High Correlation between Hydrophobic Free Energy and Molecular Surface Area Characterized by Electrostatic Potential

Shuichi HIRONO,* Qian LIU, and Ikuro MORIGUCHI

School of Pharmaceutical Sciences, Kitasato University, 5-9-1 Shirokane, Minato-ku, Tokyo 108, Japan. Received April 25, 1991

For 41 hydrocarbons including paraffins, olefins, acetylenes and aromatics, the solvent accessible surface areas partitioned according to the electrostatic molecular potential levels have been computed and correlated with the hydrophobic free energies derived from their aqueous solubilities. The results indicate that the free energy is a linear function of the areas of two parts of the solvent accessible surface, one of which has an electrostatic potential above -1 kcal/mol and the other has a potential of -1 kcal/mol or below. The coefficients of the two parameters are about 25 cal/mol/Å² and 5 cal/mol/Å² at 25°C , respectively, and the multiple correlation coefficient is 0.987 .

Keywords hydrophobic free energy; electrostatic potential; molecular surface area; hydrocarbon; QSAR

Introduction

Hydrophobic free energy, or hydrophobic Gibbs energy change ΔG° has been defined in terms of the unitary free energy of transfer of a hydrocarbon molecule from a pure hydrocarbon liquid phase to an aqueous phase by Tanford and others.¹⁻³⁾ It can be determined from the solubility (X_s in mole fraction units) of the hydrocarbon in water. If μ_{HC}° and μ_{W}° are the hydrocarbon's standard chemical potentials in mole fraction units in hydrocarbon and water, respectively, the hydrophobic free energy is given as

$$\Delta G^\circ = \mu_{\text{W}}^\circ - \mu_{\text{HC}}^\circ = -RT \ln X_s \quad (1)$$

It is well established that ΔG° defined by Eq. 1 is a linear function of the solvent accessible surface area,⁴⁾ alias solvent cavity surface area⁵⁾ (*i.e.* area measured at the distance of closest approach of water molecules) of the saturated aliphatic or the aromatic hydrocarbon molecule.^{1,2,5,6)} The slopes of their lines, however, are different as reported by Tanford *et al.* (ΔG° per unit area amounts to about 25 cal/mol/Å² for paraffins and about 20 cal/mol/Å² for aromatics).^{1,2)} So, every given family of hydrocarbon needs its own equation to represent hydrophobic free energy by the above method. The analysis presented in this paper gives a new method for representing the ΔG° of miscellaneous hydrocarbons (paraffins, olefins, acetylenes and aromatics) by one equation. The hydrophobic free energy arises from contacts between hydrocarbon and water at the solute-solvent interface and thus, the number of water molecules that can be packed around the hydrocarbon molecule is an important quantity in hydrocarbon solubility. Hermann's idea was that the number of water molecules is related to the solvent accessible surface area (solvent cavity surface area) of the hydrocarbon molecule.⁵⁾ He treated the surface of a molecule as uniform. The solvent accessible surface of a molecule, however, is not uniform from the electrostatic point of view, as shown in Fig. 1,⁷⁾ especially, if the molecule has some unsaturated bonds. In Fig. 1, cyclohexane, 1,3-butadiene, heptadiene and toluene are representative of paraffins, olefins, acetylenes and aromatics, respectively, of which the electrostatic potential levels are symbolized as depicted in the figure. It is understandable that the solvent accessible surface of a molecule can be divided into some electrostatic potential regions. Our method is to treat the accessible surface as electrostatically irregular (not uniform) and to correlate the hydrophobic free

energy (ΔG°) of a hydrocarbon molecule with the solvent accessible surface areas divided according to the electrostatic potential values calculated on the surface of the molecule.

