CLOGP - **C**alculate **L**ogP
CLOGP3 User Guide - (last modified 1989)

Abstract

The CLOGP3 User Guide will help the user under stand why the CLOGP3 program calculates log P(ow) the way it does. Although the procedure cannot be derived from first principles, we have tried to make the rules consistent with solvation theory, if for no other reason than they are more easily remembered. The method simply adds together values for structural parts of a solute molecule and correction factors dependent upon the particular way the parts are put together.

The CLOGP3 EXAMPLES section contains example CLOGP3 calculations for a variety of chemicals and is designed as a companion to the CLOGP3 User Guide. An asterisk (*) appears in the CLOGP3 User Guide when one or more examples are provided in the EXAMPLES section to illustrate aspects of the CLOGP3 computation. Examples also demonstrate DEPICT (chemical depiction) and STARLIST (Log P(ow) verification set access). An explanation of example format is included.

The ALGORITHM MANAGER is the 'heart' of the CLOGP calculating program, because it allows feedback from new measurements to fine-tune and to extend the program's capabilities. Even the most complex of new fragments can be entered in a few minutes and are then immediately retrievable by all or just a selected number of the system's users. The numerical values of both fragments and correction factors can be quickly changed as new data dictates.

14.1 Introduction

14.1.1. Measurement and Past Use of Partition Coefficients

The partition coefficient is the equilibrium concentration of solute in a non-polar solvent divided by the concentration of the same species in a polar solvent. In this and most other applications, the polar solvent is water. The logarithm of the partition coefficient, log P, has been successfully used as a hydrophobic parameter in 'extrathermo-dynamic' Hammett methodology. 1-octanol has much to recommend it as the choice for the non-polar phase(1) and log P(ow) has been used successfully in Quantitative Structure Activity Relationships (QSAR) in the following special fields: drug and pesticide design(2,3); pharmacokinetics(4); anaesthesiology(5); environmental transport and soil binding(6,7); toxicology(8); bioaccumulation(9); protein folding(10); enzyme binding(11,12); enzymic reactions in non-aqueous solvents(13); and host-guest complexation(14a,b).

In principle, the measurement of the equilibrium concentrations of solute in the octanol and water phases, after shaking in a separatory funnel, is very simple, and since good measured values are always to be preferred over calculated ones, it would seem that there should be little need for a procedure to calculate them. As it turns out, reliable shake-flask measurements are time-consuming and often difficult to make. The criteria for high reliability are: measurements over a 10-fold concentration range (with upper concentration no more than 75% of solubility or CMC, or no more than 5% of the aqueous phase, whichever is lower) and standard deviation of 0.03 or less in log terms. This often requires working at sub-micromolar concentrations, and so, with either UV spectrophotometry or gas chromatography, it means that the standard curves must be established with utmost care. Radiotracer methods seem well-suited for analyses at these low concentrations,

but impurities as well as adsorption at phase boundaries (including container walls) can introduce significant errors.

HPLC procedures provide a way around this bottleneck(15,16) and can save time if there is a limited variety of structural types, and the log P values fall in the range of 0.5 to 4.0. Most HPLC procedures which are used to develop log P(ow) values do not use octanol and thus have to be referred to that system by standard curves which can be different depending on whether the solutes do or do not contain certain basic fragments, such as a pyridine nitrogen. If the solutes do not absorb well in the UV, difficulties in detecting the elution time eliminates any advantage HPLC may have over the shake-flask method.

Procedures which employ filter probes(17) or solubility columns(18) speed up partition coefficient measurements by eliminating centrifugation as the means of phase separation. However, each has its own set of disadvantages and limit its acceptance as a method for establishing the standard values for a calculation procedure.

More efficient methods of measurement of octanol/water partition coefficients are certain to be developed in the future, but no conceivable 'breakthrough' is likely to eliminate the need for log P calculation. To put the problem in proper perspective, one need only imagine some dedicated synthetic chemist making all possible tri-substituted benzoic acids with the methods commonly available today. When finished, there would be five million analogs for which partition coefficients could be determined. And of course only by calculation is one going to have an estimate of hydrophobicity before synthesis.

The Pomona MedChem Project saw these and other arguments as reason enough to develop a method to calculate log P(ow) from structure by an additive-constitutive procedure. As it turns out, the 'constitutive' portion of the procedure was, by the very nature of the two competing solvation equilibria, very complex, and the manual method required considerable effort before it could be applied with confidence. It is the aim of the second-generation program, CLOGP3, to take most of the routine calculation burden from the user but still encourage him to study the interplay of hydrophobic and polar solvation forces which can be so crucial to the design of bioactive chemicals.

14.1.2 How to Understand CLOGP3 Calculations

The first published method for calculating log P(ow) from structure(19) was based on a 'substitution' procedure and was developed with substituent pi constants for aromatic rings in mind. Of course this method was limited to deriving a new log P from a 'parent' structure whose log P was already known. Rekker(20) was the first to publish a procedure which was more general in that it assigned 'fragmental constants' to a variety of structural pieces, and the calculated log P was the sum of the values appropriate for the molecule in question. The original pi system can be expressed as:

while the expression for Rekker's fragment system is:

The method developed by Pomona MedChem(21) follows Rekker's general formulation, but there are some important differences in the approach used to derive the actual working constants. Rekker used a 'reductionist' approach--deriving the constants for carbon and hydrogen as well as those for polar fragments from a statistical treatment of a large body of log P data which contained numerous interaction factors. Both the fragment values (f) and interaction factors (F) had to be identified and evaluated concurrently. Also, Rekker neglected to clearly define just what constitutes a fragment. Instead he provides a table in which the known constants can be found (see footnote). Rekker also treats all correction factors as some multiple of a 'Magic Number' (+0.28), but the selection of multiples was not made clear in his published work. Although his method gained some acceptance

for manual calculation, we considered it too seriously flawed to serve as the basis for a computer method.

In order to construct a dependable, verifiable algorithm suitable for log P calculations used in developing QSARs at Pomona College, we first elected to clearly define what constitutes a fragment. Next we chose a 'constructionist' approach to evaluate them; that is, we accepted as axiomatic that the hydrophobic portions of solutes were those most 'hydrocarbon-like', and defined these carbons and hydrogen fragment values as being truly constant. We gave very heavy emphasis to the carefully-measured values for three solutes: molecular hydrogen, methane and ethane, because from these we could derive fragment constants for carbon and hydrogen which would be free of obscuring interactions. For all hydrocarbon structures more complex than these, whose measured values were NOT the sum of fragment values, we attempted to define the difference in terms of universally-applicable correction factors. It appears that this approach has led not only to a workable algorithm, but has highlighted the importance of certain types of polar solvation forces which have received insufficient attention in the past.

The first attempt to reduce the 'Pomona Method' of log P calculation to computer algorithm was made in collaboration with Dr. Jack Chou and Dr. Peter Jurs of Pennsylvania State University.(22) It was called CLOGP. A great deal was learned in the process of developing this first version, and it certainly established the real need for a 'stand-alone' program to make these calculations. Nevertheless, CLOGP was difficult to install and to modify, and many well-known correction factors could not be implemented due to programming difficulties. In light of this experience, we deemed it essential to incorporate, in the second generation program, design features which would encourage its continuing evolution. To achieve that objective, the program had to be conceived as a 'modeling system' which could operate from one or more easily-revised 'value files' (see Algorithm Manager Section for details).

_____

Footnote: Not all fragments in Rekker's table are reduced to a uniform level of structural simplicity; e.g., -O-, -CH2-, and -COOH are listed as separate fragments in an aliphatic context, but -OCH2COOH is entered as an aromatic fragment because his treatment of proximity effects cannot accommodate the three fragments separately.

As an example of the ease of updating, the largest fragment encountered to date is:

It took less than two minutes to enter it into the database and begin to use it in calculations.

As CLOGP3 is being tested more thoroughly, a steady pattern of deviant structures is emerging. Some of these are recognized by the program, and the calculation is passed on with a warning message. Since many of these have been characterized qualitatively, we are considering accompanying the message with our best guess until such time as we can quantify it. (See EXAMPLES, 6.3.)

It is apparent from the structures which now give anomalous calculations that the next large improvement in CLOGP3 must await a method to interface it with a dependable molecular mechanics program. It needs to provide a measure of the distance between polar fragments through space; that is, molecular conformations in dilute solution.

14.2 Fundamental Fragments

In view of the decision to make alkane carbons and hydrogens the most fundamental fragments in the system, it is necessary to define these very carefully before defining the polar, more hydrophilic fragments.

