AFRL-RH-WP-TR-2008-0068

Comparison of the Performance of Selected Commercially Available Programs for the Estimation of the Logarithm of the Octanol-Water Partition Coefficient, Log P

Eric K. Styron
John M. Frazier
AFRL/RHPB
2729 R Street, Bldg 837, Area B
Wright-Patterson AFB OH 45433-5707

May 2001

Air Force Research Laboratory
Human Effectiveness Directorate
Biosciences and Protection Division
Applied Biotechnology Branch
Wright-Patterson AFB OH 45433-5707

NOTICE AND SIGNATURE PAGE

Using Government drawings, specifications, or other data included in this document for any purpose other than Government procurement does not in any way obligate the U.S. Government. The fact that the Government formulated or supplied the drawings, specifications, or other data does not license the holder or any other person or corporation; or convey any rights or permission to manufacture, use, or sell any patented invention that may relate to them.

This report was cleared for public release by the 88th Air Base Wing Public Affairs Office and is available to the general public, including foreign nationals. Copies may be obtained from the Defense Technical Information Center (DTIC) (http://www.dtic.mil).

AFRL-RH-WP-TR-2008-0068 HAS BEEN REVIEWED AND IS APPROVED FOR PUBLICATION IN ACCORDANCE WITH ASSIGNED DISTRIBUTION STATEMENT.

//SIGNED//	//SIGNED//
David R. Mattie	Mark M. Hoffman, Deputy Director
Work Unit Manager	Biosciences and Protection Division
Applied Biotechnology Branch	Human Effective Directorate
	Air Force Research Laboratory

This report is published in the interest of scientific and technical information exchange, and its publication does not constitute the Government's approval or disapproval of its ideas or findings.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*	2. REPORT TYPE	3. DATES COVERED *(From - To)*
May 2001	Tech Report	April 2000 – April 2001

4. TITLE AND SUBTITLE
Comparison of the Performance of Selected Commercially Available Programs for the Estimation of the Logarithm of the Octanol-Water Partition Coefficient, Log P

5a. CONTRACT NUMBER

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER
62202F

6. AUTHOR(S)
Eric K. Styron, John M. Frazier

5d. PROJECT NUMBER
2312

5e. TASK NUMBER
A2

5f. WORK UNIT NUMBER
05

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
AFRL/RHPB

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Air Force Materiel Command
Air Force Research Laboratory
Human Effectiveness Directorate
Biosciences and Protection Division
Aircrew Performance and Protection Branch
Wright Patterson Air Force Base OH 45433-7947

10. SPONSOR/MONITOR'S ACRONYM(S)
AFRL/RHPB

11. SPONSORING/MONITORING AGENCY REPORT NUMBER
AFRL-HE-WP-TR-2008-0068

12. DISTRIBUTION AVAILABILITY STATEMENT
Approved for public release; distribution is unlimited.

13. SUPPLEMENTARY NOTES

14. ABSTRACT
This report documents a comparative study on the performance of five commercially available programs that perform a total of eight different computational methods for the estimation of the logarithm of the octanol-water partition coefficient (log P). The choice of the programs for employment was based on cost and the need for a comparison of the fragment-based and nonfragment-based methodologies. A database of 150 compounds, for which the log P value is accurately known, has been accumulated and the bank of programs utilized to give eight estimated values for the log P of each compound. Each of the eight sets of 150 data points has been evaluated on the basis of the square of the correlation coefficient, and the slope and intercept of the linear best-fit line as calculated by Microsoft® Excel 97 when the data is plotted against the corresponding experimental values. This report compares the performance of the eight different computational methods, with respect to specific structural characteristics, based on the above graphical parameters. This study also found that the accuracy and precision of each of the methods is dependent on the hydrophobicity and structural classification of the compound(s) under scrutiny.

15. SUBJECT TERMS
Octanol-water partition coefficient, log P, Kow, computational methods, software programs

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			David R Mattie
U	U	U	SAR	30	19b. TELEPONE NUMBER (*Include area code*)

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI-Std Z39-18

THIS PAGE INTENTIONALLY LEFT BLANK.

PREFACE

This work was performed by Air Force Research Laboratory, Human Effectiveness Directorate, Operational Toxicology Branch, Wright-Patterson Air Force Base, Ohio. This work was supported by the Air Force Office of Scientific Research (AFOSR) Entrepreneurial Research Project (23 12A205) Toxicity of High Energy Chemicals during the period of Aug 2000 - Apr 2001. All software packages evaluated in this report are still available, although changes may have been made in the programs since the completion of this report.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

INTRODUCTION

The AFRL Propulsion Directorate has a sustained effort toward the discovery of new propellants. In an attempt to accelerate the deployment of these new propellants a collaboration with HEST was employed to evaluate available computational tools to gain more expedient and cost effective insight into the potential toxicity hazards of these chemicals. A series of twelve compounds were selected on the basis of their structures being representative of newly emerging high-energy compounds or derivatives of such compounds. As can be seen from their nomenclature, hydroxyethylhydrazine, methylhydrazine, diethylhydrazine, 1,2,4-triazole, 4-amino- 1,2,4-triazole, methoxylamine, hydroxyethylamine, ethylimidazoleamine, nitroaminoguanidine, diaminoguanidine, dihydrazinotetrazine, and 2,2-dimethyltriazanium nitrate, the compounds cover a variety of structural types. All of the chemicals were sent in the form of nitrate salts to HEST for in vitro studies.

Many of the physical-chemical properties of the selected compounds have not been determined experimentally. Therefore, it was necessary to estimate the relevant parameters by computational methods. The following software packages were selected for this purpose: Molecular Modeling Pro by ChemSW Inc., ClogP by Biobyte, ScilogP by Scivision, Kowwin by SRC, and ACD Lab's Log P suite. However, log P is not defined for ionized species; therefore only the neutral compounds were used for computational scrutiny. Since 2,2-dimethyltriazanium nitrate has no uncharged form, for the purposes of this comparative study it was eliminated and replaced with hydrazine. In recent years the parameter log D has been used to incorporate the tendency of many compounds to ionize in solution. Log D is the ratio of the equilibrium concentrations of all species (unionized and ionized) of a molecule in octanol to same species in the water phase. It differs from LogP in that ionized species are considered as well as the neutral form of the molecule. This technical report will be examining only the ability of the software to estimate values of log P.

It was found that six of the compounds had accurately determined experimental values for the log of the octanol-water partition coefficient associated with them. The calculated values were then plotted against the experimental values and the square of the correlation coefficient used as a measure of precision, and the slope and intercept of the linear best-fit lines were used to evaluate the accuracy of each of the programs. The overall performance of these types of programs was found to be somewhat unpredictable, ranging from very accurate values for some compounds to very inaccurate for others. This led to the construction of a database composed of compounds for which the logarithm of the octanol-water partition coefficient, log P, is accurately known. This database was divided into different structural classes for evaluation of each program's performance with regard to structure. The purpose of these divisions is to determine which program will yield the most dependable estimation for new compounds of a particular structural type.

A database of 150 compounds has been selected from the "starlist" of published experimental log P values and put through the battery of programs. The data has been evaluated and despite an overall acceptable correlation coefficient when all data points are included, the programs show a large variance of accuracy and precision with structural type.

SUMMARY OF PROGRAMS USED FOR THE CALCULATION OF LOG P VALUES

ClogP 4.0 for Windows is a fragment-based method by Hansch and Leo. The manner in which a fragment constant addition method works is as follows. The structure is divided into fragments

and predetermined values of each group are summed together to yield the log P estimate.[1] The compound to be estimated may be entered in SMILES notation or imported as a MolFile. ClogP is produced by Biobyte, which was cofounded by Hansch and Leo.

The Molecular Modeling Pro (MMP) package is produced by ChemSW Inc. This program possesses a molecule building module from which three-dimensional molecules may be saved in MolFile, AMPAC/MOPAC, or bitmap form. This program includes four different modules for the calculation of log P values.

The first is a fragment-based method described by the users' manual as a modified Hansch fragment constant addition methodology in which most but not all of the constants are from Substituent Constants for Correlation Analysis in Chemistry and Biology.[2] This module is not parameterized for inorganic compounds.

The second module is an atom-based constant addition methodology derived from the work of Ghose and Crippen.[3] Atom classifications may be found in reference 3.

The last two methods are integrated into the CNDO and INDO modules. These are based on a method developed by Moschner and Cece.[4] The original version uses Gasteiger-Huckel charges and other atomic properties to calculate Log P values. ChemSW Inc. modified this method to work with the closed shell CNDO method of the MMP program. According to the users' manual, "the regression coefficients were changed somewhat, terms were added for aliphatic F and Cl, and charge on sulphur."

SciLogP Ultra is a non-fragment based approach. The producers describe it as a neural net approach that uses two-dimensional molecular descriptors.[5] The compound to be estimated may be entered in SMILES notation or imported as a MolFile. SciLogP Ultra is produced by Scivision.

Kowwin uses the "atomic fragments" or "fragment constant" methodology developed by Meylan and Howard.[6] In short, this is an atom-based constant addition method in which some multi-atom fragments are used. At the time of this survey the only method of structure entry was SMILES notation, although batches were be processed by saving the SMILES notations of the compounds into a text file. The updated version allows the user to enter the CAS number if the compound has been previously reported; it then searches the internal SMILES database for the structure. Kowwin is produced by the Syracuse Research Corp. and is the program employed by the Environmental Protection Agency.

Log P DB 4.5 uses a "additive-constitutive" methodology which is similar to the techniques developed by Hansch and Rykker and is produced by ACD Labs.[7] The calculation engine adds up contributions from separate atoms, structural fragments, and intramolecular interactions between fragments. Compounds must be drawn in ChemSketch.

