Journal of Computer-Aided Molecular Design, 11 (1997) 503–515.
KLUWER/ESCOM

J-CAMD 409

Heuristic lipophilicity potential for computer-aided rational drug design

Qishi Du[a], Gustavo A. Arteca[b] and Paul G. Mezey[a,*]

[a]*Department of Chemistry, University of Saskatchewan, 110 Science Place, Saskatoon, SK, Canada S7N 5C9*
[b]*Department of Chemistry and Biochemistry, Laurentian University, Ramsey Lake Road, Sudbury, ON, Canada P3E 2C6*

Received 7 October 1996
Accepted 6 May 1997

Keywords: Molecular shape; Lipophilicity potential; Electrostatic potential; Molecular modeling; Drug design

Summary

In this contribution we suggest a heuristic molecular lipophilicity potential (HMLP), which is a structure-based technique requiring no empirical indices of atomic lipophilicity. The input data used in this approach are molecular geometries and molecular surfaces. The HMLP is a modified electrostatic potential, combined with the averaged influences from the molecular environment. Quantum mechanics is used to calculate the electron density function $\rho(\mathbf{r})$ and the electrostatic potential $V(\mathbf{r})$, and from this information a lipophilicity potential $L(\mathbf{r})$ is generated. The HMLP is a unified lipophilicity and hydrophilicity potential. The interactions of dipole and multipole moments, hydrogen bonds, and charged atoms in a molecule are included in the hydrophilic interactions in this model. The HMLP is used to study hydrogen bonds and water–octanol partition coefficients in several examples. The calculated results show that the HMLP gives qualitatively and quantitatively correct, as well as chemically reasonable, results in cases where comparisons are available. These comparisons indicate that the HMLP has advantages over the empirical lipophilicity potential in many aspects. The HMLP is a three-dimensional and easily visualizable representation of molecular lipophilicity, suggested as a potential tool in computer-aided three-dimensional drug design.

Introduction

The lipophilic or hydrophobic effect is one of the most important properties of organic and biological molecules. Molecular lipophilicity plays an important role in the study of molecular bioactivities. For computer-aided rational drug design, molecular lipophilicity is one of the key factors [1–5]. In recent years, with the help of the great progress in computational chemistry, various computational methods have become available for studies of lipophilicity, including Monte Carlo simulation [6–8], molecular dynamics simulation [9,10], quantum ab initio and semiempirical self-consistent reaction field (SCRF) methods of the continuum medium model [11–15], and combined methods of quantum mechanics and molecular mechanics [16]. Information about the role of lipophilicity has been accumulated during the last three decades concerning the physical nature of hydrophobic hydration (HH) and hydrophobic interaction (HI), the structure of hydration shells, the thermodynamic properties (enthalpic and entropic changes), and the lipophilic force between organic solute molecules in aqueous solution. Reflecting these developments, elaborate methods have been used to represent and describe molecular lipophilicity: one-dimensional (1D) scalar descriptor partition coefficients between water and organic solvent (most often, octanol is used) [17], 1D vector descriptor lipophilicity moment [18,19], two-dimensional (2D) lipophilicity maps [20–23], and three-dimensional (3D) lipophilicity potential [24,25]. The 3D molecular lipophilicity potential (MLP) provides visualization of lipophilicity and is a powerful tool in 3D quantitative structure–activity relationship (QSAR) studies and rational drug design [2–5]. However, all the earlier MLP models were empirical, which depend on an empirical parameter set of fragmental or atomic lipophilicity indices [26,27].

In the approach of 3D MLP, developed by Audry et al. [28,29] and Furet et al. [25], the logarithm of partition

*To whom correspondence should be addressed.

coefficient ($\log P_{ow}$), which is an overall measure of molecular lipophilicity, is decomposed into fragmental or atomic contributions:

$$\log P_{ow} = \sum_i f_i \tag{1}$$

where f_i is the lipophilic parameter of the individual fragment i. Based on a data set of atomic lipophilicity parameters $\{f_i\}$, Audry et al. [28,29] have suggested a formula for the calculation of lipophilicity potential. In their computer program developed for the calculation of lipophilicity potential, the formula takes the simple form [24]

$$L(\mathbf{r}) = \sum_i \frac{f_i}{1+\|\mathbf{r}-\mathbf{r}_i\|} \tag{2}$$

where $\mathbf{r}_i$ is the position of nucleus i, and summation is over all constituent atoms. If the point $\mathbf{r}$ is on atom i, $\|\mathbf{r}-\mathbf{r}_i\| = 0$. In this situation, the denominator of Eq. 2 is 1. It means that f_i is the dominant factor in the space surrounding atom i. The lipophilicity potential $L(\mathbf{r})$ defined by Eq. 2 gives us a picture: for an organic solute molecule, its lipophilic surface area exerts lipophilic force into the surrounding space to attract nonpolar molecules and repulse water molecules, whereas its hydrophilic surface area exerts hydrophilic force to attract polar molecules. It is clear that the lipophilicity potential defined by Eq. 2 is not based on a rigorous theoretical model and is not a true physical potential. There may be other empirical MLP formulas, such as one using Gaussian-type distance dependence. A number of possible functions were compared by Croizet et al. [24]. Alternative formulas are also discussed by Heiden et al. [20]. In Eq. 2, the atomic lipophilic contributions f_i decay with the distance $\|\mathbf{r}-\mathbf{r}_i\|$. Actually, we think f_i decays with a higher power, $\|\mathbf{r}-\mathbf{r}_i\|^\gamma$. In Eq. 2, the exponent γ is 1. Later we shall discuss the effects of γ.

Goodford [30] studies the binding sites of ligands on macromolecules in the way of molecular mechanics. The interaction of a probe group with a protein is computed at sample positions around the macromolecule, giving an array of energy values using molecular mechanics. The probe groups include water, methyl group, amine nitrogen, carboxyl oxygen, and hydroxyl. Contour surfaces of energy levels using a water probe give information of molecular lipophilicity surrounding a macromolecule. In his calculations, the water molecule is treated as an electrically neutral group that has no dipole but can donate two hydrogen bonds and can also accept up to two hydrogen bonds. The Lennard-Jones potential is used in the calculations of interaction energies between probe molecules and the macromolecule [30]. Goodford's method is an empirical approach too because empirical molecular mechanics is used. Kantola et al. [31] presented an atomic parametrization for the determination of hydrophobicity index that depends on the molecular conformation. Similarly, program CLOGP [32] is developed for the calculation of partition coefficients using a nonlinear regression model.

Independently of the type of formulas used in empirical MLP, atomic lipophilic parameters f_i's are the building blocks in the construction of the molecular lipophilicity potential. In earlier models one had to use experimental data to derive the f_i values [26]. Using Eq. 1 and the $\log P_{ow}$ data for 500 representative organic molecules, and employing least-squares techniques, Ghose and Crippen [26] have derived f_i values for atoms in the most common functional groups. They have classified the common atoms into 90 different 'types' or classes. These classes take into account the number and nature of atoms directly connected to the one under consideration. Improved parameter sets were found by Viswanadhan et al. [27]. In their lipophilic data set, a positive f_i means atom i is lipophilic, whereas a negative f_i means the atom is hydrophilic. The data set of atomic lipophilic parameters f_i's is still expanding. The calculation results of empirical MLP and empirical parameters of atomic lipophilicity indices need to be checked and compared with a theoretical model. However, no earlier theoretical method appears to be reliable enough to succeed in this task. Therefore, a potentially more reliable, heuristic lipophilicity potential model, if successful, is keenly needed in molecular modeling and chemical design.

A well-established and important approach for the study of molecular interactions is the molecular electrostatic potential (MEP). This 3D function has a precise quantum chemical definition. The electrostatic potential $V(\mathbf{r})$ is created in the space around a molecule by its nuclei and electrons. According to quantum mechanics, MEP is defined by the following equation:

$$V(\mathbf{r}) = \sum_\alpha \frac{Z_\alpha}{\|\mathbf{R}_\alpha - \mathbf{r}\|} - \int \frac{\rho(\mathbf{r})}{\|\mathbf{r}' - \mathbf{r}\|}\, d\mathbf{r}' \tag{3}$$

where Z_α is the charge on nucleus α located at $\mathbf{R}_\alpha$, and $\rho(\mathbf{r})$ is the electron density function [33,34]. Besides a rigorous theoretical foundation, MEP also has a solid experimental background as it can be determined by electron diffraction experiments. MEP has a simple physical meaning: $V(\mathbf{r})$ is the interaction energy between a molecule and a unit probe point charge at position $\mathbf{r}$.

