AD-A256 856

(12)

CHEMICAL
RESEARCH,
DEVELOPMENT &
ENGINEERING
CENTER

CRDEC-TR-361

PREDICTING POLYMER PROPERTIES BY COMPUTATIONAL METHODS 2: A COMPARISON OF SEMI-EMPIRICAL METHODS

A. Birenzvige
L.M. Sturdivan
G.R. Famini

RESEARCH DIRECTORATE

P.N. Krishnan
R.E. Morris

COPPIN STATE COLLEGE
Baltimore, MD 21216

DTIC
ELECTE
NOV 02 1992
S E D

September 1992

Approved for public release; distribution is unlimited.

92-28588

U.S. ARMY
ARMAMENT
MUNITIONS
CHEMICAL COMMAND

Aberdeen Proving Ground, Maryland 21010-5423

Disclaimer

The findings in this report are not to be construed as an official Department of the Army position unless so designated by other authorizing documents.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188), Washington, DC 20503.

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	1992 September	Final, 89 Jun - 89 Sep

4. TITLE AND SUBTITLE
Predicting Polymer Properties by Computational Methods 2: A Comparison of Semi-Empirical Methods

5. FUNDING NUMBERS
PR-1L162622A553

6. AUTHOR(S)
Birenzvige, A., Sturdivan, L.M., Famini G.R. (CRDEC), Krishnan, P.N., and Morris, R.E. (Coppin State College)

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
CDR, CRDEC, ATTN: SMCCR-RSL, APG, MD 21010-5423
Coppin State College, Baltimore, MD 21216

8. PERFORMING ORGANIZATION REPORT NUMBER
CRDEC-TR-361

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

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

12b. DISTRIBUTION CODE

13. ABSTRACT (Maximum 200 words)

The physical properties of various molecules calculated by MNDO, AM1, and PM3 methods are compared to experimental values using rigorous statistical analysis. The results show the best methods for calculating physical properties will depend on the property and that no one method is adequate to calculate all the physical properties. Furthermore, the results show the value of employing statistical methods in the area of computational chemistry.

14. SUBJECT TERMS
Vinyl chloride
Statistical analysis
Semi-empirical methods
Computational chemistry

15. NUMBER OF PAGES
38

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	UL

NSN 7540-01-280-5500

Standard Form 298 (Rev 2-89)
Prescribed by ANSI Std Z39-18
298-102

Blank

PREFACE

The work described in this report was authorized under Project No. 1L162622A553, CB Defense/General Investigation. This work was started in June 1989 and completed in September 1989.

The use of trade names or manufacturers' names in this report does not constitute an official endorsement of any commercial products. This report may not be cited for purposes of advertisement.

Reproduction of this document in whole or in part is prohibited except with permission of the Commander, U.S. Army Chemical Research, Development and Engineering Center, ATTN: SMCCR-SPS-T, Aberdeen Proving Ground, MD 21010-5423. However, the Defense Technical Information Center and the National Technical Information Service are authorized to reproduce the document for U.S. Government purposes.

This report has been approved for release to the public.

Accesion For
NTIS CRA&I ☒
DTIC TAB ☐
Unannounced ☐
Justification
By
Distribution /
Availability Codes
Dist | Avail and / or Special
A-1

DTIC COPY INSPECTED 1

Blank

CONTENTS

		Page
1.	INTRODUCTION	9
2.	WHY SEMI-EMPIRICAL METHODS?	9
3.	AB INITIO VERSUS SEMI-EMPIRICAL METHODS	10
4.	METHOD OF COMPUTATION	11
4.1	Computational Chemistry	11
4.2	Statistical Methods	11
5.	RESULTS AND DISCUSSION	14
5.1	Heat of Formation	14
5.2	Dipole Moment	24
5.3	Polarizability	28
5.4	Ionization Potential	33
6.	CONCLUSIONS	36
	LITERATURE CITED	37

LIST OF FIGURES

1. Comparison of Normal and Logistic Distribution 13

2. Comparison Between Experimental and Calculated Heat of Formation 16

3. Calculation Errors in Heat of Formation 16

4. Test for Normal Distribution - Heat of Formation, MNDO, All Data 17

5. Test for Normal Distribution - Heat of Formation, AM1, All Data . 18

6. Test for Normal Distribution - Heat of Formation, PM3, All Data . 18

7. Test for Normal Distribution - Heat of Formation, MNDO, Low Region 19

8. Test for Normal Distribution - Heat of Formation, AM1, Low Region 19

9. Test for Normal Distribution - Heat of Formation, PM3, Low Region 20

10. Relative Differences Versus Experimental Values - Heat of Formation 21

11. Test for Normal Distribution - Relative Difference, Heat of Formation, MNDO, All Data 21

12. Test for Normal Distribution - Relative Difference, Heat of Formation, AM1, All Data 22

13. Test for Normal Distribution - Relative Difference, Heat of Formation, PM3, All Data 22

14. Test for Normal Distribution - Relative Difference, Heat of Formation, MNDO, High Region 23

15. Test for Normal Distribution - Relative Difference, Heat of Formation, AM1, High Region 23

16. Test for Normal Distribution - Relative Difference, Heat of Formation, PM3, High Region 24

17. Calculated Versus Experimental Dipole Moment 26

18. Dipole Moment Calculation Errors 26

19. Test for Normal Distribution - Calculation Errors, MNDO 27

21. Test for Normal Distribution - Dipole Moment, PM3 28

22. Calculated Versus Experimental Polarizability 30

23. Polarizability Calculation Errors 30

24. Test for Normal Distribution, Polarizability, MNDO 31

25. Test for Normal Distribution, Polarizability, AM1 31

26. Test for Normal Distribution, Polarizability, PM3 32

27. Calculated Versus Experimental Ionization Potential 32

28. Ionization Potential Calculation Errors 34

29. Test for Normal Distribution - Ionization Potential, MNDO 34

30. Test for Normal Distribution - Ionization Potential, AM1 35

31. Test for Normal Distribution - Ionization Potential, PM3 35

LIST OF TABLES

1. Comparison of Ab Initio Versus Semi-Empirical Calculations for Vinyl Chloride 10

2. Comparison of Ab Initio (321-G) Versus Semi-Empirical Calculations for Ethyl Chloride 10

3. Comparison of Structural Parameters of Vinyl Chloride: Semi-Empirical Methods Versus Ab Initio Method 11

4. Comparison of Structural Parameters of Ethyl Chloride: MNDO, PM3, AM1 Versus Methods Versus Ab Initio Method 11

5. Molecules Investigated 14

6. Comparison Between Experimental and Calculated Heat of Formation 15

7. Bias and SD of Calculated Heat of Formation 20

8. Bias and SD of the Relative Calculated Heat of Formation 24

9. Comparison Between Experimental and Calculated Values of Dipole Moment 25

10. Bias and SD of the Calculated Dipole Moment 28

11. Comparison of Experimental and Calculated Polarizability 29

12. Bias and SD of the Calculated Polarizability 33

13. Comparison of Experimental and Calculated Ionization Potential .. 33

14. Bias and SD of the Calculated Ionization Potential 36

15. Recommended Computational Methods for Heat of Formation, Ionization Potential, Dipole Moment, and Polarizability 36

PREDICTING POLYMER PROPERTIES BY COMPUTATIONAL METHODS 2: A COMPARISON OF SEMI-EMPIRICAL METHODS

1. INTRODUCTION

In a previous report[1] the Modified Neglect Differential Overlap (MNDO) method of Dewar and Thiel[2] was used to calculate the physical properties of vinyl chloride and its homologs. The results of these calculations compared favorably to the available experimental data. Improvements to the MNDO model, namely the AM1 and PM3 methods,[3,4] have recently stimulated debate about the better method of calculating properties such as heat of formation, dipole moment, and polarizability.[5,6] In past studies that used these methods,[7,8,9] no statistical analysis has been done to establish the accuracy of each method for predicting each property. In this report, the heat of formation, dipole moment, and polarizability, calculated by using MNDO, AM1, and PM3, are compared with experimental results. The objective of this work was to statistically examine the limitation and accuracy of each method in predicting the above mentioned properties.

2. WHY SEMI-EMPIRICAL METHODS?

The physical properties of a compound can be theoretically calculated either by a semi-empirical method or by a more elaborate ab initio technique.[10,11] These approaches are based on molecular orbital theory. The ab initio model seeks the best solution of the Schrodinger wave equation, using the Hartree-Fock orbital estimating techniques, where the orbitals are estimated by linear combinations of hydrogen-like (Slater) atomic orbitals method. An approximate solution to the Schrodinger equation can be achieved only by using a basis set with a large number of orbitals. However, these high level ab initio calculations require too much computing time, even for moderate-sized (10 to 20 atoms) molecules, to be practical. The simpler-ab initio treatment uses minimum basis set too inaccurately to be chemically useful for most polyatomic molecules. Thus, to achieve the required accuracy, the higher level calculations must be used.

The second molecular orbital approach, the semi-empirical model, is based on a completely different philosophy.[12,13] Semi-empirical methods are used to avoid solving time-consuming integrals involved in the solution of the Schrodinger equation. The most popular semi-empirical methods (MNDO, AM1, PM3) use experimental data to parameterize these integrals. This is done in such a way that the solutions of the Schrodinger equation are adjusted to fit experimental data for each atom. These parameterized solutions for the atoms are used to effect a solution to the Schrodinger equation for any molecule containing the atoms for which solutions exist. Because these parameterized solutions for the atoms obviate a number of integrals, the semi-empirical methods yield reasonable and reliable estimates of the solution to the Schrodinger equation with much less computational time and can be used to find solutions for larger molecules. Dewar and co-workers have shown that for heats of formation, the accuracy of the semi-empirical method is comparable to that of quite larger set ab initio calculations.[14]

3. AB INITIO VERSUS SEMI-EMPIRICAL METHODS

To quantify the claim that semi-empirical methods yield comparable results for much less time, we have used semi-empirical methods to compute the dipole moment, polarizability, and structural data for vinyl chloride (CH_2 = CHCl) and ethyl chloride (C_2H_5Cl). The dipole moment and polarizability calculated using three semi-empirical methods and one ab initio method (321-G) is compared to the experimental values in Table 1 and Table 2. Also, the CPU time required for each method of calculation is included. The semi-empirical methods yield results that are equivalent or better than the (321-G) ab initio method and requires much less computer time. As the number of atoms in the molecule increases, the required computational time increases as n^4, where n is the number of electrons.[10] A simple calculation for the dimer (two monomer units of CH_2 = CHCl) gives a necessary CPU time of about 80 hr for this (321-G) calculation.

Table 1. Comparison of Ab Initio Versus Semi-Empirical Calculations for Vinyl Chloride

Method	Dipole Moment (Debye)	Polarizability (A3)	CPU Time
Experiment	1.45[16]	6.41[16]	
MNDO	1.71	5.84	1 min or less
AM1	1.19	3.34	1 min or less
PM3	0.93	3.41	1 min or less
321-G	1.93		4-1/2 hr

Table 2. Comparison of Ab Initio (321-G) Versus Semi-Empirical Calculations for Ethyl Chloride

Method	Dipole Moment (Debye)	Polarizability (A3)	CPU Time
Experiment	2.05[16]	6.40[16]	~2 min
MNDO	2.09	6.26	~2 min
AM1	1.69	3.32	~2 min
PM3	1.55	3.30	~2 min
321-G	2.50	4.33	~5 hr

A comparison of available experimental structural data of CH_2 = CHCl[15] and C_2H_5Cl with calculated values is given in Table 3 and 4. Again, the values from the semi-empirical calculations are comparable to the values from ab initio calculations. From a consideration of the computing time alone, the semi-empirical method is the method of choice. In some cases, the semi-empirical method results approximates the experimental data better than the ab initio calculation.

Table 3. Comparison of Structural Parameters of Vinyl Chloride: Semi-Empirical Methods Versus Ab Initio Method

Method	r (C = C) (A°)	r (C - Cl) (A°)	r (C - H) (A°)	< (C-C-Cl) (Degree)
Experiment	1.36[15]	1.73[15]	1.08[15]	121.1[15]
MNDO	1.34	1.75	1.09	122.9
AM1	1.33	1.70	1.09	123.4
PM3	1.33	1.69	1.10	121.2
321-G	1.31	1.75	1.07	122.8

Table 4. Comparison of Structural Parameters of Ethyl Chloride: MNDO, PM3, AM1 Versus Methods Versus Ab Initio Method

Method	r (C - Cl) (A°)	< (C-C-Cl) (Degree)
Experiment	1.76[17]	111.5°
MNDO	1.81	112.1°
AM1	1.78	109.0°
PM3	1.81	112.1°
321-G	1.82	110.9°

4. METHOD OF COMPUTATION

4.1 Computational Chemistry.

The calculations were carried out on a Microvax (Digital Equipment Corporation, Stanford, CA) using the MOPAC package of computer programs.[18] The three semi-empirical methods (MNDO, AM1, and PM3) were contained in the MOPAC. All structures were fully optimized using standard MMADS techniques[19] developed by the Chemometric and Biometric Modeling Branch, U.S. Army Chemical Research, Development and Engineering Center.

4.2 Statistical Methods.

As previously stated, the purpose of this study was to determine the ability of each of the three semi-empirical methods to calculate the heat of formation, dipole moment, and polarizability and ionization potential. To enable us to determine the accuracy of the calculation (i.e., the standard deviation (SD) of the error, in a statistically meaningful way), we need to show that the calculation errors are symmetrically distributed. This is done by showing that the data follow the normal distribution function. One way to show that a data set is normally distributed is to order it in an ascending order and then plot the data on a normal distribution graph paper. For example, we can take the weight of nine people (n=9) selected at random, sort them in ascending order, and then scale the linear axis so that all weights will fit. Finally, plot the cumulative fraction on the probability axis versus the weight on the linear axis, letting the denominator of the fraction

equal n+1 (for symmetry). Thus, the lightest weight would be plotted versus 0.1 (1/n+1), the next lightest versus 0.2, and so on until the heaviest would be plotted against 0.9 (n/n+1). If the weights were normally distributed, the resulting nine points would fall on a straight line. Alternatively, the normal score, that is the expected value of the normal order statistic of an ordered sample of size n can be calculated. In the statistical package MINITAB(tm), the normal score is abbreviated N-score. Plotting the N-score against normally distributed data results in the points falling about a straight line.

Calculating the N-score requires numerical solution of integral equations. The N-score calculation is available in some statistical packages on minicomputers but is not available in commonly used software packages for microcomputers. To enable us to perform the analysis on a desktop microcomputer, we need to find a distribution function that will closely resemble the normal distribution but will be easier to compute. The logistic distribution is such a distribution. Its straight line transform, which we will call the L-score, is obtainable in closed form and is simple to calculate.