RESULTS AND DISCUSSION

The master list of compounds and calculated data may be found in Appendix A along with the respective graphs. The data are arranged in ascending order of the experimental log P values. SciLog P, Clog P and Kowwin produced the best squared correlation coefficients (R^2) for all data, giving 0.9606, 0.9770, and 0.9664, respectively. Appendix B contains the same data broken up into nine structural types: aromatics and substituted benzenes, hydrazines,

guanidines and ureas, aromatic nitrogen-heterocycles, amino acids, nitro compounds, sulfur compounds, non-aromatic heterocycles, and cyclic alkanes. Although SciLog P, Clog P and Kowwin performed the best overall, this does not hold when considering an individual structural type. Demonstrating this point is the object of the next two sections.

In the next section the data is grouped and surveyed by structural type. The subsequent section reviews the performance of each program, identifying groups that appear to be problematic specifically for that program.

SURVEY BY STRUCTURAL GROUP

As expected, the aromatics/substituted benzenes group, for which all of the programs performed well, may be used to demonstrate a typical well-behaved class of compounds.

The hydrazines were best estimated by Clog P. Although log P DB 4.5 achieved a slightly higher R^2, its slope and intercept were poor. This does not appear to be a problematic structural class. The most obvious problems here are with hydrazine itself, the closest estimation was 0.4 units too high, and acetylhydrazine, for which most estimates were about a half a unit too low.

The guanidine/urea structure type is seen to be a very problematic class. Only the Crippen and SciLog P routines gave R^2 values above 0.9000 but with unacceptable slope and intercept values. The other seven programs produced only mediocre precision with poor accuracy.

It proved to be exceedingly difficult to obtain good estimates for aromatic nitrogen-heterocycles. Log P DB was the only program to produce acceptable estimates for these compounds. Excellent values for the slope and intercept were attained, giving a slope of 1.04 and an intercept of 0.04. Its precision was moderate though, with an R^2 of 0.8799. Although an R^2 of 0.8799 is not inherently impressive, it must be noted that the next highest R^2 was 0.7152 by Kowwin.

The amino acids are another problematic group. Although five programs gave moderate to good precision, all of them displayed poor accuracy. Kowwin produced the best slope and intercept by far with values of 1.12 and 0.17, respectively, its R^2 was moderate at 0.8657.

Another terribly difficult type of compound to estimate are those containing nitro groups. None of the programs produced an R^2 exceeding 0.7500 and all of the slopes and intercepts were deficient. So, unlike the aromatic nitrogen-heterocycles there was no one program that stood out as being superior.

Estimations of sulfur compounds by the Clog P, Log P DB, and Kowwin programs displayed sufficient accuracy and precision for this class to be described as well-behaved. A surprising result was the poor performance of SciLog P with this type of molecule. No results are displayed for the INDO method since it is not parameterized for atoms with atomic number greater than nine.

The non-aromatic heterocycles group turned out to be easily modeled by SciLog P, ClogP, and Log P DB. This is an unexpected merit considering the previously mentioned results for the aromatic nitrogen heterocycles group. Kowwin gave only mediocre estimations with an R^2 of 0.8846, a slope of 1.11, and an intercept of 0.14. Despite this the group can still be regarded as well-behaved.

The last group, cyclic alkanes, was well-behaved for SciLog P, Clog P, Log P DB, and the MMP fragment-based module. Although the Kowwin program gave very good precision for this group, its accuracy was very poor with a slope of 0.86 and an intercept of 0.37.

Now that the structural groups have been defined and analyzed, it is of interest to note a general trend observed in the all-data plots: the precision of estimations tends to decrease with increasing hydrophilicity of the compounds. That is to say, there is less scattering of the data points above zero than below zero. This can also be demonstrated by comparing the correlation coefficients for positive and negative experimental values, as shown in Appendix A. Of the top three programs, SciLog P shows the largest discrepancy. The R^2 falls from 0.9471 for hydrophobic compounds to 0.8539 for hydrophilic compounds. This suggests that the programs show a large variance in accuracy and precision with hydrophobic character as well as structural type.

SURVEY BY PROGRAM

It is difficult to assign a label of "best program" when the parameters are so close. When the plots are examined the first thing that stands out is the amount of scatter, that is to say the deviance of R^2 from one, therefore we begin with Clog P since it displayed the highest correlation coefficient, albeit by only a slight margin.

The overall parameters for Clog P were very good with R^2, slope and intercept values of 0.9770, 0.98, and -0.05 respectively. However, this is deceiving, as this paper aims to illustrate. When the data for the guanidine and urea derivatives are examined, we find that Clog P gives only a moderate amount of scatter, but a resultant slope of 0.84 tends to indicate that error in these values increases with distance from zero. In addition, an intercept of -0.39 shows that estimated values near zero are not reliable either. Therefore, the estimated values from Clog P for this structural type can not be relied upon.

Another problem area for Clog P is that of aromatic heterocycles containing nitrogen. The slope and intercept are irrelevant due to the irregularity of values obtained for these compounds, R^2 = 0.5727.

When dealing with amino acids, R^2 = 0.8219, some coherence is achieved. Nonetheless a slope of 0.75 and an intercept of -0.59 reveal this group to be estimated very inaccurately by Clog P.

Lastly, compounds containing nitro groups cause a high degree of scatter in Clog P calculations, giving an R^2 of 0.7134.

In order of descending correlation coefficient, the next program in line is Kowwin. With an altogether R^2 of 0.9664 Kowwin shows only a slight increase in divergence relative to Clog P. Kowwin also exhibited a minor increase in accuracy over Clog P for this data set with a slope of 1.01 and an intercept of 0.00. However, aside from an exceptional overall exhibition, Kowwin experienced trouble with certain structural types.

In addition to a low to moderate R^2 of 0.8319, Kowwin returned very poor slope and intercept values of 0.77 and -0.38, respectively, for substituted guanidines and ureas. In light of this, Kowwin log P estimations for this type of molecule should be considered questionable.

The Kowwin modeling of aromatic heterocycles also returned questionable results. Despite excellent values for the slope and intercept, an R^2 of 0.7153 indicates an unacceptable amount of scattering in the data points.

Amino acids present another modest obstacle in the use of Kowwin. None of the graphical observables can be considered poor by themselves, but moderate scattering, R^2 of 0.8657, coupled with less than average values for the slope and intercept, 1.12 and 0.17, is enough to create a troublesome area.

Nitro compounds appear at first to present a problem for Kowwin. However, Kowwin has no ability to distinguish between structural isomers, and the data set of nitro compounds contains a high percentage of structural isomers. This is simply a limit imposed by using only accurately determined experimental Log P values, that is, you have to use what's available. Therefore, no conclusions can be drawn about Kowwin's treatment of this class of compounds.

Non-aromatic heterocycles pose a problem similar to that of the amino acids for Kowwin. None of the parameters are poor, but all are on the low end of moderate.

Kowwin achieved excellent linearity for the cyclic alkanes, but the slope is poor at 0.86, and an intercept of 0.37 is very disappointing.

Next, is Scilog P. The precision of Scilog P for our data set is comparable to that of Clog P and Kowwin, R^2 equals 0.9606. But the fall in accuracy is noticeable with a slope of 0.92 and an intercept of 0.13.

Scilog P experiences very little trouble until is forced to deal with the guanidine and urea type compounds. Although the degree of scatter is average, a slope of 0.80 and an intercept of -0.30 for this group are quite unacceptable.

As with the previous programs, SciLog P also performed very poorly when dealing with aromatic nitrogen heterocycles. Just as in the case of Clog P, the inferior value for the slope is unimportant due to the notably high degree scattering, $R^2 = 0.6362$.

SciLog P behaves well when confronted with amino acids. The only concern is the deviance of the intercept to 0.87.

The conclusion that nitro compounds pose a definite problem for SciLog P is evidenced by an R^2 of 0.7454 and a slope of 0.6733.

Sulfur compounds also introduce uncertainty into the overall SciLog P performance. The low R^2 value of 0.8382 is compounded by very poor slope and intercept numbers of 0.82 and 0.34, respectively.

Following in order of R^2, we come to ACD Lab's Log P DB. The poor overall correlation coefficient, 0.8301, disguises the performance of this program. Log P DB had no more trouble areas than the top three programs and outperformed them in some of the well-behaved structural groups. However, when it performed poorly, it performed exceptionally poorly. Also, a very pronounced dependence on hydrophobicity can be seen in the graph of the Log P DB data. The R^2 deteriorates from 0.9598 for hydrophobic compounds to 0.1970 for hydrophilic compounds.

Log P DB gave the highest R^2 value for the substituted hydrazines group, 0.9656. The slope and the intercept, on the other hand, were very poor at 0.82 and 0.20.

The estimations for the guanidine type compounds appeared virtually random with a correlation coefficient of 0.3634. The slope and intercept values were equally poor.

The amino acids were also a problem for Log P DB. This is surprising since these compounds are included in the structural templates of the program and recognized by the internal database. The amino acids can be plainly seen in the overall plot forming a line with R^2 equal to 0.9150 with a slope of 1.07. Unfortunately this line is well above the normal line, giving an intercept of 2.38.

Log P DB failed to accurately model the nitro compounds. The very poor slope and intercept values are overshadowed by the randomness evident in a correlation coefficient of 0.5390.

The final program is Molecular Modeling Pro (MMP). Unlike the previous software, this package contains four different methods for the estimation of Log P. Unfortunately, the performance of each of these methods is insufficient to warrant individual scrutiny, as was done for the previous programs. It is adequate to say that all of the selected structural groups proved troublesome for these systems. However, as is evident from the dates of the references, the calculation modules are based on methods that have been developed further since being incorporated into this program. It should also be noted that in all four cases the effect of hydrophilicity is remarkably pronounced.

DOES CONVERGENCE OF THE ESTIMATIONS INCREASE ACCURACY?

