BBA 75323

A MOLECULAR ORBITAL DESCRIPTION OF THE PARTITIONING OF AROMATIC COMPOUNDS BETWEEN POLAR AND NONPOLAR PHASES

KENNETH S. ROGERS* AND ARTHUR CAMMARATA

Department of Biochemistry, Medical College of Virginia, Richmond, Va. 23219 and Laboratory of Physical Medicinal Chemistry, School of Pharmacy, Temple University, Philadelphia, Pa. 19140 (U.S.A.)

(Received April 4th, 1969)

SUMMARY

The partitioning of aromatic molecules between immiscible nonpolar–polar phases (*n*-octanol–aqueous 50 mM sodium phosphate buffer (pH 7.4)) is considered experimentally and theoretically from a molecular orbital theory point of view. An equation is developed that satisfactorily describes the partitioning process:

$$\ln \overline{P} = \left(\frac{\bar{a} - \bar{a}'}{RT}\right) \Sigma_s \left| Q_s^T \right| + \left(\frac{\bar{b} - \bar{b}'}{RT}\right) \Sigma_s S_s^E$$

where P is the chemical's partition coefficient, Q_s^T is the charge density obtained from considerations of both π and σ electron frameworks for the aromatic compound, and S_s^E is induced polarization.

Thirty aromatic molecules representing four chemical classes (hydrocarbons, heterocycles, substituted benzenes, and substituted indoles) have partition coefficients that are correlated by this model equation. These correlations provide new examples for the applicability of molecular orbital theory to studies of biological activity and chemical structure interrelationships.

INTRODUCTION

Many diverse biochemical processes such as plant growth[1], drug activity[2], enzyme inhibition[3], and erythrocyte hemolysis[4] are dependent on the lipophilicity of the aromatic substrates involved. To gain insight into these events, it is common practice to define the partition coefficient for the distribution of the substrate between an aqueous and a nonpolar phase as a measure of lipophilicity. The maximum biological responses observed for a series of related substrates may then be considered in terms of molecular partitioning between the aqueous biophase and nonpolar regions of the biophase such as membranes and interfacial barriers.

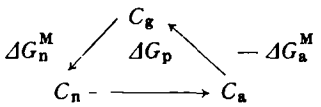
At the molecular level, it might be expected that the electronic characteristics of a substrate will control the affinity of the substrate for an aqueous or an organic

* To whom inquiries should be addressed: Department of Biochemistry, Medical College of Virginia, Richmond, Va. 23219, U.S.A.

phase, and in a previous note[5], it was reported that the partition coefficients for a series of aromatic compounds can be correlated with molecular orbital indices calculated for their π electron frameworks. In this report, it is shown that both σ and π electronic properties must be considered in order to describe satisfactorily the partitioning process. An interpretation of the correlations obtained is also presented using as a basis a recent theory of chemical reactivity[6,7] as modified for application to biological systems[8,9].

PARTITION THEORY

Partitioning of a compound between a polar and a nonpolar phase may be considered in terms of the cyclic process



where C_g, C_a, C_n is the compound as found in the gaseous, aqueous polar, and nonpolar phases, respectively. This process may be represented by the series of transformations

$$C_g = C_n \qquad \Delta G_n^M \text{ (mixing in nonpolar phase)} \tag{1}$$

$$C_n = C_a \qquad \Delta G_p \text{ (partitioning)} \tag{2}$$

$$C_a = C_g \qquad -\Delta G_a^M \text{ (mixing in aqueous phase)} \tag{3}$$

for which the overall transformation is

$$C_g = C_g \qquad \Delta G = \Delta G_n^M - \Delta G_a^M + \Delta G_p \tag{4}$$

Since the system is unchanged, no net free-energy change is possible so that

$$\Delta G = 0 = \Delta G_n^M - \Delta G_a^M + \Delta G_p \tag{5}$$

and hence

$$\Delta G_p = \Delta G_a^M - \Delta G_n^M \tag{6}$$

From the usual definition[1] of a partition coefficient $\overline{P}$

$$\Delta G_p = -RT \ln \overline{P} \tag{7}$$

it is thus found that

$$\ln \overline{P} = \frac{1}{RT}(\Delta G_n^M - \Delta G_a^M) \tag{8}$$

or, in other words, the natural logarithm of a partition coefficient should be proportional to the difference in the free energy of mixing for the equilibrium concentrations of compound in the nonpolar and the aqueous phases, respectively.