Calculation of Electrostatic Potential (EP) and Divided Solvent Accessible Surface Area According to EP Value
As the simple classical formula reproduces the quantum mechanically calculated EP quite well at van der Waals distances and beyond,⁸⁻¹¹⁾ we used the former formula to calculate the molecular potential on the solvent accessible surface. We used the value of a vacuum (1.0) as the dielectric constant to calculate the electrostatic potential. The classical representation of EP at a point x , $V(x)$, for a system of charges q_i at points r_i in a vacuum is given by

$$V(x) = \sum q_i / |r_i - x| \quad (2)$$

We used Mulliken net atomic charges obtained by a semiempirical molecular orbital method, MNDO¹²⁾ to represent q_i . The three-dimensional coordinates of the hydrocarbon molecules used here were obtained by the geometrical optimization method of MNDO using extended forms as the conformations. As most of the molecules used were small, the surface area can be thought to be little affected by the conformational change of the molecule. All calculations were carried out using the computer (HITAC M-680H) at the Computer Center of the University of Tokyo. The solvent accessible surface (SAS) of a molecule is defined as a curved surface traced out by the center of the sphere of rational radius representing a solvent molecule, as it slides over the van der Waals surface of the solute molecule.^{4,5)} We used 1.7 Å and 1.0 Å as the van der Waals radii of a carbon and a hydrogen atom, respectively, and 1.5 Å as the radius of a solvent (water) molecule as Hermann had done.

The SAS area divided according to EP values was calculated by a Monte Carlo method. To be more concrete, a number of points are uniformly generated on the SAS using three uniform random numbers¹³⁾ with a density of the points, 250 dots/Å² (0.004 Å²/dot). At each of those points, the EP value was calculated by Eq. 2 and the points on the SAS are classified into fourteen categories according to their EP values. That is to say, the 14 categories were separated by 13 boundaries of electrostatic potentials from -6 kcal/mol to $+6$ kcal/mol at an interval of 1 kcal/mol. The divided solvent accessible surface (DSAS) areas can be estimated by multiplying the number

Fig. 1. Electrostatic Potential Images on the Solvent Accessible Surface of Hydrocarbons (a) Cyclohexane, (b) 1,3-Butadiene, (c) Heptadiyne, (d) Toluene

The meanings of the symbols are as follows, *, $+3 \leq EP$; +, $+1 \leq EP < +3$; ·, $-1 < EP < +1$; -, $-3 < EP \leq -1$; Δ, $-5 < EP \leq -3$; □, $EP \leq -5$. The EP means the electrostatic potential (kcal/mol).

TABLE I. Divided Solvent Accessible Surface (DSAS) Areas According to the Electrostatic Potential Values

Notation of DSAS area (Å ²)	Electrostatic potential (kcal/mol)
$S_{HC}(+6 \leq EP)$	$+6 \leq EP$
$S_{HC}(+5 \leq EP < +6)$	$+5 \leq EP < +6$
$S_{HC}(+4 \leq EP < +5)$	$+4 \leq EP < +5$
$S_{HC}(+3 \leq EP < +4)$	$+3 \leq EP < +4$
$S_{HC}(+2 \leq EP < +3)$	$+2 \leq EP < +3$
$S_{HC}(+1 \leq EP < +2)$	$+1 \leq EP < +2$
$S_{HC}(0 \leq EP < +1)$	$0 \leq EP < +1$
$S_{HC}(-1 < EP < 0)$	$-1 < EP < 0$
$S_{HC}(-2 < EP \leq -1)$	$-2 < EP \leq -1$
$S_{HC}(-3 < EP \leq -2)$	$-3 < EP \leq -2$
$S_{HC}(-4 < EP \leq -3)$	$-4 < EP \leq -3$
$S_{HC}(-5 < EP \leq -4)$	$-5 < EP \leq -4$
$S_{HC}(-6 < EP \leq -5)$	$-6 < EP \leq -5$
$S_{HC}(EP \leq -6)$	$EP \leq -6$
$S_{HC}(-1 < EP) = S_{HC}(-1 < EP < 0) + S_{HC}(0 \leq EP < +1)$ $+ S_{HC}(+1 \leq EP < +2) + S_{HC}(+2 \leq EP < +3)$ $+ S_{HC}(+3 \leq EP < +4) + S_{HC}(+4 \leq EP < +5)$ $+ S_{HC}(+5 \leq EP < +6) + S_{HC}(+6 \leq EP)$	
$S_{HC}(EP \leq -1) = S_{HC}(-2 < EP \leq -1) + S_{HC}(-3 < EP \leq -2)$ $+ S_{HC}(-4 < EP \leq -3) + S_{HC}(-5 < EP \leq -4)$ $+ S_{HC}(-6 < EP \leq -5) + S_{HC}(EP \leq -6)$	

of the points belonging to each category by 0.004. The 14 areas are shown in Table I.