Figure 1 shows a comparison of L-score and N-score. The figure was produced as follows. The N-score of an order set of numbers from 1 to 1000 using minitab(tm) was calculated on the VAX minicomputer. The data was then downloaded into a spread sheet on a desktop personal computer. The L-score was calculated according to $\text{L-score}_{(i)} = \ln \frac{i}{n - i + i}$, where i is the order of the item in the list and n is the total number of items. The dashed line is the plot of L-score versus N-score. The solid line is a least-squares-fitted straight line through the data. As can be seen, the two lines coincide except at the ends where the slightly heavier tails of the logistic distribution causes a slight curvature away from the straight line. The correlation coefficient (R-squared) of the two measures is .994.

The calculated value was subtracted from the experimental value for each molecule. The result or its transformation was plotted against L-score, and the correlation coefficient of the least squares regression line was determined. The plot was examined visually to determine any outliers and whether the fit would improve in a limited region. The average and SD of the transformed data was calculated in the symmetry region as was the R-Square for the least square regression line. The following procedures should be followed to choose the best method for calculating the physical property in the region at a 95% confidence level.

- For each of the three different methods, plot the difference between the calculated and experimental value of the property estimated versus its L-score. (Alternatively, plot the difference of the transformed data versus L-score.)

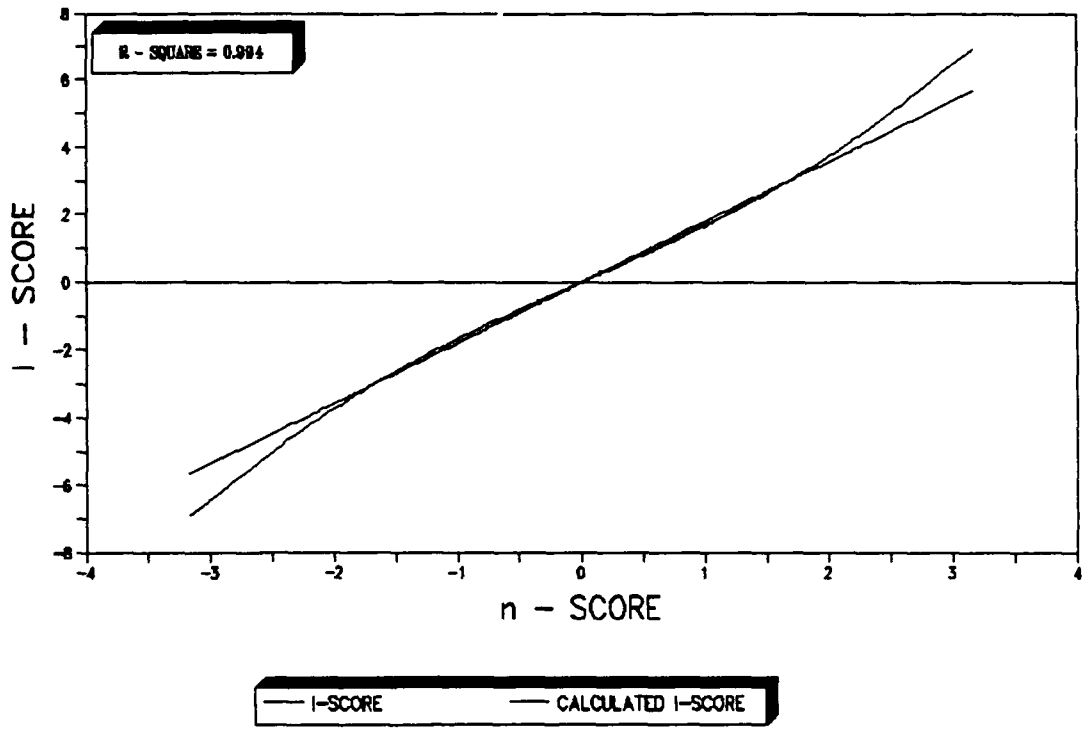

Figure 1. Comparison of Normal and Logistic Distribution

- Determine a region (magnitude of calculated and experimental values) where the L-score plot forms a straight line, indicating a symmetrical "normal" distribution of errors. Transformed and untransformed data may have to be used for different regions (e.g., the data can be normally distributed in one region and lognormal in another).

- Calculate the R-square between the difference and the L-score for the appropriate region.

- Calculate the mean and SD of the approximation error in the appropriate region.

- A method that has an R-square of .94 or larger is well approximated by the normal distribution. Among those methods that satisfy this criterion, choose the method that has the smallest SD. If the smallest SD is >2.28 times the size of the SD of another method whose R-square is <.94, choose the method with the smaller SD regardless of the value of R-square.

- Approximately 95% of the time, the true value will be in the range <calculated value - bias ±20>.

5. RESULTS AND DISCUSSION

The focus of our computations has been on predicting properties of polymers. Therefore, we have chosen a number of polymer forming molecules that contain single and double bonded carbon-atoms (ethyl and vinyl compounds). The compounds represent a variety of substituted ethylene and vanillic molecules, dictated by the availability of experimental data. Table 5 lists the compounds studied and their chemical formulas.

Table 5. Molecules Investigated

Molecule Number	Molecule Name	Molecular Formula
1	Ethylene	C_2H_4
2	Vinyl Chloride	C_2H_3Cl
3	Vinyl Bromide	C_2H_3Br
4	Ethyl Chloride	C_2H_5Cl
5	Ethyl Bromide	C_2H_5Br
6	Vinylidene Chloride	$C_2H_2Cl_2$
7	Vinyl Acetate	$C_4H_6O_2$
8	Ethyl Acetate	$C_4H_8O_2$
9	Ethyl Alcohol	C_2H_5OH
10	Vinyl Cyanide	C_3H_3N
11	Ethyl Cyanide	C_3H_5N
12	Tetrafluro Ethylene	C_2F_4

5.1 Heat of Formation.

The experimental and calculated heat of formation for the molecules investigated are listed in Table 6. The computed results for each molecule and each method of calculation are listed together with the experimental values. Figure 2 depicts the same information graphically. Note, if the calculated and experimental results were identical, all the points in Figure 2 would fall on the diagonal line. Figure 3 shows the deviation between the experimental and calculated values versus the experimental values of the heat of formation. Close examination of Figure 3 indicates that as the absolute value of the heat of formation becomes larger, the absolute value of the deviation increases, as can be expected. However, determining which of the three methods yield more reliable results from either Figures 2 or 3 is impossible.

Table 6. Comparison Between Experimental and Calculated Heat of Formation

Molecule	Heat of Formation Kcal/mole			
	Experiment[16]	Calculation MNDO	Calculation AM1	Calculation PM3
Ethylene	14.5	15.4	16.5	16.6
Vinyl Chloride	8.1	4.9	5.8	9.7
Vinyl Bromide	18.7	15.8	18.0	23.8
Ethyl Chloride	-26.8	-28.8	-26.2	-22.1
Ethyl Bromide	-15.4	-17.0	-13.1	-11.3
Vinylidene Chloride	0.6	0.0	0.0	3.1
Vinyl Acetate	-74.5	-68.9	-67.7	-68.2
Ethyl Acetate	-106.0	-98.9	-101.9	-98.8
Ethyl Alcohol	-56.2	-63.0	-62.7	-56.9
Vinyl Cyanide	43.0	43.9	45.0	50.2
Ethyl Cyanide	12.3	13.8	13.2	18.6
Tetrafluro Ethylene	-154.7	-175.7	-175.1	-168.2

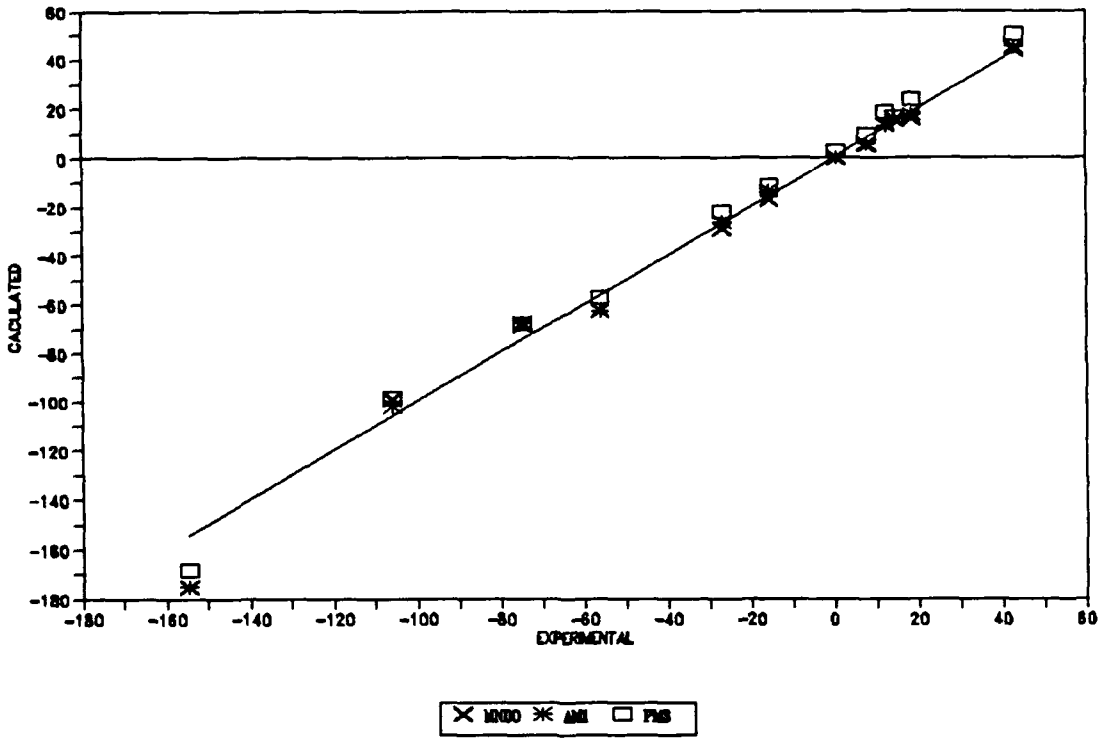

Figure 2. Comparison Between Experimental and Calculated Heat of Formation

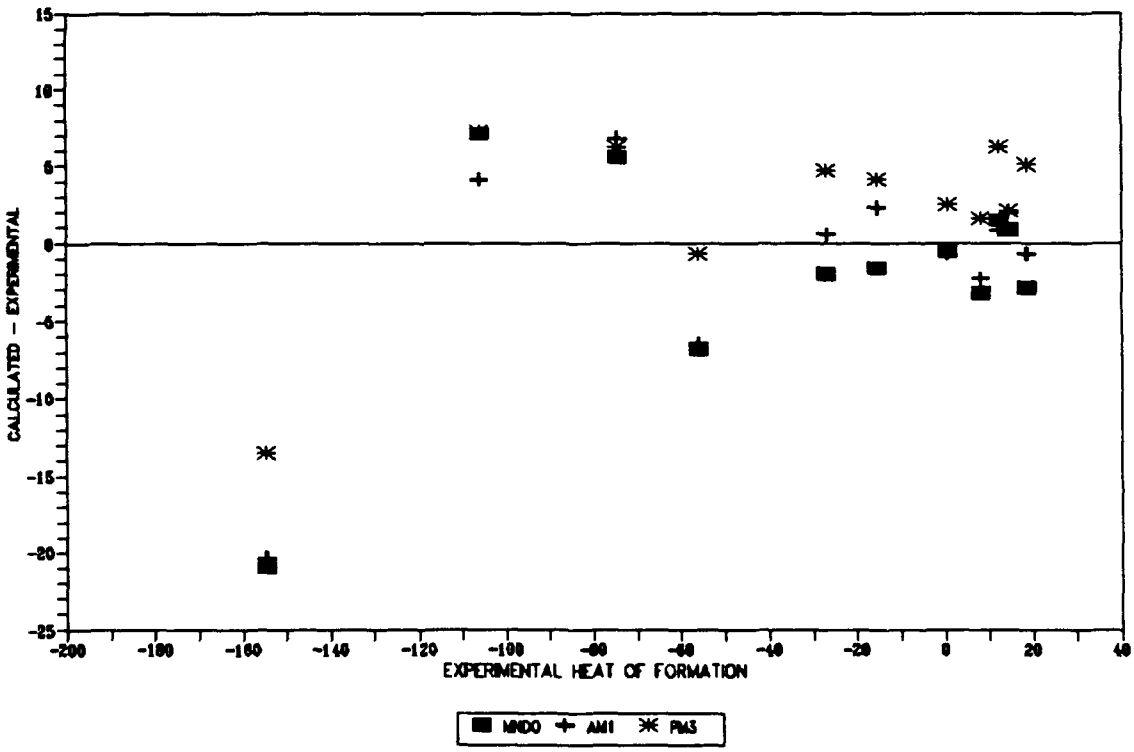

Figure 3. Calculation Errors in Heat of Formation

The calculation errors are plotted against L-score in Figures 4, 5, and 6 for MNDO, AM1, and PM3, respectively. As can be seen, the correlation coefficients for the regression line for all three methods are fairly low, which indicates that the distribution of the deviation between the calculated and experimental data are not symmetrical. Closer examination shows that in all cases there is one outlier (indicated by an arrow). The outlier in all cases is tetrafluro ethylene with a large (absclute) heat of formation (154.7 Kcal/mole). If this value is removed from the analysis, the correlation coefficient improves significantly for all three methods. The correlation coefficient improvements can be seen in Figures 7, 8, and 9 indicating that in all three methods the differences between the calculated and experimental heat of formation is symmetrically distributed for molecule that have heat of formation of about 100 kcal/mole (absolute) or lower. The mean and the SD σ and the R-Square are given in Table 7. The "mean error" indicates a systematic error in the calculation and the SD indicates the random distribution of the errors or precision. Thus, the range of heat of formation for molecules for which experimental value is not available can be estimated by (calculated velue-bias $\pm 2\sigma$) with 95% confidence.