14.2.1 Isolating Carbons

An 'Isolating Carbon' (I.C.) atom is a carbon which is NOT doubly- or triply- bonded to a hetero atom. An I.C. may be bonded to a hetero atom by a single or an aromatic bond. This definition can be made clearer from the following two examples:

1. pyrimidine 2. coumarin

In an earlier version of the manual calculation procedure (Ref. 21, p. 34) the Kekule structures for pyrimidine was considered; the earlier rule is now superceded. In coumarin, both rings are designated as aromatic, and the only carbon which is not isolating is the one in the carbonyl group, because it is doubly bonded to a hetero outside the ring.

Although the hydrophobic value of an I.C. is constant, several types must still be identified; the degree to which they delocalize electrons in any polar fragments attached to them has a great influence on overall log P. The types of I.C.s presently identified are listed below with appropriate symbols:

A ... Aliphatic
Z ... Benzyl
V .... Vinyl
Y .... Styryl
a ..... Aromatic

To be completely characterized, a polar fragment must have each of its 'valence bonds' designated with one of the above symbols (see section "Fragment Valence Types"). The numerical value of the fragment will increase roughly in the order 'A' to 'a', but must be experimentally determined for high reliability.

All hydrogens bonded to I.C.s are fragments. These two kinds of fundamental fragments are the most important members of the non-polar class. A comparison of their relative values (C = 0.20; H = 0.225) is a reminder that the measure of effective cavity size may not be as simple as using van der Waals radii or CPK models.

14.2.2 Polar Fragments

A fragment is any atom or group of atoms bounded by Isolat ing Carbon atoms, and all except hydrogen are considered polar. A fragment may have many internal bonds but those connecting it to I.C.s are called 'valence bonds'. Valence bonds are most often single, but can be aromatic, as in the case of the N fragments in pyrimidine shown above. Each hydrogen in methane is a fragment, but the hydrogens in formaldehyde are not because the carbon to which they are bonded is not isolating. This is very important to remember, for one frequently sees published calculations in which one fragment value is obtained from another by the replacement of a fragment hydrogen with another fragment of known value (see footnote). At the present time a good rule to follow is: "Never break up a Fragment; estimations can be made from values measured for different bond environments (see below), but a Fragment cannot be constructed from parts." Examples of fragments which cannot be 'broken down' further are:

_____

Footnote: An 'a priori' method of Fragment calculation is under development, but further work is needed before it can be incorporated in a computer algorithm.

Monovalent: -Cl; -CN
Divalent: -OC(=O)NH-
Trivalent: -OC(=O)N<
Tetravalent: >NC(=O)N<

As will become evident in the following sections, polar fragments can interact in various ways. To quantitate this interaction it is necessary to define several types of polar fragments:

(A) X = any halogen, but for one type of interaction fluorine must be assigned to a special subclass, 'F'.

(B) Y = all non-X fragments; these are further subdivided according to:

(1) sensitivity to halogen interaction as 'Y-1', 'Y-2', and 'Y-3'

(2) containing '-OH' or not.

14.2.3 Intrinsic Values

Here the term, 'intrinsic' fragment values, means those which would, if summed, yield the correct log P without any correction factors. It is worthwhile to examine some of the accepted hypotheses as to what solvation forces or other phenomena determine these intrinsic values.

It takes more energy to form a cavity in water than in octanol. One would predict, therefore, that increasing the size of a solute would increase its log P. Other factors being equal, this appears to be the case. However, other features of the solute can partly or completely override the effect of it size.

Water is much more capable than is octanol of accommodating localized dipoles, and it contains, on a molar basis, more hydrogen bond accepting and donating groups. So it is these three factors - size, localized dipole strength, and H-bonding ability - which largely determine the sign and magnitude of any fragment value.

14.2.3.1 Halogens

Halogens form an intense localized dipole when bonded to an aliphatic carbon atom.* This intensity is somewhat lessened if the I.C. is benzyl and greatly lessened if it is vinyl, styryl or aromatic.* Fluorine has a negative fragment value when attached to an aliphatic carbon, because the dipole effect outweighs the effect of size. Size is of greater importance with chlorine and bromine, but even bromine is less hydrophobic than a hydrogen in an aliphatic setting. As will become evident in the following section, much of this hydrophilic polar effect can be lost through 'shielding' by other halogens, or by electronic interaction with 'Y' type polar groups.

14.2.3.2 H-Polar Fragments

H-Polar Fragments ('Y') almost universally form some sort of hydrogen bonds with the donor (H) or acceptor (O) of the aqueous phase. This is thought to interrupt the peculiar 'ice-like' water shell which forms around the non-polar, hydrocarbon-like portions of each solute molecule, and thereby effectively reduces cavity size. As noted above,

this should reduce log P.

14.2.3.3 Ions

Octanol can accept some larger solutes containing a full formal charge in sufficient concentration for measurement. However, one must be careful that the species measured is the same, because water easily supports complete ionization while ion-pairing is the usual condition in octanol except at the very lowest concentrations. Consistent values can be obtained if 'standard conditions' are

adhered to: 0.1 M small counter-ion (Na+ or Cl-) and extrapolation to infinite dilution. Measured in this way a carboxylate ion is about 4.1 log units lower than the undissociated acid. No single value can be given to the positive charge on a protonatedamine or quaternary ammonium, because the charge is delocalized along the hydrocarbon chain and thus the effect is dependent on chain-length. It should be emphasized at this point that, except for zwitterions, CLOGP3 calculates values for the neutral solute only.*

14.2.3.4 Unsaturations

Double bonds in isolation have a slightly negative effect on log P.* This effect may arise from the polarity of the pi electrons or else it may be due to the shorter bond length

_____

* Each asterisk in this section indicates that an example calculation can be found in the CLOGP3 EXAMPLES section.

reducing cavity size. At any rate, it disappears if the double bonds are conjugated.* Triple bonds are decidedly hydrophilic and require a large negative correction factor.

14.3 Correction Factors

14.3.1 Structural Factors

14.3.1.1 Bonds

To properly perform its calculations, CLOGP3 needs to know the number and types of certain bonds in the solute structure. There is some reason to believe that, for the bonds in question, factors other than bond length affect the size of the solvent cavity needed to contain the hydrophobic portions of the solute molecule.

The effect of all bonds within any fragment is taken care of by the fragment value, and so it is NOT necessary to keep track of them, NOR of any bonds to hydrogen. And, as explained below, it is convenient to allow for the bond effect in aromatic rings by including it in a special aromatic I.C. type, 'aromatic carbon'; therefore, aromatic bonds also are NOT given special attention.

Bonds which DO need to be identified are the following:

14.3.1.1.1 Chain Bonds

Chain bonds are non-ring bonds between I.C.s plus any valence bonds to fragments.*

14.3.1.1.2 Ring Bonds

Ring bonds are non-aromatic ring bonds between I.C.s plus any valence bonds to fragments.*

14.3.1.1.3 Branch Bonds

Branch bonds are chain bonds emanating from 'Branched Fragments' (a fragment type presently limited to tertiary amines and phosphate esters) and counted to the last I.C. preceding any polar fragment.*

A separate count of each of these three bonds types must be made, and a negative correction applied. For chain bonds only, this correction applies to bonds AFTER the first in each chain. For example, there is no net bond correction for ethane but there is one for propane. This suggests that the correction accounts for flexing of the chain which is not possible in methane or ethane. 1,2-diethylbenzene gets only a net of two bond corrections, because each chain is counted separately. Also compatible with a 'flexing' hypothesis is the fact that the correction is greater for chains than for aliphatic rings.

6

As noted above, an isolated double bond is assigned a negative correction factor (-0.09). This factor actually becomes slightly positive if the double bonds are conjugated in a ring such as benzene. Since it is much more convenient to assign all bonds in large fused aromatic ring systems as aromatic type, rather than using the Kekule system of alternating doubles and singles, it is worthwhile to assign a special fragment value to an aromatic carbon and include all the necessary bonding effects therein. The value for aliphatic carbon is +0.20; the value for aromatic carbon which includes all bond effects associated with the aromatic ring system is +0.13.*

14.3.1.2 Branching at Isolating Carbons

14.3.1.2.1 Chain Branch

It is well-known that iso-alkanes are more water-soluble than their n-isomers.* This branching evidently does not produce a corresponding solubility increase in the octanol phase in the partitioning process, because the correction required in CLOGP3 is negative in sign.

In CLOGP3, the concept of branching was expanded, and now replaces the earlier use of the 'ring' cluster' correction(21). Fusion carbons in non-aromatic rings are considered as branched and given the same correction factor as chains; i.e., -0.13.* They are designated cluster branches.