In principle, most molecular interactions are initiated by electrostatic interaction between molecules. Many authors have studied the relationships between the electrostatic potential $V(\mathbf{r})$ and the lipophilicity potential $L(\mathbf{r})$. Rozas et al. [35] have used MEP–MLP maps to study the connections and differences between MEP and MLP on the molecular surface. In a series of research papers, Náray-Szabó and his research group [21–23] have studied the electrostatic and lipophilic complementarity of molec-

ular associations. In their model, two aspects, Coulombic and hydrophobic matching, can be formulated in terms of molecular electrostatic potentials and fields. Politzer and co-workers [36] have developed a technique called general interaction properties function (GIPF), and used it to study a number of physical properties concerning various types of molecular interactions, including solubilities and partition coefficients [37–39]. In the GIPF technique, several statistically based interaction indices are derived from the molecular surface electrostatic potential distribution, such as the average absolute deviation Π, positive, negative, and total variances σ_+^2, σ_-^2, and σ_{tot}^2, the balance coefficient ν between the positive part and the negative part, and so on. Then, molecular properties, such as partition coefficients $\log P_{ow}$, are correlated with these MEP indices together with the molecular surface area s:

$$\log P_{ow} = f(s,\Pi,\sigma_{tot}^2,\nu) \tag{4}$$

In Politzer's GIPF approach, the surface area and MEP statistical quantities are independent parameters. However, in applications of the GIPF approach, sometimes the surface area is combined with other MEP quantities, such as $\nu\sigma_{tot}^2/s$ and Πs, in order to obtain better results. Molecular lipophilicity is a property involving the molecular surface area and surface conditions. Du and Arteca [40] have combined two types of quantities into one type of parameter, surface–MEP descriptors, in order to characterize the conditions of the molecular surface:

$$B_F^+ = \left[\sum_i \Delta s_i^+ V^+(\mathbf{r}_i)\right]_F \tag{5}$$

$$B_F^- = \left[\sum_i \Delta s_i^- V^-(\mathbf{r}_i)\right]_F \tag{6}$$

$$B_R^+ = \left[\sum_i \Delta s_i^+ V^+(\mathbf{r}_i)\right]_R \tag{7}$$

$$B_R^- = \left[\sum_i \Delta s_i^- V^-(\mathbf{r}_i)\right]_R \tag{8}$$

where $V(\mathbf{r}_i)$ is the MEP at point $\mathbf{r}$, and Δs_i is the surface area element at point $\mathbf{r}$. Subscripts F and R denote the functional group and hydrocarbon moiety, respectively. Superscripts + and − stand for the positive and negative MEP, respectively. Molecular lipophilicity can be described as a function of the four surface–MEP descriptors:

$$L = f(B_F^+,B_F^-,B_R^+,B_R^-) \tag{9}$$

Surface–MEP descriptors (B_F^+, B_F^-, B_R^+, and B_R^-) have been used to correlate partition coefficients $\log P_{ow}$ successfully [40]. However, this approach has a serious shortcoming: one has to designate lipophilic surface areas and hydrophilic surface areas in a molecule artificially, and it does not give a 3D representation of the lipophilicity potential.

Heuristic molecular lipophilicity potential

Hydrophobic effects are complex phenomena. The term usually refers to both hydrophobic hydration (HH) and hydrophobic interaction (HI) [6,41,42]. HH concerns the thermodynamic and structural changes that are associated with the solvation of a nonpolar solute in water [7,8]. HH is conveyed by thermodynamic properties: free energy, enthalpy, and entropy of hydration. HI refers to the interactions between two organic molecules dissolved in aqueous solution [10,41]. HI is more useful in chemistry and HH is the basis to understand the nature of HI and to do qualitative and quantitative prediction for HI. Heuristic molecular lipophilicity potential (HMLP) is a tool for the study of HH, and not directly for the study of hydrophobic force law. However, it can be used in the research of HI indirectly.

A complete lipophilicity potential model should include both lipophilic effects and hydrophilic effects. The nature of lipophilicity has been illustrated by many authors [7,10,41,43]. Water has some unique properties, such as a very small size, two hydrogen-bond donors and two acceptors, and tetrahedral charge distributions. Water is a highly associated liquid by hydrogen bonds. On average, every water molecule has 3.0–3.5 hydrogen bonds. Around an inert solute molecule, the water molecules actually have higher coordination (of four hydrogen bonds) and thus are even more ordered than the bulk liquid [41]. Both theoretical and experimental studies indicate that the reorientation, or restructuring, of water around nonpolar solutes or surfaces is entropically very unfavorable, since it disrupts the existing water structure and imposes a new and more ordered structure on the surrounding water molecules [41]. On the other hand, the meaning of hydrophilicity is not so clear as lipophilicity. Israelachvili [41] gives an explanation for the hydrophilicity: 'While there is no phenomenon actually known as the hydrophilic effect or the hydrophilic interaction, such effects can be recognized in the property of certain molecules and groups to be soluble and to repel each other strongly in water, in contrast to the strong attraction exhibited by hydrophobic groups'. In the HMLP, hydrophilicity effects include the interactions of dipole and multipole moments, charged atoms in a molecule, and hydrogen bonding. In other words, it is a unified lipophilicity and hydrophilicity potential in the same measure system. Figure 1 shows a unified lipophilicity and hydrophilicity measure system.

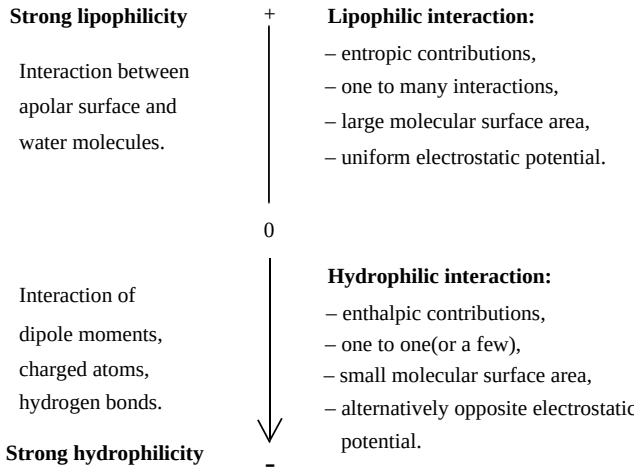


Fig. 1. A measure system for unified lipophilicity and hydrophilicity potential.

We follow the conventions of empirical MLP [24,28,29] and atomic lipophilicity indices [26,27]; in HMLP the positive values are used to represent lipophilicity and the negative values are used for hydrophilicity.

Studies of rare gases and hydrocarbons in water show that although ΔH of the solution is negative, such compounds are insoluble [17,41]. The reason is the large negative change of ΔS in this process [41,43,44]. Hydrophobic interaction is an entropy-dominant phenomenon. According to Frank and Evans [44], when organic compounds are placed in water, the water molecules arrange themselves around the nonpolar parts in what is termed 'iceberg' structures. The arrangement of water molecules in the hydration shell is more ordered than in the bulk of the solvent, a property analogous to that of ice. However, the density in the hydration shell is higher than in the bulk, which is not true for ice [17]. Frank and Evans made their conclusion 50 years ago; however, all recent studies support their conclusion [6,7]. The investigation results of Monte Carlo simulations and molecular dynamic studies provide a good insight into the structure of the hydration shell [6–10]. These studies show that, in the hydration shell, water molecules are arranged in a more ordered manner than in the bulk of the solvent.

Chemists quite often think that a molecule consists of charged atoms. A simple picture for molecular interactions is that of positively charged atoms attracting the negatively charged atoms and repulsing the positively charged atoms. In some cases (for example, in the qualitative studies of hydrogen bonding, dipole interactions, nucleophilic and electrophilic attacks), this is a good approximation. In these cases, the atoms under consideration are strongly charged. However, this model is not always true. Particularly in the study of molecular lipophilicity, this simple approximation may give a wrong picture.

