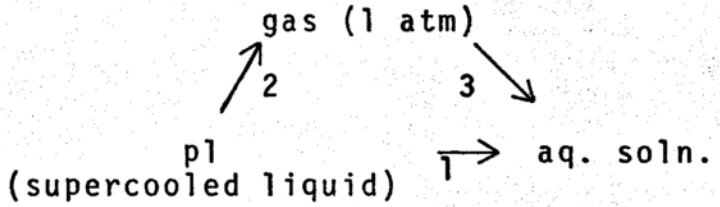
A THERMODYNAMIC ANALYSIS OF THE AQUEOUS SOLUBILITY AND HYDROPHOBICITY OF HYDROCARBONS USING MOLECULAR SURFACE AREA

Shabbir Tyabji Anik

Under the supervision of Associate Professor

Gordon L. Amidon

The solution process, pure liquid to aqueous solution has been partitioned into solute-solute and solute-solvent interactions using the thermodynamic cycle shown below.



In an earlier study (S. T. Anik, M.S. Thesis, University of Wisconsin, 1976), it was shown that for the aliphatic and aromatic hydrocarbons, although the free energy change per $\AA^2$ for step 1 was nearly the same, the partitioning process provided a distinctly different view. Thus for the aliphatic hydrocarbons only 20% (6 cal/$\AA^2$) of the $\delta\Delta G°$ for step 1 was due to solute-solvent interactions (step 3), whereas for the aromatic hydrocarbons the $\delta\Delta G°$ was negative indicating that the aromatic residue has a net favorable interaction with water.

The free energy changes for the cycle steps have been further partitioned into enthalpic and entropic

contributions. The results indicate that when comparing the $\delta\Delta G°$, $\delta\Delta H°$ and $\delta T\Delta S°$ for the cycle steps for these two classes of hydrocarbons, the only similarity lies in the $\delta\Delta G°$ for the pure liquid to aqueous solution process. The remaining thermodynamic parameters are different in magnitude and in several cases sign, strongly emphasizing the differing nature of the interactions of these hydrocarbons with water.

The free energy data for the cycle steps for the alkyl aromatic hydrocarbons have also been analyzed by regression analysis using the group surface area approach. The model which gives the best estimates of the data is

$$\Delta G° = \theta_0 + \theta_1 ArC + \theta_2 ArH + \theta_3 ALov + \theta_4 Remalp + \theta_5 I$$

where ArC is aromatic carbon, ArH is aromatic hydrogen, Alov is aliphatic overlap, and Remalp is remaining aliphatic areas and I is an indicator variable. For the pure liquid to aqueous solution process the model simplifies to

$$\Delta G° = \theta_0 + \theta_1 AromSA + \theta_2 AlSA + \theta_3 I$$

The coefficients of this model are identical to those obtained for the individual hydrocarbons indicating the compensatory nature of steps 2 and 3. For the gas phase to aqueous solution process the coefficients suggest that the aromatic carbons are intrinsically hydrophilic while the aliphatic hydrogens are more intrinsically hydrophobic than

the aliphatic hydrogens. The higher gas phase solubility of the methyl benzenes compared to the corresponding n-alkyl benzenes is a consequence of the large positive coefficient of the ArH term. The model also predicts reasonably well the free energy changes for the alkyl naphthalenes and higher polycyclic aromatics, compounds not included in the original analysis. The observed data for the free energy of solvation have also been compared to values calculated by the scaled particle theory. The SPT generally tends to overestimate the interaction energy however the trends in the $\delta\Delta G°$ and particularly the $\delta\Delta S°$ of solvation are correctly reproduced.

Finally the results for the $\delta\Delta G°_{solvation}$ for the aromatic hydrocarbons has been used to estimate the solvent contribution to the binding of the series benzene, naphthalene, and anthracene to the active site of α-chymotrypsin. The solvent contribution to the free energy of binding becomes increasingly unfavorable in going from benzene to anthracene and the observed order of binding: anthracene > naphthalene > benzene is a result of the increasing Van der Waals or non-bonded interactions between the ligand molecule and the active site.

APPROVED: Gordon L. Amidon

Gordon L. Amidon

4/10/78

Date

A THERMODYNAMIC ANALYSIS OF THE AQUEOUS SOLUBILITY AND HYDROPHOBICITY OF HYDROCARBONS USING MOLECULAR SURFACE AREA

BY

SHABBIR TYABJI ANIK

A thesis submitted in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

(Pharmacy)

at the

UNIVERSITY OF WISCONSIN-MADISON

1978

To

My Parents

For The Most Wonderful Upbringing

Any Son Would Want

and

My Sister

Who Always Put My Interests Above Her Own,

With Love.

ACKNOWLEDGEMENTS

I would like to express my deep appreciation to Professor Gordon L. Amidon for some of the most stimulating and enjoyable years of my educational career. He has contributed immeasurably to my development, both scientifically and as an individual, through the many long and varied discussions we have had and for which he so generously gave his time. And above all his patience to listen to the most trivial and his support and encouragement which have been a tremendous boost to my self confidence. My association with him shall always be remembered.

I would also like to thank:

Prof. S. K. Pradhan who was a source of inspiration throughout my undergraduate career,

Dr. Robert S. Pearlman for his help with the computer programs, his suggestions and friendship,

Dr Sam H. Yalkowsky for his interest in my work and his many valuable suggestions and discussions,

The Pharmaceutics Faculty for their interest in my development as a graduate student,

All my colleagues in Pharmacy for their help and friendship,

The J. N. Tata Endowment and the R. D. Sethna Scholarship Funds which made my visit to the U.S. possible,

The National Institutes of Health (GM 20437) and the

Graduate School for financial support through my graduate studies.

TABLE OF CONTENTS

Page

Page

Page

I. INTRODUCTION

A. General Considerations

The aqueous solubility of a compound is of central importance in pharmaceutical development due to its effect on such diverse phenomena as dissolution, partitioning, absorption, stability, sustained release, pain on injection and taste. Hence structural modification of a compound so as to alter these properties through changes in solubility, and the ability to *a priori* estimate the solubility of a given compound or its derivative is of considerable value in improving drug product efficacy. The study of the solubility process also provides, at a more fundamental level, an understanding of the complex processes in solution. A convenient approach to elucidating solute-solvent interactions for use in predicting solubility behavior is from an analysis of the solution properties of simpler molecules. Research in this area has extended to a variety of solutes and solvents, however the aqueous solution properties of a solute have received particular attention because of their relevance to the biological sciences and an interest in understanding the unique physical properties of the solvent water.

There has been considerable attention focused in the past toward the development of theoretical and semiempirical models for characterization of the solution process.

The success of theoretical models has been almost entirely restricted to interactions of a nonpolar solute in a nonpolar solvent. These represent the simplest of the three types of solute-solvent interactions, the other two being nonpolar-polar and polar-polar. The nonpolar-polar systems are of considerable interest in the pharmaceutical field, in particular the behavior of nonelectrolytes in water.

B. Regular Solution Approach

For the system of a nonpolar solute in a nonpolar solvent the most successful theory is the regular solution theory proposed by Hildebrand and coworkers (1). The main assumptions of the theory are: i) The excess entropy of mixing is zero; ii) The excess free energy change is a direct result of the excess enthalpy of mixing; and, iii) The interaction energy between the two components is given by the geometric mean of the cohesive energy densities of the individual components. On this basis the following expression can be obtained (2)

$$RT \ln\left(\frac{a_2}{x_2}\right) = V_2 \phi_1^2 (\delta_1 - \delta_2)^2 \qquad \text{Eq. 1}$$

where subscripts 1 and 2 stand for solvent and solute respectively, a_2 is the activity, x_2 is mole fraction, V_2 is the molar volume, ϕ_1 is the volume fraction and $\delta = (E/V)^{1/2}$ is the cohesive energy density and is commonly known

as the solubility parameter. The equation gives good results for systems obeying the basic assumptions (3,4). This approach however fails for polar compounds due to its inability to account for specific interactions such as hydrogen bonding, nonideal entropies of mixing and failure of the geometric mean rule. To overcome some of these difficulties several methods for calculating polar and hydrogen bonding contributions to the solubility parameter giving two and three component solubility parameters have been proposed. These methods have been discussed in a recent review (5).

C. Group Contribution Method

1. Background

The development of theoretical models for solutions of polar substances, particularly in water, has been extremely difficult and due to i) the complex structural properties of water itself and ii) an increased complexity in accounting for specific solute-solvent and solute-solute interactions. As a result many attempts to predict aqueous solubility have been of a semi-empirical nature and have met with varying degrees of success. In the group contribution methods the basic assumption is that the free energy of solution of a solute is additively composed of the independent contributions of the constituent functional groups. The methods differ mainly in the definition of a

group or groups used to partition the molecule into its constitutive parts.

One of the early investigators in the field of solution thermodynamics was Langmuir (6). He proposed the "Principle of independent surface action" where the force field around a group which is accessible to other molecules is characteristic of that group and is independent of the nature of the rest of the molecule. Butler (7) further developed this concept and calculated free energy contributions to the solubility on a group basis. He observed a constant decrease in the solubility of the normal aliphatic alcohols with increasing chain length and from determinations of the activity coefficients of their aqueous solutions at infinite dilution calculated a value of 800 cal/ CH_2 group. In a series of papers (8,9,10,11) he further analyzed the behavior of aqueous solutions of alcohols in terms of branching, the properties of the pure solute phase, and solute-solvent interactions based on the free energy of solvation. Using an approach similar to Langmuir's, he determined functional group values for ethers, alcohols, ketones and amines. He also noted the shortcomings of the additivity assumption for polyfunctional compounds using the series ethanol, ethylene glycol, propanol and glycerol to illustrate the non-additivity in the free energy of solution with each additional hydroxyl group. Since then there have been a large number of group

contribution methods proposed for calculating various physical properties including solubility. Carbon numbers (12,13), functions of the number of carbon atoms to account for interaction terms (14,15), parachor values (16), types of atoms and bonds (17), π values (18), F-values (19), molar volumes (20), cavity surface area (21,22), and molecular connectivity (23) have all been used in correlations with aqueous solubilities, partition coefficients and activity coefficients of different classes of compounds. A comprehensive review by Davis et al. (24) discusses the merits of the different approaches to the group contribution method.

2. The Hansch Approach

In recent years the group contribution method has received considerable attention in the pharmaceutical literature due to extensive work by Hansch and his coworkers (25-29) on partition coefficients and their relation to biological activity. Since the partition coefficient does not involve the lattice energy of the crystalline solute, a major complexity in the case of polyfunctional compounds is reduced and as a result the method has been very successful.

Hansch defined the substituent parameter π as

$$\pi x = \log Kp(RX) - \log Kp(RH) \qquad \text{Eq. 2}$$

where Kp is the partition coefficient between octanol and water and RX and RH are the substituted and parent compound respectively. The partition coefficient is then calculated as the sum of the π values of the different structural parts of the molecule. Extensive tables of partition coefficients and π group contributions are available. Following this initial approach several modifications have been introduced into the equation relating partition coefficients to π values, which include interaction terms, steric parameters, polarizabilities, etc. to improve correlations (18,28). Recognizing the close relationship between the aqueous solubility of a molecule and its lipophilicity, Hansch was also able to show good correlation between the solubilities and partition coefficients for a variety of organic liquids (30). Yalkowsky (31) has extended this to include organic solids. He found a reasonably good correlation of the aqueous solubility with partition coefficient and melting point given by the relation

$$\log S = 0.5 - 1.0 \log PC - .01\ T_m \qquad \text{Eq. 3}$$

where S is the molar solubility, PC is the partition coefficient and T_m is the melting point in °C.

3. Cavity Surface Area Method

Along with the principle of independent surface action, Langmuir suggested that the energetics of forming a cavity of suitable size in the solvent to accomodate the solute molecule may play an important role in the solution thermodynamic process. This has been further discussed by Uhlig (32) and Eley (33). From a consideration of interfacial energetics, the surface area of the cavity seems a natural parameter to use in determining free energy changes associated with dissolving a solute in a solvent. However due to lack of available methods for accurately calculating molecular surface area, previous workers used molar volume (20) or related quantities such as parachor (16) which reflect the cavity size, to correlate with solubility and partition coefficients. Thus McAuliffe found good correlation between molar volume and solubility for several classes of hydrocarbons. Recently the cavity model for understanding solution thermodynamics is finding increasing use as a result of i) the development of the scaled particle theory and ii) a surface area computational program developed by Hermann (21) which allows accurate and consistent calculation of the molecular surface area.

a. The scaled particle theory. The scaled particle theory was originally developed as a statistical mechanical theory for a liquid of dense hard spheres (34).

The theory is based entirely on geometric considerations and gives an expression for the work required to create a cavity in a hard sphere liquid. Pierotti (35,36) has extended its applications to real liquids including water.

From a consideration of the relationships between the second viral coefficient, the Henry's law constant and the Boltzmann equation, he derived the following expression for the Henry's law constant of a solute in a solvent given by

$$\ln K_h = \frac{W}{RT} + \ln \frac{RT}{V_1} \qquad \text{Eq. 4}$$

where K_h is the Henry's law constant, W is the reversible work required to dissolve 1 mole of solute in an infinite amount of solvent at constant temperature and pressure and V_1 is the molar volume of the solvent. To evaluate the reversible work of solution, the process of introducing a solute molecule into the solvent is viewed as consisting of two steps i) formation of a cavity in the solvent to accomodate the solute molecule and ii) a charging process of the hard sphere to bring it to the required potential of the solute molecule. The reversible work or the partial molar Gibbs free energy change $\overline{G}_c$ for step i) is equivalent to that required to introduce one mole of hard spheres into solution. The free energy change $\overline{G}_i$ for step ii) is the result of dispersion, induction and dipole interactions of the solute molecules with the solvent. The expression for

the Henry's law constant can now be written as

$$\ln K_H = \frac{\overline{G}_c}{RT} + \frac{\overline{G}_i}{RT} + \ln \frac{RT}{V_1} \qquad \text{Eq. 5}$$

The free energy of cavity formation as calculated by the scaled particle theory is

$$\overline{G}_c = -RT \ln(1 - y) + RT\left(\frac{3y}{1 - y}\right)D +$$

$$RT\left[\frac{3y}{1 - y} + \frac{9}{2}\left(\frac{y}{1 - y}\right)^2\right]D^2 + y\frac{P}{\rho}D^3 \qquad \text{Eq. 6}$$

where $y = (\pi\rho\sigma_1^3)/6$ is the reduced number density, $D = \sigma_2/\sigma_1$, σ_1 and σ_2 are the hard sphere diameters of the solvent and solute respectively, ρ is the number density, P is pressure, R is the gas constant and T is temperature. To calculate the free energy of interaction the following approximations were made:

i) spherically symmetric, pairwise additive interactions,

ii) a unit radial distribution independent of temperature (i.e., $\overline{S}_i$, the entropy of interaction is zero),

iii) rotationally averaged terms for the inductive and the dipole-dipole interactions.

The individual contributions to $\overline{G}_i$ then are

$$\bar{G}_{DISP} = - \frac{8\pi\rho}{9\sigma_{12}^3} C_{DISP} \qquad \text{Eq. 7}$$

where C_{DISP} for the Lennard-Jones Potential is $C_{DISP} = C_{LJ} = 4(\varepsilon_1\varepsilon_2)^{1/2}[(\sigma_1 + \sigma_2)/2]^6$.

$$\bar{G}_{IND} = - \frac{4\pi\rho}{3\sigma_{12}^3}(\mu_1^2\alpha_2 + \mu_2^2\alpha_1) \qquad \text{Eq. 8}$$

where $\mu_{1,2}$ are the dipole moments and $\alpha_{1,2}$ are the polarizabilities for the solvent and solute respectively.

$$\bar{G}_{DIP} = \frac{8\pi\rho}{9\sigma_{12}^3}\left(\frac{\mu_1^2\mu_2^2}{kT}\right) \qquad \text{Eq. 9}$$

Using these expressions, Pierotti as well as other investigators (37,38) have obtained fairly good agreement for solutions of inert gases and hydrocarbons in water and some nonpolar solvents and have emphasized the ability of the theory to account for the solute properties in diverse solvents.

b. Molecular surface area approach. A semi-empirical cavity model has more recently been proposed by Hermann (21). Using geometrical parameters of the molecule to calculate molecular surface areas, he found good correlation with aqueous solubilities of hydrocarbons. The surface area also correctly accounted for the solubilities of branched and cyclic compounds and hence appears to be a

better parameter than molar volume for estimating solubility. Amidon et al. (39) found good correlation for a series of monofunctional compounds and calculated the contributions for several functional groups. The surface area method in comparison with other empirical approaches reflecting molecular topology was found to be the best overall parameter in solubility correlations (40).

The cavity model has also been applied to understanding hydrophobic interactions, conformational equilibria, complexation and protein-ligand binding. For example, Sinanoglu (41) has estimated the free energy for the helix coil transformation of DNA molecules in several solvents. His calculations suggest that the energy required to form a cavity in the solvent constitutes a dominant part of the free energy change of the transformation. The relative free energy changes for the different solvents were in good agreement with their denaturing abilities. Surface area considerations have also been used in complex stability (42) and in determinations of the solvent contribution in protein-ligand binding (43).

D. Aqueous Solution Properties of Aliphatic and Aromatic Hydrocarbons

1. Standard State Considerations

Most considerations of the thermodynamics of the solubility process have been based on the pure liquid (or

solid) as the standard state and hence focused their attention on the process

pure liquid (or solid) → aqueous solution

to compare the effects of structural modification such as branching or functional groups on solute-solvent interactions. The free energy for this process, given by

$$\Delta G^{\circ}_{soln} = -RT \ln X_s \qquad \text{Eq. 10}$$

is however a composite quantity and reflects the changes in both solute-solute interactions in the solute phase and solute-solvent interactions in the solution phase. These different contributions to the free energy of solution may be progressively separated as follows.

a. Lattice energy contributions. The solution process for a solid or liquid may be represented as in Scheme I.

Scheme I

$$\text{Solid} \xrightarrow{\Delta G^{\circ}_{s}} \text{Pure Liquid or Supercooled Liquid (SCL)} \longrightarrow \text{Aqueous Solution}$$

The free energy change for the process solid → solution may

then be written as

$$\Delta G_{solid \rightarrow soln} = \Delta G^{\circ}_{s} + \Delta G^{\circ}_{SCL \rightarrow soln} \qquad \text{Eq. 11}$$

ΔG°_{s} is the free energy change for transformation of the solid to its supercooled liquid at the temperature of interest and may be calculated from the expression (neglecting heat capacity effects)

$$\Delta G^{\circ}_{s} = \Delta H^{\circ}_{f} \frac{(T_m - T)}{T_m} \qquad \text{Eq. 12}$$

where ΔH°_{f} is the heat of fusion and T_m is the melting point in °K. It is a measure of the lattice energy of the solid (crystal) and the importance of its consideration in solute-solvent interactions may be illustrated by the following examples.

i) In a study of the effect of chain length on the solubility for a homologous series of compounds, the higher homologs may be solids. The change in solubility of the solid homolog may be a result of the additional lattice energy of the crystal rather than any dramatic change in its behavior towards the solvent. Correction for this lattice energy by ΔG°_{s} then allows comparison of the lower liquid homologs with the higher solid homologs (44).

ii) Anthracene and phenanthrene are isomers but have aqueous solubilities differing by a factor of 100. Their activity coefficients in water however are very similar (45). The difference in solubility is almost entirely due to the difference in their interactions with water and can be accounted for by the above correction.

b. Solute-solute interactions in the pure liquid (supercooled liquid). As early as 1933, Butler (8) noted the additional complexity of solute-solute interactions in the pure liquid phase in interpreting the aqueous solution behavior of normal and branched aliphatic alcohols. In order to separate the solute-solute interactions, he determined the Henry's law constants of their aqueous solutions from measurements of vapor pressure as a function of concentration. The solute-solvent interaction free energy is then given by

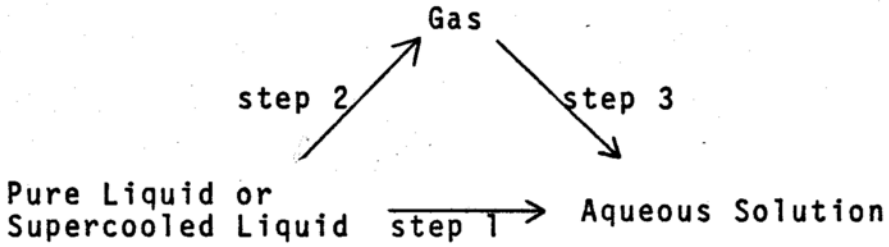
$$\Delta G^{\circ}_{\substack{\text{solvation} \\ \text{(hydration)}}} = RT \ln K_h \qquad \text{Eq. 13}$$

where K_h is the Henry's law constant, and is equivalent to using the gas phase (1 mm or 1 atm vapor pressure) as the standard state. In this manner he obtained $\delta\Delta G^{\circ}_{\text{solvation}}$ values for the methylene group and various other polar functional groups. From this analysis he arrived at the

interesting conclusion that in the case of the methylene group, although the free energy of transfer from pure liquid to aqueous solution was 800 cal per mole, only 160 cal were a result of solute-solvent interaction, the remaining 640 cal being due to interactions in the pure liquid phase.

The separation of the solute-solute and solute-solvent interactions may also be formulated by partitioning the pure liquid to aqueous solution process into two contributing steps to give the thermodynamic cycle shown in Scheme II.

Scheme II



The free energy change for the pure liquid to gas process is a measure of the solute-solute interactions while that for the gas to aqueous solution process measures the solute-solvent interactions. Hine and Mookerjee in a recent publication (46) have suggested that the gas phase to aqueous solution process (step 3) provides a more accurate measure of the hydrophilicity (or 'intrinsic

hydrophilicity') of a compound. However, due to the difficulty in determining Henry's law constants for solutes with exceedingly low vapor pressures, the gas phase standard state has not been widely used.

