit of alkanols (LC_i), electrostatic potential at each atom (EP_i), difference between LUMO and HOMO orbital energy (HOMO-LUMO gap), Molecular dipole moment (MDP), Molecular polarizability (MP), Hardness ($\eta=0.5(E_{HOMO}+E_{LUMO})$), Softness ($S=1/\eta$), Electronegativity ($\chi=-0.5(E_{HOMO}-E_{LUMO})$), Electrophilicity ($\omega=\chi^2/2\eta$), Total free energy in solution with all non electrostatic terms
AIM descriptors	Electron density (Rho) on critical points, Rho on surface between two atoms, Hamiltonian kinetics energy density (K (r)), Lagrangian kinetics energy density (G(r))
Physiochemical descriptors	Molecular volume (V), Molecular surface area (MSA), Molar refractivity (MR), n-octanol/water partition coefficient (log P), torsion angle of C-C-O-H

Table 3 Calculated values of the descriptors used in different MLR

Entry	pMAC	Local charge H	Log p	Polarizability	K(r) OH	G(r) OH	G(r) CO	K(r) CC
1	3.00	0.33	3.42	38.51	6.62	6.63	2.01	2.05
2	3.15	0.34	3.05	38.91	6.61	6.04	2.12	2.36
3	3.28	0.35	2.52	54.28	6.58	6.23	2.09	2.03
4	4.24	0.34	3.29	54.27	6.63	6.09	2.11	2.25
5	3.28	0.34	2.25	54.62	6.61	6.03	2.10	2.25
6.5	4.35	0.38	2.42	55.37	6.59	5.67	2.17	2.12
7	3.88	0.32	3.15	70.60	6.67	6.51	1.99	2.05
8	3.64	0.41	1.67	69.77	6.49	5.79	2.11	2.00
9	4.35	0.34	2.91	70.03	6.64	6.11	2.12	2.23
10	3.40	0.34	1.73	70.57	6.61	6.04	2.09	2.22
11	3.30	0.35	1.36	86.18	6.64	6.12	3.40	1.98
12	4.70	0.34	2.81	85.83	6.62	6.05	2.10	2.22
13	4.67	0.32	2.96	102.96	6.66	6.52	1.99	2.05
14	3.34	0.34	0.65	102.70	6.61	6.04	2.09	2.20
15	4.11	0.35	2.22	117.89	6.62	6.04	2.09	2.22
16	3.37	0.35	0.19	118.98	6.61	6.18	2.08	2.21
17	5.93	0.32	2.85	135.34	6.67	6.52	1.99	5.62
18	3.65	0.20	0.39	134.08	6.65	6.32	2.01	1.96
19	2.70	0.32	3.58	22.34	6.70	6.60	2.09	……..
20	3.49	0.15	3.41	54.34	6.62	6.64	2.01	2.05
21	1.99	0.09	0.43	77.39	1.87	4.37	4.41	1.94
22	2.38	0.14	0.35	93.05	1.88	4.34	4.32	1.93
23	2.76	0.14	0.09	109.53	1.87	4.34	4.32	1.93
24	3.29	0.14	-0.20	125.80	1.88	4.34	4.32	1.93

data matrix. The remaining descriptors were used to construct the MLR model, in accordance with the stepwise and genetic algorithm selection methods. Cross-validation procedure (leave-one-out (Q_{LOO}^2) and leave-five-out (Q_{LFO}^2) was applied to measure the predictive capabilities of the models by using Matlab 7 program [29]. Furthermore, five splits of test and calibration sets were prepared in order to check predictivity of models which were shown in Table 5. In order to assess the risk of chance correlation [30, 31], input scrambling was performed [32]. According to the results there was not any risk for chance correlation (R_{max}^2=0.332, Q_{max}^2=0.448).

Results and discussion

The chemical structures of the molecules used in this study are shown in Fig. 1 and biological activity of alkanol(thiol)s derivatives are presented in Table 1.