As interactions between molecules in the gas phase may be considered negligible, *i.e.* the gaseous molecules behave ideally, the predominant contribution to the free energy of mixing must be due to interactions between molecules of compound and

molecules of solvent. Consider the interaction between a single molecule of compound and a single molecule of solvent. Second-order quantum perturbation theory may be used to describe the interaction, since in the case of weak interactions this order of approximation in the present treatment leads to exact results[10]. In particular, the formalism giving this interaction in terms of the ground-state molecular orbitals of the interacting species, expressed respectively as a linear combination of atomic orbitals[6,7], may be used in its modified form[8,9].

According to this approach, the free energy of mixing may be thought of as consisting of ionic E^{I} and non-ionic E^{C} contributions due to interactions between the atoms of solvent (r) and solute (s)

$$\Delta G^{M} = \Sigma_r \Sigma_s (E^{I}_{rs} + E^{C}_{rs}) \tag{9}$$

If in solution a number of molecules of solvent interact in a random fashion with a solute molecule, each atom of the solute experiences an averaged effect due to the solvent atoms. Under these conditions, the ionic and non-ionic components of Eqn. 9 may each be factored[9] into two parts*, and those factors representing averaged effects due to solvent may be abbreviated $\bar{a}$ and $\bar{b}$,

$$\Delta G^{M} = \bar{a} (\Sigma_s E^{I}_{s}) + \bar{b} (\Sigma_s E^{C}_{s}) \tag{10}$$

Each atom of the solute may have its ionic contribution E^{I}_{s} given by the calculated sum of the net σ and π charges Q^{T} on the atom. The absolute value of the sum will be used, since in interacting randomly with solvent, a solute atom can experience either positive or negative charges of the solvent atoms. Hence,

$$E^{I}_{s} = | Q^{T}_{s} | \tag{11}$$

For the non-ionic contribution of each atom of solute E^{C}_{s}, the molecular orbital

* The interaction of a compound with a molecule of solvent may also be expressed in the following manner

$$\Delta G^{M} = \Sigma_r \Sigma_s \left[\frac{Q^{T}_{r} Q^{T}_{s} e^{2}}{\varepsilon_{rs} D_{rs}} + \Sigma^{\text{unocc.}}_{m} \Sigma^{\text{occ.}}_{n} \frac{c^{2}_{mr} c^{2}_{ns} \beta^{2}}{E_m - E_n} \right]$$

where atom r of solvent interacts with atom s of compound. The charge densities of atoms r and s are Q^{T}_{r} and Q^{T}_{s}, respectively. The effective dielectric between r and s is ε_{rs}; e is the electrostatic unit of charge; and D_{rs} is the distance between r and s. $\Sigma^{\text{occ.}}_{n}$ and $\Sigma^{\text{unocc.}}_{m}$ refer to summation over the occupied and unoccupied orbitals of compound and solvent, respectively. The coefficient for atom r is c^{5} when the linear combination of atomic orbitals molecular orbital m has the energy E_m, and c_{ns} is the coefficient for atom s when the molecular orbital n has the energy E_n.

If in solution, a number of molecules interact randomly with the chemical so that each atom s of solute experiences an average effect due to each atom r of solvent, then those terms referring to the atoms of solvent may be replaced by an averaged value and factored from the summation. Hence,

$$\Delta G^{M} = \Sigma_s \left[\bar{a} Q^{T}_{s} + \bar{b} \Sigma^{\text{occ.}}_{n} \frac{c^{2}_{ns}}{\bar{E} - E_n} \right]$$

$$\bar{a} = \left(\Sigma_r \frac{Q^{T}_{r} e^{2}}{\varepsilon_{rs} D_{sr}} \right)_{\text{av.}} ; \quad \bar{b} = (\Sigma_r \Sigma^{\text{unocc.}}_{m} c^{2}_{mr} \beta^{2})_{\text{av.}}$$

index S_s^E, often termed superdelocalizability or delocalizability[11,12] may be used as a measure of induced polarization[13]. Hence,

$$E_s^C = S_s^E \tag{12}$$

Combining Eqns. 8, 10, 11 and 12 affords an expression that relates partition coefficients to indices that may be calculated using molecular orbital approaches

$$\ln \bar{P} = \frac{(\bar{a} - \bar{a}')}{RT} \Sigma_s |Q_s^T| + \frac{(\bar{b} - \bar{b}')}{RT} \Sigma_s S_s^E \tag{13}$$

In Eqn. 13, a primed coefficient is taken to identify terms associated with a polar or aqueous solvent. From the coefficients it can be seen that if in partitioning between an aqueous and a nonpolar phase the distribution of a compound in the aqueous phase is favored then $\bar{a} < \bar{a}'$ and the first term of Eqn. 13 should be negative. On the other hand, if distribution of the compound in the nonpolar phase is favored then $\bar{b} > \bar{b}'$ and the second term of Eqn. 13 should be positive. Thus, correlations obtained by the use of this relation represent the hydrophilic–lipophilic balance for the series of compounds investigated.