2. Standard State Considerations and the Hydrophobic Effect

The use of the gas phase standard state in interpreting solute-solvent interactions has important consequences in understanding the nature of the 'hydrophobic effect'. Most interpretations of the hydrophobic effect are based on the large negative entropy of solution with a consequently large positive free energy of solution for some inert gases and particularly the aliphatic hydrocarbons as shown in Table 1. These considerations have led to the picture of the hydrocarbons being 'squeezed' out of water. However when considering the date on a per CH_2 group basis, it is evident from Butler's analysis, that of the 800 cal per CH_2, only 20% is a result of interaction of the methylene group with water, the remaining free energy change being accounted for by the solute-solute interactions. Similar analyses by Mukerjee (47), Amidon *et al.* (39) and Wolfenden (48) have further reinforced the conclusion that the contribution of the solvent may be smaller than was previously assumed.

Recently Cramer (50) has strongly criticized the

Table 1. Thermodynamic Parameters for Transfer of Hydrocarbons From Pure Liquid Hydrocarbon to Water.[a,b]

	ΔG°	ΔH°	ΔS°
C_2H_6	3,900	-2,500	-21
C_3H_8	4,900	-1,700	-22
C_4H_{10}	5,900	-800	-23

[a] Data from ref. 49.

[b] Values of parameters in cal/mole.

concept of the 'hydrophobic interaction' based on similar considerations. On determining the gas phase solubilities (1 atm) of the inert gases, he found that the solubility actually increased with increasing molar volume contrary to expectation. Comparison of graphs of gas phase solubilities in water and octanol vs. molar volume for the inert gases and aliphatic hydrocarbons led him to conclude that the linearity in the graphs was probably a result of compensation in the lipid and water phases. The solvation energies in octanol and water suggested that the liquid phase also played an important role in the transfer process, the free energy of transfer not being solely due to a driving force by water. Further evidence for this view is the behavior of the aromatic hydrocarbons in aqueous solution (51) where the gas phase (1 atm) aqueous solubility of the aromatic hydrocarbons increases with increasing surface area indicating that the aromatic residue has a net favorable interaction per unit area with water and is not intrinsically hydrophobic. The contrasting behavior of the aromatic and aliphatic hydrocarbons are compared in Table 2, and suggests that further analysis of the enthalpic and entropic contribution to the cycle steps would be of considerable value in comparing aliphatic and aromatic hydrocarbons.

The differing behavior of the aliphatic and aromatic hydrocarbons in aqueous solution directs attention to the

Table 2. $\delta\Delta G/\mathring{A}^2$ for the Cycle Steps for the Aliphatic and Aromatic Hydrocarbons.[a,b]

	Aliphatic	Aromatic
pl → aq soln	29	28
pl → gas	23	55
gas → aq soln	6	-27

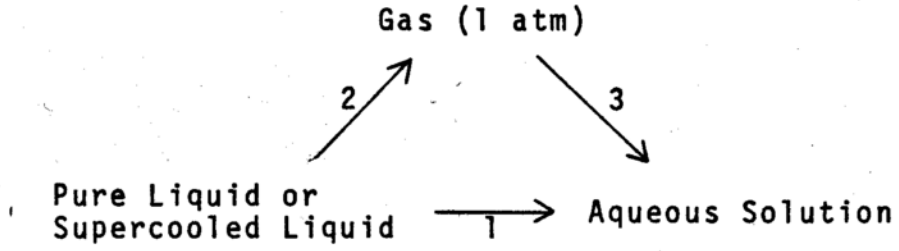
[a] Values in cal/$\mathring{A}^2$.

[b] Data from ref. 51.

aqueous solution properties of the alkyl aromatic hydrocarbons. Although considerable data on the aqueous solubilities of the alkyl benzenes and naphthalenes have been published there has not been any systematic thermodynamic analysis, particularly in light of the above results. The surface area approach and the free energy partitioning scheme would provide a convenient method for analyzing the solution properties of these compounds in terms of group contributions and also allow estimation of the solvent contribution in hydrophobic processes such as the binding of ligand molecules to proteins.

E. Purpose of Research

1. Study the partitioning of the solubility process into solute-solute and solute-solvent interactions using the thermodynamic cycle



with particular emphasis on the aqueous solution behavior of the aliphatic, aromatic, and alkyl aromatic hydrocarbons using the molecular surface approach.

2. Partitioning of the free energy changes of the

cycle steps for the aliphatic and aromatic hydrocarbons into enthalpic and entropic contributions to further elucidate the nature of their differing behavior in aqueous solutions.

3. a. Developing a general model using the surface area approach and regression analysis for estimating the free energy changes of the 3 steps for the alkyl aromatics.

b. Study the effects of alkyl substitution on the solvation free energy.

c. Differentiation between the contributions of the various surface area group terms to the free energy change for the 3 steps.

d. Determine the effect of the choice of solvent radius used in the surface area calculations on parts (a) and (c).

4. Consider the applicability of the scaled particle theory to the aromatic and alkyl aromatic hydrocarbons.

5. Determint the contribution of the solvent to the free energy of protein-ligand interactions, in particular in the binding of benzene, naphthalene and anthracene to the α-chymotrypsin active site.

II. METHODS

A. Experimental

1. Materials

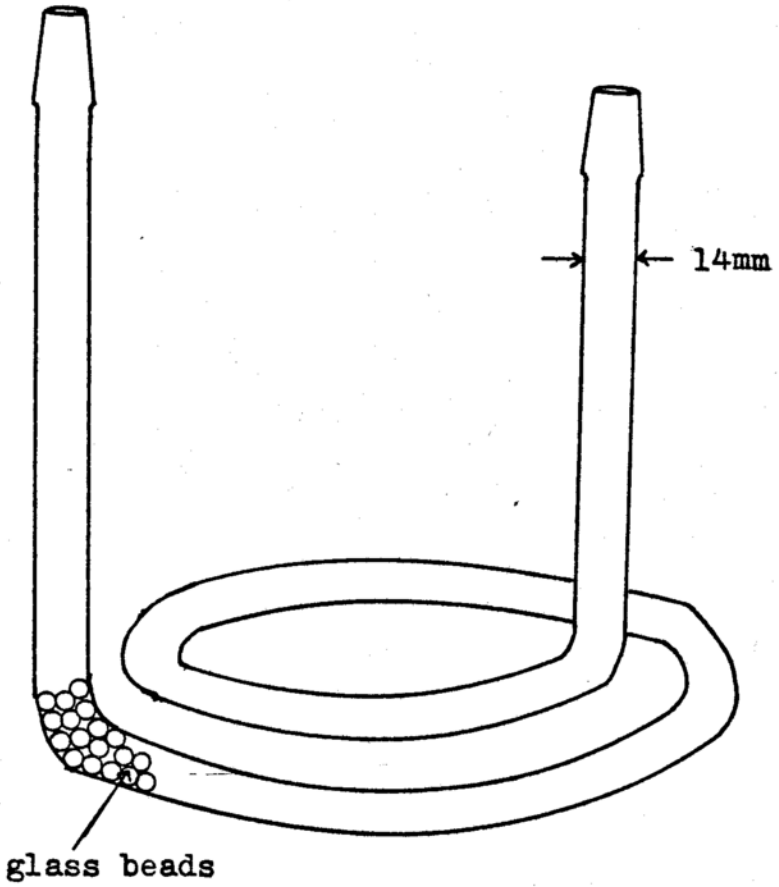
1,4 dimethyl naphthalene (99%), 2,3 dimethyl naphthalene (99%), 2,6 dimethyl naphthalene (99%) and 9,10 dimethyl anthracene (99%) were obtained from Aldrich Chemical Co. These were used without further purification.

Ethanol (95%) (Commercial Solvents Inc.) was distilled before use. All water used was double distilled with final distillation from alkaline permanganate.

2. Apparatus and Methods

a. Vapor pressure measurement. The vapor pressures for both solids and liquids were determined using the gas saturation method, the final analysis being done spectrophotometrically. The gas saturation method and apparatus used is described in detail in previous work (51). For the liquid 1,4 dimethyl naphthalene the column used in the gas saturation apparatus is shown in Figure 1. The columns were packed with glass beads and contained from 5-10 grams for the solids and about 2 grams for the liquid. At least 5 readings were obtained for each compound.

b. Solubility determinations. 9,10 dimethyl anthracene. The solid was equilibrated in 15 ml ampoules



Total length of column = 150cm

Fig.1. Saturation column for vapor pressure measurements of liquids.

containing from 8 to 10 ml of water by end on end rotation in a water bath maintained at 25°C. The water in the ampoule was deaerated by a stream of nitrogen before sealing. The solutions were sampled at different periods of time to check if equilibration was complete. Eighteen days were required to achieve equilibrium. The number of days may be a function of the container and the method of equilibration. Samples were obtained by filtration through a glass wool plug on a pipette tip and were assayed by fluorimetric analysis.

1,4 dimethyl naphthalene. This is a liquid of density approximately 1 and poses greater difficulties in separating the aqueous phase from the hydrocarbon phase in order to obtain a sample for assay purposes. Initially a glass wool plug was tried but this did not prevent fine droplets from entering the pipette giving erratic absorbance values. To overcome this the ampoules were centrifuged at 6,000 rpm for 10 minutes in a temperature controlled Beckman centrifuge. Ten minutes was sufficient time to separate the 2 phases although a fine film still remained at the surface. Higher centrifugation times or speeds resulted in broken ampoules. After centrifugation 5 ml of solution was withdrawn from the center of the ampoule by means of a syringe taking care not to touch the sides or the bottom of the ampoule. The 5 ml was immediately diluted to 10 ml by rinsing the syringe with 95%

ethanol into a volumetric flask. The solution was then assayed by measuring the absorbance at 288.5 nm. The samples reached a constant value after 10 days of equilibration as described above.

c. Spectral Methods. U.V. 2,3 Dimethyl naphthalene, 2,6 dimethyl naphthalene, and 1,4 dimethyl naphthalene were measured by U.V. absorption on the Cary 16 spectrophotometer. Standard solutions were prepared in 95% ethanol. Linear graphs were obtained giving molar absorptivities of 4.968×10^3 (λ_{max} = 278), 4.677×10^3 (λ_{max} = 273), and 6.839×10^3 (λ_{max} = 288.5) respectively. In the case of 1,4 dimethyl naphthalene a standard curve for solutions in 47.5% ethanol was also prepared for use in the solubility measurement and gave a molar absorptivity of 6.532×10^3 (λ_{max} = 288.5).

Fluorimetry. 9,10 Dimethyl anthracene was assayed by fluorimetric analysis on a Perkin Elmer MP-4 spectrophotofluorimeter. The difficulty that arises is the preparation of the standard curve for aqueous solutions. However Schwarz (52) has shown that the quantum yield of these polycyclic aromatics is about the same in both water and ethanol provided the solutions are purged with nitrogen to remove all dissolved oxygen before taking the reading. Hence a standard curve was prepared using 95% ethanol solutions of dimethyl anthracene. The λ_{ex} is 259 nm and λ_{em} = 402 nm. A linear graph was obtained for fluorescence peak

height vs. concentration. A solution of concentration = 1×10^{-7}% w/v (.001 µg/ml) gave 20% of full scale deflection at 402 nm with coarse sensitivity setting at 3 and fine sensitivity at 5 and both slits open at 10 nm.

d. Heats of fusion. Heats of fusion of 2,3 dimethyl naphthalene, 2,6 dimethyl naphthalene and 9,10 dimethyl anthracene were measured on a Mettler TA2000 differential scanning calorimeter which was previously calibrated with an Indium standard. Typically the sample size was between 5-15 mg. Peak areas were measured using a planimeter and at least three readings were obtained for each compound. The heat of fusion may be calculated by the following expression:

$$\frac{K \times R \times AUC \times MW}{E_{rel} \times V \times w} \qquad \text{Eq. 14}$$

where K is a constant which accounts for the conversion factors, R is the range in µV, AUC is the area under the peak in planimeter units, MW is the molecular weight, E_{rel} is the relative calorimetric sensitivity in µV sec/mcal, V is the chart speed and w is the weight of the compound in mgms.

B. Data and Statistical Analysis

1. Molecular Surface Area Calculation

The molecular surface area is calculated as the sum of the surface area of each atom accessible to the solvent. The accessible surface area of an atom is defined as the area on the surface of a sphere of radius R on each point of which the center of a solvent molecule can be placed in contact with the atom without penetrating any other atoms of the molecule. The radius R is comprised of the Van der Waals radius of the atom and the solvent radius. In this study the Van der Waals radii used were: aliphatic carbon 1.6 Å, aliphatic hydrogen 1.2 Å, aromatic carbon 1.7 Å. Surface areas are calculated using a solvent radius of 1.5 Å for water unless otherwise specified. Figure 2 shows a cross section of a hydrocarbon molecule and the smoothed area generated by the solvent molecule.

Surface areas were calculated using a modified version of the program written by Hermann (21)*. The calculation involves the following steps. First the coordinates of the atoms of the molecule (with respect to each other) are calculated using standard values for bond lengths (53), bond angles (53) and twist angles, the latter being based on an all trans conformation wherever possible. The coordinates together with the Van der Waals radius for each atom then

*The program was modified by Dr. Robert Pearlman.

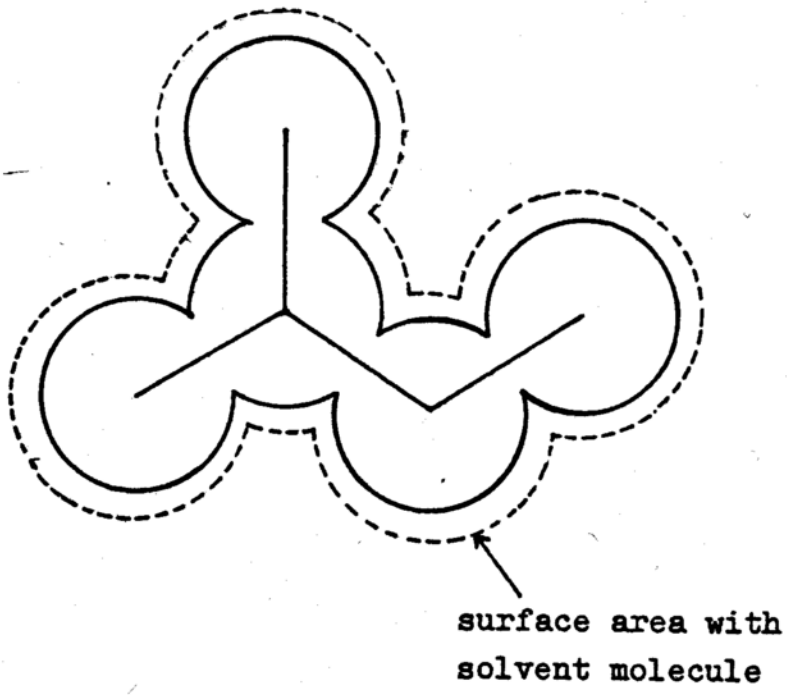


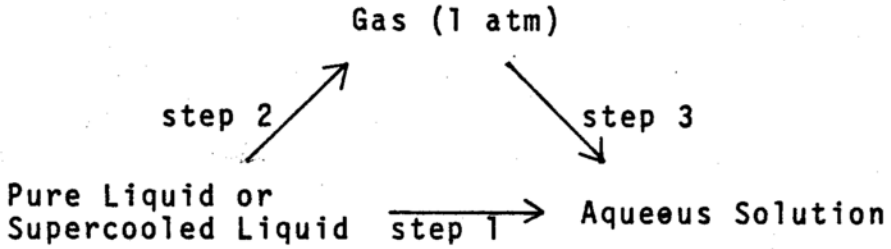
Fig.2. Cross section of hydrocarbon molecule.

form the input data for the surface area calculation. The radius to be used for the solvent molecule is also specified. The accessible surface area of each atom is calculated individually and this allows separation of the atomic or group contributions of a particular functional group to the total surface area of the molecule.

2. Free Energies and Group Surface Areas for the Aromatic, Alkyl Aromatic and Aliphatic Hydrocarbons

The free energy of solution is partitioned into two contributing steps, the free energy of vaporization and the free energy of solvation, and is represented by the thermodynamic cycle shown in Scheme III.

Scheme III



The free energy changes for these 3 steps are given by the expressions

$$\Delta G^\circ = -RT \ln X_s \text{ (step 1)} \qquad \text{Eq. 15}$$

$$\Delta G^\circ = -RT \ln P^\circ \quad \text{(step 2)} \qquad \text{Eq. 16}$$

$$\Delta G^\circ = -RT \ln \frac{X_s}{P^\circ} = RT \ln K_h \quad \text{(step 3)} \qquad \text{Eq. 17}$$

where X_s is the mole fraction solubility, P° is the vapor pressure of the pure liquid solute in atmospheres and K_h is the Henry's Law constant. In the case of solids, the supercooled liquid is used as the standard state. The expression for step 3 is valid provided the aqueous solutions of the solute follows Henry's Law up to the solubility limit. This assumption was found to be valid for aqueous solutions of benzene and naphthalene (51) and is assumed here for the alkyl aromatic hydrocarbons.

The free energies of solution for the polycyclic and alkyl aromatics were determined using average solubilities from those reported in the literature. For the alkanes, free energies of solution were calculated from solubilities reported by McAuliffe (20). The solubility data, average solubilities and standard errors as well as the free energy changes associated with individual solubility measurements are listed in Appendix I. The average free energy change calculated from the average solubility value and from averaging the individual free energy changes are almost identical. Free energies of vaporization were obtained from standard thermochemical tables (54) or from vapor pressure measurements (listed in Appendix A) and the free

energies of solvation were determined by difference.

From previous results of the analysis of aqueous solubilities using the molecular surface area approach (39) we can relate the free energy change with different surface area group contributions by the general equation

$$\Delta G = \theta_0 + \sum_{i=1}^{n} \theta_i (GSA)_i \qquad \text{Eq. 18}$$

where the θ_i's are parameters with units of cal/Å^2 and $(GSA)_i$ is the surface area of the i'th kind of group or atoms. For the alkyl aromatics, the total surface area (TSA) may be partitioned into an aromatic surface area, which is the sum of the surface areas of the aromatic carbons and hydrogens, and an aliphatic surface area, which may be divided into an aliphatic overlap area and a remaining aliphatic area. The overlap area is defined as the area of the aliphatic carbon atom with its attached hydrogens which is adjacent to the aromatic nucleus. The different group and atomic surface areas are shown in Figure 3.

Table 3 lists the free energy data and the surface area contributions for all the compounds used in the regression analysis.

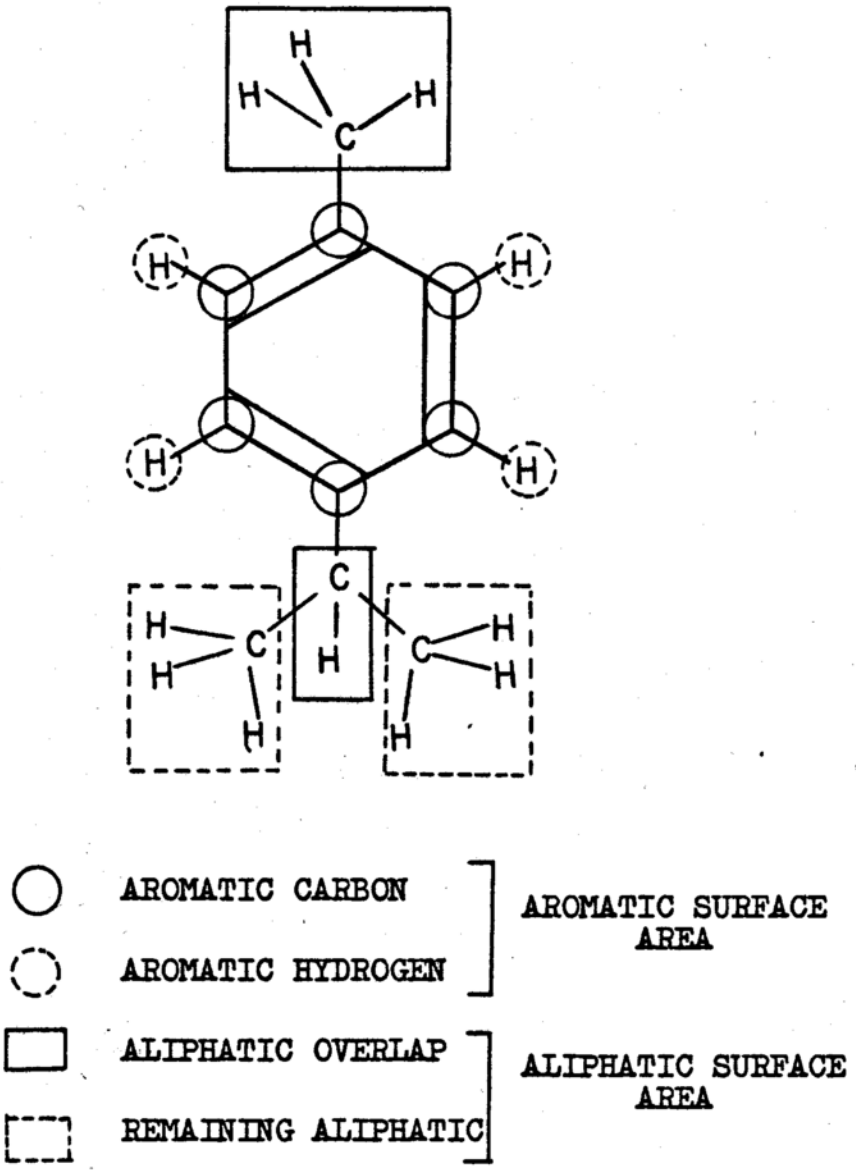


Fig.3. Partitioning of total surface area into group surface areas.

sec-Butyl benzene	59.4	123.0	182.4	2.0	183.0	185.0	367.4	7.48	3.58	3.90
t-Butyl benzene	57.3	110.4	167.7	0.0	185.0	185.0	352.7	7.32	3.48	3.84
1-Methyl,4-iso-propyl benzene	46.9	79.4	126.3	100.0	143.0	243.0	369.3	7.28	3.68	3.60
Naphthalene	120.2	203.1	323.3	0.0	0.0	0.0	323.3	6.55	4.68	1.87
Anthracene	154.9	236.0	390.9	0.0	0.0	0.0	390.9	8.34	8.63	-0.29
Phenanthrene	155.1	229.2	384.3	0.0	0.0	0.0	384.3	8.51	8.15	0.36
Pyrene	165.5	236.0	401.5	0.0	0.0	0.0	401.5	9.55	9.88	-0.33
Butane	0.0	0.0	0.0	0.0	255.0	255.0	255.0	5.90	-0.52	6.42
Pentane	0.0	0.0	0.0	0.0	287.0	287.0	287.0	6.80	0.26	6.54
Hexane	0.0	0.0	0.0	0.0	319.0	319.0	319.0	7.80	0.98	6.82
Heptane	0.0	0.0	0.0	0.0	351.0	351.0	351.0	8.60	1.67	6.93
Octane	0.0	0.0	0.0	0.0	383.0	383.0	383.0	9.50	2.37	7.13

[a]For solids, the free energy changes refer to the supercooled liquid.

