

ity for these molecules. Chloroform was not included because of some uncertainties about the chlorine parameter.

DISCUSSION

Two observations come boldly into view: the correlations of the connectivity index with cavity surface area, polarizability, and biological activity are very good, suggesting a fundamental significance for the connectivity index; and the connectivity index value for each molecule is very simply derived, almost by inspection, from no more than four numbers (1, 2, 3, and 4) describing connections.

The mathematical derivation of the connectivity index is extremely simple and is based on discrimination of bond types in a molecule according to the number of its nearest neighbors. The physical significance of the connectivity index is more fundamental than may appear. The individual connectivity index values reflect molecular connectivity and present ultimate fragmentation of the molecular skeleton. Any function of a molecule can be represented as a power expansion of bond contributions. At this initial stage, attention was confined to the linear term of such bond additive schemes.

One may look at connectivity index values as distinctive weighting of bond contributions. It is very significant that the weighting is based solely on the valency of vertexes of molecular graphs and, hence, is nonempirical.

It is plausible to expect that the connectivity index relates to such molecular properties as enclosure volume and surface area. The predictive value of this simply computed index is demonstrated to the biological scientist. In subsequent papers, additional correlations will be described.

REFERENCES

- (1) L. B. Kier, *J. Pharm. Sci.*, **61**, 1394(1972).
- (2) H. D. Holtje and L. B. Kier, *J. Med. Chem.*, **17**, 814(1974).
- (3) H. D. Holtje and L. B. Kier, *J. Pharm. Sci.*, **63**, 1435(1974).
- (4) H. D. Holtje and L. B. Kier, *J. Theoret. Biol.*, **48**, 197(1974).
- (5) F. London, *Discuss. Faraday Soc.*, **1936**, 8.
- (6) P. Claverie and R. Rein, *Int. J. Quantum Chem.*, **3**, 535(1969).
- (7) D. Agin, L. Hersh, and D. Holtzman, *Proc. Nat. Acad. Sci. USA*, **53**, 952(1965).
- (8) C. P. Smyth, "Dielectric Structure and Behavior," McGraw-Hill, New York, N.Y., 1955.
- (9) M. Randić, *J. Amer. Chem. Soc.*, in press.
- (10) R. B. Hermann, *J. Phys. Chem.*, **76**, 2754(1972).
- (11) G. L. Amidon, S. H. Yalkowsky, and S. Leung, *J. Pharm. Sci.*, **63**, 1858(1974).

ACKNOWLEDGMENTS AND ADDRESSES

Received January 27, 1975, from the **Massachusetts College of Pharmacy, Boston, MA 02115*, and [†]*Tufts University, Medford, MA 02155*

Accepted for publication March 28, 1975.

The authors acknowledge the assistance of the Eastern Nazarene College Computer Center. L. H. Hall is grateful for the support of Eastern Nazarene College during his sabbatical leave.

[‡] On sabbatical leave from the Department of Chemistry, Eastern Nazarene College, Quincy, MA 02170

* To whom inquiries should be directed.

Molecular Connectivity II: Relationship to Water Solubility and Boiling Point

LOWELL H. HALL *, LEMONT B. KIER [‡], and WALLACE J. MURRAY

Abstract □ The connectivity index, easily computed by arithmetic and based upon the degree of connectedness at each vertex in the molecular skeleton, is shown to give highly significant correlations with water solubility of branched, cyclic, and straight-chain alcohols and hydrocarbons as well as with boiling points of alcohols. These correlations are superior to those based on well-founded theory relating to solvent cavity surface area. An empirical modification to the connectivity index gave an improved correlation for both solubilities and boiling points.