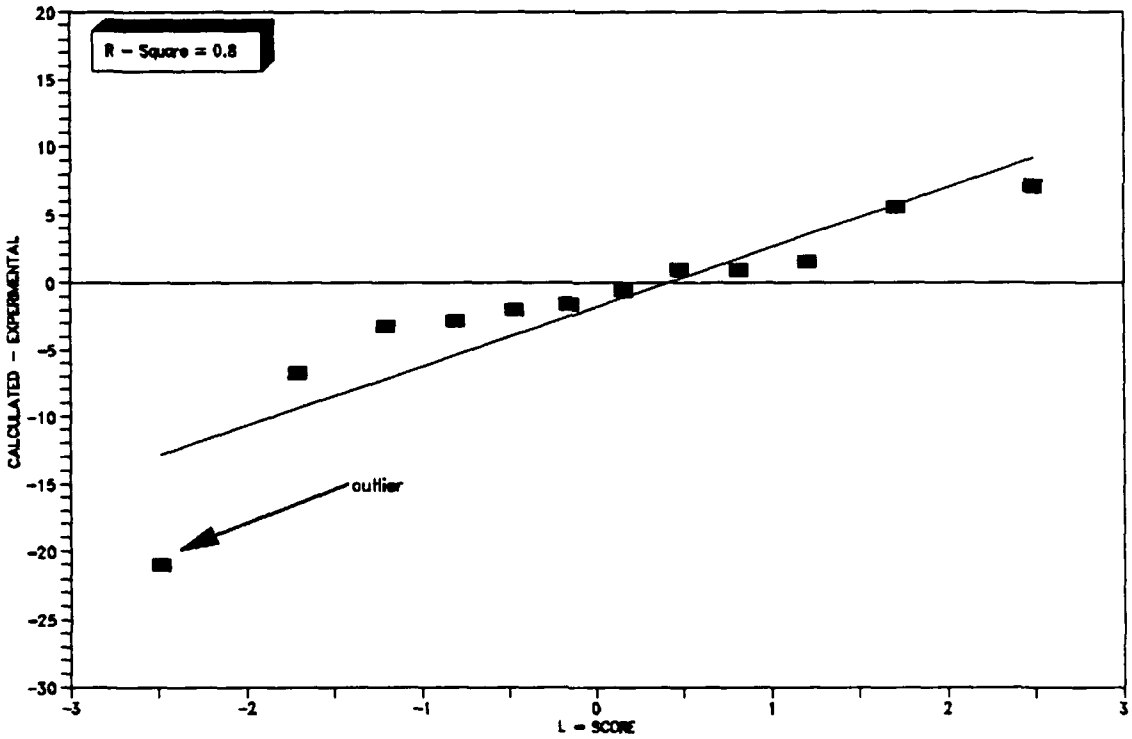

Figure 4. Test for Normal Distribution - Heat of Formation, MNDO, All Data

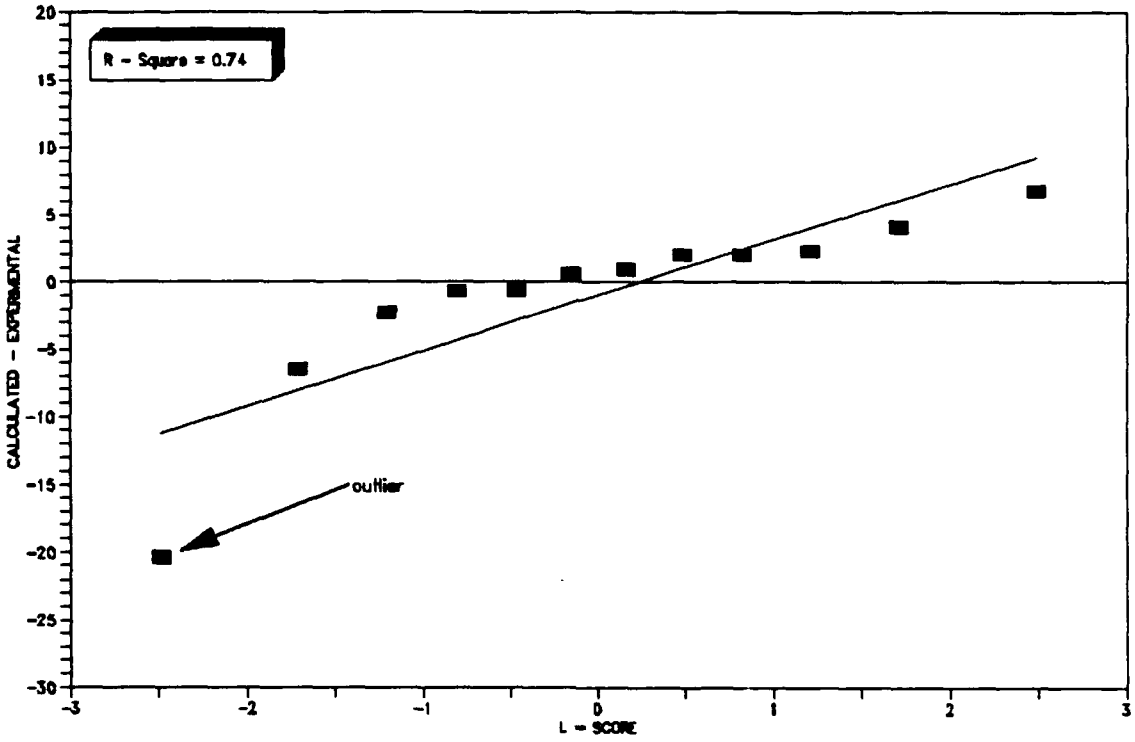

Figure 5. Test for Normal Distribution - Heat of Formation, AM1, All Data

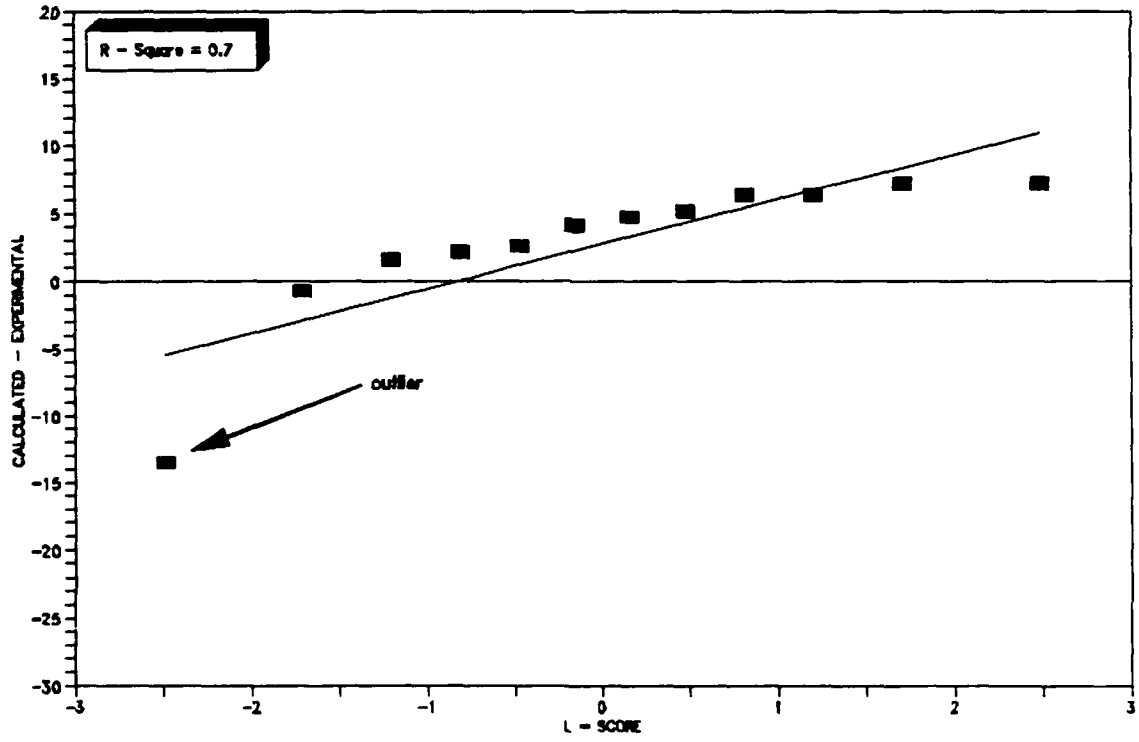

Figure 6. Test for Normal Distribution - Heat of Formation, PM3, All Data

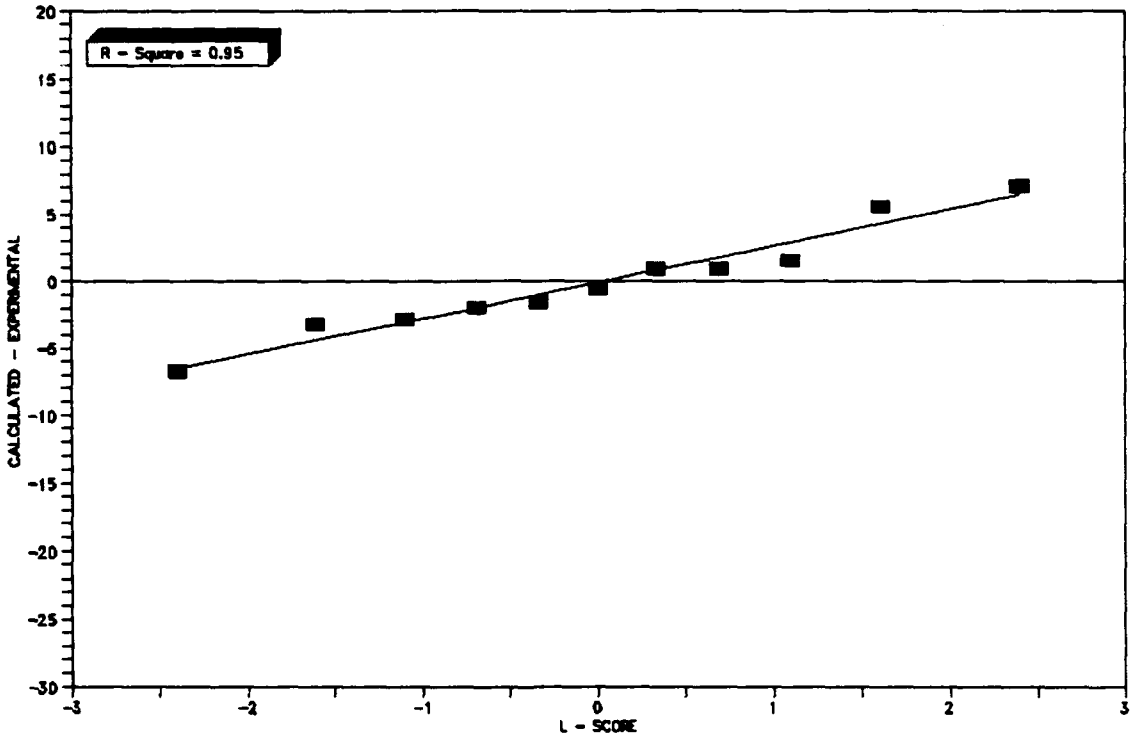

Figure 7. Test for Normal Distribution - Heat of Formation, MNDO, Low Region

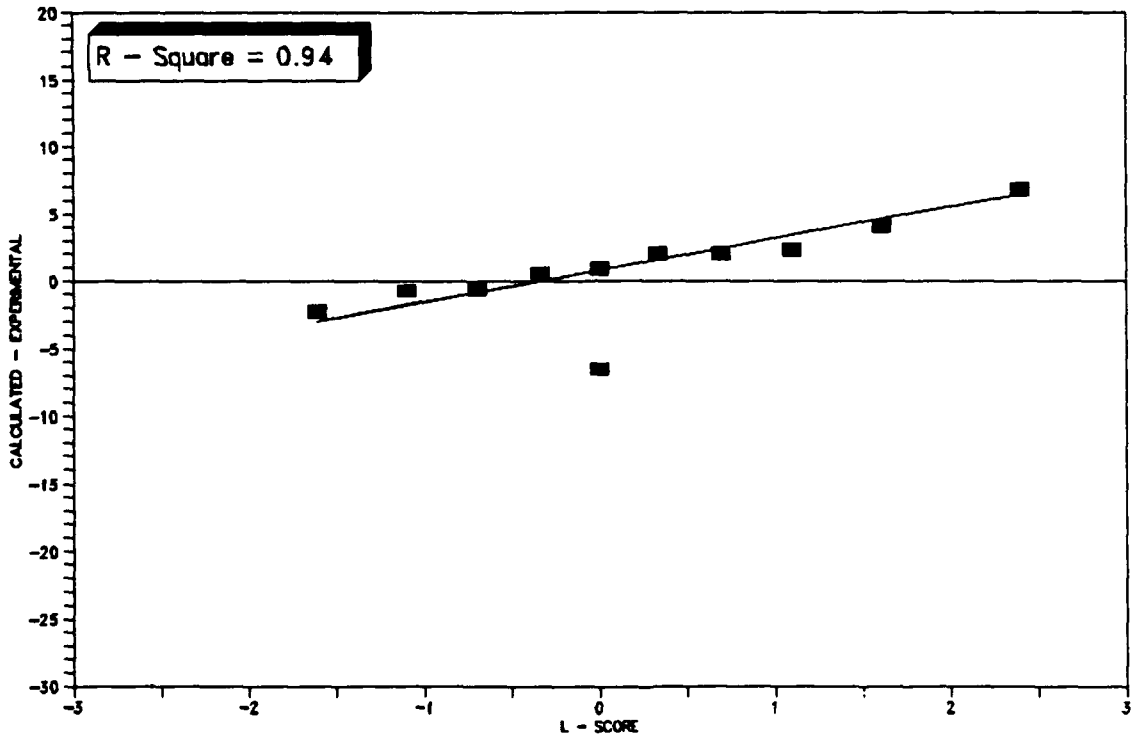

Figure 8. Test for Normal Distribution - Heat of Formation, AM1, Low Region

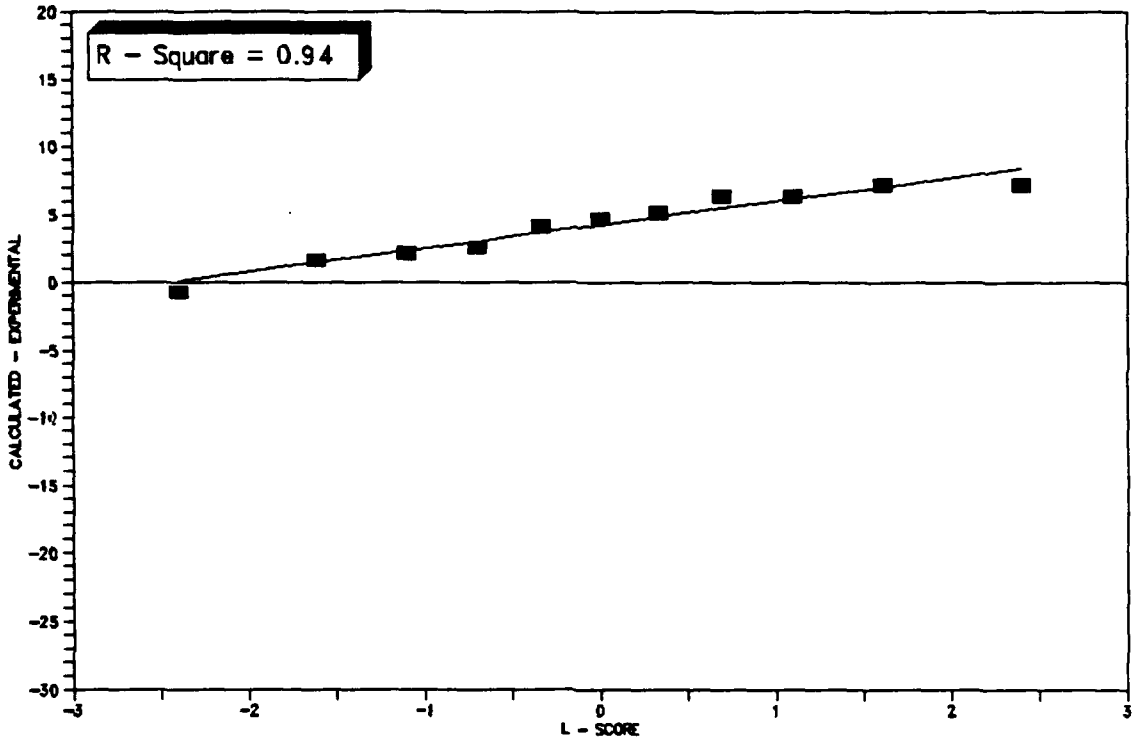

Figure 9. Test for Normal Distribution - Heat of Formation, PM3, Low Region

Table 7. Bias and SD of Calculated Heat of Formation

	MNDO	AM1	PM3
Bias	-2.00	-1.26	2.74
σ	7.07	6.86	5.43
R - Square	0.95	0.94	0.94

Figure 10 shows the relative difference of the calculated and experimental values as a function of the experimental values. Again, it is not possible to determine which of the three methods would yield better results. Figures 