3. Regression Analysis and Criteria for Selecting a Model

Multiple linear regression analysis was used to determine the best fit of the data for a given model. The regression analyses were performed on a Univac 1110 computer using the regression routine 'REGAN 3' of the statistical package 'STATJOB' at the University of Wisconsin.

In developing a model, a method essentially similar to the stepwise regression procedure (55) was followed. The criteria used in judging the adequacy of the model were the following:

1) Multiple correlation coefficient
2) Standard error of regression
3) Examination of residual plots
4) Number of parameters in the model
5) Partial correlation coefficients
6) Partial F values of parameters
7) Standard errors of parameters
8) Correlation between variables
9) Physical interpretation of the regression coefficients for the given data.

These criteria provide a general basis for the selection of the "best" model. However the selection of a particular model as "best" is ultimately dependent on the relative importance assigned to the above criteria as well as the more subjective criterion of physical acceptability of the

model with respect to the process being modelled.

A high multiple correlation coefficient, a low standard error of regression and a randomized residual plot are of primary concern in any regression analysis since they provide a measure of how well the model as a whole, is able to account for the given data. Although the multiple correlation coefficient can be made to approach the value of 1 by increasing the number of variables and parameters in the regression equation, the gain in the multiple correlation coefficient with each additional variable quickly decreases. The partial correlation coefficient and the partial F values both indicate the extent to which the new variable accounts for the residual sum of squares. Furthermore the parameters of the equation become less meaningful due to the presence of correlations amongst the variables. This is particularly important when the model is being developed not only for predictive purposes but also to understand the behavior of the variable with respect to the physical process involved. Here the standard errors of the parameters and the correlation matrix for the variables assume important roles in determining the inclusion of a particular parameter in the model.

The correlation matrix is initially used to choose the independent variable most highly correlated with the dependent variable. The matrix however also lists the correlation coefficients between the independent variables. As is

often the case, 2 or more variables which show good correlation with the dependent variable are also highly correlated with each other. In the event that the correlation coefficient between the two independent variables in the model is 1 the coefficients of these variables in the regression equation can assume an infinite number of combinations which will give the same regression statistics. The coefficients then lose significance and are not amenable to interpretation. Hence in developing a model particularly for interpretive purposes it is important to have a minimum amount of correlation between the independent variables. The correlation amongst the variables in the regression equation is also reflected by the magnitudes of the standard errors of the regression coefficients, which will be large, and the coefficients may not be significantly different from zero at the 95% confidence level when there is high correlation between the variables. Finally the choice of a regression model is governed by how meaningful the model is when related to the physical process. This decision is subjective and is based upon an understanding of the process and the credibility of the explanation suggested by the model.

III. RESULTS AND DISCUSSION

A. Free Energy, Enthalpic and Entropic Contributions to the Cycle Steps for the Aliphatic and Aromatic Hydrocarbons

1. Free Energy Results

In a previous study (51) on the free energy of solution of the aliphatic and aromatic hydrocarbons, it was observed that the free energy change per $\mathring{A}^2$ for the pure liquid to aqueous solution process for the aliphatic and aromatic hydrocarbons was nearly identical ($\sim$30 cal per $\mathring{A}^2$) suggesting similar 'hydrophobicities' for these two classes of compounds. However, the partitioning of this process into the two contributing steps resulted in a distinctly different view. The results of this study are summarized in Figures 4 and 5. The most contrasting feature apparent in Figures 4 and 5 is the slope of the lines for the gas phase to aqueous solution process. For the aliphatic hydrocarbons the slope is small and positive (6 cal per $\mathring{A}^2$) however for the aromatic hydrocarbons it is large and negative (-27 cal per $\mathring{A}^2$) suggesting a net favorable interaction of the aromatic residue with water. This large negative $\delta\Delta G°$ is however compensated by a large positive $\delta\Delta G°$ for the pure liquid to gas process such that the $\delta\Delta G°$ for the overall process is similar to that for the aliphatic hydrocarbons.

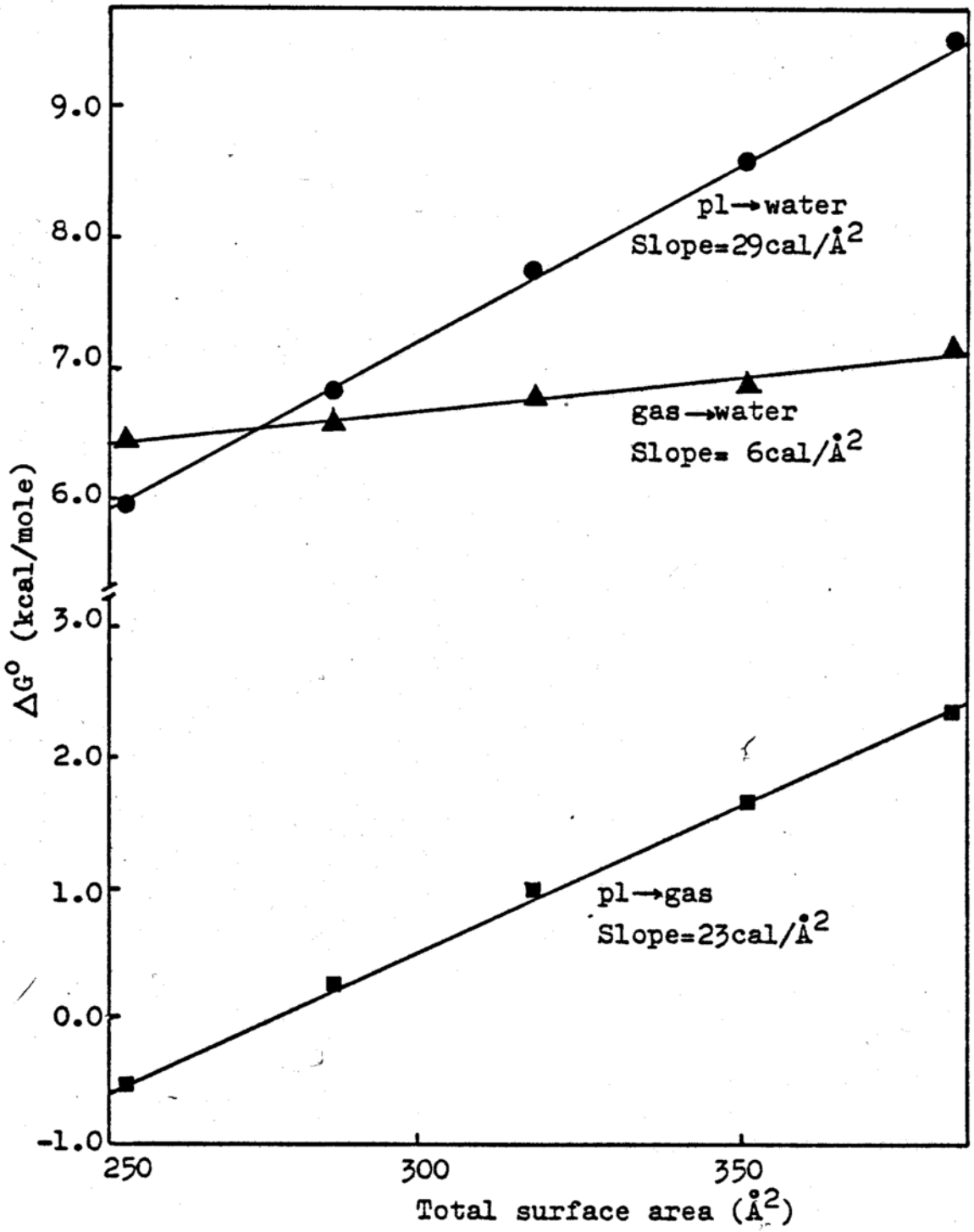


Fig.4. Free energy vs surface area for the cycle steps. Aliphatic hydrocarbons.

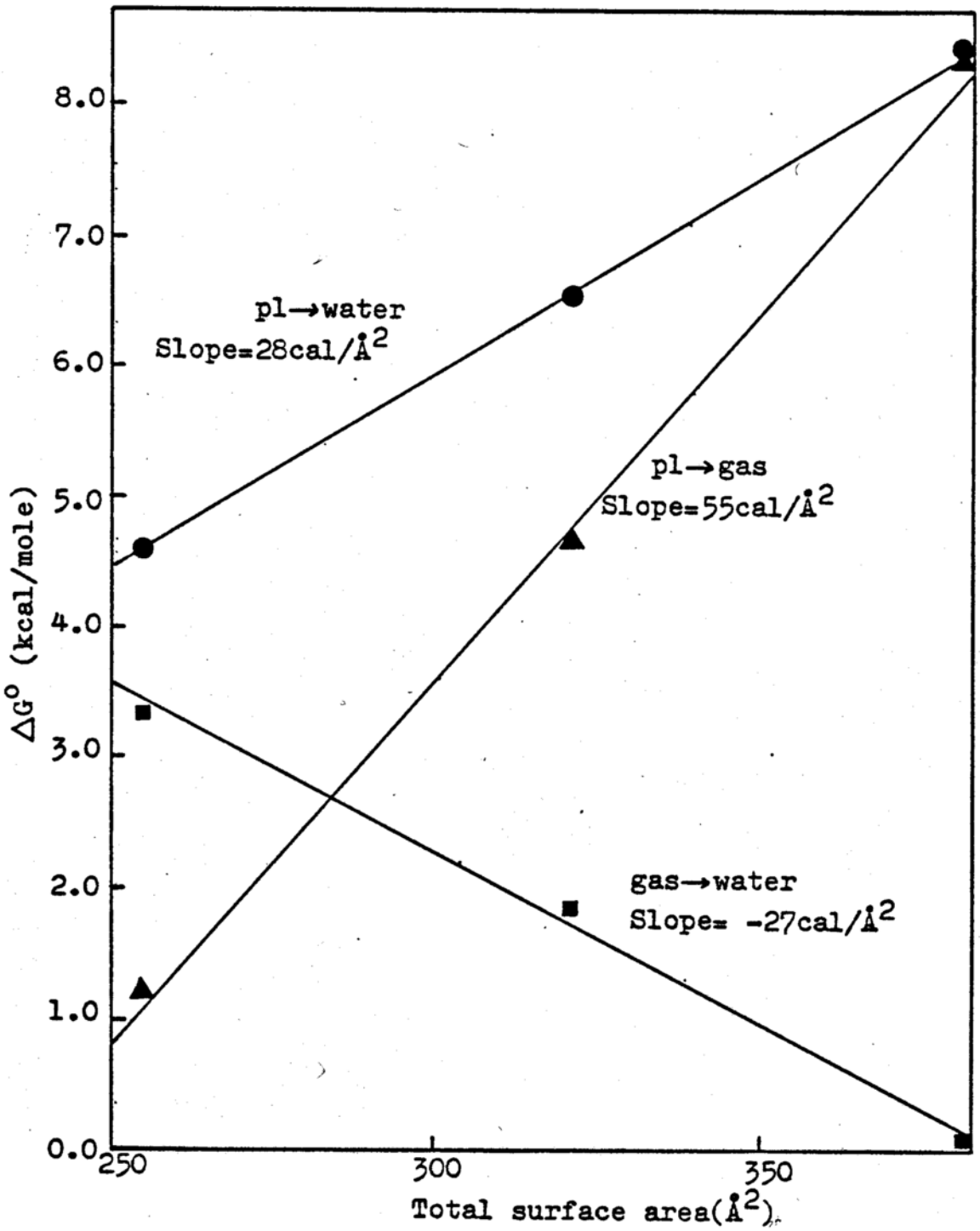


Fig.5. Free energy vs surface area for the cycle steps. Aromatic hydrocarbons.

2. Enthalpy and Entropy Results

The previous partitioning analysis was restricted to the free energy change. Although the data is more limited, this partitioning analysis can be applied to the enthalpic and entropic contributions. The enthalpic and entropic contributions for steps 1 and 2 for the alcohols and alkanes, and the aromatic hydrocarbons are listed in Tables 4 and 5 respectively. The values for step 3 are obtained by difference.

In determining the change in $\Delta H°$ of solution per $\mathring{A}^2$ for the aliphatic hydrocarbons, the data for the alcohols and alkanes were combined. However a graph of the heat of solution vs. TSA in Figure 6 exhibits considerable curvature for the smaller chain length alcohols and calculations based on this data would give a much smaller enthalpic contribution per CH_2 group to the solution process than the composite data suggests. Consequently the data for the C_4-C_6 alcohols only were combined with the C_3-C_6 hydrocarbons and regressed vs. total surface area using an indicator variable for the alcohols (to account for the effect of the OH group) to estimate the $\delta\Delta H°$ per $\mathring{A}^2$ contribution. The regression equation obtained is

$$\Delta H° = -5.0 + .0157TSA + 1.6IOH \qquad \text{Eq. 19}$$

$$r = .98,\ s = .02,\ n = 7$$

Table 4. Enthalpic and Entropic Contributions for the Cycle Steps for Normal Aliphatic Alcohols and Hydrocarbons.

	Process[a,b]					
	pl → aq. soln.			pl → gas		
Compound	ΔH°	TΔS°	Ref.	ΔH°	TΔS°	Ref.
Aliphatic						
Ethanol	-2.43	-3.19	59	10.08	8.53	64
n-Propanol	-2.42	-4.00	59	11.24	9.13	64
n-Butanol	-2.25	-4.65	59	12.53	9.73	64
n-Pentanol	-1.87	-5.09	59	13.60	10.20	64
n-Hexanol	-1.44	-5.47	60	14.80	10.8	64
Ethane	-2.5	-6.4	61, 62			
Propane	-1.7	-6.6	61, 62			
n-Butane	-0.8	-6.7	61, 62	5.10	5.62	54
n-Pentane	-0.5	-7.3	63	6.4	6.14	54
n-Hexane	0.0	-7.8	63	7.55	6.57	54

[a] Values in kcal/mole. TΔS° was calculated from ΔG° and ΔH°.

[b] $\Delta G°_{soln}$ for C_2-C_5 alcohols, ref. 49, p. 20; for hexanol, ref. 22; for C_2-C_4 hydrocarbons, ref. 61, 62; for C_5-C_6 hydrocarbons, ref. 63.

Table 5. Enthalpic and Entropic Contributions for the Cycle Steps for the Aromatic Hydrocarbons

	Process[a,b]					
	pl → aq. soln.			pl → gas		
Compound	ΔH°	TΔS°	Ref.	ΔH°	TΔS°	Ref.
Benzene	0.5, 0.6	-4.0, -4.1	65, 66	8.10	6.87	54
Naphthalene	2.72, 2.80	-3.82, -3.74	66, 67, 68	12.89	8.21	54
Anthracene	4.79, 3.60	-3.55, -4.74	67, 68, 69	17.40	8.77	54
Phenanthrene	6.52, 4.21	-1.99, -4.3	67, 68, 69	17.24	9.09	54
Pyrene	7.44	-2.11	67, 68	18.84	54	

[a] Values in kcal/mole. TΔS was calculated from ΔG° and ΔH°. ΔG° and ΔH° for solid aromatics refer to supercooled liquid.

[b] ΔG°_{soln} from refs. 66, 67, 17, 70.

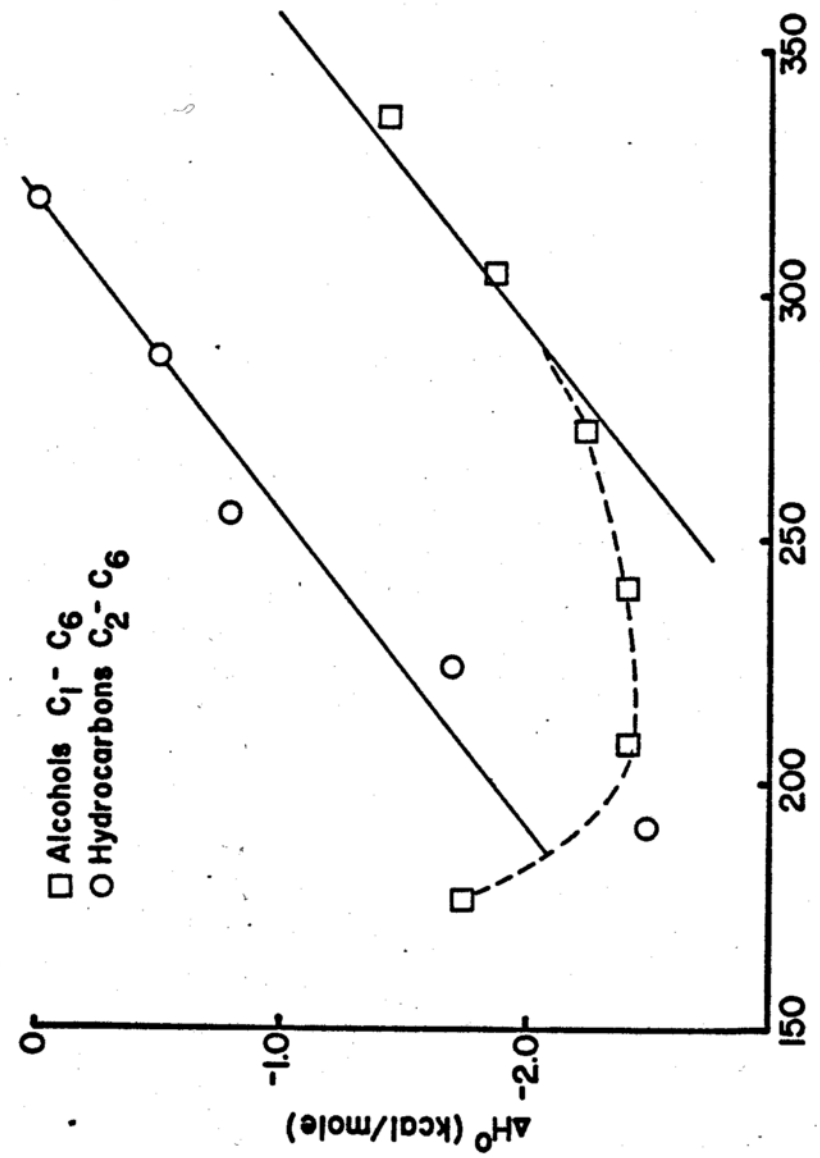


Fig.6. Heat of solution vs surface area for the aliphatic hydrocarbons and alcohols.

The slope of 15.7 cal per $\mathring{A}^2$ is equivalent to 500 cal per CH_2 group (CH_2 = 31.8 $\mathring{A}^2$) and this value is in good agreement with those reported by Nelson and deLigny (56) from alkane data and by Corkill et al. (57) for the CH_2 group in α,ω diols, but higher than the value obtained by Aveyard and Mitchell (58) from alcohol data.

The determination of $\delta\Delta H^\circ$ per $\mathring{A}^2$ for the aromatic hydrocarbons poses more difficulty due to the considerable scatter in the experimental data as seen in Table 6 and Figure 7. Some of the data however are not as reliable as the others. Thus the values reported by Schwarz (Ref. 4 in table) are subject to some uncertainty due to questionable experimental procedures. For example he used molar absorptivities determined in 100% ethanol to calculate solubilities from absorbance values of solutions containing only 20% ethanol. Furthermore he used this single value as a one point calibration for his fluorescence measurements which were used for the solubility determinations at all other temperatures. The equilibration time of 24 hours is small and although he asserts that his solubility and heats of solution values are in excellent agreement with reported ones this does not appear to be the case [for example in the case of naphthalene (66, 67)].

However in order to get an unbiased estimate of $\delta\Delta H^\circ$ per $\mathring{A}^2$, several estimates of $\delta\Delta H^\circ$ per $\mathring{A}^2$ were obtained by linear regression of ΔH°_{soln} vs. TSA using different

Table 6. Enthalpy of Solution for the Aromatic Hydrocarbons.[a]

Compound	Reference 1	2	3	4	5
Benzene	0.5	0.6	- -	- -	- -
Naphthalene	- -	2.72	2.80	.98	- -
Anthracene	- -	- -	4.79	1.42	3.60
Phenanthrene	- -	- -	6.52	4.22	4.21
Pyrene	- -	- -	7.44	7.74	- -

[a]Values in kcal/mole. ΔH°_{soln} refers to the supercooled liquid.

[1]Ref. 65.

[2]Ref. 66.

[3]Ref. 67.