TABLE I

HYDROCARBONS: PARTITION COEFFICIENTS AND CALCULATED MOLECULAR ORBITAL INDICES

Compound	$\Sigma_s\|Q_s^T\|$	$\Sigma_s S_s^E$	$\ln \bar{P}$, *obs.*	$\ln \bar{P}$, *calc.*	*Ratio* $\frac{\ln \bar{P}, \text{calc.}^*}{\ln \bar{P}, \text{obs.}}$
Biphenyl	1.076	10.292	7.272	7.630	1.049
Naphthalene	0.866	8.874	6.932	6.611	0.954
Indene	1.084	8.464	6.430	6.243	0.971
Benzene	0.636	4.998	3.584	3.735	1.042

* Ratio expresses the correlation between calculated and observed values; 1.000 is a perfect correlation.

TABLE II

HETEROCYCLES: PARTITION COEFFICIENTS AND CALCULATED MOLECULAR ORBITAL INDICES

Compound	$\Sigma_s\|Q_s^T$	$\Sigma_s S_s^E$	$\ln \bar{P}$, *obs.*	$\ln, \bar{P}$ *calc.*	*Ratio* $\frac{\ln \bar{P}, \text{calc.}^*}{\ln \bar{P}, \text{obs.}}$
Carbazole	1.799	12.177	7.576	6.136	0.810
Thianaphthene	1.812	14.698	7.113	7.776	1.093
Indole	1.438	9.114	5.182	4.929	0.951
Quinoline	1.487	8.333	4.754	4.300	0.905
Benzothiazole	2.299	12.000	4.673	4.879	1.044
Indazole	1.245	8.881	4.190	5.215	1.245
Benzoxazole	1.817	8.114	3.663	3.403	0.929
Benzimidazole	2.177	8.443	2.773	2.794	1.008
Oxindole	2.206	8.843	2.639	3.000	1.137
Quinoxaline	2.352	7.954	1.946	2.078	1.068

* Ratio expresses the correlation between calculated and observed values; 1.000 is a perfect correlation.

TABLE III

SUBSTITUTED INDOLES: PARTITION COEFFICIENTS AND CALCULATED MOLECULAR ORBITAL INDICES

Indole	$\Sigma_s\|Q_s^T\|$	$\Sigma_s S_s^E$	*ln* $\bar{P}$, *obs.*	*ln* $\bar{P}$, *calc.*	*Ratio*	$\frac{ln\ \bar{P},\ calc.^*}{ln\ \bar{P},\ obs.}$
5-Bromo	1.488	11.481	6.906	6.557	0.949	
1,2-Dimethyl	1.538	12.051	6.498	6.614	1.018	
5-Chloro	1.492	11.343	6.269	6.458	1.030	
5-Methyl	1.546	10.427	6.180	5.652	0.915	
3-Methyl	1.433	10.691	5.984	6.403	1.070	
5-Fluoro	1.638	9.973	5.268	4.906	0.931	
Unsubstituted	1.438	9.114	5.182	5.484	1.058	
5-Methoxy	1.855	11.883	4.754	4.833	1.017	
5-Hydroxy	2.136	10.524	2.434	2.569	1.055	

* Ratio expresses the correlation between calculated and observed values; 1.000 is a perfect correlation.

TABLE IV

SUBSTITUTED BENZENES: PARTITION COEFFICIENTS AND CALCULATED MOLECULAR ORBITAL INDICES

Benzene derivate	$\Sigma_s\|Q_s^T\|$	$\Sigma_s S_s^E$	*ln* $\bar{P}$, *obs.*	*ln* $\bar{P}$, *calc.*	*Ratio*	$\frac{ln\ \bar{P},\ calc.^*}{ln\ \bar{P},\ obs.}$
Anilyl	1.873	12.302	7.427	6.927	0.933	
Phenyl	1.076	10.292	7.272	7.247	0.997	
Dimethylamino	1.198	10.266	6.026	6.866	1.139	
Methyl	0.826	6.253	4.852	4.324	0.891	
Methoxy	1.169	7.674	4.700	4.679	0.995	
Unsubstituted	0.636	4.998	3.584	3.712	1.036	
Hydroxy	1.457	6.335	3.277	2.732	0.834	
1-Hydroxy, 2-amino	2.343	8.274	2.542	2.114	0.832	
Amino	1.636	6.804	1.775	2.676	1.508	
1-Hydroxy, 2-carboxy	3.716	8.468	−1.514	−1.337	0.883	

* Ratio expresses the correlation between calculated and observed values; 1.000 is a perfect correlation.