It is always better to have several programs produce one answer than to have several programs produce several answers. However, increased precision does not infer increased accuracy. This holds true in the case of Log P estimations. For example, consider the treatment of pyridazine by Clog P, SciLog P, and Kowwin. SciLog P and Kowwin both yield values close to zero, 0.01 and -0.06. Clog P gives a more hydrophilic estimation of -0.73. The experimental value of -0.72 demonstrates that convergence of the majority of a battery of programs does not necessarily give the correct answer. The one program that appears to deviate just may be correct.

To demonstrate the generality of this principle, the data calculated from these three programs was compiled in Appendix C and the variance among them was calculated for all 150 data points. Data points were then eliminated to give subsequent data sets in which the maximum variance was incrementally reduced from 0.11 to 0.04, thereby introducing a more stringent definition of convergence from data set to data set. These limits were chosen to be reflective of the standard deviation ranges that some programs claim to achieve. The graphs of the slope and intercept data against the corresponding variance produce more or less flat lines, indicating there is no correlation between convergence and accuracy. If more accurate estimations are attained when divergent calculations are eliminated then the slope data should approach unity as the variance is reduced, likewise the intercept data should approach zero.

Although the convergence of a bank of programs on a single value cannot be used as a basis for increased confidence, information can still be gained from the comparison of the calculated values. Divergence of the calculated values is a red flag. When estimations do not agree, it is evident that the type of compound submitted is a problem area for at least one of the programs.

CONCLUSIONS

It has been demonstrated that although a program may perform well for a large data set, erroneous values may be obtained for a certain structure type. The examples which best illustrate this are compounds containing the amino acid and nitro structural groups.

Therefore, it is suggested that the best technique to employ when an accurate estimation of Log P is desired is to put together a small database of compounds as has been done for the nine structural classes here.

To obtain meaningful results with this methodology, the compounds must be structurally similar to the compound being estimated or at least share the functional groups that are suspected to be problematic. This may be accomplished by utilizing the substructure search functions of Clog P or Log P DB 4.5. However, the list of compounds thus yielded will most likely need to be reduced through the elimination of unsuitable examples. For instance, when building a database of hydrazine compounds phenylhydrazine is an acceptable data point, but p-nitrophenylhydrazine is not. The latter is eliminated due to its content of a nitro group, which is itself a problematic functional group and would skew the results. Therefore it is ideal that the compounds comprising the database consist only of well-behaved functional groups and the functional group of interest. They must also have accurately determined experimental Log P values associated with them.

Once an appropriate database of compounds is compiled, it is then submitted to the program or battery of programs for estimation. The calculated data set(s) are plotted against the experimental values and the square of the correlation coefficient and the slope and intercept of the best-fit line are derived. These parameters are then used to determine if the calculated value is dependable, in the case of a single program, or which program produces the most dependable value for this type of compound, in the case of a bank of programs.

REFERENCES

1. Hansch, C.; Leo, A.; Hoekman, D. Exporing QSAR. American Chemical Society, Washington, D.C., 1995.
2. Hansch, C.; Leo, A. Substituent Constants for Correlation Analysis in Chemistry and Biology. Wiley and Sons, 1979.
3. Viswanadhan, V.; Ghose, A.; Revankar, G.; Robins, R. *J. Chem. Inf. Comput. Sci.* 1989, 29, 163.
4. Moschner, K.; Cece, A. "Development of a General QSAR for Predicting Octanol- Water Partition Coefficients and its Application to Surfactants," Environmental Toxicology and Risk Assessment - Third Volume, ASTM STP 1218, J.S. Hughes, G.R. Biddinger, and E. Mones, Eds., American Society for Testing and Materials, Philadelphia, 1995.
5. Devillers, J. Neural Networks in QSAR and Drug Design Academic Press, 1996.
6. Meylan, W.; Howard, P. *J. Pharm. Sci.* 1995, 84, 272.
7. Spessard, G. *J. Chem. Inf. Comput. Sci.* 1998, 38, 6.

APPENDIX A: COMPOUND NAMES, EXPERIMENTAL LOG P VALUES, CALCULATED LOG P VALUES, AND GRAPHS OF CALCULATED VALUES AGAINST EXPERIMENTAL VALUES

TABLE A1. Experimental and Estimated Log P Values with Graphical Parameters

Compound	Exp LogP	Scilog P	Clog P	Log P DB 4.5	Kowwin	MMP	Ghose	CNDO	INDO
Arginine	-4.20	-3.85	-3.52	-1.78	-4.00	-5.24	-0.81	-1.96	-2.35
Asparagine	-3.82	-3.07	-3.54	-1.51	-4.99	-3.96	-1.72	-2.70	-3.07
Glutamic acid	-3.69	-3.08	-2.69	-1.43	-3.83	-2.43	-0.78	-2.06	-2.56
Glutamine	-3.64	-3.60	-3.37	-1.60	-4.49	-3.43	-1.47	-2.27	-2.63
Histidine	-3.32	-2.46	-3.73	-1.18	-3.22	-2.02	-1.11	-3.23	-3.50
Glycine	-3.21	-2.34	-3.21	-1.03	-3.41	-1.81	-0.82	-3.20	-3.57
Serine	-3.07	-2.76	-2.74	-1.58	-3.46	-2.86	-1.19	-0.63	-1.13
Lysine	-3.05	-3.10	-3.42	-1.04	-2.99	-1.80	-0.64	-1.73	-2.12
Threonine	-2.94	-2.48	-2.43	-1.23	-3.04	-2.49	-0.78	-0.19	-0.69
Alanine	-2.85	-2.03	-3.12	-0.68	-2.99	-1.50	-0.41	-2.71	-3.08
Aminourea	-2.75	-2.40	-2.75	-1.76	-2.64	-3.01	-1.51	-2.16	-2.41
Proline	-2.54	-1.35	-2.41	-0.57	-2.15	-1.13	-0.02	-2.12	-2.50
Cysteine	-2.49	-1.84	-2.35	0.24	-3.05	-1.82	-0.61	-2.59	
Tyrosine	-2.26	-1.23	-2.22	0.38	-1.76	-0.60	0.99	-1.06	-1.43
Hydrazine	-2.07	-1.52	-1.68	-1.19	-1.47	-1.66	-1.41	-1.33	-1.34
Formylhydrazine	-2.05	-1.78	-2.05	-1.11	-1.97	-2.33	-1.58	-1.35	-1.61
Ethylenediamine	-2.04	-1.74	-2.02	-1.24	-1.62	-1.41	-1.40	-0.55	-0.68
2-Hydroxyethyl hydrazine	-1.98	-1.89	-1.59	-1.49	-1.98	-2.26	-1.61	-1.03	-1.41
Methionine	-1.87	-1.27	-1.73	0.37	-2.41	-1.36	-0.42	-1.97	
Isoleucine	-1.70	-0.97	-1.76	0.73	-1.59	-0.05	0.86	-1.33	-1.72
Guanazole	-1.61	-1.23	-1.33	-1.51	-1.48	-2.31	-0.41	-2.86	-2.95
Acetylhydrazine	-1.58	-2.14	-2.54	-1.37	-2.24	-2.70	-0.82	-2.02	-2.43
Leucine	-1.52	-0.87	-1.67	0.73	-1.59	-0.05	0.79	-1.33	-1.72
Piperazine	-1.50	-1.18	-1.48	-1.17	-0.80	-0.81	-0.76	0.42	0.23
Methylurea	-1.40	-1.21	-1.30	-1.23	-1.09	-1.30	-0.55	-1.27	-1.55
Phenylalanine	-1.38	-0.78	-1.56	1.11	-1.28	0.06	1.28	-0.80	-1.07
Dimethylsulfoxide	-1.35	-0.19	-1.38	-1.35	-1.22	-1.38	-1.43	0.04	
Dimethylsulfone	-1.34	-0.52	-1.50	-1.19	-1.11	-1.42	-1.24	-0.06	
Cyanoguanidine	-1.15	-1.57	-1.68	-1.64	-1.34	-4.99	-0.08	-0.57	-0.84
3-Amino-1-propanol	-1.12	-1.23	-0.96	-1.07	-1.12	-1.59	-1.00	-0.26	-0.65
Methylhydrazine	-1.05	-1.11	-1.05	-0.86	-1.00	-1.51	-1.16	-0.59	-0.63
Tryptophane	-1.05	-0.39	-1.57	1.04	-1.22	-0.33	0.93	-1.18	-1.47
Propionic Acid Hydrazide	-1.00	-1.25	-1.10	-0.71	-1.03	-2.26	-0.78	-0.50	-0.75
1-Amino-2-propanol	-0.96	-1.20	-0.99	-0.96	-1.19	-1.02	-0.64	-0.29	-0.68
2-Methylaminoethanol	-0.94	-0.93	-0.99	-0.97	-1.15	-1.12	-0.65	-0.18	-0.58
Nitroguanidine	-0.89	-1.31	-1.68	-1.31	-1.72	-3.67	-0.24	-0.65	-0.84
1-Methyl-2-cyanoguanidine	-0.77	-0.65	-0.77	-0.82	-0.79	-3.73	0.16	-0.12	-0.41
Methanol	-0.77	-0.50	-0.76	-0.72	-0.63	-0.76	-0.27	-0.89	-1.27
Ethylurea	-0.74	-0.78	-0.78	-0.70	-0.60	-0.78	-0.21	-0.79	-1.08
Pyridazine	-0.72	0.01	-0.73	-0.77	-0.06	-0.71	0.96	-0.14	-0.10
18-Crown-6	-0.68	-0.68	-1.18	-0.81	-1.42	0.93	-0.99	2.62	1.49
Bis(2-hydroxyethyl)sulfide	-0.63	-0.82	-0.81	-0.60	-0.62	-0.81	-0.43	0.20	
Butyric Acid Hydrazide	-0.62	-0.79	-0.57	-0.18	-0.54	-1.73	-0.39	-0.03	-0.28
Tetrazole	-0.60	-1.02	-0.96	-1.16	-1.58	-2.46	-0.99	-3.00	-3.05

