Chemical Reviews

Volume 93, Number 4 June 1993

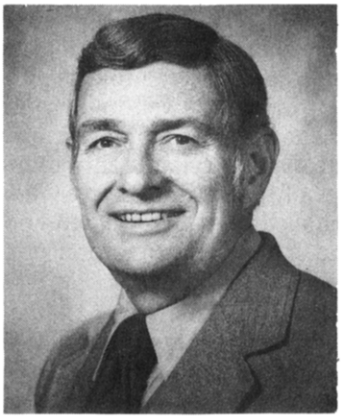
Calculating log P_{oct} from Structures

Albert J. Leo

Medicinal Chemistry Project, Pomona College, Claremont, California 91711

Received May 22, 1992 (Revised Manuscript Received January 22, 1993)

Contents



Albert Leo was born 1925 in Winfield, IL, and was educated in Southern California. He spent two years in the U.S. Army Infantry, serving in the ETO during 1944–5. He received his B.S. in Chemistry from Pomona College (1948; Phi Beta Kappa, Magna Cum Laude) and M.S. and Ph.D. in Physical Organic Chemistry from the University of Chicago, studying reaction kinetics under Prof. Frank Westheimer. After 15 years in industrial research and development in the area of food chemistry, he returned to Pomona College to initiate and direct the Medchem Project under the leadership of his former professor, Dr. Corwin Hansch. The Project provides software and databases useful in the design of bioactive chemicals and is distributed worldwide. His study of partition coefficients as a measure of hydrophobicity resulted in a 1971 paper in *Chemical Reviews* which has become a "Citation Classic". Dr. Leo was given an "Excellence in Science" award by Sigma Xi in 1980 and was chairman of the Gordon Conference on QSAR in Biology in 1981.

I. Introduction

Over 20 years ago an article entitled "Partition Coefficients and Their Uses" appeared in *Chemical Reviews*.[1] It has since became a "Citation Classic".[2] Since a parameter which describes the manner in which a bioactive solute partitions between polar and nonpolar phases has been found to be essential in predicting the transport and activity of drugs, pesticides, and various xenobiotics,[3,4] this review spurred on efforts to calculate an appropriate hydrophobic parameter from structure. The octanol/water solvent system has become the standard model for this work, and use of log P_{oct}, which puts this equilibrium constant in free energy terms, has been widely used in regression analysis of bioactivities. Measurement of this parameter is always recommended, but even though simple in principle, it is time-consuming and impossible, of course, if the prospective compound has yet to be synthesized. Many calculation methods have been proposed since the first by Fujita, Iwasa, and Hansch[5a,b] which was based on the replacement of a hydrogen atom on a parent of known log P. More recent methods are based on the addition of fragment values together with factors which take into account any fragment interaction.[6-13] Entirely different approaches depend more heavily on mathematical techniques; i.e. principal components[14,15] or properties calculated by molecular orbital methods.[16-19] Justification of each of these latter methods has been based on calculation of relatively small sets of solutes, often having rather limited structural variety. This review will analyze them, together with the results of the CLOGP program developed by the Pomona Me-

0009-2665/93/0793-1281$12.00/0

dicinal Chemistry Project, in calculating the values for 8000 structures. These comprise the *entire* current "Starlist" of carefully measured solutes collected by the Project from the literature as well as many unpublished values obtained through private communication and those measured in our laboratories. Using *every* reliably measured value is the ultimate test of a calculation method, and it is the surest way to clearly demonstrate its strengths and weaknesses. We have noted that when CLOGP fails badly, there are strong indications that it is *conformational* information which is lacking. Whenever this can reasonably be inferred from the amount and direction of the deviation of measured vs calculated values, the information may be even more valuable than the hydrophobic parameter itself.

A. Sources of Partitioning Data

Literature sources for the partition coefficients collected in the Pomona Medchem database (MASTERFILE) over a period of 24 years constitute impressive evidence of the widespread utility of the hydrophobic parameter derived from this type of data. Besides the 450 *different* journals represented, there are many books, monographs, and postgraduate theses covering rather specialized subjects. One expects to find this sort of data in the well-known journals devoted to medicinal chemistry, chromatography, pharmacology, pesticide chemistry, and environmental science, but the degree of specialization in each of these areas, with a journal devoted to each subspecialty, is amazing—and it makes acquisition of the data very challenging.

It has been known for some time that hydrophobicity plays an important role in environmental transport and toxicity, and it is no surprise that *Environmental Science & Technology* is a good data source. Other more specialized journals in this area which have contributed to the database are *Bulletin of Environmental Contamination & Toxicology*, *Journal of Great Lakes Research*, *Toxicology & Industrial Health*, *Toxicology & Environmental Chemistry*, *Annals of Occupational Hygeine*, *Ecotoxicity & Environmental Safety*, and *Environmental Chemistry* (China).

Partly due to the pioneering work of Fujita, hydrophobicity has played an important role in pesticide design, and *Pesticide Science* is a prime source of log P_{oct} values. Other more specialized journals in this field which are good sources are *Australian Journal of Plant Physiology*, *Australian Journal of Soil Research*, *Phytopathology*, *Journal of Economic Entomology*, *Pesticide Biochemistry & Physiology*, and *Plant & Cell Physiology*.

Medicinal chemists were quick to make use of partitioning data to help direct their products to the desired sites of action, and thus it is not surprising that the well-known journals in this field have been fertile sources of data. The *Journal of Medicinal Chemistry* and *European Journal of Medicinal Chemistry* are obvious examples which are readily accessible, but others are harder to locate: *Journal of Taiwan Pharmaceutical Association*, *Journal of Labelled Compounds & Radiopharmaceuticals*, *Medical Industry* (China), *Southeast Asia Journal of Tropical Medicine & Public Health*, *International Journal of Immunopharmacology*, *Journal of Italia Medicina Lavoro*, *Japanese Journal of Opthalmology*, *Scandanavian Journal of Rheumatology*, *Archives of Geschulstforshung* (Tumor Research), *Journal of Radiation Oncology*, *Agents & Actions*, *Trends in Pharmaceutical Science (TiPS)*, *Circulation Research*, *Journal of Pharmacobio-Dynamics*, *Toxicology & Applied Pharmacology*, *Klinical Wochenschrift*, *Advances in Pharmacology & Chemotherapy*, *Proceedings of the Society for Experimental Biology & Medicine*, *Journal of Neuropharmacology*, *Bioenergetics*, *Research Journal of Reticuloendothelial Society*, and *Yakugaku Zasshi*. Quite a bit of effort has been devoted to the use of hydrophobicity in predicting both anesthesia and rates of skin penetration. These are reported in *Anesthesiology*, of course, but also in *Anesthesiology & Analgesia Current Research*, *Advances in the Biology of Skin*, and *Journal of Investigative Dermatology*.

We are deeply indebted to colleagues in far-flung places for spotting and translating articles containing very useful measurements. Some of these are *Kakuriken Hokoku*, *Shika Rikogaku Zasshi* (*Journal of Dental Engineering*) *Kagaku Sochi*, *Reports of Yamanouchi Research Labs*, *Bunseki Kagaku* (Analytical Chemistry), *Yaoxue Xuebao* (China), *Yakuzaigaku*, *Acta Sci. Circumstantiae* (China), and *Nehezvegyipari Kutato intezet Kozlemenyei*.

Ph.D. or M.S. dissertations from colleges of pharmacy are well represented in the reference section of the database, such as Berlin Academy Wissenschaften, Rijksuniversity, University of Leyden, University of Nijmegen, and University of Wisconsin. Some of these data "sees the light of day" in accessible journals, but some are published only in the university bulletins. Examples of such bulletins sources are *Bulletin of Meiji College of Pharmacy* (Japan), *Johns Hopkins Hospital Bulletin*, *Bulletin of University of Modena* (Italy), *Bulletin of Moscow University*, *Bulletin of Kyoto University*, and *Bulletin of University Jagiellonskiego*.

The relationship of partitioning and chromatography is well recognized, and the *Journal of Chromatography* ranks high as a source of log P values. Other journals in this vein which must be consulted are *Separation Science & Technology*, *Journal of Colloidal and Interfacial Science*, *Afinidad*, and *Journal of Controlled Release*.

Journals which may not immediately come to mind as good sources of partitioning data, but which, nevertheless appear in our reference section are *Journal of Molecular Catalysis*, *Journal of Molecular Evolution*, *Phosphorus & Sulfur*, *Dyes & Pigments*, *Food & Cosmetic Toxicology*, *Drug & Alcohol Dependency*, *Steroids*, *Precambrian Research*, *Refrigeration Engineering*, *Journal of Lipid Research*, *Radiochemica & Radioanalytical Letters*, and *Chemical Senses*.

Many of the "core" journals are taken by the Pomona Science Library, but we acknowledge the absolutely essential role of the following libraries in assembling the Medchem database: Biomedical Library at the University of California, Los Angeles, the Medical Library at the University of California San Francisco, the Agricultural and Biochemical Library at the University California Riverside, the Linda Hall Library, Kansas City, and the National Library of Medicine, Washington D.C. Since most of the data appearing in

MASTERFILE was *not* simply found under the Chemical Abstract Index under the heading "Partition Coeffecients", we should also mention the invaluable aid of Citation Index in providing the necessary leads.

B. Standard Dataset for Judging Calculations

The 1971 article in *Chemical Reviews*[1] listed roughly 5 000 partition coefficients in hundreds of solvent systems. A later book by Hansch and Leo[20] listed almost 15 000 log P values and proposed a manual method of calculation of log P_{oct} by fragments. It soon became apparent that, with the database growing in size and the calculation method growing in complexity, full computerization was essential. The problem of providing a simple system for unique computer-readable structural entry, suitable for the storage and retrieval of both fragments and complete solute structures, was nicely solved by the development of the SMILES[21] language and CANGEN.[22] The mechanics of overall entry and retrieval were managed by a thesaurus-oriented database system called THOR.[23] By early 1992 the MASTERFILE database contained over 40 000 log P values measured in over 300 solvent systems. The octanol/water system is the most frequent one employed, having 18 000 measured values. Over 25 000 different structures were represented, but a great number of measurements were of "apparent log P" (often called log D) made at a pH where the solute was partly in ionic form. When a reliable aqueous pK_a was available, some of these were "corrected" for ionization, if that amounted to 1.5 log units or less. The assumption was made that the concentration of ionic species in the octanol phase was negligible unless the pH was nearly 4 log units on the ionized side of the pK_a. However, for most ionizable solutes, the values in Starlist are those measured at a pH which suppressed ionization (usually pH 1.0 for acids and 13.0 for bases).

The great majority of the 8 162 select log P values (Starlist) were obtained using the standard shake-flask procedure.[24] Some HPLC values were accepted when the stationary phase was octanol (on a silica support),[25] and when octanol is used in centrifugal partition chromatography (CPC).[26] Reverse-phase HPLC, usually with octadecyl–silyl stationary phase and methanol–water elutant, has been frequently used to estimate log P_{oct} values by regressing relative retention times of a standard solute set against shake-flask measured values. In our opinion, this is only appropriate when the standard set has structural features very similar to the unknowns, and so we rarely included these measurments in Starlist. The one exception is for several tetrahydrocannabinol analogs of very high log P.[27] In this case the standard set contained a number of the most polar analogs (in the log P 2–3 range), and HPLC methodology appeared to be the only procedure to check a calculation procedure for solutes with an expected log P above 9.0. Using the data of Hermens and deBrujin, Leo[28] showed rather conclusively with polychlorinated biphenyls that there was an upper limit of about 8.5 for the partition coefficients measured in the octanol/water system. Most probably this is due to the 10^{-3} M octanol in the water phase which acts as a "detergent" for the extremely small concentration of hydrophobic solute present. It should be noted here that if log P_{oct} is being used to model transport of environmental toxicants, and since aquifers, river water, and even pristine lakes contain natural solubilizing materials, a measured log P may give a more linear response than will CLOGP.

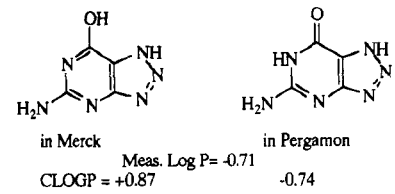


Figure 1. 8-Azaguanine.

A problem, which is often given insufficient attention, concerns tautomers. Most often both tautomeric structures are entered in Starlist, but the data accompanies the one which is considered to predominate, when that is known. In Figure 1, 8-azaguanine in the 6-OH form (as shown in Merck Index, 9th ed.) is calculated by CLOGP as +0.87 but as −0.74 in the 6-one form. The latter is the one shown in the 11th ed. and also in *Comprehensive Medicinal Chemistry* Vol. 6 (Pergamon Press, 1991). Since the measured value is −0.71, we consider the 6-one tautomer to be the effective form as seen in the partitioning equilibrium. There is no reason to suppose that the tautomeric equilibrium cannot be influenced by solvent polarity, however; i.e., the predominant form in water may not be predominant in octanol. The requirement first enunciated by Nernst[29] that a constant partition ratio only applies when considering the *same* species in each phase certainly complicates any method of calculation of tautomers from structure.

The hydrophobicity of organic ions is a matter of great interest to those designing bioactive compounds, but calculation as log P_{oct} is not meaningful unless standard conditions are specified.[30] Even so, the manner in which a positive charge is delocalized over both aliphatic and aromatic environments[20] is still not fully understood and is left for manual estimation. (Values for ion pairs are entered in Masterfile as "Log P-Good".) For zwitterions CLOGP estimates the value where the charges roughly cancel; i.e., the value which would be measured near the isoelectric point. Di-, tri-, and tetrapeptides, both as zwitterions and blocked as acetylamides, are calculated by CLOGP which uses appropriate corrections for the steric effect of the side chains.[31,32]

A complete listing of current Starlist structures and values is too extensive to report here. In a book soon to be published[33] they will appear as a part of the entire Masterfile. Some short lists of measured and calculated values will be found in the following: 67 aromatic amines;[34] 63 nitroaromatics;[35] and assorted CNS agents.[3] Some of the general classes of solutes which often cause difficulties in calculation and which are represented in Starlist are shown in Table I.

II. Results and Discussion

The value of a method of calculating a hydrophobic parameter certainly should be judged by how closely it predicts a carefully measured value. Beyond that,

Table I. Some Classes of Starlist Solutes

aromatic heterocyclics	2 569
pyrimidines	458
purines	56
aliphatic heterocyclics	1 488
steroids	133
polyhalogenated hydrocarbons	>100
phosphates, phosphonates, and amidates	80
β-lactams	59
benzodiazepines	87
aryloxypropanolamines (β-blockers)	44
sulfonureas	34
di- and tripeptides[a]	
"free"	68
"blocked"	59
tetrapeptides[a]	32

[a] To invoke the special peptide corrections, UDRIVE must be entered via SMILES/Macro; e.g. /Gly//Phe/O, where the initiating H_2N is assumed in the first AA designation, but the OH of the terminal carboxyl is specified. An incorrect calculation will result from a normal SMILES entry.

however, the details of the calculation can often reveal some important information about the competing solvation forces as well as suggesting possible conformations of the solute in each phase. While it cannot be denied that a reliable log P_{oct} generated by a "black box" is valuable, knowledge of *why* a certain solute is much more hydrophobic or hydrophilic than expected might be of equal or greater value.

A. Review of Other Methodologies

1. By Substituents

The first widely-accepted methodology for calculating log P was proposed by Fujita, Iwasa, and Hansch in 1964.[5a] They considered log P to be an additive-constitutive, free-energy-related property which was numerically equal to the sum of the log P of the "parent" solute plus a π-term which represented the difference in log P between a particular substituent and the hydrogen atom which it replaced. Thus, π for substituent "X" was defined as

$$\pi_{(X)} = \log P_{(R-X)} - \log P_{(R-H)} \quad (1)$$

It is obvious, then, that *by definition* $\pi_H = 0$. For many log P calculations, the relationship might be expressed as

$$\log P_{(Y-R-X)} = \log P_{(H-R-H)} + \pi_{(Y)} + \pi_{(X)} \quad (2)$$

An example would be

$$\log P_{NO_2-(C_6H_4)-CH_3} = \underset{2.13}{\log P_{C_6H_6}} + \underset{-0.28}{\pi_{NO_2}} + \underset{+0.56}{\pi_{CH_3}} = 2.41; \text{ measured} = 2.45 \quad (3)$$

An elementary consideration, but one often overlooked, is that log P is *not* the sum of π-values. π-values must be *added to a log P*.

Initially the π-system of Fujita et al. was applied only to substitution on aromatic rings, and then only when the hydrogen atom being replaced was definitely of a "hydrocarbon" nature. Some investigators later attempted to apply this methodology to calculations in which the hydrogen was clearly part of a polar moiety, such as a hydroxyl or amine. The results were often, but not always, in error. But even in their first publication on this method, Fujita et al. warned that not even all "aromatic hydrogens" could be substituted without some correction factor. Π for a substituent which is a hydrogen-bond acceptor is more positive when it replaces a hydrogen on an electron-deficient ring than when it replaces one on benzene. An example would be the amino substituent on either nitrobenzene or pyridine. To deal with this clearly electronic effect, the early π-calculation methodology[5a] provided eight sets of π-values which served as models for almost any aromatic system, either electron rich or electron deficient. These will be designated as suggested by Fujita with the substituent followed by a slash and the parent *solute* to which the measurement refers. For example, the substituent constant for the amino group when attached to nitrobenzene is designated: $\pi_{NH_2/PhNO_2}$.