11, 12, and 13 show the relative differences as a function of their respective L-score. As can be seen, the relation is linear except for compounds with small (absolute) heat of formation. This can be expected since a small absolute error for these compounds will be large, relative to the heat of formation absolute value. When the outliers are removed, the linearity of the line become apparent as can be seen in Figures 14, 15, and 16. Table 8 gives the average, SD and R-square of the relative error in the applicable region.

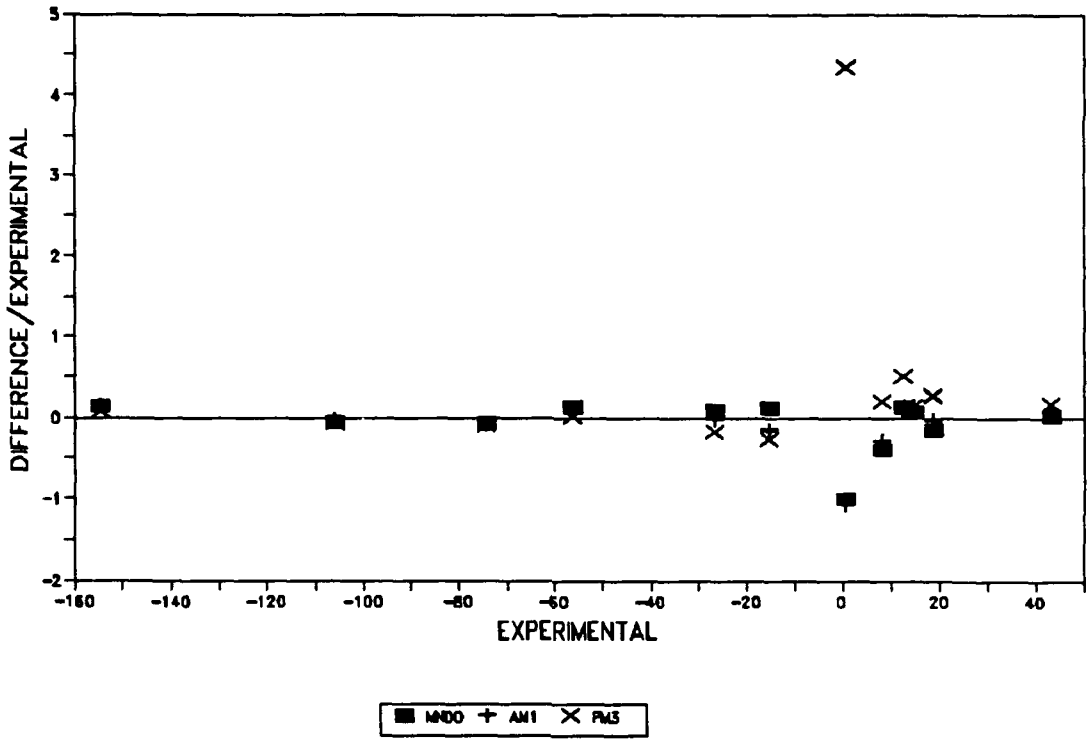

Figure 10. Relative Differences Versus Experimental Values - Heat of Formation

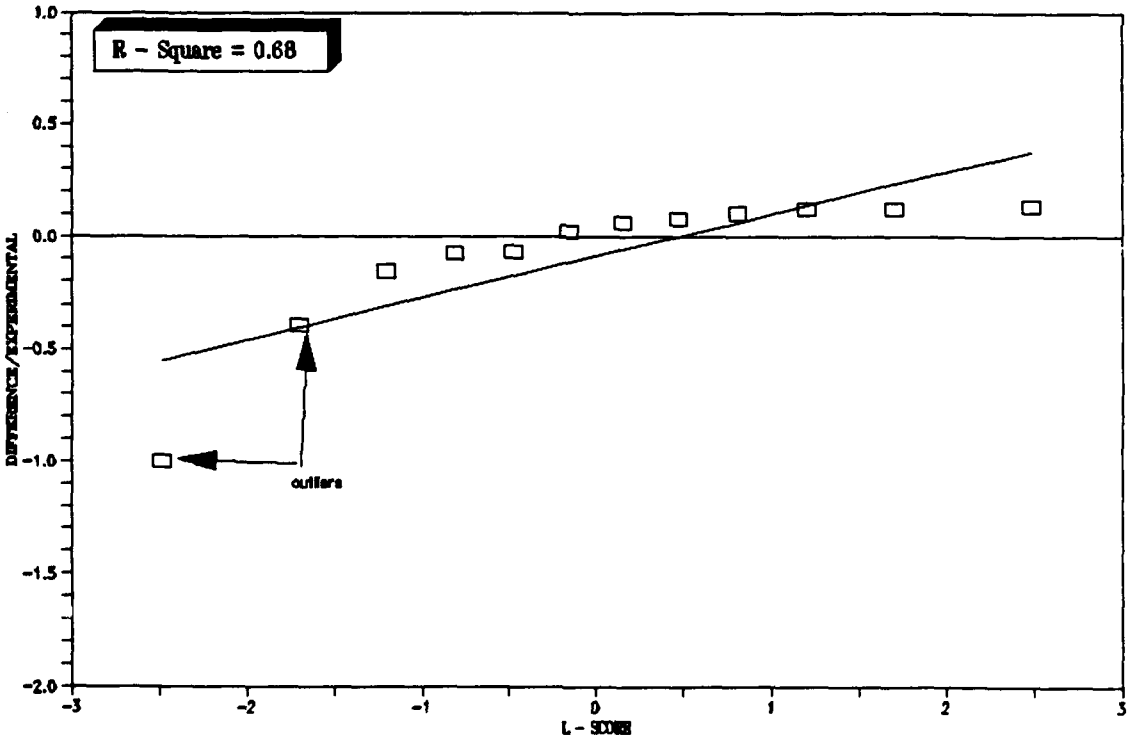

Figure 11. Test for Normal Distribution - Relative Difference, Heat of Formation, MNDO, All Data

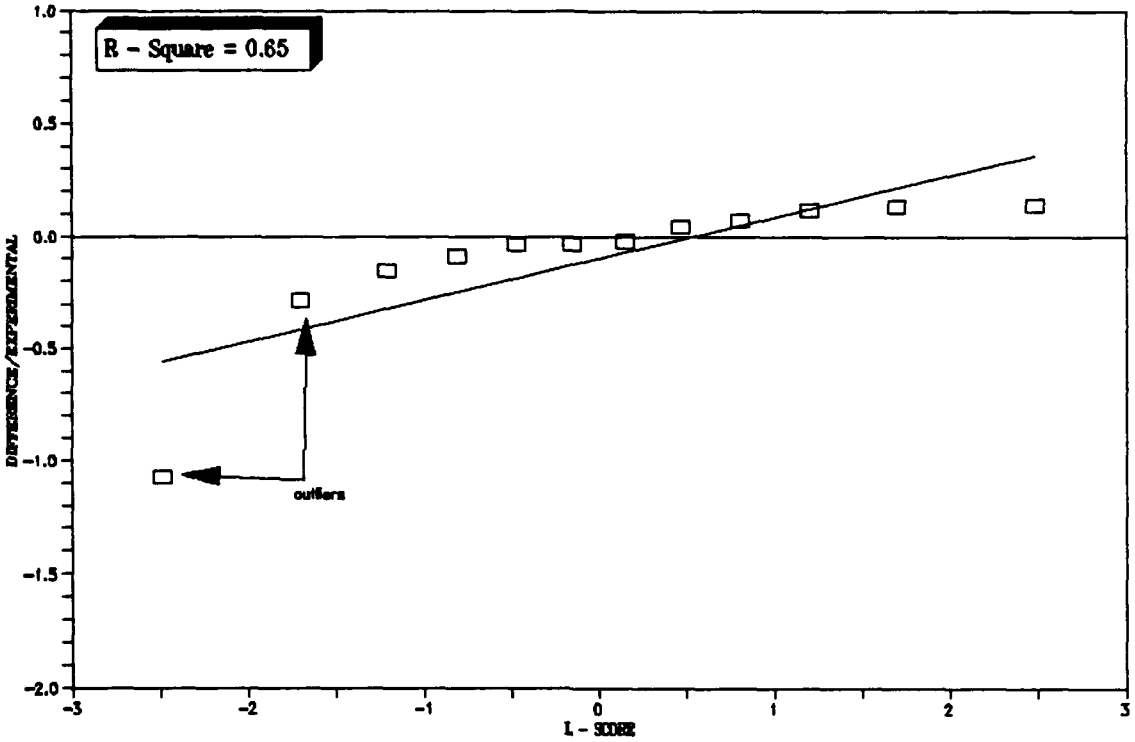

Figure 12. Test for Normal Distribution - Relative Difference, Heat of Formation, AM1, All Data

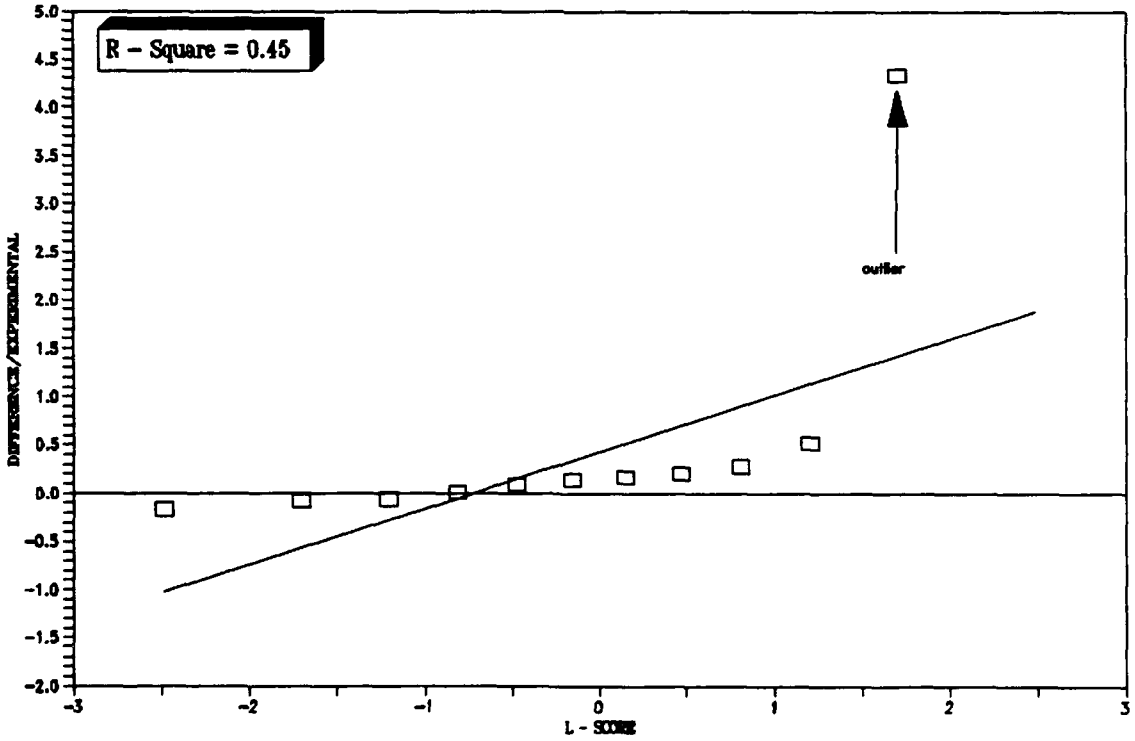

Figure 13. Test for Normal Distribution - Relative Difference, Heat of Formation, PM3, All Data