Diethylsulfone	-0.59	0.13	-0.44	-0.13	-0.12	-0.42	0.13	0.83	
Methylamine	-0.57	-0.66	-0.66	-0.66	-0.64	-0.66	-0.61	-0.74	-0.88
1,3-Dimethylurea	-0.49	-0.96	-0.55	-1.02	-0.62	-0.55	-0.14	-0.77	-1.06
1-Methyl-4-nitro-1H- imidazole	-0.44	-0.07	-0.14	0.22	0.42	0.23	0.65	-0.71	-0.81
Trioxane	-0.43	-0.74	-0.48	-1.17	-0.56	-0.56	0.61	-0.69	-1.49
Nitroethanol	-0.42	-0.26	-1.01	-0.42	-1.01	-0.76	-0.40	0.10	-0.20
Pyrimidine	-0.40	-0.11	-0.31	-0.33	-0.06	-0.07	0.90	-0.29	-0.32
1,4-Dimethylpiperazine	-0.40	0.17	-0.74	-0.61	-0.38	0.98	-0.04	1.49	1.25
Dimethylamine	-0.38	-0.09	-0.52	-0.43	-0.17	-0.52	-0.20	-0.01	-0.18
Nitromethane	-0.35	0.07	-0.28	-0.20	-0.04	0.28	-0.08	-0.10	-0.27
1,3-Diaminobenzene	-0.33	0.11	-0.31	-0.82	-0.39	-0.31	0.48	0.64	0.69
1,4-Diaminobenzene	-0.30	-0.10	-0.31	-0.85	-0.39	-0.31	0.48	0.80	0.84
1,2,3-triazole	-0.29	-0.70	-0.60	0.23	-0.01	-1.16	-0.09	-1.89	-1.91
p-Dioxane	-0.27	-0.20	-0.39	-0.27	-0.32	0.31	-0.33	0.49	-0.05
Thioactetamide	-0.26	-0.14	-0.25	-0.46	-0.83	-0.39	-0.13	-0.49	
Pyrazine	-0.23	-0.08	-0.31	-0.28	-0.06	-0.07	1.28	-0.04	-0.04
1-Methyl-2,4-dinitro-1H-imidazole	-0.17	0.42	-0.12	-0.06	0.24	-0.02	1.51	0.08	-0.03
1-Methyl-2-nitro-1H-imidazole	-0.17	0.15	0.10	0.21	0.42	-0.02	1.33	-0.73	-0.83
Ethylamine	-0.13	-0.17	-0.13	-0.13	-0.15	-0.14	-0.27	-0.17	-0.31
4-Nitroimidazole	-0.11	0.00	-0.09	0.35	-0.12	-0.01	0.40	-1.23	-1.33
Imidazole	-0.08	-0.31	-0.67	-0.08	0.06	0.06	0.22	-0.83	-0.91
2-Aminopyrazine	-0.07	-0.42	-0.37	0.15	-0.12	-0.57	0.73	-0.97	-0.89
1-Carbamyl-2-phenylhydrazine	0.13	0.21	0.13	0.33	-0.38	-0.71	0.42	-0.12	-0.26
2-Methyl-4-nitroimidazole	0.13	0.36	0.18	0.06	0.42	0.64	0.69	-0.66	-0.80
2-Nitroimidazole	0.15	0.12	0.15	0.09	-0.12	-0.20	1.08	-1.13	-1.27
2-Methyl-5-nitroimidazole	0.15	0.36	0.42	0.06	0.42	0.61	0.40	-0.50	-0.67
1,2-Diaminobenzene	0.15	0.18	-0.31	0.05	0.16	0.09	0.48	0.79	0.83
1,3,5-Trihydroxybenzene	0.16	0.35	0.14	0.06	0.55	0.14	1.19	0.59	0.18
1-Methyl-5-nitro-1H-imidazole	0.16	0.05	0.10	0.22	0.42	0.26	0.36	-0.55	-0.65
Nitroethane	0.18	0.42	0.25	0.34	0.45	0.81	0.60	0.38	0.20
Tetramethylurea	0.19	-0.05	-0.11	-0.98	-0.20	-0.43	0.58	0.31	0.23
2,4-Dinitro-1H-Imidazole	0.20	0.52	-0.05	-0.18	-0.31	-0.27	1.27	-0.32	-0.44
Propylamine	0.26	0.25	0.39	0.40	0.34	0.39	0.20	0.36	0.21
Pyrazole	0.26	-0.22	-1.09	0.32	0.06	0.07	0.71	-1.01	-1.15
4-Nitropyridine	0.33	0.75	0.57	0.70	0.62	0.41	1.37	0.56	0.49
1,1,3,3-Tetramethylguanidine	0.41	0.17	-0.29	-1.22	-0.28	-2.49	0.75	0.69	0.38
1-Methyl-4,5-dinitro-1H-imidazole	0.41	0.51	-0.12	0.81	0.24	0.19	0.55	0.15	0.03
Thiazole	0.44	0.53	0.49	0.44	0.99	0.79	0.74	-0.18	
Pyrrolidine	0.46	0.30	0.38	0.37	0.70	0.00	0.12	0.67	0.47
Tetrahydrofuran	0.46	0.67	0.53	0.33	0.94	0.33	0.34	0.54	0.13
2-Nitropyridine	0.48	0.86	0.57	0.52	0.62	0.78	1.60	0.45	0.40
Phthalazine	0.57	1.62	0.44	0.46	1.12	0.26	2.22	1.06	1.11
1,4-Dihydroxybenzene	0.59	0.61	0.81	0.64	1.03	0.81	1.48	0.92	0.63
4-Nitropyrazole	0.59	0.07	0.46	0.59	-0.12	0.01	0.79	-1.25	-1.41
3-Nitropyridine	0.60	0.77	0.57	0.56	0.62	0.63	1.49	0.48	0.43
Pyridine	0.65	0.65	0.64	0.73	0.80	0.67	1.42	0.68	0.63

4,5-Dinitro-1H-Imidazole	0.65	0.59	-0.05	0.94	-0.31	-0.15	0.30	-0.24	-0.37
2-Nitrofuran	0.66	1.23	1.06	1.23	1.18	1.09	0.88	0.73	0.58
Pyrrole	0.75	0.44	0.29	0.75	0.88	0.78	0.80	0.50	0.58
2-Methyl-4,5-dinitro-1H-imidazole	0.77	0.82	0.22	0.65	0.24	0.50	0.58	0.24	0.08
1,3-Dihydroxybenzene	0.80	0.96	0.81	0.76	1.03	0.81	1.48	0.83	0.52
2-Nitropropane	0.80	0.73	0.55	0.68	0.87	1.11	1.36	0.84	0.66
Cyclohexanone	0.81	0.77	0.86	0.76	1.13	0.59	1.53	1.73	1.56
Cyclohexanone oxime	0.84	0.76	1.19	0.78	0.91	1.36	2.61	1.97	1.74
Piperidine	0.84	0.72	0.93	0.93	1.19	0.55	0.52	1.19	1.01
1-Nitropropane	0.87	1.26	0.77	0.87	0.95	1.33	1.00	0.85	0.68
1,2-Dihydroxybenzene	0.88	0.92	0.88	0.88	1.03	0.81	1.48	0.97	0.71
Aniline	0.90	0.94	0.91	0.94	1.08	0.92	1.26	1.06	1.12
Methylpyrrolidine	0.92	0.77	0.84	0.76	0.91	0.84	0.48	1.29	1.05
Methylisothiocyanate	0.94	1.20	1.17	0.70	1.30	2.23	0.52	1.09	
Tetrahydropyran	0.95	1.14	0.95	0.89	1.43	0.89	0.73	1.08	0.65
Hydrazinophthalazine	1.00	1.23	0.97	0.59	0.77	-0.58	0.92	0.16	0.21
3-Methyl-4-nitropyrazole	1.02	0.44	0.65	1.05	0.42	0.66	0.50	-0.74	-0.93
Benzylamine	1.09	0.99	1.09	1.09	1.07	1.09	1.16	1.49	1.45
Benzyl alcohol	1.10	1.14	1.10	1.03	1.08	1.10	1.51	1.28	0.98
Diethylsulfate	1.14	0.16	1.09	1.19	1.14	1.15	0.74	0.02	
2-Methyl-2-nitropropane	1.17	0.86	0.95	1.03	1.33	1.75	1.78	1.41	1.22
2-Nitrobutane	1.20	1.16	1.08	1.21	1.36	1.64	1.75	1.32	1.14
Cyclohexanol	1.23	1.21	1.27	1.34	1.64	1.05	1.32	1.25	0.84
Phenylhydrazine	1.25	0.69	1.26	1.25	0.79	0.84	0.52	0.78	0.90
Isonitrobutane	1.40	1.27	1.17	1.21	1.36	1.73	1.33	1.33	1.15
4-Nitrophenylhydrazine	1.41	1.22	1.66	1.45	1.19	0.58	0.47	0.45	0.32
Phenol	1.46	1.24	1.47	1.48	1.51	1.48	1.76	1.05	0.86
1-Nitrobutane	1.47	1.26	1.30	1.40	1.44	1.86	1.39	0.93	0.73
Ethylisothiocyanate	1.47	1.76	1.69	1.23	1.79	2.76	0.86	1.50	
Cyclohexylamine	1.49	1.23	1.37	1.40	1.63	1.37	0.97	1.52	1.37
Ethylmethylsulfide	1.54	1.78	1.37	1.42	1.41	1.37	0.80	1.21	
3-Nitrothiophene	1.55	1.55	1.62	1.63	1.63	1.53	1.40	1.42	
2-Nitrothiophene	1.58	1.53	1.62	1.85	1.63	1.53	1.22	1.44	
Cyclopropane	1.72	1.57	1.68	1.69	1.70	1.68	1.19	1.42	1.44
Dimethyldisulfide	1.77	1.60	1.74	1.77	1.87	2.15	1.07	0.66	
Thiophene	1.81	1.84	1.79	1.90	1.81	1.79	1.33	0.76	
Benzoic Acid	1.87	1.73	1.88	1.89	1.87	1.89	1.75	1.00	0.67
2-Phenylpyrrolidine	1.89	2.03	1.93	1.94	2.33	2.28	1.97	2.60	2.50
Diethylsulfide	1.95	2.35	1.90	1.95	1.90	1.90	1.14	1.58	
Benzene	2.13	1.95	2.14	2.22	1.99	2.14	2.05	2.42	2.45
1-Butanethiol	2.28	2.61	2.23	2.31	2.25	2.23	1.53	1.96	
Toluene	2.73	2.66	2.64	2.68	2.54	2.79	2.51	2.91	2.92
1,2-Diphenylhydrazine	2.94	2.52	2.97	2.94	3.06	2.95	2.45	3.09	3.27
Cyclopentane	3.00	2.87	2.79	2.82	2.68	2.80	1.98	2.47	2.35
o-Xylene	3.12	3.22	3.09	3.14	3.09	3.44	2.98	3.34	3.36
Cyclohexyl methacrylate	3.13	3.06	3.14	3.41	3.54	3.18	2.48	2.82	2.35
p-Xylene	3.15	3.11	3.14	3.14	3.09	3.44	2.98	3.36	3.35
m-Xylene	3.20	3.17	3.14	3.14	3.09	3.44	2.98	3.34	3.37
Naphthalene	3.30	3.32	3.32	3.45	3.17	3.12	3.05	3.62	3.70
1,3,5-Trimethylbenzene	3.42	3.59	3.64	3.60	3.63	4.09	3.45	3.71	3.80
Cyclohexane	3.44	3.40	3.35	3.39	3.18	3.35	2.38	2.95	2.83

