Theoretical Derivation of Heuristic Molecular Lipophilicity Potential: A Quantum Chemical Description for Molecular Solvation

Qishi Du,*[,†,‡] Peng-Jun Liu,[‡] and Paul G. Mezey[§]

Tianjin Normal University, Institute of Bioinformatics and Drug Discovery (IBDD), Department of Chemistry, 241 Weijin Road, Hexi District, Tianjin, 300074 China, Hainan Normal University, Department of Chemistry, Haikou, Hainan, 571158 China, and Memorial University of Newfoundland, Scientific Modeling and Simulation Laboratory (SMSL), Department of Chemistry, St. John's, Newfoundland, Canada A1B 3X7

Received September 28, 2004

We present the theoretical derivation of a heuristic molecular lipophilicity potential (HMLP), which gives a structure-based and quantum chemical description of an important aspect of molecular solvation. The quantum mechanical electrostatic potential (ESP) $V(\mathbf{r})$ on a formal molecular surface is calculated, and then the molecular lipophilicity potential $L(\mathbf{r})$ is constructed by comparing the local electron density with the ESP on the surrounding atoms using a screening function. The screening function is derived from statistical mechanical theory treating the polar solvent molecules as dipoles. HMLP is able to describe the main interactions of solute molecules with polar and nonpolar solvent molecules. HMLP is a unified lipophilicity and hydrophilicity potential: its positive values represent lipophilicity, and its negative values represent hydrophilicity. In this paper, several examples show that HMLP gives more reliable descriptions for the molecular solvation than some other methods, such as atomic partial charges and the empirical lipophilicity potential.

INTRODUCTION

All chemical and physical properties of molecules and chemical reactions in solutions are related to solvation, specifically, in aqueous solution, to hydration.[1−5] Organic molecules in aqueous solutions experience either hydrophobic hydration or hydrophilic hydration, and if one considers local regions, often partially hydrophobic hydration and partially hydrophilic hydration.[3−7] Molecular lipophilicity and hydrophilicity are among the most frequently used concepts in chemistry, biology, medicine, and many other research fields in the life sciences.[1,2] Although the concepts of lipophilicity and hydrophilicity are familiar to most chemists, nevertheless, in chemistry there are no quantum chemically sound, precise theoretical definitions and no rigorous, theoretical calculation methods for these two properties.

Molecular lipophilicity is one of the controlling factors in the relationship between structures and functions of proteins and functional peptides,[8−10] including protein folding, signal peptides, water channels, and ion channels. The signal peptides are called the "zip codes" of proteins[8] because they control the entrances of proteins and lead the proteins to specific subcellular locations in the cell. The secret that signal peptides can pass through membranes leading proteins into a selected organelle lies in the specific composition of lipophilic and hydrophilic amino acid residues.[9,10] The reason that the proteins fold in various patterns in aqueous solution arises in part directly from the various degrees of lipophilicity of amino acid residues in the sequences.[11,12] The discoveries of water channels[13−19] and ion channels[20] in cell membranes have increased interest in the study of lipophilicity and hydrophilicity of proteins. The lipophilic and hydrophilic parameters[21,22] of amino acid residues are widely used in various molecular modeling techniques for proteins, peptides, and other biological systems.

Molecular electrostatic potential is a powerful tool in the study of molecular interactions and has been used in the studies of a number of molecular interactions by many authors.[23−25] However, due to the complexity and difficulty of molecular lipophilcity, electrostatic potential has not been used successfully in the explanation for molecular lipophilicity. The heuristic molecular lipophilicity potential (HMLP)[26−29] has been suggested for the description of molecular lipophilicity using an empirical screening function. It is a step toward the solution of the more general problem. In this study, we present a theoretical derivation for HMLP based on quantum mechanics and statistical mechanics, revealing its physical meaning and demonstrating its usefulness by several examples.

MOLECULAR LIPOPHILIC AND HYDROPHILIC SURFACES

One of the central concepts used in our model is that of a molecular surface since many of the local properties we calculate refer to a segment or the whole of a formal molecular surface. Whereas the molecular surface concept is a useful one, it is clearly a somewhat artificial concept, since molecular electron densities are fuzzy entities, with no actual boundaries, and in a rigorous sense no finite closed surface exists that would contain the entire molecule with all of its electron density cloud.[30−33] Nevertheless, in an approximate sense molecular surfaces can be defined, for example, using molecular isodensity contour (MIDCO) surfaces for some low density value, or by the more conventional van der Waals surfaces, commonly used in biochemical applications. Some of the relevant fundamental problems of approximate molecular surfaces have been discussed.[30−33] In the present

* Corresponding author tel: +086-22-2354-0997; fax: +086-22-2354-0187; e-mail: duqishi@yahoo.com.
† Tianjin Normal University.
‡ Hainan Normal University.
§ Memorial University of Newfoundland.

10.1021/ci049707l CCC: $30.25
Published on Web 02/22/2005

Table 1. Atomic Charges and Surface ESP Descriptors of Pentanoic Acid, Calculated Using Gaussian 98[34]

	HF/6-31G(d)		AM1		DFT/6-31G(d)		ESP
	q_i^{Mull}	q_i^{ESP}	q_i^{Mull}	q_i^{ESP}	q_i^{Mull}	q_i^{ESP}	b_i[b]
C_1(COOH)	0.755	0.886	0.363	0.190	0.564	0.737	−0.154
O(CO)	−0.558	−0.635	0.381	−0.256	−0.453	−0.548	−2.321
O(OH)	−0.703	−0.704	−0.370	−0.339	−0.561	−0.613	−0.915
H(OH)	0.466	0.462	−0.278	0.286	0.407	0.425	1.668
C_2	−0.423	−0.271	−0.287	−0.503	−0.358	−0.233	0.861
H	0.203	0.091	0.170	0.241	0.173	0.080	0.457
H	0.203	0.091	0.170	0.241	0.173	0.080	0.457
C_3	−0.303	−0.008	−0.255	−0.316	−0.243	−0.012	0.650
H	0.178	0.027	0.150	0.212	0.149	0.027	0.186
H	0.178	0.027	0.150	0.212	0.149	0.027	0.186
C_4	−0.309	0.169	−0.262	−0.412	−0.255	0.147	0.622
H	0.155	−0.021	0.128	0.250	0.132	−0.017	0.315
H	0.155	−0.021	0.128	0.250	0.132	−0.017	0.316
C_5	−0.479	−0.327	−0.348	−1.177	−0.445	−0.304	0.447
H	0.161	0.077	0.122	0.367	0.145	0.073	0.245
H	0.161	0.077	0.122	0.367	0.145	0.073	0.233
H	0.163	0.080	0.122	0.385	0.144	0.075	0.233