[4]Ref. 68

[5]Ref. 69

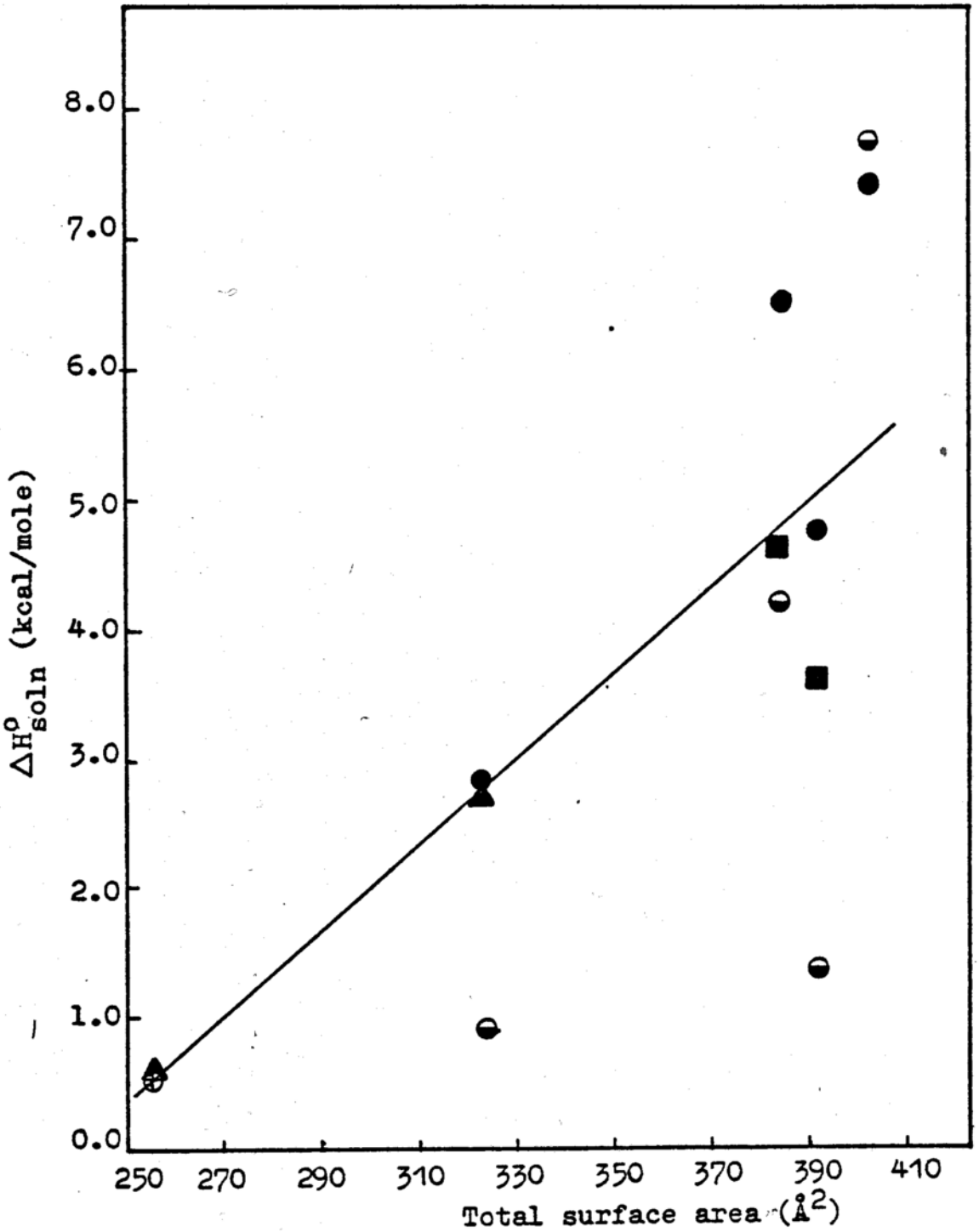


Fig.7. Heat of solution vs surface area for the aromatic hydrocarbons. (⊕)ref.65, (▲)ref.66,(●)ref.67 (◒)ref.68, (■)ref.69.

combinations of the data sets. The results are summarized in Table 7. The inclusion or exclusion of Schwarz's data does not greatly affect the value of $\delta\Delta H°$ per $\mathring{A}^2$ and for the above reasons his data were not included in the subsequent analysis. The $\delta\Delta H°$ per $\mathring{A}^2$ was calculated using the data for benzene, naphthalene, anthracene and phenanthrene. Pyrene shows considerable deviation from this line (Fig. 7) and either the value for pyrene does not fit the surface area model or is in error. However the trends in the $\delta\Delta H°$ and $\delta T\Delta S°$ values are not significantly changed so as to affect the interpretation of the results discussed later.

It should be noted that for the solids, the heats of solution and vaporization data are corrected to the supercooled liquid using heats of fusion at the melting point due to unavailability of ΔH_f° at 25°C. This correction does not affect the value of $\delta\Delta H°$ per $\mathring{A}^2$ for step 3.

Discussion. The partitioning of the free energy data into enthalpic and entropic components is generally done for the purpose of invoking a mechanistic interpretation at the molecular level and this has been particularly true in the case of the hydrophobic effect. This effect is widely associated with the large negative entropy of solution and the structuring of water on introduction of a nonpolar solute into aqueous solution. In this analysis however we are concerned mainly with the differences in behavior of

Table 7. Comparison of $\delta\Delta H^{\circ}_{soln}$ Using Different Data Sets for ΔH°_{soln} for the Cycle Steps.[a]

	All Data		Excluding Schwarz's Data		Excluding Pyrene and Schwarz's Data	
	ΔH°	r	ΔH°	r	ΔH°	r
Step 1	.036	.79	.037	.89	.032	.90
Step 2	.072	.99	.072	.99	.070	.99
Step 3	-.036	.81	-.035	.91	-.038	.94

[a]Data in Table 6. Values in cal/mole/Å^2. r = multiple correlation coefficient.

the aliphatic and aromatic hydrocarbons and will restrict ourselves to a thermodynamic comparison in evaluating the results.

Table 8 summarizes the $\delta\Delta G°$, $\delta\Delta H°$ and $\delta T\Delta S°$ values for the three steps for the aliphatic and aromatic hydrocarbons. We consider first the aliphatic hydrocarbons. As with the analysis for the free energy changes, we can compare the relative magnitudes of the enthalpic and entropic contributions of steps 2 and 3 to step 1. Thus the negative entropic contribution for step 1 is a consequence of the large negative entropy change for the solvation process. The latter however is compensated to a considerable extent by the negative enthalpy of solvation term so that the contribution of the $\delta\Delta G°$ of solvation to the pl → aq. soln. process is small. Thus the $\delta\Delta G°$ of solution is largely composed of the $\delta\Delta G°$ of vaporization which is a direct result of the $\delta\Delta H°$ of vaporization and reflects the non-bonded interactions in the solute phase.

Considering now the aromatic hydrocarbons, we observe that for the pl → aq. soln. process the $\delta T\Delta S°$ term is actually positive so that the positive $\delta\Delta G°$ for this process is entirely due to the enthalpic contribution. Thus while for the aliphatic hydrocarbons the pl → aq. soln. process is characterized by a significant negative entropic contribution, the solubility of the aromatic hydrocarbons is enthalpically controlled. The gas phase to aq. soln.

Table 8. Comparison of the Thermodynamic Parameters for Aliphatic and Aromatic Hydrocarbons.

Process	$\delta(\Delta G°)/\mathring{A}^2$[a]	$\delta(\Delta H°)/\mathring{A}^2$[a]	$\delta(T\Delta S°)/\mathring{A}^2$[a]
Aliphatic			
pl → aq. soln.	29	16	-13
pl → gas	23	37	14
gas → aq. soln.	6	-21	-27
Aromatic[b]			
pl → aq. soln.	28	32	4
pl → gas	55	70	15
gas → aq. soln.	-27	-38	-11

[a] cal/mole/$\mathring{A}^2$.

[b] Analysis refers to benzene, naphthalene, anthracene and phenanthrene (pl refers to pure supercooled liquid except for benzene).

process has a much smaller negative entropy change and is dominated by the large negative $\delta\Delta H°$ resulting in an overall negative $\delta\Delta G°$ and a favorable interaction with water. Hence the decrease in solubility with increasing surface area for the aromatic hydrocarbons is a consequence of the increasing nonbonded interactions in the solute phase, as evidenced by the large $\delta\Delta H°$ of vaporization, and not due to a phobia towards water.

From the above analysis it is apparent that the only similarity in the interactions of the aliphatic and aromatic hydrocarbons with water is the free energy change for the pure liquid to aqueous solution process. The enthalpic and entropic changes for this process are quite different, and the mechanistic interpretation associated with the hydrophobic effect from the data for the alkanes may not extend to aromatic residues. Partitioning of these enthalpic and entropic changes in a thermodynamic cycle further emphasizes the differing behavior of these hydrocarbons in aqueous solution.

B. Analysis of the Hydrophobicity of Alkyl Aromatic Hydrocarbons

1. Analysis of Data

The first step in the analysis of the solution properties of the alkyl aromatics is to compare their behavior with the aliphatic and aromatic hydrocarbons. Figures 8, 9

and 10 present the free energy change with total surface area of these three classes of compounds for the three steps of the thermodynamic cycle and Table 9 lists the regression statistics of the free energy change vs. TSA for the alkanes and aromatics. For the alkyl aromatics the line is drawn using only the n-alkyl substituted benzenes. Several interesting observations may be noted. i) The free energy change for the pure liquid to aqueous solution step shows good correlation with the TSA of the molecule for all three classes of compounds and the alkyl aromatics and the polycyclic aromatics lie on essentially the same straight line with a slope similar to that of the alkanes; ii) the correlation between the free energy change and TSA for the alkyl aromatics is much poorer for the pure liquid to gas and the gas to solution cycle steps; iii) this suggests that the linearity observed for step 1 is a result of considerable free energy compensation; and, iv) as one might expect the slopes of the lines for the n-alkyl substituted benzenes are similar to those for the n-alkanes indicating the independence of the CH_2 group at greater distances from the aromatic nucleus.

Perhaps the most striking observation however, is the 'hydrophobic' nature of the methyl groups attached to the aromatic nucleus which is quite different from that commonly observed for methyl groups. From Figure 10 it can be noted that all the polymethyl benzenes fall below the line

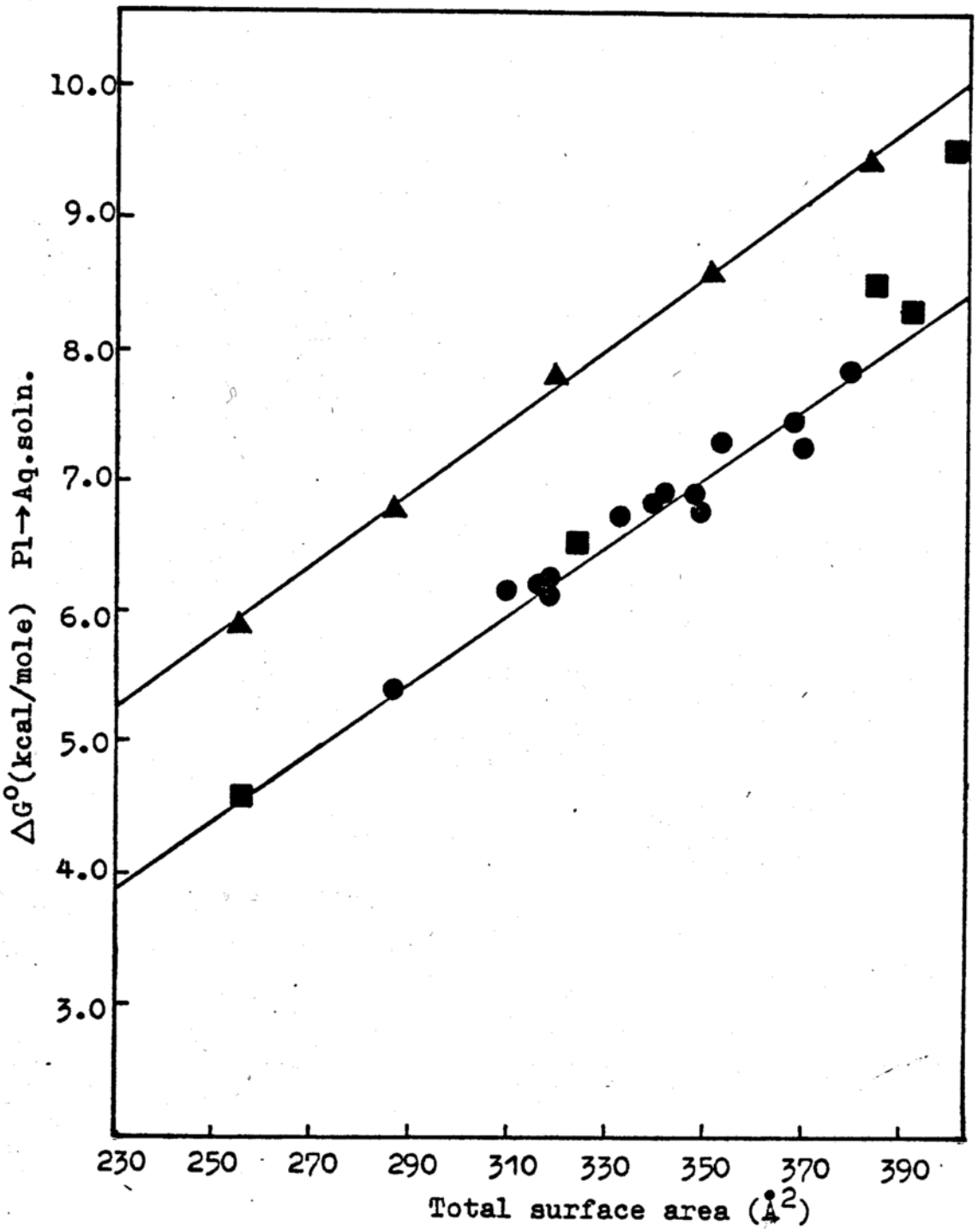


Fig.8. Free energy of solution vs total surface area for Aliphatic(▲), Aromatic (■), and Alkyl Aromatic(●) hydrocarbons.

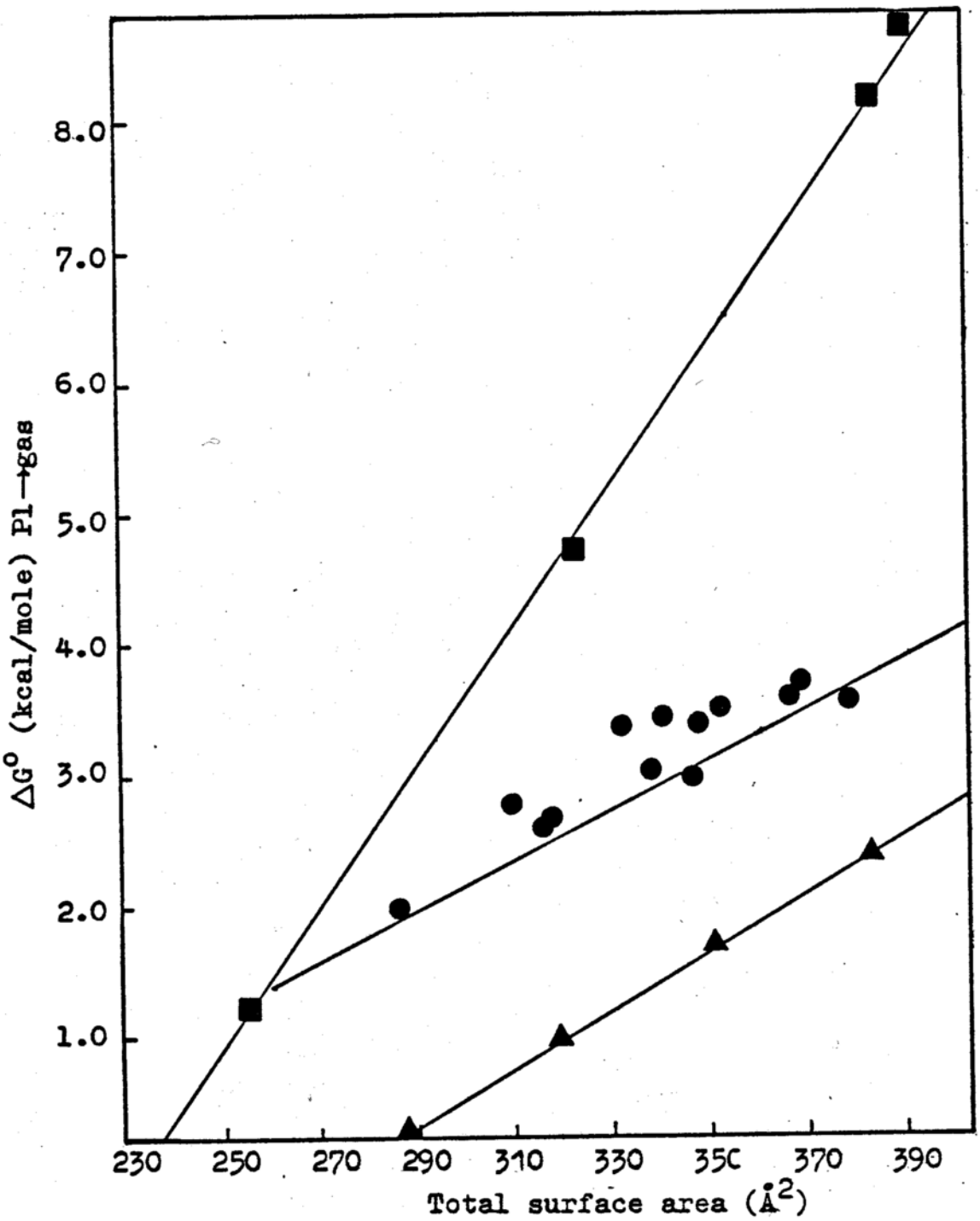


Fig.9. Free energy of vaporisation vs total surface area for Aliphatic(▲), Aromatic(■), and Alkyl Aromatic(●) hydrocarbons .

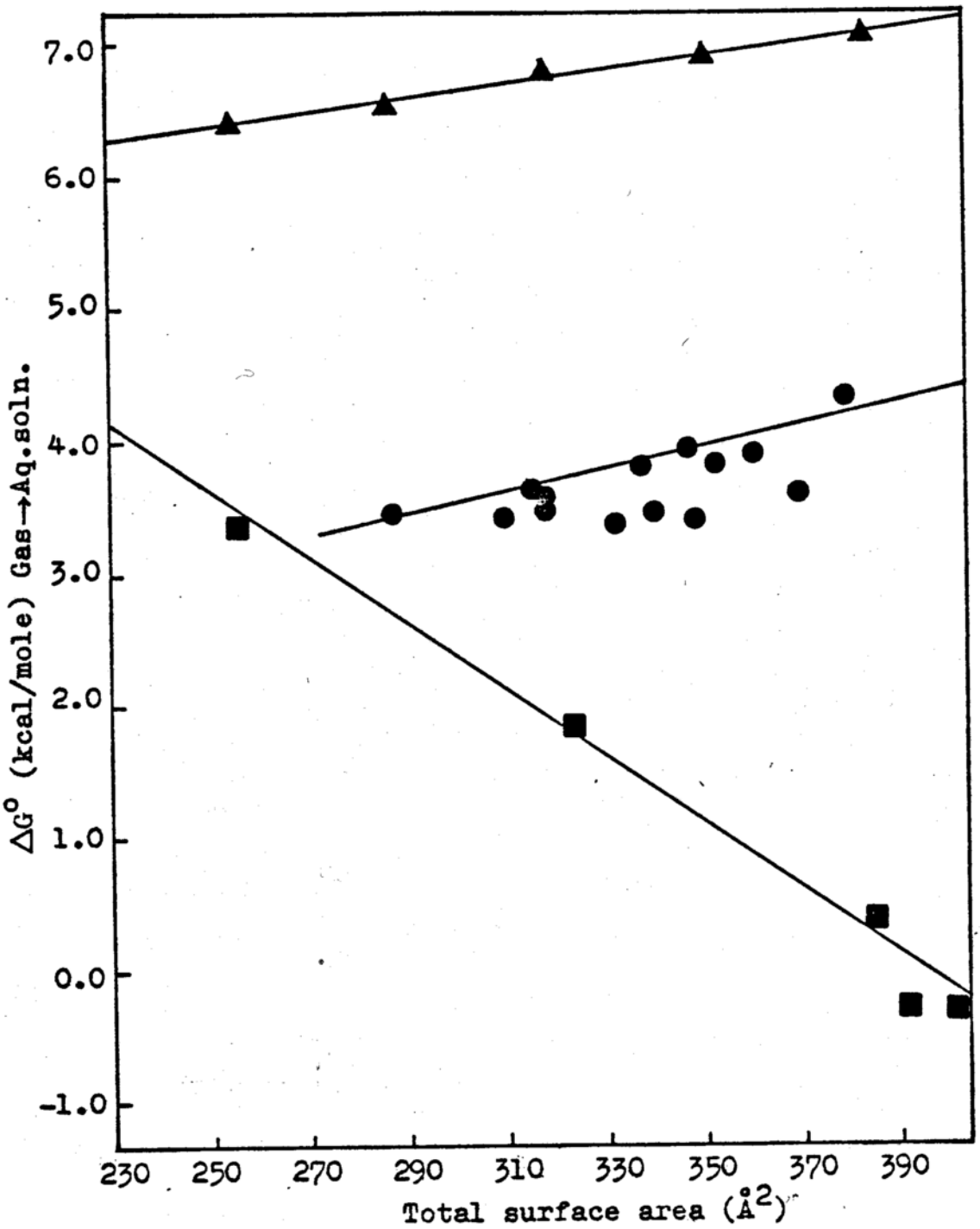


Fig.10. Free energy of hydration vs total surface area for Aliphatic(▲),Aromatic(■) and Alkyl aromatic (●) hydrocarbons.

Table 9. Analysis of the Data for the Aliphatic and Aromatic Hydrocarbons Using the Equation $\Delta G^\circ = C_0 + C_1 \cdot TSA.$[a,b]

Process	C_0	C_1	n	r	s
Aliphatic hydrocarbons					
pl → aq. soln.	-1.1 (.1)	0.029 (.0004)	5	1.0	.041
pl → gas	-6.2 (.08)	0.023 (.0003)	5	1.0	.025
gas → aq. soln.	5.1 (.1)	0.006 (.0005)	5	.98	.046
Aromatic hydrocarbons[c]					
pl → aq. soln.	-3.5 (1.1)	0.031 (.003)	5	.98	.36
pl → gas	-13.6 (1.1)	0.057 (.003)	5	.99	.37
gas → aq. soln.	10.1 (.7)	-0.026 (.002)	5	.99	.23

[a] Values in parentheses are standard errors.

[b] n = number of data points, r = correlation coefficient, s = standard error of regression.

[c] pl refers to pure supercooled liquid except for benzene.