The following are examples of calculations which *incorrectly* use π-values taken from a benzene parent for use on electron-deficient rings:

incorrect

$$\text{(A)}\quad \log P_{3\text{-aminopyridine}} = \underset{0.65}{\log P_{pyridine}} + \underset{-1.23}{\pi_{NH_2/PhH}} = -0.58; \text{ measured} = +0.11 \quad (4)$$

$$\text{(B)}\quad \log P_{3\text{-cyanophenol}} = \underset{2.13}{\log P_{benzene}} + \underset{-0.67}{\pi_{OH/PhH}} + \underset{-0.57}{\pi_{CN/PhH}} = 0.89; \text{ measured} = 1.70 \quad (5)$$

Recognizing that both pyridine and benzonitrile have electron-deficient rings, a better choice of "parent" solute from which to take π-values of the electron-donating substituents would have been nitrobenzene, for which $\pi_{NH_2/PhNO_2} = -0.48$ and $\pi_{OH/PhNO_2} = +0.15$. Then

$$\text{(C)}\quad \log P_{3\text{-aminopyridine}} = 0.65 + (-0.48) = +0.17; \text{ measured} = +0.11 \quad (6)$$

$$\text{(D)}\quad \log P_{3\text{-cyanophenol}} = 2.13 + 0.15 - 0.57 = 1.71; \text{ measured} = 1.70 \quad (7)$$

In 1983 Fujita offered a much more detailed treatment of the "bidirectional" method of calculating log P_{oct} for disubstituted aromatic rings which also included ortho effects.[5b]

2. By Fragments Using C_M

The "fragmental" method of calculating log P_{oct} from structure was pioneered by Rekker and his colleagues.[6,7] Using the largest database of measured values available at that time, they employed statistical methods to determine the average contribution of simple fragments, such as C, CH, CH_2, CH_3, OH, NH_2, etc. It is important to realize that the Rekker methodology gives no clear indication as to what constitutes a valid fragment. If it can be evaluated, it can be used. Although the atomic values for carbon and hydrogen were determined (0.15

and 0.175, respectively), they are not normally used, for then a branching factor would be needed. Only the combined hydrocarbon fragments as shown above will give correct computations. Likewise, for aromatic hydrocarbons, only combined fragments are listed; i.e., C_6H_5 = 1.866; C_6H_4 = 1.688; C_6H_3 = 1.431. Very early in their work Rekker's group appreciated the need to assign different values to each polar fragment depending on whether the carbon atom to which it was attached was aliphatic or aromatic. (Hereafter "A" will be used to indicate the aliphatic attachment of a fragment and "a" the aromatic.) Also they found it necessary to introduce "proximity" corrections if two polar fragments were separated by only one or two aliphatic carbons. Clearly the fragment methodology, like the π-system of Hansch and Fujita,[5] has both additive and constitutive components. Thus Rekker's formula takes the form

$$\log P = \sum a_n f_n + \sum b_m F_m \qquad (8)$$

where "a" is the number of occurrences of fragment "f" of type "n" and "b" is the number of occurrences of correction factor "F" of type "m". Rekker makes the rather daring proposal that all constitutive factors can be directly attributed to a fundamental property of the water in the first solvation shell. For example, the two major corrections (polar proximity and aliphatic vs aromatic bonding) are treated as a product of a "magic constant" (C_M originally = 0.28; revised[7] = 0.22) times a key number, k_n, which is postulated to be the number of "displaced" water molecules.[6] For a proximity with one carbon separation, k_n is taken as 3, and the "ideal" PE-1 would be 3(0.28) = 0.84. The value in Rekker's table[6] is 0.861, which is very close to the average of observed values. For the aromatic–aliphatic difference, k_n is also taken as 3, but this value is actually seen to vary between 1.25 for $N(Me)_2$ and 0.54 for NH_2. This "quantum-displacement" of water molecules seems to bear some similarity to Dunn's ISA/HSA hypothesis discussed later. The precision of the data presently available may not be adequate to clearly support or refute this "quantum" correction hypothesis, because, with very little to direct one's choice of k_n, the maximum deviation in this calculation would be ±0.14 by the original method or ±0.11 by the later; i.e., one-half of C_M. The deviation of measured values, when taken from the variety of sources used by Rekker, is nearly this large. Pursuring this approach even further, van de Waterbeemd and Testa[36] proposed a hydration factor, ω, which is just one-fourth of Rekker's C_M. The physical reality of ω is even more difficult to justify, since the average precision of measurement in a single laboratory with good technique is often ±0.03 or more. Although the concept of a "magic constant" and the "ω factor" have been received with some skepticism, Rekker's calculation methods are widely used and quoted.

Rekker's method can currently be applied with computer assistance.[37] Full computerization would be difficult if not impossible, because the breakdown of solute structure into fragments is a matter of operator choice. One might suppose that it would not matter how this "fragmentation" is accomplished, as long as all the necessary fragment values could be found in the tables furnished. This is not true in all cases, as is illustrated in the following example:

$$\log P_{C_6H_5\text{-}O\text{-}CH_2\text{-}CO_2H} = \underset{1.866}{f_{C_6H_5}} + \underset{-0.433}{f^a_O} + \underset{+0.53}{f_{CH_2}} + \underset{-0.954}{f^A_{CO_2H}} + \underset{+0.861}{\text{PE-1}} = 1.87;\ \text{measured} = 1.34 \qquad (9)$$

In this case a proximity effect (PE) of fixed value, taken from an average of a number of polar fragments in purely *aliphatic* environments, greatly overcompensates. To meet this problem, Rekker adds to his tables a new fragment, $f^a_{-OCH_2CO_2H}$, which combines two simpler polar fragments with the intervening methylene and assigns it a value which takes into consideration an "anomalous" proximity correction. While it might be possible to construct an algorithm which, when the computer was faced with alternate ways of constructing a solute structure from fragments, always picked the largest one, such a method glosses over the apparent fact that the more polar fragments lose more hydrophilicity than the less polar ones when they are placed in close proximity.

Another task for the human operator in Rekker's computer-assisted method, one which often requires a great deal of experience, is the choice of the "key number", k_n, which is the multiplier of the magic constant, C_M. If the key number could be related directly to a scale, such as Wolfenden's hydration potential,[38] and incorporated into the program, it would be a great improvement.

3. By Atomic Contribution and/or Surface Area

One of the first calculation methods to be based on atomic contributions was proposed by Broto and his colleagues.[9] From an early database of about 6000 measured values, they chose 1868 solutes as their basis set. The specific structures were not disclosed, but an important factor restricting this set was their stated inability to deal with any structures containing "potential internal hydrogen bonds". This greatly simplifies calculation problems but excludes a great many solutes of biological interest. Using Monte Carlo and regression methods, they developed a set of 222 descriptors which consisted of combinations of up to four atoms with specific bonding pathways up to four in length. They claimed a precision of about 0.4 log units for this method and found it led to some insights into solvation forces. For instance, they conclude a quaternary carbon is more lipophilic than a teritary, and that an ethylenic carbon (>C=) is actually hydrophilic. The latter is somewhat surprising even if the π electrons of the double bond are credited with some hydrogen-bond acceptor ability. Having selected fragments with bond pathways up to four they were able to see some strong electronic interaction in conjugated systems and some evidence for proximity effects in aliphatic systems.

Ghose and Crippen[10] automated a similar atomic contribution procedure, but reduced the descriptors to 110, while maintaining a standard deviation of 0.4. For hydrocarbons, their method assigns a greater hydrophobicity to hydrogen and lower level to carbon than is usually thought to be the case. This poses no particular problem for the hydrocarbons themselves,

but one questions the insight gained for diethyl ether where the oxygen is considered slightly hydrophobic (+0.04) while the flanking carbons are hydrophilic (−0.95). The work of Kamlet, Taft, et al.[39] and Leahy[40a] clearly established the strong hydrophilic influence of oxygen's lone pair electrons acting as a hydrogen-bond base (i.e. its β strength), and this is indicated by its negative contribution to log P in most methodologies.

Viswanadhan et al.[40b] have extended the atom contribution method of Ghose and Crippen, using 120 atom types with 44 devoted to carbon having different states of hybridization and oxidation. The method still lacks the capability of dealing with any long-range interactions, such as found in *p*-nitrophenol for example, but performs rather well in other instances. The authors clearly point out its present shortcomings, listing 45 solutes where the deviation ranges from 1.0 to 2.8 log units. From a user's viewpoint, the greatest difficulty may arise from some ambiguity in the classification of carbon types and the fact that, through a typographical error, the symbols for aliphatic and aromatic bonds are mislabeled in Table 1.[40b]

An earlier method of log P calculation proposed by Iwase, Moriguchi, et al.[12] was based on solvent-accessible surface areas (SASA) of solutes covered with a 1.4 Å solvent "coat" and designated S_A. In nonpolar solutes, this factor was, of course, positive. On the other hand, polar groups with appreciable surface area make a negative contribution. Their method of evaluating this hydrophilic effect (S_H) is essentially the same as that used by Rekker[6] in evaluating polar fragment values and requires the same distinction between aliphatic and aromatic attachment. For 138 rather simple structures a two parameter equation gave excellent results (r = 0.995; s = 0.13) but it has not been tested with solutes with strongly interacting groups, such as the nitrophenols and nitroanilines, or with intramolecular hydrogen-bonding groups, such as in salicylic acid. How well this method treats proximity effects in aliphatic systems can only be seen in their values for ethylene glycol and 2,3-butanediol. Unfortunately the measured value for the former (−1.92) was taken from an early compilation, and the recent accepted value of −1.36 does not agree well with their calculated value. Repeating their calculation for 2,3-butanediol, using the SAVOL program,[41a] yields a value almost 1 log unit lower than the measured value of −0.92. It should be pointed out that a reduction in SASA has been shown adequate to rationalize the small *reduction* in hydrophobicity of adjoining chlorines in PCBs,[41b] but efforts to explain the opposite kind of proximity effect for polar groups have not been successful, as will be seen in the following paragraphs.

Solute surface area was also deemed important in an approach using principal component analysis developed by Dunn and co-workers.[14,15] An initial analysis of log P data from six different aqueous/nonpolar solvent systems (octanol, ether, chloroform, benzene, carbon tetrachloride, and hexane) indicated that nearly 60% of the variance could be related to the *aqueous* solution properties of the solute, since this principal component was essentially the same for all six systems. This had been noted previously (but only *qual*itatively) and had been ascribed to some aspect of solute size. The unique contribution of Dunn and co-workers was to restrict this factor to the surface area associated with the nonpolar portion of the solute molecule and to provide a means for its calculation by computer.

To compute this "isotropic surface area" (ISA) for a given solute, the first step is to create a model of its hydrated form, which is termed a "supermolecule". The rules for arranging these water molecules on the polar portions of the solute molecule are crucial for the success of the method and are admittedly empirical, but they can take note of hydration information gleaned from X-ray crystallography. Dunn's hydration geometries allow one "fixed" water molecule for solute groups such as nitro, —N= in pyridine, aniline, ketones, and tertiary amines, while two waters are assigned to other amines; a total of three to carboxyls and five to amides. This is in general agreement with Wolfenden's hydration potential scale.[38] Dunn et al. propose that the second principal component, which accounts for another 35% of the variance, relates to the fraction of the total accessible surface area which is hydrated in this manner. This term is designated as f(HSA). Equations for predicting partition coefficients in th six solvent systems are given. The one for octanol is

$$\log P_{oct} = 0.01\ (\pm 0.001)\mathrm{ISA} - 0.26\ (\pm 0.51)\mathrm{f(HSA)} \qquad (10)$$
$$n = 69;\ r^2 = 0.82;\ s = 0.44$$

It is very disturbing to note that the second component, the fraction of the total surface which has polar hydration, has *no* significance in predicting octanol/water log P; that is, the 95% confidence limits cover four times the coefficient for this component. Taken together with the rather high standard deviation, this suggests that the number of waters assigned to a carbonyl group, for instance, may not be the same in acetic acid (entry 8 in Table I, ref 15) as it is in dichloroacetic acid (entry 14). Most certainly the H-bond acceptor strength, β, of the carbonyl oxygen is reduced in the latter. The same problem must exist for the phenol group when comparing the parent (entry 26) with the *p*-nitro analog (entry 38). If the utility of this method is to be extended to complex drug molecules, where the conformation in solution may not be accurately derived from current molecular mechanics programs, the calculation of meaningful ISA and HSA values certainly will face the same problems which all the other methods of calculation from structure must deal: namely, that polar groups that are distant topologically may still influence each other's immediate solvation shell.

A recent effort by Moriguchi et al.[13] uses atom-type descriptors together with factors for proximity effects, unsaturation, intramolecular hydrogen bonds, ring structures, and amphoteric properties. Certain structures require specific descriptors, such as nitro, isocyanato, and β-lactam moieties. Using a large but unspecified selection of structures from the Pomona Masterfile database, these authors present a 14-parameter regression equation which predicts 1230 solutes with a standard deviation, s, of 0.411 and a regression coefficient of 0.952. Without knowing the exact composition of the test set, it is not possible to judge how effectively this method deals with the important electronic, steric, and conformational effects which will be discussed in some detail later. They do

compare their method with that proposed in 1981 by Klopman.[11b] Using the same 195 solutes and 13 parameters, they obtained the same regression coefficient (0.975) as Klopman did with 9 parameters.

In a recent article Klopman and Wang[11a] described their computer-automated structure evaluation (CASE) methodology of log P calculation. This method uses 10 atomic parameters and develops 76 "star-centered" fragments to account for immediate bonding environment. Two indicator variables, one for saturated aliphatic hydrocarbons and one for amino acids, complete the needed parameters. However, for the 935 structures in the test set only 39 of the 88 parameters above were found significant. The statistical results seem quite favorable: $r = 0.965$; $s = 0.385$. Their conclusion that this "approach yielded excellent approximations to the experimental log P values for a rather large database" ... and "seemed to provide accurate log P estimations even for complex molecules" needs to be examined closely, especially when accompanied by the claim that it "is found to give better results than previously described techniques".

Klopman admits that the large indicator variable for saturated hydrocarbons is somewhat of an embarrassment since it results in large negative deviations for short-chain molecules (−0.6 for ethane) and larger positive deviations for the long chain (+0.9 for octane). This would be disturbing enough if it applied only to alkanes, but all homologous series suffer from the same defect; i.e. the deviation range from methanol to octanol also spans 1.5 log units. The CASE methodology performs best on a large set of structures and acts in a "reductionist" [20] fashion. Probably this is why it fails to find a significant hydrophobic contribution by hydrogen. This is indeed disturbing considering the carefully determined log P of hydrogen gas of 0.45. Until these troublesome problems can be resolved, it may be difficult for researchers to place much confidence in the calculations for dicloxacillin, methotrimeprazine, and haloperidol, even though these and 14 others are given as examples of complex drugs which can be predicted with a deviation of less than ±0.5 log units.

4. By Molecular Properties

As every organic chemist knows, a molecule is rarely a simple sum of its parts. For this reason, it would seem that all the methods of calculation from parts—whether the π-methodology of Fujita and Hansch[5a] or any of the fragment or atom-based methods discussed so far—should eventually be displaced by a method based on calculated *molecular* properties. Some experts in the MO field believe that the time has already arrived, since "This approach, using the most advanced methods, like MNDO or AM-1 gives reliable molecular properties including energies, conformations, ionization potentials, and dipole moments." [18] Encouraged, perhaps, by successes in fields other than solvent equilibria, the case for MO has been put even more strongly with the statement that calculations based on the contribution of structural fragments "did have their use 20 years ago" but are "purely empirical, simplistic methods which are based on two-dimensional chemistry" and are now "obsolete".[42] It is worthwhile, therefore, to carefully examine what these advanced MO methods can really offer when the crucial quantity needed is the energy of the *solvated* molecule in each of two immiscible solvents.

Rogers and Cammarata[16] developed a method of calculating the log P_{oct} of aromatic solutes using a charge density term, Q_s^T, together with an induced polarization term, S_s^E. They postulated that partitioning into the aqueous phase was "charge-controlled", while incorporation into the nonpolar phase was "polarizability-controlled". Judging from an admittedly small set of 30 solutes, this method held forth some promise, even though the measured values they used for benzene and carbazole are highly suspect. Apparently the method was not developed any further, and judging from the lack of literature references, is not widely used.

A method of "solvent-dependent conformational analysis" (SCAP) was developed by Hopfinger and Battershell[17] using semiempirical procedures. SCAP could make calculations hundreds or even thousands of times faster than the conventional MO techniques available at that time. The software component, CAMSEQ, requires essentially only a connection table input. For simple aliphatic or aromatic hydrocarbons and for monofunctional solutes, the error in the original SCAP estimations of log P was only slightly greater than with the π-method of Fujita et al.[5] However, the following shortcomings were evident: The size of the 1-octanol molecule makes complete configurational analyses impractical, and so its solvation shell parameters must be estimated by extrapolating those from the lower alkanols. Also, SCAP must neglect the water present in the saturated octanol phase (2 M), and this water surely plays an important role in the overall structure of the nonpolar phase.[43] Finally, SCAP can take into account only the first hydration shell layer, which is probably sufficient for the hydrocarbon portions of a solute molecule but which may be inadequate for strongly polar groups. For whatever reasons, the SCAP method does not work as well as the purely empirical procedures (π or fragment) for solutes containing polar groups that can interact electronically or are in close proximity to each other.

An equation proposed by Bodor et al.[18] in 1989 used 15 parameters and an intercept to calculate log P_{oct} for 118 solutes. These "molecular descriptors" are an indicator variable for alkanes; the molecular weight; the molecular surface area together with its square; the "ovality" of the molecule together with its square; the sum of the absolute values of the charges on oxygen and nitrogen atoms; the square root of the sum of squared charges on oxygen atoms together with the square and the fourth power of that number; a repeat of the last steps for nitrogen; and finally the calculated dipole moment. The equation employing these parameters calculates 118 solutes with a regression coefficient of 0.9388 and a standard error of 0.296. In a later paper Bodor and Huang[19a] calculate a larger set of 302 solutes, with the standard error rising to 0.397 for the additions and to 0.324 33 overall. Adding three more parameters—the number of carbon atoms, the fourth power of ovality, and the sum of absolute values of atomic charges on each atom—reduces the standard deviation to 0.305 79.