Methylcyclohexane	3.61	3.76	3.87	3.88	3.59	3.87	2.71	3.44	3.30
1,2,4-Trimethylbenzene	3.63	3.74	3.59	3.60	3.63	4.09	3.45	3.85	3.83
1,2,3-Trimethylbenzene	3.66	3.78	3.54	3.60	3.63	4.09	3.45	3.88	3.84
4-Vinylcyclohexene	3.93	2.91	3.43	3.44	3.73	3.31	2.63	3.40	3.50
Cycloheptane	4.00	3.92	3.91	3.95	3.67	3.91	2.77	3.43	3.30
Cyclooctane	4.45	4.31	4.47	4.51	4.16	4.47	3.17	3.90	3.76
Anthracene	4.45	4.67	4.49	4.68	4.35	4.09	4.05	4.68	4.87
Phenanthrene	4.46	4.61	4.49	4.68	4.35	4.09	4.05	4.73	4.86
Pyrene	4.88	5.06	4.95	5.17	4.93	4.35	4.37	5.23	5.45
	Rsqrd	0.96	0.98	0.83	0.97	0.83	0.77	0.80	0.84
	slope	0.92	0.98	0.76	1.01	0.95	0.61	0.82	0.89
	int	0.13	-0.05	0.35	0.00	0.02	0.56	0.18	0.00
	Rsqrd X > 0	0.9471	0.9650	0.9598	0.9434	0.8589	0.7585	0.8530	0.8689
	Rsqrd X < 0	0.8539	0.9180	0.1970	0.9000	0.4102	0.2860	0.3840	0.4546