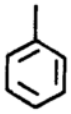
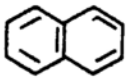
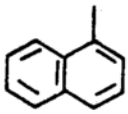
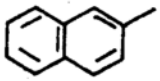
of the n-alkyl benzenes indicating a more favorable interaction with water. Thus in going from benzene to toluene there is an increase of only 50 calories in the free energy of solvation whereas 1,2,3 trimethyl benzene and benzene have about the same solvation free energy. The effect of methyl groups is further exemplified by Table 10 which lists the free energy of hydration for 1 and 2 methyl naphthalenes. It is observed that 1 methyl naphthalene is more soluble than naphthalene whereas 2 methyl naphthalene is less soluble than naphthalene when considering the gas phase to solution process. These considerations clearly suggest that methyl groups adjacent to aromatic residues must be treated as distinct units in any general model.

The simplest model other than using the total surface area is to partition it into aromatic and aliphatic surface areas and can be written as

$$\Delta G = \theta_0 + \theta_1 \text{AromSA} + \theta_2 \text{ALSA} + \theta_3 I \qquad \text{Model 1}$$

This however relies on the additivity of the separate contributions of the alkyl and aromatic portions and on the basis of the above considerations will not be a satisfactory model for cycle steps 2 and 3. Given the data for the polymethyl aromatic compounds any model must include a term or terms to account for the different nature of the methyl or methylene group adjacent to the aromatic nucleus. The

Table 10. Comparison of the Free Energies of Hydration of 1 and 2 Methyl Naphthalenes.

Compound	Structure	Free Energy of Hydration (Kcal/mole)
Benzene		3.40
Toluene		3.45
Naphthalene		1.87
1-Methyl Naphthalene		1.68
2-Methyl Naphthalene		1.97

terms in model 1 can be partitioned further to give the hierarchy of models below,

$$\Delta G = \theta_0 + \theta_1 \text{AromSA} + \theta_2 \text{ALov} + \theta_3 \text{Remalp} + \theta_4 I \qquad \text{Model 2}$$

$$\Delta G = \theta_0 + \theta_1 \text{ArC} + \theta_2 \text{ArH} + \theta_3 \text{ALSA} + \theta_4 I \qquad \text{Model 3}$$

$$\Delta G = \theta_0 + \theta_1 \text{ArC} + \theta_2 \text{ArH} + \theta_3 \text{ALov} + \theta_4 \text{Remalp} + \theta_5 I \qquad \text{Model 4}$$

The various group areas are defined as in Figure 3. 'I' is an indicator variable and accounts for the presence of an aromatic group. It has the value of 1 for the alkyl and polycyclic aromatics and 0 for the alkanes and is similar to the IFG coefficient discussed in earlier work (39). Several other models were also tested incorporating dipole moments, polarizabilities and various definitions for the overlap area such as sum of alkyl and attached aromatic carbon and 2 carbon contributions of the alkyl chain. These additional terms did not significantly improve the correlation (see Appendix B). The alkanes and the polycyclic aromatics were included in the data set for 2 reasons: i) to uncouple correlations between some of the group surface area terms and ii) the polycyclic aromatics give a wider range of values for the free energy changes and hence allow extension of the model to alkyl derivatives of the higher aromatics such as the alkyl naphthalenes,

anthracenes and polycyclic aromatics.

2. Analysis of Models

a. Pure liquid to aqueous solution process. Table 11 lists the relevant statistics for the 4 models applied to the pure liquid to solution process. All the models give very good correlation as might be expected from the previous discussion. The coefficients of the ArSA and AlSA terms in model 1 are identical to those obtained when the aliphatic and aromatic hydrocarbons are regressed independently versus total surface area suggesting independent additive contributions for these two classes of compounds. These values however are misleading since the nonadditives are evident when the solubility process is partitioned into its contributing steps. The negative value of the indicator variable is consistent with the observation that a hydrocarbon with an aromatic residue is more soluble than an alkane with the same surface area.

Although further partitioning of the ArSA and AlSA terms as in models 2, 3 and 4 does not improve the regression statistics it allows comparison of the relative magnitudes of the different surface area contributions. Thus the ArC coefficients in both models 3 and 4 are larger than the ArH coefficients, but these are not significantly different at the 95% confidence level, whereas the AlSA, Alov and Remalp coefficients are almost identical. The

Table 11. Results of Regression Analysis Using Different Models for the Process Pl → Aq. Soln. Surface Areas Calculated Using Solvent Radius = 1.5 Å.[a]

Model	r	s
1) $\Delta G° = \theta_0 + \theta_1 ArSA + \theta_2 AlSA + \theta_3 I$		
-1.13 0.030 0.028 -1.82	.989	.19
(0.36) (0.001) (0.001) (0.14)		
2) $\Delta G° = \theta_0 + \theta_1 ArSA + \theta_2 Alov + \theta_3 Remalp + \theta_4 I$		
-1.12 0.030 0.028 0.027 -2.0	.990	.19
(0.35) (0.001) (0.001) (0.001) (0.2)		
3) $\Delta G° = \theta_0 + \theta_1 ArC + \theta_2 ArH + \theta_3 AlSA + \theta_4 I$		
-0.94 0.036 0.025 0.027 -1.78	.990	.18
(0.36) (0.004) (0.003) (0.001) (0.14)		
4) $\Delta G° = \theta_0 + \theta_1 ArC + \theta_2 ArH + \theta_3 Alov + \theta_4 Remalp + \theta_5 I$		
-0.59 0.046 0.014 0.024 0.026 -1.18	.991	.18
(0.45) (0.009) (0.009) (0.003) (0.001) (0.50)		

[a]Values in parentheses are standard errors. r = correlation coefficient, s = standard error of regression, no. of data points = 24.

magnitudes of the coefficients however must be interpreted with caution due to the high correlation between some of the area contributions (vide infra).

b. Gas phase to aqueous solution process. Analysis of the gas phase to aqueous solution cycle step is of primary interest since it provides the clearest basis for discussing the interaction of molecules and functional groups with water. Table 12 lists the models and statistical results for this process. In contrast to the previous results for the pure liquid to aqueous solution process, model 1 fairs poorly. Model 2 shows a significant improvement over 1 both in a higher correlation coefficient and smaller standard error. Intuitively one might expect the inclusion of the aliphatic overlap area term to account for the observed solution behavior of the alkyl aromatics and hence model 2 to be the most suitable, since it is the polymethyl benzenes which show the greatest deviations in Figure 10. However the statistics for model 3, which involves partitioning of the ArSA rather than the AlSA, are better than model 2. Furthermore the coefficient of the Remalp term in model 2 is negative. This is in direct conflict with the results obtained with alkanes alone indicating that this model is not a realistic one.

Models 3 and 4 then are possible choices for representing the free energy change for this process. Both

Table 12. Results of Regression Analysis Using Different Models for the Process Gas → Aq. Soln. Surface Areas Calculated Using Solvent Radius = 1.5 Å.[a]

	Model	r	s
1)	$\Delta G° = \theta_0 + \theta_1 ArSA + \theta_2 AlSA + \theta_3 I$		
	7.77 -0.016 -0.003 -1.3	.958	.62
	(1.16) (0.003) (0.003) (0.45)		
2)	$\Delta G° = \theta_0 + \theta_1 ArSA + \theta_2 Alov + \theta_3 Remalp + \theta_4 I$		
	7.68 -0.065 0.023 0.002 -1.39	.985	.38
	(0.71) (0.002) (0.003) (0.002) (0.40)		
3)	$\Delta G° = \theta_0 + \theta_1 ArC + \theta_2 ArH + \theta_3 AlSA + \theta_4 I$		
	6.22 -0.065 0.023 0.002 -1.39	.993	.26
	(0.51) (0.005) (0.004) (0.002) (0.19)		
4)	$\Delta G° = \theta_0 + \theta_1 ArC + \theta_2 ArH + \theta_3 Alov + \theta_4 Remalp + \theta_5 I$		
	5.50 -0.086 0.046 0.009 0.004 -2.62	.994	.27
	(0.61) (0.012) (0.013) (0.004) (0.002) (0.68)		

[a]Values in parentheses are standard errors. r = correlation coefficient, s = standard error of regression, no. of data points = 24.

models seem satisfactory with similar correlation coefficients and standard errors. In choosing a model which might adequately represent the process under consideration, the model must be physically acceptable and should also possess the ability to estimate the solution properties of compounds not included in the original analysis. With respect to the latter criterion, both models do reasonably well in estimating the free energy changes for the alkyl naphthalenes (vide infra).

In considering the relative magnitudes of the coefficients for these two models and their group contribution implications, it is first necessary to examine the correlation between the variables included in the model. Table 13 lists the correlation matrix of the variables included in the model. The ArC, ArH and AlSA terms all are highly correlated. Partitioning of the AlSA term however into Alov and Remalp terms reduces the correlation between these latter variables and the ArC and ArH terms. In this respect model 4 is preferable to model 3.

If we assume that the magnitude of the coefficients reflects the free energy of interaction ($\delta\Delta G°/\mathring{A}^2$) of a particular group, then the coefficients of model 4 suggest the following: The aromatic carbons are intrinsically hydrophilic and the aromatic hydrogens are more intrinsically hydrophobic than the aliphatic hydrogens. Although the ArC and ArH terms are highly correlated, the standard errors of

Table 13. Correlation Matrix for Variables in the Regression Equation. Surface Areas Calculated Using 1.5 Å Solvent Radius.

	ArSA	ArC	ArH	AlSA	AlOV	Remalp	I
ArSA	1						
ArC	.991	1					
ArH	.996	.975	1				
AlSA	-.944	-.921	-.950	1			
AlOV	-.176	-.113	-.214	.158	1		
Remalp	-.704	-.725	-.684	-.764	-.517	1	
I	.716	.691	.726	-.689	.377	-.844	1

the coefficients are not excessively large and both coefficients are significant at the 95% confidence level. The partial correlation coefficient for the ArH term is also reasonably large (0.65). This interpretation is consistent with the observed data, where replacement of an aromatic hydrogen by a methyl group tends to increase the gas phase solubility. The Alov and Remalp coefficients are not significantly different and are within the range of the value obtained for the alkanes. For model 3, the coefficients are smaller in magnitude but indicate similar trends. The AlSA coefficient is much smaller than that observed for the alkanes.

The above interpretation is based entirely on surface area considerations. One of the limitations of this approach is that it does not directly account for the influence of substituents on the electronic changes in the molecule. This might be particularly relevant for the alkyl benzenes where increasing methyl substitution increases the Lewis basicity or hydrogen bonding ability of the aromatic nucleus (71). Table 14 lists the relative Lewis basicities measured using HCl as the acceptor molecule (72), and the corresponding ArH areas. The correlation coefficient between these two variables is -0.86 indicating that the ArH area reflects to some extent the Lewis basicity, and the improvement in the statistics of the model with the inclusion of the ArH term may in part be

Table 14. Lewis Basicity and ArH Surface Area for the Alkyl Benzenes.

Compound	ArH ($\AA^2$)	Rel. Basicity (HCl)
Benzene	170.1	.61
Toluene	129.8	.92
Ethyl benzene	125.5	1.06
o-Xylene	101.8	1.13
m-Xylene	90.2	1.26
p-Xylene	90.2	1.00
Isopropyl benzene	119.4	1.24
1,2,3 Trimethyl benzene	73.4	1.41
1,2,4 Trimethyl benzene	61.8	1.31
1,3,5 Trimethyl benzene	50.2	1.59
t-Butyl benzene	110.4	1.36

due to the correlation. Hence the favorable $\delta\Delta G°$ of solvation with increasing methyl substitution can also be explained as being a result of changes in the π character of the aromatic nucleus. The surface area method however cannot differentiate between these two alternative explanations.

c. Pure liquid to gas phase process. The results for this cycle step are given in Table 15. Model 4 has the smallest standard error of regression, is the only model with the correct sign for the indicator variable and more satisfactorily accounts for the alkane data.

As in the previous discussion of the contributing cycle steps, the relative magnitudes of the coefficients may be interpreted in terms of the contribution of the different group surface areas to the free energy of vaporization. The positive coefficient for the aromatic carbon area may be attributed to the strong interaction between the aromatic residues in the solute phase. The ArH coefficient is negative and suggests a decrease in the free energy of vaporization with increasing ArH area while the Alov and Remalp coefficients are positive and result in increasing free energies of vaporization with increasing group surface area. The coefficient of the Remalp term is identical to that for the alkanes and suggests an independence of the methylene and aromatic group at longer chain

Table 15. Results of Regression Analysis Using Different Models for the Process Pl → Gas. Surface Areas Calculated Using Solvent Radius = 1.5 Å.[a]

Model	r	s
1) $\Delta G° = \theta_0 + \theta_1 ArSA + \theta_2 AlSA + \theta_3 I$		
-8.91 0.046 0.031 -0.79	.965	.69
(1.28) (0.003) (0.004) (0.50)		
2) $\Delta G° = \theta_0 + \theta_1 ArSA + \theta_2 Alov + \theta_3 Remalp + \theta_4 I$		
-8.81 0.052 0.040 0.031 -2.71	.987	.43
(0.79) (0.003) (0.003) (0.002) (0.45)		
3) $\Delta G° = \theta_0 + \theta_1 ArC + \theta_2 ArH + \theta_3 AlSA + \theta_4 I$		
-7.17 0.101 0.001 0.025 -0.38	.995	.26
(0.50) (0.005) (0.004) (0.002) (1.19)		
4) $\Delta G° = \theta_0 + \theta_1 ArC + \theta_2 ArH + \theta_3 Alov + \theta_4 Remalp + \theta_5 I$		
-6.1 0.133 -0.032 0.015 0.022 1.45	.997	.21
(0.53) (0.010) (0.011) (0.004) (0.002) (0.59)		

[a]Values in parentheses are standard errors. r = correlation coefficient, s = standard error of regression, no. of data points = 24.

lengths.

The contribution of the different group areas to the change in the free energy of vaporization may be illustrated by the example for benzene and toluene in Table 16. The vapor pressure of toluene and the other polymethyl benzenes is a function of all the group surface area terms rather than the area of the methyl group alone. A consequence of the different contributions of the group surface areas is that the polymethyl benzenes have a higher free energy of vaporization compared to corresponding n-alkyl benzene. Finally it may be noted that the major portion of the change in the free energy of solution (step 1) with increasing methyl substitution is largely a result of the increased interaction of the polymethyl benzenes in the solute phase.

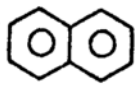
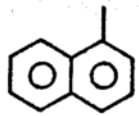
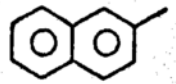
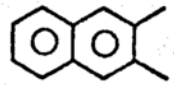
3. Extension of the Model to Higher Alkyl and Polycyclic Aromatics

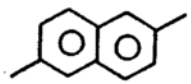
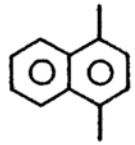
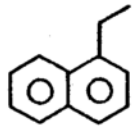
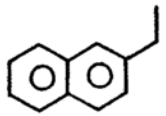
The value of the above analysis is further enhanced by its extension to estimating the solution properties of compounds which were not included in the original analysis. Table 17 lists surface area contributions and observed and predicted free energies of hydration for naphthalene and some of its alkyl derivatives, and Figure 11 shows a graph of observed vs. predicted values for the free energy changes for all 3 steps, where the predicted free energies

Table 16. Group Contributions to the Free Energy of Vaporization for Benzene and Toluene.

		ArC	ArH	Alov	Remalp	Intercept and Indicator Variable	Obsv.	Pred.
Surface Areas	Benzene	85.6	170.1	0	0			
	Toluene	72.7	129.8	84.0	0			
ΔG° (Kcal)	Benzene	11.38	-5.44	0	0	-4.65	1.23	1.29
	Toluene	9.67	-4.15	1.26	0	-4.65	1.97	2.13

Table 17. Group Surface Areas and Observed and Predicted Values of the Free Energy of Hydration for the Alkyl Naphthalenes.

Compound	Structure	ArC	ArH	AlOV	Remalp	TSA	Free Energy of Hydration (Kcal/mole) Obsv.	Pred.
Naphthalene		120.2	203.1	0.0	0.0	323.3	1.87	1.85
1 Methyl naphthalene		108.0	164.6	74.5	0.0	347.4	1.68	1.84
2 Methyl naphthalene		107.4	162.7	84.1	0.0	354.1	1.97	1.88
2,3 Dimethyl naphthalene		97.5	134.4	145.8	0.0	377.6	1.94	1.99

2,6 Dimethyl naphthalene		94.6	122.4	168.0	0.0	385.0	2.31	1.88
1,4 Dimethyl naphthalene		95.7	126.1	149.0	0.0	370.8	1.53	1.79
1 Ethyl naphthalene		103.6	154.0	38.5	76.5	373.6	1.78	1.71
2 Ethyl naphthalene		100.0	158.4	44.3	80.4	383.1	2.05	2.28

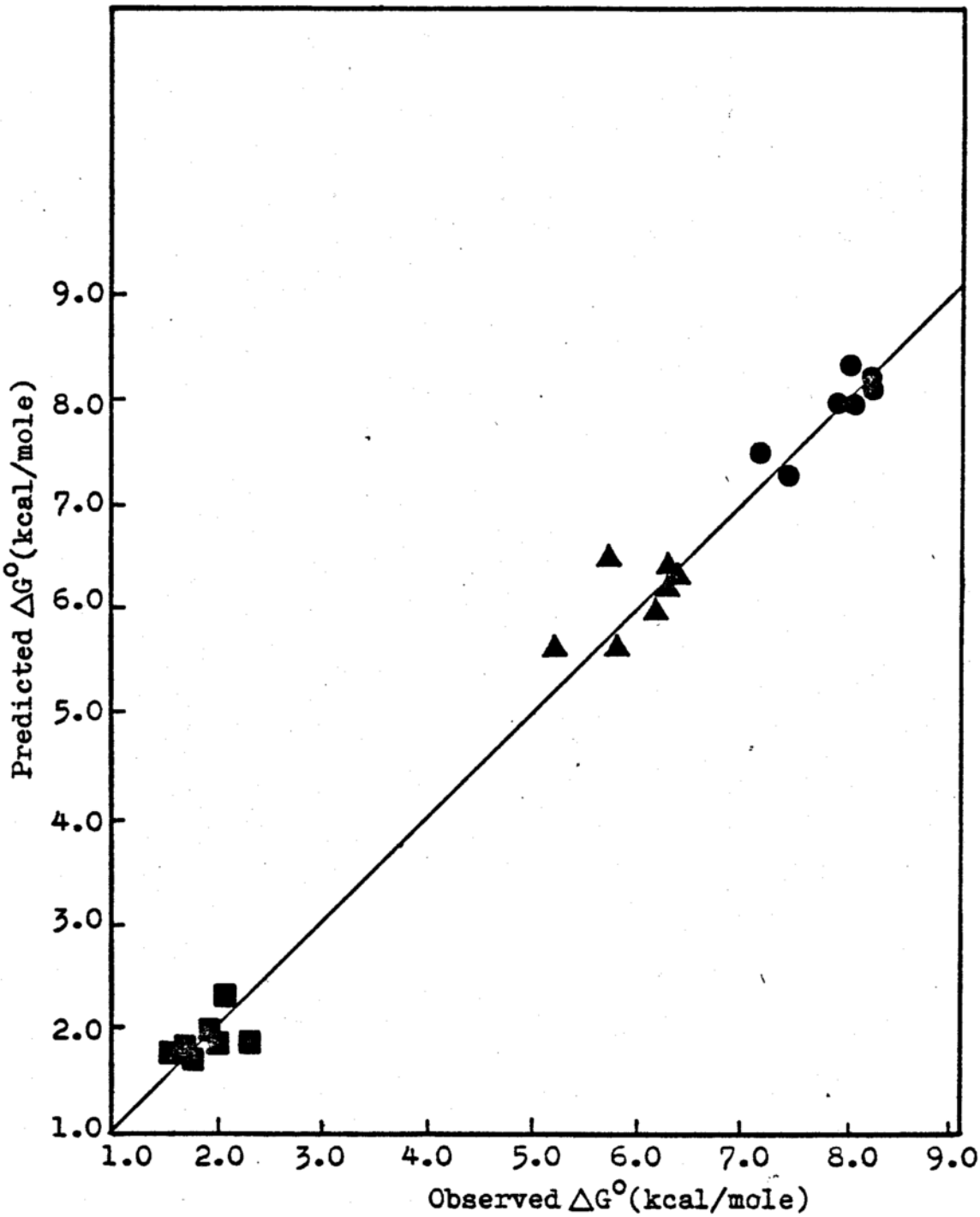


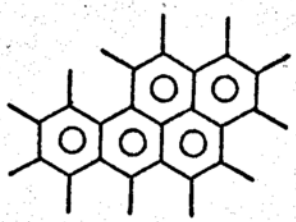
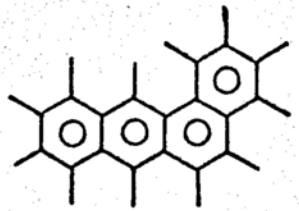
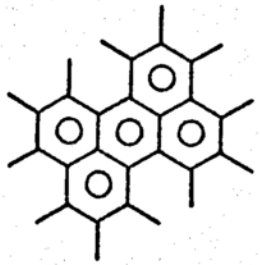
Fig.11. Predicted ΔG° vs Observed ΔG° for the three steps Pl—Aq.soln.(●), Pl—gas(▲), and Gas—Aq.soln.(■) for the alkyl naphthalenes using the model

$$\Delta G^{\circ} = \theta_0 + \theta_1 \mathrm{ArC} + \theta_2 \mathrm{ArH} + \theta_3 \mathrm{Alov} + \theta_4 \mathrm{Remalp} + \theta_5 \mathrm{I}$$

were calculated using model 4. In most cases the model is qualitatively correct in predicting the effects of the methyl group on the gas phase solubility. Thus in going from 1 to 2 methyl naphthalene the solubility decreases whereas in going from 2,3 dimethyl naphthalene to 1,4 dimethyl naphthalene the solubility increases. Table 18 shows observed and predicted free energy changes for the pure liquid to aqueous solution process for some higher polycyclic aromatics. The observed and predicted values are again in good agreement in most cases indicating that the model is of general applicability in estimating the solubilities of a variety of polycyclic aromatic and alkyl aromatic compounds.