Results

The experimental data used for the analysis were the solubilities of hydrocarbons in water measured by McAuliffe.¹⁴ Those solubility data were used to calculate ΔG° at 25 °C using Eq. 1 for a variety of hydrocarbon compounds having carbon numbers of four or above. Table II gives a list of the 41 compounds with their SAS areas ($S_{HC}(\text{total})$) and two DSAS areas ($S_{HC}(-1 < EP)$, $S_{CH}(EP \leq -1)$) and ΔG° values calculated by the above method. The electrostatic potentials on the SAS of these molecules were in the range of -7 kcal/mol to $+5$ kcal/mol. 90% of the surface areas of the molecules used here belonged to the range of -5 — $+1$ kcal/mol potential, 3% of the areas to the -7 — -5 kcal/mol potential range and 2% to the $+1$ — $+5$ kcal/mol range. The values of ΔG° for these compounds are plotted against $S_{HC}(\text{total})$ in Fig. 2. Figure 2 shows that ΔG° values of every family of hydrocarbon (paraffins, olefins, acetylenes and aromatics) may follow a straight line with a similar slope but the intercepts of these lines are respectively different. It can also be seen from Fig. 2 that the diynes (1,6-heptadiyne and 1,8-nonadiyne) follow a different line from the other acetylenes (monoynes). In order to represent ΔG° of the above hydrocarbons by one equation, we examined a correlation between ΔG° and various combinations of the DSAS areas stated above and we have found that ΔG° is a linear function of a SAS area

TABLE II. Hydrophobic Free Energies and Electrostatic Potential Surface Areas for 41 Hydrocarbons

Hydrocarbon	$\Delta G^{\circ(a)}$	S_{HC} (total)	S_{HC} ($-1 < EP$) ^{b)}	S_{HC} ($EP \leq -1$) ^{b)}
Paraffins				
Isobutane	5.85	243.5	243.5	0.0
<i>n</i> -Butane	5.97	248.4	248.4	0.0
Cyclopentane	6.00	249.5	249.5	0.0
Isopentane	6.72	270.6	270.6	0.0
Cyclohexane	6.72	272.1	272.1	0.0
<i>n</i> -Pentane	6.84	278.8	278.8	0.0
Methylcyclopentane	6.88	277.0	277.0	0.0
2,2-Dimethylbutane	7.39	288.3	288.3	0.0
2-Methylpentane	7.56	302.2	302.1	0.0
Methylcyclohexane	7.63	299.6	299.6	0.0
<i>n</i> -Hexane	7.78	309.8	309.8	0.0
<i>n</i> -Heptane	8.57	344.1	344.2	0.0
2,2,4-Trimethylpentane	8.75	339.9	339.9	0.0
<i>n</i> -Octane	9.53	337.7	337.7	0.0
Olefins				
1,3-Butadiene	4.34	229.5	181.5	48.0
1-Butene	5.05	240.2	216.8	23.4
1,4-Cyclohexadiene	5.19	258.4	202.1	56.3
1,4-Pentadiene	5.23	263.1	214.2	48.9
Cyclopentene	5.25	242.9	211.5	31.4
2-Pentene	5.84	275.0	245.3	29.7
Cyclohexene	5.91	265.8	235.6	30.2
1-Pentene	6.03	270.7	246.7	24.0
1-Hexene	6.78	302.9	279.7	23.2
1-Methylcyclohexene	6.84	295.2	261.0	34.2
4-Vinylcyclohexene	6.93	316.3	281.4	34.9
Acetylenes				
1-Butyne	4.12	232.0	149.1	82.9
1-Pentyne	4.61	263.3	179.3	84.0
1,6-Heptadiyne	4.76	311.1	164.6	146.5
1-Hexyne	5.60	294.8	211.6	83.2
1,8-Nonadiyne	6.45	376.6	219.3	157.3
1-Heptyne	6.49	327.8	243.0	84.8
1-Octyne	7.38	361.6	278.2	83.4
Aromatics				
Benzene	4.62	247.0	181.1	65.9
Toluene	5.45	277.6	207.6	70.0
<i>p</i> -Xylene	6.10	307.1	232.2	74.9
<i>m</i> -Xylene	6.11	307.9	233.1	74.8
<i>o</i> -Xylene	6.18	301.2	227.4	73.8
Ethylbenzene	6.26	307.6	237.5	70.1
Propylbenzene	6.88	339.6	268.0	71.0
1,2,4-Trimethylbenzene	6.91	330.5	252.0	78.5
Isopropylbenzene	6.99	326.4	253.9	72.5

a) Hydrophobic free energy (kcal/mol) calculated by Eq. 1 from McAuliffe's solubility data, except for the values of *m*-xylene and *p*-xylene which are taken from ref. 2. b) Surface area ($\text{\AA}^2$) referred to Table I.