As mentioned in the Introduction section, MEP is the best physical property to describe molecular interactions. However, so far it has not been so successful in describing lipophilicity, as in the studies of hydrogen bonds [45–47] and nucleophilic and electrophilic attacks, although a number of promising efforts have been made by some authors [21–23]. The reason is that molecular lipophilicity is an entropy-dominated phenomenon, dealing with interactions with a huge number of water molecules. Molecular lipophilicity cannot be illustrated just based on the MEP distributions on individual atoms, unlike hydrogen bonding; in that case the maximum and minimum of MEP are sufficient for the qualitative description. However, the description of lipophilicity needs a large neighborhood in the molecular environment. Here we show an example of MEP distribution on the pentanoic acid mol-

TABLE 1
ATOMIC CHARGES, SURFACE AREAS, AND SURFACE–MEP DESCRIPTORS OF PENTANOIC ACID

Atom	q_i	S^+	S^-	S_{total}	b^+	b^-	b_i
C_1(COOH)	0.3099	34.48	29.73	64.21	0.4490	–0.6028	–0.1538
=O	–0.2722	0	35.54	35.54	0	–2.3205	–2.3205
O (OH)	–0.2951	3.064	26.65	29.72	0.0374	–0.9528	–0.9154
H (OH)	0.2013	30.34	0	30.34	1.668	0	1.6682
C_2	–0.1262	43.78	4.196	47.98	0.8951	–0.0346	0.8605
H	0.0738	13.73	0	13.73	0.4571	0	0.4571
H	0.0738	13.73	0	13.73	0.4572	0	0.4572
C_3	–0.0948	26.82	16.78	43.60	0.3817	–0.3170	0.0647
H	0.0596	12.38	0.6037	12.98	0.1868	–0.0013	0.1855
H	0.0596	12.38	0.6037	12.98	0.1870	–0.0013	0.1857
C_4	–0.0941	47.98	0	47.98	0.6223	0	0.6223
H	0.0527	13.73	0	13.73	0.3154	0	0.3154
H	0.0527	13.73	0	13.73	0.3156	0	0.3156
C_5	–0.1776	69.68	16.97	86.65	0.4647	–0.0174	0.4473
H	0.0592	13.13	0	13.13	0.2446	0	0.2446
H	0.0587	12.98	0	12.98	0.2326	0	0.2326
H	0.0587	12.98	0	12.98	0.2332	0	0.2332

Calculated by Gaussian 92 at the 6-31G* level, in atomic units.

ecule in Table 1. The data in Table 1 are calculated by Gaussian 92 at the 6-31G* level. Geometry is optimized at the STO-3G level, and the molecular surface is the van der Waals fused-sphere surface. In Table 1, S_{total} represents the exposed atomic total area on the molecular surface. S^+ and S^- are the surface areas of positive and negative MEP, respectively, and b^+ and b^- are atomic positive and negative MEP descriptors defined by the following equations:

$$b_i^+ = \sum_k V^+(\mathbf{r}_k)\Delta s_k \tag{10}$$

$$b_i^- = \sum_k V^-(\mathbf{r}_k)\Delta s_k \tag{11}$$

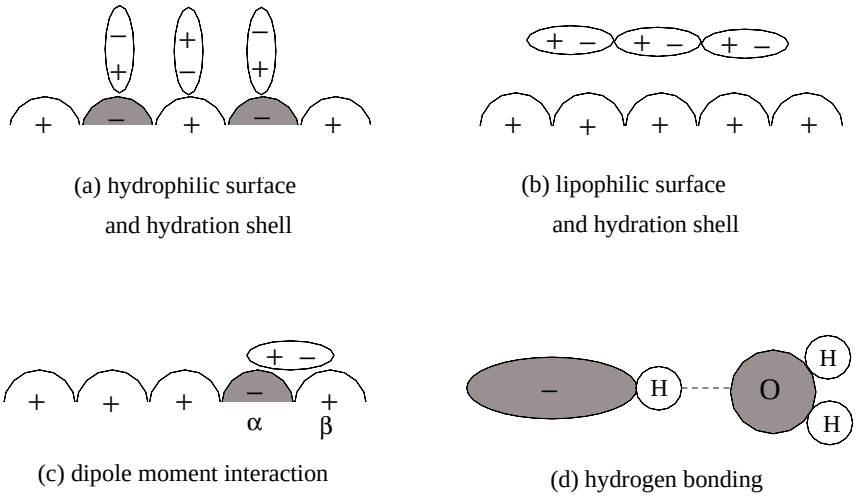
$$b_{total} = b_i^+ + b_i^- \tag{12}$$

Atomic charge q_i is the sum of nuclear charge Z_i and electronic charge $q_i^{(e)}$ on the atom i, $q_i = Z_i + q_i^{(e)}$. In quantum chemistry, electronic charges are usually obtained based on Mulliken population analysis. As shown in Eq. 3, there are two different contributions for $V(\mathbf{r})$: nuclear charges and electron density $\rho(\mathbf{r})$. There is the possibility that on the surface of a negatively charged carbon, $V(\mathbf{r})$ is positive. As shown in Table 1, except for C_1, which is in the carboxyl group -COOH, all the other carbons have negative net atomic charges q_i. However, except for C_1, all the other carbons have positive MEP descriptors b_{total}. All hydrogen atoms on the hydrocarbon chain have positive b_{total} too.

This example shows that for the lipophilic hydrocarbon chain, net atomic charges and MEP–surface descriptors tell us different stories. In the hydrocarbon chain, both carbon and hydrogen atoms are weakly charged and have weak interaction with water molecules. If one thinks that carbon and hydrogen atoms are negatively and positively charged, alternating in a lipophilic hydrocarbon chain, the structure of water molecules surrounding this lipophilic surface is the same as that surrounding the hydrophilic surface, as shown in Fig. 2a. However, based on the surface–MEP descriptors, the b_{total}'s of both carbon and hydrogen atoms in the hydrocarbon chain are positive, and water molecules arrange themselves tangentially to the lipophilic surface, as shown in Fig. 2b. If one uses the MEP-equivalent formal charges, one will find that both carbons and hydrogens in the lipophilic hydrocarbon chain have positive formal charges.

The lipophilic effect is mainly dominated by negative entropic change. Usually, chemists think that interactions of dipole and multipole moments, charged atoms, dispersions, and hydrogen bonding are electrostatic interactions, and lipophilic interaction has a nonelectrostatic origin and is dominated entropically [41]. From the macroscopic point of view, this is true; however, from the microscopic point of view, lipophilic interaction is also electrostatic interaction.

Figure 2 shows four types of interactions between various molecular surfaces and water molecules, where the signs + and – stand for positive and negative MEP, respectively. In Fig. 2a, the hydrophilic surface consists of atoms with opposite MEP alternatively, such as the silica and fiber surfaces. In the hydration shell around the hydrophilic surface, water molecules bind on the surface in an energy-favorable way: the positive and negative ends of water molecules stick on the surface alternatively. This arrangement is similar to the structure of water molecules in the bulk of the aqueous solution. Figure 2b shows the lipophilic surface, which consists of atoms with uniform MEP of the same sign [40]. In the hydration shell surrounding the lipophilic surface, water molecules are placed tangentially on the lipophilic surface. This is more favorable for energy than having water molecules arrange themselves parallel and perpendicular to the surface,



(a) hydrophilic surface and hydration shell

(b) lipophilic surface and hydration shell

(c) dipole moment interaction

(d) hydrogen bonding

Fig. 2. (a) Hydrophilic surface and hydration shell, (b) lipophilic surface and hydration shell, (c) local dipole moment, and (d) hydrogen bonding. Signs + and – are for MEP and not for charge.

resulting in strong repulsive interactions between water molecules. However, this structure is unfavorable with respect to entropy. This model of the lipophilic surface and the structure of the hydration shell is supported by Monte Carlo simulation conducted by Guillot et al. [7]. They find that in the hydration shell around a lipophilic surface, 'water molecules are arranged, on average, tangentially to the (lipophilic) solute molecule.' Figure 2c shows a local dipole moment on the molecular surface where the atom α is bordering another atom β with the opposite MEP. As shown in Fig. 2d, hydrogen bonding is observed in circumstances where a hydrogen atom covalently bonds to an electronegative atom, such as oxygen, nitrogen or a halogen. It is clear that at a point **r** on the molecular surface, the lipophilicity potential is not decided only by the atom the point **r** belongs to, but by the molecular environment.

A complete theoretical derivation of the free energy change from first principles, explicitly including a huge number of water molecules, is not easy. No rigorous theoretical method is available for this task. However, a heuristic lipophilicity potential model is likely to be sufficient for some tasks in molecular modeling and chemical design. Here we suggest a unified lipophilicity and hydrophilicity potential model, as follows:

$$L(\mathbf{r}) = V(\mathbf{r}) \sum_{\mathbf{r}\in S_\alpha} \sum_{i\neq\alpha} M_i(\mathbf{r};\mathbf{R}_i,b_i) \tag{13}$$

where $V(\mathbf{r})$ is the MEP at point **r**, and **r** is on the surface S_α of atom α. In the sum, $M_i(\mathbf{r};\mathbf{R}_i,b_i)$ is the screening function on position **r** from atom i. In Eq. 13 summation is over all constituent atoms except atom α. In the screening function $M_i(\mathbf{r};\mathbf{R}_i,b_i)$, $\mathbf{R}_i$ is the nuclear position of atom i, and b_i is the atomic surface–MEP descriptor of atom i [40]:

$$b_i = \sum_{k\in S_i} V(\mathbf{r}_k)\Delta s_k \tag{14}$$

where Δs_k is the area element on atom i. Summation is over all exposed surfaces of atom i. The atom-based screening function $M_i(\mathbf{r};\mathbf{R}_i,b_i)$ can take a number of forms. Three possible formulas of $M_i(\mathbf{r};\mathbf{R}_i,b_i)$ are as follows:

$$M_i(\mathbf{r};\mathbf{R}_i,b_i) = \frac{r_0^\gamma}{b_0}\frac{b_i}{\|\mathbf{R}_i-\mathbf{r}\|^\gamma} = \zeta\frac{b_i}{\|\mathbf{R}_i-\mathbf{r}\|^\gamma} \tag{15}$$

$$M_i(\mathbf{r};\mathbf{R}_i,b_i) = \frac{b_i}{b_0}\exp\left(-\frac{\|\mathbf{R}_i-\mathbf{r}\|}{d_0}\right) \tag{16}$$

$$\begin{aligned} M_i(\mathbf{r};\mathbf{R}_i,b_i) &= \frac{b_i}{|b_i|}\exp\left(-\frac{b_0}{\lambda_0}\frac{\|\mathbf{R}_i-\mathbf{r}\|}{|b_i|}\right) \\ &= \mathrm{sign}(b_i)\exp\left(-\xi\frac{\|\mathbf{R}_i-\mathbf{r}\|}{|b_i|}\right) \end{aligned} \tag{17}$$