MATERIALS AND METHODS

Partition coefficients for the materials listed in Tables I–IV were determined following the procedure described previously[4]. The ratio of the equilibrium concentrations of compound in *n*-octanol and aqueous 50 mM sodium phosphate buffer (pH 7.4) is the value used for the respective partition coefficients. For the compounds included in this study, the maximum error in determination of concentration is $\pm$ 5 %. All calculations were performed on an IBM 1130 computer. The simple Hückel method[12] was used for the π-system calculations and the method of DEL RE and co-workers[14,15] as modified by BERTHOD AND PULLMAN[16] was used for the σ-system calculations. The parameters employed were those suggested by the references cited; methyl substituents were treated as heteroatoms. Standard package programs provided by IBM were used for the statistical analyses.

RESULTS AND DISCUSSION

Multiple regression analyses of the data presented in Tables I–IV were performed using Eqn. 13 as a model. Each set of data is correlated by the following equations:

Hydrocarbons (Table I)

$$\ln P = -0.266\ (\pm 2.963)\ \Sigma_s \mid Q_s^T \mid + 0.758\ (\pm 0.279)\ \Sigma_s S_s^E + 0.116 \qquad (14)$$
$$(t_1 = -0.089;\ t_2 = 2.695;\ r = 0.983;\ F = 14.2;\ \text{S.E.} = 0.537)$$

Heterocycles (Table II)

$$\ln \overline{P} = -2.279\ (\pm 0.656)\ \Sigma_s \mid Q_s^T \mid + 0.663\ (\pm 0.111)\ \Sigma_s S_s^E + 2.168 \qquad (15)$$
$$(t_1 = -3.469;\ t_2 = 5.930;\ r = 0.931;\ F = 22.7;\ \text{S.E.} = 0.764)$$

Substituted indoles (Table III)

$$\ln \overline{P} = -5.319\ (\pm 0.580)\ \Sigma_s \mid Q_s^T \mid + 0.566\ (\pm 0.142)\ \Sigma_s S_s^E + 7.974 \qquad (16)$$
$$(t_1 = -9.163;\ t_2 = 3.974;\ r = 0.969;\ F = 46.5;\ \text{S.E.} = 0.381)$$

Substituted benzenes (Table IV)

$$\ln \overline{P} = -2.639\ (\pm 0.236)\ \Sigma_s \mid Q_s^T \mid + 0.887\ (\pm 0.095)\ \Sigma_s S_s^E + 0.958 \qquad (17)$$
$$(t_1 = -11.302;\ t_2 = 9.460;\ r = 0.980;\ F = 86.1;\ \text{S.E.} = 0.613)$$

The quantities in parentheses appearing after each coefficient in the equation are the standard errors associated with the estimate of the coefficients. Statistics for the equation are also indicated in parentheses. These are the t-tests for the first and second coefficients; the multiple correlation coefficient; the F-test; and the standard error of the estimate provided by the equation, respectively.

Each correlation is acceptable on statistical grounds, with the possible exception of the equation derived using the data of Table I. In this instance the correlation is compromised because of the limited number of data points. Nevertheless, induced polarization (S^E) is indicated by t-test as the primary factor influencing the partitioning of hydrocarbons. Of greater importance, the signs of the coefficients in each of the Eqns. 14–17 are as expected from previous considerations.

A comparison of the coefficients appearing in Eqns. 14–17 suggests that partitioning of a compound between an aqueous and a nonpolar phase will be controlled to varying degrees depending on the relative total charge and polarizability of the compound. The close correspondance between the coefficients of the respective equations indicates that while the structure of an aromatic compound may vary widely, its partitioning characteristics are not unlike those for other compounds. Hence any property which correlates with partition coefficients should be essentially non-specific with regard to the structure of the aromatic compounds involved. In the present case, all thirty compounds found in Tables I–IV have their partition coefficients correlated by the relation

$$\ln \overline{P} = -2.705\ (\pm 0.231)\ \Sigma_s \mid Q_s^T \mid + 0.708\ (\pm 0.062)\ \Sigma_s S_s^E + 2.467 \qquad (18)$$
$$(t_1 = -11.710;\ t_2 = 11.039;\ r = 0.943;\ F = 107.9;\ \text{S.E.} = 0.732)$$

The correlation may be improved and the coefficients made more specific if attention is restricted to a congeneric series of compounds, but an improvement of roughly 0.1–0.4 ln unit in the estimate of ln P is all that could be expected using the present treatment.