[a] Results are in atomic units. [b] Atomic ESP descriptors b_i on van der Waals surface are calculated using Gaussian 98,[34] at the HF/6-31G(d) level. Atomic radii are optimized in our earlier work.[29] The surface point density is 25 points/Å^2.

study, we shall assume that such approximate surfaces are appropriate tools for our purposes.

To elucidate the background, we start with a specific example to illustrate the difference between lipophilic surfaces and hydrophilic surfaces. Table 1 lists the atomic charges of pentanoic acid calculated using ab initio Hartree−Fock method, DFT, and the semiempirical AM1 method. Each method gives the atomic charges from Mulliken population analysis and from the ESP (electrostatic potential) fitting using Gaussian 98.[34] In the HF and DFT/B3LYP calculations we use the 6-31G(d) basis set. Table 1 shows that hydrophilic carboxyl COOH is a polar atomic group, in which C and H are positively charged and two oxygen atoms O(CO) and O(OH) are negatively charged. However, in the lipophilic hydrocarbon chain C_4H_9, the carbon atoms have negative charges and hydrogen atoms have positive charges. This illustrates that the lipophilic hydrocarbon chain C_4H_9 is also a polar atomic group. Therefore, atomic partial charges are not suitable to provide fundamentally different descriptions for the polar and nonpolar atomic groups, and such charges by themselves cannot simply distinguish between these two types.

In Table 1, the last column lists the atomic ESP descriptors[26−29] b_i, defined as

$$b_i = \int_{S_i} V(\mathbf{r})\, \mathrm{d}s \approx \sum_{k \in S_i} V(\mathbf{r}_k) \Delta s_k \qquad (1)$$

where Δs_k is the area element on the exposed surface of atom i, $V(\mathbf{r}_k)$ is electrostatic potential in the center $\mathbf{r}_k$ of area element Δs_k. Summation is over all the exposed surface area S_i of atom i, where the atomic surface concept is interpreted in terms of a semiclassical molecular surface model.[30−33] The electrostatic potential $V(\mathbf{r}_k)$ is calculated using ab initio quantum chemistry:

$$V(\mathrm{r}) = \sum_{\alpha} \frac{Z_\alpha}{|R - r|} - \int_{\infty} \frac{\rho(\mathrm{r}')}{|r' - r|} \mathrm{d}r' \qquad (2)$$

where Z_α is the atomic charge on nucleus α located at R_α, and $\rho(\mathbf{r}')$ is the electron density function. This atomic ESP descriptor gives a better description for electrostatic interactions than the atomic maximum and minimum[23,24] ESP values, because it considers two important factors: atomic exposed surface area and ESP distribution on the formal atomic surface. The electrostatic interaction energy $E^{elec}(\mathbf{r})$ at point $\mathbf{r}$ on the molecular surface between the solute and solvent charges may be evaluated as follows:

$$E^{elec}(\mathbf{r}) = V^{(slu)}(\mathbf{r}) q^{(slv)} \qquad (3)$$

where the superscripts slu and slv refer to solute and solvent, respectively. Unlike classical Coulomb law, eq 3 is a quasi-quantum mechanical equation, because the electrostatic potential $V(\mathbf{r})$ contains the information of electron density $\rho(\mathbf{r})$ through eq 2. Therefore, atomic ESP descriptors are more reliable than atomic partial charges.

The calculation results in the last column of Table 1 show that on the surface of lipophilic hydrocarbon chain C_4H_9, ESP descriptors of all carbon and hydrogen atoms have the same positive sign. This means that the electrostatic potentials on lipophilic surfaces are uniformly positive. For the nonpolar molecular surfaces atomic ESP descriptors give a different description from the description of atomic partial charges. On the other hand, Table 1 shows that the atomic ESP descriptors of the polar and hydrophilic group COOH have positive as well as negative signs. The same conclusion is supported by our calculations on several examples[26−29] and by the calculations of others.[23−25]

HEURISTIC MOLECULAR LIPOPHILICITY POTENTIAL (HMLP)

In Table 1 all lipophilic atoms in the hydrocarbon chain C_4H_9 have positive ESP descriptors. However, we cannot say that the necessary condition for a lipophilic atom is the positive ESP descriptor because the polar atom H(OH) has positive ESP descriptor and it is a hydrophilic atom. In some cases, for example, in the lipophilic fluorocarbon chains, ESPs of all carbon and fluorine atoms have the same negative sign. In this context, "nonpolar surface" means that the ESPs on the molecular surface are uniform in sign. On the other hand, the atomic ESP descriptors on a "polar surface" have different signs.