Since both the substituent and fragment constant methods of calculating log P were described as "essentially empirical" and "having no scientific basis",[18] it is

Table II. Correlation of log *P* of Alkanes with Size Parameters

no.	compd	*V*-0[a]	*V*-1.5[b]	*A*-0[c]	*A*-1.5[d]	log *P*	est.[e]	Dev.[e]
1	methane	28.22	171.25	47.90	152.24	1.09	1.132	−0.04
2	ethane	44.96	234.12	69.99	191.08	1.81	1.800	+0.01
3	propane	61.59	290.17	91.41	222.55	2.36	2.341	+0.019
4	butane	78.23	346.29	112.86	254.06	2.89	2.883	+0.007
5	pentane	94.89	402.41	134.30	285.58	3.39	3.425	−0.035
6	hexane	111.52	458.53	155.75	317.09	3.90	3.967	−0.067
7	isobutane	78.10	340.74	112.16	247.99	2.76	2.779	−0.019
8	2,3-dimethylbutane	110.16	432.85	147.10	292.92	3.42	3.552	−0.132
9	neopentane	94.49	387.53	132.25	268.76	3.11	3.136	−0.026
10	cyclopropane[f]	54.12	262.69	79.60	206.73	1.72	2.069	−0.349
11	cyclopentane	84.82	359.01	115.30	257.98	3.00	2.950	+0.05
12	cyclohexane	100.94	405.25	133.86	280.77	3.44	3.343	+0.097
13	methylcyclohexane	117.51	484.19	155.09	307.29	3.61	3.799	−0.189
14	cycloheptane	117.47	447.02	152.27	300.05	4.00	3.674	+0.326

[a] Solute volume "bare". [b] Solute volume with 1.5 Å surface. [c] Solute area "bare". [d] Solute area with 1.5 Å surface. [e] Predicted log *P* and deviation from eq 14. [f] Outlier deleted from eqs 11–15.

reasonable to ask, "What sounder scientific basis exists for the 15 'molecular properties' used in the first set or the 18 used in the latter?" The sole justification, it appears, for either set of parameters is through regression analysis, but this hardly qualifies them as nonempirical. To ascertain if a particular parameter has "fundamental significance" and does not merely covary with a significant factor, one can begin by analyzing the simplest of solutes, the alkanes. There is general agreement that the free energy of solvation of an alkane, in either the water or octanol phase, ought to be largely determined by the solute's size, since the London forces acting in the two phases will not be greatly different. (For example, see section II.A.5.) Table II contains log *P* values for the 14 normal, branched, and cyclic alkanes in Starlist, which are considered as reliably measured, and also lists their molecular volumes and surface areas as calculated by SAVOL.[41]

Except for cyclopropane, where ring strain may impart some polarity, alkane log *P*'s correlate very well with *all* measures of solute size, but especially well with the area of the solute as covered with a layer 1.5 Å thick. This is in essential agreement with the findings of Camilleri et al.[44a] who found that calculated surface area took care of branching and cyclization effects. The following regression equations were developed from the data in Table II:

$$\log P = 0.029(0.004)V\text{-}0 + 0.462(0.343) \quad (11)$$

$$n = 13;\ r = 0.981;\ s = 0.166$$

$$\log P = 0.009(0.001)V\text{-}1.5 - 0.285(0.480) \quad (12)$$

$$n = 13;\ r = 0.978;\ s = 0.182$$

$$\log P = 0.024(0.003)A\text{-}0 + 0.059(0.359) \quad (13)$$

$$n = 13;\ r = 0.984;\ s = 0.152$$

$$\log P = 0.017(0.002)A\text{-}1.5 - 1.487(0.450) \quad (14)$$

$$n = 13;\ r = 0.989;\ s = 0.128$$

In these and all following equations "*n*" is the number of data points in the regression; "*r*" is the coefficient of regression; and "*s*" is the standard deviation from regression.

The concentration of cycloheptane could not be measured accurately in the water phase and had to be determined by reextraction. Therefore its measured partition coefficient is not as dependable as the others. If it is also treated as an outlier, the following regression equation is obtained:

$$\log P = 0.017(0.001)A\text{-}1.5 - 1.372(0.275) \quad (15)$$

$$n = 12;\ r = 0.996;\ s = 0.076$$

The positive fragment values assigned to aliphatic carbons and their attached hydrogens in either the Rekker or Leo methodology reflect this unalloyed size effect. In the case of polar fragments with negative values, the effect of the size contribution is outweighed by the greater affinity for water afforded by hydrogen bonding and/or dipolarity. Evaluation of these opposing forces is not possible by the fragment method alone, but promising results are being attained in combination with solvatochromic methodology. (See the next section.)

In addition to the alkane indicator variable and molecular weight, Bodor's method[18] uses two parameters for area and two for shape. With *six* parameters devoted to the nonpolar contribution to octanol/water partitioning (when the data above indicate that *one* is sufficient) it is still true that neither butane nor neopentane are very well predicted (deviation = +0.17 and −0.24), and the deviation for both cyclopentane and cyclohexane are about +0.80. Unless some clear evidence supports a need for the "extra five", it seems Occam's razor should be applied. The poor performance with the simple alkanes also raises some doubts about the claim (ref 18, Table III) that the method has the power to deal with solutes as complex as penicillin, phenytoin, triamcinolone, etc.

Following the SCAP methodology referred to above, some efforts have been made to extend MO methods to include the solvent effects. The work which most directly addresses partitioning phenomena was reported by Jorgensen et al.[44b] Three features of this important work should be carefully noted when considering its relation to calculating log P_{oct}: (1) The nonpolar solvent chosen was chloroform which is much simpler than

octanol since it has no H-bond acceptor capability and contains only 3% as much water at saturation; (2) only the *difference*, $\Delta\log P$, between solute pairs was calculated; and (3) the solute pairs were small (e.g. acetic acid/acetamide or methanol/ethylamine), and generally had only a single polar group; i.e. conformational and group interaction effects could be ignored. These constraints make the problems addressed by Jorgensen orders of magnitude simpler than calculating log P_{oct} for a drug like methotrexate partitioned between octanol and water. It is important to note, therefore, that in referring to calculating absolute log P's, even in the simpler chloroform system, he states that "the computational requirements would be much greater than for the calculation of $\Delta\log P$ for solutes of similar size and the results would be prone to greater imprecision".

It seems accurate to state, therefore, that at this point in time the fundamental properties of solvated structures, as derived from MO calculations, are not sufficiently precise or well-defined that they can be used to dependably calculate octanol/water partition coefficients. It is pertinent, therefore, to ask: How much understanding of the competing forces which determine the hydrophilic/hydrophobic balance can be inferred from any of the techniques discussed so far? To account for the nonadditivity of fragment (or substituent) values, workers using these methods have carefully examined possible contributions from conformation, ionization, hydration, stereoisomerism, ion-pair formation, keto–enol tautomerism, intra- and intermolecular hydrogen-bond formation, folding, etc.[5-8,20,30,45] These have not been "ignored", as some have suggested.[18] On the contrary, considerable effort has been expended by those using either the π or fragment methodologies to verify these "constitutive" contributions by other physical–chemical methods, such as NMR shifts, which measure hydrogen-bonding strength. The exact nature of these "correction factors", which allows for specific ways different fragments can be assembled into a whole solute molecule, may never be verified to everyone's satisfaction. But at least they have always appeared in the computer output,[46] and thus for every solute the user is made aware of a reasonable hypothesis of how a given hydrophobicity level has been attained.

Most researchers agree that molecular conformation, especially for solutes with flexible chains, is difficult to deal with, and we can consider "folding" as just one aspect of the problem. MO calculations are just now attempting to predict the conformation of small flexible solutes when surrounded by water. Extending these to large drug molecules and those surrounded by wet octanol may never be practical. As will be discussed in the next section, partitioning data appears to support the existence of an H-bonded ring (in octanol at least) between a strong H-bond acid (α) and base (β) even when the ring contains as many as 10 atoms. (See Figure 7, parts B and C.) Certainly the "ordinary" MO calculations, which predict molecular properties in a vacuum, would hardly be appropriate to calculate the energy of such a 'folded' conformation when fully solvated by either phase. NMR measurements on pure liquids or in the vapor state are informative, but as yet few if any have been reported for solutions in either solvent.

Even in the case of well-documented H-bonds between ortho substituents on a benzene ring, the resultant effects on the competing solvation forces in water and wet octanol are not predicted with assurance. For example, there is ample evidence that *o*-nitrophenol forms an intramolecular bond in the aqueous phase (e.g., it can be steam-distilled from water solutions). MO calculations are not of much help in telling us why this results in a slight *reduction* in log P_{oct} but over a 3.0 log *increase* in log P_{alkane} (as judged by a comparison of ortho and para isomers). Proponents of the fragment methodology have long recognized that the log P_{oct} for salicylic acid is higher than expected (i.e., higher than the meta or para isomers) and have "empirically" ascribed this to intramolecular hydrogen bonding, even though the O–H–O angle is far from ideal. In the following section on CLOGP results, this will be discussed along with other examples which has led us to propose a mechanism which combines conformational effects which have opposing hydrophobic implications.

Recently Sasaki et al.[19b] proposed a method based on "nonempirical" parameters derived from molecular structure and based on molecular mechanics and molecular orbital methods. To determine accessible surface area, they add 1.4 Å to each van der Waals atomic radius to account for a water layer, but do not describe what method they use to fix the coordinates of flexible structures. They recommend using ab initio molecular orbital calculations to determine electrostatic surface potentials (ESP), but do not mention the difficulties and the time consumed in carrying these out for complex flexible solutes. A parameter for surface tension (S) is calculated in such a way that assigns a zero value to surfaces with ESP below a cutoff of $\epsilon 1$ so that surface tension is only attributed to nonpolar areas of the solute cavity. An electrostatic interaction term (ES) is calculated by multiplying the total sum of ESP value (P_{es}) by the surface area of the polygon on which ESP is greater than a cutoff value, $\epsilon 2$. The third needed parameter, a charge-transfer interaction term (CT), is estimated from "simple perturbation theory" by a very complex equation. It is difficult to see just how CT relates to the solvation forces competing between the octanol and water phases. It contains a term (S_k/S_{k0}) called solvent accessibility, which, it would seem, ought to be covered by S and/or ESP since they both use solvent-accessible surface area. And another term, C_{ip} is "the coefficient of pth atomic orbital when ith molecular orbital is expanded under LCAO approximation". While these terms may indeed be useful in the calculation of log P for a specific structure, they provide little guidance for the drug or pesticide designer who is pondering over what general direction to take in altering his lead structure to bring it into a specific log P range.

Multiple regression analysis is used to relate these three parameters, S, ES, and CT, to log P_{oct}. A regression coefficient of 0.983 and a standard deviation of 0.260 were obtained for 63 solutes. Surface tension (S) was found most significant followed by electrostatic interaction (ES) and charge transfer (CT). This does not readily fit with the widely-held view (see the next section) that log P_{oct} depends mostly on solute size and hydrogen-bond basicity. However, the surface tension term, S, contains a large "size component", and it is

possible that ES and CT might correlate well with hydrogen-bond basicity, β. This is worthy of further investigation.

Sasaki et al. list another 37 solutes where the average deviation in calculation is 0.48. They note that their method does very poorly with benzoic acid (difference = −2.39) and with many heterocycles (difference for uracil = +0.87). They ascribe this to dimerization in the octanol phase. There is ample evidence that dimerization in the octanol phase is negligible[1] and even when it occurs, such as in the benzene/water system, it can be accounted for by the dependence of P on concentration. In comparing their results with CLOGP much is made of the fact that their early version of CLOGP could not complete calculations for 2*H*-1,2,3-triazole and *N*,*N*-dipropylthiocarbamic acid, *S*-ethyl ester. Version 3.4 and higher calculate these solutes very well. Furthermore, v3.5 takes into consideration the lower hydrophobicity of a methyl substituent on an electron-deficient ring, which they have noted occurs in 3-hydroxy-5-methylisoxazole. See section II.B.2 and Figure 4.

5. By Solvatochromic Parameters

The "solvatochromic" approach to log P calculation, as first proposed by Kamlet, Taft, Abraham, et al.[47a,b] was, in essence, a molecular properties methodology. In its simplest form it could be expressed as

$$\log P_{oct} = aV + b\pi^* + c\beta_H + d\alpha_H + e \qquad (15a)$$

where V is a solute volume term; Π^* a term for solute polarity/polarizability; β_H is an independent measure of solute hydrogen-bond acceptor strength; α_H is the corresponding hydrogen-bond donor strength; and e is the intercept. This team's early efforts employed a measured molar volume term, V, to account for the positive contribution of solute size, but this was later replaced by a molecular volume calculated by computer.[40a] In early work, Π^* was obtained from shifts in ν_{max} of nitroaromatic primary indicators.[47b] While there is nothing wrong with a spectrophotometric method in principle, later efforts focused on trying to base Π^* on molecular dipole moment, with Hildebrand's solubility parameter, δ, as a variable correction factor.[48] However, repeated attempts to establish an unequivocal role for dipole moment in eq 15a have been disappointing.[49] Dipole moment in the haloalkanes, where it is largely localized in the carbon–halogen bond, does correlate with the polarity which counteracts the difference in size between the halogen and the hydrogen which it replaces. For monochloroalkanes Kamlet et al.[47c] propose the relationship: $\pi^* = 0.03 + 0.23\mu$; but for aromatic compounds the importance of dipole moment is much less, and the expression becomes: $\pi^* = 0.56 + 0.11\mu$. This reinforces our contention that when the dipole extends over the distance of an aromatic ring it no longer correlates with the polar contribution to log P. Our earliest database contained one of many examples supporting this hypothesis. Measured dipole moments for 1,2-dichlorobenzene and 1,4-dichlorobenzene are 2.27 and 0.0 respectively; yet their measured log P values are, within experimental error, identical (3.43 vs 3.44). The SASA for the 1,2-analog is slightly lower than for the 1,4, and thus would tend to reinforce any log P lowering caused by dipole moment, but none is seen. Further evidence of a poor correlation is apparent when comparing *m*- and *p*-nitroanilines (μ = 4.9 and 6.2, respectively) where the log P values are 1.37 and 1.39; or the *o*- and *p*-nitrophenols (μ = 3.14 and 5.07 respectively) where the log P's are 1.79 and 1.91. Nevertheless, at least in the one instance of phenanthraquinone (discussed in the next section) an increased dipole moment does seem to be the most logical explanation of a negative correction factor needed by CLOGP. It remains to be seen, therefore, if the dipole moment D, as calculated by MO methods,[18] retains its significance when larger sets are investigated.

Abraham has recently proposed[50] that "excess" molecular polarizability, R_2 (but designated here as XMR) may be the proper evaluation of the polarizability component of the original Π^* term. Here "excess" means the greater polarizability exhibited by a polar solute over an alkane of equal size. XMR is a parameter which lends itself to direct calculation. CMR,[51] a companion program to CLOGP, uses a combination of atom and bond polarizabilities to calculate molecular refractivity from the same SMILES structural input. Another simple algorithm calculates McGowen solute volume,[52] and it can produce volume estimations for hypothetical alkanes with a fractional number of carbon atoms. XMR for any polar solute can be calculated in these few steps: CMR calculates the molar refractivity and McGowen volume of the solute; the fractional number of carbons is then calculated for an alkane to equal this volume; and finally the CMR value for that "fractional" alkane is computed and subtracted from the solute CMR. Abraham et al.[53] also have proposed a new procedure for evaluating the polarity portion of the old Π^* term, which he now designates as π_2^H. Polarity is determined from the difference in measured gas/liquid partition coefficients, one liquid being polar and the other an alkane.

It has been fairly well established that the *solvent* β-strength of water and octanol are nearly equal, and therefore the coefficient, d, in eq 15a is close to zero, since they both compete equally well for solute H-donors. (Note this is not true for other partitioning systems, such as chloroform/water or alkane/water where α is about −4.0.) The Π^* term for many solutes appears to be not as important as first believed, and if the XMR portion of it can be readily calculated (see above), this may leave only the solute β_H strength to be evaluated in order to calculate log P_{oct}. Several research groups have made careful measurements of β_H[54] for solutes with single acceptor sites by determining their equilibrium constants with a standard H-donor (e.g. *p*-fluorophenol) in an inert solvent, such as carbon tetrachloride. However, these measurements are at least as difficult and time consuming as a measurement of an octanol/water partition coefficient, and so it is obvious that one would not use measured β_H's in eq 15a as an *easy* path to log P_{oct} values.

Once eq 15a has been well established, its real value will be the understanding it affords of the relative contribution of each type of solvating force: solute size, solute polarizability/polarity, and solute H-acceptor strength. Only when the β_H value of each and every

polar group is known, both in isolation and in combination with others, could eq 15a become an economical way to calculate log P_{oct} values. At the present time, it makes more sense to reverse the calculation order. In other words, the most expeditious route to the determination of the effective $\Sigma\beta_H$ for solutes with multiple interacting polar groups is to back-calculate them from measured octanol/water and alkane/water partition coefficients.[49,55] Gas–liquid chromatography has also recently been proposed[53] as a method for obtaining, not only π_2^H values, but β_H values by "back-calculation". Another approach to its calculation is by a program called SPARC developed by Karcikoff[55b] at the Athens GA, Environmental Research Laboratory. By using a large database of infinite dilution activity coefficients for binary solutions, it shows promise in its early stages, but before it can deal effectively with very complex solutes, it may become as "empirical" as the fragmental approach.

B. CLOGP Methods

1. Implications of Programmed Factors

It is not possible to give a complete description of the CLOGP methodology of calculating log P_{oct} in this review, but the more pertinent features will be explained with examples. Further details can be found in previous publications.[8,30] CLOGP is basically a fragment method which can be expressed in the equation first published by Rekker. See eq 8. This simply states that the log P value of a solute can be estimated from the sum of the contribution of each fragment type, times the number of times it occurs; plus the sum of factors (which account for fragment interaction) times the number of times each factor occurs. The ways that CLOGP accounts for steric, electronic and hydrogen bonding interactions are quite different from Rekker's, and these will be covered with examples in the discussion section which follows.

The CLOGP program follows a set of simple rules to break the solute structure into fragments. Fragmentation is *not* left to user discretion. An "isolating carbon" atom (IC) is defined as one not doubly or triply bonded to a hetero atom. It and its attached hydrogens (ICHs) are considered hydrophobic fragments. All atoms or groups of covalently bonded atoms which remain after removal of ICs and ICHs are polar fragments, even though some, like iodine attached to a phenyl ring, have positive values. It will be noted that a polar fragment contains no ICs but each has one or more bonds to ICs, and these are considered its "environment". This definition results in the following types of fragments: monovalent, —Cl, —CN; divalent, —OC(=O)NH—; trivalent, —OC(=O)N<; tetravalent, >NC(=O)N<. Also note that a carbon can be aromatically bonded to a hetero atom and still be considered "isolating", such as the carbon at the 2-position in pyrimidine.