In the above three functions (r_0,b_0,γ), (b_0,d_0) and (b_0,λ_0) are parameters. The unit of b_0 is the same as b_i (energy × area); r_0, d_0, and λ_0 have a unit of length. Therefore, $M_i(\mathbf{r};\mathbf{R}_i,b_i)$ is a dimensionless function in all the above three equations. In Eq. 15, $\zeta = (r_0)^\gamma/b_0$ is a simple scale factor. In the later calculations, we take $\zeta = 1$. Exponent γ in Eq. 15 is the parameter which decides how strong the influence is and how rapidly the influence decays with distance. It will be optimized in this study. In Eq. 16, b_0 is a simple scale factor, and d_0 plays the same role as γ in Eq. 15. In Eq. 17, $\xi = b_0/\lambda_0$ makes the exponent a dimensionless quantity, and affects the behaviors of the screening function. ξ and d_0 will be optimized in later research. In this study, we test the screening function 1 (Eq. 15) and optimize parameter γ based on the experimental partition coefficients $\log P_{ow}$. The heuristic MLP defined by Eq. 13 and the properties of the screening function $M_i(\mathbf{r};\mathbf{R}_i,b_i)$ can be interpreted as follows:

(1) The lipophilicity potential $L(\mathbf{r})$ is an average, or modified, electrostatic potential. $\Sigma_{i\neq\alpha}M_i(\mathbf{r};\mathbf{R}_i,b_i)$ is the modifying factor which represents the influence from all surrounding atoms at point **r**.

(2) If $\Sigma_{i\neq\alpha}M_i(\mathbf{r};\mathbf{R}_i,b)$ has the same sign as $V(\mathbf{r})$, then the lipophilicity potential $L(\mathbf{r})$ is positive, and the point **r** is lipophilic. However, if $\Sigma_{i\neq\alpha}M_i(\mathbf{r};\mathbf{R}_i,b_i)$ has a sign opposite to that of $V(\mathbf{r})$, then the lipophilicity potential $L(\mathbf{r})$ is negative, and the point **r** is hydrophilic.

(3) The influences from all other atoms decay with the distance $\|\mathbf{R}_i-\mathbf{r}\|$. In Eqs. 15–17, γ, d_0, and ξ are parameters which decide how strong the influences are and how rapidly the influences decay with distance.

(4) The atomic surface–MEP descriptor b_i defined by Eq. 14 represents the MEP distribution and surface condition of atom i. The effect of b_i is similar to the empirical lipophilic parameters f_i in Eq. 2; however, b_i's are theoretical and structural parameters instead of empirical parameters.

(5) It is best to think of $\Sigma_{i\neq\alpha}M_i(\mathbf{r};\mathbf{R}_i,b_i)$ as a unitless modifying factor. Therefore, the unit of the lipophilicity potential $L(\mathbf{r})$ is the same as that of the electrostatic potential $V(\mathbf{r})$. However, there is no direct, simple connection between the values of $L(\mathbf{r})$ and the free energy of the solution system.

The heuristic lipophilicity potential developed in this research is a structure-based potential. In the calculation of the heuristic MLP based on Eqs. 13–17, the input data are molecular geometries and molecular surfaces. This technique can be implemented for realistic contour surfaces determined from ab initio electron densities [48] and can also be used for various empirical molecular surfaces. In the case of a van der Waals surface, it is a fused-sphere surface. Quantum chemical methods are used to calculate the electron density function $\rho(\mathbf{r})$ and the electrostatic potential $V(\mathbf{r})$ following the definition of Eq. 3. For the shape analysis of the heuristic lipophilicity potential, the

shape group methods are applicable [48–50]. Whereas these methods are designed for small molecules, a new technique for the calculation of the electron density function $\rho(\mathbf{r})$ of large biomolecules was developed by Walker and Mezey [51,52]. This technique, called molecular electron density Lego assembler (MEDLA), has been used to construct ab initio quality electron densities at the 6-31G** level for proteins containing more than 1000 atoms. The macromolecular density matrix method, the adjustable density matrix assembler (ADMA) [53,54], is a method that no longer needs an extensive numerical density database, and appears advantageous for MEP applications. Therefore, there is no insurmountable difficulty for the application of HMLP in drug design and QSAR studies. In the HMLP there are no empirical parameters of atomic lipophilicity indices used; therefore, it is a nonempirical MLP. However, it is not a rigorous theoretical model because there is no rigorous theoretical derivation of all aspects of the model from first principles. Therefore, we regard it as a heuristic model.

Simple examples and tests of HMLP

In this section, we show simple examples and the tests applied to the calculation of results for several small molecules: ethanol (C_2H_5OH), *n*-propylamine ($C_3H_7NH_2$), and *n*-propanoic acid (C_2H_5COOH). In the following calculations, the screening function $M_i(\mathbf{r};\mathbf{R}_i,b_i)$ takes the form of Eq. 15. Based on Eq. 13, we can define the atomic lipophilic index l_a as follows:

$$l_a = \sum_{i \in S_a} L(\mathbf{r}_i)\Delta s_i \tag{18}$$

where summation is over the exposed area S_a of atom a. If $l_a > 0$, then atom a is lipophilic; if $l_a < 0$, then atom a is hydrophilic. The molecular lipophilic index (L_M) and the hydrophilic index (H_M) are the sum of the corresponding values for all lipophilic atoms and hydrophilic atoms, respectively:

$$L_M = \sum_{\substack{a \\ (l_a>0)}} l_a \tag{19}$$

TABLE 2
HEURISTIC AND EMPIRICAL LIPOPHILICITY INDICES OF ETHANOL

Atom	Heuristic indices l_a (hartree bohr^2)	Empirical indices $l_a^{(em)}$ (no unit)
(-CH_2-)		
C_1	0.07855	2.839
H	0.00667	0.9255
H	0.00668	0.9248
(-CH_3)		
C_2	0.02297	11.50
H	0.00318	2.068
H	0.00057	1.777
H	0.00058	1.798
(-OH)		
O	–0.1722	0.1106
H	–0.0707	–1.367
L_M	0.1192	21.94
H_M	–0.2429	–1.367

$$H_M = \sum_{\substack{a \\ (l_a<0)}} l_a \tag{20}$$

Table 2 shows the atomic lipophilic indices l_a of ethanol. Molecular geometry is optimized with the STO-3G basis set, and a fused-sphere van der Waals surface [55] is used ($R_C = 2.00$ Å, $R_H = 1.17$ Å, $R_O = 1.39$ Å). Electron density is calculated by the Gaussian 92 program at the 6-31G* level. The surface grid point density is chosen as 25 points Å^2 [56–58]. In this example, the exponent of Eq. 15 is taken as $\gamma = 2.50$. In Table 2, the empirical lipophilic indices of ethanol are calculated using Eqs. 18–20, based on the empirical lipophilicity potential of Eq. 2 [24] and empirical atomic lipophilic indices [27]. As expected, the calculations of HMLP show that the hydrocarbon part takes positive values (lipophilic) and the hydroxyl group takes negative values (hydrophilic). Oxygen is the most hydrophilic atom and C_1 is the most lipophilic atom in ethanol. However, in the empirical calculations, the atomic lipophilic index of oxygen is (surprisingly) positive (lipophilic), and the molecular lipophilic index L_M is much

TABLE 3
COMPARISON OF DIFFERENT BASIS SETS AND SURFACE POINT DENSITIES USING ETHANOL AS AN EXAMPLE

Point density	Basis set	Total points	Index l_O	Index l_H	Index L_M	Index H_M
10	6-31G*	738	–0.1729	–0.0717	0.1125	–0.2446
25	6-31G*	1730	–0.1722	–0.0707	0.1192	–0.2429
50	6-31G*	2483	–0.1737	–0.0718	0.1179	–0.2456
25	4-31G*	1730	–0.1689	–0.0714	0.1144	–0.2402
25	3-21G*	1730	–0.1940	–0.0877	0.1255	–0.2825
25	4-31G	1730	–0.2253	–0.0896	0.1566	–0.3149
25	STO-3G	1730	–0.0964	–0.0426	0.0634	–0.1396

Results in atomic units: hartree bohr^2.