When we consider the interaction between an atom of a solute molecule and a polar solvent molecule, we have to consider the manner of solute−solvent interaction of surrounding atoms and the interactions between solvent molecules. Based on the above analysis, we have suggested the following heuristic molecular lipophilicity potential (HMLP) in our previous works:[26−29]

$$\underset{\mathrm{r}\in S_\alpha}{L(\mathbf{r})} = V(\mathbf{r})\sum_{i\neq\alpha} M_i(\mathbf{r};\mathbf{R}_i, b_i) \tag{4}$$

where $L(\mathbf{r})$ is HMLP at point $\mathbf{r}$ on the surface S_α of atom α, $M_i(\mathbf{r};\mathbf{R}_i, b_i)$ is the screening function of atom i. In eq 4, the 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 ESP descriptor of atom i defined by eq 1. In our previous works we used an empirical distance-dependent screening function:

$$M_i(\mathbf{r};\mathbf{R}_i, b_i) = \frac{r_0^\gamma}{b_0}\frac{b_i}{||\mathbf{R}_i - \mathbf{r}||^\gamma} = \varsigma\frac{b_i}{||\mathbf{R}_i - \mathbf{r}||^\gamma} \tag{5}$$

In eq 5, exponent γ is the parameter that describes the influence from the surrounding atoms and how rapidly the influence decays with the distance $||\mathbf{R}_i - \mathbf{r}||$. Parameter b_0 has the same unit as b_i, and r_0 has the unit of length. Parameter $\zeta = (r_0)^\gamma/b_0$ is the ratio of the γth power of the atomic radius r_0 and ESP descriptor b_0 of reference atom "0". Therefore, the screening function $M_i(\mathbf{r};\mathbf{R}_i, b_i)$ is dimensionless, and the unit of HMLP $L(\mathbf{r})$ is the same as that of ESP $V(\mathbf{r})$. If the screening function in eq 4 is substituted by eq 5, then HMLP $L(\mathbf{r})$ takes the form:

$$\underset{\mathrm{r}\in S_\alpha}{L(\mathbf{r})} = V(\mathbf{r})\sum_{i\neq\alpha}^{M}\varsigma\frac{b_i}{||R_i - r||^\gamma} \tag{6}$$

In our previous work,[26] the two parameters took the values $\zeta = 1$ and $\gamma = 2.5$. On the basis of HMLP, we construct atomic lipophilicity index as follows:

$$l_a = \sum_{k\in S_\alpha} L(r_k)\Delta_{S_k} = \sum_{k\in S_\alpha} V(\mathbf{r}_k)\Delta_{S_k}\sum_{i\neq a}^{\text{atoms}}\varsigma\frac{b_i}{||R_i - r_k||^\gamma} \tag{7}$$

where the summation is over all the exposed surface area S_a of atom a. If $l_a > 0$, then atom a has the same polarity as its environment and atom a is a lipophilic atom, whereas, if $l_a < 0$, atom a has polarity opposite to its environment and atom a is a hydrophilic atom. In the same way, we can define molecular lipophilicity index L_M as the sum of lipophilicity indices of all lipophilic atoms ($l_a > 0$):

$$L_M = \sum_{\substack{a\\(l_a>0)}} l_a \tag{8}$$

and molecular hydrophilic index (H_M) as the sum of lipophilicity indices of all hydrophilic atoms ($l_a < 0$) in a molecule:

$$H_M = \sum_{\substack{a\\(l_a>0)}} l_a \tag{9}$$

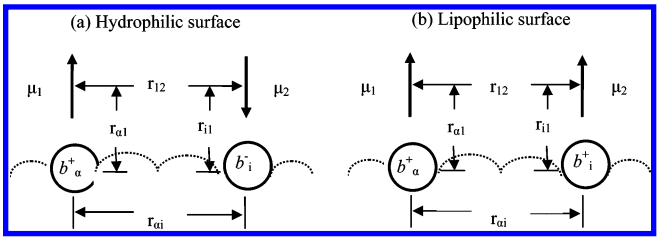

Figure 1. Charge−dipole and dipole−dipole electrostatic interactions. (a) On the hydrophilic surface, the atomic ESP descriptors b_α and b_i have the opposite signs. (b) On the lipophilic surface, the atomic ESP descriptors b_α and b_i have identical signs.

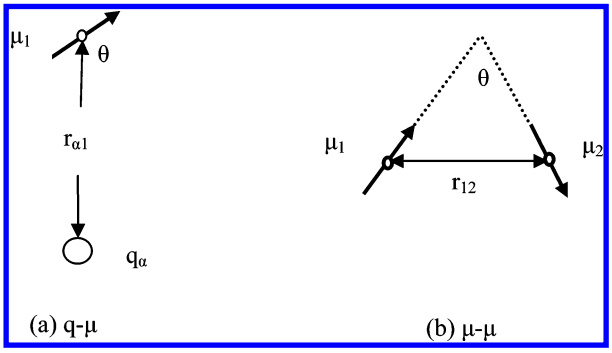

Figure 2. (a) Electrostatic interactions between charge and dipole. (b) Electrostatic interaction between two dipoles.

Similarly, we can define lipophilic and hydrophilic indices, l_s and h_s, for the substitutents (or fragments) in a set of molecular derivatives.

THEORETICAL DERIVATION OF HMLP

Figure 1 illustrates the orientation of polar solvent molecules on lipophilic and hydrophilic surfaces. In Figure 1a, atomic ESP descriptors b_α and b_i have opposite signs on the hydrophilic surface. The dipoles of two polar solvent molecules attracted by atoms α and i of the solute molecule are parallel in opposite directions. Figure 1b shows the molecular lipophilic surface where the atomic ESP descriptors b_α and b_i have the same signs. The dipoles of the two polar solvent molecules attracted by atoms α and i of the solute molecule are parallel in the same direction.

Figure 2a shows the electrostatic interaction between a point charge and a dipole, and Figure 2b shows the electrostatic interaction between two dipoles. The electrostatic interaction between point charge q_α and dipole μ_1 is evaluated by

$$u_{\alpha-\mu} = \frac{\mu_1 q_\alpha}{\epsilon r_{\alpha1}^2}\cos\theta = V_\alpha(r_{\alpha1})\frac{\mu_1}{\epsilon r_{\alpha1}}\cos\theta \tag{10}$$

where θ is the angle formed by dipole μ_1 and the distance $r_{\alpha1}$ from point charge q_α to the center of dipole μ_1. In eq 10, $V_\alpha(r_{\alpha1}) = q_\alpha/r_{\alpha1}$ is the electrostatic potential generated by charge q_α in the center of dipole μ_1. When $\theta = 180°$, then the electrostatic interaction energy $u_{\alpha-\mu}$ takes the maximum value:

$$u_{\alpha-\mu}^0 = \frac{\mu_1 q_\alpha}{\epsilon r_{\alpha1}^2} = V_\alpha(r_{\alpha1})\frac{\mu_1}{\epsilon r_{\alpha1}} \tag{11}$$