4. Choice of Solvent Radius in Surface Area Calculations and its Effect on the Regression Equation

The surface areas used in the previous regression analysis were calculated using a 1.5 Å solvent radius for water. There has been some controversy in the past (39,75) regarding the choice of zero vs. 1.5 Å solvent radius in the use of surface area as a correlative parameter. Previous results have shown a linear relationship between the TSA (1.5 Å solvent radius) and TSA (zero solvent radius) for the normal alkanes. A similar relation is obtained for the alkyl benzenes given by

1,2 Benzo-pyrene		237.0	234.2	0.0	0.0	471.1	11.66	11.12
1,2 Benzanth-racene		189.8	262.1	0.0	0.0	451.9	10.63	10.63
Perylene		200.6	255.4	0.0	0.0	456.0	10.91	11.04

[a]Solubilities from refs. 17, 73. ΔH_f° and T_m from ref. 74.

[b]The observed solubility appears to be rather high.

[c]Predicted solubilities calculated using Model 4.

$$TSA(1.5) = 1.30\ TSA(0.0) + 115.1 \qquad \text{Eq. 20}$$

$$r = .99$$

Hence the regression statistics obtained with the previous models using 1.5 Å solvent radius should not significantly change when using surface areas calculated with zero solvent radius.

Tables 19, 20 and 21 (which are analogous to Tables 11, 12 and 15) list the regression results obtained using zero solvent radius. The regression statistics are very similar. The difference in magnitudes of the coefficients may be attributed to two factors: i) Since the TSA and the contributing group surface areas are smaller, the free energy/unit area increases in order to account for the same free energy change for a particular compound (see Table 22). ii) Although the total surface areas with zero and 1.5 Å solvent radius are linearly related by eq. 2, the same relation does not hold between the contributing group surface areas. Table 23 shows the relationships between the ArC, ArH, ArSA and TSA terms. The contribution of the ArSA to the total surface area is about the same for both solvent radii but the relative contribution of the ArC and ArH terms are quite different. For the zero solvent radius the ratio of ArC/ArH is always greater than 1 whereas with 1.5 Å solvent radius, it is always less than 1. Thus the coefficients of the ArC terms in the zero solvent radius

Table 19. Results of Regression Analysis Using Different Models for the Process Pl → Aq. Soln. Surface Areas Calculated Using Zero Solvent Radius.[a]

	Model	r	s
1)	$\Delta G° = \theta_0 + \theta_1 ArSA + \theta_2 AlSA + \theta_3 I$		
	1.75 0.043 0.037 -1.86	.991	.17
	(0.23) (0.001) (0.001) (0.13)		
2)	$\Delta G° = \theta_0 + \theta_1 ArSA + \theta_2 Alov + \theta_3 Remalp + \theta_4 I$		
	1.78 0.044 0.039 0.037 -2.1	.992	.16
	(0.22) (0.001) (0.002) (0.001) (0.18)		
3)	$\Delta G° = \theta_0 + \theta_1 ArC + \theta_2 ArH + \theta_3 AlSA + \theta_4 I$		
	1.89 0.054 0.025 0.036 -1.89	.993	.15
	(0.22) (0.005) (0.008) (0.001) (0.12)		
4)	$\Delta G° = \theta_0 + \theta_1 ArC + \theta_2 ArH + \theta_3 Alov + \theta_4 Remalp + \theta_5 I$		
	1.96 0.060 0.012 0.034 0.036 -1.73	.993	.16
	(0.26) (0.010) (0.024) (0.005) (0.002) (0.34)		

[a]Values in parentheses are standard errors. r = correlation coefficient, s = standard error of regression, no. of data points = 24.

Table 20. Results of Regression Analysis Using Different Models for the Process Pl → Gas. Surface Areas Calculated Using Zero Solvent Radius.[a]

Model	r	s
1) $\Delta G° = \theta_0 + \theta_1 ArSA + \theta_2 AlSA + \theta_3 I$		
-4.65 0.071 0.035 -1.44	.982	.50
(0.68) (0.004) (0.004) (0.38)		
2) $\Delta G° = \theta_0 + \theta_1 ArSA + \theta_2 Alov + \theta_3 Remalp + \theta_4 I$		
-4.44 0.079 0.050 0.033 -3.01	.997	.21
(0.29) (0.002) (0.002) (0.002) (0.24)		
3) $\Delta G° = \theta_0 + \theta_1 ArC + \theta_2 ArH + \theta_3 AlSA + \theta_4 I$		
-3.80 0.135 -0.036 0.030 -1.63	.998	.16
(0.23) (0.005) (0.008) (0.001) (0.12)		
4) $\Delta G° = \theta_0 + \theta_1 ArC + \theta_2 ArH + \theta_3 Alov + \theta_4 Remalp + \theta_5 I$		
-3.89 0.127 -0.018 0.033 0.030 -1.87	.998	.16
(0.26) (0.012) (0.025) (0.005) (0.002) (0.34)		

[a]Values in parentheses are standard errors. r = correlation coefficient, s = standard error of regression, no. of data points = 24.

Table 21. Results of Regression Analysis Using Different Models for the Process Gas → Aq. Soln. Surface Areas Calculated Using Zero Solvent Radius.[a]

	Model	r	s
1)	$\Delta G° = \theta_0 + \theta_1 ArSA + \theta_2 AlSA + \theta_3 I$		
	6.39 -0.028 0.002 -0.43	.979	.44
	(0.61) (0.003) (0.003) (0.34)		
2)	$\Delta G° = \theta_0 + \theta_1 ArSA + \theta_2 Alov + \theta_3 Remalp + \theta_4 I$		
	6.21 -0.034 -0.011 0.004 0.089	.994	.24
	(0.32) (0.002) (0.003) (0.002) (0.26)		
3)	$\Delta G° = \theta_0 + \theta_1 ArC + \theta_2 ArH + \theta_3 AlSA + \theta_4 I$		
	5.68 -0.081 0.061 0.007 -0.27	.995	.21
	(0.31) (0.007) (0.001) (0.002) (0.17)		
4)	$\Delta G° = \theta_0 + \theta_1 ArC + \theta_2 ArH + \theta_3 Alov + \theta_4 Remalp + \theta_5 I$		
	5.82 -0.068 0.035 0.002 0.006 0.08	.995	.22
	(0.35) (0.016) (0.033) (0.006) (0.002) (0.46)		

[a]Values in parentheses are standard errors. r = correlation coefficient, s = standard error of regression, no. of data points = 24.

Table 22. Comparison of $\delta\Delta G°/\mathring{A}^2$ for Surface Areas Calculated With Zero and 1.5 Å Solvent Radius.

	Alkanes		Aromatics	
	0.0	1.5	0.0	1.5
Step 1	.040	.029	.045	.028
Step 2	.032	.023	.081	.055
Step 3	.008	.006	-.036	-.027

tert Butyl benzene	.29	.20	1.4	.49	.16	.31	.52	.47
1 Methyl 4 Isopropyl benzene	.26	.16	1.6	..42	.13	.22	.59	.35
Naphthalene	.59	.41	1.4	1.0	.37	.63	.59	1.0
Anthracene	.60	.40	1.5	1.0	.40	.60	.67	1.0
Phenanthrene	.62	.38	1.6	1.0	.40	.60	.67	1.0
Pyrene	.63	.37	1.7	1.0	.41	.59	.70	1.0

models will be smaller than those which would have been obtained if the ArC/ArH ratio was the same as in surface areas calculated using a 1.5 Å solvent radius. This however does not mean that the coefficients must necessarily be smaller in magnitude than those obtained using a 1.5 Å solvent radius (see i). Similar considerations also apply to the ArH coefficient.

The trends in the relative magnitudes of the coefficients are similar to those previously considered and the interpretations suggested by the models do not change. Model 4 (zero solvent radius) also correctly predicts the free energy changes for the alkyl naphthalenes indicating that the choice of the surface area to be used as a correlative parameter is to some extent arbitrary. In this study we have focused on the use of surface areas calculated with a 1.5 Å solvent radius in order to extend these results to solvation and desolvation processes in the binding of ligand molecules to proteins (vide infra).

C. Scaled Particle Theory (SPT) Results

1. Estimation of the Lennard Jones (6,12) Pair Potential Parameters and the Polarizabilities of the Aromatic and Alkyl Aromatic Hydrocarbons

Several methods have been proposed for the determination of the Lennard Jones pair potential parameters ε (the energy minimum) and σ_H (the hard sphere diameter).

The most frequently cited methods use gas solubility data (76), gas phase virial coefficients and viscosities (77), heats of vaporization (37), as well as empirical correlations with critical temperatures and volumes (78). The gas solubility method was used to determine ε and σ_H for benzene, toluene and m-xylene (75). However, the corresponding data for the remaining alkyl aromatic compounds is not available hence in order to determine these parameters for the remaining alkyl benzenes suitable correlations must be developed. The hard sphere diameter σ_H correlates well with the σ_{SA} determined from the molecular surface area (calculated using a zero solvent radius) for the alkanes as well as the aromatics (Fig. 12). The hard sphere diameters for the alkanes were taken from Nelson and deLigny (78). Both the aliphatic and aromatic compounds lie on the same line. The linear relationship is given by

$$\sigma_H = 0.6234\sigma_{SA} + 1.542 \qquad r = .996 \qquad \text{Eq. 21}$$

and was used to calculate σ_H for the remaining alkyl aromatic compounds (Table 24).

The determination of the ε or ε/k (where k is the Boltzmann constant) value provides a more difficult problem. There is considerable variation amongst the values reported in the literature (37) depending on the method used to determine them. Nelson and deLigny (78) employed the

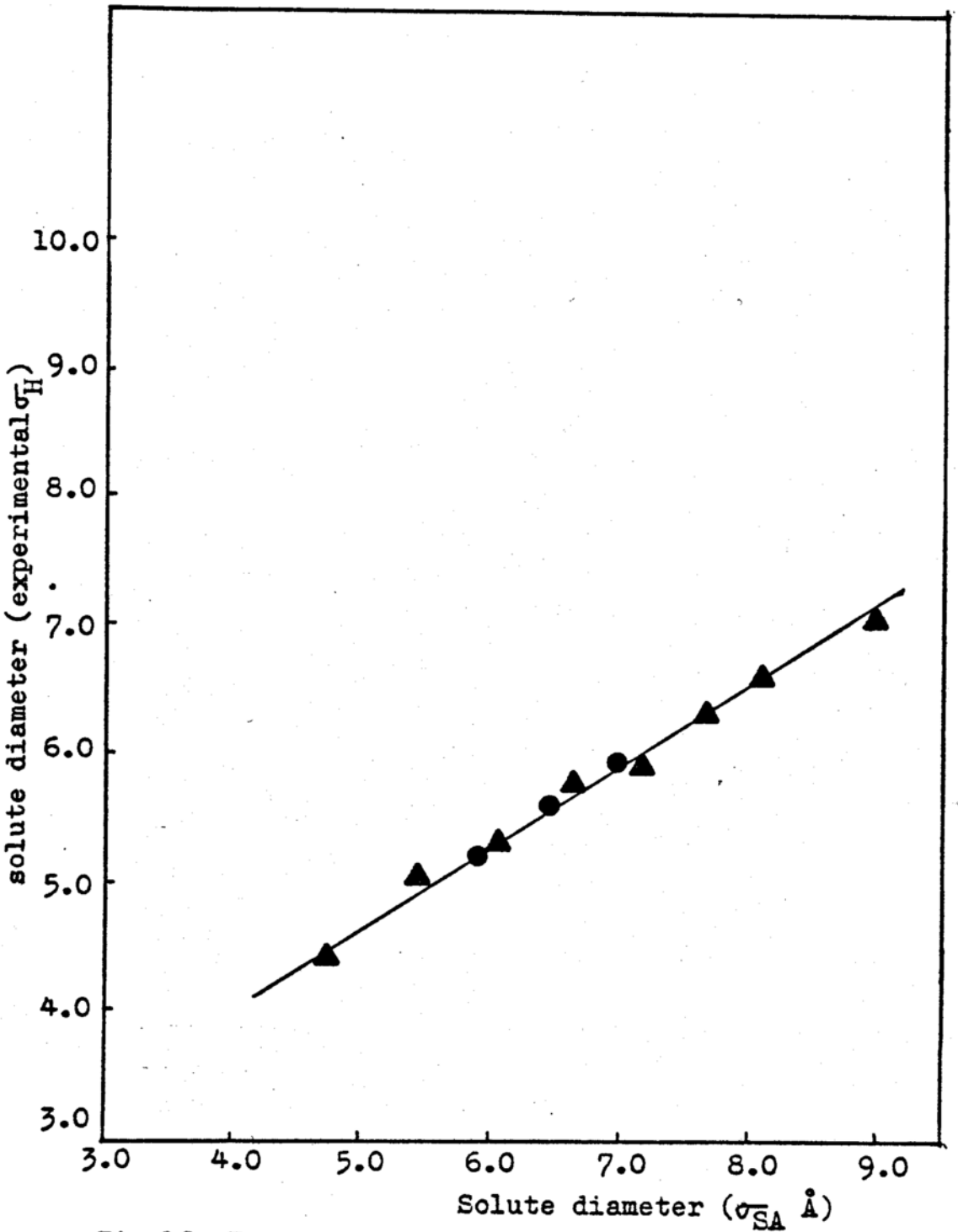


Fig.12. Experimental solute hard sphere diameter vs solute diameter calculated from surface area. ▲Alkanes(Ref.78), ● Alkyl aromatics (Ref.75).

empirical correlation

$$\varepsilon/k = .74T_c \qquad \text{Eq. 22}$$

for the alkanes, where T_c is the critical temperature. However their values are not in good agreement with those determined by the gas solubility method (76). The ε/k for benzene, toluene and m-xylene exhibited a reasonable correlation with their ionization potentials (Fig. 13). For lack of a better method, the relation

$$\varepsilon/k = -85.8IP + 1325.5 \qquad r = .98 \qquad \text{Eq. 23}$$

where IP is the ionization potential in eV, was used to calculate $\varepsilon/$ for the alkyl and polycyclic aromatics. The average polarizability was determined by summation of the individual bond contributions (79). The bond contributions are average values obtained by averaging α_θ, the polarizability of the molecule when the electric field makes an angle θ with the chemical bond, over all orientations and is given by

$$\alpha = \frac{1}{3}(\alpha_{||} + 2\alpha_{\perp}) \qquad \text{Eq. 24}$$

where $\alpha_{||}$ is the polarizability parallel to the bond and $\alpha_{\perp}$ is the polarizability perpendicular to the bond.

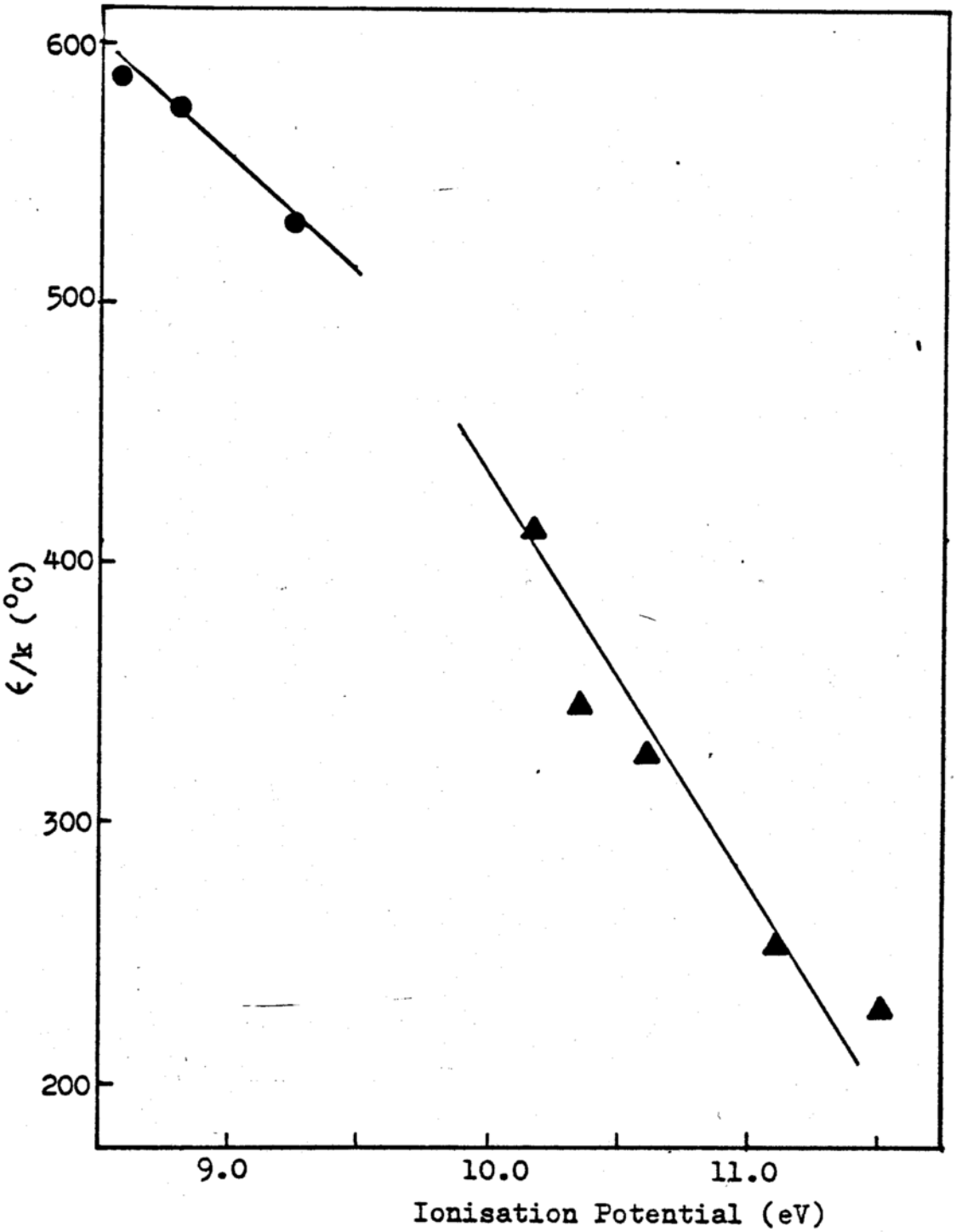


Fig. 13. ϵ/k vs Ionisation potential.
▲ Alkanes(C_2-C_6), ● (benzene, toluene, m-xylene).

Table 24 lists the values of the parameters used in the calculations. For the alkanes, the parameters reported by Nelson and deLigny were used.

2. Calculation of the Free Energies and Entropies of Solvation

Table 25 lists the calculated and observed free energies and entropies of solvation for the alkyl and polycyclic aromatics. The free energies of solvation were calculated using equations 5-9 described earlier. The entropy of solvation is given by

$$\Delta S_{SPT} = [\Delta H_c - \Delta G_c - RT \ln \frac{RT}{\bar{V}_1}]/T \qquad \text{Eq. 25}$$

where ΔH_c is the enthalpy of cavity formation and the entropy contribution of the term 'RT ln $(RT/\bar{V}_1)$' is due to the confinement of the solute from the gas phase to the molar volume of the solvent. The SPT assumes that ΔS_i, the entropy change due to solute-solvent interaction is small. The expression for the enthalpy of cavity formation is given by (37)

$$\Delta H_c = y\alpha_p RT^2(1 - y)^{-3}[(1 - y)^2 + 3(1 - y)D + 3(1 + 2y)D^2] + yPV_1D^3 \qquad \text{Eq. 26}$$

Naphthalene	156.8	7.07	5.95	17.0	8.12	629	0.0
Anthracene	203.5	8.05	6.56	23.6	7.55	678	0.0
Aliphatic[f] Hydrocarbons					IP		
Ethane	70.6	4.74	4.42	4.33	11.5	230	0
Propane	93.2	5.44	5.05	6.15	11.1	254	0
Butane	115.8	6.06	5.32	7.97	10.63	328	0
Pentane	138.4	6.64	5.77	9.80	10.35	345	0
Hexane	161.1	7.16	5.91	11.62	10.18	413	0
Hepane	183.7	7.66	6.33	13.44	9.90	416	0

[a] For aromatic hydrocarbons calculated from eq. 21 except for benzene, toluene, m-xylene.

[b] For aromatic hydrocarbons calculated from bond polarisabilities (79).

[c] Handbook of Phys. & Chem., 53rd Ed., 1972-73.

[d] For aromatic hydrocarbons calculated from eq. 23 except for benzene, toluene, m-xylene.

[e] Tables of Experimental Dipole Moments, A. L. McClellan, W. H. Freeman & Co., San Francisco, 1973.

[f] Data for aliphatic hydrocarbons from ref. 78.

Aliphatic Hydrocarbons									
Ethane	6.98	4.64	.63	0	5.27	5.98	6.15	33.9	33.6
Propane	8.75	6.26	.70	0	6.96	6.06	6.22	38.9	40.0
Butane	9.58	7.90	.82	0	8.72	5.13	6.42	41.2	41.3
Pentane	11.03	9.58	.86	0	10.44	4.86	6.54	45.2	45.1
Hexane	11.51	10.94	.97	0	11.91	3.87	6.82	46.5	48.2
Heptane	12.99	12.65	.97	0	13.62	3.64	6.93		

[a] ΔG in Kcal/mole, ΔS in cal/°/mole.

[b] $\Delta G_{solv} = \Delta G_c + \Delta G_i + RT \ln \frac{RT}{\bar{V}_l} = \Delta G_c + \Delta G_i + 4.27$

where α_p is the thermal expansion coefficient of the solvent, R is the gas constant and the other variables are as defined earlier. Appendix C gives an example of a calculation of the thermodynamic parameters for benzene in water.

3. Discussion. As is apparent from Table 25, the agreement between the calculated and observed solvation energies is for the most part poor (Fig. 14). This may be partly due to the empirical methods used in determining the parameters. However some generalizations can be made from Table 25. Considering the series benzene, naphthalene and anthracene the SPT correctly accounts for the decrease in the free energy of solvation with increasing surface area. This is a direct result of the proportionately larger increase in dispersion interactions over the cavity energy with increasing solute size. However the SPT also predicts a decrease in solvation energy (although smaller) with surface area for the alkanes so that the SPT apparently tends to overestimate the interaction energy with increasing solute size. This is also true for the alkyl benzenes where for example ethyl benzene has a smaller solvation free energy than toluene. However for a set of compounds with the same carbon number, e.g., the three carbon alkyl benzenes, SPT predicts a lower free energy of solvation of the polymethyl benzenes as compared to the corresponding n-alkyl benzene.