14.3.1.2.2 Group Branch

If an H-Polar group branches from an I.C. the increase in water solubility, compared to the n-isomer, is even greater than with chain branching. Again this carries over to partitioning equilibrium; H-Polar group branching requires a larger correction than does chain branching (see footnote). For this reason isopropyl alcohol is given one group branch correction and no chain branch is considered.* Tertiary butyl alcohol gets one of each type.*

If a fragment has more than two external (valence) bonds, it could be considered a branching point. However, in all cases except the 'Branched Fragments' noted above (t-amines and phosphate esters), the entire negative branching effect is included in the fragment value itself. Only in the case of the 'Branched Fragments' is the effect chain-length dependent.

_____

Footnote: The current algorithm allows for NO branching of either kind for halogens (X) on I.C.s. This may seem in conflict with the log P for 2-chloropropane where a chain branch correction would help the calculation. However, halogenation is very often multiple and would be very sensitive to small differences in the branching correction.* The current algorithm works quite well by including the halogen branching factor in the X-C-X and X-C-C-X corrections.*

14.3.2 Interaction Factors

14.3.2.1 Aliphatic Proximity (Measured Topologically)

14.3.2.1.1 Halogen vs. Halogen (X vs. X)

The positive correction to log P for this interaction is thought to result from dipole shielding and is limited to halogens on the same (geminal) or adjacent (vicinal) I.C.s. The geminal interaction is designated 'X-C-X', and the corrections can be thought to arise as follows: adding a second halogen to an I.C. which already has one creates the first X-C-X pair, and the correction required is +0.60.* Adding the third halogen to the same I.C. creates two more such pairings, each of which requires a correction of +0.5.* If the fourth halogen is added, the dipole is almost completely shielded, and the three additional pairings require corrections of +0.40 each.* For carbon tetrachloride the total geminal halogen correction would be:

For the vicinal halogen correction, X-C-C-X, the bond between carbons must not be double.* The correction is evaluated by subtracting one from the number of halogens meeting the structural requirement and multiplying by the factor 0.28. (Again Rekker's Magic Constant pops up!)

14.3.2.1.2 H-Polar vs. H-Polar (Y vs. Y)

As noted in the section on 'Intrinsic Values', the negative sign on 'Y' fragments is thought to result from their 'structure-breaking' (and thus cavity-reducing) ability in the water phase. 'Y' fragments appear to eliminate the cavity requirement for two or more I.C.s to which they are attached. Obviously if two 'Y' fragments are located on the same or adjacent I.C.s some of this cavity reduction is going to be counted twice. Thus a positive correction factor is called for when the topological separation is less than three I.C.s. The CLOGP3 algorithm is an improvement over the original Rekker procedure(20) in that, in place of the same correction for every Y-C-Y or Y-C-C-Y, it makes the correction proportional to how much hydrophilic character (negative fragment value) there is lost. This proportionality appears to apply even if one of the fragments is charged and has a highly negative value, but as previously noted, the CLOGP3 algorithm currently does not treat ions.

If one of the Y fragments in a Y-C-Y interaction contains an -OH moiety (e.g. -NHOH, -COOH, or -OH itself), a greater proportion of the hydrophilic character of the pair is lost. The coefficient by which the fragment sum is multiplied increases from 0.32 to 0.42.*

If both the 'Y' fragments and the carbons of Y-C-C-Y are in a ring, the hydrophilicity loss is not as great as if they are all in a chain (coefficient 0.26 vs. 0.20).* If one 'Y' is a substituent on the ring while the other is in the ring, the correction coefficients are averaged (0.23).* If one of the carbons has two 'Y' fragments, the geminal correction is applied first; then, for the (Y-C-C<Y'Y") correction, both pairings are calculated and averaged.* If both I.C.s have geminal 'Y' fragments, then the vicinal correction is not applicable.* (See penicillin in EXAMPLES.)

14.3.2.1.3 Halogen vs. H-Polar (X vs. Y)

The interaction being considered at this point is limited to that which takes place across single bonds. It is, therefore, probably due to an inductive or field effect. (The electronic interaction between fragments on or in aromatic rings is discussed in the following section.) In evaluating the X-C-Y correction factor, all halogens can be treated alike. However, there are at least three levels of sensitivity shown by 'Y' type fragments. In CLOGP3 the most sensitive class, 'Y-3', is restricted to the structural type: -SO2-R.* 'Y-2' consists of the types: -CONH-R, -O-R, -S- R,and -NH-R; and 'Y-1' of all other, H-polar fragments.* The correction for the first alpha-halogen (i.e. X-C-Y) is the same for all three Y-types (+0.9). For 'Y-3' fragments, this correction factor is doubled when there are two alpha halogens (X{2}-C-Y3) and tripled when there are three (X{3}-C-Y3). For 'Y-2' fragments, the second and third alpha- halogens need much less correction, and for 'Y-1', virtually none. In the case of multiple halogenation, the X-C-X and X-C-Y corrections are additive.

The CLOGP3 algorithm makes no separation of 'Y' types to make the X-C-C-Y correction, but needs to distinguish fluorine from the other 'X' halogens.*

14.3.2.2 Electronic (through Pi-bonds)

14.3.2.2.1 Fragment Valence Type

As previously noted, all fragments (X or Y type) are assigned the most negative values when bonded to aliphatic I.C.s (designated as 'A'). This can be considered as the 'base' or 'intrinsic' level. If the fragment value when attached to a benzyl I.C. (Z) has not been measured, it can be estimated as the base level plus 0.2.* If the value when attached to a vinyl I.C. (V) has not been measured it can be estimated as the average of the base and aromatic-bonded (a) values.* Likewise the value for the styryl-attached fragment can be estimated as two-thirds the way from the base to the aromatic

value.* CLOGP3 will only make these estimations when measured values have not been entered in the database.

14.3.2.2.2 Extension of Aromaticity

The extension of the aromatic ring system through fusion (as in naphthalene or direct substitution (as in biphenyl) appears to increase log P, especially if a heteroaromatic atom is next to the juncture.* If the ring-joining carbons are attached only to other aromatic carbons, the electron delocalization is minimal and so is the correction: +0.10 for each I.C. If the I.C.s are also attached to a polar (fused-in) fragment, such as in quinoline or 2-phenylpyrimidine, the correction is greater, +0.31.* (See appendix.)

14.3.2.2.3 Sigma/Rho Fragment Interaction

When two or more X and/or Y type fragments are attached to an aromatic ring system, the correction factors can be calculated by a method very similar to that used by Hammett(23) to calculate the electronic effects in other equilibria, such as acid ionization.* This requires the assignment of a measure of electronic 'strength' (sigma) and 'susceptibility' (rho). In dealing with electronic effects on partitioning equilibria, a few fragments appear to act 'bidirectionally' and require both sigma and rho values, although they cannot, of course, act upon themselves. Most of the details of sigma and rho assignment to X and Y type fragments can be found elsewhere(24), but it should be pointed out that the latest version of the program (CLOGP3) follows a newer procedure for 'fused-in' fragments. Fragments fused in aromatic rings (e.g. -N= or >C=O) may also be assigned both rho and sigma constants and treated together with 'on-ring' fragments instead of requiring a separate treatment.*

Fragments on different rings in an aromatic ring system interact with one another but the effect is attenuated.* If the two rings in the system are fused, as in 5-acetyl-1- naphthylamine, the 'intrinsic' effect is only half the sigma rho product just as it is in a biphenyl system such as in 4-(m-chlorophenyl)aniline.

One frequently encounters aromatic ring systems containing several fragments with rho and sigma values assigned, and the potential correction from all cross products of sigma/rho could be very large. Since these multiple effects are NOT additive, some scaling down procedure was indicated. The one chosen for CLOGP3 takes the following steps:

a) The full potential sigma/rho product for each possible interaction is calculated and placed in descending value order, AFTER considering if the fragment pair are on the same or separate rings.

b) Except for the sigma for a pyridine type nitrogen, each use of sigma or rho causes it to 'age'. The first interaction at the top of the list is entered at full potential because the current age of its sigma and rho components is 'zero' for each. Each use reduces the effective sigma or rho value to 1/2 its previous value, and so if each were at 'age 1', the increment to the correction would only be 1/4 as much as a 'fresh' interaction. The mechanics of this computation are best understood by looking at the detailed output of some complex structures. Two such examples are given in the appendix.*

14.3.2.3 Special Ortho

As noted in the previous section, aromatic substituent (fragment) pairs, if they have sigma and rho values assigned to them, are given the same correction factor regardless of their relative position on the ring. It is important to keep in mind that if the fragment pair are on adjacent positions (i.e. ortho), an additional correction may be required.