When $u_{\alpha-\mu}^0 \ll kT$, we have to take the statistical average of

$\mu_1 \cos\theta$ over all possible directions:

$$\bar{\mu}_1 = \frac{\int_0^{\pi}(\mu_1 \cos\theta) e\mu_1 V_\alpha(r_{\alpha 1}) \cos\theta/\epsilon r_{\alpha 1} kT \sin\theta \, d\theta}{\int_0^{\pi} e\mu_1 V_\alpha(r_{\alpha 1}) \cos\theta/\epsilon r_{\alpha 1} kT \sin\theta \, d\theta} \cong \frac{\mu_1^2}{3kT\epsilon r_{\alpha 1}} V_\alpha(r_{\alpha 1}) \quad (12)$$

If in eq 11 μ_1 is replaced by $\bar{\mu}_1$ in eq 12, then we obtain the average charge−dipole electrostatic interaction energy:

$$u_{\alpha-\mu} = \frac{V_\alpha^2(r_{\alpha 1})\mu_1^2}{3kT\epsilon^2 r_{\alpha 1}^2} \quad (13)$$

When the length of dipole is much smaller than the distance between two dipoles $l \ll r_{12}$, the dipole−dipole electrostatic interaction in Figure 2b is evaluated by the following equation:

$$u_{\mu-\mu} = \frac{1}{\epsilon r_{12}^3}\left[\vec{\mu}_1 \cdot \vec{\mu}_2 - \frac{3(\mu_1 \cdot r_{12})(\vec{\mu}_2 \cdot \vec{r}_{12})}{r_{12}^2}\right] \quad (14)$$

When the two dipoles are parallel, then the electrostatic interaction takes the maximum value, and eq 14 is simplified to

$$u_{\mu-\mu} = \pm\frac{\mu_1\mu_2}{\epsilon r_{12}^3} \quad (15)$$

In eq 15, the sign ± means that when dipoles μ_1 and μ_2 point in the same direction, then the interaction energy $u_{\mu-\mu}$ is positive, otherwise, the interaction energy $u_{\mu-\mu}$ is negative. If the dipoles μ_1 and μ_2 in eq 15 are replaced by $\vec{\mu}_1$ and $\vec{\mu}_2$ defined by eq 12, then we get the average dipole−dipole electrostatic interaction energy:

$$u_{\mu-\mu} = \left[\frac{\mu_1^2}{3kT\epsilon r_{\alpha 1}} V_\alpha(r_{\alpha 1})\right]\left[\frac{\mu_2^2}{3kT\epsilon r_{i2}} \frac{V_i(r_{i2})}{\epsilon r_{12}^3}\right] \quad (16)$$

In eq 16, there is no ± sign, and the sign of the interaction energy $u_{\mu-\mu}$ is decided by the signs of $V_\alpha(r_{\alpha 1})$ and $V_i(r_{i2})$. If $V_\alpha(r_{\alpha 1})$ and $V_i(r_{i2})$ have the same sign, then the interaction energy $u_{\mu-\mu}$ is positive, otherwise, it is negative. In eq 16, $r_{\alpha 1} = r_{i2} \approx d_o$ is the thickness d_o of the first solvation layer, and $r_{12} \approx r_{\alpha i}$ is the distance between atom α and atom *i*. If the two dipoles μ_1 and μ_2 are the dipoles of the polar solvent molecules, then they are equal and replaced by μ. Equation 16 describes the electrostatic interaction between two polar solvent molecules attracted by atom α and *i* of the solute molecule. If we consider the electrostatic interaction between dipole μ_1 (attracted by atom α) and all other dipoles mi attracted by surrounding the atoms, we get

$$u_{\mu-\mu}(r) = \frac{\mu^4}{(3kT)^2 d_o^2 \epsilon^3} V_\alpha(r)\left[\sum_{i\neq\alpha}^{M} \frac{V_i(r_i)}{r_{\alpha i}^3}\right] \quad (17)$$

In eq 17, $V_\alpha(\mathbf{r})$ is the electrostatic potential generated by atom α at point **r** on the surface of atom α, $V_i(\mathbf{r}_i)$ is the electrostatic potential of atom *i* at point $\mathbf{r}_i$ on the surface of atom *i*, and *M* is the number of atoms in the solute molecule.

Comparing the HMLP definition, eq 6 to eq 17, the electrostatic potential $V_i(\mathbf{r}_i)$ of atom *i* in eq 17 is replaced by atomic ESP descriptor b_i of eq 6. Equation 17 reveals the physical meaning of heuristic molecular lipophilicity potential: HMLP $L(\mathbf{r})$ is the electrostatic interaction energy $u_{\mu-\mu}(\mathbf{r})$ produced by a solvent dipole μ_1 attracted by atom α with all surrounding solvent dipoles on the molecular surface. If the electrostatic potential $V_\alpha(\mathbf{r})$ on the surface of atom α and $V_i(\mathbf{r}_i)$s on the surfaces of surrounding atoms have opposite signs, as shown in Figure 1a, then the interaction energy is negative and **r** is a hydrophilic point, otherwise, if $V_\alpha(\mathbf{r})$ on the surface of atom α and $V_i(\mathbf{r}_i)$s of the surrounding atoms have the same signs, as shown in Figure 1b, then the interaction energy is positive and **r** is a lipophilic point. Comparing eq 17 with the HMLP definition (eq 6), we find the values of two parameters ζ and γ in HMLP. Parameter ζ is decided by temperature *T* and the dipole μ of solvent molecules:

$$\varsigma = \frac{\mu^4}{(3kT)^2 d_o^2 \epsilon^3} \quad (18)$$

Therefore, HMLP is directly proportional to the fourth power of the dipole μ of the solvent and inversely proportional to the square of temperature *T*. Ideally, the value of parameter γ should be 3. However, in the derivation of eq 17 the water molecule is treated as a dipole, and eq 17 holds only subject to the condition that the length of the dipole is much smaller than the distance between the two dipoles, $l \ll r_{12}$. It is better to treat the parameter ζ and γ as empirical parameters and leave their values to be decided by comparisons with experiments. HMLP $L(\mathbf{r})$ defined by eq 17 does not include all solvation terms; however, it is the most discriminant term that describes the difference between lipophilic and hydrophilic interactions.