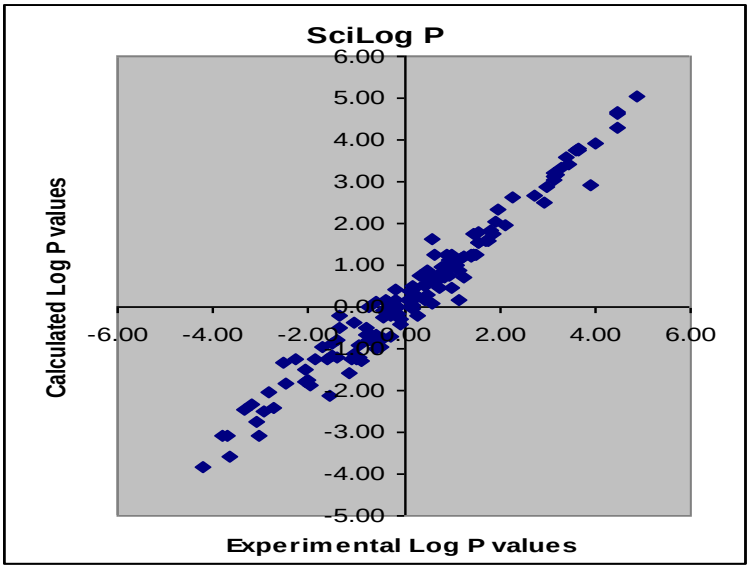

FIGURE A1. SciLog P Estimated Values vs. Experimental Values

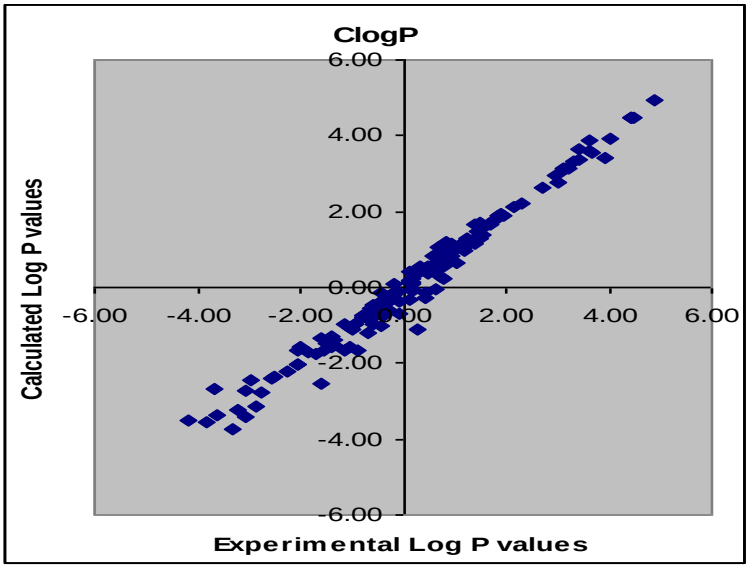

FIGURE A2. Clog P Estimated Values vs. Experimental Values

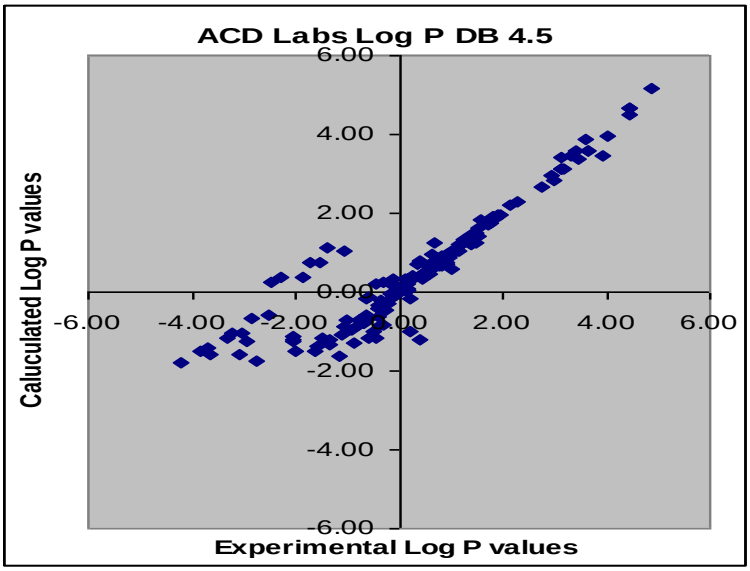

FIGURE A3. Log P DB 4.5 Estimated Values vs. Experimental Values

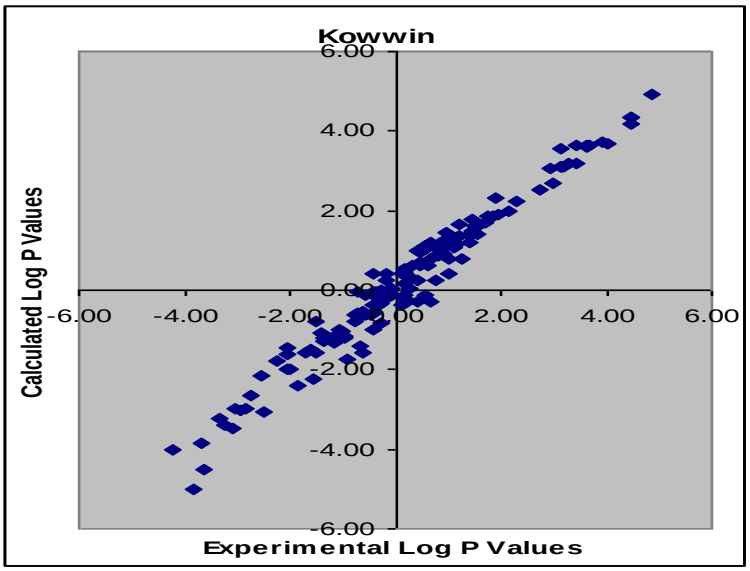

FIGURE A4. Kowwin Estimated Values vs. Experimental Values

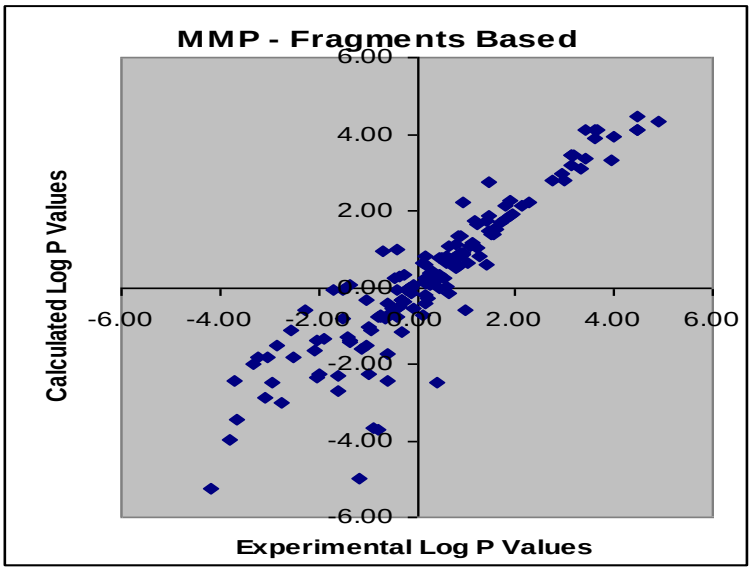

FIGURE A5. MMP Fragments Method Estimated Values vs. Experimental Values

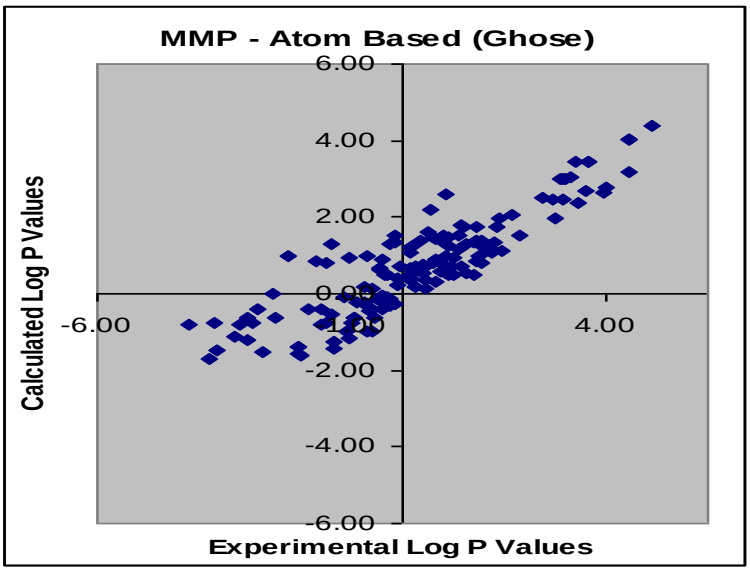

FIGURE A6. MMP Atom Method Estimated Values vs. Experimental Values

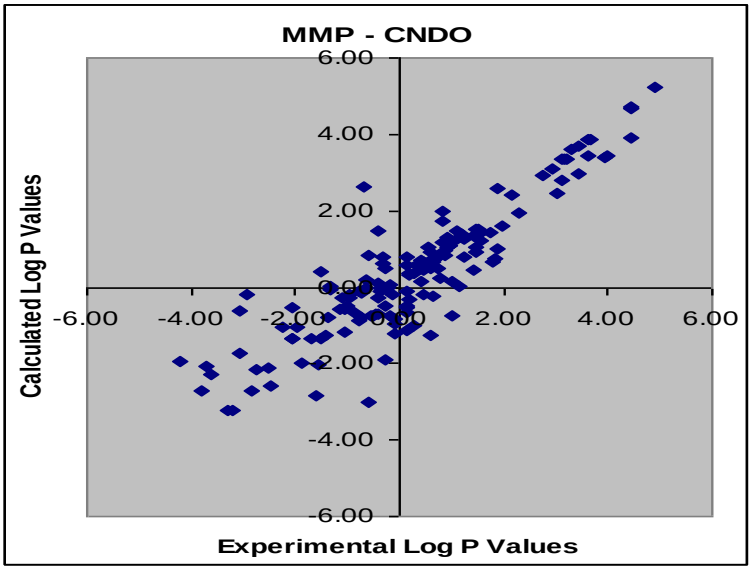

FIGURE A7. CNDO Estimated Values vs. Experimental Values

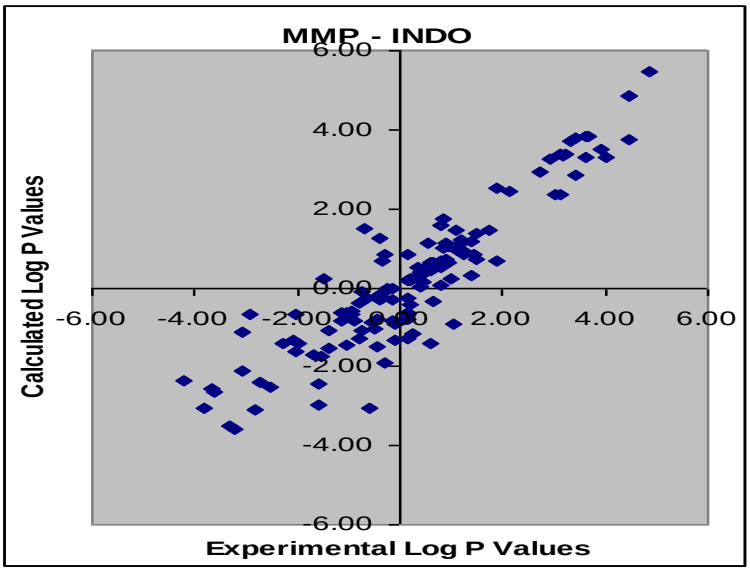

FIGURE A8. INDO Estimated Values vs. Experimental Values

APPENDIX B: DATA DIVIDED INTO STRUCTURAL TYPES

TABLE B1. Aromatics and Substituted Benzenes

Aromatic/Substituted Benzenes	Exp log P	Scilog P	Clog P	Log P DB 4.5	Kowwin	MMP	Ghose	CNDO	INDO
1,3-Diaminobenzene	-0.33	0.11	-0.31	-0.82	-0.39	-0.31	0.48	0.64	0.69
1,4-Diaminobenzene	-0.30	-0.10	-0.31	-0.85	-0.39	-0.31	0.48	0.80	0.84
1,2-Diaminobenzene	0.15	0.18	-0.31	0.05	0.16	0.09	0.48	0.79	0.83
1,3,5-Trihydroxybenzene	0.16	0.35	0.14	0.06	0.55	0.14	1.19	0.59	0.18
1,4-Dihydroxybenzene	0.59	0.61	0.81	0.64	1.03	0.81	1.48	0.92	0.63
1,3-Dihydroxybenzene	0.80	0.96	0.81	0.76	1.03	0.81	1.48	0.83	0.52
1,2-Dihydroxybenzene	0.88	0.92	0.88	0.88	1.03	0.81	1.48	0.97	0.71
Aniline	0.90	0.94	0.91	0.94	1.08	0.92	1.26	1.06	1.12
Benzylamine	1.09	0.99	1.09	1.09	1.07	1.09	1.16	1.49	1.45
Benzyl alcohol	1.10	1.14	1.10	1.03	1.08	1.10	1.51	1.28	0.98
Phenol	1.46	1.24	1.47	1.48	1.51	1.48	1.76	1.05	0.86
Benzoic Acid	1.87	1.73	1.88	1.89	1.87	1.89	1.75	1.00	0.67
Benzene	2.13	1.95	2.14	2.22	1.99	2.14	2.05	2.42	2.45
Toluene	2.73	2.66	2.64	2.68	2.54	2.79	2.51	2.91	2.92
o-Xylene	3.12	3.22	3.09	3.14	3.09	3.44	2.98	3.34	3.36
p-Xylene	3.15	3.11	3.14	3.14	3.09	3.44	2.98	3.36	3.35
m-Xylene	3.20	3.17	3.14	3.14	3.09	3.44	2.98	3.34	3.37
Naphthalene	3.30	3.32	3.32	3.45	3.17	3.12	3.05	3.62	3.70
1,3,5-Trimethylbenzene	3.42	3.59	3.64	3.60	3.63	4.09	3.45	3.71	3.80
1,2,4-Trimethylbenzene	3.63	3.74	3.59	3.60	3.63	4.09	3.45	3.85	3.83
1,2,3-Trimethylbenzene	3.66	3.78	3.54	3.60	3.63	4.09	3.45	3.88	3.84
Anthracene	4.45	4.67	4.49	4.68	4.35	4.09	4.05	4.68	4.87
Phenanthrene	4.46	4.61	4.49	4.68	4.35	4.09	4.05	4.73	4.86
Pyrene	4.88	5.06	4.95	5.17	4.93	4.35	4.37	5.23	5.45
	Rsqrd	0.9916	0.9945	0.9940	0.9906	0.9722	0.9748	0.9447	0.9182
	slope	1.00	1.01	1.08	0.96	0.99	0.74	0.94	1.00
	int	0.07	-0.04	-0.19	0.11	0.06	0.69	0.38	0.20

TABLE B2. Hydrazine Derivatives

Hydrazines	Exp log P	Scilog P	Clog P	Log P DB 4.5	Kowwin	MMP	Ghose	CNDO	INDO
Hydrazine	-2.07	-1.52	-1.68	-1.19	-1.47	-1.66	-1.41	-1.33	-1.34
Formylhydrazine	-2.05	-1.78	-2.05	-1.11	-1.97	-2.33	-1.58	-1.35	-1.61
2-Hydroxyethyl hydrazine	-1.98	-1.89	-1.59	-1.49	-1.98	-2.26	-1.61	-1.03	-1.41
Acetylhydrazine	-1.58	-2.14	-2.54	-1.37	-2.24	-2.70	-0.82	-2.02	-2.43
Methylhydrazine	-1.05	-1.11	-1.05	-0.86	-1.00	-1.51	-1.16	-0.59	-0.63
Propionic Acid Hydrazide	-1.00	-1.25	-1.10	-0.71	-1.03	-2.26	-0.78	-0.50	-0.75
Butyric Acid Hydrazide	-0.62	-0.79	-0.57	-0.18	-0.54	-1.73	-0.39	-0.03	-0.28
1-Carbamyl-2-phenylhydrazine	0.13	0.21	0.13	0.33	-0.38	-0.71	0.42	-0.12	-0.26
Phthalazine	0.57	1.62	0.44	0.46	1.12	0.26	2.22	1.06	1.11
Hydrazinophthalazine	1.00	1.23	0.97	0.59	0.77	-0.58	0.92	0.16	0.21
Phenylhydrazine	1.25	0.69	1.26	1.25	0.79	0.84	0.52	0.78	0.90
4-Nitrophenylhydrazine	1.41	1.22	1.66	1.45	1.19	0.58	0.47	0.45	0.32
1,2-Diphenylhydrazine	2.94	2.52	2.97	2.94	3.06	2.95	2.45	3.09	3.27
	Rsqrd	0.9183	0.9593	0.9656	0.9449	0.8837	0.8190	0.8463	0.8468
	slope	0.94	1.01	0.82	0.96	0.97	0.78	0.76	0.85
	int	-0.01	-0.01	0.20	-0.06	-0.63	0.12	0.07	-0.02