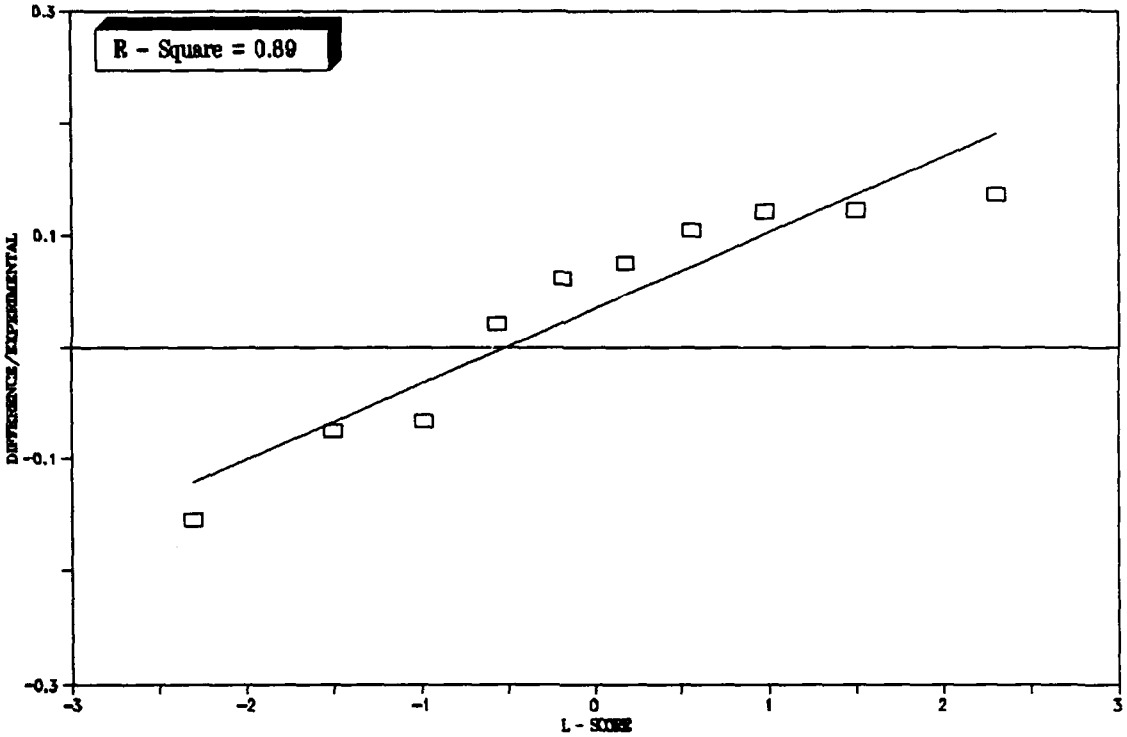

Figure 14. Test for Normal Distribution - Relative Difference, Heat of Formation, MNDO, High Region

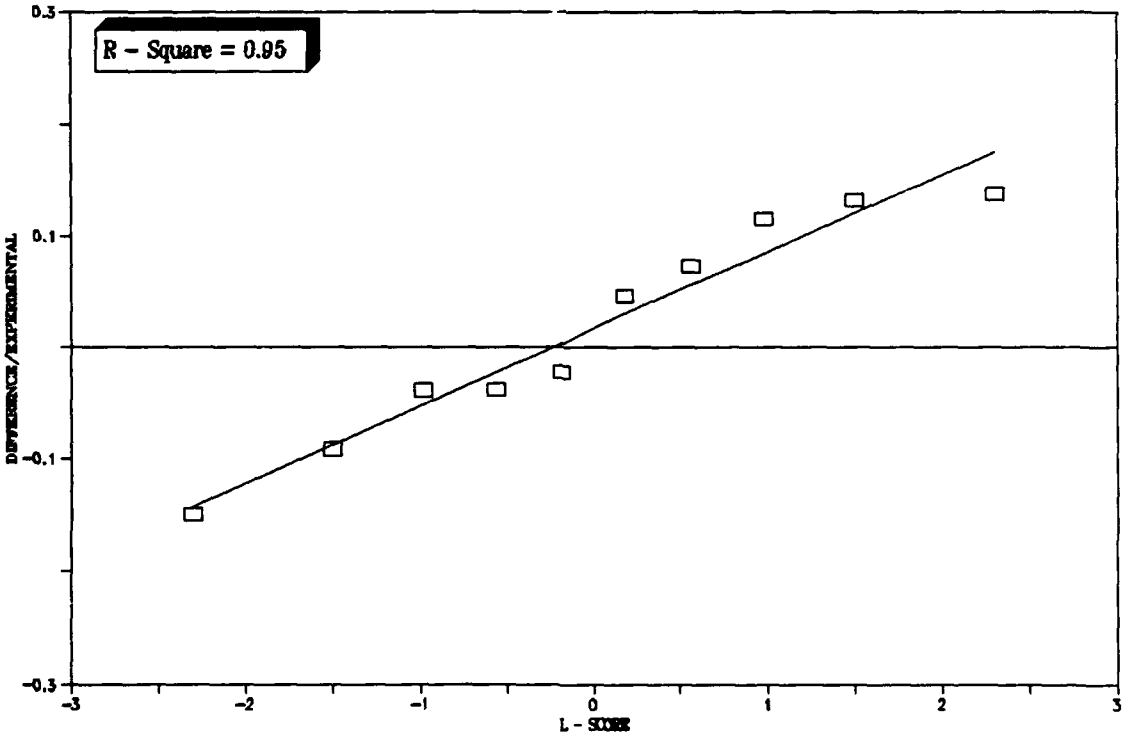

Figure 15. Test for Normal Distribution - Relative Difference, Heat of Formation, AM1, High Region

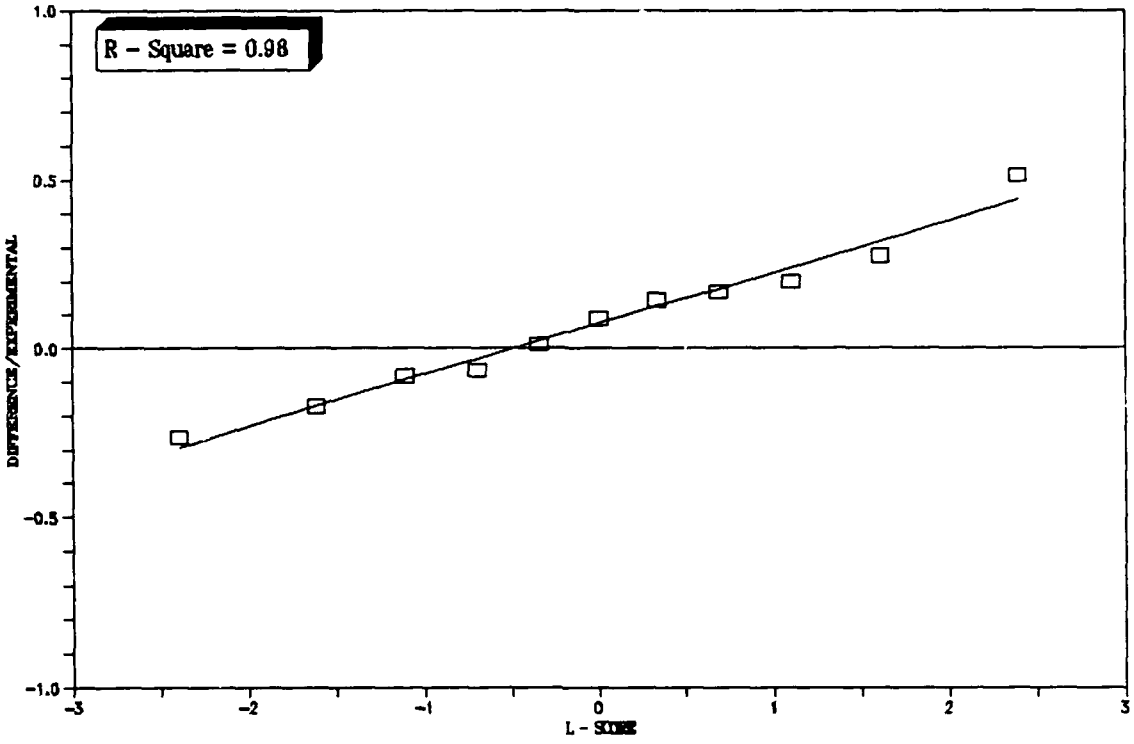

Figure 16. Test for Normal Distribution - Relative Difference, Heat of Formation, PM3, High Region

Table 8. Bias and SD of the Relative Calculated Heat of Formation

	MNDO	AM1	PM3
Bias	0.03	0.02	0.07
σ	0.10	0.09	0.21
R - Square	0.9	0.94	0.98

5.2 Dipole Moment.

The experimental and calculated values of the dipole moment are listed in Table 9 and are plotted in Figure 17. The calculation errors are plotted against the dipole moment in Figure 18. From these figures, it is not possible to determine which is the better method to calculate this property. The calculation errors are plotted against their respective L-score in Figures 19, 20, and 21 and the statistics data are given in Table 10.

Table 9. Comparison Between Experimental and Calculated Values of Dipole Moment

Molecule	Dipole Moment Debye (D)			
	Experiment[16]	Calculation MNDO	Calculation AM1	Calculation PM3
Ethylene	0.0	0.0	0.0	0.0
Vinyl Chloride	1.45	1.71	1.19	0.93
Vinyl Bromide	1.36	1.31	1.31	1.33
Ethyl Chloride	2.05	2.08	1.69	1.55
Ethyl Bromide	1.90	1.66	1.66	1.84
Vinylidene Chloride	1.28	1.85	1.21	0.78
Vinyl Acetate	1.79	1.66	1.73	1.77
Ethyl Acetate	1.82	1.85	1.80	1.84
Ethyl Alcohol	1.66	1.40	1.55	1.45
Vinyl Cyanide	3.67	3.00	3.00	3.25
Ethyl Cyanide	3.50	2.71	2.94	3.25
Tetrafluro Ethylene	0.00	0.00	0.00	0.00