Actually, HMLP $L(\mathbf{r})$ is the local description of lipophilicity and hydrophilicity on the molecular surface. HMLP is linked to a traditional experimental value, the partition coefficient log $P_{o/w}$ in 1-octanol and in water:[35]

$$\log P_{o/w} = \log\frac{C_o}{C_w} = -\frac{\Delta G_{tr}}{2.303RT} = \sum_i n_i f_i \quad (19)$$

In eq 19, f_i is the empirical lipophilicity parameter of atom *i*. If we add a hydrophilic atom *i* to the solute molecule, then the concentration of solute in the water phase increases and it has a negative contribution ($f_i < 0$) to the partition coefficient log $P_{o/w}$. By contrast, the addition of a lipophilic atom *i* has a positive contribution ($f_i > 0$) to the partition coefficient log $P_{o/w}$.

On a lipophilic surface where the electrostatic potential is uniformly distributed, as shown in Figure 1b, the parallel orientation of polar solvent molecules (such as water) with the same direction is very unfavorable for energy because of the strong repulsive interactions between water molecules. Actually, the polar solvent molecules leave away the lipophilic surface and form an "ice shell" in a highly ordered manner surrounding the lipophilic surface and producing a strong negative entropic effect. In this case the driving force

acting on the polar solvent molecules to take a lipophilic arrangement surrounding the lipophilic surface is the repulsive interaction energy described by eq 17.

CALCULATION RESULTS

Propionic Acid. Figure 3 shows the $V(\mathbf{r})$ versus $L(\mathbf{r})$ maps on the molecular surface of propionic acid, C_2H_5COOH. The molecular van der Waals surface is generated using Connolly's program MS,[36,37] at a point density is 25/Å^2; there are a total of 2283 points. The electrostatic potential is calculated using Gaussian 98[34] at the HF/6-31G(d) level. In Figure 3, the *x*-axis corresponds to ESP, and the *y*-axis represents HMLP. The upper side of the map is the lipophilic region, and the lower side is the hydrophilic region. The left-hand side of the map corresponds to the negative ESP region, and the right-hand side corresponds to the positive ESP region. All polar atoms O(CO), O(OH), and H(OH) in the hydrophilic group COOH are found in the hydrophilic region (lower side). The two negatively charged oxygen atoms O(CO) and O(OH) belong to the left-hand side, and the positively charged atom H(OH) belongs to the right-hand side of the map. In the lipophilic group C_2H_5, the nonpolar carbon atoms and the hydrogen atoms are found in the lipophilic region (upper side). The carbon atom in COOH is the link between lipophilic and hydrophilic atoms. The carboxyl carbon atom C(COOH) is in the center of the $V(\mathbf{r})-L(\mathbf{r})$ map, and it shows some transitional character. Half of its surface area is in the negative ESP region, and the other half of its surface area is in the positive ESP region. Correspondingly, some part of its surface is lipophilic and the other part is hydrophilic. All atoms directly connecting with C(COOH) have some points joining at the center of the $V(\mathbf{r})$-$L(\mathbf{r})$ map. The nonpolar atoms H and C in the C_2H_5 group are lipophilic and bear positive ESP. They are found in the upper right corner of the map.

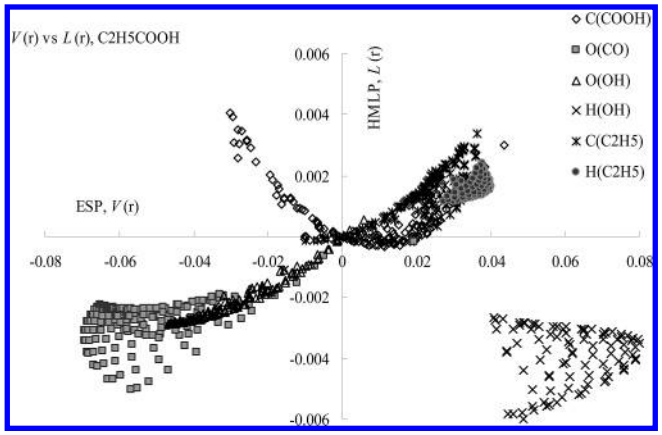

Figure 3. ESP and HMLP distribution map of propionic acid C_2H_5-COOH on the van der Waals surface, $V(\mathbf{r})-L(\mathbf{r})$.

Alcohols. Table 2 lists the atomic charges, ESP descriptors, and atomic lipophilicity indices of ethanol. In the nonpolar group C_2H_5, atoms C and H have negative and positive charges q_i^{ESP}, respectively. Atomic partial charge cannot describe lipophilic groups correctly. On the other hand, all H and C in the lipophilic group C_2H_5 have positive ESP descriptors b_i and positive atomic lipophilicity indices l_a. The two polar atoms O and H in the hydroxyl group OH have negative lipophilicity indices, and they are hydrophilic

Table 2. Atomic Charges, ESP Descriptors, and Lipophilicity Indices of Ethanol

	atomic charge	ESP descriptor	lipophilicity index	
atom	q_i^{ESP}	b_i	f_i (empirical)[a]	l_a (HMLP)
(−OH) O	−0.7092	−2.111	−0.0517	−0.1799
H	0.4146	1.359	−0.3703	−0.0783
(−CH_2−) C_1	0.4433	0.8559	−0.9481	0.0596
H	−0.0528	0.3253	0.3722	0.0085
H	−0.0528	0.3284	0.3722	0.0085
(−CH_3) C_2	−0.2405	0.0820	−0.6327	0.0219
H	0.0363	0.2829	0.4307	0.0046
H	0.0807	0.1362	0.4307	0.0012
H	0.0807	0.1379	0.4307	0.0012
L_M	----	----	2.0365	0.1055
H_M	----	----	−2.0028	−0.2582

[a] Refs 38−40.

atoms. For ethanol, the total molecular hydrophilicity index H_M (−0.2582) is 2.5 times the value of the lipophilicity index L_M (0.1055). This means that ethanol is a strong hydrophilic compound. However, the value (−2.0028) of empirical H_M and the value (2.0365) of empirical L_M are almost equal in magnitude. This is unreasonable based on chemical knowledge. The empirical atomic lipophilicity index[16] of C2 in the lipophilic group C_2H_5 has a negative value of −0.6327. This is apparently not correct.