Table 4 The correlation coefficient existing between the variables used in different MLR

	pMAC	Local charge H	Log p	Polarizability	K(r) OH	G(r) OH	G(r) CO	K(r) CC
pMAC	1.000							
Local charge H	0.47	1.00						
Log p	0.44	0.50	1.00					
Polarizability	0.30	-0.30	-0.64	1.00				
K(r) OH	0.53	0.66	0.72	-0.28	1.00			
G(r) OH	0.49	0.80	0.66	-0.39	0.74	1.00		
G(r) CO	-0.56	-0.75	-0.68	0.27	0.71	0.75	1.00	
K(r) CC	0.64	0.16	0.24	0.35	0.26	0.20	-0.23	1.00

Table 5 The results of random splitting of the data to three sets for equations of all descriptor sets

	All descriptors					
Model	$R_{calibration}^2$	$R_{predction}^2$	$R_{calibration}^2$	$R_{predction}^2$	$R_{calibration}^2$	$R_{predction}^2$
1	0.935	0.819	0.775	0.761	0.885	0.775
2	0.762	0.606	0.727	0.751	0.750	0.638
3	0.920	0.816	0.974	0.987	0.943	0.812

In order to quantitatively obtain the effects of the structural parameters of the alkanol derivatives on their anesthetics activity, QSAR analysis with different types of molecular descriptors was operated. The octanol-water partition coefficient (log P) has been considered as the descriptor for the hydrophobic effect. The steric effect has been described by means of the surface area and volume. The electronic descriptors such as electron densities have been derived from AIM calculations. The quantum chemical descriptors were calculated by DFT method. A total of 35 descriptors were calculated for each molecule. After checking colinearity, the number of descriptors reduced to 23. Some of them are related to the properties of the individual atoms in the basic structure of the alkanol derivatives. Other descriptors were related to the whole structural properties of the alkanol derivatives.

MLR analysis

MLR analysis with the stepwise selection and elimination of variables was employed to model the structure-activity relationships with a different set of descriptors. The first QSAR model was derived by using the physiochemical and quantum descriptors and the following equation was obtained:

$$\begin{aligned} \mathrm{pMAC} = {} & 2.789(\pm 0.996)\text{Local charge H} \\ & + 0.026(\pm 0.003)\text{polarizability} \\ & + 0.610(\pm 0.088)\text{Log P}_{+}0.510(\pm 0.430) \end{aligned} \quad (1)$$

where $N = 24$, $R^2 = 0.835$, $F = 33.7$, $Q^2_{LOO} = 0.780$, $Q^2_{LFO} = 0.720$, $SEE = 0.367$.

The numbers in parentheses are the standard deviation of the coefficients. This equation has an acceptable quality. As is expected, Log P, as indicator of lipophilicity, was entered to the equation, since our previous studies about halogenated ether confirmed that lipohilicity and polarizability have a very significant role in the anesthesia [33].

Second QSAR model was obtained by using AIM descriptors. This equation is the following:

$$\begin{aligned} \mathrm{pMAC} = {} & -0.443(\pm 0.131)\mathrm{G(r)(CO)} \\ & + 0.603(\pm 0.158)\mathrm{K(r)(CC)} \\ & + 3.393(\pm 0.550) \end{aligned} \quad (2)$$

where $N = 23$, $R^2 = 0.630$, $F = 17.06$, $Q^2_{LOO} = 0.389$, $Q^2_{LFO} = 0.390$, $SEE = 0.535$.

This equation does not have high quality especially in cross-validation process. It can not be discussed about the meaning of included parameters.