In correlating the partition coefficients of compounds, it is important to note that σ charges must be included if the compounds contain groups such as carboxy, hydroxy, amino, or methoxy. All of these groups can hydrogen bond with an aqueous solvent, and the formation of a hydrogen bond to or from the electronegative atom is influenced greatly by the σ charge on the atom[13]. For this reason, use of only π molecular orbital indices ordinarily provides poor correlations with partition coefficients.

CONCLUSIONS

A number of rigorous molecular orbital approaches to describe the nature of secondary interactions between molecules have appeared in the recent literature[17,18]. With few exceptions[19], the application of the results of these studies to real systems is unwieldy or presently impossible. The approach taken in this article, while not entirely rigorous from a molecular orbital standpoint, is relatively easy to apply and provides results which are very satisfactory considering the simplicity of the methods used.

One of the most important conclusions arising from this work is that in describing the partitioning process many of the terms appearing in more rigorous treatments of secondary interactions seem to subtract out or at least are almost negligible relative to coulombic and induced polarization interactions. Partitioning into an aqueous phase may thus be considered as 'charge-controlled', while partitioning into a nonpolar phase may be considered as 'polarizability-controlled'. It is the relative contribution of each controlling factor which therefore predominantly determines the partition coefficient observed for a compound.

Secondly, since many biochemical processes are found to be related to the partition coefficients of the substrates involved[1,3,4], it seems possible that further insight into the molecular events occurring during these biological processes might be gained by use of the approach described. Investigations bearing on this possibility are in progress.

ACKNOWLEDGEMENT

The authors wish to express their appreciation to the Department of Biometry at the Medical College of Virginia for their guidance during the computational aspects of this work. This investigation was supported in part by grants from A. H. Robins Co., National Science Foundation, and John A. Hartford Foundation.

REFERENCES

1 C. HANSCH, R. M. MUIR, T. FUJITA, P. P. MALONEY, F. GEIGER AND M. STREICH, *J. Am. Chem. Soc.*, 85 (1963) 2817.
2 C. HANSCH AND T. FUJITA, *J. Am. Chem. Soc.*, 86 (1964) 1616.
3 C. HANSCH AND E. W. DEUTSCH, *Biochim. Biophys. Acta*, 126 (1966) 117.
4 K. S. ROGERS, *Proc. Soc. Exptl. Biol. Med.*, 130 (1969) 1140.
5 K. S. ROGERS AND A. CAMMARATA, *J. Med. Chem.*, 12 (1969) 692.
6 G. KLOPMAN AND R. F. HUDSON, *Theoret. Chim. Acta*, 8 (1967) 165.

7 G. KLOPMAN, *J. Am. Chem. Soc.*, 90 (1968) 223.
8 A. CAMMARATA, *J. Med. Chem.*, 11 (1968) 1111.
9 A. CAMMARATA, *J. Med. Chem.*, 324 (1969) 12.
10 H. MARGENAU, *Rev. Mod. Phys.*, 11 (1939) 1.
11 K. FUKUI, T. YONEZAWA AND C. NAGATA, *J. Chem. Phys.*, 27 (1957) 1247.
12 A. STREITWIESER, JR., *Molecular Orbital Theory for Organic Chemists*, Wiley, New York, 1961, pp. 52, 330.
13 A. CAMMARATA AND R. L. STEIN, *J. Med. Chem.*, 11 (1968) 829.
14 G. DEL RE, *J. Chem. Soc.*, (1958) 4031.
15 G. DEL RE, B. PULLMAN AND T. YONEZAWA, *Biochim. Biophys. Acta*, 75 (1963) 153.
16 H. BERTHOD AND A. PULLMAN, *J. Chim. Phys.*, (1965) 942.
17 H. O. HERSHFELDER, *Advan. Chem. Phys.*, 12 (1968) 1.
18 B. PULLMAN, *Molecular Associations in Biology*, Academic Press, New York, 1968, p. 230.
19 R. REIN AND M. POLLAK, *J. Chem. Phys.*, 47 (1967) 2089.