Table 3 lists the molecular lipophilicity indices L_M and hydrophilicity indices H_M of eight alcohols. The hydrophilicity indices H_M are nearly constant (~ -0.241); however, the lipophilicity indices L_M increase with the increasing numbers of carbon atoms. This example shows that HMLP indices are approximately transferable and additive in a family of derivatives.

If we suppose that there exists a linear correlation relationship between HMLP indices and water−octanol partition coefficients, we may write

$$\log P_{o/w} = \frac{-\Delta G^\circ_{tr}}{2.3026RT} = \frac{1}{2.3026RT}(aL_M + bH_M) + c \qquad (20)$$

When $R = 8.3144$ kJ/mol and $T = 298.3$ K, we get a good linear correlation equation, with $a = 1.269 \times 10^5$, $b = -5.637 \times 10^5$, and $c = -27.01$, correlation coefficient $r = 0.997$, and a standard deviation $\sigma = 0.130$. According to chemical thermodynamics, the transferring free energy consists of transferring enthalpy and entropy, $\Delta G^\circ_{tr} = \Delta H^\circ_{tr} - T\Delta S^\circ_{tr}$. The HMLP hydrophilicity index H_i makes the main contribution to the enthalpy part ΔH°_{tr} and the lipophilicity index L_i makes the main contribution to the entropy part ΔS°_{tr}.

Polymethylene Oxide and Poly(ethylene Oxide). In polymethylene oxide $[-CH_2-O-]_n$ there are more "hydrophilic" oxygen atoms than in poly(ethylene oxide) $[-CH_2-CH_2-O-]_n$, and in poly(ethylene oxide) there are more "lipophilic" segments CH_2 than in polymethylene oxide. However, the former is lipophilic and the latter is hydrophilic.[1] This phenomenon is quite strange from the usual chemical point of view, and it might appear difficult to get a reasonable explanation. Our model seems to provide one.

Figure 4 shows two models, a short segment of polymethylene oxide, $CH_3-O-CH_2-O-CH_2-O-CH_3$, and that

Table 3. Lipophilicity Parameter L_M, Hydrophilicity Parameter H_M, and Experimental[a] Log $P_{o/w}$ Values of Alcohols

	CH_3OH	C_2H_5OH	C_3H_7OH	C_4H_9OH	$C_5H_{11}OH$	$C_6H_{13}OH$	C_7H15_5OH	$C_8H_{17}OH$
L_M	0.136	0.119	0.158	0.179	0.211	0.234	0.258	0.279
H_M	−0.235	−0.243	−0.241	−0.242	−0.242	−0.241	−0.241	−0.241
log P_{cal}	−0.798	−0.387	0.281	0.876	1.562	2.080	2.504	2.954
log P_{exp}	−0.817	−0.315	0.342	0.817	1.457	1.982	2.542	3.062

[a] Ref 41.

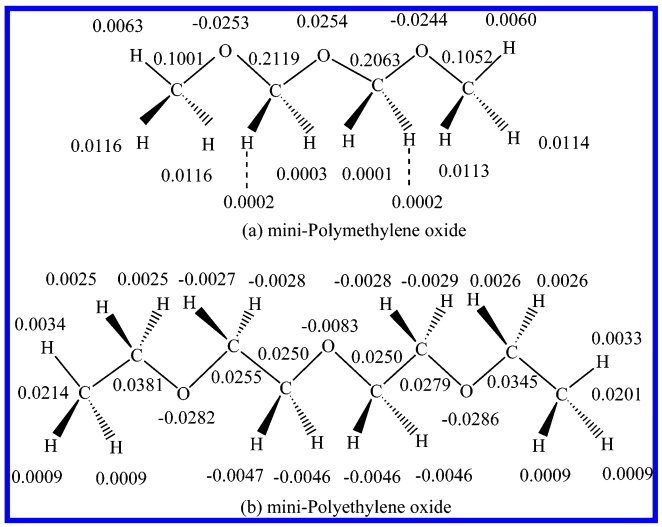

Figure 4. Structures and atomic HMLP lipophilicity indices of polymethylene oxide and poly(ethylene oxide).

of poly(ethylene oxide), $CH_3-CH_2-O-CH_2-CH_2-O-CH_2-CH_2-O-CH_2-CH_3$. The geometries of the two molecules are optimized using Gaussian 98[34] at the HF/STO-3G level. The HMLP indices are calculated at the HF/6-31G* level and shown on the atoms in Figure 4.

In the model polymethylene oxide molecule, the atomic lipophilicity indices of the two oxygen on the two ends are negative, the atomic lipophilicity indices of all carbon and hydrogen atoms are positive, and the atomic lipophilicity index of the oxygen atom in the central position is positive (0.0254). With an increase of the degree of polymerization, there will be more lipophilic oxygen atoms. Therefore, polymethylene is a lipophilic molecule. In the model poly(ethylene oxide) molecule, the atomic lipophilicity indices of all three oxygen atoms are negative. Except for two C_2H_5 groups on the two ends, the lipophilicity indices of all hydrogen atoms of the C_2H_5 groups in the central part of poly(ethylene oxide) are negative. Therefore, poly(ethylene oxide) is a hydrophilic molecule. The reason is that in polymethylene oxide the central oxygen atom has two oxygen atoms at the second neighbor positions, and all three oxygen atoms possess similar negative electrostatic potentials. On the other hand, in poly(ethylene oxide) the central oxygen atom has two oxygen atoms at the third neighbor position, and the oxygen atoms are separated by C_2H_5 groups that have electrostatic potential opposite to that of the oxygen atoms. Therefore, oxygen is not necessarily a hydrophilic atom, and C_2H_5 is not necessarily a lipophilic group. The atomic lipophilicity and hydrophilicity are decided not only by the local group but also by its environment.