14.3.2.3.1 Crowding

'Crowding' of certain fragment types can effectively lower their aromatic-attached values. This is most apparent in the case of fragments attached to the aromatic ring through a hetero atom which possesses an electron pair, such as -NHCOCH3.* A reasonable explanation of this observation is that the lone pair can no longer remain in the plane of the ring, making the fragment attachment resemble aliphatic (A) rather than true aromatic (a). The magnitude of the correction appears to depend on both steric and electronic (field) effects(25).

If this explanation is valid, one would expect the correction to vary continuously up to a maximum characteristic of each fragment type. It was surprising, therefore, to find that the rather large data set used in the original evaluation of the 'negative ortho' effect(24) seemed to fit multiples of Rekker's Magic Constant(20). This is handled in CLOGP3 by assigning integers to a matrix which has generalized fragment types for coordinates.

More recent data provides many examples which do not support this 'quantized' correction. Nevertheless, it is being retained for the present because of simplicity and because its maximum error is only 0.14.

14.3.2.3.2 Intra-Molecular Hydrogen Bonding

Hydrogen bonding is known to occur intramolecularly between two ortho substituents if one is a donor and the other an acceptor. A classical example of such an H-bond is that in o-nitrophenol. As might be expected, an intramolecular H-bond reduces water's ability to accommodate that solute, and the log P of o-nitrophenol is over two log units higher than the m- and p-isomers in the heptane and carbon tetrachloride solvent systems. One must always keep in mind, however, that the octanol phase possesses both H-donor and H-acceptor capability, not only because it is an alcohol, but because of the 2M water present at saturation. In actuality, the presence of the intramolecular H-bond in o-nitrophenol penalizes solvation in octanol slightly more than it does solvation in water, and its log P is 0.09 log units lower than the m- and p-isomers.

In terms of intramolecular H-bonding between aromatic ortho substituents, the octanol/water system appears to be sensitive to a very restricted class. The only clear-cut cases seem to result from a carbonyl group directly attached to the ring acting as acceptor, and a directly-attached -OH or -NH- acting as donor.* In all of the cases observed so far, the correction is very close to +0.63, and is stored in the same matrix used for the 'negative ortho' corrections. Thus the 'crowding' and H-bonding ortho effects never are applied simultaneously, but a sigma/rho correction can be added to either.

14.4 Summary

The fragment method of calculating log P(ow) has been proved valuable in many fields, including drug design and hazard assessment. However, manual calculations require a great deal of instructions and become very lengthy for complex structures and thus are error-prone. The computer program, CLOGP3, enables the method to be applied by non-experts and includes an estimate of error, which is not possible with a manual method. Considerable effort will be made to improve and extend the program in the future. Regular users can avail themselves of an annual update which will bring them current with all newly measured fragment values and improved correction factors. Versions with the Unified Driver compare the calculation from structure with a measured value from log P(ow) for neutral solutes. Starlist is also included in the annual updating service. We plan to make available in the near future a searching program, GENIE, which will extend the search of Starlist to close analogs.

The current literature contains many examples of QSAR accompanied by calculations of hydrophobicity which have not been made according to a consistent application of the rules they purport to follow. This has caused some confusion and cast doubt upon the entire approach. Perhaps, if the use of CLOGP3 becomes more widespread, published calculations will become more comparable, especially if reference is made to the program version.

Any prediction of the future is risky, but, judging from the recent past, we can expect an increasing demand for log P(ow) values. It is inconceivable that CLOGP3 will be perfected to such an extent that it supplants partition coefficient measurement. The two methods should remain as they are now: mutually complimentary.

With the STARLIST module in CLOGP3, the user will be able to check the calculated value against an acceptable measured value if the solute structure entered is one of over 4,000 contained in that special file which is limited to non- tautomeric structures measured at a pH where the neutral form predominates.

14.5 Bibliography

1. Smith, R., Hansch, C. Ames, M., J. Pharm. Sci., 64, 599 (1975).
2. Martin, Y., Quantitative Drug Design, Medicinal Research Series No.8, Marcel Dekker, New York, 1978.
3. Magee, P., Chemtech, 11, 378 (1981).
4. Smyth, R., Pfeffer,M., Van Harken, D., Cohen, A. and Hottendorf, G., Antimicrob. Agents Chemother., 19,1004, (1981).
5. Koblin, D., Eger II, E., Johnson, B., Collins, P., Terrell, R., and Spears, L., Anesth. Analg., 60, 464 (1981).
6. Brown,D. and Flagg,E., J. Environ.Qual., 10, 382 (1981).
7. Ellinghausen, H., Guth, J. and Eser, H., Ecotox. Environ. Safety, 4, 26 (1980.
8. Levitan, H., Proc. Nat. Acad. Sci. USA, 74, 2914 (1977).
9. Neeley, W., Branson, D. and Blau, G., Environ. Sci. Technol., 8, 1113 (1974).
10. Tanford, C., The Hydrophobic Effect, Wiley, New York, 1980, 2nd Ed., Ch.13.
11. Coats, E., Genther, G., Dietrich, S., Guo, Z., and Hansch, C., J. Med. Chem., 24, 1422 (1981).
12. Smith, R., Hansch, C. and Langridge, R., Arch. Giochem. Biophys., 215, 319 (1982).
13. Martinek, K. and Semenov, A., J. Appl. Biochem., 3, 93 (1981).
14a. Newcomb, M., Moore, S. and Cram, D., J. Am. Chem. Soc.,99, 6405 (1977).
14b. Hansch, C., J. Org. Chem., 43, 4889 (1978).
15. Unger, S. and Feuerman, T., J. Chromatog., 176, 426 (1979).
16. Mirrlees, R. and Taylor, P., J. Med. Chem., 19, 615 (1976).
17. Tomlinson, E., Proceedings of A.Ph.A., Symposium on Partition Coefficients, San Diego, CA, 1982.
18. Tewari,Y., Miller, M., Wasik, S. and Martire, D., J. Chem. Eng. Data, 27, 451 (1982).
19. Fujita, T., Iwasa, J. and Hansch, C.,J. Am. Chem. Soc., 86, 5157 (1964).
20. Rekker, R., The Hydrophobic Fragmental Constant, Elsevier, Amsterdam (1977).
21. Hansch, C. and Leo, A., Substituent Constants for Correlation Analysis in Chemistry and Biology, Wiley Interscience New York (1979).
22. Chou, J. and Jurs, P.,J. Chem. Inf. Comput. Sci.,19, 172 (1979).
23. Hammett, L., Physical Organic Chemistry: Reaction Rates, Equilibria and Mechanism, McGraw-Hill, New York, 1970, 2nd Ed.
24. Leo, A.,J. Chem. Soc., Perkin Trans. II, 825 (1983)
25. Ogino, A., Matsumura, S. and Fujita, T., J. Med. Chem., 23, 437 (1980).
26. Leo, A. and Hansch, C., J. Org. Chem, 36, 1539 (1971).

14.6. Example CLOGP3 Calculations

The calculations shown in this section illustrate how CLOGP3 treats the basic fragment types and correction factors discussed in the reference section. No attempt was made to illustrate every fragment in the database, of course. The output is not exactly as seen with UDRIVE but combines the DEPICT output from the primary UDRIVE panel with the detailed calculations and the structure maps from the C(LOGP3) panel. Additional output was generated to illustrate the unusually complex sigma/rho electronic correction factor for atrazine and adenine.

The examples are arranged according to the section number of the CLOGP3 User Guide text which deals with the main feature illustrated. Since some structures illustrate more than one feature, some of these "secondary features" are also indicated by section number in the output.

14.7. Interpretation of Output

14.7.1. Maps

The first line in the "Map Box" gives the SMILES as entered by the user. Fragment ordering, therefore, is NOT unique, and so to completely understand the tabular results, one should become familiar with the map. The second line in the map indicates Isolating Carbon type under their respective SMILES notations. The third line numbers the polar fragments (i.e.,all those NOT I.C. or hydrogen) in the order entered. The fourth line shows the locations of hydrogen atoms, including those contained in fragments. The last lines indicate the location of atoms in rings. Even in simple structures these maps can be of help in interpreting the calculations. For example, in the four calculations in section 2.3.1., the value of the -Br fragment varies from 0.200 to 1.090. The reason is apparent from the maps which show the variation in type of I.C. attachment.

14.7.2. Picture

Each example is accompanied by a picture of the chemical structure as generated by the DEPICT algorithm. DEPICT indicates aromaticity by drawing circles inside aromatic rings and suppressing all aromatic carbons symbols. More information about how aromaticity is defined and how the picture is generated may be found in the SMILES and DEPICT sections of the MedChem Software manual.

CLOGP3 Examples

14.7.3. Tabular Results

Most of the nomenclature in the "calculation details" is understandable after reading the main body of the CLOGP3 User Guide, but a few terms could use further explanation:

1) Under Class one may find "SCREEN". This is further documented under Description section of the CLOGP3 User Guide. For instance, it may note "possibly anomalous steroid".