TABLE B3. Substituted Guanidines and Ureas

Guanadine Type Compounds	Exp log P	Scilog P	Clog P	Log P DB 4.5	Kowwin	MMP	Ghose	CNDO	INDO
Aminourea	-2.75	-2.40	-2.75	-1.76	-2.64	-3.01	-1.51	-2.16	-2.41
Methylurea	-1.40	-1.21	-1.30	-1.23	-1.09	-1.30	-0.55	-1.27	-1.55
Cyanoguanidine	-1.15	-1.57	-1.68	-1.64	-1.34	-4.99	-0.08	-0.57	-0.84
Nitroguanidine	-0.89	-1.31	-1.68	-1.31	-1.72	-3.67	-0.24	-0.65	-0.84
1-Methyl-2-cyanoguanidine	-0.77	-0.65	-0.77	-0.82	-0.79	-3.73	0.16	-0.12	-0.41
Ethylurea	-0.74	-0.78	-0.78	-0.70	-0.60	-0.78	-0.21	-0.79	-1.08
1,3-Dimethylurea	-0.49	-0.96	-0.55	-1.02	-0.62	-0.55	-0.14	-0.77	-1.06
Tetramethylurea	0.19	-0.05	-0.11	-0.98	-0.20	-0.43	0.58	0.31	0.23
1,1,3,3-Tetramethylguanidine	0.41	0.17	-0.29	-1.22	-0.28	-2.49	0.75	0.69	0.38
	Rsqrd	0.9074	0.8685	0.3634	0.8318	0.1121	0.9352	0.8856	0.8822
	slope	0.80	0.84	0.23	0.77	0.60	0.69	0.86	0.87
	int	-0.30	-0.39	-0.99	-0.38	-1.82	0.44	0.13	-0.10

TABLE B4. Aromatic Heterocycles Containing Nitrogen

Aromatic Heterocycles	Exp log P	Scilog P	Clog P	Log P DB 4.5	Kowwin	MMP	Ghose	CNDO	INDO
Guanazole	-1.61	-1.23	-1.33	-1.51	-1.48	-2.31	-0.41	-2.86	-2.95
Pyridazine	-0.72	0.01	-0.73	-0.77	-0.06	-0.71	0.96	-0.14	-0.10
Tetrazole	-0.60	-1.02	-0.96	-1.16	-1.58	-2.46	-0.99	-3.00	-3.05
Pyrimidine	-0.40	-0.11	-0.31	-0.33	-0.06	-0.07	0.90	-0.29	-0.32
1,2,3-Triazole	-0.29	-0.70	-0.60	0.23	-0.01	-1.16	-0.09	-1.89	-1.91
Pyrazine	-0.23	-0.08	-0.31	-0.28	-0.06	-0.07	1.28	-0.04	-0.04
Imidazole	-0.08	-0.31	-0.67	-0.08	0.06	0.06	0.22	-0.83	-0.91
2-Aminopyrazine	-0.07	-0.42	-0.37	0.15	-0.12	-0.57	0.73	-0.97	-0.89
Pyrazole	0.26	-0.22	-1.09	0.32	0.06	0.07	0.71	-1.01	-1.15
Pyridine	0.65	0.65	0.64	0.73	0.80	0.67	1.42	0.68	0.63
Pyrrole	0.75	0.44	0.29	0.75	0.88	0.78	0.80	0.50	0.58
	Rsqrd	0.6362	0.5727	0.8799	0.7153	0.6985	0.3189	0.4989	0.4924
	slope	0.68	0.67	1.04	0.99	1.37	0.64	1.33	1.36
	int	-0.13	-0.35	0.04	0.07	-0.23	0.64	-0.61	-0.63

TABLE B5. Amino Acids

	Exp			Log P					
Amino Acids	log P	Scilog P	Clog P	DB 4.5	Kowwin	MMP	Ghose	CNDO	INDO
Arginine	-4.20	-3.85	-3.52	-1.78	-4.00	-5.24	-0.81	-1.96	-2.35
Asparagine	-3.82	-3.07	-3.54	-1.51	-4.99	-3.96	-1.72	-2.70	-3.07
Glutamic acid	-3.69	-3.08	-2.69	-1.43	-3.83	-2.43	-0.78	-2.06	-2.56
Glutamine	-3.64	-3.60	-3.37	-1.60	-4.49	-3.43	-1.47	-2.27	-2.63
Histidine	-3.32	-2.46	-3.73	-1.18	-3.22	-2.02	-1.11	-3.23	-3.50
Glycine	-3.21	-2.34	-3.21	-1.03	-3.41	-1.81	-0.82	-3.20	-3.57
Serine	-3.07	-2.76	-2.74	-1.58	-3.46	-2.86	-1.19	-0.63	-1.13
Lysine	-3.05	-3.10	-3.42	-1.04	-2.99	-1.80	-0.64	-1.73	-2.12
Threonine	-2.94	-2.48	-2.43	-1.23	-3.04	-2.49	-0.78	-0.19	-0.69
Alanine	-2.85	-2.03	-3.12	-0.68	-2.99	-1.50	-0.41	-2.71	-3.08
Proline	-2.54	-1.35	-2.41	-0.57	-2.15	-1.13	-0.02	-2.12	-2.50
Cysteine	-2.49	-1.84	-2.35	0.24	-3.05	-1.82	-0.61	-2.59	xxxx
Tyrosine	-2.26	-1.23	-2.22	0.38	-1.76	-0.60	0.99	-1.06	-1.43
Methionine	-1.87	-1.27	-1.73	0.37	-2.41	-1.36	-0.42	-1.97	xxxx
Isoleucine	-1.70	-0.97	-1.76	0.73	-1.59	-0.05	0.86	-1.33	-1.72
Leucine	-1.52	-0.87	-1.67	0.73	-1.59	-0.05	0.79	-1.33	-1.72
Phenylalanine	-1.38	-0.78	-1.56	1.11	-1.28	0.06	1.28	-0.80	-1.07
Tryptophane	-1.05	-0.39	-1.57	1.04	-1.22	-0.33	0.93	-1.18	-1.47
	Rsqrd	0.9169	0.8219	0.9149	0.8657	0.8122	0.7422	0.2061	0.3021
	slope	1.09	0.75	1.07	1.12	1.42	0.87	0.44	0.52
	int	0.87	-0.59	2.38	0.17	2.01	2.01	-0.66	-0.74

TABLE B6. Nitro Compounds

Nitro Compounds	Exp log P	Scilog P	Clog P	Log P DB 4.5	Kowwin	MMP	Ghose	CNDO	INDO
1-Methyl-4-nitroimidazole-5-amine	-1.00	0.02	0.13	1.20	-0.49	-0.29	-0.19	-0.95	-1.11
1-Methyl-4-nitro-1H- imidazole	-0.44	-0.07	-0.14	0.22	0.42	0.23	0.65	-0.71	-0.81
Nitroethanol	-0.42	-0.26	-1.01	-0.42	-1.01	-0.76	-0.40	0.10	-0.20
Nitromethane	-0.35	0.07	-0.28	-0.20	-0.04	0.28	-0.08	-0.10	-0.27
1-Methyl-2,4-dinitro-1H-imidazole	-0.17	0.42	-0.12	-0.06	0.24	-0.02	1.51	0.08	-0.03
1-Methyl-2-nitro-1H-imidazole	-0.17	0.15	0.10	0.21	0.42	-0.02	1.33	-0.73	-0.83
4-Nitroimidazole	-0.11	0.00	-0.09	0.35	-0.12	-0.01	0.40	-1.23	-1.33
2-Methyl-4-nitroimidazole	0.13	0.36	0.18	0.06	0.42	0.64	0.69	-0.66	-0.80
2-Nitroimidazole	0.15	0.12	0.15	0.09	-0.12	-0.20	1.08	-1.13	-1.27
2-Methyl-5-nitroimidazole	0.15	0.36	0.42	0.06	0.42	0.61	0.40	-0.50	-0.67
1-Methyl-5-nitro-1H-imidazole	0.16	0.05	0.10	0.22	0.42	0.26	0.36	-0.55	-0.65
Nitroethane	0.18	0.42	0.25	0.34	0.45	0.81	0.60	0.38	0.20
2,4-Dinitro-1H-Imidazole	0.20	0.52	-0.05	-0.18	-0.31	-0.27	1.27	-0.32	-0.44
4-Nitropyridine	0.33	0.75	0.57	0.70	0.62	0.41	1.37	0.56	0.49
1-Methyl-4,5-dinitro-1H-imidazole	0.41	0.51	-0.12	0.81	0.24	0.19	0.55	0.15	0.03
2-Nitropyridine	0.48	0.86	0.57	0.52	0.62	0.78	1.60	0.45	0.40
4-Nitropyrazole	0.59	0.07	0.46	0.59	-0.12	0.01	0.79	-1.25	-1.41
3-Nitropyridine	0.60	0.77	0.57	0.56	0.62	0.63	1.49	0.48	0.43
4,5-Dinitro-1H-Imidazole	0.65	0.59	-0.05	0.94	-0.31	-0.15	0.30	-0.24	-0.37
2-Nitrofuran	0.66	1.23	1.06	1.23	1.18	1.09	0.88	0.73	0.58
2-Methyl-4,5-dinitro-1H-imidazole	0.77	0.82	0.22	0.65	0.24	0.50	0.58	0.24	0.08
2-Nitropropane	0.80	0.73	0.55	0.68	0.87	1.11	1.36	0.84	0.66
1-Nitropropane	0.87	1.26	0.77	0.87	0.95	1.33	1.00	0.85	0.68
3-Methyl-4-nitropyrazole	1.02	0.44	0.65	1.05	0.42	0.66	0.50	-0.74	-0.93
2-Methyl-2-nitropropane	1.17	0.86	0.95	1.03	1.33	1.75	1.78	1.41	1.22
2-Nitrobutane	1.20	1.16	1.08	1.21	1.36	1.64	1.75	1.32	1.14
Isonitrobutane	1.40	1.27	1.17	1.21	1.36	1.73	1.33	1.33	1.15
1-Nitrobutane	1.47	1.26	1.30	1.40	1.44	1.86	1.39	0.93	0.73
3-Nitrothiophene	1.55	1.55	1.62	1.63	1.63	1.53	1.40	1.42	
2-Nitrothiophene	1.58	1.53	1.62	1.85	1.63	1.53	1.22	1.44	
	Rsqrd	0.7454	0.7134	0.5390	0.6329	0.6803	0.3554	0.5182	0.4243
	slope	0.67	0.77	0.65	0.81	0.91	0.54	0.94	0.86
	int	0.28	0.06	0.33	0.12	0.18	0.65	-0.32	-0.45