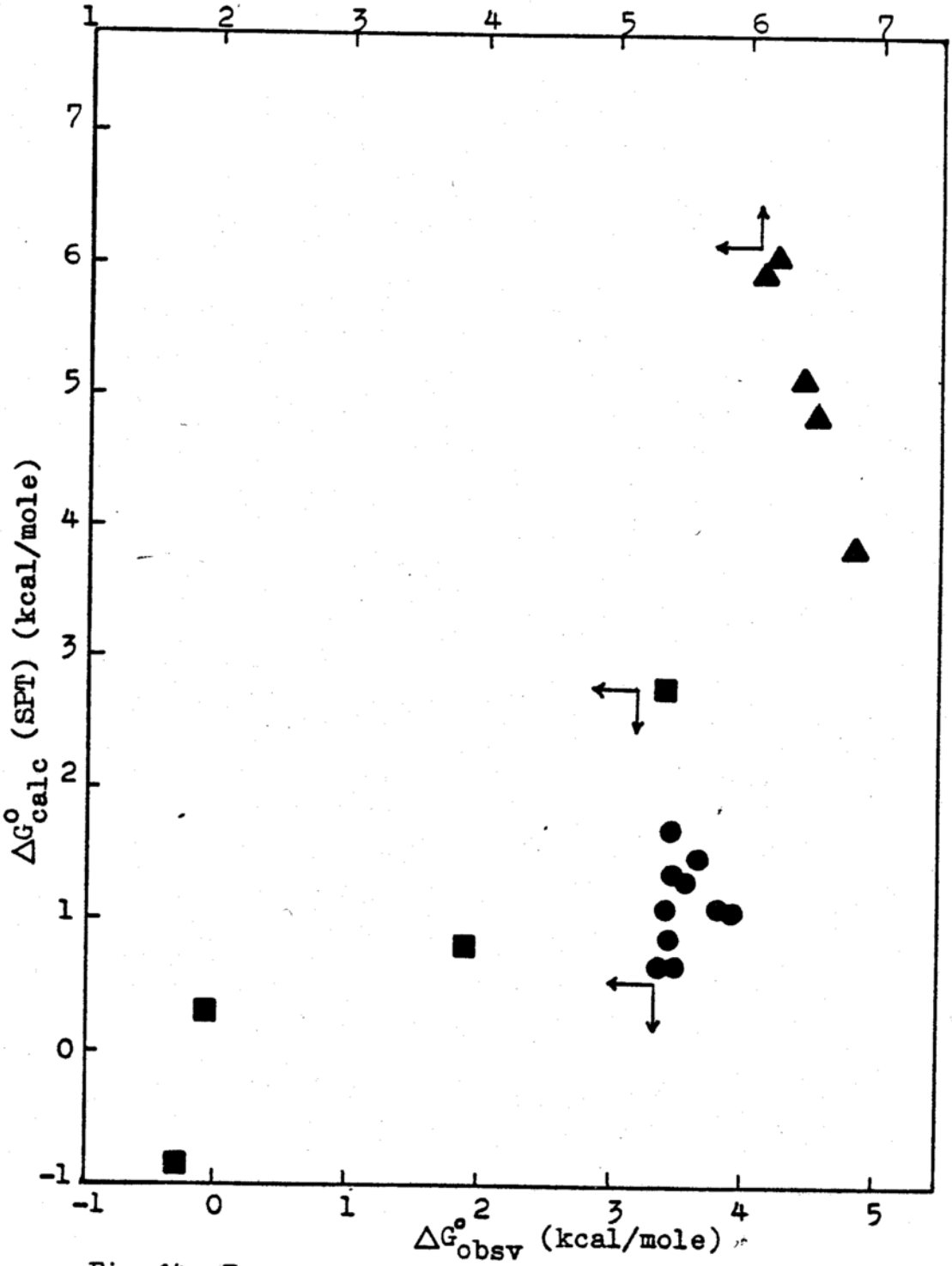


Fig.14. Free energy of solvation.Predicted(SPT) vs Observed. (see table 25). ▲Alkanes, ● Alkyl aromatics, ■ Polycyclic aromatics.

The entropies of solvation calculated by the SPT, as noted by Nelson and deLigny are in very good agreement with the observed values for the alkanes. This led them to suggest that the poor agreement between the observed and predicted free energies of solvation was due to an overestimation of the interaction energies. For the alkyl aromatics although the entropies of solvation calculated by the SPT are consistently more negative than the observed, the trends in $\delta\Delta S^{\circ}$ are again correctly reproduced. If as suggested by Nelson and deLigny we assume that the poor agreement in the solvation energies is a result of overestimation of the interaction energy, we can calculate a correction factor to the interaction energy required to get agreement with the observed results. Table 26 lists such correction factors. For the alkyl aromatics, this factor is remarkably constant. Nelson and deLigny from their results on n-alkanes have interpreted this in terms of a reduction in the density of water around the solute molecule. The reasons however may be more complex than suggested by this explanation, given the approximations in the SPT.

To summarize, within the context of the theory, SPT qualitatively accounts for the differing behavior of these three classes of hydrocarbons. The decrease in solvation energy for the aromatic hydrocarbons is ascribed to the increasing interaction of the aromatic residues with water.

Table 26. Correction Factor to Interaction Energy (See Text).

	$\bar{G}_{INT}$		Correction Factor
	Calc.	Required	
Alkanes			
Ethane	5.27	5.1	.97
Propane	6.96	6.8	.97
Butane	8.72	7.53	.86
Pentane	10.44	8.76	.84
Hexane	11.91	8.96	.75
Heptane	13.62	10.33	.76
Aromatics			
Benzene	10.84	10.20	.94
Naphthalene	15.09	14.04	.93
Anthracene	19.0	18.41	.97
Toluene	13.13	11.42	.87
Ethyl benzene	13.92	11.81	.85
o-Xylene	14.48	12.18	.84
m-Xylene	14.60	12.39	.85
p-Xylene	14.36	12.27	.85
Propyl benzene	15.82	13.0	.82
Isopropyl benzene	15.60	12.89	.83
1,2,3 Trimethyl benzene	15.81	13.12	.83
1,2,4 Trimethyl benzene	15.95	13.16	.83
1,3,5 Trimethyl benzene	15.89	13.38	.84

Although the trend for the alkanes is not correctly reproduced, the difference in solvation energies for the alkanes and aromatics is a consequence of the larger interaction energy for the aromatic hydrocarbons. Thus for example butane and benzene have the same σ_H but benzene has a larger interaction energy than butane resulting in a higher solubility. Finally for the alkyl aromatics the interaction energy is again overestimated, but the trend within a group of compounds of equivalent carbon number are similar to that experimentally observed. As is the case for the alkanes the interaction energy for the alkyl aromatics is overestimated by a constant factor.

D. The Intrinsic Solvent Contribution to the Free Energy of Protein-Ligand Interactions

1. General Considerations

The binding process may be partitioned in a manner analogous to the previous analysis into the two contributing steps as shown in Figure 15. The intrinsic solvent contribution refers to the free energy required to desolvate those portions of the protein and the ligand made inaccessible to the solvent on complexation and may be written as

$$\Delta G_{solv} = \Delta G_{P\text{-solvent}} + \Delta G_{L\text{-solvent}} \qquad \text{Eq. 27}$$

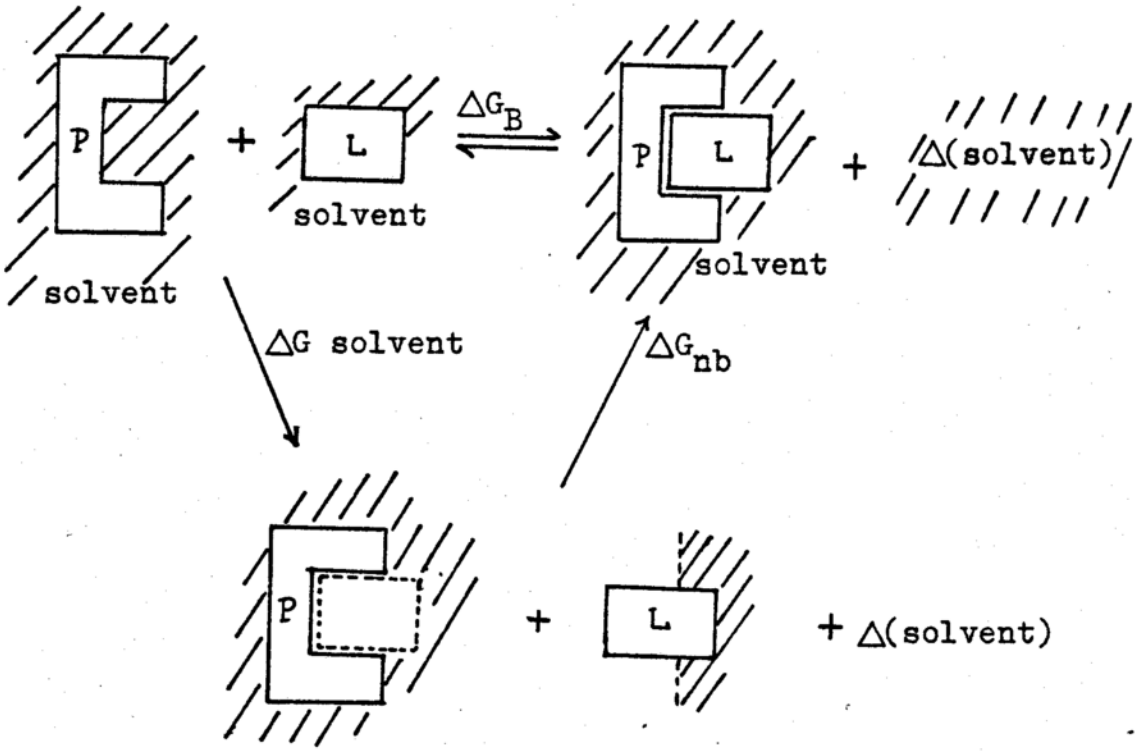


Fig.15. Partitioning of the binding process into two contributing steps.

where P and L stand for protein and ligand respectively. Focusing on the ligand-solvent contribution, the free energy change for this term is given by

$$\Delta G_{L\text{-solvent}} = k\Delta TSA \qquad \text{Eq. 28}$$

where ΔTSA is the accessible surface area of the bound ligand minus the surface area of the free ligand and k is a free energy per unit area term equivalent to the $\delta\Delta G^{\circ}_{solvation}$ per $\mathring{A}^2$. For aliphatic hydrocarbons it is about 6 cal per $\mathring{A}^2$ and for aromatic hydrocarbons, -27 cal per $\mathring{A}^2$.

Previous estimates have used the value of k derived from the free energy of solution of aliphatic hydrocarbons (25 cal per $\mathring{A}^2$) resulting in an overestimate of the solvent contribution since this value represents the net result of the two cycle steps. Thus for example the loss in accessible surface area in the trypsin-PTI complex is about 1400 $\mathring{A}^2$ (80). The free energy change for this process based on the value of 25 cal per $\mathring{A}^2$ would be 35 kcal per mole. However of this 35 kcal per mole, only 8 kcal (20%) is associated with the solvent contribution assuming the excluded surface area to be entirely aliphatic (6 cal per $\mathring{A}^2$). The remaining free energy change being due to the Van der Waals interaction between the ligand and the binding site assuming that this contribution is analogous to the

pure liquid to gas phase cycle step. Furthermore the value of 8 kcal (6 cal/$\mathring{A}^2$) is an upper limit since most of the groups involved in the binding process are less hydrophobic than the CH_2 group. Hence these general considerations suggest that the major driving force in the binding of ligand molecules to proteins is the Van der Waals interactions between the ligand and the binding site rather than the 'squeezing' out of the ligand by the solvent.

2. Aromatic Inhibitor Binding to α-Chymotrypsin

The calculation of the solvent contribution may be illustrated by the binding of the aromatic molecules benzene, naphthalene and anthracene to α-chymotrypsin. The preferred binding position of the aromatic inhibitor was determined by maximizing the nonbonded interaction between the ligand and binding site (81). The surface area of the ligand excluded on binding was determined as the difference between the surface area of the free and bound ligands exposed to the solvent. It is necessary to use a solvent radius in these surface area calculations to account for exclusion of the solvent molecules as a result of binding. This is illustrated in Figure 16 where although parts of the ligand may not be in Van der Waals contact with the binding site, this area is excluded from the solvent. The results of this calculation using a 1.5 $\mathring{A}$ solvent radius are listed in Table 27. The surface area lost on binding

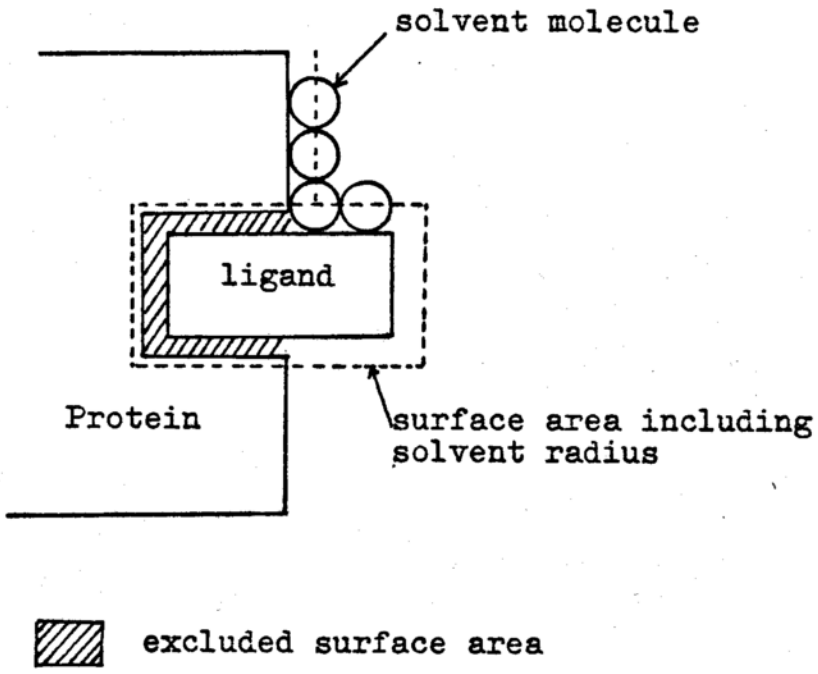


Fig.16. Solvent radius considerations in excluded surface area calculation.

Table 27. Results for Inhibitor Binding to α-Chymotrypsin.

	Exposed Surface Area Å^2			ΔG° (kcal/mole)	
Inhibitor	Free	Bound	ΔTSA_L	Ligand-Solvent[a]	Binding[b]
Benzene	256	22	234	6.3	-2.84
Naphthalene	323	64	259	7.0	-5.36
Anthracene	391	105	286	7.7	-7.16

[a] Intrinsic solvent contribution, see text.

[b] Experimental values from ref. 82.

is 92%, 80% and 73% respectively for benzene, naphthalene and anthracene. The ligand solvent contribution is given by

$$\Delta G_{L\text{-solvent}} = k\Delta TSA \qquad \text{Eq. 29}$$

where k = -27 cal per $\mathring{A}^2$ for the aromatic hydrocarbons. From these results it is evident that the solvent contribution becomes increasingly unfavorable for the series benzene, naphthalene and anthracene and hence the binding of these aromatic inhibitors is not solvent driven.

The remaining free energy terms associated with the binding process (Fig. 15) are ΔG_{nb} and $\Delta G_{P\text{-solvent}}$. The latter term is the solvent contribution of the excluded protein surface area on binding and an upper limit for this would be $\Delta G_{P\text{-solvent}} = k\Delta TSA$ where k = -6 cal/$\mathring{A}^2$ for an entirely aliphatic site. However the active site of α-chymotrypsin is known to be more polar than this and the $\Delta G^{\circ}_{P\text{-solvent}}$ contribution to the binding free energy is likely to be small, if not unfavorable. Hence the order of binding (Table 27): anthracene > naphthalene > benzene is probably a consequence of the nonbonded interaction free energy (ΔG_{nb}) between the ligand and the binding site rather than a free energy contribution of the solvent.

IV. SUMMARY

The pure liquid to aqueous solution process for the aliphatic, aromatic and alkyl aromatic hydrocarbons has been partitioned by means of a thermodynamic cycle into two contributing steps which reflect the contribution of the solute-solute interactions in the pure liquid phase and solute-solvent interactions in the solution phase. Results based on this partitioning process suggested that for the aliphatic hydrocarbons, 80% of the $\delta\Delta G°$ (per $\mathring{A}^2$) of solution was a result of solute-solute interactions in the solute phase, the remaining 20% being due to solute-solvent interactions. In contrast, the $\delta\Delta G°$ for the aromatic hydrocarbons for this latter process is actually negative, indicating a net favorable interaction of the aromatic residue with water.

The $\delta\Delta G°$ for the cycle steps has been further partitioned into enthalpic and entropic contributions for the aliphatic and aromatic hydrocarbons with the following results:

i) The $\delta\Delta G°$ for the pure liquid to aqueous solution process for the aliphatic hydrocarbons has a large negative $\delta T\Delta S°$ contribution.

ii) The $\delta\Delta G°$ for the pure liquid to aqueous solution process for the aromatic hydrocarbons has a large positive enthalpic component and a smaller

but positive $\delta T\Delta S°$ contribution.

iii) For the gas phase to aqueous solution process, the aromatic hydrocarbons have a larger negative $\delta\Delta H°$ of solvation and a much smaller negative $\delta T\Delta S°$ than the aliphatic hydrocarbons resulting in a negative $\delta\Delta G°$ of solvation with increasing surface area.

Hence with increasing surface area, the pure liquid (or solid) to aqueous solution process is entropically controlled for the aliphatic hydrocarbons while it is enthalpically controlled for the aromatic hydrocarbons. These results clearly indicate that when comparing the various $\delta\Delta G°$, $\delta\Delta H°$ and $\delta T\Delta S°$ for the three cycle steps for these two classes of hydrocarbons, the only similarity lies in the $\delta\Delta G°$ for the pure liquid to aqueous solution process. The remaining thermodynamic parameters are different in magnitude and in several cases sign, strongly emphasizing the differing interactions of these hydrocarbons with water.

The free energy data for the cycle steps for the alkyl aromatic compounds were analyzed using the group surface area approach. For the pure liquid to aqueous solution process the model

$$\Delta G° = \theta_0 + \theta_1 \mathrm{AromSA} + \theta_2 \mathrm{AlSA} + \theta_3 I$$

satisfactorily accounts for the data with the coefficients

similar to those obtained for the aromatics and alkanes independently. However for steps 2 and 3 the simplest model which is most satisfactory in accounting for the free energy data is of the form

$$\Delta G^\circ = \theta_0 + \theta_1 ArC + \theta_2 ArH + \theta_3 Alov + \theta_4 Remalp + \theta_5 I$$

where the coefficients for the individual steps are

$$\Delta G^\circ = -6.1 + .133ArC - .032ArH + .015Alov + .022\ Remalp + 1.45I \qquad pl \rightarrow gas \quad Step\ 2$$

$$\Delta G^\circ = 5.5 - .086ArC + .046ArH + .009Alov + .004Remalp - 2.62I \qquad gas \rightarrow aq.\ soln. \quad Step\ 3$$

The coefficients of these two steps are complementary and the simplified model for step 1 is a result of the compensatory nature of these two steps. The individual coefficients suggest that for the pure liquid to gas process, increase in the ArC, Alov and Remalp areas increases the free energy of vaporization whereas increasing the ArH area tends to decrease the free energy of vaporization. The reverse is true for the gas phase to aqueous solution process. Here increasing the ArC area reduces the free energy of solvation whereas increasing the ArH area opposes the solvation process. The Alov and Remalp terms contribute to

a much lesser extent. The overall consequence of the different group surface area contributions is that the polymethyl benzenes have a smaller free energy of solvation than the corresponding n-alkyl benzene.

The observed results for the free energies of solvation for these three classes of hydrocarbons have been compared with those calculated from the Scaled Particle Theory (SPT). Although agreement between the observed and calculated results is poor, the SPT reproduces some of the general trends observed in the data. Thus the aromatic hydrocarbons are predicted to be more soluble than the aliphatic hydrocarbons with the same surface area due to the increased free energy of interaction of the aromatic residue with water. In the case of the alkyl aromatics, the methyl substituted benzenes are predicted to have a smaller free energy of solvation when compared with the corresponding n-alkyl benzene. The SPT however tends to overestimate the interaction free energies with increasing solute size.

Finally the results of the surface area analysis of the free energy of solvation data for the aromatic hydrocarbons have been used to estimate the solvent contribution to the free energy of protein-ligand interactions. In particular, the results for the binding of the aromatic inhibitors benzene, naphthalene and anthracene to the active site of α-chymotrypsin indicate that the solvent

contribution becomes progressively unfavorable for this series. Consequently the major driving force for binding likely resides in the nonbonded interactions between the aromatic inhibitor residue and the binding site.

V. BIBLIOGRAPHY

1. J. H. Hildebrand, J. M. Prausnitz and R. L. Scott, "Regular and Related Solutions," Van Nostrand Reinhold Company, N.Y., 1970.
2. Ibid., page 87.
3. R. L. Scott, Ann. Rev. Phys. Chem., 7, 43(1956).
4. C. Black, E. L. Derr and M. N. Papadopoulos, Ind. Eng. Chem., 55(8), 40(1963).
5. A. F. M. Barton, Chem. Rev., 76(6), 731(1975).
6. I. Langmuir, Colloid Symposium Monograph, Vol. III (48), 166(1925).
7. J. A. V. Butler, D. W. Thomson and W. H. MacLennan, J. Chem. Soc., 674(1933).
8. J. A. V. Butler, C. N. Ramchandani and D. W. Thomson, J. Chem. Soc., 280(1935).
9. J. A. V. Butler and C. N. Ramchandani, J. Chem. Soc., 952(1935).
10. J. A. V. Butler and W. S. Reid, J. Chem. Soc., 1171 (1936).
11. J. A. V. Butler, Trans. Faraday Soc., 33, 229(1937).
12. J. L. Copp and J. H. Everett, Discuss. Faraday Soc., 15, 268(1953).
13. G. Saracco and E. Spaccamela Marchett, Ann. Chim. (Rome), 48, 1357(1958).