2) Under Type one may find the ring number for an Ortho interaction, for example, because there may be more than one ring on which such a correction might potentially be applicable.

3) Under Description one finds (ZW-) or (ZW+) after the name of each fragment which can participate in zwitterion formation. However, only when strong enough pairs are present is the correction actually entered (near the bottom of the table.) A sulfonic acid is strong enough even if both it and the amine are aromatic; for a carboxylic acid, both it and the amine must be aliphatic.

14.8 Examples of Anomalies

The final set of calculations illustrate some of the present shortcomings of CLOGP3. For the first five of these, some rational explanation can be given for the discrepancy, and even an estimation of the amount.

1) Overlapping hydrophilic rings: adenosine; this type is usually 1.0 to 2.0 log units underpredicted. The program gives no warning at present.

2) Steroids: cortisone acetate; the program gives a warning that steroids are often calculated low. However, the amount depends on the substitution pattern. Progesterone is calculated well, but polar groups at position 11 are not as hydrophilic as usual; if H-bonding groups are on both the alpha- and beta-position at 17, another positive correction is required.

3) Overlap of hydrophobic chains: 2-amino-1-phenylethanol analogs are well-predicted when the two alkyl chains on the amine are no longer than two carbons. Beyond the diethyl, much of the added hydrophobicity is lost.

4) Shielding of aliphatic nitrogen: morphine; also common in other alkaloids, such as strychnine. The pKa expected for these aliphatic amines is over 9.0, but is observed to be 7.5 to 8.0. Perhaps the same phenomena is responsible for their greater hydrophobicity.

5) Weakened amine in zwitterion: ampicillin; An aliphatic amine in close proximity to carbonyl groups appears much less basic (pKa about 7.0 instead of 9.5). Unless the difference between acid and amine pKas is at least 7.0, the full zwitterion correction does not apply. CLOGP-3.2 does NOT warn of this occurence. An ineffectual zwitterion appears to be present in hydrophilic peptides; e.g. triglycine. Many other penicillins and amino acids are calculated satisfactorily.

6) Aliphatic proximity beyond two I.C.s: dibromodeoxy-galactitol; probably an example of intramol-H-bonding between fragments on alkyl chains.

7) Conformationa 1 proximity: grayanotoxin; probably an example of interaction of polar fragments in proximity through conformation.

Algorithm Manager User Guide

14.9 Algorithm Manager

14.9.1 Introduction

As explained in CLOGP3 Section 14.1.2, the advanced versions are designed as modeling systems, with the final calculated results dependent upon easily accessible files of values for both fragments and interaction factors. This section will explain how to access and, when appropriate, change these 'value files' using the ALGORITHM MANAGER utility. The first of these files, called the Fundamental Value File, contains the values established for Hydrogen and for Isolating Carbons, and also for the correction factors for the interactions of polar fragments. We suggest that this file be used for reference purposes only. The MedChem Project will continue to optimize these values using the annually updated STARLIST as a basis. The second Value File contains the negative corrections for ortho 'decoupling' and also those for intramolecular hydrogen bonding. Users will need to refer to this file to properly characterize any new fragment they wish to enter. Alteration of existing values should be made only after careful consideration of new pertinent data. The third File contains information pertaining to polar fragments and we anticipate it will be altered and extended by many users.

The ALGORITHM MANAGER (AM) can be called whenever the cursor is at the command position of the primary UDRIVE panel by entering 'A' or 'a'. The menu returned in the primary AM panel is:

Enter b(rowse), c(onstant), o(rtho), m(odify), f(iler), r(edraw), exit:

Each of these choices will be explained in the following text.

14.9.2 The Fundamental Value File

The response 'c(onstants)' to the primary AM panel will return a panel with the Fundamental Values and the prompts to modify, delete, forget or update. This file appears in the following format:

```
+--------------------------------------------------------------------------+
|                                                                          |
|   CLOGP3 model constants (User updating not recommended):                |
|                                                                          |
|   Enter m(odify), d(elete), forget, update:                              |
|                                                                          |
|                                                                          |
| HYDROGEN       .227  SR-FUSED2     .250  Y1-C-X(1)     .900  YCY(R2,00)  -.320 |
| IC-ALIPHAT     .195  SR-JOINED     .200  Y1-C-X(2)    1.150  YCY(R2,00)  -.320 |
| IC-AROMAT      .130  SIGMA-DROP    .500  Y1-C-X(3)    1.150  YCY(R2,00)  -.320 |
| FUSION         .100  RHO-DROP      .500  Y2-C-X(1)     .900  YCCY(R0,01) -.260 |
| BIPHENYL       .100  ORTHOVAL     -.280  Y2-C-X(2)    1.300  YCCY(R0,00) -.230 |
| HET-FUSION     .310  HBONDVAL      .630  Y2-C-X(3)    1.700  YCCY(R2,00) -.150 |
| CHAINBRANC    -.130  OCL-AROMAT  21.000  Y3-C-X(1)     .900  Z-APPROX     .200 |
| GROUPBRANC    -.220  OCL-BENZYL  20.000  Y3-C-X(2)    1.800  _____  _____ |
| DOUBLEBOND    -.090  XCX(2)        .600  Y3-C-X(3)    2.700  _____  _____ |
| TRIPLEBOND    -.500  XCX(3)       1.590  YCY(R0,00)   -.320  _____  _____ |
| CHAINBOND     -.120  XCX(4)       2.800  YCY(R0,01)   -.420  _____  _____ |
| RINGBOND      -.090  XCCXVAL       .280  YCY(R0,02)   -.420  _____  _____ |
| ZWITTERION   -2.300  FCCXVAL       .280  YCY(R1,00)   -.320  _____  _____ |
| FBRANCHMAX    7.000  XCCYVAL       .350  YCY(R1,01)   -.370  _____  _____ |
| SR-FUSED       .500  FCCYVAL       .450  YCY(R1,02)   -.370  _____  _____ |
|                                                                          |
+--------------------------------------------------------------------------+
```

As explained above, we suggest that after referring to these values, the user should normally enter the command 'forget'. If the data is to be changed, then after the user enters 'm(odify)', the Algorithm Manager will prompt for the name of the fragment or factor needing modification. The cursor moves to that position, and after the correct value is entered, the 'update' command returns the user to the primary AM panel where there is the option to 'exit'. At this point, the program will ask the file name for storage. The procedure for storing corrections (and/or additions) will also apply to the Ortho Value and Fragment Value Files, and should be thoroughly understood.

To keep the altered data separate from the 'official' MedChem values, the user should enter it in a personal 'delta' file, such as 'JDoe.del'. When the CLOGP3 program is re-entered through UDRIVE, the additions and corrections incorporated in 'JDoe.del' can be utilized by immediately accessing the AM and entering 'f(iler)'. Only the first item on the resulting menu, 'rf' (read formatted), need concern the average user--this will ask for a file of new values which will take precedence over any it finds in the 'official' MedChem values originally supplied. If 'JDoe.del' is supplied at this point, and UDRIVE re-entered, the altered calculations will be made during that session only. If, on the other hand, no personal delta file was created after the 'update' command, then the 'return' key will default the changes to the System Delta File, and they will be part of the calculations of all users from then on. Obviously, as soon as it is apparent that 'JDoe.del' results in better calculations, then that file should be incorporated in the System Delta File. Hopefully, when this occurs, that user will see fit to feed back that valuable information to Pomona MedChem.

The other prompts returned for 'f(iler)' concern the systems manager more than the average user, and are explained in the reference pages of this section.

The only Fragment values appearing in the Fundamental Value File are for: hydrogen, aliphatic and aromatic Isolating Carbon atoms. The definitions of these three fragments are given in the CLOGP3

User Manual. These values have been modified slightly (in the third place past the decimal) to minimize the deviation in calculating STARLIST solutes.

Most of the names of the Factors in this file are abbreviated in a manner which is obvious from reading CLOGP3 Section 14.2. Not so obvious are the following:

FBRANCHMAX (=7) means that the extra (negative) bond correction for branched fragments is dropped after seven bonds (see CLOGP3 14.3.1.1.3).

SR-FUSED (=.5) means the intrinsic sigma/rho interaction for fragments on different (but fused) rings is half that applied to interactions of fragments on the same ring.

SR-FUSED2 (=.25) reduces the sigma/rho correction by a factor of 4 if there is a fused aromatic ring between the two interacting fragments.

SR-JOINED (=.2) reduces the sigma/rho correction by a factor of 5 if the interacting fragments are on different rings of an aromatic system like biphenyl. SIGMA-DROP (=.5) each use of a sigma value in a sigma/rho product 'ages' it so that it's value is cut in half for the next use--unless it is designated (in the 'name' field) as: 'never ages'.