DISCUSSION AND CONCLUSION

The theoretical derivation of HMLP presented in this work reveals new aspects of the physical meaning of the heuristic molecular lipophilicity potential, $L(\mathbf{r})$, and gives the theoretical values of two parameters ζ and γ in the screening function. Theoretically, HMLP holds for aqueous solution as well as for other solutions of polar solvents. Due to the complexity of molecular lipophilicity, eq 17 is derived under some approximate conditions; therefore, we still consider HMLP as a "heuristic" model and leave the two parameters ζ and γ to be decided by comparisons with experiments. In eq 17, the exponent γ is expected to be 3. A rough optimization for exponent γ based on the experimental data of partition coefficients log $P_{o/w}$ shows that in practice 2.5 is a proper value.[13] In the charge−dipole interaction of eq 13, the exponent γ of distance-dependent function is 2, and in the dipole−dipole interaction of eq 15, the exponent γ of distance-dependent function is 3. The fitted value of 2.5 of the exponent γ is between the above two values.

Molecular lipophilicity is a collective property between solute molecule and large number of solvent molecules. When we consider the interaction of an atom in the solute molecule with a solvent molecule, then we have to consider the manner of the interaction of the surrounding atoms of the solute with other solvent molecules and the interactions among the solvent molecules. The local lipophilicity and hydrophilicity of an atom in solute molecule is decided not only by itself but also by its environment. In HMLP the screening function serves in this role.

We think that the three examples given in this paper provide convincing evidence for the ability of HMLP to describe the molecular and fragmental lipophilicity and hydrophilicity of various molecules. Using optimized parameters ζ and γ, HMLP is expected to provide quantitative results for the description of molecular lipophilicity and hydrophilicity. The HMLP indices defined by eqs 7−9 may extend the application of HMLP to a wider range of problems. In a recent work, we have calculated the lipophilic and hydrophilic indices for 20 natural amino acid side chains, which may be useful in the studies of proteins and peptides. In principle, HMLP can be applied to aqueous solutions as well as to other solutions of polar solvents, and it may provide a new computational technique for more general solvation studies.

ACKNOWLEDGMENT

This work is supported by grants from Chinese National Science Foundation under Contract 20373048, by the Tianjin Commission of Sciences and Technology under Contract 023618211 and 033801911, and by the Canadian Foundation for Innovation as well as the Natural Sciences and Engineering Research Council of Canada.

REFERENCES AND NOTES

(1) Israelachvili, J. N. *Intermolecular and Surface Forces*; Academic Press: London, 1992.