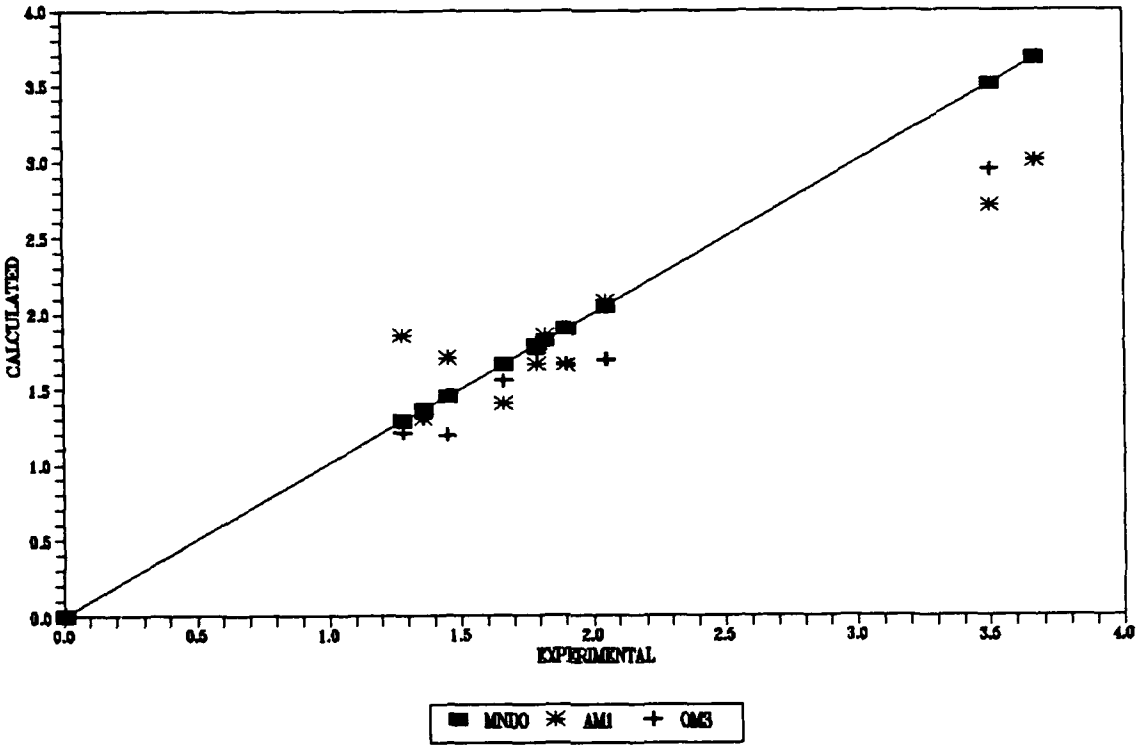

Figure 17. Calculated Versus Experimental Dipole Moment

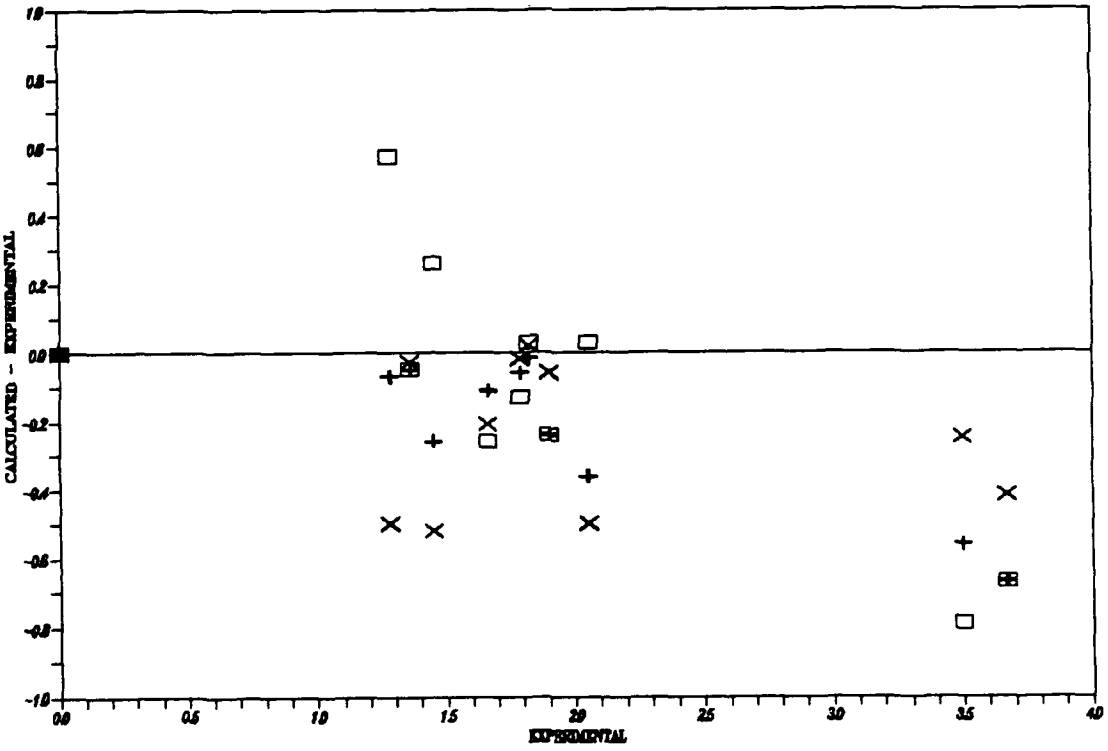

Figure 18. Dipole Moment Calculation Errors

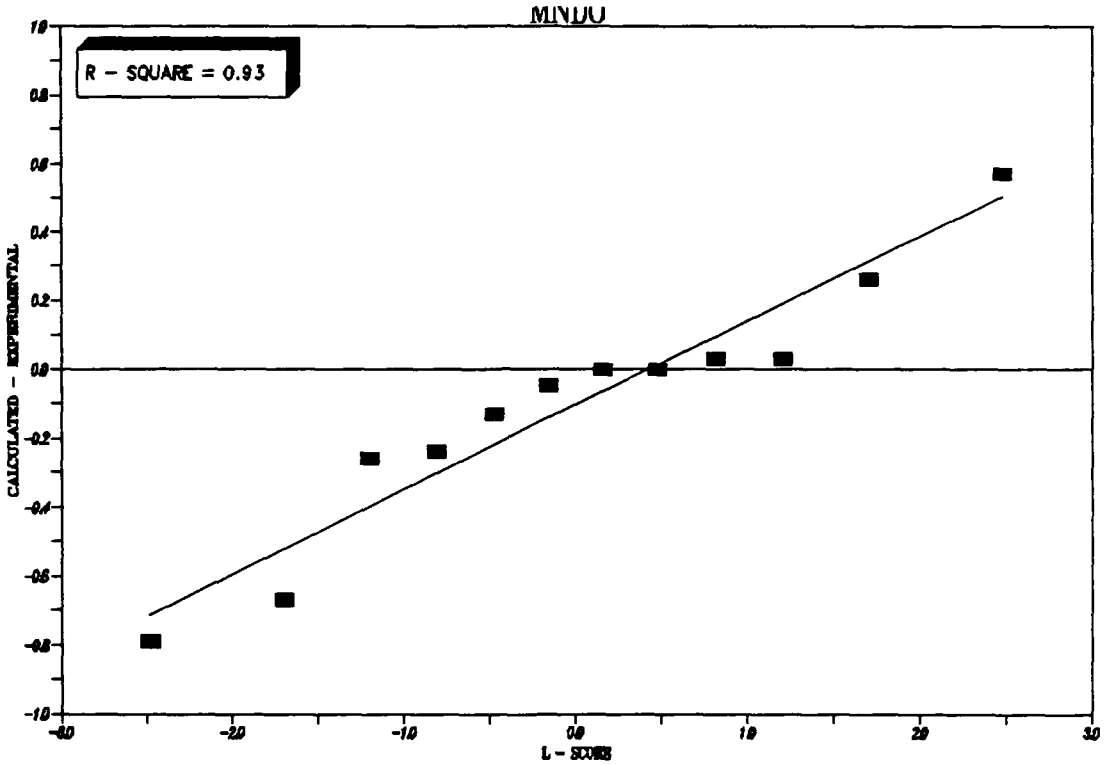

Figure 19. Test for Normal Distribution - Calculation Errors, MNDO

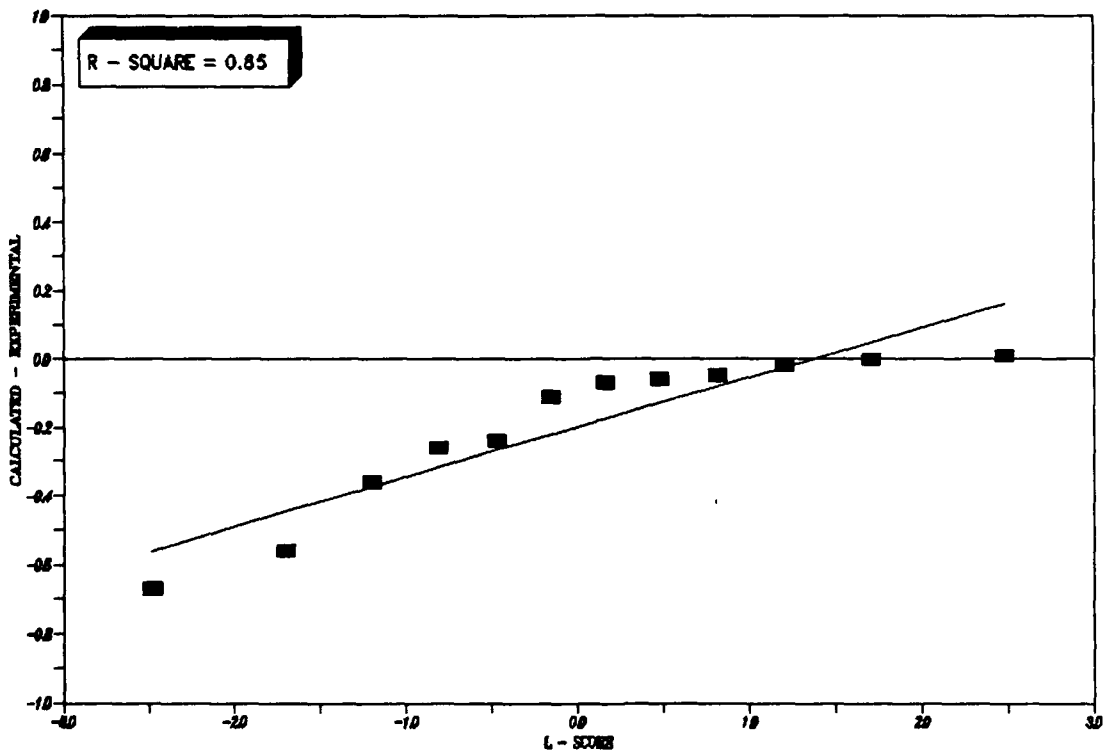

Figure 20. Test for Normal Distribution - Dipole Moment, AM1

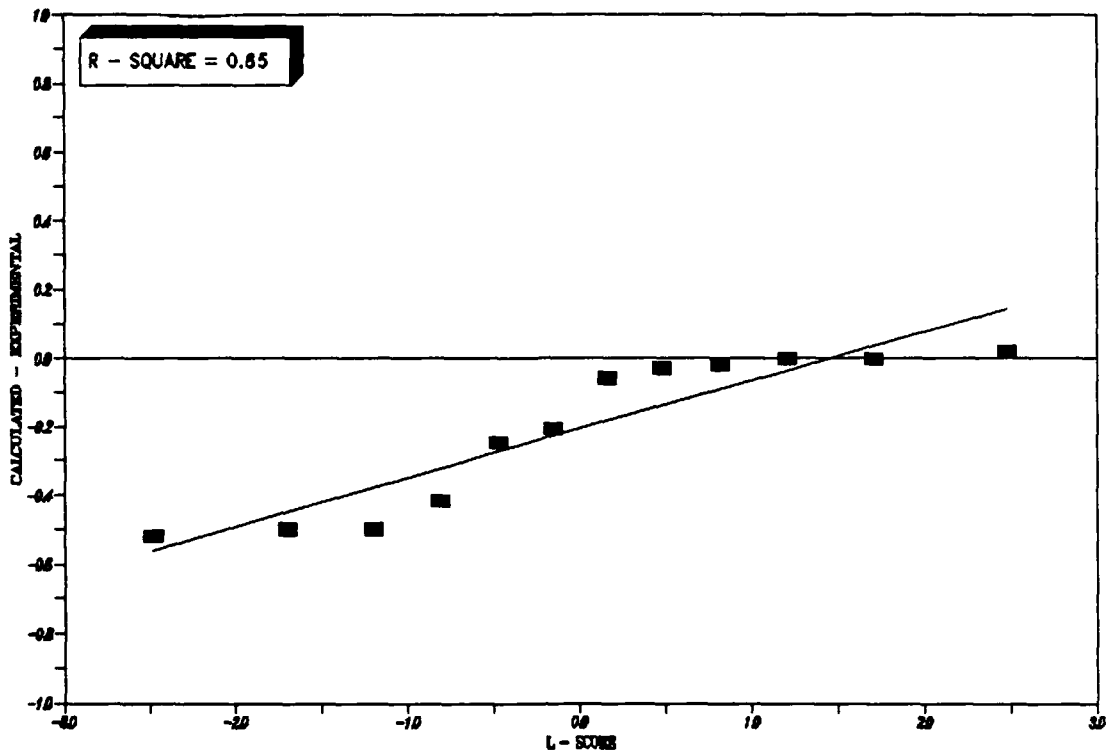

Figure 21. Test for Normal Distribution - Dipole Moment, PM3

Table 10. Bias and SD of the Calculated Dipole Moment

	MNDO	AM1	PM3
Bias	-0.10	-0.29	-0.21
σ	0.36	0.22	0.21
R - Square	0.93	0.85	0.85

5.3 Polarizability.

The experimental and calculated values of the polarizability are listen in Table 11 and are plotted in Figure 17. The calculation errors are plotted against the polarizability in Figure 22. From examining Figures 22 and 23, the polarizability results, calculated by the MNDO method given appears to be closest to the experimental values. However, closer examination (Figures 24-27) indicates that the calculation errors are not symmetrically distributed (i.e., the line of the calculated error against the L-score has low R-Square) giving any estimate low confidante level. On the other hand, the results obtained by PM3 are biased (Table 12), but the errors are distributed symmetrically around the calculated values giving the estimate a high degree of confidence.

Table 11. Comparison of Experimental and Calculated Polarizability

Molecule	Polarizability			
	Experiment[16]	Calculation MNDO	Calculation AM1	Calculation PM3
Ethylene	4.25	3.88	2.47	2.23
Vinyl Chloride	6.41	5.84	3.34	3.41
Vinyl Bromide	7.57	6.99	3.67	3.83
Ethyl Chloride	6.40	6.26	3.32	3.30
Ethyl Bromide	8.05	7.44	3.70	3.85
Vinylidene Chloride	7.89	7.90	4.30	4.70
Vinyl Acetate	8.20	8.87	6.30	5.80
Ethyl Acetate	9.70	9.05	5.97	5.40
Ethyl Alcohol	5.11	5.02	3.05	2.70
Vinyl Cyanide	8.05	6.04	4.28	4.23
Ethyl Cyanide	6.24	6.15	4.02	3.82