RHO-DROP (=.5) as with sigma-drop, but all types of rho fragments age.

OCL-AROMAT (=21) the ortho class of an aromatic ring adjacent to another fragment; i.e. the 2-position in biphenyl.

OCL-BENZYL (=20) the ortho class of a methyl group. (Note that both these ortho classes are special because the attachment atom is an I.C. rather than an atom which is a part of a polar fragment.)

XCX(2) (=.6) correction for 2 vicinal halogens

XCX(3) (=1.59) correction for 3 vicinal halogens

XCX(4) (=2.80) correction for 4 vicinal halogens

FCCXVAL now has same value as XCCX (=.28), but can be differentiated if future data indicates the need.

FCCYVAL (=.45) fluorine different from other halogens in geminal interaction with H-bonding fragment, Y.

Y#-C-X(#) 'Y' fragments are of three types as indicated by suffix # (see CLOGP3 Section 14.3.2.1.3); (#) refers to the number of halogens on the alpha carbon.

YCY(R#,O#) for vicinal interaction of two 'Y' type fragments, one must specify how many are in a ring as opposed to on a ring or in a chain, and also how many contain -OH; the # following the R gives the number IN a ring and the number following the O gives the number which contain -OH.

YCCY(R#) for geminal interaction of 'Y' fragment pairs, only the number IN a ring need be specified; this is the # following 'R' (0, 1, or 2).

Z-A APPROX is the correction applied to the aliphatic value to get a benzyl approximation (only used if the value for the benzyl environment not known). See CLOGP3 Section 14.3.2.2.1.

14.9.3. The Ortho Value File

The interaction of fragments in ortho positions on aromatic rings can be displayed as a matrix (see CLOGP3 ref. 24). To access these values from the first panel of AM, the user enters 'o(rtho)' and then 'm(odify)'. AM then prompts for the first number of the pair from a menu of fragments. An

example panel for fragment #12, with the values for the corrections of all other fragments ortho to carboxamido, is shown below:

```
+---------------------------------------------------------------------------+
|                                                                           |
|    CLOGP3 Ortho Correction data:                                          |
|                                                                           |
|    Enter first  ortho class i:   12                                       |
|    Enter second ortho class j:                                            |
|                                                                           |
|       #1 CN                  .000      #16 OH                 .630         |
|       #2 NO2               -.840       #17 NH-(Z1)            .630         |
|       #3 CF3                 .000      #18 NH-(Z2)            .630         |
|       #4 I                 -.840       #19 NH2                .630         |
|       #5 Br                -.840       #20 Benzyl-C          -.560         |
|       #6 Cl                -.840       #21 Aromatic-C        -.560         |
|       #7 F                 -.280       #22                    .000         |
|       #8 SO2-NH-(V)          .000      #23                    .000         |
|       #9 C(O)-(W)            .000      #24                    .000         |
|      #10 CHO                 .000      #25                    .000         |
|      #11 CO2H             -1.120       #26                    .000         |
|      #12 CONH-(X)         -1.400       #27                    .000         |
|      #13 O-(Y)             -.280       #28                    .000         |
|      #14 SCH3                .000      #29                    .000         |
|      #15 SH                  .000      #30                    .000         |
|                                                                           |
+---------------------------------------------------------------------------+
```

In order to change any of these values, the number of the second member of that pair must be entered. The cursor moves to the position to accept the new value. Entering the new value with the 'return' key results in the prompts: 'forget' or 'update'. If 'update' is entered, the user may make another modification or else exit. Choosing 'exit' gives the further choice of sending the new value to the private delta file or else to the System Delta File. This choice is explained in the previous section. Note that changing one pair changes the converse pair also; e.g. a new value entered for #12 vs. #17 automatically enters that value for #17 vs. #12.

14.9.4. The Fragment Data Worksheet

Because the Fragment Value File is too large and complex to comprehend in its entirety, the values and other data it contains are accessed individually. If the user merely wants to peruse these values, then the response to the initial 'algorithm manager' panel query should be 'b(rowse)'. To enter a new fragment with its associated data or to change an existing entry AM then prompts for the entry (via SMILES) of a complete solute structure containing the desired fragment. If the Algorithm Manager was entered from the calculation routine of UDRIVE because a SMILES was entered which returned a 'Missing Fragment' message, that SMILES can be retrieved with the entry of '='. This shortcut can only be used if the 'missing' fragment is the only one in the SMILES notation.

To manage the CLOGP3 algorithm properly, the user should enter several kinds of data into the Fragment Data Worksheet, and in order to do this correctly, the user must understand the meaning of: Generic SMILES, Name Specific Data, Orientation Specific Data, and Environment Specific Data.

14.9.4.1. Generic SMILES

The program produces the Generic SMILES for a fragment in which all fragment bonds to Isolating Carbon atoms are replaced with asterisks and the notation is put into a unique order by the CANGEN routine. It is important to realize that the user may encounter the message 'Missing

Fragment Value' in the CLOGP3 output even when a number of values have been entered for that Generic Fragment. This occurs when the bond environment required has not been entered into the Value File and it cannot be estimated from others which are present.

14.9.4.2. Name Specific Data

Each Generic fragment should be given a name to identify it in the CLOGP3 output. This data can be entered after the Fragment Data Worksheet has accepted the SMILES for the structure which contains the new fragment and after using the command, 'n(ame)'. On the following example panel, which resulted on entry of the SMILES CONC(=O)N, the data entry positions (which appear in bold on the terminal) appear in block outline.

```
+--------------------------------------------------------------------------+
|         ===============================================================  |
|         Pomona College MedChem Software -- CLOGP3 Algorithm Manager      |
|         ===============================================================  |
|                                                                          |
| Enter b(rowse), C(onstant), o(rtho), m(odify), f(iler), r(edraw), exit: b|
|                                                                          |
|                                                                          |
| Generic Smiles  *ONC(=O)N                                                |
| Name..........                                                           |
|      NAME SPECIFIC                       ENVIRONMENT SPECIFIC            |
|  Branch correction   .00   environ value err zwit   environ value err zwit |
|  Proximity type... NA      _____ _____ ___ ____   _____ _____ ___ ____ |
|  OH containing.... N       _____ _____ ___ ____   _____ _____ ___ ____ |
|  Sigma never ages. N       _____ _____ ___ ____   _____ _____ ___ ____ |
|   ORIENTATION SPECIFIC     _____ _____ ___ ____   _____ _____ ___ ____ |
|   ori oclas sigma  rho     _____ _____ ___ ____   _____ _____ ___ ____ |
|   ___ ____ _____ ____      _____ _____ ___ ____   _____ _____ ___ ____ |
|   ___ ____ _____ ____      _____ _____ ___ ____   _____ _____ ___ ____ |
|   ___ ____ _____ ____      _____ _____ ___ ____   _____ _____ ___ ____ |
|   ___ ____ _____ ____      _____ _____ ___ ____   _____ _____ ___ ____ |
|   ___ ____ _____ ____      _____ _____ ___ ____   _____ _____ ___ ____ |
|   ___ ____ _____ ____      _____ _____ ___ ____   _____ _____ ___ ____ |
+--------------------------------------------------------------------------+
```

After the name has been entered, the cursor moves to prompt entry of other data which does NOT depend on symmetry or bonding environment, namely:

Size of Branch Correction = 0.0, -0.08, or -0.19; see CLOGP3 Section 3.1.1.3.

OH Containing = 'Y' or 'N'; see CLOGP3 Section 14.3.2.1.2.

Sigma Never Ages = 'Y' or 'N'; 'Y' is for fused-in aromatic fragments containing an unsubstituted nitrogen; e.g. *n*, *nn*, *[nH]n* etc.

14.9.4.3. Orientation Specific Values

Upon entry of the 'o(rientation)' command, the program will return a panel with numbers above the bonding atoms for multivalent fragments. These numbers show the sequence in the Generic SMILES (in which asterisks are counted). If the fragment is symmetrical, the SAME data should be entered in the columns headed by 'oclas', 'sigma', and 'rho'. If the fragment is asymmetrical, as in the following carbamate example, then the values in each column will depend on the attachment atom. If no values are entered, the program will assume that no correction of this type is necessary. In the example panel, values have already been entered in these columns; the cursor is asking for an entry of either '2' or '5' so that a value change can be made.