TABLE B7. Sulfur Compounds

Sulfur Compounds	Exp log P	Scilog P	Clog P	Log P DB 4.5	Kowwin	MMP	Ghose	CNDO	INDO
Dimethylsulfoxide	-1.35	-0.19	-1.38	-1.35	-1.22	-1.38	-1.43	0.04	xxxx
Dimethylsulfone	-1.34	-0.52	-1.50	-1.19	-1.11	-1.42	-1.24	-0.06	xxxx
Bis(2-hydroxyethyl)sulfide	-0.63	-0.82	-0.81	-0.60	-0.62	-0.81	-0.43	0.20	xxxx
Diethylsulfone	-0.59	0.13	-0.44	-0.13	-0.12	-0.42	0.13	0.83	xxxx
Thioactetamide	-0.26	-0.14	-0.25	-0.46	-0.83	-0.39	-0.13	-0.49	xxxx
Thiazole	0.44	0.53	0.49	0.44	0.99	0.79	0.74	-0.18	xxxx
Methylisothiocyanate	0.94	1.20	1.17	0.70	1.30	2.23	0.52	1.09	xxxx
Diethylsulfate	1.14	0.16	1.09	1.19	1.14	1.15	0.74	0.02	xxxx
Ethylisothiocyanate	1.47	1.76	1.69	1.23	1.79	2.76	0.86	1.50	xxxx
Ethylmethylsulfide	1.54	1.78	1.37	1.42	1.41	1.37	0.80	1.21	xxxx
Dimethyldisulfide	1.77	1.60	1.74	1.77	1.87	2.15	1.07	0.66	xxxx
Thiophene	1.81	1.84	1.79	1.90	1.81	1.79	1.33	0.76	xxxx
Diethylsulfide	1.95	2.35	1.90	1.95	1.90	1.90	1.14	1.58	xxxx
1-Butanethiol	2.28	2.61	2.23	2.31	2.25	2.23	1.53	1.96	xxxx
	Rsqrd	0.8382	0.9898	0.9806	0.9501	0.8907	0.9098	0.4997	xxxx
	slope	0.82	1.01	0.96	0.97	1.09	0.69	0.42	xxxx
	int	0.34	-0.01	0.03	0.12	0.14	-0.05	0.38	xxxx

TABLE B8. Non-Aromatic Heterocycles

Non-Aromatic Heterocycles	Exp log P	Scilog P	Clog P	Log P DB 4.5	Kowwin	MMP	Ghose	CNDO	INDO
Piperazine	-1.50	-1.18	-1.48	-1.17	-0.80	-0.81	-0.76	0.42	0.23
18-Crown-6	-0.68	-0.68	-1.18	-0.81	-1.42	0.93	-0.99	2.62	1.49
Trioxane	-0.43	-0.74	-0.48	-1.17	-0.56	-0.56	0.61	-0.69	-1.49
1,4-Dimethylpiperazine	-0.40	0.17	-0.74	-0.61	-0.38	0.98	-0.04	1.49	1.25
p-Dioxane	-0.27	-0.20	-0.39	-0.27	-0.32	0.31	-0.33	0.49	-0.05
Pyrrolidine	0.46	0.30	0.38	0.37	0.70	0.00	0.12	0.67	0.47
Tetrahydrofuran	0.46	0.67	0.53	0.33	0.94	0.33	0.34	0.54	0.13
Piperidine	0.84	0.72	0.93	0.93	1.19	0.55	0.52	1.19	1.01
Methylpyrrolidine	0.92	0.77	0.84	0.76	0.91	0.84	0.48	1.29	1.05
Tetrahydropyran	0.95	1.14	0.95	0.89	1.43	0.89	0.73	1.08	0.65
2-Phenylpyrrolidine	1.89	2.03	1.93	1.94	2.33	2.28	1.97	2.60	2.50
	Rsqrd	0.9302	0.9760	0.9315	0.8846	0.5120	0.7476	0.1405	0.2384
	slope	0.94	1.08	1.01	1.12	0.63	0.73	0.38	0.52
	int	0.08	-0.10	-0.10	0.14	0.39	0.09	0.99	0.55

TABLE B9. Cyclic Alkanes

Cyclic Alkanes	Exp log P	Scilog P	Clog P	Log P DB 4.5	Kowwin	MMP	Ghose	CNDO	INDO
Cyclohexanone	0.81	0.77	0.86	0.76	1.13	0.59	1.53	1.73	1.56
Cyclohexanone oxime	0.84	0.76	1.19	0.78	0.91	1.36	2.61	1.97	1.74
Cyclohexanol	1.23	1.21	1.27	1.34	1.64	1.05	1.32	1.25	0.84
Cyclohexylamine	1.49	1.23	1.37	1.40	1.63	1.37	0.97	1.52	1.37
Cyclopropane	1.72	1.57	1.68	1.69	1.70	1.68	1.19	1.42	1.44
Cyclopentane	3.00	2.87	2.79	2.82	2.68	2.80	1.98	2.47	2.35
Cyclohexyl methacrylate	3.13	3.06	3.14	3.41	3.54	3.18	2.48	2.82	2.35
Cyclohexane	3.44	3.40	3.35	3.39	3.18	3.35	2.38	2.95	2.83
Methylcyclohexane	3.61	3.76	3.87	3.88	3.59	3.87	2.71	3.44	3.30
4-Vinylcyclohexene	3.93	2.91	3.43	3.44	3.73	3.31	2.63	3.40	3.50
Cycloheptane	4.00	3.92	3.91	3.95	3.67	3.91	2.77	3.43	3.30
Cyclooctane	4.45	4.31	4.47	4.51	4.16	4.47	3.17	3.90	3.76
	Rsqrd	0.9530	0.9752	0.9770	0.9686	0.9572	0.5497	0.8723	0.8686
	slope	0.95	0.94	0.99	0.86	0.96	0.40	0.65	0.68
	int	-0.03	0.14	-0.01	0.37	0.05	1.08	0.82	0.57

APPENDIX C: DATA FOR CONVERGENCE OF ESTIMATIONS

TABLE C1. Data for Accuracy as a Function of Convergence

Variance	ScilogP slope	ScilogP int	ClogP slope	ClogP int	Kowwin slope	Kowwin int
0.11	0.9822	0.0392	1.0007	-0.035	0.974	0.0654
0.1	0.9816	0.0438	0.9989	-0.023	0.9726	0.0736
0.09	0.9838	0.047	0.9995	-0.0178	0.976	0.0701
0.08	0.9893	0.046	0.9999	-0.019	0.9764	0.0721
0.07	0.9922	0.0346	0.9994	-0.0162	0.9773	0.0655
0.06	0.9832	0.0364	0.9984	-0.0083	0.9739	0.0729
0.05	0.9943	0.0296	1.0058	-0.011	0.9689	0.078
0.04	0.9843	0.0484	0.9960	-0.0041	0.9605	0.0724

FIGURE C1. Slope as a Function of Convergence

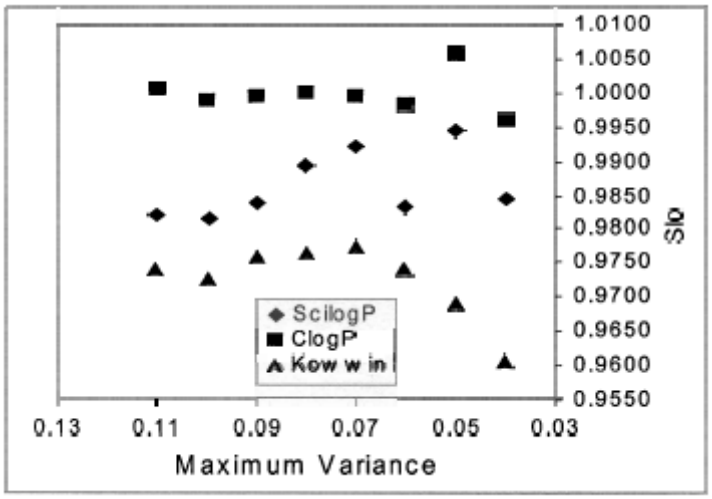

FIGURE C2. Intercept as a Function of Convergence