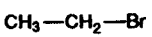
After compiling a list of the fragments it encounters, CLOGP consults the "Fragment Database" to see if all have been evaluated in the required bonding environment. We have preferred a "constructionist" approach; that is giving very heavy weight, when evaluating any fragment, to careful measurements of the simplest solute(s) showing it in the proper environment and when it is the only fragment present. Examples A–D in Figure 2 show this for the bromine fragment in aliphatic (A), benzyl (z), vinyl (v), and aromatic (a) environments. In example C the vinyl bromide fragment value was not determined from the vinyl bromide measurement alone but is a compromise with measurements of other rather simple occurrences of this type. Frequently a fragment has been evaluated only in aliphatic and aromatic environments, which can be considered as the least delocalizing and the most delocalizing, respectively. The other environments are intermediate between these two, and the program is allowed to estimate the other values using a formula provided to it, and it will so indicate this fact in the output.

CLOGP uses the shorthand notation of "X" for halogens and "Y" for fragments capable of hydrogen bonding. The —C≡ of an alkyne is referred to as a "pseudohalogen". Currently CLOGP accounts for six types of fragment interaction in aliphatic systems: X–X, X–Y, and Y–Y with either one or two intervening ICs. Examples E–G in Figure 2 illustrate some of these factors in an approximate representation of the actual computer screen display. This *topological* measure of interaction distance seems to adequately account for polar and dipolar interactions *within* the solute itself, but, as will be noted later, low-energy conformations are possible in which two such groups are much closer through solvent space than is apparent from their topological distance.

Fragment interactions in aromatic systems can operate over relatively large distances, and both electronic and steric effects must be considered. CLOGP uses a modified Hammett[56a] approach in accounting for electronic effects. It consults a table of σ values to assess electron-withdrawing power of any fragment and also a table of ρ values to assess susceptibility to electron withdrawal. It does not differentiate between ortho, meta, and para interactions, and this leads to a certain level of error especially with the fluoro and methoxy fragments. The effect of multiple electronic interactions is not additive, and the present CLOGP v.3.5 allows them to "age" or "fade", starting with the strongest σ/ρ pair of fragments regardless of whether they are present as substituents or as fragments fused in an aromatic ring. It is now apparent that this "fading" should be considered separately for those fragments *on* a hetero ring and those *in* the ring, and this change will be programmed in later versions.

Another proposed revision, which will reduce the number of existing positive deviations, will allow a somewhat reduced electronic factor when the ρ-fragment is on a benzyl carbon. Some electronic effect might be expected in view of the σ° scale defined by Taft which is based on the ionization of substituted phenylacetic acids.[56b] However, the electronically-induced change in pK_a of a carboxylic acid does not truly parallel the change in hydrophilic character of all "Y-type" (H-bonding) fragments. More data is needed to define the effect more exactly, but as can be seen in Table III, part A, it appears that the difference in

A. SMILES: CCBr
NAME: ETHYL BROMIDE

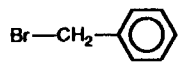


Command? S

PROPERTY	MEASURED	ESTIMATE	ERROR LEVEL
LOGP	1.61	1.605	All Fragments measured

Class	Type	Contrib.Descript.	Comment	Value
FRAGMENT	#1	Bromide	MEASURED	0.200
ISOLATING CARBON		2 Aliph. ICs		0.390
EXFRAGMENT HYDROG.		5 Hydrog. on I.C.		1.135
EXFRAGMENT BONDS		1 chain & 0 alicyclic		-.120
RESULT	v3.4	All Frag. Meas.	CLOGP	1.605

B. SMILES: BrCc1ccccc1
NAME: BENZYL BROMIDE

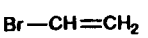


Command? S

PROPERTY	MEASURED	ESTIMATE	ERROR LEVEL
LOGP	2.92	2.924	All Fragments measured

Class	Type	Contrib.Descript.	Comment	Value
FRAGMENT	#1	Bromide	Meas.	0.480
ISOLATING CARBON		1 Aliphatic I.C.		0.195
ISOLATING CARBON		6 Aromatic I.C.		0.780
EXFRAGMENT HYDROG.		7 Hydrogens on I.C.s		1.589
EXFRAGMENT BONDS		1 chain & 0 alicyclic		-.120
RESULT	v3.4	All Frag. measured	CLOGP	2.924

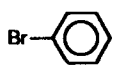
C. SMILES: BrC=C
NAME: VINYL BROMIDE



PROPERTY	MEASURED	ESTIMATE	ERROR LEVEL
LOGP	1.57	1.611	All Fragments Measured

Class	Type	Contrib.Descript.	Comment	Value
FRAGMENT	#1	Bromide	MEASURED	0.690
ISOLATING CARBON		2 Aliphatic I.Cs		0.390
EXFRAGMENT HYDROG		3 Hydrogens on I.C.s		0.681
EXFRAGMENT BONDS		1 chain & 0 alicyclic		-.120
EXFRAGMENT MBONDS		1 double, 0 triple		-.030
RESULT	v3.4	All Frag. measured	CLOGP	1.611

D. SMILES: Brc1ccccc1
NAME: BROMOBENZENE

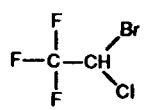


Command? S

PROPERTY	MEASURED	ESTIMATE	ERROR LEVEL
LOGP	2.99	3.005	All Fragments Measured

Class	Type	Contrib.Descript.	Comment	Value
FRAGMENT	#1	Bromide	MEASURED	1.090
ISOLATING CARBON		6 Aromatic I.C.s		0.780
EXFRAGMENT HYDROG.		5 Hydrogens on I.C.s		1.135
RESULTS	v3.4	All Frag. measured	CLOGP	3.005

E. SMILES: FC(F)(F)C(Cl)Br
NAME: HALOTHANE

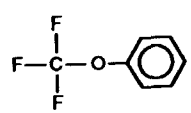


Command? f

PROPERTY	MEASURED	ESTIMATE	ERROR LEVEL
LOGP	2.30	2.447	All Fragments Measured

Class	Type	Contrib. Descript.	Comment	Value
FRAGMENT	#1	Fluoride	MEASURED	-0.380
FRAGMENT	#2	Fluoride	MEASURED	-0.380
FRAGMENT	#3	Fluoride	MEASURED	-0.380
FRAGMENT	#4	Chloride	MEASURED	+0.060
FRAGMENT	#5	Bromide	MEASURED	+0.200
ISOLATING CARBON		2 Aliphatic I.C.s		+0.390
EXFRAGMENT HYDROG.		1 Hydrogen on I.C.		+0.227
EXFRAGMENT BONDS		5 chain & 0 alicyclic (net)		-0.600
PROXIM.	X-C-X	5 interacting fragments		+2.190
PROXIM.	X-CC-X	3 F and 2 non-F frag. 4 used		+1.120
RESULTS	v3.4	All frag. measured	CLOGP	+2.447

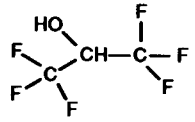
F. SMILES: FC(F)(F)Oc1ccccc1
NAME: BENZENE,TRIFLUOROMETHOXY



PROPERTY	MEASURED	ESTIMATE	ERROR LEVEL
LOGP	3.17	3.17	All Fragments Measured

Class	Type	Contrib.Descript.	Comment	Value
FRAGMENT	#1	Fluoride	MEASURED	-0.380
FRAGMENT	#2	Fluoride	MEASURED	-0.380
FRAGMENT	#3	Fluoride	MEASURED	-0.380
FRAGMENT	#4	Ether	MEASURED	-0.610
ISOLATING CARBON		1 Aliphatic I.C.		+0.195
ISOLATING CARBON		6 Aromatic I.C.s		+0.780
EXFRAGMENT HYDROG.		5 Hydrogens on I.C.s		+1.135
EXFRAGMENT BONDS		4 chain & 0 alicyclic (net)		-0.480
PROXIM.	X-C-X	3 interacting frag.		+1.590
PROXIM.	Y2-X(3	Effect for Y-frag.#4		+1.700
RESULT	v.3.4	All frag. measured	CLOGP	+3.170

G. SMILES: FC(F)(F)C(O)C(F)(F)F
NAME: HEXAFLUORO-2-PROPANOL

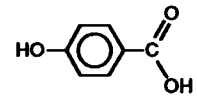


Command? S

PROPERTY	MEASURED	ESTIMATE	ERROR LEVEL
LOGP	1.66	1.592	All Fragments Measured

Class	Type	Contrib. Descript.	Comment	Value
FRAGMENT	#1	Alcohol (Hydroxy)	MEASURED	-1.640
FRAGMENT	#2	Fluoride	MEASURED	-0.380
FRAGMENT	#3	Fluoride	MEASURED	-0.380
FRAGMENT	#4	Fluoride	MEASURED	-0.380
FRAGMENT	#5	Fluoride	MEASURED	-0.380
FRAGMENT	#6	Fluoride	MEASURED	-0.380
FRAGMENT	#7	Fluoride	MEASURED	-0.380
ISOLATING CARBON		3 Aliphatic I.C.s		+0.585
EXFRAGMENT BRANCH		1 Non-X.polar group branch		-0.220
EXFRAGMENT HYDROG.		1 Hydrogen on I.C.		+0.227
EXFRAGMENT BONDS		8 chain & 0 alicyclic (net)		-0.960
PROXIM.	X-C-X	6 interacting fragments		+3.180
PROXIM.	X-CC-Y	6 F and 0 non-F interactions		+2.700
RESULTS	v3.4	All frag. meas.	CLOGP	+1.592

H. SMILES: OC(=O)c1ccc(O)cc1
NAME: P-HYDROXYBENZOIC ACID

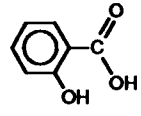


Command? S

PROPERTY	MEASURED	ESTIMATE	ERROR LEVEL
LOGP	1.58	1.557	All fragments measured

Class	Type	Contrib. Descript.	Comment	Value
FRAGMENT	#1	Carboxy (ZW-)	MEASURED	-0.030
FRAGMENT #2	Alcohol (Hydroxy)		MEASURED	-0.440
ISOLATING CARBON		6 Aromatic I.C.s		+0.780
EXFRAGMENT HYDROG.		4 Hydrogens on I.C.s		+0.903
ELECTRONIC SIGRHO		1 Potential; 1.00 used	In Ring	+0.339
RESULT	v3.4	All frag. measured	CLOGP	+1.557

I. SMILES: OC(=O)c1ccccc1O
NAME: SALICYLIC ACID

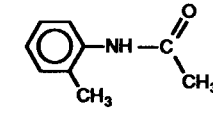


Command? S

PROPERTY	MEASURED	ESTIMATE	ERROR LEVEL
CLOGP	2.26	2.187	All fragments measured

Class	Type	Contrib. Descript.	Comment	Value
FRAGMENT	#1	Carboxy (ZW-)	MEASURED	-0.030
FRAGMENT	#2	Alcohol (Hydroxy)	MEASURED	-0.440
ISOLATING CARBON		6 Aromatic I.C.s		+0.780
EXFRAGMENT HYDROG		4 Hydrogens on I.C.s		+0.908
ELECTRONIC SIGRHO		1 Potential; 1.00 used	In Ring	+0.339
H-BONDING	RING 1	Frag-pair: (1 & 2)		+0.630
RESULT	v3.4	All frag. measured	CLOGP	+2.187

J. SMILES: CC(=O)Nc1ccccc1C
NAME: O-METHYLACETANILIDE



Command? S

PROPERTY	MEASURED	ESTIMATE	ERROR LEVEL
LOGP	0.86	0.900	All Fragments Measured

Class	Type	Contrib. Descript.	Comment	Value
FRAGMENT	#1	NH-Amide	MEASURED	-1.510
ISOLATING CARBON		2 Aliphatic I.C.s		+0.390
ISOLATING CARBON		6 Aromatic I.C.s		+0.780
EXFRAGMENT HYDROG		10 Hydrogens on I.C.s		+2.270
EXFRAGMENT BONDS		1 chain & 0 alicyclic (net)		-0.120
BENZYLBOND SIMPLE		1 Benzyl to simple aromatic		-0.150
ORTHO	RING 1	1 Normal ortho interaction		-0.760
RESULTS	v.3.4	All frag. measured	CLOGP	+0.900

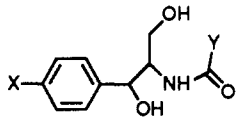
Figure 2. Fragment values for bromine from simplest analogs.

Table III

A. Electronic Factor on Benzyl Fragments

solute	position	measd log *P*	CLOGP	difference
pyridine–methanol	2	+0.06	−0.39	+0.45
	3	−0.02	−0.39	+0.37
	4	+0.06	−0.39	+0.45
pyridine–methylamine	2	−0.21	−0.40	+0.19
	3	−0.32[a]	−0.40	+0.08
	4	−0.38	−0.40	+0.02
nitrobenzyl alcohol	2	+1.24	+0.77	+0.47
	3	+1.21	+0.85	+0.36
	4	+1.26	+0.85	+0.41
nitrobenzyl amine	4	+1.06	+0.84	+0.22

B. Electronic Factors in Chloramphenicols



no.	X	Y	CLOGP dev.
1	$NHCONH_2$	$CHCl_2$	−0.25
2	H	$CHCl_2$	−0.04
3	Ph	$CHCl_2$	+0.03
4	$CONH_2$	$CHCl_2$	+0.06
5	Br	$CHCl_2$	+0.06
6	I	$CHCl_2$	+0.15
7	$COCH_3$	$CHCl_2$	+0.19
8	NO_2	$CHBr_2$	+0.39
9	CN	$CHCl_2$	+0.40
10	NO_2	propyl	+0.45
11	NO_2	$CHCl_2$	+0.45
12	NO_2	$CH(Et)_2$	+0.45
13	NO_2	CH(Cl)Me	+0.48
14	NO_2	CH_2Ph	+0.49
15	NO_2	CH_2I	+0.53
16	NO_2	CH_2Br	+0.55
17	NO_2	$CH(Me)_2$	+0.59
18	NO_2	CF_3	+0.60
19	NO_2	CH_3	+0.61
20	NO_2	CHF_2	+0.61
21	NO_2	CH_2F	+0.62
22	NO_2	CH_2Cl	+0.62
23	NO_2	$C(Me)_3$	+0.73
24	NO_2	CH_2CN	+0.76
25	NO_2	CH(CN)Ph	+0.81

[a] Another apparently valid measurement = −0.10; then difference = +0.30.

sensitivity (ρ) between OH and NH_2 is greater on a benzyl carbon than when they are attached directly on the ring; i.e., where ρ-OH = 1.06 and ρ-NH_2 = 1.00.

Electronic effects acting through a benzyl carbon atom are also apparently responsible for the positive deviation currently noted for most chloramphenicol analogs, as can be seen from Table III, part B.

It will be noted that the unsubstituted phenyl analog (entry 2) is calculated well. On the one hand the electron-releasing ureido analog (entry 1) needs a modest negative factor while the moderately electron-attracting bromo, iodo, or acetyl analogs need small positive factors. This is in sharp contrast to the strongly electronegative CN and NO_2 analogs which need correction factors ranging between +0.4 and +0.8 log units. The deviation for nine of the analogs having comparable electronic aromatic interactions fall within a narrow range of +0.39 to +0.55. The higher deviations are most likely due to the fact that the interaction of the "Y" group with the amido function is slightly underestimated in CLOGP.

Table IV. Group Branching Factors

no.	Frag. Type	log *P* normal	"iso"	"sec"	"tert"
1	**BuOH**	0.88	0.76	0.61	0.35
2	**BuNH_2**	0.97	0.73	0.74	0.40
3	**BuNO_2**	1.47			1.17
4	**BuOCH_3**	1.66			0.94
5	**Bu$OCONH_2$**	0.85	0.65		0.48
6	**BuCO_2H**	1.39	1.16	1.18	1.47
7	**BuCO_2Me**	1.96	1.82		1.83
8	**BuOCOMe**	1.78	1.78	1.72	1.76
9	**BuCN**	1.12	1.07	1.10	1.08
10	Bu(C=O)Bu	2.97			3.00[a]

[a] For di-*tert*-butyl.

Large interactions can often be seen between fragments ortho to one another. They can result in positive corrections if the pair have suitable hydrogen bond donor–acceptor capabilities (Figure 2, example I), or they can be negative if at least one of the pair is "twisted" out of the ring plane and is less delocalized thereby (Figure 2, example J).

A consistent deviation of the first member of any homologous series indicated that chain rotation, which was possible after the first C–C bond, might result in a flexibility that slightly reduced log *P*. This was accounted for in aliphatic systems by subtracting one from the bond count *outside* of the fragments. Flexibility inside the fragments was already taken into account in the fragment value. The flexibility in alicyclics was accounted for with a less negative bond factor.

When branching occurs, either in an alkane chain at ICs (as in isobutane) or at fragments (as in 2-propanol), there appears to be a reduction in log *P* as compared to the straight chain analogs. The early data on alcohols, amines, methyl ethers, carbamates, and nitro compounds led to the conclusion that this effect was general, and it was treated as a chain branch factor of −0.13 and a group branch factor of −0.22. The first five entries in Table IV show that this was a reasonable compromise for the normal and branched butyl analogs containing these fragments. Recently it became apparent that some polar fragments were *not* more hydrophilic when attached to a branching carbon. These are the carboxylic acids and esters, nitriles, and ketones as seen in rows 6–10 of Table IV.