14.9.4.3.1. Worksheet for Carbamate Fragment

```
+--------------------------------------------------------------------------+
|  ORIENTATION:  Ortho class, sigma and rho are dependent on               |
|     orientation of the fragment as it is attached to a ring.             |
|     Symmetric orientations normally contain identical data.              |
|                                                                          |
| Enter number corresponding to the orientation:                          |
|                                                                          |
| Orientation...  2      5                                                 |
|                 |      |                                                 |
| Generic Smiles  *NC(=O)O*                                                |
| Name..........  NH-Carbamate                                             |
|      NAME SPECIFIC                      ENVIRONMENT SPECIFIC             |
|  Branch correction   .00    environ value err zwit    environ value err zwit |
|  Proximity type... Y2       VV      -1.10  20   0     _____ _____ ___ ____ |
|  OH containing.... N        Aa      -1.46   0   0     _____ _____ ___ ____ |
|  Sigma never ages. N        AA      -1.79   0   0     _____ _____ ___ ____ |
|   ORIENTATION SPECIFIC       aa       -.44   0   0     _____ _____ ___ ____ |
|   ori oclas sigma  rho      aA       -.91   0   0     _____ _____ ___ ____ |
|    5  13     .17   .50      YA      -1.10  40   0     _____ _____ ___ ____ |
|    2  17     .00  1.08      _____ _____ ___ ____     _____ _____ ___ ____ |
|                                                                          |
|   ___ ____ _____ ____      _____ _____ ___ ____     _____ _____ ___ ____ |
|   ___ ____ _____ ____      _____ _____ ___ ____     _____ _____ ___ ____ |
|   ___ ____ _____ ____      _____ _____ ___ ____     _____ _____ ___ ____ |
|                                                                          |
+--------------------------------------------------------------------------+
```

If the user does not have partitioning data to establish these three values for the new fragment being entered, data can often be estimated using the tables found at the end of this Section (which is modified slightly from those appearing in CLOGP3 ref. 24). In this example, the program tells us that the nitrogen has the '2' orientation, and from the tables in this section the other entries are: oclas = 17, sigma = 0, and rho = 1.08. For the oxygen, they are: an orientation of '5', oclas = 13, sigma = .17, and rho = .50.

14.9.4.4. Environment Specific Values

In the calculation portion of UDRIVE a user may have entered O-styryl-N-methylcarbamate and received a 'missing fragment' message. After entering AM the above panel shows that the Name, Name Specific, and Orientation Specific data have been previously entered, as well as many values for specific bonding environments, including the one with the styryl and alkyl bonding reversed. The computer cannot calculate the desired value from those already entered, and so a new environment-dependent value must be entered. The user should enter 'e(nvironment)', and, following the prompt asking for a complete SMILES, enter either '=' or 'c1ccccc1C=COC(=O)NC'. Then the following panel appears:

14.9.4.4.1. Worksheet for O-styryl-carbamate

```
+------------------------------------------------------------------------------+
|   SPECIFIC ENVIRONMENT:  Fragment hydrophobicity value and zwitterion class  |
|      are dependent on the electronic (bonding) environment of a fragment.    |
|      Specify this by entering a COMPLETE structure containing the fragment   |
|      in the context of interest; the fragment of interest must be the ONLY   |
|      fragment.  A response of "=" means "same as the last SMILES looked up". |
|                                                                              |
| Complete Smiles ANC(=O)OY                                                    |
|                      |      |                                                 |
| Generic Smiles  *NC(=O)O*                                                    |
| Name.........   NH-Carbamate                                                 |
|      NAME SPECIFIC                          ENVIRONMENT SPECIFIC             |
|  Branch correction   .00    environ value err zwit    environ value err zwit |
|  Proximity type... Y2       VV    -1.10  20   0       _____ _____ ___ ____ |
|  OH containing.... N        Aa    -1.46   0   0       _____ _____ ___ ____ |
|  Sigma never ages. N        AA    -1.79   0   0       _____ _____ ___ ____ |
|   ORIENTATION SPECIFIC      aa     -.44   0   0       _____ _____ ___ ____ |
|   ori oclas sigma  rho      aA     -.91   0   0       _____ _____ ___ ____ |
|    5   13    .17   .50      YA    -1.10  40   0       _____ _____ ___ ____ |
|    2   17    .00  1.08      AY    -1.20 ___ ____       _____ _____ ___ ____ |
|                                                                              |
|   ___ ____ _____ ____       _____ _____ ___ ____     _____ _____ ___ ____ |
|   ___ ____ _____ ____       _____ _____ ___ ____     _____ _____ ___ ____ |
|   ___ ____ _____ ____       _____ _____ ___ ____     _____ _____ ___ ____ |
+------------------------------------------------------------------------------+
```

Note that the 'complete SMILES' has been replaced with 'ANC(=O)OY' and the environment appears with the cursor at the position for entry of the appropriate value. Note also that this is different from the entry: YA = -1.10. We may wish to estimate the AY environment as slightly more hydrophilic than the YA (the oxygen's electrons not being as easily delocalized), and enter a value of -1.20. The cursor then prompts us to enter a value for 'err', which in this case is 40 for 'a priori'. Normally a user following this procedure would be entering a value from a MEASURED solute, and therefore the 'err' entry would be 0.0. In the column headed 'zwit', a 0.0 is entered when the fragment in that particular environment cannot form a zwitterion with even the strongest of the opposite species, which is the case with the carbamate in the example above. An aromatic amine would be given a 'zwit' value of -1 (weak) and an aliphatic amine -2. Conversely, an aliphatic or aromatic carboxylic acid would be given a value of +1 while a sulfonic acid would get +2.

14.9.5. Help

If, while in the AM, the user is not sure of the proper procedure, a Help panel may be called by entering '?' at any point where the cursor expects either a value or a symbol such as 'Y' or 'N'. An invalid entry automatically returns a help panel to the user.

14.9.6. Deleting Data from the AM

Mistakes in entering information into the Algorithm Manager are bound to occur. The most frequent of these is likely to be an erroneous value or classification. These can be corrected using the 'm(odify)' command as explained above. However, a user may enter a completely spurious SMILES for a 'private' fragment but with valid data. In this case all the information entered should be copied before beginning the deletion procedure. To delete a fragment enter the AM as usual and enter 'm(odify)' at the first prompt. The next panel gives the option 'd(elete)'. As soon as that is entered, enter the option 'n(ame)'. This entry will delete ALL information connected with this fragment if all the following steps are carried out: In the following panels the responses should be: 'update', 'exit' and then the entry of the proper delta file.

A more common deletion might remove an incorrect designation of an environment. As an example, we might imagine that the value for the O-styryl carbamate shown in the previous section properly applied to the N,N-dimethyl derivative, which is a completely different fragment. Deletion of the erroneous AY entry would proceed as follows: After using the 'm(odify)' and 'd(elete)' commands as in the previous paragraph, the next entry is 'e(nvironment)'. The next panel asks for the environment you wish to delete, which is 'AY'. When this is entered, and followed by 'update', 'exit', and entry of the proper delta file, the deletion of AY is complete.