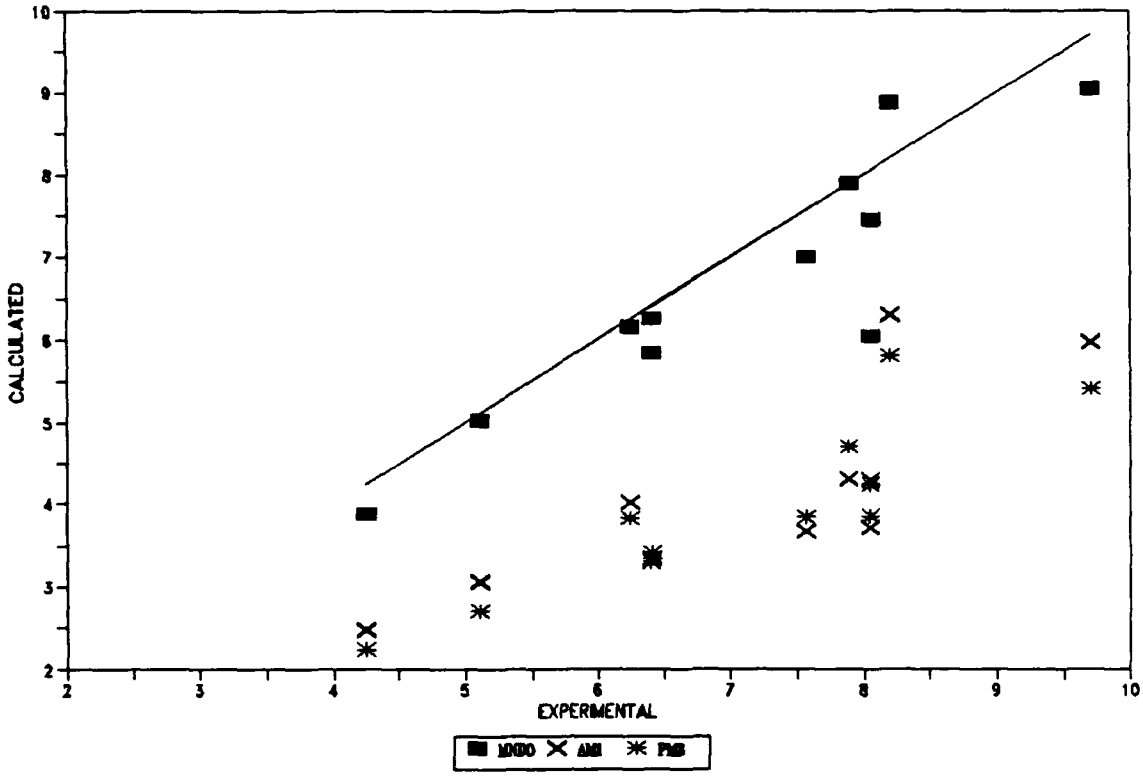

Figure 22. Calculated Versus Experimental Polarizability

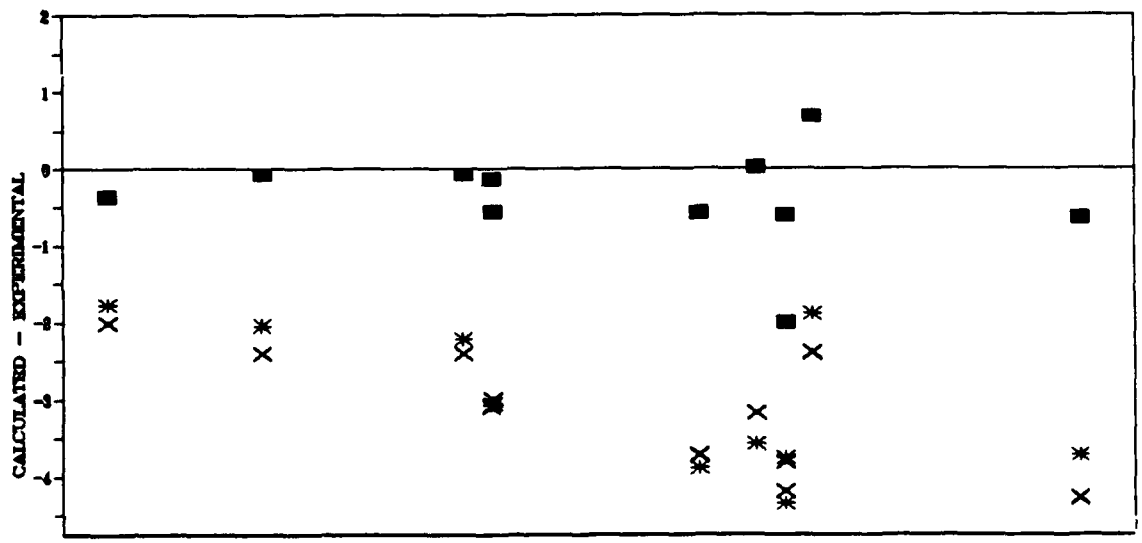

Figure 23. Polarizability Calculation Errors

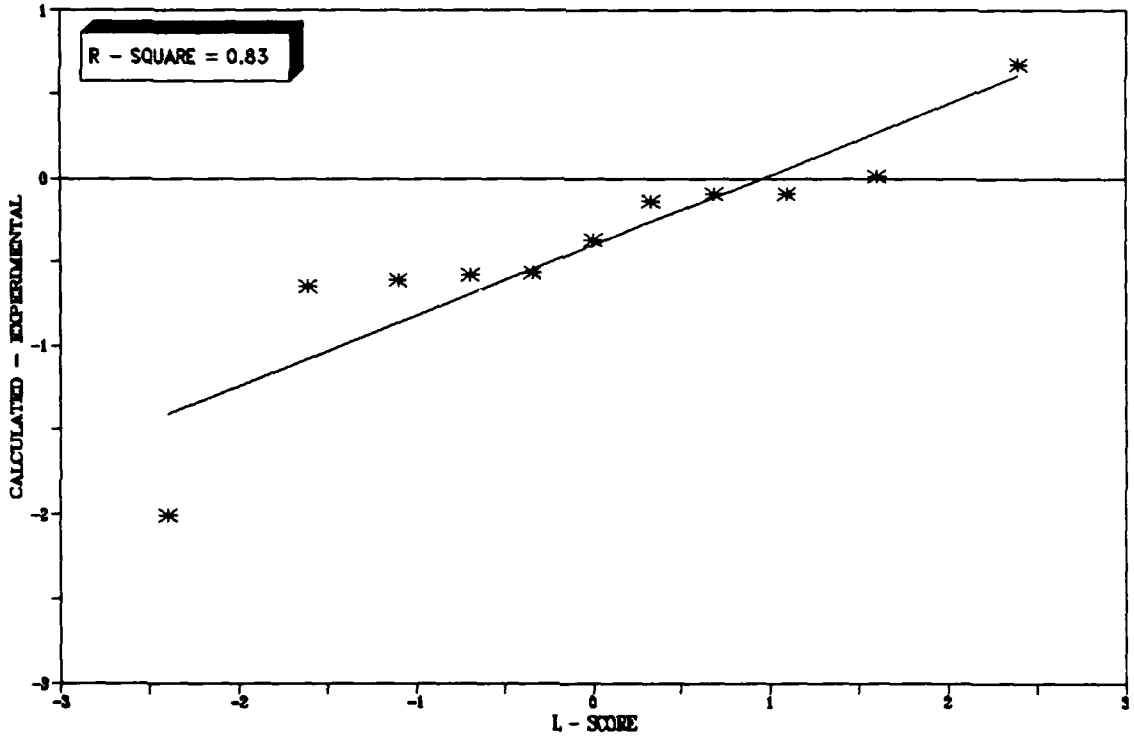

Figure 24. Test for Normal Distribution, Polarizability, MNDO

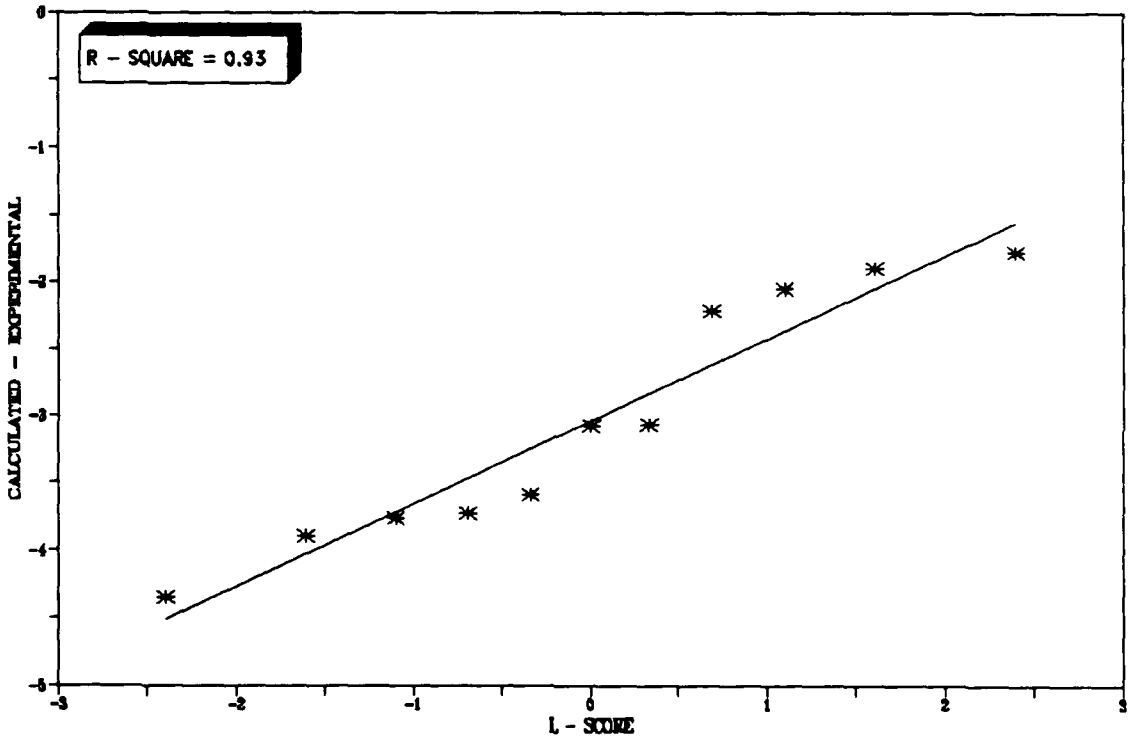

Figure 25. Test for Normal Distribution, Polarizability, AM1

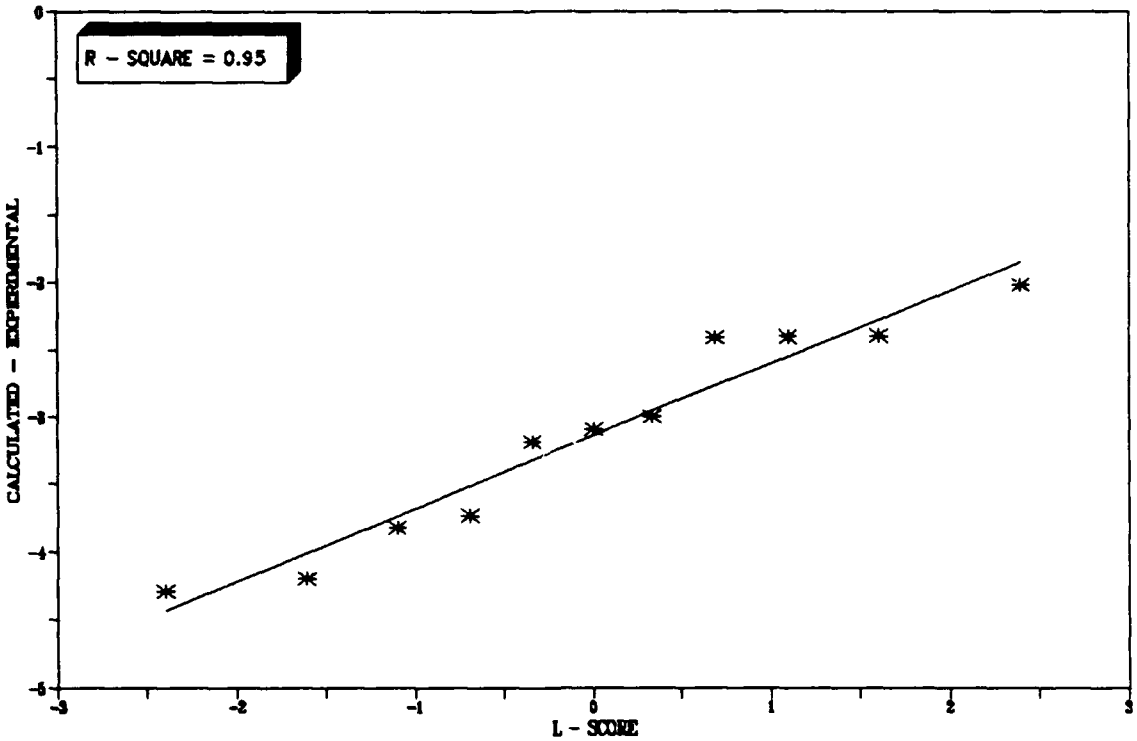

Figure 26. Test for Normal Distribution, Polarizability, PM3

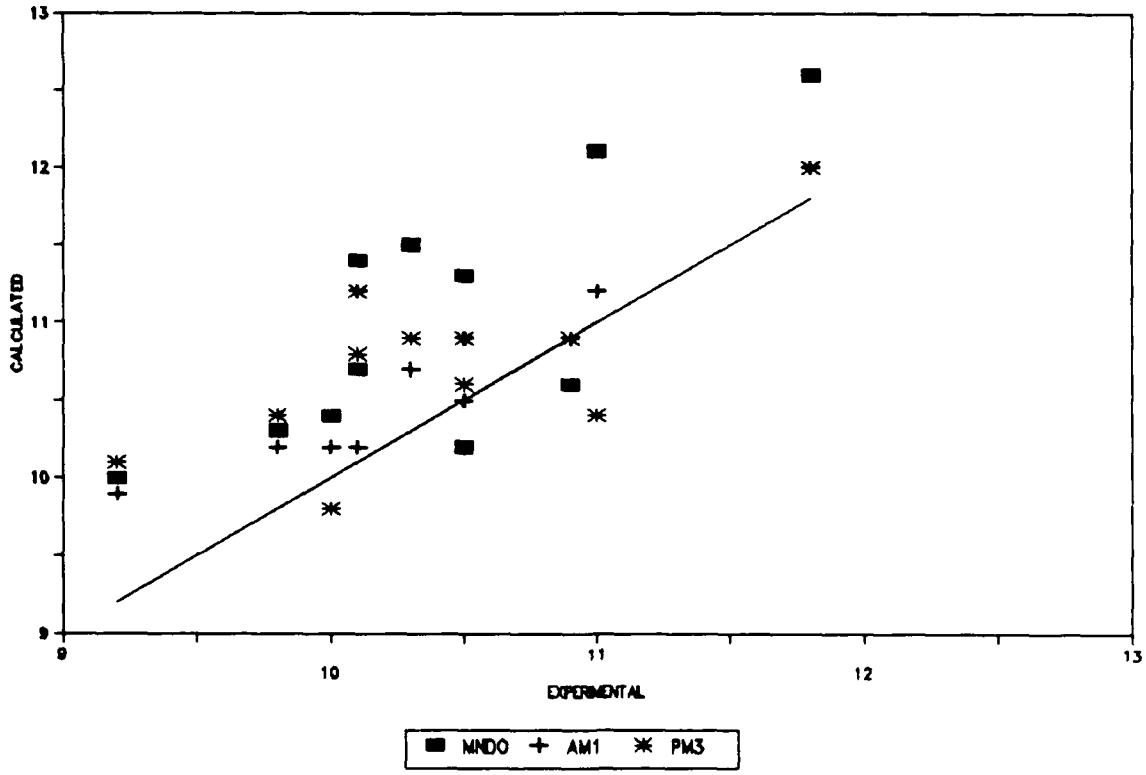

Figure 27. Calculated Versus Experimental Ionization Potential

Table 12. Bias and SD of the Calculated Polarizability

	MNDO	AM1	PM3
Bias	-0.40	-3.04	-3.15
σ	0.63	0.87	0.75
R - Square	0.83	0.93	0.95

5.4 Ionization Potential.

The experimental and calculated values of the dipole moment are listed in Table 13 and are plotted in Figure 17. The calculation errors are plotted against the dipole moment in Figure 28. It is not possible, from these figures, to determine which is the better method to calculate these properties. The calculation errors are plotted against their respective L-score in Figures 29, 30, and 31 and the calculation errors are given in Table 14.

Table 13. Comparison of Experimental and Calculated Ionization Potential

	Ionization Potential (EV)			
Molecule	Experiment[16]	Calculation MNDO	Calculation AM1	Calculation PM3
Ethylene	10.5	10.2	10.5	10.6
Vinyl Chloride	10.0	10.4	10.2	9.8
Vinyl Bromide	9.8	10.3	10.2	10.9
Ethyl Chloride	11.0	12.1	11.2	10.4
Ethyl Bromide	10.3	11.5	10.7	10.9
Vinyl Acetate	9.2	10.0	9.9	10.1
Ethyl Acetate	10.1	11.4	11.2	11.2
Ethyl Alcohol	10.5	11.3	10.9	10.9
Vinyl Cyanide	10.9	10.6	10.9	10.9
Ethyl Cyanide	11.8	12.6	12.0	12.0
Tetrafluro Ethylene	10.1	10.7	10.2	10.8