having EP above -1 kcal/mol, $S_{HC}(-1 < EP)$ (see Table I) as shown in Fig. 3. The following best-fit equation is found using linear regression analysis through an origin.

$$\Delta G^{\circ} = 0.025S_{HC}(-1 < EP) \quad (3)$$

$N=41, r=0.967, s=0.302$

where the unit of ΔG° is kcal/mol and that of $S_{HC}(-1 < EP)$ is $\text{\AA}^2$. As the maximum potential value of the molecules used here is less than $+5$ kcal/mol, $S_{HC}(-1 < EP)$ equals to $S_{HC}(-1 < EP \leq +5)$. Furthermore, since DSAS areas between $+1$ and $+5$ kcal/mol were only 2% of the total SAS areas, $S_{HC}(-1 < EP) \approx S_{HC}(-1 < EP < +1)$.

The contribution of the rest DSAS area, $S_{HC}(EP \leq -1)$ (see Table I) to ΔG° is evaluated using multivariate regression analysis through an origin and an equation

Fig. 2. Unitary Free Energies of Transfer of Hydrocarbons from the Pure Liquid to Aqueous Solution at 25°C from the Data Given in Table II against Solvent Accessible Surface Area

—□—, paraffins; —△—, olefins; —*—, acetylenes; —○—, aromatics.

Fig. 3. The Same Data as in Fig. 2, Plotted as a Function of Solvent Accessible Surface Area with $EP > -1$ kcal/mol

□, paraffins; △, olefins; *, acetylenes; ○, aromatics.

giving a better correlation than Eq. 3 is obtained as follows,

$$\Delta G^{\circ} = 0.025S_{HC}(-1 < EP) + 0.005S_{HC}(EP \leq -1) \quad (4)$$

$N=41, r=0.986, s=0.203$

In Fig. 1, the surface area represented by the symbols ($\cdot, +, *$), corresponds to $S_{HC}(-1 < EP)$ and that of moiety represented by ($-, \Delta, \square$) corresponds to $S_{HC}(EP \leq -1)$. As the minimum potential value of the molecules used here is greater than -7 kcal/mol, $S_{HC}(EP \leq -1)$ equals to $S_{HC}(-7 < EP \leq -1)$. Furthermore, since DSAS areas between -7 and -5 kcal/mol were only 3% of the total SAS areas, $S_{HC}(EP \leq -1) \approx S_{HC}(-5 < EP \leq -1)$.

The values of ΔG° calculated by Eq. 4 are plotted against those estimated from McAuliffe's solubility data

Fig. 4. Calculated (by Eq. 4) versus Observed Hydrophobic Free Energies from the Data Given in Table II

□, paraffins; △, olefins; *, acetylenes; ○, aromatics.

in Fig. 4.

Discussion

Tanford *et al.* concluded that the hydrophobic free energy per unit area of the SAS amounts to about 25 cal/mol/Å² at 25°C for saturated aliphatic hydrocarbons and to 20 cal/mol/Å² for aromatic hydrocarbons.¹⁾ As can be seen from Fig. 1 and Table II, the electrostatic potential values on the SAS are absolutely small (limited to the range of $-1 < EP < +1$ kcal/mol) for paraffins to result in $S_{HC}(EP \leq -1)$ being zero, while the other families (olefins, acetylenes and aromatics) have regions with bigger and smaller electrostatic potential values. The surface areas belonged to the range of -5 — $+1$ kcal/mol potential, -7 — -5 kcal/mol and $+1$ — $+5$ kcal/mol were 90, 3 and 2% of the total surface areas, respectively. Thus, the Eq. 4 can be regarded approximately the same as follows.

$$\Delta G^\circ \approx 0.025S_{HC}(-1 < EP < 1) + 0.005S_{HC}(-5 < EP \leq -1) + 0.000S_{HC}(-7 < EP \leq -5) + 0.000S_{HC}(+1 \leq EP < +5) \quad (5)$$