TABLE 4
TEST CALCULATIONS OF THE EXPONENT γ USING *n*-PROPYLAMINE

Exponent γ	l_N (-NH_2)	l_H (-NH_2)	L_M	H_M
1.00	–1.3227	0.1663	1.8444	–1.3227
1.50	–0.6073	0.0384	0.8151	–0.6073
2.00	–0.2909	0.01326	0.3762	–0.2909
2.40	–0.1658	–0.00490	0.2195	–0.1756
2.50	–0.1445	–0.00544	0.1930	–0.1554
2.60	–0.1262	–0.00571	0.1698	–0.1376
2.70	–0.1102	–0.00578	0.1497	–0.1218
2.80	–0.0963	–0.00569	0.1320	–0.1078
3.00	–0.0740	–0.00525	0.1031	–0.0845

In atomic units: hartree $bohr^2$.

TABLE 5
COMPARISON OF THE USE OF DIFFERENT ATOMIC RADII FOR *n*-PROPYLAMINE (γ = 2.50)

Atomic radii (Å)	l_N (-NH_2)	l_H (-NH_2)	L_M	H_M
$R_C = 2.00$, $R_N = 1.46$, $R_H = 1.17$	–0.1445	–0.00544	0.1930	–0.1554
$R_C = 1.75$, $R_N = 1.46$, $R_H = 1.17$	–0.1870	0.00637	0.3266	–0.1870
$R_C = 1.75$, $R_N = 1.55$, $R_H = 1.17$	–0.1914	–0.00262	0.3047	–0.1967
$R_C = 2.00$, $R_N = 1.55$, $R_H = 1.17$	–0.1452	–0.01008	0.1891	–0.1653

In atomic units: hartree $bohr^2$.

higher than the hydrophilic index H_M. Based on the chemical intuitions for ethanol, the result of empirical lipophilicity potential is not satisfactory.

Table 3 gives the comparison using different basis sets and surface point densities in the calculation of ethanol. All the other conventions are the same as in Table 2. As shown in Table 3, there is not much difference using point densities 10, 25, and 50 points/$Å^2$. All four types of basis sets give qualitatively the same results; however, when the poor basis set STO-3G is used, the numerical results are very different from the others. The basis set in this type of calculation should include polarization functions.

The exponent γ in Eq. 15 is an important parameter. Table 4 summarizes the results of calculations for *n*-propylamine using different values of γ. A point density of 25 and the basis set 6-31G* are used. Atomic radii are R_C = 2.00 Å, R_H = 1.17 Å, and R_N = 1.46 Å. Molecular geometry is optimized at the STO-3G level. In the amino group -NH_2, there are two positively charged hydrogen atoms and one negatively charged nitrogen atom; therefore, the lipophilic indices of the two hydrogen atoms are very sensitive to the exponent γ. As shown in Table 4, when $\gamma \leq 2.00$, l_H is positive, and for $2.40 \leq \gamma \leq 3.00$, l_H is negative. Beyond 2.40, l_H remains essentially constant; however, l_N decreases remarkably. In further calculations we have used the value γ = 2.50. However, this is not a rigorous optimization, and γ = 2.50 may not be the best value.

A molecular surface can be regarded as the molecular interaction interface [48]. The generation of a molecular surface is a key step in this approach. Table 5 lists the results of calculations for *n*-propylamine using different atomic radii. Exponent γ is taken as 2.50. All the other conditions are the same as in Table 4. From Table 5, we know that for a fused-sphere van der Waals surface, atomic radii affect the atomic lipophilic indices to a certain degree. For simplicity, in this research we use van der Waals fused-sphere surfaces [56–58]. Atomic radii are optimized based on MEP criteria by Du and Arteca [55]. These results are very close to the results of Ooi et al. [59]. However, our optimizations did not include all atomic types. It should be pointed out that theoretical electron isodensity surfaces [48] might be more reliable in this type of research.

Politzer and his research group [45] have studied hydrogen bonds using MEP successfully in an extended region. Table 6 lists the heuristic and empirical atomic lipophilic indices l_a's of functional groups (-NH_2, -OH, and -COOH) of *n*-propanol, *n*-propylamine, and *n*-propa-

TABLE 6
ATOMIC LIPOPHILIC INDICES OF FUNCTIONAL GROUPS (-NH_2, -OH, AND -COOH)

Functional groups	Heuristic indices (hartree $bohr^2$)	Empirical indices
-NH_2 ($C_3H_7NH_2$)	$l_N = -0.1445$	$l_N^{(em)} = -2.8713$
	$l_H = -0.0054$	$l_H^{(em)} = -3.6734$
-OH (C_3H_7OH)	$l_O = -0.1693$	$l_O^{(em)} = 0.4957$
	$l_H = -0.0717$	$l_H^{(em)} = -0.8826$
-COOH (C_2H_5COOH)	$l_{=O} = -0.0915$	$l_{=O}^{(em)} = -0.5734$
	$l_O = -0.0653$	$l_O^{(em)} = 0.9178$
	$l_H = -0.1074$	$l_H^{(em)} = -1.3931$
	$l_C = 0.0569$	$l_C^{(em)} = 1.4863$

TABLE 7
HYDROGEN-BONDING ENERGIES AND LIPOPHILIC INDICES OF CARBOXYLIC ACID

Type	Hydrogen bonds	6-31G(d)[a] (kcal/mol)	HMLP indices (atomic units)
a. CH_3–C(=O)–OH···O(H)H	Donor	–8.5	$l_H = -0.1052$
b. CH_3–C(=O···H–O–H)–OH	Acceptor	–5.5	$l_{=O} = -0.0930$
c. CH_3–C(=O)–O(H)···H–O–H	Acceptor	–2.4	$l_O = -0.0614$

[a] Reference 46.

noic acid, which can be used to study hydrogen bonds. We find that heuristic indices l_H's in the three types of functional groups are –0.0054, –0.0717, and –0.1075, respectively, in reasonable order with respect to their hydrogen-bonding donor strength.

For carboxylic acid, the hydrogen-bond energies of three possible types of hydrogen bonds and lipophilic indices are listed in Table 7. The hydrogen-bond energies are taken from the results of Gao [46]. Our results of lipophilic indices are in good agreement with the hydrogen-bond energies. Compared with other hydrogen-bond indices, such as MEP [45], MFP [47] and Mulliken population, one advantage of lipophilicity indices l_a's is that both hydrogen-bond donors and acceptors use the same indices l_a.

Partition coefficients and HMLP

Partition coefficients $\log P_{ow}$ are experimental data. For a long time, $\log P_{ow}$ has been used as the overall measure