(2) Ducker, W. A.; Xu, Z.; Israelachvili, J. N. Measurements of hydrophobic and DLVO forces in bubble-surface interactions in aqueous solutions. *Langmuir* **1994**, *10*, 3279−3289.
(3) Kuharski, R. A.; Rossky, P. J. Molecular dynamics study of solvation in a urea-water solution. *J. Am. Chem. Soc.* **1984**, *106*, 5786.
(4) Kuharski, R. A.; Rossky, P. J. Solvation of hydrophobic species in aqueous urea solution: a molecular dynamics study. *J. Am. Chem. Soc.* **1984**, *106*, 5794.
(5) Lee, C. Y.; McCammon, J. A.; Rossky, P. J. The structure of liquid water at an extended hydrophobic surface. *J. Chem. Phys.* **1984**, *80*, 4448.
(6) Bagchi, B. Water solvation dynamics in the bulk and in the hydration layer of proteins and self-assemblies. *Annu. Rep. Prog. Chem., Sect. C.* **2003**, *99*, 127.
(7) Srinivas, G.; Bagchi, B. Study of pair contact formation among hydrophobic residues in a model HP-36 protein: relationship between contact order parameter and rate of folding and collapse. *J. Phys. Chem. A* **2003**, *107*, 11768.
(8) Chou, K.-C. Prediction of protein signal sequences. *Curr. Protein Pept. Sci.* **2002**, *3*, 615−622.
(9) Chou, K.-C. Using subsite coupling to predict signal peptides. *Protein Eng.* **2001**, *14*, 75−79.
(10) Chou, K.-C.; Elord, D. W. Protein subcellular location prediction. *Protein Eng.* **1999**, *12*, 107−118.
(11) Srinivas, G.; Bagchi, B. Study of the dynamics of protein folding through minimalistic models. *Theor. Chem. Acc.* **2003**, *109*, 8.
(12) Srinivas, G.; Bagchi, B. Foldability and the funnel of HP-36 protein sequence: use of hydropathy scale in protein folding. *J. Chem. Phys.* **2002**, *116*, 8579−8587.
(13) Agre, P.; King, L. S.; Yasui, M.; Guggino, W. B.; Ottersen, O. P.; Fujiyoshi, Y.; Engel, A.; Nielsen, S. Water transport controversies−Aquaporin water channels: from atomic structure to clinical medicine, *J. Physiol.* **2002**, *542*, 3−16.
(14) Dutzler, R.; Campbell, E. B.; MacKinnon, R. Gating the selectivity filter in CIC chloride channels−an hourglass-shaped pore, gated by a glutamate carboxyl gro. *Science* **2003**, *300*, 108−111.
(15) Fu, D.; Libson, A.; Miercke, L. J. W.; Weitzman, C.; Nollert, P.; Krucinski, J.; Stroud, R. M. Structure of a glycerol-conducting channel and the basis for its selectivity. *Science* **2000**, *290*, 481−486.
(16) Murata, K.; Mitsuoka, K.; Hirai, T.; Walz, T.; Agre, P.; Heymann, J. B.; Engel, A.; Fujiyoshi, Y. Membrane proteins: nanomachines with optimized structure and function. *Nature* **2000**, *407*, 599−605.
(17) Ren, G.; Reddy, V. S.; Cheng, A.; Melnyk, P.; Mitra, A. K. Visualization of a water-selective pore by electron crystallography in vitreous ice. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 1398−1403.
(18) Sui, H.; Han, B. G.; Lee, J. K.; Walian, P.; Jap, B. K. Structural basis of water-specific transport through the AQP1 water channel. *Nature* **2001**, *414*, 872−878.
(19) de Groot, B. L.; Grubmüller, H. Water permeation across biological membranes: mechanism and dynamics of aquaporin-1 and GlpF. *Science* **2001**, *294*, 2353−2357.
(20) Tajkhorshid, E.; Nollert, P.; Jensen, M. O.; Miercke, L. J.; O'Connell, J.; Stroud, R. M.; Schulten, K. Control of the selectivity of the aquaporin water channel family by global orientational tuning. *Science* **2002**, *296*, 525−530.
(21) Kyte, J.; Doolittle, R. A simple method for displaying the hydrophobic character of protein. *J. Mol. Biol.* **1982**, *157*, 105−132.
(22) Radzicka, A.; Wolfenden, R. Hydrophilicity of side chains in kJ/mol. *Biochemistry* **1988**, *27*, 1664−1670.
(23) Politzer, P.; Truhlar, D. G. In *Chemical Applications of Atomic and Molecular Electrostatic Potential*; Politzer, P., Ed.: Plenum: New York, 1981.
(24) Lipkowitz, K. B.; Boyd, D. B. In *Reviews in Computational Chemistry*, Vol. 2; Politzer, P., Murray, J. S., Eds.: VCH: New York, 1991.
(25) Sen, K. In *Molecular Similarity II*; Topics in Current Chemistry 174; Mishra, P. C., Kumar, A., Eds.: Springer-Verlag: New York, 1995.
(26) Du, Q.; Mezey, P. G. Heuristic molecular lipophilicity potential for computer-aided rational drug design: optimizations of screening functions and parameters. *J. Comput.-Aided Mol. Des.* **1998**, *12*, 451−471.
(27) Du, Q.; Arteca, G. A.; Mezey, P. G. Heuristic molecular lipophilicity potential for computer-aided rational drug design. *J. Comput.-Aided Mol. Des.* **1997**, *11*, 503−515.
(28) Du, Q.; Arteca, G. A. Modeling lipophilicity from the distribution of electrostatic potential on a molecular surface. *J. Comput.-Aided Mol. Des.* **1996**, *10*, 133−139.
(29) Du, Q.; Arteca, G. A. Derivation of fused-sphere molecular surfaces from the electrostatic potential, distribution. *J. Comput. Chem.* **1996**, *17*, 1258−1268.
(30) Mezey, P. G. Molecular Surfaces. In *Reviews in Computational Chemistry*, Vol. 1; Lipkowitz, K. B., Boyd, D. B., Eds.; VCH: New York, 1990; pp 265−294.
(31) Mezey, P. G. *Shape in Chemistry: An Introduction to Molecular Shape and Topology*; VCH Publishers: New York, 1993.
(32) Mezey, P. G. Quantum chemistry of macromolecular shape. *Int. Rev. Phys. Chem.* **1997**, *16*, 361−388.
(33) Mezey, P. G. Shape analysis. In *Encyclopedia of Computational Chemistry*, Vol. 4; Schleyer, P. v. R., Allinger, N. L., Clark, T., Gasteiger, J., Kollman, P. A, Schaefer, H. F., III, Schreiner, P. R., Eds.; John Wiley & Sons: Chichester, UK, 1998; pp 2582−2589.
(34) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Zakrzewski, V. G.; Montgomery, J. A., Jr.; Stratmann, R. E.; Burant, J. C.; Dapprich, S.; Millam, J. M.; Daniels, A. D.; Kudin, K. N.; Strain, M. C.; Farkas, O.; Tomasi, J.; Barone, V.; Cossi, M.; Cammi, R.; Mennucci, B.; Pomelli, C.; Adamo, C.; Clifford, S.; Ochterski, J.; Petersson, G. A.; Ayala, P. Y.; Cui, Q.; Morokuma, K.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Cioslowski, J.; Ortiz, J. V.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Gomperts, R.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Gonzalez, C.; Challacombe, M.; Gill, P. M. W.; Johnson, B. G.; Chen, W.; Wong, M. W.; Andres, J. L.; Head-Gordon, M.; Replogle, E. S.; Pople, J. A. *Gaussian 98*, revision A.9; Gaussian, Inc.: Pittsburgh, PA, 1998.
(35) Leo, A. J.; Hansch, C.; Elkins, D. *Chem. Rev.* **1971**, *71*, 525.
(36) Connolly, M. L. Solvent-accessible surfaces of proteins and nucleic acids. *Science* **1983**, *221*, 709−713.
(37) Connolly, M. L. Analytical molecular surface calculation. *J. Appl. Crystallogr.* **1983**, *16*, 548−558.
(38) Viswanadhan, V. N.; Ghose, A. K.; Revankar, G. R.; Robins, R. K. *J. Chem. Inf. Comput. Sci.* **1989**, *29*, 163−172.
(39) Doweyko, A. M.; Mattes, W. B. An application of 3D-QSAR to the analysis of the sequence specificity of DNA alkylation by uracil mustard. *Biochemistry* **1992**, *31*, 9388−9392.
(40) Audry, E.; Dubost, J. P.; Dallet, P.; Langlois, M. H.; Colleter, J. C. Le potentiel de lipophilie moleculaire: applications a une serie d'amines beta-adrenolytiques. *Eur. J. Med. Chem.* **1989**, *24*, 155.
(41) Smith, B. J. Solvation parameters for amino acids. *J. Comput. Chem.* **1999**, *20*, 428−442.

CI049707L