Secondary amines appear to require the branching factor as it is presently programmed:

t-BuNH-*i*-Pr

measured log *P* = 1.56 calcd = 1.56

using $\sum F_{branch} = -0.57$

It appears that, for group branching to require a negative correction factor, the fragment must be attached through a hetero atom, but even then it must have a certain residual H-bond basicity. Support for this postulate would come if the attaching oxygen in the carbamate (entry 5) could be shown to have a higher β-value than that in the ester (entry 8). In any event, it seems apparent that the negative branching factor

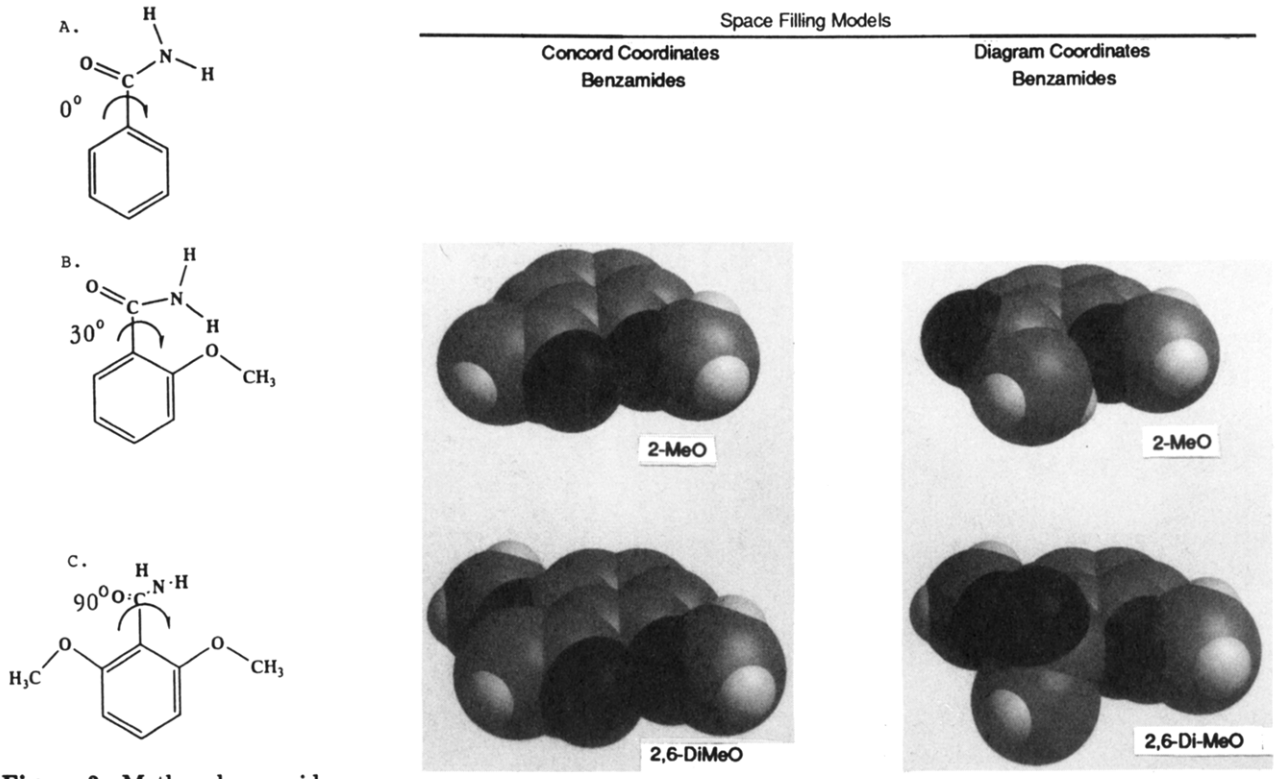


Figure 3. Methoxybenzamides.

required for fragments 1–5 does *not* arise from a decrease in solute size (area or volume).

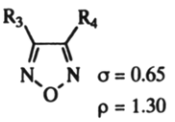
As mentioned in a previous section, the need for a positive H-bond factor in ortho-substituted benzene solutes, such as salicylic acid, was appreciated very early in the development of a calculation methodology based on fragments.[20,45] A 'common type' of H-bond factor was programmed with a value of +0.63, and was applied to occurrences of a carbonyl fragment ortho to either OH or NH. It is very important to note that the factor for log P_{oct} could *not* be based on the strengths of such an intramolecular H-bond as determined by other physical measurements currently available. As noted before, the difference in log P between the ortho and para isomers is a good measure of the H-bond contribution, and for nitrophenol in octanol/water it would be slightly negative (–0.12), while it is strongly positive in much less polar solvent systems: +2.17 in benzene/water; +3.29 in cyclohexane/water; and +3.57 in heptane/water.

2. Implications of Current Deviations

The CLOGP correction of +0.63 for the "common type" intramolecular H-bond is now thought to be the *net* result of two "events": a twisting out of ring plane of the carboxyl group (having a negative effect on log P), but which then allows a stronger hydrogen bond to form which adds as much as 1.0 unit to log P. Further evidence for the dual effect of this kind of ortho substitution comes from 2-methoxy- and 2,6-dimethoxybenzamides.[57] In Figure 3 the first diagram shows the amido fragment in benzamide planar to the ring (0° twist), and in the second, twisted about 30° to optimize a hydrogen bond with the 2-methoxy. With methoxys at both 2 and 6, the third diagram indicates the amido fragment completely decoupled from resonance, making it very much more hydrophilic than the standard taken for aromatic attachment. This is supported by the measured value for 2,6-dimethoxybenzamide which is –0.22 or about 1.0 log unit below that of the 2-methoxy or 4-methoxy analogs.

On a benzene "parent", the methoxy fragment has a π-value of –0.02, but on benzamide in either the 2- or 4-position the π-value is +0.22, the increase arising from an electronic interaction. The observed π-value for the 6-methoxy (with the 2-methoxy already in place) is –1.09. The simplest explanation which best fits this data is that given in the diagrams in Figure 3, with the positive correction needed for the H-bond in the 2-methoxybenzamide balanced by the negative correction needed by amide twisting. Adding the 6-methoxy fragment increases the degree of negative twist and in addition breaks the H-bond, thus creating the need for *two negative* corrections. The result is the unusually low π-6-MeO observed for the 2,6-dimethoxy analog. The space-filled models show that the molecular modeling program, CONCORD,[41] prefers the planar configuration for all three substituents, even though this does press the oxygens' lone-pair electrons close to one another. This may not be energetically unfavorable in the aqueous phase where a rather stable water-bridge may form between oxygen lone pairs, but there may not be extra water available to allow "water bridging" in wet octanol. The right-hand pair of space-filling models depict the configurations for the 2-methoxy and 2,6-dimethoxy analogs with varying twist of the carboxy moiety as indicated by the partitioning data.

It seems difficult to reconcile the configurations developed by AM-1 for the 2-methoxy- and the 2,6-dimethoxybenzamides with the partitioning data cited above. AM-1[58] (ver. 4.10) calculates the carbonyl to be twisted 58.6° out of ring plane for the 2-methoxybenzamide, and the amino group is pointed away from the methoxy so that the O–N distance is 3.8 Å. This is



furazan: log P

#	R_3	R_4	measd	calcd	diff
A	Ph	Ph	3.83	4.36	−0.53
B	Me	Ph	2.59	2.53	+0.06
C	Me	NO_2	0.97	0.95	+0.02
D	NH_2	Ph	1.81	1.88	−0.07

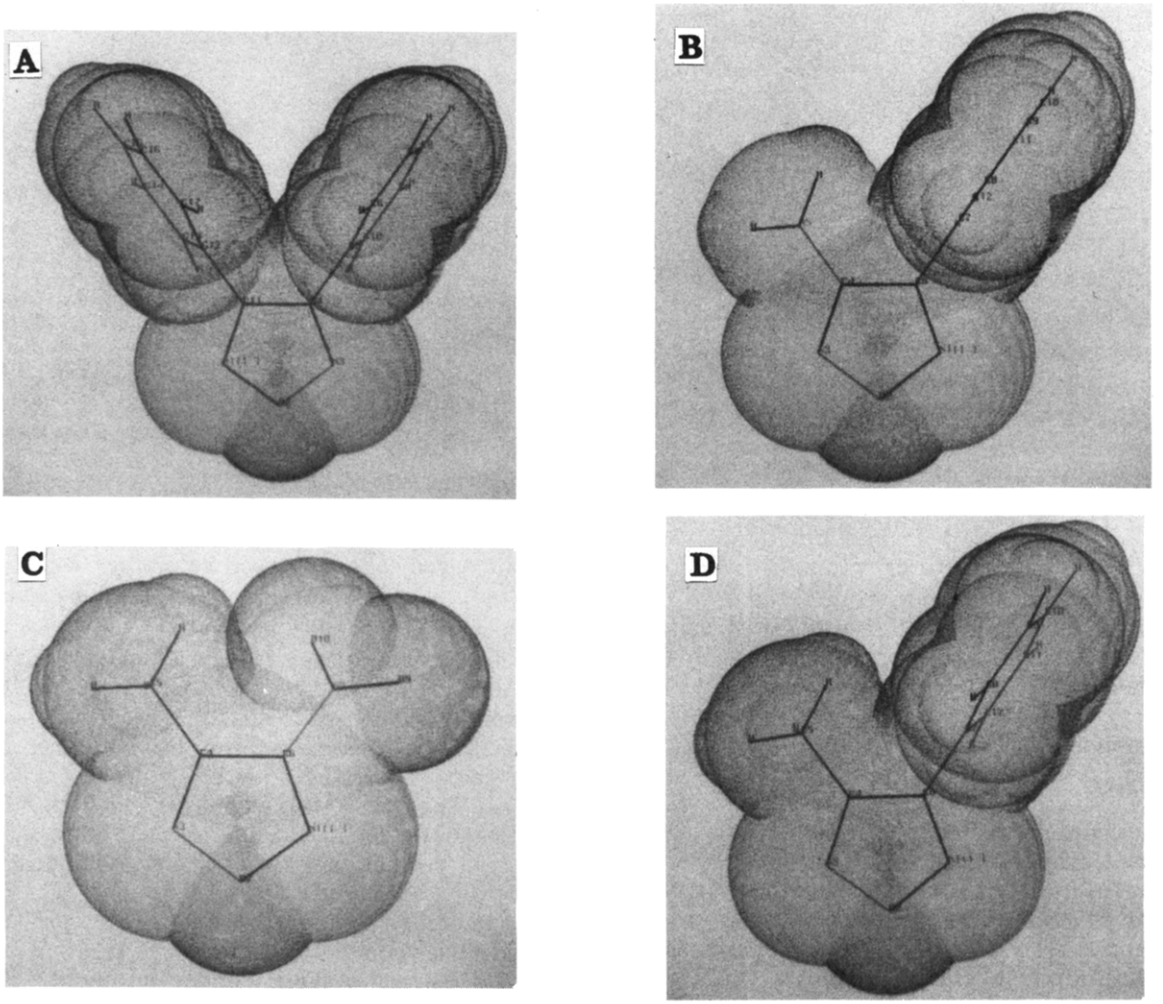
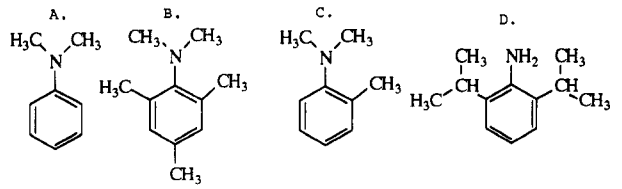


Figure 4. Furazans.

greater than ideal for H-bonding, and so one would have predicted, incorrectly, that a negative factor would be required. For the 2,6-dimethoxybenzamide, the AM-1 coordinates has the carbonyl 32° out of plane and the N–H–O distance at 2.8 Å which is about ideal for H-bonding. From this information one would predict that the calculation should take H-bonding into account and require a positive correction for the 2,6-dimethoxy analog. As noted above, just the reverse is found—it is much more negative than predicted. This is part of the basis for the earlier statement that MO calculations must predict solvent-directed conformations before they can be depended upon for partition coefficient calculations.

Partitioning data can help clarify the electronic nature of some unusual aromatic heterocycles, such as the furazans and furoxans. Figure 4 shows four of the 62 furazans and furoxans reported by Calvino et al.[59] with their measured and calculated log *P* values. Also shown are the space-filling models using coordinates developed by CONCORD.[41] This study confirmed the need of a rather large negative factor (–0.38) to account for the electron release by alkyl groups to such an electron-deficient ring. It also established the furazan fragment as slightly stronger than nitro in electron-withdrawing power (σ = 0.65) but also very sensitive itself to electronegative substituents (ρ = 1.35). As expected, the furoxan fragment is a little more hydrophilic than the furazan (–0.92 vs –0.55) but it is surprising that it seems more sensitive to electronegative substituents (ρ = 1.60). Two interesting inferences regarding conformation can be made from the CLOGP deviations. Attaching one phenyl ring to another, as in biphenyl, extends the aromatic system somewhat, and this is accounted for in CLOGP by a factor of +0.10 for each of the linking carbons. (The same factor is used for the two fusion carbons in naphthalene.) The aromatic extension takes on greater importance when the link is



Log P				
Meas.	2.31	4.41	2.85	3.18
Calc.	2.34	3.83	2.84	3.67
Diff.	-.03	+.58	+.01	-.49

Figure 5. Anilines.

next to an aromatic hetero atom, as in 2-phenylpyridine, and the factor is increased to +0.31. In compound A of Figure 4 the CONCORD coordinates show the two phenyl rings at 90° to the furazan ring, and X-ray crystallography indicates they are twisted at least 80°. This is undoubtedly the reason that CLOGP overestimates the log P, because the phenyl rings probably are much less conjugated than in biphenyl. CONCORD also shows the phenyl ring in compound B at 90°, but in this case the factor for aromaticity extension seems justified, and models indicate that the 3-methyl group does not prevent near planarity. Even with the slightly larger 3-amino group in compound D the aromatic extension factor appears needed, and so the 4-phenyl seems to be nearly planar here also.

Any simple interpretation of AM-1 coordinates and dihedral angles does not help explain partitioning data observed with dimethylanilines. The nitrogen in the parent dimethylaniline is calculated by AM-1 to be trigonal and planar to the ring. See Figure 5A. Adding a 2-methyl group is seen to twist the amine–ring bond by 66° and the nitrogen is now tetrahedral (Figure 5C). With methyls in both the 2 and 6 positions (Figure 5B), the nitrogen reverts to trigonal, but now its plane is 90° to that of the ring. CLOGP uses several simple analogs of *N,N*-dimethylaniline to evaluate the $^{a}f_{N<}$ fragment, but aggrement with measured is, of course, very good (difference = −0.03). As seen in Figure 5C, it is even closer with the *N,N*,2-trimethyl analog (difference = +0.01). This is surprising, since the tetrahedral nitrogen would be expected to have a greater H-bond acceptor strength, β_H, and be more hydrophilic since there would be less delocalization by the phenyl ring. More surprising still is the large positive deviation (+0.58) for the *N,N*,2,4,6-pentamethyl analog. Compared to the parent, from which the fragment was evaluated, one would expect that the 90° twist out of plane would require a negative correction factor. If one wishes to explain the positive deviation by invoking steric blocking of the nitrogen by the hydrophobic methyls, then it must be accepted that this is entirely opposite from the N-unsubstituted aniline where the 2,6-diisopropyl analog (Figure 5D) requires a negative correction. Possibly the moderate H-donor strength of the NH_2 in aniline (α_H = 0.26) is the basis for this difference.

It is certainly possible that other conclusions could be reached in these cases by those more familiar with MO calculations, but it is clear that any simple interpretation can be misleading. At any rate these sorts of problems *are* being addressed in fragment calculation methodology, and it is within the realm of possibility that, for the near future, there may be as much useful information flowing from fragment calculations to MO calculations as the other way around, especially if interest in SCAP[17] is revived.

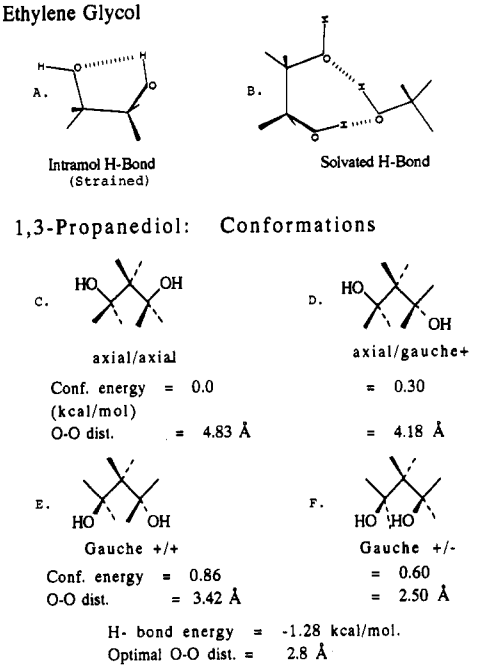
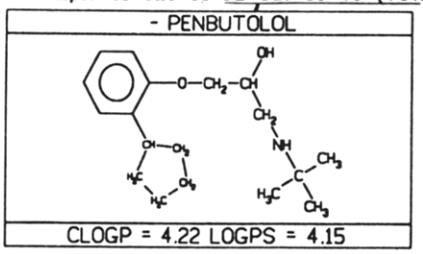


Figure 6. Diols.

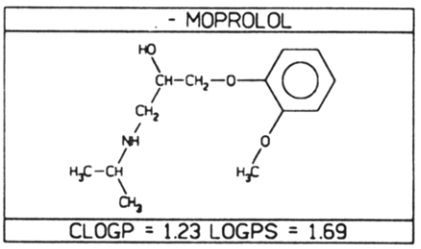
If the effect on log P_{oct} caused by intramolecular H-bonds forming in flexible chains was consistent whenever such an H-bonded ring contained 5 or 6 non-H atoms, programming it in CLOGP would not be a great challenge. However, reality is not that simple. To match the log P of ethylene glycol (−1.36) CLOGP needs a YCCY interaction factor of +0.85. This seems to be the result of a "pure" field/inductive effect of the hydroxyls on each other, making the effective β_H less than additive. An H-bonded ring with only four large atoms (Figure 6A) would be too strained to be stable. Van Duin et al.[60] propose that a solvent-supplied hydroxyl bridge occurs when ethylene glycol is dissolved in a protic solvent, as seen in Figure 6B. However, this would seem to favor solvation by water over octanol, since the former has a much higher concentration of hydroxyl groups. Obviously this is opposite to the observed need of a positive factor, and a decreased $\Sigma\beta_H$ seems a more likely cause than H-bonding.