TABLE 8
MOLECULAR LIPOPHILIC AND HYDROPHILIC INDICES OF 41 MOLECULES

	L_M	H_M	$L_M^{(em)}$	$H_M^{(em)}$	Exptl $\log P_{ow}$	Calcd $\log P_{ow}$
Hydrocarbons						
CH_4	0.0128	0.0000	65.55	0.0000	1.09	1.07
C_2H_6	0.0417	0.0000	94.56	0.0000	1.81	2.03
C_3H_8	0.0474	0.0000	144.0	0.0000	2.36	2.22
C_4H_{10}	0.0643	0.0000	202.4	0.0000	2.89	2.78
C_5H_{12}	0.0803	0.0000	263.0	0.0000	3.39	3.32
C_6H_{14}	0.1028	0.0000	336.3	0.0000	3.90	4.06
C_7H_{16}	0.1155	0.0000	404.8	0.0000	5.18	5.13
C_8H_{18}	0.1348	0.0000	487.5	0.0000	–	–
C_9H_{20}	0.1492	0.0000	564.6	0.0000	–	–
$C_{10}H_{22}$	0.1684	0.0000	657.4	0.0000	–	–
Alcohol						
CH_3OH	0.1358	–0.2350	1.316	–3.856	–0.817	–0.798
C_2H_5OH	0.1192	–0.2429	21.95	–1.367	–0.315	–0.387
C_3H_7OH	0.1577	–0.2410	42.62	–0.883	0.342	0.281
C_4H_9OH	0.1792	–0.2422	68.07	–0.4017	0.817	0.876
$C_5H_{11}OH$	0.2105	–0.2421	97.23	0.0000	1.457	1.562
$C_6H_{13}OH$	0.2338	–0.2421	129.9	0.0000	1.982	2.080
$C_7H_{15}OH$	0.2582	–0.2409	165.5	0.0000	2.542	2.504
$C_8H_{17}OH$	0.2789	–0.2408	203.6	0.0000	3.062	2.954
$C_9H_{19}OH$	0.2999	–0.2407	244.1	0.0000	–	–
$C_{10}H_{21}OH$	0.3172	–0.2405	286.0	0.0000	–	–
Amine						
CH_3NH_2	0.06829	–0.1768	0.000	–27.53	–0.564	–0.607
$C_2H_5NH_2$	0.1394	–0.1497	5.470	–14.28	–0.253	–0.560
$C_3H_7NH_2$	0.1930	–0.1554	20.48	–11.50	0.010	0.317
$C_4H_9NH_2$	0.2434	–0.1500	40.65	–8.916	0.779	0.796
$C_5H_{11}NH_2$	0.2804	–0.1500	65.82	–7.336	1.080	1.27
$C_6H_{13}NH_2$	0.3160	–0.1499	94.83	–6.009	1.526	1.73
$C_7H_{15}NH_2$	0.3457	–0.1499	127.2	–4.859	2.183	2.12
$C_8H_{17}NH_2$	0.3738	–0.1499	162.3	–3.727	2.789	2.48
$C_9H_{19}NH_2$	0.3983	–0.1499	200.0	–2.715	–	–
$C_{10}H_{21}NH_2$	0.4217	–0.1499	240.0	–1.736	–	–
Acid						
HCOOH	0.1227	–0.2434	0.0000	–29.57	–0.539	–0.657
CH_3COOH	0.2160	–0.2596	17.67	–3.790	–0.310	–0.019
C_2H_5COOH	0.2247	–0.2642	35.44	–1.967	0.253	–0.057
C_3H_7COOH	0.3141	–0.2634	59.69	–1.042	0.766	1.07
C_4H_9COOH	0.3593	–0.2702	87.48	–0.4851	1.411	1.41
$C_5H_{11}COOH$	0.4098	–0.2701	119.4	–0.0523	2.017	2.04
$C_6H_{13}COOH$	0.4428	–0.2701	154.0	0.0000	2.642	2.44
$C_7H_{15}COOH$	0.4773	–0.2700	191.9	0.0000	–	–
$C_8H_{17}COOH$	0.5033	–0.2700	231.5	0.0000	–	–
$C_9H_{19}COOH$	0.5299	–0.2699	274.0	0.0000	–	–
$C_{10}H_{21}COOH$	0.5517	–0.2699	317.9	0.0000	–	–

In atomic units: hartree $bohr^2$.

of molecular lipophilicity. The quantity $\log P_{ow}$ is related to the molar standard free energy ΔG°_{tr} of transfer from the aqueous phase to the organic phase:

$$\log P_{ow} = \Delta G^{\circ}_{tr} / 2.3026RT \quad (21)$$

HMLP is the 3D representation of molecular lipophilicity. It gives the details of lipophilicity information in the molecular space or surface. On the other hand, partition coefficients are 1D scalar descriptors. Many authors [60,61] point out that $\log P_{ow}$ values are not very sensitive to the change of molecular lipophilicity. The reason is that sometimes the changes of two components of transfer free energy ($\Delta G^{\circ}_{tr} = \Delta H^{\circ}_{tr} - T\Delta S^{\circ}_{tr}$), enthalpy and entropy, cancel each other. Partition coefficients are also affected by intramolecular interactions such as hydrogen bonding. The conversion from 3D HMLP to $\log P_{ow}$ is not so straightforward. Here we do not try to present a general method for the calculation of $\log P_{ow}$ from HMLP, and to compete with other methods. Our purpose is to check the reasonableness of HMLP based on experimental data log P_{ow} using several families of compounds.

In this section, we use heuristic lipophilicity potential to calculate $\log P_{ow}$, and to test our method. Also, we make a comparison between heuristic MLP and empirical MLP [25,28,29]. Table 8 lists the heuristic and empirical molecular lipophilic and hydrophilic indices of 41 molecules, including linear hydrocarbons, aliphatic amines, alcohols, and acids. The data of partition coefficients log P_{ow} are from Leo et al. [17]. If one assumes that the relationship between $\log P_{ow}$ and lipophilic indices L_M and hydrophilic indices H_M is linear, then the following approximation can be used:

$$\log P_{ow} = C_0 + C_1 L_M + C_2 H_M \quad (22)$$

The linear coefficients are determined by least-squares fitting to experimental data for $\log P_{ow}$. We show below the correlations restricted for families of compounds. The results include the correlation coefficient (r), the standard error (σ), the number of compounds in the fitting (n), and the standard errors in the linear coefficients for the compounds.

(1) Linear hydrocarbons ($n = 7$, $r = 0.995$, $\sigma = 0.152$):

$$\log P_{ow} = (0.6444 \pm 0.1197) + (33.28 \pm 1.519)L_M \quad (23)$$

(2) Aliphatic alcohols ($n = 8$, $r = 0.997$, $\sigma = 0.130$):

$$\log P_{ow} = (-27.01 \pm 4.85) + (22.22 \pm 0.87)L_M + (-98.71 \pm 20.29)H_M \quad (24)$$

(3) Aliphatic amines ($n = 8$, $r = 0.980$, $\sigma = 0.282$):

$$\log P_{ow} = (-7.18 \pm 2.78) + (12.94 \pm 1.45)L_M + (-32.18 \pm 16.29)H_M \quad (25)$$

(4) Aliphatic carboxylic acids ($n = 7$, $r = 0.978$, $\sigma = 0.305$):

$$\log P_{ow} = (5.53 \pm 6.81) + (12.34 \pm 2.34)L_M + (31.63 \pm 28.23)H_M \quad (26)$$

For hydrocarbons, all hydrophilic indices (H_M) are zero; therefore, there is only one parameter L_M in Eq. 23. Considering that only two parameters (L_M and H_M) are used, the correlation results are encouraging. In particu-

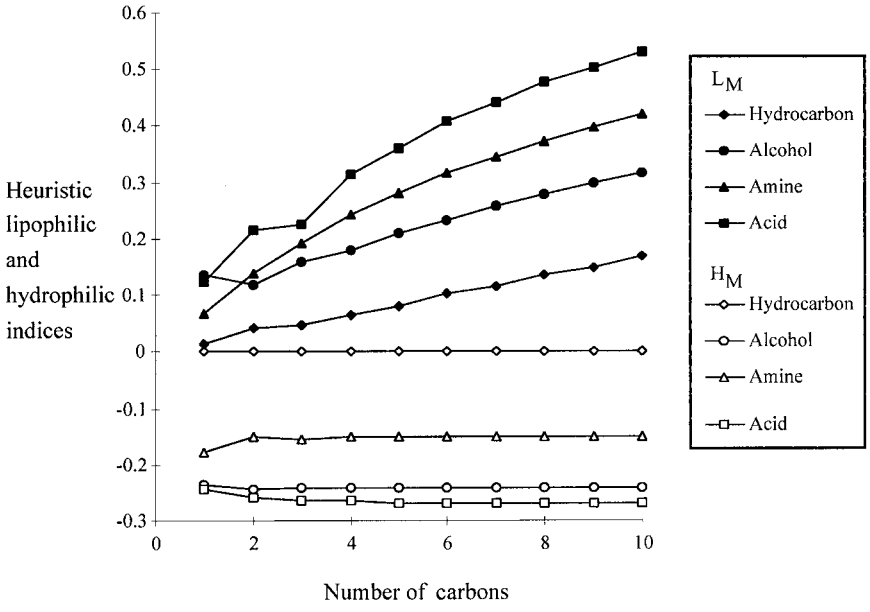


Fig. 3. Molecular heuristic lipophilic indices L_M (top four curves) and hydrophilic indices H_M (bottom four curves) as a function of the number of carbon atoms for linear hydrocarbons, aliphatic alcohols, amines, and acids.