Sigma and Rho constants of generalized structures

```
-------------------------------------------------------------------------
No.     Sigma   Rho    Generalized                 Examples
                       Structure
-------------------------------------------------------------------------
1.      1.00    1.17   =N(=O)-     pyridine-N-oxide
2.      0.84    0.21   -N=         pyridine, quinoline
3.      0.71    0.00   -SO2F
4.      0.65    0.00   -SO2X       X=alk,    N(Me)2
5.      0.65    0.00   -CN
6.      0.60    0.00   -NO2
7.      0.49    0.00   -CF3
8.      0.28    0.00   halogens
9.      0.58    0.44   -CHO
10.     0.51    0.27   -C(=O)-X    X=alk,    OCH3,    C6H5,     N(Me)2
11.     0.32    0.35   -CO2H
12.     0.32    0.72   -CONH-X     X=H,      NH2,     C6H5,     Alk
13.     0.17    0.50   -O-X        X=alk,    CONHCH3, CON(Me)2, CH2CO2H,
                                     PO(O-alk)2,      NOT C6H6
14.     0.25    0.88   -SO2NH-X    X=H,      C6H5
15.     0.00    0.50   -S-X        X=H
16.     0.00    0.30   -S-X        X=alk
17.     0.00    0.61   -NX2        -N(Me)2,-N=NN(Me)2
18.     0.00    1.06   -OH
19.     0.00    1.08   -NH-X       X=alk,    COMe,    CON(Me)2, CHO,
                                     C6H5,    SO2CF3,  CONHC6H6
20.     0.00    0.90   -N(X)NH2    subst. hydrazine
21.     1.34    0.00   arom =N-N=  diazine
22.     0.50    0.00   arom =N-O-  isoxazole
23.     0.00    0.50   arom -O-    o1cccc1 furan
24.     0.00    0.40   arom -S-    s1cccc1 thiophene
25.     0.00    0.70   arom -NH-N= pyrazole
26.     0.00    0.80   arom -NH-   Hn1cccc1 pyrrole
27.     0.00    1.00   -N(C(=O)X)2 X=alk ring
```

```
                    Matrix of ortho value corrections by ortho class (1-21)
                         Tabulated values are multiples of -0.28.


-------------------------------------------------------------------------------
1  2  3  4   5   6   7  8  9  10 11 12 13 14 15 16 17 18 19 20 21 22
-  -  -  -   -   -   -  -  -  -  -  -  -  0  -  -  0  -  -  -  -  -  -  (1)  -CN
   0  0 (1)(1)  1  (1) 1  1  0  1  3  1  -  -  0  -  2  H* 0  1  -  (2)  -NO2
      -  -   -   -   -  -  -  -  -  -  -  -  -  0  -  -  -  -  -  2  (3)  -CF3
         -   -   -   - (2)(1) - 2 (3) 1  -  -  1 (2) -  -  -  -  2  (4)  -I
             1   -   - (2)(1) - 2  3  1  -  -  1  2 (2) -  -  -  2  (5)  -Br
                 0   -  2  1  -  2  3  1 (2) -  1  2  2  0  0  - 1.5 (6)  -Cl
                     0  (1) 0  -  1  1  0  -  -  1  1 (1) 0  -  -  .5 (7)  -F
                        -  -  -  -  -  -  -  -  - (H)(H)(H) H  0  -  -  (8)  -SO2N-(V)
                           3  -  3  -  1  -  -  H  H  H  H  0  -  -  (9)  -CO-(W)
                              -  -  -  -  -  -  H  H  H  H  0  -  - (10)  -CHO
                                 4 (4) 1* -  0  H  H  H  H  1  -  - (11)  -CO2H
                                    5  1  -  -  H  H  H  H  2  2  - (12)  -CONH-(X)
                                       2  -  -  1  -  2 (0) 0  -  0 (13)  -O-(Y)
                                          -  -  -  -  -  1  -  -  - (14)  -SCH3
                                             -  -  0  -  0  -  -  - (15)  -SH
                                                0  -  -  H* 0  -  - (16)  -OH
                                                   -  -  0  3  -  - (17)  -NH-(Z1)
                                                      -  -  3  -  - (18)  -NH-(Z2)
                                                         -  -  0  - (19)  -NH2
                                                            -  -  1 (20)  -CH3
                                                               -  - (21)  -Phenyl
                                                                  - (22)  -N(Z3)2


   Symbol          Meaning              Symbol          Meaning
   ------          -------------        ------          --------------------
   Normal  ->   Measured                (V)      ->   -H
   (X)     ->   Estimated               (W)      ->   -Me,  -OMe,  -N(Me)2
   -       ->   No data                 (X)      ->   -H,   -NH2
   0       ->   Zero effect             (Y)      ->   -Me, -COMe, -CONHMe,
   *       ->   5 if Y=phenyl                          -CON(Me)2,  -CH2CO2H
   H*      ->   +.40 to +50             (Z1)     ->   -CONH2
   H       ->   +.63 (H-bond)           (Z2)     ->   -COMe, -SO2(a)
                                        (Z3)     ->   -C(=O)R, R=alk ring
```

14.10. Algorithm control files

This reference section is intended to explain the external file conventions used by the CLOGP3 Algorithm Manager. AM permits changes in the CLOGP3 algorithm. Changes may be temporary (in core only) or permanent (by modification of the CLOGP3 control files). Proper use of AM's external files allows original distribution, distributed updates, system, public, and private changes to the CLOGP3 algorithm to be accessed with any priority desired.

The Algorithm Manager (AM) controls the CLOGP3 modeling algorithm by virtue of the fact that CLOGP3 obtains all constants, corrections, and fragment information from a special "database", FRAGDB. In normal use, FRAGDB reads all modeling information from file(s) at the start of a run and passes it on to CLOGP3 as required. The primary FRAGDB data file is named FRAGDB.DAT. FRAGDB.DAT contains all the information pertaining to log(P) computations in an original MedChem distribution. Since FRAGDB.DAT is static and quite large it is converted to a binary image (FRAGDB.BIN), which is the actual file used. If FRAGDB.BIN is not available, FRAGDB will read in FRAGDB.DAT and create FRAGDB.BIN for future use.

Once the distribution set is established, a default update file is read (UPDATE.DEL). Delta files (names end .DEL) are formatted files which describe the CLOGP3 algorithm in the same format as FRAGDB.DAT. The main thing to remember about .DEL files is that later entries supercede earlier ones. Therefore anything described in UPDATE.DEL will override the same information in the original distribution set FRAGDB.DAT. Pomona MedChem will distribute updates as UPDATE.DEL files.

A delta file may include another delta file at any point. The syntax of that instruction line is:

READ|FILENAME.DEL

This allows free specification of changes in any order you like. For instance, imagine that your UPDATE.DEL file reads:

READ|TIM.DEL

READ|SYSTEM.DEL

READ|JOE.DEL

READ|MEDCHEM.DEL

READ|SUPERHACKER.DEL

All entries in all files will supercede the distribution version as read in FRAGDB.BIN. Since SUPERHACKER's delta file is last, entries in SUPERHACKER.DEL will supercede anything that conflicts with it in other files. In this case, the MedChem update MEDCHEM.DEL has higher priority than all others except SUPERHACKER. JOE's delta file is considered to be more reliable than the SYSTEM's file, but not as reliable as MEDCHEM's. TIM isn't sure of his delta file at all; if anybody disagrees with TIM, TIM loses.

Algorithm Manager Reference

Everytime you exit AM after changing anything, AM asks if you want to save the changes in a delta file. A default file name is suggested, but any name at all can be entered. The default file name is the local name specified as SYSTEM.DEL in RENAME.DAT.

As if all the above flexibility isn't enough, recall that all "standard" MedChem filenames are defined as local files using the RENAME facility. FRAGDB.DAT, FRAGDB.BIN, UPDATE.DEL, and SYSTEM.DEL are all defined in RENAME.DAT. See the PMCIO Reference Manual for information on how to change the names.

Note that information in all files referenced by update.del will be available to all users on the system.

14.11. Algorithm Manager interface

AM has access to all algorithm control files, FRAGDB variables, and global areas that the FRAGDB (and thus CLOGP) has. Anything they can read, AM can read. AM can also append onto the algorithm control files. The "filer" command in the top AM level allows flexible and powerful control via this access.

14.11.1. Write Unformatted -- "Create snapshot"

There are four kinds of I/O operations that one might want to do via AM. Writing an unformatted file (MEDCHEM.BIN) is done by Pomona MedChem before each major release. The unformatted file thus created contains the complete state of the FRAGDB at the time of writing. Reading that .BIN file will reset FRAGDB to the state it was in at the time it was written. Writing unformatted files is strongly discouraged. It is opposed because it is almost always easier to write and use a

system delta file. Unless you have >100-200 changes it will be quicker to use the formatted read facility.

Algorithm Manager Reference

14.11.2. Read Unformatted -- "Start from scratch"

An unformatted read is how FRAGDB first reads in the primary distribution version at the start of each run. Doing an unformatted read of FRAGDB.BIN unconditionally resets the CLOGP3 algorithm to the original distribution state. No updates then apply.

14.11.3. Write Formatted -- "Dump the works"

A "wf" (write formatted) request causes the entire contents of FRAGDB to be written to a file you specify in .DEL file formatted form. It is slow, not particularly readable, and in no sense better than an unformatted write (unless you feel better with an instantly-obsolete obese backup). In short, don't.

On the other hand, it is very wise and reasonable to write short amounts of formatted data. In fact, this is what is done every time you leave AM. The difference is that saving changes in delta files is 1000 times faster than saving the whole works in delta file format.

14.11.4. Read Formatted -- "Learn from a file"

Reading a formatted file is the recommended way to access the information on private files, and the only use of the AM "filer" facility that 99%+ of users want or need.

Note that the file specifications of the last section (via UPDATE.DEL) force public access. (Each file read in supercedes previous ones for all users.) This is fine for updating purposes, but there are occasions when private access might be preferred. Private files are required for some development and security purposes. You may wish to make a preliminary test on a set of fragment values in which you have no confidence at all. Or possibly the data you are entering is meant to be unavailable to other users of CLOGP3 on your machine (confidential or proprietary). In either case you will not want to slip it into the SYSTEM.DEL stream.

To read a formatted file, simply select "rf" (read formatted) and specify the .DEL file name desired. All the infor- mation in that file will then be available; any conflicts will be superceded by the new information. The data will be integrated into FRAGDB in memory but will not be written out nor made accessible to others. If you decide to include the file for public access, simply include the name in UPDATE.DEL in a position reflecting the priority you feel it should deserve.