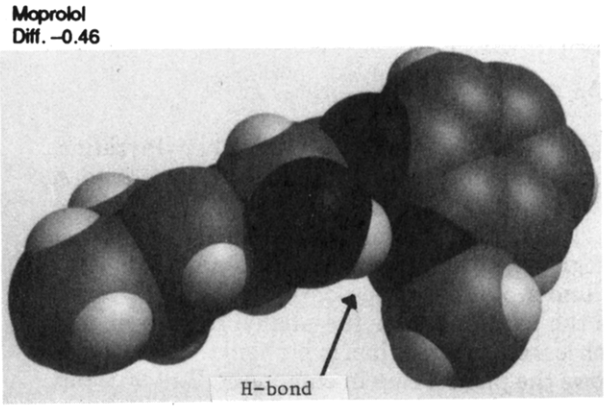
With 1,3-propanediol (M = −1.04; C = −1.69), where a factor of +0.65 is needed and intramolecular H-bonding would form a five-membered ring, the situation is quite different as shown in Figure 6C–F. MM2 empirical force field and the DELPHI molecular mechanics programs gave a modest energy difference[60] of 0.6 Kcal/mol between the axial/axial form (Figure 6C) and the gauche +/− (Figure 6F). One would expect that this difference would be more than made up by a strong H-bond (>1.3 Kcal/mol) since the O–H–O distance is nearly optimal in the latter conformation. Probably the log P factor of +0.65 is the net result of some such trade-off. Even if the basic conformational energies of other 1,3 dipolar propanes could be assumed to be the same as the diol, one would need to know the H-bond donor/acceptor strength of each pair. Evidently 1,3-propanediamine does not form an intramolecular H-bond for its log P is calculated well without any HB factor: measd = −1.43; CLOGP = −1.49.

A. Example of one of 32 out of 46 (70%) which are calculated within ±0.3

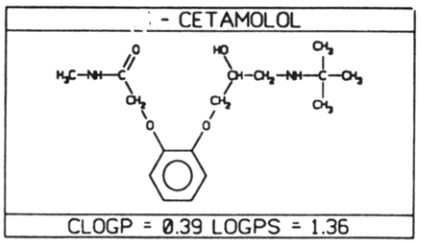


B. Example of one with 7-atom H-bonded ring:





C. Example of one with 'choice' of one 7-atom or two 10-atom H-bond rings



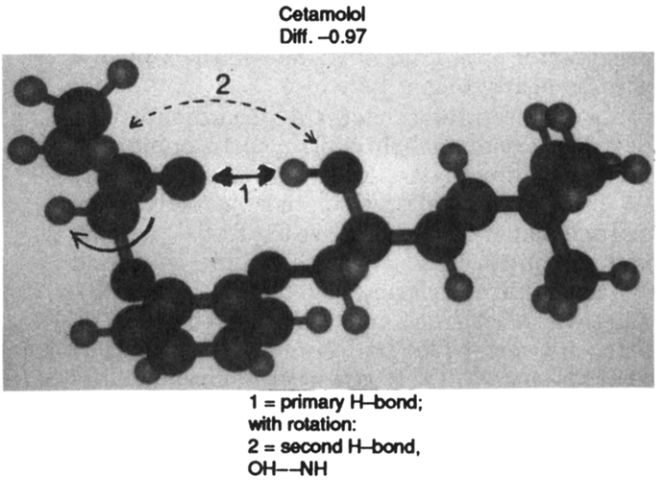
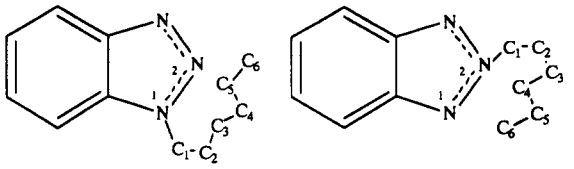


Figure 7. Conformation of (aryloxy)propanolamine β-blockers as inferred from CLOGP deviations.

Other cases of long-range interactions, briefly referred to in an earlier section, involve two flexible chains ortho to each other on a benzene ring and are illustrated in Figure 7. Since 70% of the 46 measured (aryloxy)propanolamine β-blockers[61] are predicted within the limits of ±0.3, one can conclude that most of the normal interaction factors (proximity of the hydroxyl with the oxygen and amide and the branching of the amine) are accounted for in the present version of CLOGP. Some of the analogs with ortho substituents, such as A in Figure 7, are in the well-predicted category but others are not, such as B and C. In every one of the poorly-predicted analogs, the substituent ortho to the oxypropanolamine moiety contains one or more H-bond acceptors. This strongly suggests that when a 7-membered ring can form via H-bonding, such as in moprolol (Figure 7B), a factor of about +0.5 is called for but has *not* been programmed into the current version of CLOGP (v3.5). When such an ortho alkoxy chain also contains other H-bond acceptors and donors, and these offer the possibility of larger H-bonded rings, then the positive factor must be increased. In the case of cetamolol, Figure 7C, the correction is roughly twice that for the single, smaller but more ideally-sized ring. Data from other physical measurements, such as far-IR and NMR, would be welcomed to either support or refute these tentative mechanisms.

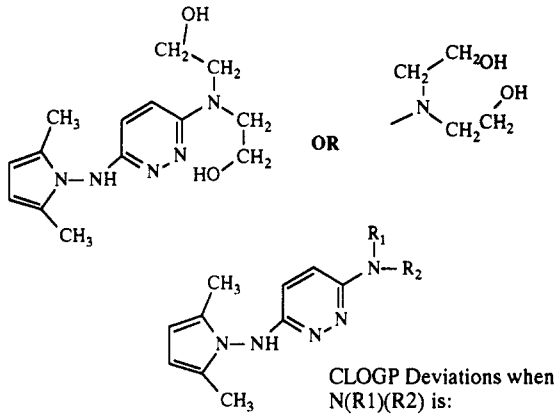
A number of cases have been found where a positive correction factor seems to be required where a flexible chain might be interacting with an aromatic ring heteroatom. An especially clear-cut case involves separate sets of 1- and 2-substituted benzotriazoles[62] illustrated in Figure 8. When the side chain contains no H-bonding group, the CLOGP deviations are negligible. The table lists the interaction factors required when OH or $CONH_2$ is placed at varying distances from the attaching nitrogen. Presently CLOGP invokes no such factor beyond C2, and it is apparent that one is needed even up to C5.



Total Interaction Factor

Interv. Carb.->		C1	C2	C3	C4	C5	C6
1-Chain							
OH	Obs	-	0.89	0.77	0.37	0.32	-0.01
	CLOGP	-	0.87	0	0	0	0
CONH2	Obs	1.18	1.13	0.71	0.39	0.07	-
	CLOGP	1.31	0.97	0	0	0	-
2-Chain							
OH	Obs	-	0.82	0.77	0.42	0.25	0.05
	CLOGP	-	0.75	0	0	0	0
CONH2	Obs	1.06	1.03	0.62	0.33	0.20	-
	CLOGP	1.16	0.86	0	0	0	-

Figure 8. 1- or 2-substituted benzotriazoles.

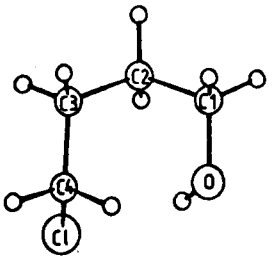


CLOGP Deviations when N(R1)(R2) is:

R1	R2	Dev.
CH_3	CH_3	+0.35
CH_2CH_3	CH_2CH_3	0.00
$-CH_2-CH_2-CH_2-CH_2-$		-0.17
$-CH_2-CH_2-CH_2-CH_2-CH_2-$		-0.05
$-CH_2-CH_2-NH-CH_2-CH_2-$		+0.13
$-CH(CH_3)-CH_2-O-CH_2-CH(CH_3)-$		+0.01
----------	----------	
CH_2-CH_2-OH	CH_2-CH_2-OH	+1.13
$CH_2-CH(CH_3)OH$	$CH_2-CH(CH_3)OH$	+1.08
CH_2-CH_2-OH	$CH_2-CH(CH_3)OH$	+1.14
CH_3	$CH_2-CH(CH_3)OH$	+0.62

Figure 9. H-Bonding in pyridazineamines.

A final example of H-bonding between a donor on a flexible chain and a ring nitrogen is seen in the pyridazinamines[63] of Figure 9. It will be noted that when neither of the two chains on the 3-amino substituent contain OH, the log P value is predicted very well; i.e. average deviation of ±0.07, except for the *N,N*-dimethyl. When one of the N,N chains contains a hydroxyl, the CLOGP deviation is +0.62. Note that this is after it applies the YCCY interaction factor of +0.67, which has proven adequate in hundreds of other instances. When both N,N chains contain a hydroxyl, the positive factor needed is about twice as great even though two YCCY factors are used (+1.34). This is rather compelling evidence for an H-bond between the hydroxyl and the azo ortho to the chain. When two hydroxyls are present, there is not only added opportunity for an O–H–N bond to form but also an O–H–O between chains.



4-Chlorobutanol

$G^-G^+G^+$

Log P	Meas.	CLOGP	Diff.
	0.85	0.54	+0.31

H-Bond?

Figure 10. 4-Chlorobutanol, $G^-G^+G^+$.

In an earlier section it was noted that the topological measure of fragment proximity used by CLOGP was inadequate in some cases of X–Y interaction as well as with the Y–Y interactions just discussed. One example, 4-chlorobutanol, is shown in Figure 10 where a factor of +0.31 is needed. Bastiansen et al.[64] used electron diffraction in the gas phase and MM2 calculations[65] to see if the only one of 14 conformers which would be favorable for an O–H···Cl bond ($G^-G^+G^+$) was actually present in greater proportion than expected if calculated without such a bond. The answer was "yes", and they concluded that the conformation shown in Figure 10 was favored in the gas phase. Perhaps it is also favored in the octanol phase and thus leads to a higher concentration there than if both polar moieties were independently exposed. This rationalizes the need for the positive factor, but it does not help much in quantifying it for use in real time log P calculations.

Early in this section it was stated that once the aliphatic vs aromatic difference in fragment value was known the intermediate environments—benzyl, vinyl, and styryl—could be estimated by the computer. In a few special cases a single value for all kinds of aromatic attachments may not be adequate. For example, benzophenone is taken as the prototype structure to evaluate the carbonyl fragment in a diaromatic environment. However, doing so takes it in its most hydrophilic condition. See Figure 11A. Both CONCORD and AM-1 show the two phenyl rings at the normal 120° angle, and the carbonyl is 90° to both of them. Notice that the carbonyl fragment in 9-fluorenone is also diaromatic bonded (Figure 11B), but now it and the phenyl rings are planar. In this state one expects the H-bond acceptor strength, β_H, of the oxygen atom to be reduced; i.e. the planar carbonyl should be more hydrophobic. This is reflected in the deviation in the calculation which amounts to +0.53. Note that the calculated dipole sum is nearly the same for both, but if it had an effect, it would render 9-fluorenone *less* rather than more hydrophobic than calculated. Undoubtedly the same situation is responsible for the +0.77 deviation in the CLOGP calculation for anthraquinone (not shown). It is not readily apparent how the atomic charge or dipole information derived from AM-1 can be used as a *quantitative* predictor of this effect. It surely is simpler (and perhaps safer) to assign a more positive fragment value to the diaromatic carbonyl when it is in a ring than when it is not.

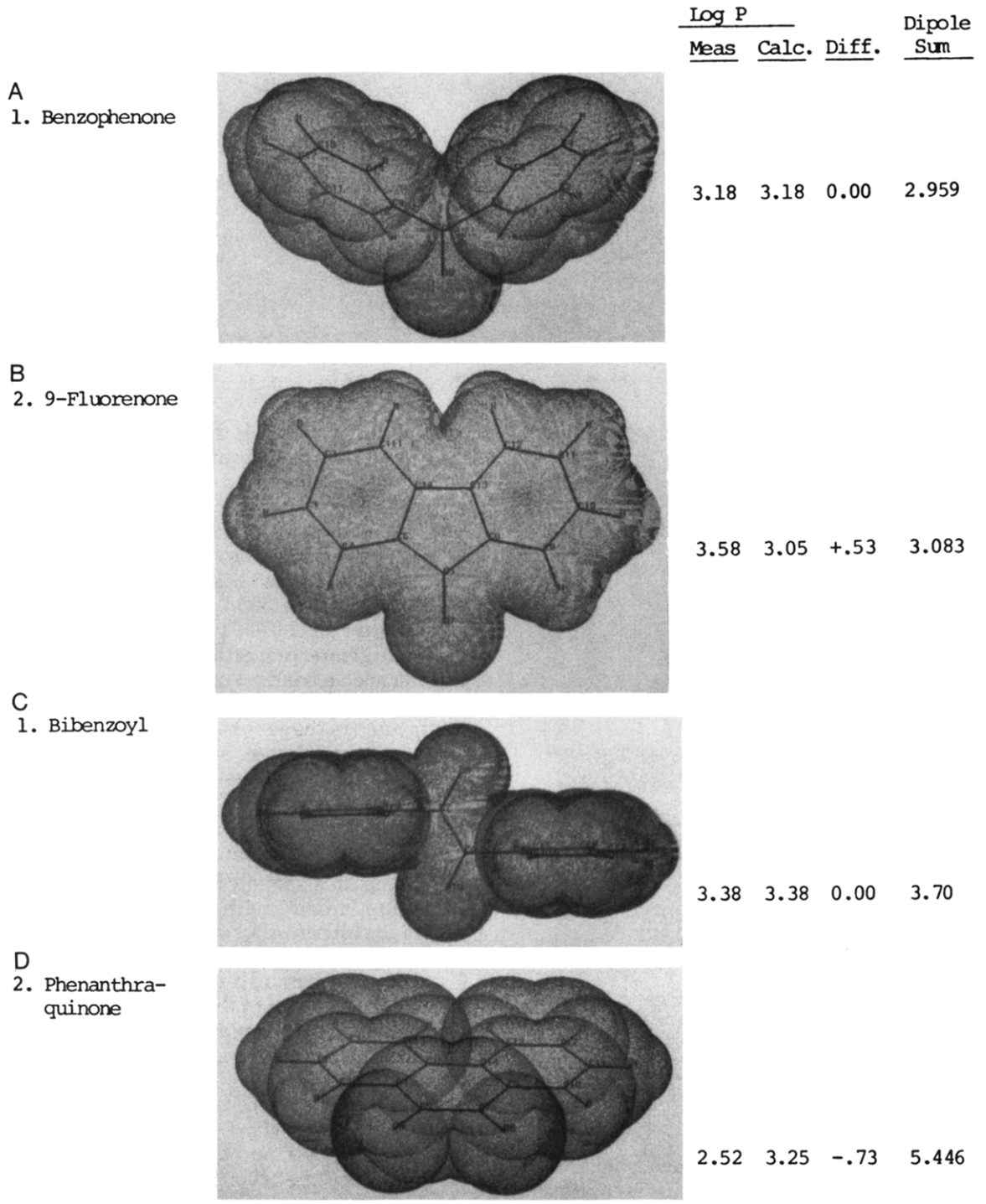


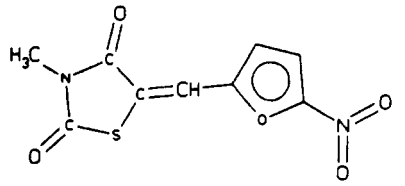
Figure 11. Diaromatic fragment evaluation.

A somewhat different problem seems to be involved with the diaromatic diketone fragment. It has been evaluated from bibenzoyl, where CONCORD and AM-1 produce the reasonable conformation shown in Figure 11C. Here the carbonyls are both at 90° to the ring to which they are bonded and should be in their most polar configuration. In phenanthraquinone (Figure 11D) the carbonyls are in the plane of the ring, and presumably their polarity is more delocalized. Nevertheless the CLOGP deviation is not positive, as in the case of fluorenone discussed above, but is −0.73. Phenanthraquinone has a very high dipole moment (measured in benzene = 5.33μ), about 1.5μ higher than that of bibenzoyl. (In the liquid state, bibenzoyl's dipole moment is very low 0.35μ). It would appear in this case that dipole moment *does* play a role in overcoming what otherwise would be expected as a positive factor when the carbonyl groups are made to be planar with the ring.

3. CLOGP Results

In almost every instance, a solute whose measured log P appears in Starlist will have its fragments evaluated and listed in the Fragment Database. Exceptions occur when the newly-encountered fragment is conjugated with a strongly electrophilic group whose σ value[66] has not been reliably determined. In such cases the interaction factor could influence the fragment value by a log unit or more. One of 166 such instances where

SMILES: CN2C(=O)SC(=Cc1ccc(o1)N(=O)=O)C2=O
NAME:



MASTER
1 SMILES
1 WLN
1 NAME
1 TIMESTAMP
1 LOCAL NAME
1 LOGP
1 LOGPSTAR
1 CMR
1 MOLFORM

Command? s

PROPERTY	MEASURED	ESTIMATE	ERROR LEVEL
CLOGP	1.78	(ERR)	INVALID due to missing fragment value

Class	Type	Log(P) Contribution Description	Comment	Value
FRAGMENT	# 1	--> AN1C(=O)YSC1=O	MISSING	0.000
FRAGMENT	# 2	Aromatic oxygen	MEASURED	-0.110
FRAGMENT	# 3	Nitro	MEASURED	-0.030
ISOLATING	CARBON	3 Aliphatic isolating carbon(s)		0.585
ISOLATING	CARBON	4 Aromatic isolating carbon(s)		0.520
EXFRAGMENT	BRANCH	1 chain and 0 cluster branch(es)	(Chain)	-0.130
EXFRAGMENT	HYDPOG	6 Hydrogen(s) on isolating carbons		1.362
EXFRAGMENT	BONDS	1 chain and 2 alicyclic (net)	(COMBINED)	-0.300
RESULT	v3.4	INVALID due to missing fragment value	CLOGP	1.897

Figure 12. Missing fragment.