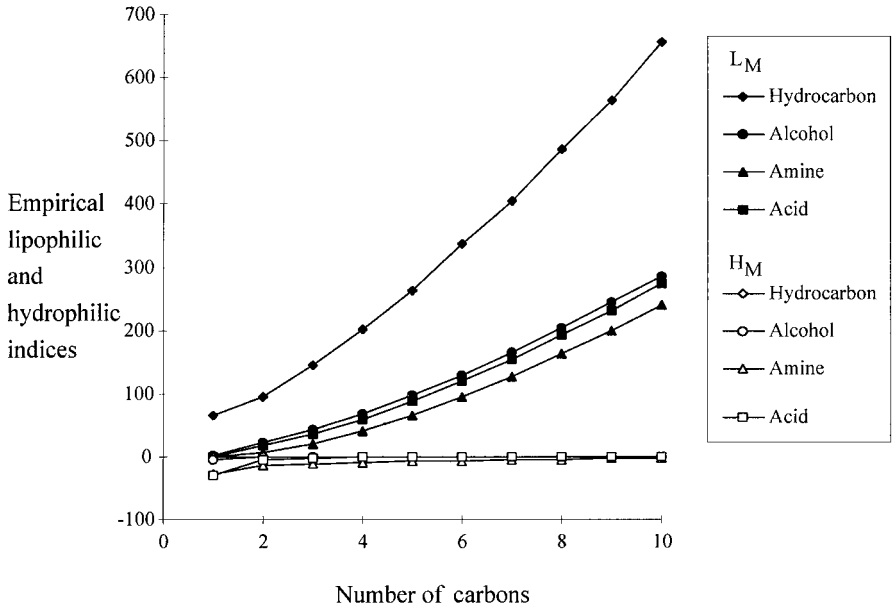


Fig. 4. Molecular empirical lipophilic indices $L_M^{(em)}$ (top four curves) and hydrophilic indices $H_M^{(em)}$ (bottom four curves, some nearly coincident) as a function of the number of carbon atoms for linear hydrocarbons, aliphatic alcohols, amines, and acids.

lar, for aliphatic alcohols the correlation coefficient (r = 0.997) is very good.

Molecular lipophilic indices L_M and hydrophilic indices H_M are functions of molecular structure. Figure 3 shows the behaviors of L_M and H_M of HMLP as a function of the number of carbon atoms for four types of compounds (linear hydrocarbons, aliphatic amines, alcohols, and acids). In Fig. 3, the top four curves are the theoretical lipophilic indices L_M of four types of compounds, and the bottom four curves are the theoretical hydrophilic indices H_M. Figure 4 shows the empirical indices $L_M^{(em)}$ and $H_M^{(em)}$.

We can see in Fig. 3 that for a family of compounds, the heuristic lipophilic indices L_M increase with an increase in the number of carbon atoms. However, these functions are not exactly linear. The increase becomes less noticeable with an increase in the number of carbon atoms. On the other hand, H_M stays almost constant. The empirical indices $L_M^{(em)}$ increase sharply with an increasing number of carbon atoms. The hydrophilic index H_M should have a negative value if there is a hydrophilic group in the molecule. Indices H_M of heuristic MLP give reasonable results. Except for hydrocarbons (which have no hydrophilic group; therefore all H_M are zero), all the other three types of compounds (alcohols, acids, and amines) have constant negative H_M, which are the contributions from hydrophilic functional groups. However, empirical MLP gives unreasonable results. Only formic acid and methylamine have negative $H_M^{(em)}$. When there are more than two carbon atoms, indices $H_M^{(em)}$ of all acids, amines, and alcohols are zero. The values of $H_M^{(em)}$ for the three types of compounds do not follow chemical intuition. These results show that the atomic lipophilic parameters $\{f_i\}$ provided by Ghose and Crippen [26] and Viswanadhan et al. [27] might need modifications. Also, the exponent $\gamma = 1$ in Eq. 2 is too small. Correlation between experimental data partition coefficients $\log P_{ow}$ and molecular lipophilic and hydrophilic indices L_M and H_M provides a criterion for the optimization of parameter γ in Eq. 15. It also gives a criterion for the selection of the form of the screening function $M_i(\mathbf{r};\mathbf{R}_i,b_i)$.

Discussion and Conclusions

A major goal in chemical research is to predict the behavior of new compounds based on their molecular structures. Quantitative correlations of molecular structures of ligands with the binding constants and, subsequently, the predictions for novel compounds are the task of QSAR. The heuristic lipophilicity potential, defined by Eqs. 13–20, is based on structural information. The input data are molecular geometries and molecular surfaces. Quantum mechanics is used to calculate the electron density function $\rho(\mathbf{r})$ and the electrostatic potential $V(\mathbf{r})$. The examples of this study show that this model gives qualitatively and quantitatively correct, chemically reasonable results, and it works in cases not well described by the empirical lipophilicity potential [24,28,29]. Here our goal is to explore the possibility of this new technique being used in computer-aided rational 3D drug design.

In studies of protein–ligand complexes, there are three main types of factors: steric, electrostatic, and lipophilic. In its original form, 3D QSAR methods use as features direct measurements of the 3D shapes of molecules and the 3D distribution of charges in the molecular space [4]. The 3D method can relate predicted activity to specific portions of a molecule, and directly guide the design of

an improved ligand. Later more molecular properties were included in the 3D QSAR studies, such as MEP and empirical MLP [3]. In the first stage of its development, the comparative molecular field analysis (CoMFA) approach uses in its standard implementation only Lennard-Jones and Coulomb potentials [3]. Calculation results show that these potentials describe the energetic contributions to the binding constants [62]. Later some fields considering the differences in hydrophobic surface contributions are included in order to describe the entropic effects [3]. Kellogg and Abraham [5] have described hydrophobic fields and developed the program HINT. Furthermore, Goodford [30] uses a water probe to measure the hydrophobic surface regions in terms of field, and presents the program GRID. The heuristic lipophilicity potential is a modified electrostatic potential, taking the averaged influences from the environment. Therefore, HMLP is a unified lipophilic-electrostatic potential, it keeps the information of both MEP and MLP, and has a 3D form. These characters appear to make the HMLP useful in 3D drug design. HMLP is not a technique comparable to 3D QSAR and CoMFA. We suggest that HMLP may provide new complementing features to CoMFA and 3D QSAR. Just like empirical MLP [24], heuristic MLP, possibly in combination with a computer graphic technique to show the distribution of MLP on the molecular surface in different colors, provides more detailed information, and gives the expressive visualization of lipophilicity. This technique is expected to be valuable for the study of complementary lipophilic and electrostatic maps on the molecular surfaces in both direct and indirect drug design [5,21–23]. Unlike the method of surface–MEP descriptors [40], in the heuristic MLP approach it is not necessary to assign the hydrophilic functional groups and lipophilic hydrocarbon chains, or designate the lipophilic region and hydrophilic region in a molecule artificially.

Lipophilicity potential is a real physical potential of solute molecules in aqueous solution. It has a complex physical and chemical nature [41], and further research is needed for a complete theoretical treatment of lipophilicity potential from first principles. The heuristic lipophilicity potential developed in this work is just one step towards this goal. The interaction between water molecules and an organic solute molecule at the position around an atom in the molecular surface is affected by all its neighbor atoms. The atom-based screening function $M_i(\mathbf{r};\mathbf{R}_i,b_i)$ describes the influence from atom i. In the model of heuristic lipophilicity potential, the key is to find a good screening function. Besides the atomic surface–MEP descriptor b_i, one might need to include more properties, such as atomic shape parameters, which describe the interfering effect of atoms in a molecule to the water structure. Unlike empirical MLP, the screening function of HMLP has a certain physical meaning. However, in the first stage, we regard it as a mathematical function, and optimize its parameter based on physical and chemical facts, such as experimental data $\log P_{ow}$ or solvation free energies through correlation calculations. The experimental technique that is best suited to provide the answers to the calculation results of HMLP is NMR [63]. HMLP provides information of 3D distributions of lipophilicity in molecular space. NMR gives the details of water structure near certain atomic groups [64–66]; therefore, it may provide a good criterion for further improvements of HMLP.

In this research, HMLPs are calculated by ab initio quantum chemistry. It may appear that this technique is somehow cumbersome and impractical in real drug research. Actually, in the calculation of HMLP, the only property we have to get from quantum chemical calculation is the electron density function $\rho(\mathbf{r})$; see Eq. 3. The technique MEDLA, so-called 'computational microscope' [67], developed earlier [51,52], shows that there is no difficulty in obtaining the electron density functions $\rho(\mathbf{r})$ of common biomolecules, even for protein molecules containing more than 1000 atoms. The example of HIV (human immunodeficiency virus) protease shows that the calculation of electron density at the 6-31G** level needs only 35 min. The more advanced and elaborate method ADMA can give the analytical electron density function $\rho(\mathbf{r})$ [53,54]. Therefore, it is possible to apply HMLP in practical drug research.

Acknowledgements

This research was supported by the Natural Science and Engineering Research Council of Canada (NSERC).