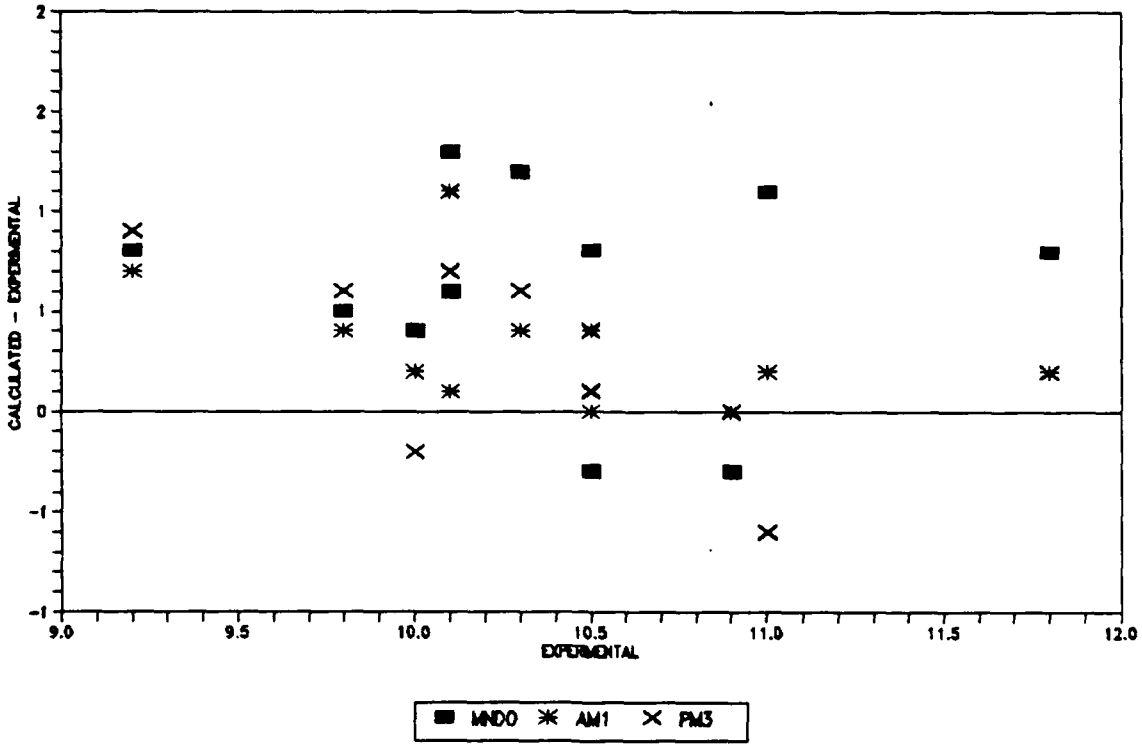

Figure 28. Ionization Potential Calculation Errors

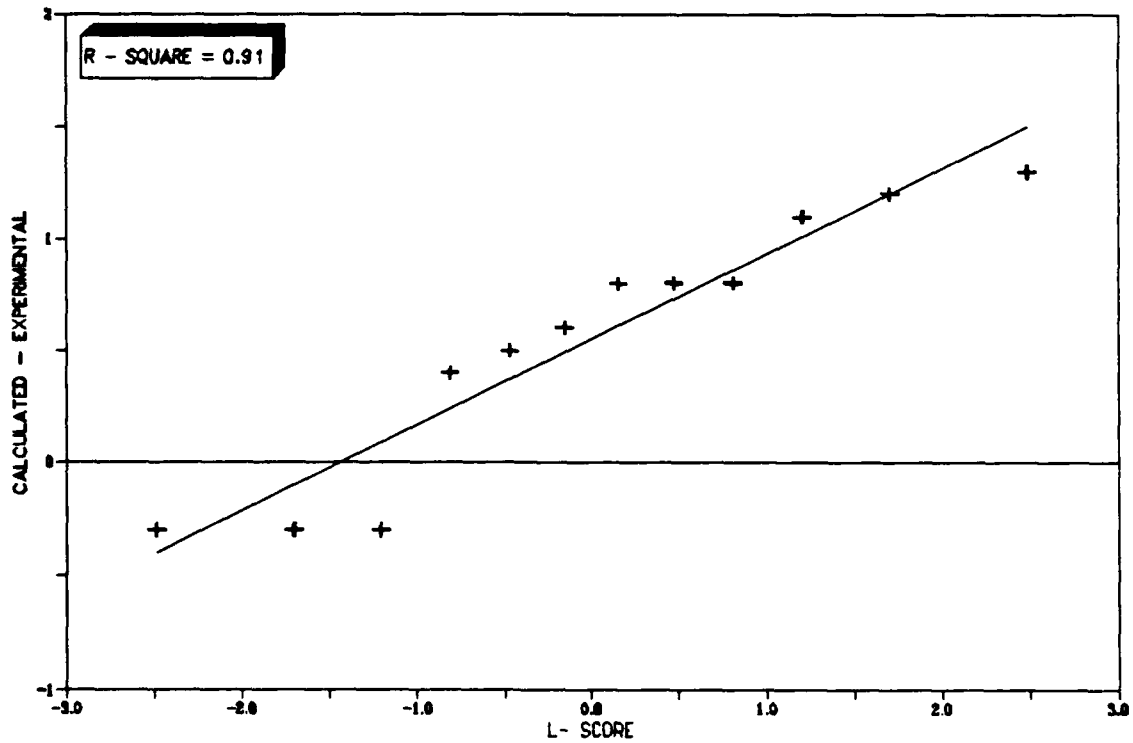

Figure 29. Test for Normal Distribution - Ionization Potential, MNDO

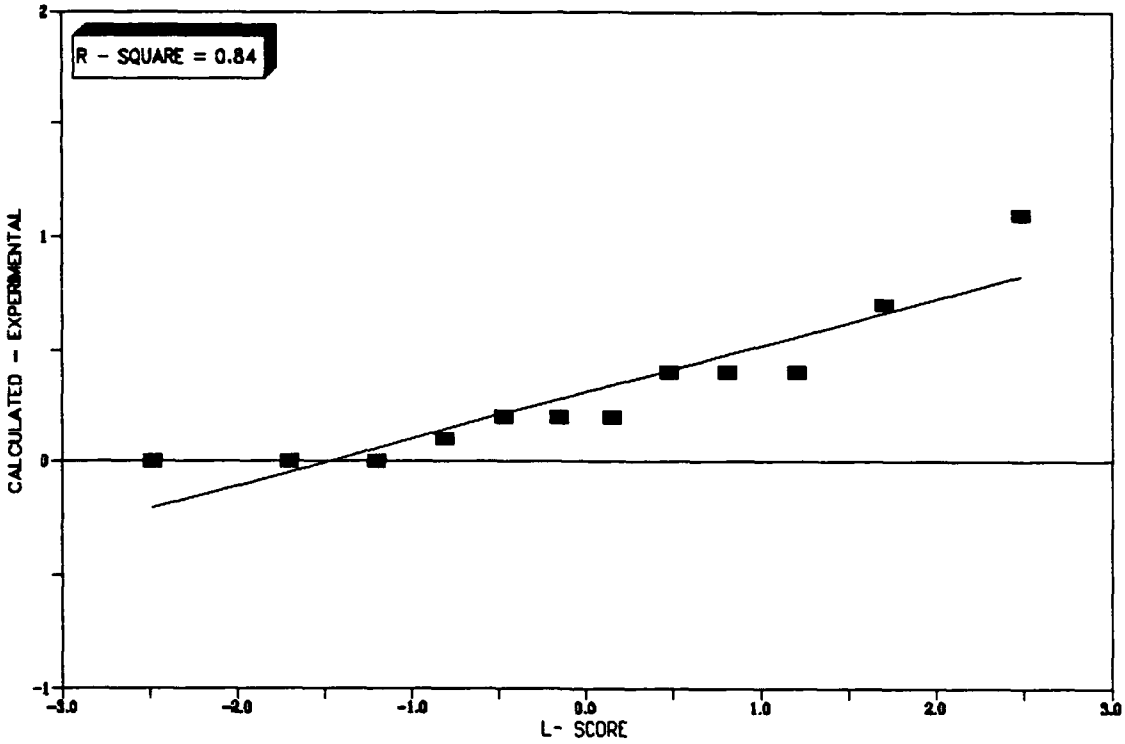

Figure 30. Test for Normal Distribution - Ionization Potential, AM1

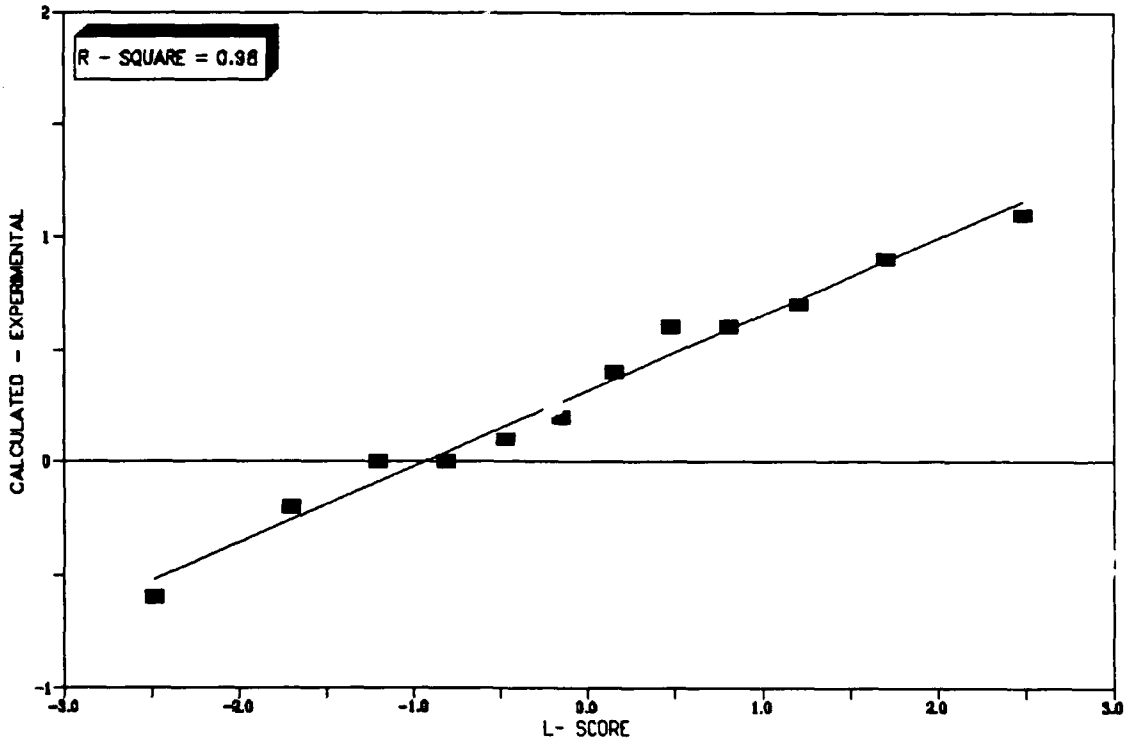

Figure 31. Test for Normal Distribution - Ionization Potential, PM3

Table 14. Bias and SD of the Calculated Ionization Potential

	MNDO	AM1	PM3
Bias	0.55	0.31	0.32
σ	0.55	0.31	0.47
R - Square	0.91	0.84	0.98

6. CONCLUSIONS

This report clearly shows the value of employing rigorous statistical methods when evaluating the adequacy of semi-empirical molecular orbital methods. We showed that by employing the right methods we were able to separate systematic and random errors in the calculation. Table 15 summarizes the results obtained in this study with a limited set of data. The table provides the recommended calculation method for each of the four physical properties studied together with the bias and SD of the calculation errors. In the near future, we plan to extend the analysis to a much larger data set to validate the methodology developed here.

Table 15. Recommended Computational Methods for Heat of Formation, Ionization Potential, Dipole Moment, and Polarizability

Physical Property	Recommended Method	Bias	SD (σ)
Heat of Formation	For molecules with heat of formation below 10 Kcal/mole (absolute) use MNDO	-2 (Kcal/mole) for MNDO	7.1 (Kcal/mole) for MNDO
	For molecules with heat of or reaction above 100 Kcal/mole (absolute) use PM3	0.03 for the ratio of Calculated / Experimental for PM3	0.1 for the ration of Calculated / Experimental for PM3
	For molecules with heat of formation between 10 and 100 Kcal/mole use either one		
Ionization Potential	PM3	0.32 (ev)	0.47 (ev)
Dipole Moment	MNDO	-0.1 (deby)	0.36 (deby)
Polarizability	PM3	-3.15 (A^3)	0.75 (A^3)

LITERATURE CITED

1. Krishnan, P.N., Morris, R.E., Famini, G.R., and Birenzvige, A., Predicting Polymer Properties by Computational Methods I: Poly Vinyl Chloride and Its Homologs, CRDEC-TR-254, U.S. Army Chemical Research, Development and Engineering Center, Aberdeen Proving Ground, MD, January 1991, UNCLASSIFIED Report.

2. Dewar, M.J., and Thiel, W., "Ground States of Molecules, 38: The MN DO Methods. Approximations and Parameters," J. Am. Chem. Soc. Vol. 99, p 4899 (1977).

3. Dewar, M.J., Zoebisch, E.G., Healy, E.F., and Stewart J.J.P., "AM1: A New General Purpose Quantum Mechanical Model," J. Am. Chem. Soc. Vol. 107, p 3902 (1985).

4. Stewart, J.J.P., "Optimization of Parameters for Semi-Empirical Methods, I. Method," J. Am. Chem. Soc. Vol. 10, p 209 (1989).

5. Dewar, N.J.S., Healy, E.F., Holder, A.J., and Yate-Chin, Y., "Comments on a Comparison of AM1 with a Recently Developed PM3 Method", J. Comp. Chem. Vol. 11, p 541 (1990).

6. Stewart, J.J.P., Reply to "Comments on a Comparison of AM1 with a Recently Developed PM3 Method," J. Comp. Chem. Vol. 11, p 543 (1990).

7. Stewart, J.J.P., "Optimization of Parameters for Semi Empirical Methods II: Applications," J. Comp. Chem. Vol. 10, p 221 (1989).

8. Dewar, M.J., Thiel, W. "Ground State of Molecules. 39 MNDO Results for Molecules Containing Hydrogen, Carbon, Nitrogen, and Oxygen," J. Am. Chem. Soc. Vol. 99, p 4907 (1977).

9. Famini, G.R., "Applying Computational Chemistry Techniques to Simulant Technology II: Estimation and Prediction of Physical and Chemical Properties," In Proceedings of the Second International Simulant Workshop, CRDEC-SP-002, U.S. Army Chemical Research, Development and Engineering Center, Aberdeen Proving Ground, MD, December 1988, UNCLASSIFIED Report.

10. Hehre, W.J., Radem, L., Schleyer, P.V.R., and Pople, J.A., Ab Initio Molecular Orbital Theory, John Wiley and Sons, New York, NY, 1986.

11. Shaefer, H.F., Electronic Structure of Atoms and Molecules, Addison-Wesley, Reading, MA, 1972.

12. Dewar, M.J.S., "Computing Calculated Reactions," Chem. Br. Vol. 11, p 97 (1975).

13. Dewar, M.J.S., "Quantum Organic Chemistry," Science Vol. 187, p 1037 (1975).

14. Dewar, M.J.S., and Starch, D.M., "Comparative Tests of Theoretical Procedures for Studying Chemical Reactions," J. Am. Chem. Soc. Vol. 107, p 3898 (1985).

15. Ivey, R.C., and Davis, M.I., "On Gas Phase Electron Diffraction Study of Vinyl Chloride," J. Chem. Phys. Vol. 57, p 1909 (1972).

16. CRC Handbook of Chemistry and Physics, 68th Edition, CRC Press, Inc., Boca Raton, FL (1987-88).

17. Beach, J.Y., and Stevenson, D.P., "Electron Differentiation of the Molecular Structures of Ethyl Chloride, Ethyl Bromide, Isopropyl Chloride, Isopropyl Bromide, Methyl Chloroform and Isocrotyl Chloride," J. Am. Chem. Soc. Vol. 61, p 2643 (1939).

18. Stewart J.J.P., MOPAC: A General Molecular Orbital Package, FJSR L-TR-88-0007, Frank J. Seiler Research Laboratory, U.S. Air Force Academy, Colorado Springs, CO, December 1988, UNCLASSIFIED Report.

19. Leonard, J.M., A User's Guide to the Molecular Modeling Analysis and Display System, CRDEC-TR-030, U.S. Army Chemical Research, Development and Engineering Center, Aberdeen Proving Ground, MD, January 1989, UNCLASSIFIED Report.