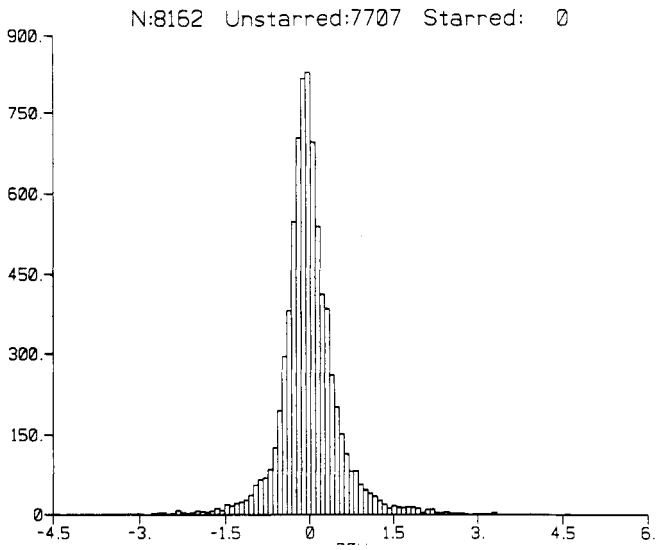


Figure 13. Distribution of deviations.

there is a Starlist value but CLOGP still reports a "missing fragment" is shown in Figure 12. It is depicted in the format returned by the computer, except on color monitors the missing fragment is shown in red. This is the only appearance of the 2-sulfothiazolidin-4-one fragment, and its susceptibility (ρ) to electronic effects would have to be estimated. Furthermore, the effective electronegativity of the nitro group through the furfurilidene linkage is largely a matter of guesswork. It is possible that either MNDO or AM-1 can provide reliable estimates of this effect when "through-resonance" of this type is involved, but we have not yet successfully applied it in this manner. There are 166 solutes in Starlist which return such a "missing fragment" notice, and their removal reduces the effective list to 7996.

It can be seen from Figure 13 that the deviations produced by CLOGP in calculating the entire Starlist of 8000 solutes forms a near-Gausian distribution centered near zero. Examining 100 or so of the large deviants in both right-hand and left-hand "shoulders" reveals some very interesting information. In the area where CLOGP needs a large positive "correction factor"

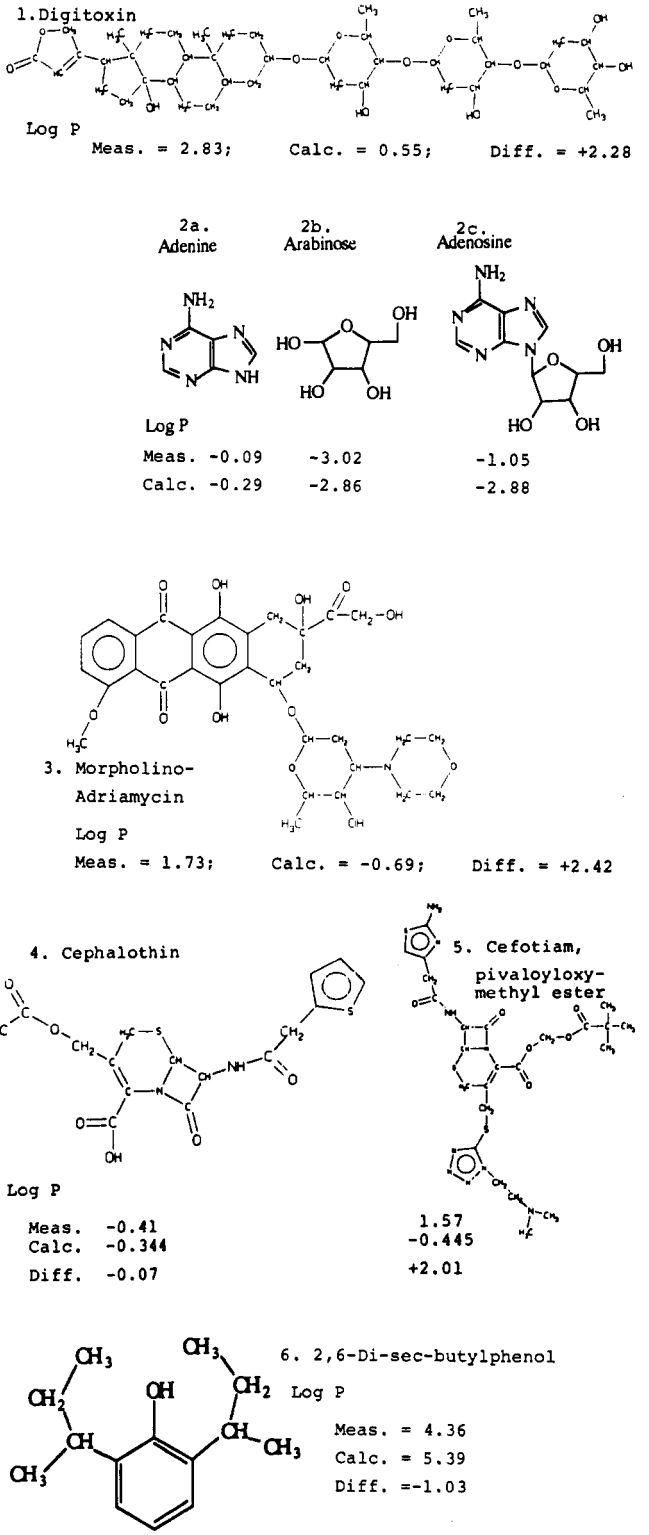


Figure 14. Glycoside and other anomalies.

there are many digitalis analogs which contain a glycoside side chain, as in example 1 in Figure 14. The deviation is not surprising if one is aware of CLOGP's strengths and weaknesses. It has a reasonably good ability to calculate a simple sugar, such as ribose or arabinose, but its performance with sucrose is poor (deviation +2.0). A large positive deviation (i.e. what CLOGP needs to duplicate the measured value) is also seen in the nucleosides, even though the separate base and sugar components calculate well. See example 2, a–c, in Figure 14. This is also the case with the adriamycin analogs where an aminohexopyranose is

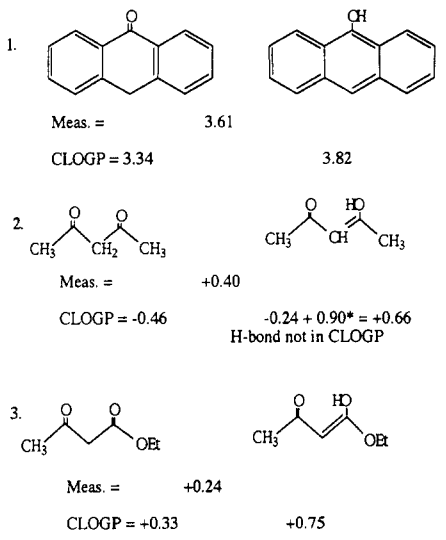


Figure 15. Tautomers.

attached to a tetracycline moiety, as seen in example 3 of Figure 14. This could well be a "folding" problem, and CLOGP could be programmed with an arbitrary correction term to account for its occurrence. In this case it seems preferable to wait until the source of the difficulty is more fully understood.

Further searching the large positive deviants revealed that, although most β-lactam antibiotics are estimated well (Figure 14, no. 4), a dozen cephalosporins were badly calculated (deviations +1.3 to +2.0; e.g. no. 5, Figure 14). Most of these contain either the pivaloyloxy or the tetrazole moiety, or both. A simple tetrazole is calculated well (deviation −0.14), but, as noted in the previous section, we have recently discovered that the negative "group branching factor" for acids and esters is inappropriate. Pivalic acid needs a correction of +0.47, since CLOGP has applied −0.35 in branching factors. These facts are still insufficient to account for the magnitude of the deviations of cephalosporins containing these moieties. Although we could insert an arbitrary correction into CLOGP, we prefer wait until we see how such a factor can be given a more general rationalization.

Tautomerism seems to be a further cause of positive deviations. In most cases of keto/enol tautomerism the keto form predominates in water and is the one commonly depicted. The enol form is generally more lipophilic and may exist in the octanol phase at higher levels than expected. The CLOGP values for the two forms generally bracket the measured value found in STARLIST, but only the calculation for the keto forms are included in the statistical study. In Figure 15, example 1, the measured value of 3.61 for 9-anthrone is consistent with the presence of a sizeable fraction of 9-anthracenol, since the CLOGP values are 3.34 and 3.82 respectively. In Figure 15, example 2, the measured log P for acetylacetone is seen to lie between the CLOGP value for the dione structure and the hydrogen-bonded vinyl alcohol. Since the vinyl alcohol fragment value cannot be determined in isolation, the latter CLOGP calculation is only approximate, and it cannot be used to reliably estimate the percentage of enol in either phase. In the case of ethyl acetoacetate, Figure 15, example 3, the calculation as dione is slightly higher than the measured, and this agrees with the finding that little or no enol is present since that structure calculates much higher.

Removing such highly deviant structures for which there was an apparent rationale (although not necessarily one which was easily converted to an algorithm) the "reduced" Starlist contained 7800 selected log P_{oct} values. Regressing the measured values, log P^* against CLOGP gives

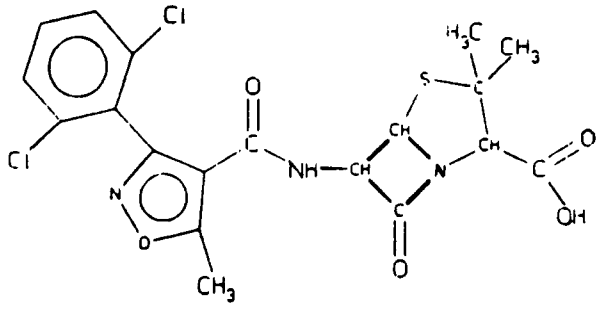
$$\log P^* = 0.903(\pm 0.005)\text{CLOGP} + 0.209(\pm 0.012) \quad (16)$$

$$n = 7800;\ r = 0.970;\ s = 0.398$$

Although a standard deviation of under 0.4 is quite remarkable for such a large and diverse solute set, it is a little surprising to see a slope appreciably less than 1.0 and an intercept of over 0.2. It first appeared that this might have been the result of the fixing of carbon and hydrogen fragment values early on in a "contructionist" approach,[20] using very careful measurements of the simplest sort of solutes. Indeed, changing a few of the "fundamental" values for aromatic carbon (0.13 to 0.15), for hydrogen (0.227 to 0.217) and reducing the bond factors by 0.01 did reduce the intercept to zero (within the 95% confidence limits) while maintaining the r and s values. From a statistician's viewpoint this would seem justified, even though it does raise the deviation of the simpler structures where the prediction was previously excellent and makes the measured values more solid. Thinking as physical chemists, this hardly seems like an improvement, and furthermore, looking at the nature of the solutes with high residuals, one can reach another conclusion. There are more solutes with large positive corrections needed as compared to large negative ones. The large positives are almost always associated with structures in which there seems to be long-range fragment interactions (probable H-bonding)—a condition with which CLOGP, using a topological measure of distance, cannot deal at present. (See examples in Figures 7–9). At such time that CLOGP can be integrated with programs which develop *solvent-directed* three-dimensional conformations, it may be possible to include correction factors which will partly or completely eliminate this intercept.

An example of a solute with a large negative deviation (−1.03) is 2,6-di-*sec*-butylphenol, which is shown as no. 6 in Figure 14. A steric effect resulting in a negative deviation (−0.49) was previously noted in 2,6-diisopropyl aniline (Figure 5D). One might first assume that a nonpolar blockage of the polar hydrogen-bonding fragments in these solutes would *raise* log P, but of course the partition coefficient is a result of *competing* solvent forces. In 2,6-di-*sec*-butyl phenol a SAVOL[41] calculation indicates the solvent accessible surface area (SASA) of the hydroxyl has been reduced to 48% of what it is in the parent phenol. This may not "penalize" solvation by water as much as it penalizes octanol. Another apparent example of this phenomenon is seen in comparing borneol with isoborneol. SASA for the oxygen in the former is 44% greater than for the latter, but even with more polar surface exposed its log P is 0.40 units higher. Correlating these negative deviations with some parameter which can be calculated (such as SASA) has proved to be difficult, but at least they seem to be of lower magnitude than the positive deviations.

SMILES: Cc1onc(c1C(=O)NC3C2SC(C)(C)C(N2C3=O)C(O)=O)c4c(Cl)cccc4Cl
NAME: DICLOXACILLIN



MASTER
1 SMILES
1 WLN
1 CAS NUMBER
1 NAME
1 TIMESTAMP
1 LOCAL NAME
10 LOGP
1 LOGPSTAR
1 CLOGP
1 CMR
1 MOLFORM
2 OTHER TYPES

Command? f

PROPERTY	MEASURED	ESTIMATE	ERROR LEVEL
CLOGP	2.91	2.734	All fragments measured

Class	Type	Log(P) Contribution Description	Comment	Value
FRAGMENT	# 1	Isoxazolyl	MEASURED	-0.950
FRAGMENT	# 2	NH-Amide	MEASURED	-1.810
FRAGMENT	# 3	Sulfide	MEASURED	-0.790
FRAGMENT	# 4	Amide	MEASURED	-3.190
FRAGMENT	# 5	Carboxy (ZW-)	MEASURED	-1.110
FRAGMENT	# 6	Chloride -	MEASURED	0.940
FRAGMENT	# 7	Chloride	MEASURED	0.940
ISOLATING	CARBON	7 Aliphatic isolating carbon(s)		1.365
ISOLATING	CARBON	9 Aromatic isolating carbon(s)		1.170
FUSION	CARBON	1 Extended aromatic iso-C(s)		0.100
FUSION	CARBON	1 Extended hetero-aromatic iso-C(s)		0.310
EXFRAGMENT	BRANCH	2 chain and 1 cluster branch(es)	(COMBINED)	-0.390
EXFRAGMENT	BRANCH	2 Non-halogen, polar group branch(es)	(Group)	-0.440
EXFRAGMENT	HYDROG	15 Hydrogen(s) on isolating carbons		3.405
EXFRAGMENT	BONDS	1 chain and 7 alicyclic (net)	(COMBINED)	-0.750
BENZYLBOND	MHSMCY	1 Benzyl bonds to multihetero-5-rings		-0.380
PROXIMITY	Y-C-Y	Frags 2 and 4: -.32 (-1.810+ -3.190)		1.600
PROXIMITY	Y-C-Y	Frags 3 and 4: -.32 (-0.790+ -3.190)		1.274
PROXIMITY	Y-C-Y	Frags 4 and 5: -.37 (-3.190+ -1.110)		1.591
PROXIMITY	Y-CC-Y	2 pairs over bond 16-13 (AvWt=-.190)		0.517
ELECTRONIC	SIGRHO	4 Potential interactions; 2.39 used	WithinRing	0.786
ELECTRONIC	SIGRHO	6 Potential interactions; 0.15 used	JoinedRing	0.006
ORTHO	RING 3	2 Normal ortho interaction(s)		-0.960
ORTHO	RING 4	2 Normal ortho interaction(s)		-0.500
RESULT	v3.4	All fragments measured	CLOGP	2.734

Figure 16. Dicloxacillin.

Reducing the Starlist set a little further by removing deviants to the same absolute value level (±1.2) we obtain the following regression for 7500 solutes:

$$\log P^* = 0.911(\pm 0.004)\ \text{CLOGP} + 0.191(\pm 0.011) \quad (17)$$

$$n = 7500;\ r = 0.978;\ s = 0.336$$

As expected, the slope is closer to unity and the intercept slightly reduced. Paring down Starlist still further eliminates other structural features which are more difficult to parameterize, and, of course, improves the statistics:

$$\log P^* = 0.914(\pm 0.004)\ \text{CLOGP} + 0.184(\pm 0.010) \quad (18)$$

$$n = 7250;\ r = 0.982;\ s = 0.300$$

We believe that any calculation method which improves on any of the three equations above merits careful consideration, but it should also be judged on how much other information it provides with each calculation. Take, for example, the details of the calculation for dicloxacillin seen in Figure 16. The six fragment values (isoxazolyl, sulfide, carboxyl, 2 chloros, and 2 amidos—one cyclic) must be appropriate for thousands of appearances among the 7250 other structures in eq 18. The electronic interactions (+0.79) and proximity effects (Σ = +4.982) must also be appropriate for thousands of other pairings. Negative corrections (–0.5 and –0.96) for "twisting" by the two chloros ortho to the isoxazole ring and the methyl ortho to the amido fragment must also work in many other situations. Perhaps the labels which we have applied to these factors are not theoretically justified, and so they remain "empirical" and "simplistic", but they still may be valuable in deciding which few of thousands of possible compounds one should synthesize next. Rationally optimizing the lead structures of promising drugs and pesticides depends heavily on such "fragment" and "interaction-factor" information.

III. Conclusions

Any method of calculating log P_{oct} by summation of fragment values (atomic or larger clusters) *must* include factors which take into account the fact that the polar ones are *not additive* in most instances. If these factors are accounted for by a set of "clusters", then the bond pathways must include those up to five in length to

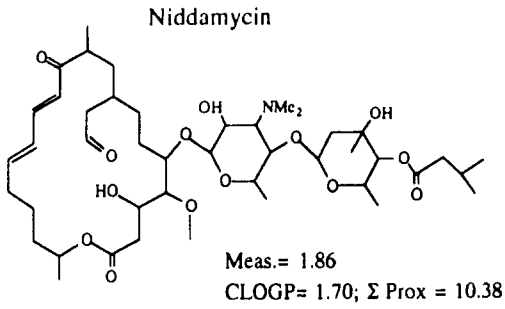


Figure 17. Compensating errors?

account for a simple interaction such as seen in *p*-nitrophenol. To adequately handle a test set of the size presented in eq 16, this would require hundreds or even thousands of "cluster parameters". Such a system would need to be computerized, of course, but even if it delivered accurate predictions it would be more difficult to interpret each component contribution to the final log *P*.