We conclude from Eqs. 3, 4 and 5 that the hydrophobic repulsive free energy between water and various hydrocarbons per unit area on the SAS having the electrostatic potential above -1 kcal/mol is 25 cal/mol/Å² at 25°C, is the same as Tanford's conclusion for paraffins. In addition, it is concluded that the hydrophobic free energy per unit area on the SAS with the electrostatic potential of -1 or below amounts to 5 cal/mol/Å², which means the surface of $EP \leq -1$ kcal/mol is less hydrophobic by a factor of about 5 than that of $EP > -1$ kcal/mol. It can be seen from Table II that $S_{HC}(-1 < EP)$ and $S_{HC}(EP \leq -1)$ are in the ratio of about three to one for aromatics. The average free energy per unit area of the two moiety of the SAS for

aromatics, thus, is calculated to be 20 cal/mol/Å² (i.e. $(3 \times 25 + 5)/4 = 20$), which is consistent with Tanford's apparent value for aromatics. It is also understandable from the larger area of $EP \leq -1$ kcal/mol that the diynes are different from the other acetylenes. Consequently, our method of ΔG° estimation by the DSAS areas is thought to be a more general method applying to miscellaneous hydrocarbons. The idea that the hydrophobic free energy is proportional to the solvent accessible surface ($S_{HC}(\text{total})$) of a solute molecule is correspondent to that the contribution of the enthalpy term to the hydrophobic free energy is regarded as almost zero for hydrocarbons. That is true of paraffins, but with regard to unsaturated hydrocarbons the solvent accessible surfaces are electrostatically not uniform as shown in Fig. 1, and hence the contribution of the enthalpy term to ΔG° might not be negligible. It might be thought that the contribution of the enthalpy term is taken into account by the use of the DSAS areas. It is also interesting from the standpoint of understanding hydration of a solute molecule that the hydrophobic free energy is little affected by the plus potential region on the SAS (25 cal/mol/Å² for the SAS of $EP > -1$ kcal/mol) but is affected by the minus potential region (5 cal/mol/Å² for the SAS of $EP \leq -1$ kcal/mol), which suggests that the hydrogen atoms of water work more predominantly on the interaction of water molecules with a solute than the oxygen atom. The DSAS area parameter is expected to be applied to a variety of molecules including heteroatoms as well as hydrocarbons and to be a new measure of hydrophobicity of compounds.

References

- 1) J. A. Reynolds, D. B. Gilbert, and C. Tanford, *Proc. Natl. Acad. Sci.*, **71**, 2925 (1974).
- 2) C. Tanford, "The Hydrophobic Effect," ed. by W. John and Sons, New York, 1979, Chapter 2.
- 3) S. J. Gill and I. Wadso, *Proc. Natl. Acad. Sci.*, **73**, 2955 (1976).
- 4) B. Lee and F. M. Richards, *J. Mol. Biol.*, **55**, 379 (1971).
- 5) R. Hermann, *J. Phys. Chem.*, **76**, 2754 (1972).
- 6) M. J. Harris, T. Higuchi, and J. H. Rytting, *J. Phys. Chem.*, **77**, 2694 (1973).
- 7) S. Hirono, H. Umeyama, and I. Moriguchi, *Chem. Pharm. Bull.*, **32**, 3061 (1984).
- 8) D. M. Hayes and P. A. Kollman, *J. Am. Chem. Soc.*, **98**, 3335 (1976).
- 9) D. M. Hayes and P. A. Kollman, *J. Am. Chem. Soc.*, **98**, 7811 (1976).
- 10) P. A. Kollman and D. M. Hayes, *J. Am. Chem. Soc.*, **103**, 2955 (1981).
- 11) P. K. Weiner, R. Langridge, J. M. Blaney, R. Schaefer, and P. A. Kollman, *Proc. Natl. Acad. Sci.*, **79**, 3754 (1982).
- 12) M. J. S. Dewar and W. Thiel, *J. Am. Chem. Soc.*, **99**, 4899 (1977).
- 13) M. Shibuya, *Ann. Inst. Stat. Math., Jpn.*, **14**, 81 (1962).
- 14) C. McAuliffe, *J. Phys. Chem.*, **70**, 1267 (1966).