The final MLR model was obtained by all descriptors in pool data. This equation is the following:

$$\begin{aligned} \mathrm{pMAC} = {} & 0.029(\pm 0.003)\text{Polarizability} \\ & + 0.818(\pm 0.097)\log \mathrm{p} \\ & - 1.313(\pm 0.292)\mathrm{k(r)(OH)} \\ & + 0.538(\pm 0.138)\mathrm{G(r)(OH)} \\ & + 4.274(\pm 1.043) \end{aligned} \quad (3)$$

where $N = 23$, $R^2 = 0.974$, $F = 42.01$, $Q^2_{LOO} = 0.953$, $Q^2_{LFO} = 0.910$, $SEE = 0.295$.

Among the quantum chemical, AIM and physiochemical descriptors used, polarizability, log p and G(r) and K(r) on surface between two atoms (O, H) appeared in Eq. 3.

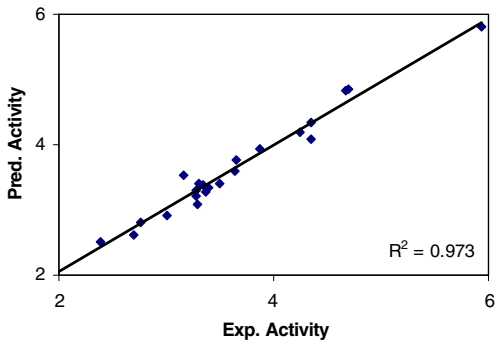

Fig. 2 Plot of predicted activity against the corresponding experimental activity

As can be seen, this equation has high quality and the variables used in this equation can explain 93.4% of the variance in the biological activity of alkanol(thiol) derivatives (Fig. 2). As the model was built, it was observed that compound No.15 had a significant deviation (more than 3SD) from the regression line; therefore, it was assumed as an outlier and deleted from the modeling procedure.

For simplicity, these standardized regression coefficients are not reported in this paper but the equations of the proposed models report the variables in descending relative importance. Therefore, the trend of the appearance of variables in Eq. 3 suggests that the polarizability of drugs is the most important electronic property influencing the binding of this type of alkanol(thiol) derivatives with their receptors. As is obvious, the positive coefficient of polarizability, log P and G(r) on the surface between two atoms (O, H) reveals that the effective concentration of the molecule decreases when increasing these descriptors. Meanwhile, the negative coefficient of K(r) on the surface between two atoms (O, H) indicated that the effective concentration of the molecule increases when increasing this descriptor.

Conclusions

A quantitative structure-activity relationship study was performed to study the anesthetics activity of 24 alkanol (thiol)s derivatives. The DFT theory was used to optimize the 3D geometry of the molecules and to calculate a diverse set of quantum chemical descriptors. AIM derived descriptors in combination with some chemical descriptors were also used. In the MLR procedure, the quantum chemical descriptors concerning the molecular properties (polarizability) and that of the individual atom in the molecule (local charge H), and the chemical and AIM descriptors, including log P and Rho surface OH, were found to have the same importance in controlling the anesthetics behavior of the whole molecule. This results may imply the importance of the OH bond in the anesthetic activity of alkanol(thiol)s.