In principle, a knowledge of molecular properties, as influenced by the polar solvent (water) and the nonpolar solvent ("wet" octanol), would be sufficient to calculate log P_{oct} with great confidence. Some workers think MO methods have already reached this level of performance. Using the 8000 measured values contained in Starlist as a test set (or a substantial randomized subset of it) could determine just how well alternative methods meet expectations. Most experts in the MO field acknowledge a certain weakness in calculations for phosphorous and sulfur.[19] There are 80 examples of solutes containing the P=O moiety and 485 with S=O in the 7500 solutes used in eq 17. Inability to handle such solutes would be a severe handicap in the eyes of most drug and pesticide designers.

Any method, CLOGP included, is bound to produce some calculations in which compensating errors produce a fortuitous result. In Figure 17 niddamycin is surely one of these. The sum of proximity effects in this case is 10.38 log units, which is the amount of hydrophilic character CLOGP sees as "lost" by placing 14 oxygens and one nitrogen in this particular close environment. One should not expect that other solutes needing this many proximity factors will be calculated with less than a 0.2 deviation. However, such "lucky accidents" will not carry much weight in a test set of more than 7000 solutes.

Many chemists are understandably skeptical of "fragmenting" an aromatic heterocycle and trying to make sense of the parts. Fujita and his coworkers have found it quite a daunting task to predict the π-values on diazines from those of the simple azines.[67] Taylor, for theoretical reasons, has expressed doubts that any 'bidirectional' method of treating electronic interaction between fragments can succeed in the long run.[68] Even if true, this may not bar its practical utility. But it is not surprising that one of the most difficult tasks in constructing CLOGP has been to take the isolated fragment values of —NH— from pyrrole, —N= from pyridine, —O— from furan, and —S— from thiophene and combine them with reasonable σ/ρ values for electronic interaction to come up with satisfactory estimates for the diazines, triazines, oxazoles, thiazoles, pyrimidines, and purines. In Figure 18 one can observe the results of the "constructionist" approach to the

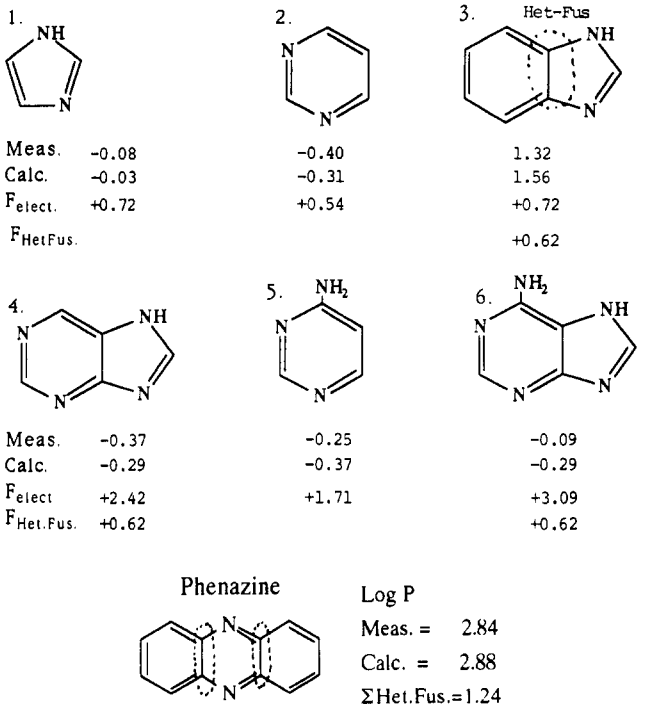


Figure 18. Stepwise "construction" of diazine and purine log *P*'s.

fragmental calculation of diazines and substituted pyrimidines and purines. The fragment values for —NH— and —N= (−0.68 and −1.14) are taken directly from pyrrole and pyridine, and the σ/ρ values are an average taken from the simplest (and most carefully measured) substituted analogs. They are

pyrrole —NH—	$\sigma = 0.00$	$\rho = 0.80$
pyridine —N=	$\sigma = 0.90$	$\rho = 0.30$

In purine (no. 4) +0.62 of the electronic effect is credited to fragment interaction *across* the ring fusion as opposed to +1.80 effect *within* each ring, and in adenine (no. 6) it is +0.66 vs +2.43. The worst anomaly seems to be benzimidazole where the "heterofusion" factor is applied twice (for +0.62) and seems to overcompensate. It may also be partially responsible for the overprediction in adenine. Yet the factor is proper for purine and seems proper when applied four times in phenazine ($F = +1.24$; deviation = −0.04). When these results are taken with thousands of calculations for other heterocycles (see Table I) it seems reasonable to put some faith in the separate values assigned to these polar interactions. There is just too much agreement to be the result of "lucky accident" as was the case with niddamycin.

Of course, much improvement still needs to be made, including, as was mentioned above, a revamping of the method to account for nonadditivity of electronic interactions *on* and *in* the rings ("ageing" or "fading"). Perhaps as important it is to take into account conjugation operating through vinyl, styryl, and quinone systems.

Some critics of the CLOGP program believe it is unnecessarily complex and ask why a simplified method cannot be offered which can be applied manually even to complex drug and pesticide structures. Rekker's latest "cookery book",[7] which uses the same basic fragmental approach as CLOGP, appears to offer these people hope that this is an achievable goal. But it

remains to be seen how well a novice can deal with choosing the correct multipier (k_n) for the "magic constant" for complex drugs. It appears that for ampicillin $k_n = 21$, which results in a correction factor of +4.60, while furosemide is assigned only one for +0.219.[7]

One of the objectives of this review is to clearly state the position that there is an innate complexity in the competing forces which determine the octanol/water partition coefficient. To best face up to this complexity one should work with the largest reliable database of measured values. Knowing from the start that log P of a solute could not be a simple sum of its parts, one can try to categorize the factors which depend on the arrangement of the parts. The labels that have been assigned to these factors are reasonable and informative, but no one can guarantee their accuracy. Matters have not changed a great deal since 1985 when Wolfenden stated:[69] "Because of remaining uncertainties about the structure of water, alone and in the neighborhood of solutes, it is easier to appreciate the importance of solvation effects than to be sure of their physical origins". This becomes even more relevant for estimating log P_{oct} when one considers the lack of knowledge about the solvation mechanisms in wet octanol. At any rate it is hoped that researchers in the field will agree that the current Starlist database of over 7800 measured log P_{oct} values provides the most demanding test of performance for any new or improved procedure for estimation of this valuable parameter.

IV. Acknowledgments

Some idea of the size of the data-gathering task represented by the 40 000 values in MASTERFILE can be gained from section I.A of this review. The lion's share of credit for this monumental "sleuthing" effort belongs to Prof. Corwin Hansch who followed up the early efforts of Myer and Overton (1899) and Collander (1950) and convinced the world of the importance of the hydrophobic parameter. A number of measurements of partition coefficients of crucial solute structures were made at the Pomona College laboratories by, among others, Mr. George Gould, Ms. Priscilla Jow, Dr. Gargi Debnath with occasional aid from visiting scientists, such as Prof. Carlo Silipo and Prof. Antonio Vittoria. Without the insight of Dr. David Weininger, who provided much of the computer program "tools", the CLOGP methodology reported in this review would never have advanced much beyond the "experimental" stage. For the analysis of the 8000 solutes in Starlist, much credit is due to Mr. David Hoekman. Having CLOGP in use in many laboratories around the globe has resulted in invaluable "feedback" and constructive comments when it failed to match a newly measured log P. Although these "advisors" are not mentioned by name, their help is gratefully acknowledged. Finally, the many measurements contributed by "private communication" attest to the cooperative nature of Medchem Project, and a great deal of credit for any valuable service it has provided must be widely shared.

V. References

(1) Leo, A.; Hansch, C.; Elkins, D. *Chem. Rev.* **1971**, *71*, 525.
(2) *Current Contents* **1983**, *23*, 14.
(3) Hansch, C.; Bjorkroth, J. P.; Leo, A. *J. Pharm. Sci.* **1987**, *76*, 663.
(4) Leo, A. *Environmental Health Chemistry*; McKinney, J. D., Ed.; Ann Arbor Science: Ann Arbor, MI, 1981; p 323.
(5) (a) Fujita, T.; Iwasa, J.; Hansch, C. *J. Am. Chem. Soc.* **1964**, *86*, 5175. (b) Fujita, T. *Prog. Phys. Org. Chem.* **1983**, *14*, 75.
(6) Rekker, R. *The Hydrophobic Fragmental Constant*; Elsevier Scientific Publishing: Amsterdam, 1977.
(7) Rekker, R.; Mannhold, R. *Calculation of Drug Lipophilicity*; VCH: Weinheim, 1992.
(8) Leo, A. *Methods in Enzymology*; Langone, J., Ed.; Academic Press: San Diego, CA, 1991; Vol. 202, p 544.
(9) Broto, P.; Moreau, G.; Vandycke, C. *Eur. J. Med. Chem.* **1984**, *19*, 71.
(10) Ghose, A.; Crippen, G. *J. Comput. Chem.* **1986**, *7*, 565.
(11) (a) Klopman, G.; Wang, S. *J. Comput. Chem.* **1991**, *12*, 1025. (b) Klopman, G.; Iroff, L. *J. Comput. Chem.* **1981**, *2*, 157.
(12) Iwase, K.; Komatsu, K.; Hirono, S.; Nakagawa, S.; Moriguchi, I. *Chem. Pharm. Bull.* **1985**, *33*, 2114.
(13) Moriguchi, I.; Hirono, S.; Liu, Q.; Nakagome, I.; Matsushita, Y. *Chem. Pharm. Bull.* **1992**, *40*, 127.
(14) Dunn, W., III; Grigoras, S.; Koehler, M. *J. Med. Chem.* **1987**, *30*, 1121.
(15) Koehler, M.; Grigoras, S.; Dunn, W., III. *Quant. Struct-Act. Relat.* **1988**, *7*, 150.
(16) Rogers, K.; Cammarata, A. *Biochim. Biophys. Acta* **1969**, *193*, 22.
(17) Hopfinger, A.; Battershell, R. *J. Med. Chem.* **1976**, *19*, 569.
(18) Bodor, N.; Gabanyi, Z.; Wong, C-K. *J. Am. Chem. Soc.* **1989**, *111*, 3783.
(19) (a) Bodor, N.; Huang, M.-J. *J. Pharm. Sci.* **1992**, *81*, 272. (b) Sasaki, Y.; Kubodera, H.; Matuszaki, T.; Umeyama, H. *J. Pharmacobio-Dyn.* **1991**, *14*, 207.
(20) Hansch, C.; Leo, A. *Substituent Constants for Correlation Analysis in Chemistry and Biology*; Wiley Interscience: New York, 1979; p 19.
(21) Weininger, D. *J. Chem. Inf. Comput. Sci.* **1988**, *28*, 31.
(22) Weininger, D.; Weininger, A.; Weininger, J. *J. Chem. Inf. Comput. Sci.* **1989**, *29*, 97.
(23) Weininger, D. *User's Manual V 3.54*; Pomona College Medchem Project: Claremont, CA, 1989.
(24) Purcell, W.; Bass, G.; Clayton, J. *Strategy of Drug Design*; Wiley-Interscience: New York, 1973; Appendix I.
(25) Mirrlees, M.; Moulton, S.; Murphy, C.; Taylor, P. *J. Med. Chem.* **1976**, *19*, 615.
(26) El Tayar, N.; Tsai, R.-S.; Carrupt, P.-A.; Testa, B. *J. Chem. Soc., Perkin Trans. 2* **1992**, *2*, 79.
(27) Thomas, B.; Compton, D.; Martin, B. *J. Pharmacol. Exp. Ther.* **1990**, *255*, 624.
(28) Leo, A. *Sci. Total Environ.* **1991**, *109/110*, 121.
(29) Nernst, W. *Z. Phys. Chem.* **1891**, *8*, 110.
(30) Leo, A. *Comprehensive Medicinal Chemistry*; Hansch, C., Ed., Pergamon Press: Oxford, 1990; Vol. 4, p 315.
(31) Akamatsu, M.; Yoshida, Y.; Nakamura, H.; Asao, M.; Iwamura, H.; Fujita, T. *Quant. Struct.-Act. Relat.* **1989**, *6*, 195.
(32) Leo, A. *QSAR: Rational Approaches to the Design of Bioactive Compounds*; Silipo, C., Vittoria, A., Eds.; Elsevier Science Pub.: Amsterdam, 1991; p 349.
(33) Hansch, C.; Leo, A. *Correlation Analysis in Chemistry and Biology*; Submitted to American Chemical Society Books.
(34) Debnath, A.; Debnath, G.; Shusterman, A.; Hansch, C. *Environ. Mol. Mutagen.* **1992**, *19*, 37.
(35) Debnath, A.; Compadre, R.; Debnath, G.; Shusterman, A.; Hansch, C. *J. Med. Chem.* **1991**, *34*, 786.
(36) van de Waterbeemd, H.; Testa, B. *Int. J. Pharm.* **1983**, *14*, 29.
(37) van de Waterbeemd, H. *Hydrophobicity of Organic Compounds: How to Calculate it by Personal Computers*; Compudrug International: Vienna, 1986.
(38) Wolfenden, R. *Science* **1983**, *222*, 1087.
(39) Kamlet, M.; Doherty, R.; Fiserova-Bergerova, V.; Carr, P.; Abraham, M.; Taft, R. *J. Pharm. Sci.* **1987**, *76*, 13.
(40) (a) Leahy, D. *J. Pharm. Sci.* **1986**, *75*, 629. (b) Viswanadhan, V.; Ghose, A.; Revankar, G.; Robins, R. *J. Chem. Inf. Comput. Sci.* **1989**, *29*, 163.
(41) (a) SAVOL and CONCORD programs written by R. Pearlman, Univ. of Texas, Austin; furnished courtesy of Tripos Assoc., St. Louis Mo. (b) De Bruijn, J.; Hermens, J. *Quant. Struct-Act. Relat.* **1990**, *9*, 11.
(42) Bodor, N. Private communication.
(43) Smith, R.; Hansch, C.; Ames, M. *J. Pharm. Sci.* **1975**, *64*, 599.
(44) (a) Camilleri, P.; Watts, S.; Boraston, J. *J. Chem. Soc. Perkin Trans. 2* **1988**, 1699. (b) Jorgensen, W.; Briggs, J.; Contreras, M. *J. Phys. Chem.* **1990**, *94*, 1683.
(45) Leo, A. *J. Chem. Soc., Perkin Trans. 2* **1983**, 825.
(46) CLOGP results are quoted in ref 18.
(47) (a) Kamlet, M.; Abboud, J.-L.; Abraham, M.; Taft, R. *J. Org. Chem.* **1983**, *48*, 2877. (b) Kamlet, M.; Abboud, J.-L.; Taft, R. *J. Am. Chem. Soc.* **1977**, *99*, 6027. (c) Kamlet, M.; Doherty, R.; Abraham, M.; Taft, R. *Quant. Struct.-Act. Relat.* **1988**, *7*, 71.
(48) Kamlet, M.; Doherty, R.; Abboud, J.-L.; Abraham, M.; Taft, R. *J. Pharm. Sci.* **1986**, *75*, 338.
(49) Taft, R.; Abraham, M. Private communication.

(50) Abraham, M. 1st CRDEC Meeting on Solute/Solvent Interactions, Aberdeen Proving Grounds: Maryland; May 1991.
(51) *Medchem Software Manual, V.3.54*; Daylight Chemical Information Systems, Inc., 1989.
(52) Abraham, M.; McGowan, J. *Chromatographia* **1987**, *23*, 243.
(53) Abraham, M.; Whiting, G.; Doherty, R.; Shuely, W. *J. Chromatog.* **1991**, *587*, 213; 229. And private communication.
(54) Laurence, C.; et al. *J. Chem. Soc. Perkin Trans. 2* **1988**, 1491.
(55) (a) El Tayar, N.; Testa, B.; Carrupt, P.-A. *J. Phys. Chem.* **1992**, *96*, 1455. (b) Karickhoff, S. 1st CREDEC Meeting on Solute/Solvent Interactions, Aberdeen Proving Grounds, Maryland; May 1991.
(56) (a) Hammett, L. P. *Physical Organic Chemistry*, 2nd ed.; McGraw-Hill: New York, 1970. (b) Taft, R., Jr. *J. Phys. Chem.* **1960**, *64*, 1805.
(57) Nakagawa, Y.; Izumi, K.; Oikawa, N.; Sotomatsu, T.; Shigemura, M.; Fujita, T. *Environ. Toxicol. Chem.* **1992**, *11*, 901.
(58) Dewar, M. J. S.; Zoebisch, E. G.; Healy, E. F.; Stewart, J. J. P. *J. Am. Chem. Soc.* **1985**, *107*, 3902.
(59) Calvino, R.; Gasco, A.; Leo, A. *J. Chem. Soc., Perkin Trans. 2* **1992**, 1643.
(60) van Duin, M.; Baas, J.; van de Graaf, B. *J. Org. Chem.* **1986**, *51*, 1298.
(61) Recanitini, M. Private communication.
(62) Sparatore, F.; Rotonda, M.; Caliendo, G.; Novellino, E.; Silipo, C.;Vittoria, A. *Il Farmaco, Ed. Sci.* **1988**, *43*, 29; 141.
(63) Bellasio, E.; Campi, A.; DiMola, N.; Baldoli, E. *J. Med. Chem.* **1984**, *27*, 1077.
(64) Bastiansen, O.; Fernholt, L.; Hedberg, K.; Seip, R. *J. Am. Chem. Soc.* **1985**, *107*, 7836.
(65) Allinger, N. *J. Am. Chem. Soc.* **1977**, *99*, 8127.
(66) Hansch, C.; Leo, A.; Taft, R. *Chem. Rev.* **1991**, *91*, 165.
(67) Yamagami, C.; Takao, N.; Fujita, T. *Quant. Struct.-Act. Relat.* **1990**, *9*, 313.
(68) Taylor, P. *Comprehensive Medicinal Chemistry*; Hansch, C., et al., Eds.; Pergamon Press: Oxford, 1990; Vol. 4, Chapter 18.7.
(69) Wolfenden, R. *Proc. Indian Acad. Sci.* **1985**, *94*, 121.