References

1 Hansch, C., In Ariens, E.J. (Ed.) Drug Design, Vol. 1, Academic Press, New York, NY, U.S.A., 1971, p. 271.
2 Loew, G.H., Villar, H.O. and Alkorta, I., Pharm. Res., 10 (1993) 475.
3 Klebe, G., Abraham, U. and Mietzner, T., J. Med. Chem., 37 (1994) 4130.
4 Jain, A.N., Koile, K. and Chapman, D., J. Med. Chem., 37 (1994) 2315.
5 Kellogg, G.E. and Abraham, D.J., J. Mol. Graph., 10 (1992) 212.
6 Forsman, J. and Jönsson, B., J. Chem. Phys., 101 (1994) 5116.
7 Guillot, B., Guissani, Y. and Bratos, S., J. Chem. Phys., 95 (1991) 3643.
8 Tanaka, H. and Nakanishi, K., J. Chem. Phys., 95 (1991) 3719.
9 Haile, J.M., Molecular Dynamics Simulation, Wiley, New York, NY, U.S.A, 1992.
10 Smith, D.E. and Haymet, A.D.J., J. Chem. Phys., 98 (1992) 6444.
11 Miertus, S. and Tomasi, J., Chem. Phys., 65 (1982) 239.
12 Miertus, S. and Moravek, R., Coll. Czech-Slov. Chem. Commun., 55 (1990) 2430.
13 Wong, M.W., Frisch, M.J. and Wiberg, K.B., J. Am. Chem. Soc., 113 (1991) 4776.

14 Bonaccorsi, R., Floris, F., Palla, P. and Tomasi, J., Thermochim. Acta, 162 (1990) 213.
15 Cramer, C.J. and Truhlar, D.G., J. Comput. Chem., 13 (1992) 1089.
16 Field, M.J., Bash, P.A. and Karplus, M., J. Comput. Chem., 11 (1990) 700.
17 Leo, A., Hansch, C. and Elkins, D., Chem. Rev., 71 (1971) 525.
18 Eisenberg, D., Weiss, R.M. and Terwilliger, T.C., Nature, 299 (1982) 371.
19 Eisenberg, D. and McLachlan, A.D., Nature, 319 (1986) 199.
20 Heiden, W., Moeckel, G. and Brickmann, J., J. Comput.-Aided Mol. Design, 7 (1993) 503.
21 Náray-Szabó, G. and Nagy, P., Int. J. Quantum Chem., 35 (1989) 215.
22 Náray-Szabó, G., J. Mol. Graph., 7 (1989) 76.
23 Náray-Szabó, G., J. Mol. Struct. (THEOCHEM), 138 (1986) 197.
24 Croizet, F., Langlois, M.H., Dubost, J.P., Braquet, P., Audry, E., Dallet, Ph. and Colleter, J.C., J. Mol. Graph., 8 (1990) 153.
25 Furet, P., Sele, A. and Cohen, N.C., J. Mol. Graph., 6 (1988) 182.
26 Ghose, A.K. and Crippen, G.M., J. Comput. Chem., 7 (1986) 565.
27 Viswanadhan, V.N., Ghose, A.K., Revankar, G.R. and Robins, R.K., J. Chem. Inf. Comput. Sci., 29 (1989) 163.
28 Audry, E., Dubost, J.-P., Colleter, J.-C. and Dallet, P., Eur. J. Med. Chem.-Chim. Thér., 21 (1986) 71.
29 Audry, E., Dubost, J.-P., Dallet, P., Langlois, M.H. and Colleter, J.-C., Eur. J. Med. Chem.-Chim. Thér., 24 (1989) 155.
30 Goodford, P.J., J. Med. Chem., 28 (1985) 849.
31 Kantola, A., Villar, H.O. and Loew, G.H., J. Comput. Chem., 12 (1991) 681.
32 Bodor, N., Gabanyi, Z. and Wong, C.K., J. Am. Chem. Soc., 111 (1989) 3783.
33 Politzer, P. and Truhlar, D.G. (Eds.) Chemical Applications of Atomic and Molecular Electrostatic Potentials, Plenum, New York, NY, U.S.A., 1981.
34 Politzer, P. and Murray, J.S., In Lipkowitz, K.B. and Boyd, D.B. (Eds.) Reviews in Computational Chemistry, VCH, New York, NY, U.S.A., 1991.
35 Rozas, I., Du, Q. and Arteca, G.A., J. Mol. Graph., 13 (1995) 98.
36 Murray, J.S., Brinck, T., Lane, P., Paulsen, K. and Politzer, P., J. Mol. Struct. (THEOCHEM), 307 (1993) 55.
37 Politzer, P., Lane, P., Murray, J.S. and Brinck, T., J. Phys. Chem., 96 (1992) 7938.
38 Murray, J.S., Brinck, T. and Politzer, P., J. Phys. Chem., 97 (1993) 13807.
39 Brinck, T., Murray, J.S. and Politzer, P., J. Org. Chem., 58 (1993) 7070.
40 Du, Q. and Arteca, G.A., J. Comput.-Aided Mol. Design, 10 (1996) 133.
41 Israelachvili, J.N., Intermolecular and Surface Forces, Academic Press, London, U.K., 1992.
42 Head-Gordon, T., J. Am. Chem. Soc., 117 (1995) 501.
43 Dogonadze, R.R., Kornyshev, A.A. and Ulstrup, J., In Dogonadze, R.R., Kálmán, E., Kornyshev, A.A. and Ulstrup, J. (Eds.) The Chemical Physics of Solvation, Part A: Theory of Solvation, Plenum, New York, NY, U.S.A., 1985.
44 Frank, H. and Evans, M., J. Chem. Phys., 13 (1945) 507.
45 Murray, J.S., Ranganathan, S. and Politzer, P., J. Org. Chem., 56 (1991) 3734.
46 Gao, J., In Smith, D.A. (Ed.) Modeling the Hydrogen Bond, American Chemical Society, Washington, DC, U.S.A., 1994.
47 Mishra, P.C. and Kumar, A., In Sen, K. (Ed.), Topics in Current Chemistry (174), Molecular similarity II, Springer, New York, NY, U.S.A., 1995.
48 Mezey, P.G., In Lipkowitz, K.B. and Boyd, D.B. (Eds.) Reviews of Computational Chemistry, Vol. 1, VCH, New York, NY, U.S.A., 1990.
49 Mezey, P.G., Int. J. Quantum Chem., Quantum Biol. Symp., 12 (1986) 113.
50 Mezey, P.G., Shape in Chemistry: An Introduction to Molecular Shape and Topology, VCH, New York, NY, U.S.A., 1993.
51 Walker, P.D. and Mezey, P.G., J. Am. Chem. Soc., 115 (1993) 12423.
52 Walker, P.D. and Mezey, P.G., J. Am. Chem. Soc., 116 (1994) 12022.
53 Mezey, P.G., Struct. Chem., 6 (1995) 261.
54 Mezey, P.G., J. Math. Chem., 18 (1995) 141.
55 Du, Q. and Arteca, G.A., J. Comput. Chem., 17 (1996) 1258.
56 Connolly, M.L., Science, 221 (1983) 709.
57 Connolly, M.L., J. Appl. Crystallogr., 16 (1983) 548.
58 Connolly, M.L., J. Am. Chem. Soc., 107 (1985) 1118.
59 Ooi, T., Oobatake, M., Némethy, G. and Scheraga, H.A., Proc. Natl. Acad. Sci. USA, 84 (1987) 3086.
60 Cabani, S. and Gianni, P., J. Chem. Soc. Faraday Trans. I, 75 (1979) 1184.
61 Cabani, S., Gianni, P., Mollica, V. and Lepori, L., J. Solution Chem., 10 (1981) 563.
62 Klebe, G. and Abraham, U., J. Med. Chem., 36 (1993) 70.
63 Lee, S.H. and Rossky, P.J., J. Chem. Phys., 100 (1994) 3334.
64 Piculell, L., J. Chem. Soc. Faraday Trans., 82 (1986) 387.
65 Piculell, L. and Halle, B., J. Chem. Soc. Faraday Trans., 82 (1986) 401.
66 Halle, B. and Piculell, L., J. Chem. Soc. Faraday Trans., 82 (1986) 415.
67 Borman, S., C. & EN, August (1995) 29.
68 Dean, P.M., Molecular Foundations of Drug–Receptor Interactions, Cambridge University Press, Cambridge, U.K., 1987.
69 Frank, R., Theoretical Drug Design Methods, Akademie Verlag, Berlin, Germany, 1984.
70 Geladi, P. and Kowalski, B., Anal. Chim. Acta, 185 (1986) 1.
71 Geladi, P. and Kowalski, B., Anal. Chim. Acta, 185 (1986) 19.
72 Cramer III, R.D., Patterson, D.E. and Bunce, J.D., J. Am. Chem. Soc., 110 (1988) 5959.
73 Marshall, G.R. and Cramer III, R.D., Trends Pharmacol. Sci., 9 (1988) 285.
74 Cramer III, R.D., Patterson, D.E. and Bunce, J.D., J. Am. Chem. Soc., 110 (1988) 5959.