14. G. J. Pierotti, C. H. Deal and E. L. Derr, J. Am. Chem. Soc., 78, 2989(1956).

15. G. J. Pierotti, C. H. Deal and E. L. Derr, Ind. Eng. Chem., 51, 95(1959).

16. J. C. McGowan, Rec. Trav. Chim. Pays-Bas, 75, 193 (1956).

17. F. Irmann, Chem. Ing. Tech., 37, 789(1965).

18. A. Leo, C. Hansch and D. Elkins, Chem. Rev., 71(6), 525(1971).

19. G. G. Nys and R. F. Rekker, Eur. J. Med. Chem., 8(5), 521(1973).

20. C. McAulliffe, J. Phys. Chem., 70, 1267(1966).

21. R. B. Hermann, J. Phys. Chem., 76, 2754(1972).

22. G. L. Amidon, S. H. Yalkowsky and S. Leung, J. Pharm. Sci., 63(12), 1858(1974).

23. L. B. Kier and L. H. Hall, "Molecular Connectivity in Chemistry and Drug Design," Academic Press, N.Y., 1976.

24. S. S. Davis, T. Higuchi and J. H. Rytting, "Advances in Pharmaceutical Sciences," Ed. by H. S. Bean, A. H. Beckett and J. E. Carless, Academic Press, New York, 1974.

25. C. Hansch, P. P. Maloney, T. Sujita and R. M. Muir, Nature, 194, 178(1962).

26. C. Hansch, R. M. Muir, T. Fujita, P. P. Maloney, C. F. Geiger and M. Streich, J. Am. Chem. Soc., 85, 2817(1963).

27. C. Hansch and T. Fujita, J. Am. Chem. Soc., 86, 1616 (1964).

28. T. Fujita, J. Iwasa and C. Hansch, J. Am. Chem. Soc., 86, 5175(1964).

29. J. Iwasa, T. Fujita and C. Hansch, J. Med. Chem., 8, 150(1965).

30. C. Hansch, J. E. Quinlan and G. L. Lawrence, J. Org. Chem., 33(1), 307(1968).

31. S. H. Yalkowsky in "Design of Biopharmaceutical Properties Through Prodrugs and Analogs," E. B. Roche, Ed., published by American Pharmaceutical Association, Washington.

32. H. H. Uhlig, J. Phys. Chem., 41, 1215(1937).

33. D. D. Eley, Trans. Faraday Soc., 35, 1281(1939).

34. H. Reiss in "Advances in Chemical Physics," Vol. IX, page 1, Ed. by I. Prigogine, Interscience Publishers, N.Y., 1965.

35. R. A. Pierotti, J. Phys. Chem., 67, 1840(1963).

36. R. A. Pierotti, J. Phys. Chem., 69, 281(1965).

37. R. A. Pierotti, Chem. Rev., 76(6), 717(1976).

38. P. R. Philip and C. Jolicoeur, J. Soln. Chem., 4, 105 (1975).

39. G. L. Amidon, S. H. Yalkowsky, S. T. Anik and S. C. Valvani, J. Phys. Chem., 79, 2239(1975).

40. G. L. Amidon and S. T. Anik, J. Pharm. Sci., 65, 801 (1976).

41. D. Sinanoglu in "Molecular Associations in Biology," B. Pullman, Ed., Academic Press, New York, N.Y., 1968, p. 427.

42. J. L. Cohen and K. A. Connors, J. Pharm. Sci., 59, 1271(1970).

43. G. L. Amidon, R. Pearlman and S. T. Anik, in preparation.

44. S. H. Yalkowsky, G. L. Flynn and T. G. Slunick, J. Pharm. Sci., 61, 853(1972).

45. C. Tsonopoulos and J. M. Prausnitz, Ind. Eng. Chem. Fundam., 10(4), 593(1971).

46. J. Hine and P. K. Mookerjee, J. Org. Chem., 40, 3 (1975).

47. P. Mukerjee, Adv. Colloid Interface Sci., 1, 241(1967).

48. R. Wolfenden and C. A. Lewis, J. Theor. Biol., 59, 231 (1976).

49. C. Tanford, "The Hydrophobic Effect," Wiley Interscience, N.Y., 1973, p. 18.

50. R. D. Cramer III, J. Am. Chem. Soc., 99, 5408(1977).

51. S. T. Anik, M.S. Thesis, University of Wisconsin (1976).

52. F. P. Schwarz and S. P. Wasik, Analytical Chemistry, 48, 524(1976).

53. J. A. Pople and D. L. Beveridge, "Approximate Molecular Orbital Theory," McGraw Hill, New York, 1970.

54. D. R. Stull, E. F. Westrum and G. L. Sinke, "The Chemical Thermodynamics of Organic Compounds," John Wiley & Sons, 1969.

55. N. R. Draper and H. Smith, "Applied Regression Analysis," John Wiley & Sons, New York, 1966.

56. H. D. Nelson and D. L. deLigny, Rec. Trav. Chim., 87, 528(1968).

57. J. M. Corkill, J. F. Goodman and J. R. Tate, Trans. Faraday Soc., 63, 773(1967).

58. R. Aveyard and R. W. Mitchell, Trans. Faraday Soc., 64, 1757(1968).

59. E. M. Arnett, W. B. Kover and J. V. Carter, J. Amer. Chem. Soc., 91, 4028(1969).

60. H. Fühner, Ber., 57, 510(1924).

61. W. F. Claussen and M. F. Polglase, J. Amer. Chem. Soc., 73, 1571(1951).

62. G. C. Krescheck, H. Schneider and H. A. Scheraga, J. Phys. Chem., 69, 3132(1965).

63. S. J. Gill, N. F. Nichols and I. Wadsö, J. Chem. Thermodynamics, 8, 445(1976).

64. R. C. Wilhoit and B. Zwolinski, J. Phys. Chem. Ref. Data, 2, Suppl. 1, 1973.

65. S. J. Gill and I. Wadsö, Proc. Natl. Acad. Sci. U.S.A., 73, 2955(1976).

66. R. F. Bohon and W. F. Claussen, J. Amer. Chem. Soc., 74, 1571(1951).

67. D. R. Wauchope and F. W. Getzen, J. Chem. Engg. Data, 17, 38(1972).

68. F. P. Schwarz, J. Chem. Engg. Data, 22, 273(1977).

69. W. E. May, S. P. Wasik and D. H. Freeman, Analytical Chemistry, 50, 175(1978).

70. P. P. Eganhouse and J. A. Calder, Geochimica et Cosmochimica Alta, 40, 555(1976).

71. Z. Yoshida and E. Ōsawa, J. Amer. Chem. Soc., 87, 1467 (1965).

72. H. C. Brown and J. D. Brady, J. Amer. Chem. Soc., 74, 3570(1952).

73. D. Mackay and W. Y. Shiu, J. Chem. Engg. Data, 22, 399 (1977).

74. F. Casellato, C. Vecchi, A. Girell and B. Casv, Thermochimica Acta, 6, 361(1973).

75. J. A. Reynolds, D. B. Gilbert and C. Tanford, Proc. Natl. Acad. Sci. U.S.A., 71, 2925(1974).

76. E. Wilhelm and R. Battino, J. Chem. Phys., 55, 4012 (1971).

77. J. O. Hirschfelder, C. F. Curtiss and R. B. Bird, "Molecular Theory of Gases and Liquids," John Wiley & Sons, New York, 1954.

78. H. D. Nelson and C. L. deLigny, Rec. Trav. Chim., 87, 623(1968).

79. Reference 77, p. 949.

80. J. Janin and C. Chothia, J. Mol. Biol., 100, 197 (1976).

81. G. L. Amidon and R. S. Pearlman, to be published.

82. J. L. Miles, D. A. Robinson and W. J. Canady, J. Biol. Chem., 238, 2932(1963).

APPENDIX A

Solubility, Free Energy, Heats of Fusion and Melting Point Data for the Alkyl and Polycyclic Aromatic Hydrocarbons

	1.22	5.34	5
	1.01	5.45	6
	1.05	5.43	9
	.977	5.47	10
	1.05	5.42	11
	Ave. = 1.06×10^{-4}	Ave. = 5.42	
	S.E. = $.012 \times 10^{-4}$	S.E. = .014	
Ethyl benzene	2.70×10^{-5}	6.23	1
	2.22	6.35	2
	2.85	6.20	12
	2.57	6.21	6
	2.80	6.21	9
	2.95	6.18	10
	2.74	6.22	11
	Ave. = 2.69×10^{-5}	Ave. = 6.24	
	S.E. = $.09 \times 10^{-5}$	S.E. = .021	

	2.65	6.24	11
	Ave. = 3.09×10^{-5}	Ave. = 6.16	
	S.E. = $.174 \times 10^{-5}$	S.E. = .034	
Propyl benzene	8.25×10^{-6}	6.93	1
	8.32	6.93	12
	Ave. = 8.28×10^{-6}	Ave. = 6.93	
	S.E. = $.035 \times 10^{-6}$		
Isopropyl benzene	7.25×10^{-6}	7.01	2
	11.0	6.76	3
	7.5	6.99	6
	9.8	6.83	11
	12.1	6.71	13
	Ave. = 9.52×10^{-6}	Ave. = 6.86	
	S.E. = $.95 \times 10^{-6}$	S.E. = .06	

sec-Butyl benzene	4.16×10^{-6}	7.34	12
	2.36	7.68	11
	Ave. = 3.26×10^{-6}	Ave. = 7.51	
	S.E. = $.9 \times 10^{-6}$	S.E. = .17	
tert-Butyl benzene	4.57×10^{-6}	7.28	12
	3.96	7.37	11
	Ave. = 4.27×10^{-6}	Ave. = 7.33	
	S.E. = $.31 \times 10^{-6}$	S.E. = .064	
1 Methyl 4 isopropyl	4.57×10^{-6}	7.28	14
Naphthalene	4.43×10^{-6}	7.30	3
	4.82	7.25	5
	4.39	7.31	15
	4.40	7.31	16
	4.73	7.21	17
	Ave. = 4.55×10^{-6}	Ave. = 7.29	
	S.E. = $.092 \times 10^{-6}$	S.E. = .013	

2,3 Dimethyl naphthalene	2.30×10^{-7}	9.06	16
	3.47	8.81	18
	Ave. = 2.89×10^{-7}	Ave. = 8.94	
	S.E. = $.59 \times 10^{-7}$	S.E. = .13	
2,6 Dimethyl naphthalene	1.50×10^{-7}	9.31	16
	2.33	9.05	18
	Ave. = 1.91×10^{-7}	Ave. = 9.18	
	S.E. = $.42 \times 10^{-7}$	S.E. = .13	
1,4 Dimethyl naphthalene	1.65×10^{-6}	7.89	This work
	1.14	8.11	18
	Ave. = 1.39×10^{-6}	Ave. = 8.0	
	S.E. = $.26 \times 10^{-6}$	S.E. = .11	
1,5 Dimethyl naphthalene	1.50×10^{-7}	9.31	16
	2.33	9.05	18
	Ave. = 1.91×10^{-7}	Ave. = 9.18	
	S.E. = $.42 \times 10^{-7}$	S.E. = .13	

	S.E. = $.11 \times 10^{-7}$	S.E. = .05	
Pyrene	1.43×10^{-8}	10.7	19
	1.47	10.68	21
	1.56	10.65	10
	1.78	10.57	20
	1.32	10.75	15
	1.20	10.80	18
	Ave. = 1.46×10^{-8}	Ave. = 10.69	
	S.E. = $.08 \times 10^{-8}$	S.E. = .032	
Chrysene	1.18×10^{-10}	13.5	21
	4.73	12.72	10
	1.58	13.37	18
	Ave. = 2.50×10^{-10}	Ave. = 13.20	
	S.E. = $.324 \times 10^{-10}$	S.E. = 0.24	
1,2:5,6 Dibenzo anthracene	3.24×10^{-11}	14.31	21

2. Heats of Fusion and Melting Points

	T_m °K	ΔH_f°	Ref.
Naphthalene	353	4.54	14
2 Methyl naphthalene	308	2.90	22
2,3 Dimethyl naphthalene	375	4.35	This work
2,6 Dimethyl naphthalene	382	5.72	This work
Anthracene	489	6.90	14, 23
9,10 Dimethyl anthracene	455	5.85	This work
Phenanthrene	374	4.46	14
Pyrene	429	3.66	14
Chrysene	531	6.25	23
1,2:5,6 Dibenzo anthracene	544	7.45	23
1,2 Benzopyrene	454	3.96	23
1,2 Benzanthracene	434	5.11	23
Perylene	554	7.59	23
Benzo(g,h,i)perylene	554	4.15	23

3. Vapor Pressures (mm Hg) at 25°C

	P° (mm Hg)	ΔG°_{v}	S.E. (cal.)	Ref.
2,3 Dimethyl naphthalene	4.7×10^{-3}	7.11	12	This work
2,6 Dimethyl naphthalene	5.6×10^{-3}	7.00	10	This work
1,4 Dimethyl naphthalene	1.7×10^{-2}	6.36	7	This work
Anthracene	3.7×10^{-6}	11.33	21	24
Phenanthrene*	1.7×10^{-4}	9.06	10	25
Pyrene*	6.6×10^{-6}	10.99	11	25

*Extrapolated from higher temperature studies.

REFERENCES TO APPENDIX A

1. F. Irmann, Chem. Ing. Tech., 37, 789(1965).
2. L. C. Price, "The Solubility of Hydrocarbons and Petroleum in Water as Applied to the Primary Migration of Petroleum," University of California, Riverside, Ph.D. Thesis, 1973.
3. L. F. Andrews and R. M. Keefer, J. Amer. Chem. Soc., 71, 3644(1949).
4. D. S. Arnold, C. A. Plank, E. E. Erikson and F. Pike, Chem. Eng. Data Series, 3, 253(1958).
5. R. L. Bohon and W. F. Claussen, J. Amer. Chem. Soc., 73, 1571(1951).
6. C. McAuliffe, J. Phys. Chem., 70, 1267(1966).
7. W. F. McDevit and F. A. Long, J. Amer. Chem. Soc., 74, 1773(1952).
8. F. Franks, M. Gent and H. H. Johnson, J. Chem. Soc. (London), 2716(1963).
9. Morrison and Billet, J. Chem. Soc., 3819(1952).
10. H. R. Klevens, J. Phys. Colloid Chem., 54, 283(1950).
11. C. Sutton and J. A. Calder, J. Chem. Engg. Data, 20, 321(1975).
12. L. F. Andrews and R. M. Keefer, J. Amer. Chem. Soc., 72, 5034(1950).
13. D. N. Glew and R. E. Robertson, J. Phys. Chem., 60, 332(1956).

14. C. Tsonopolous, Ph.D. Dissertation, University of California, Berkeley, 1970.

15. R. D. Wauchope and F. W. Getzen, J. Chem. Engg. Data, 17, 38(1972).

16. R. P. Eganhouse and J. A. Calder, Geochimica et Cosmochimica Acta, 40, 555(1976).

17. J. E. Gordon and R. L. Thorne, J. Phys. Chem., 71, 4390(1967).

18. D. Mackay and W. Y. Shiu, J. Chem. Engg. Data, 22, 399 (1977).

19. E. Boyland and B. Green, Brit. J. Cancer, 16, 347 (1962).

20. H. W. Malherbe, Biochem. J., 40, 351(1946).

21. W. W. Davie, M. E. Krahl and G. H. A. Clowes, J. Amer. Chem. Soc., 64, 108(1942).

22. McCullough and Finke, J. Phys. Chem., 61, 289(1957); ibid., 61, 1105(1957).

23. E. Casellato, C. Vecchi, A. Girell and B. Casu, Thermochimica Acta, 6, 361(1973).

24. S. T. Anik, M.S. Thesis, University of Wisconsin, 1976.

25. R. S. Bradley and J. G. Cleasby, J. Chem. Soc., 1690 (1953).

APPENDIX B

Results of Regression Analysis for Models Incorporating Dipole Moments and Polarizability Densities

1. Results of Regression Analysis Using the Model[a]

$$\Delta G = \theta_0 + \theta_1 \text{ArSA} + \theta_2 \text{AlSA} + \theta_3 \text{Dipole Moment} + \theta_4 I$$

Process	θ_0	θ_1	θ_2	θ_3	θ_4	r	s
Pure Liqd → Aq. Soln.	-1.13	.030	.028	.42	-2.04	.99	.18
	(.34)	(.001)	(.001)	(.23)	(.17)		
Pure Liqd → Gas	-8.86	.045	.031	-.69	-.46	.97	.70
	(1.29)	(.004)	(.004)	(.87)	(.66)		
Gas → Aq. Soln.	7.28	-.015	-.003	1.11	-1.58	.96	.61
	(1.14)	(.003)	(.003)	(.76)	(.58)		

[a] r = correlation coefficient, s = standard error, no. of data points = 24. Values in parentheses are standard errors.

2. Results of Regression Analysis Using the Model[a]

$$\Delta G = \theta_0 + \theta_1 \text{ArSA} + \theta_2 \text{AlSA} + \theta_3 \text{Polarizability Density} + \theta_4 I$$

Process	θ_0	θ_1	θ_2	θ_3	θ_4	r	s
Pure Liqd → Aq. Soln.	-1.11	.026	.025	.24	-1.83	.997	.10
	(.19)	(.001)	(.001)	(.03)	(.07)		
Pure Liqd → Gas	-8.93	.036	.024	.58	-.79	.98	.58
	(1.07)	(.005)	(.004)	(.18)	(.42)		
Gas → Aq. Soln.	7.82	-.010	.001	-.34	-1.04	.96	.60
	(1.11)	(.005)	(.004)	(.20)	(.43)		

[a]Values in parentheses are standard errors, r = correlation coefficient, s = standard error and no. of data points = 24.

Polarizability Density = $\frac{\text{Total Polarizability}}{\text{Total Surface Area}} \times 10^2$.

APPENDIX C

Calculation of the Free Energy of Hydration of Benzene by the Scaled Particle Theory

Benzene	Water
$\sigma_2 = 5.24 \times 10^{-8}$ cm	$\sigma_1 = 2.77 \times 10^{-8}$ cm
$\alpha_2 = 10.32 \times 10^{-24}$ cm^3	$\alpha_1 = 1.47 \times 10^{-24}$ cm^3
$\varepsilon_2/k = 531$	$\varepsilon_1/k = 79.3$
$\mu_2 = 0$	$\mu_1 = 1.84$D
	$V_1 = 18.07$

$$y = \frac{\pi\rho\sigma_1^3}{6} = .371$$

$$D = \frac{\sigma_2}{\sigma_1} = 1.89$$

$$\sigma_{12} = \frac{\sigma_1 + \sigma_2}{2} = \frac{(5.24 + 2.77)}{2} \times 10^{-8} = 4.005 \times 10^{-8} \text{ cm}$$

$$\frac{\pi\rho}{6\sigma_{12}^3} = 2.717 \times 10^{44}$$

CAVITY ENERGY

$$\bar{G}_c = -RT \ln(1 - y) + RT\left(\frac{3y}{1 - y}\right)D + RT\left[\frac{3y}{1 - y} + \frac{9}{2}\left(\frac{y}{1 - y}\right)^2\right]D^2 + \frac{y}{\rho} P D^3 \longrightarrow \text{negligible}$$

$$\bar{G}_c = RT\{\ln .629 + \frac{3 \times .371}{.629} \cdot 1.89 + [\frac{3 \times .371}{.629} + 4.5(\frac{.371}{.629})^2](1.89)^2$$

$$\bar{G}_c = 9.33 \text{ kcal}$$

LENNARD JONES DISPERSION ENERGY

$$C_{DIS} = 4 \times \left(\frac{\varepsilon_1}{k}\frac{\varepsilon_2}{K}\right)^{1/2} \cdot \frac{R}{N} \cdot \sigma_{12}^{6}$$

$$= 4 \times (79.3 \times 531)^{1/2} \cdot \frac{R}{N} \cdot (4.005)^6 \times 10^{-48}$$

$$C_{DIS} = 3.387 \cdot \frac{R}{N} \cdot 10^{-42}$$

$$\bar{G}_{DISP} = \frac{-8}{9}\frac{\pi N\rho}{\sigma_{12}^{3}} \cdot C_{DISP}$$

$$= \frac{-8}{9} \times \pi \times \frac{N}{V} \times \frac{1}{\sigma_{12}^{3}} \times 3.387 \times R \times 10^{-42}$$

$$= \frac{-8}{9} \times \pi \times \frac{6.023 \times 10^{23}}{18.07} \times \frac{1}{(4.005)^3 \times 10^{-24}} \times$$

$$3.387 \times 1.987 \times 10^{-42}$$

$$= 9.75 \text{ kcal}$$

INDUCTION ENERGY

$$\bar{G}_{IND} = \frac{-4\pi N\rho}{3\sigma_{12}^{3}}\{\mu_1^{2}\alpha_2 + \mu_2^{2}\alpha_1\}$$

$$= \frac{-4 \times \pi \times (6.023 \times 10^{23})^2}{3 \times (4.005)^3 \times 10^{-24} \times 18.07} \times (1.84)^2 \times 10^{-36} \times$$

$$10.32 \times 10^{-24}$$

$$= 4.57 \times 10^{10} \text{ ergs}$$

$$= -1.09 \text{ kcal}$$

$$\bar{G}_{INT} = \bar{G}_{DISP} + \bar{G}_{IND} = -9.75 - 1.09 = -10.84$$

$$\bar{G}_{SOLV} = \bar{G}_{c} + \bar{G}_{i} + RT \ln \frac{RT}{V}$$

$$= 9.33 - 10.84 + 4.27 = 2.76 \text{ kcal}$$