References

1. Takano Y, Houk HN (2005) Chem Theory Comput 1:70–77
2. Cossi M, Rega N, Scalmani G, Barone V (2003) J Comput Chem 24:669–681
3. Barone V, Cossi M (1998) J Phys Chem 102:1995–2001
4. Safarpour MA, Hemmateenejad B, Miri R, Jamali M (2003) QSAR Comb Sci 22:997–1005
5. Hemmateenejad B, Safarpour MA, Miri R, Taghavi F (2004) J Comput Chem 25:1495–1503
6. Hemmateenejad B, Safarpour MA, Taghavi F (2003) THEOCHEM 635:183–190
7. Putta S, Eksterowicz J, Lemmen C, Stanton R (2003) J Chem Inf Comput Sci 43:1623–1635
8. Todeschini R, Consonni V (2000) Handbook of Molecular Descriptors. Wiley, Weinheim
9. Horvath D, Mao B (2003) QSAR Comb Sci 22:498–509
10. Carbo-Dorca R, Amat L, Besaqlu E, Girones X, Robert D (2000) THEOCHEM 504:181–228
11. Karelson M, Lobanov VS, Katritzky AR (1996) Chem Rev 96:1027–1044
12. Thanikaivelan P, Subramanian V, Rao JR, Nair BU (2000) Chem Phys Lett 323:59–70
13. Par RG, Yang W (1989) In: Density-functional theory of atoms and molecules. Oxford Univ Press, Oxford, pp 47–69
14. Bernardi F, Bottoni A, Garavelli M (2002) Quant Struct Act Relat 21:128–148
15. Sulpizi M, Folkers G, Rothlisberger U, Carloni P, Scapozza L (2002) Quant Struct Act Relat 21:173–181
16. Chermette H (1999) J Comput Chem 20:129–154
17. Geerlings P, De Proft F, Langenaeker W (2003) Chem Rev 103:1793–1874
18. Chattaraj PK, Nath S, Maiti B (2003) Reactivity descriptors. In: Tollenaere J, Bultinck P, Winter HD, Langenaeker W (eds) Computational medicinal chemistry and drug discovery. Dekker, New York, pp 295–322
19. Bader RFW (1990) Atoms in molecules: A quantum theory. Oxford Univ Press, Oxford, pp 179–195
20. Eger EI II, Ionescu P, Laster MJ, Gong D, Hudlicky T, Kendig JJ, Harris RA, Trudell JR, Pohorille A (1999) Anesth Analg 88:867–876
21. Fang Z, Ionescu P, Chortkoff BS, Kandel L, Sonner J, Laster MJ, Edmond EI II (1997) Anesth Analg 84:1042–1048
22. Zhang Y, Trudell JR, Mascia MP, Laster MJ, Gong DH, Harris RA, Eger EI II (2000) Anesth Analg 91:1294–1299
23. Frisch MJ, Trucks GW, Schlegel HB, Scuseria GE, Robb MA, Cheeseman JR, Zakrzewski VG, Montgomery JA Jr, Stratmann RE, Burant JC, Dapprich S, Millam JM, Daniels AD, Kudin KN, Strain MC, Farkas O, Tomasi J, Barone V, Cossi M, Cammi R, Mennucci B, Pomelli C, Adamo C, Clifford S, Ochterski J, Petersson GA, Ayala PY, Cui Q, Morokuma K, Malick DK, Rabuck AD, Raghavachari K, Foresman JB, Cioslowski J, Ortiz JV, Baboul AG, Stefanov BB, Liu G, Liashenko A, Piskorz P, Komaromi I, Gomperts R, Martin RL, Fox DJ, Keith T, Al-Laham MA, Peng CY, Nanayakkara A, Gonzalez C, Challacombe M, Gill PMW, Johnson B, Chen W, Wong WM, Andres JL, Gonzalez C, Head–Gordon M, Replogle ES, Pople JA (1998) Gaussian 98, Revision A.7. Gaussian Inc, Pittsburgh, PA
24. Becke AD (1993) J Chem Phys 98:5648–5652
25. Lee C, Yang W, Parr RG (1988) Phys Rev 37:785–789
26. Mulliken RS (1955) J Chem Phys 23:1833–1840
27. Reed AE, Weinhold F (1983) J Chem Phys 78:4066–4073
28. Levine IN (2000) Quantum Chemistry, 5th ed. Prentice Hall, Upper Saddle River, NJ 07458:508–509
29. Version 7 (2005), Mathwork Inc, http://www.mathworks.com, USA
30. Topliss JG, Costello RJ (1972) J Med Chem 15:1066–1068
31. Salt DW, Ajmani S, Crichton R, Livingstone D (2007) J Chem Inf Comput Sci 47:143–149
32. Kubinyi H, Hamprecht FA, Mietzner T (1998) J Med Chem 41:2553–2564
33. Mehdipour AR, Hemmateenejad B, Miri R (2007) Chem Biol Drug Des 69:362