Keyphrases □ Molecular connectivity—correlation with water solubility and boiling point, hydrocarbons, alcohols □ Solubility—correlation with molecular connectivity, hydrocarbons, alcohols □ Boiling point—correlation with molecular connectivity, hydrocarbons, alcohols

It has long been known that branched- and straight-chain organic molecules differ in their properties. For example, branched-chain alcohols and hydrocarbons generally have lower boiling points and higher solubilities than the corresponding straight-chain isomers. However, until the recent suggestion by Randić (1), there was no simply computed quantitative scheme for correlating physicochemical data with such topological characteristics as branching. The earlier ideas of Hosoyea (2) and Smolenski (3)

have not received acceptance as means for correlating properties. The Hosoyea index shows only a qualitative correlation, and the Smolenski additivity function becomes complex to apply.

Since the manner in which organic groups are connected to form a molecule appears to influence molecular properties, the relationship between a connectivity index and various physicochemical and biological properties was studied. Success in correlating the connectivity index (χ) with polarizability and biological activity of certain nonspecific local anesthetics was reported in Part I (4).

The great utility of a simply computed index that could be used to predict physicochemical properties led to the study of the water solubilities of aliphatic alcohols and hydrocarbons. By applying the significant structure theory to the solubilities of these compounds, several studies showed that the log of the solubilities correlates well with the calculated surface area of the cavity required to accommodate the solute in the solvent water (5, 6).

In Part I, a highly significant correlation of the cavity surface area to the connectivity index was shown (4). This paper explores the correlation of the

Table I—Boiling Points of Alcohols and Water Solubilities of Alcohols and Hydrocarbons *versus* Connectivity Index (χ)

Compound	χ	Tb_{obs}^a	Tb_{calc}^b	S_{obs}^a , molal	S_{calc}^c , molal
1-Butanol	2.414	117.7°	116.6°	1.006	0.748
2-Methylpropanol	2.269	107.9°	111.0°	1.023	1.079
2-Butanol	2.269	99.5°	100.0°	1.068	1.805
1-Pentanol	2.914	137.8°	135.8°	0.26	0.211
3-Methylbutanol	2.807	131.2°	130.3°	0.311	0.305
2-Methylbutanol	2.769	128.7°	131.7°	0.347	0.277
2-Pentanol	2.769	119.0°	119.2°	0.53	0.510
3-Pentanol	2.807	115.3°	120.6°	0.615	0.463
3-Methyl-2-butanol	2.641	111.5°	114.3°	0.667	0.705
2-Methyl-2-butanol	2.561	102.0°	104.6°	1.403	1.170
1-Hexanol	3.414	157.0°	155.0°	0.0614	0.0597
2-Hexanol	3.269	139.9°	138.4°	0.136	0.144
3-Hexanol	3.307	135.4°	139.9°	0.160	0.131
3-Methyl-3-pentanol	3.122	122.4°	126.2°	0.436	0.283
2-Methyl-2-pentanol	3.061	121.4°	123.9°	0.327	0.331
2-Methyl-3-pentanol	3.179	126.5°	134.9°	0.2	0.181
3-Methyl-2-pentanol	3.179	134.2°	134.9°	0.194	0.181
2,3-Dimethyl-2-butanol	2.943	118.6°	119.3°	0.427	0.446
3,3-Dimethylbutanol	3.061	143.0°	141.5°	0.075	0.146
3,3-Dimethyl-2-butanol	2.943	120.0°	125.9°	0.244	0.329
4-Methylpentanol	3.269	151.8°	149.5°	0.102	0.0861
4-Methyl-2-pentanol	3.124	131.7°	132.8°	0.163	0.208
2-Ethylbutanol	3.345	146.5°	152.4°	0.0616	0.0710
Cyclohexanol ^d	3.393	161.0°	143.2°	0.383	0.373
1-Heptanol	3.914	176.3°	174.2°	0.0155	0.0169
2-Methyl-2-hexanol	3.561	142.5°	143.1°	0.0843	0.0934
3-Methyl-3-hexanol	3.622	142.4°	145.4°	0.104	0.0801
3-Ethyl-3-pentanol	3.683	142.5°	147.7°	0.147	0.0687
2,3-Dimethyl-2-pentanol	3.481	139.7°	140.0°	0.135	0.114
2,3-Dimethyl-3-pentanol	3.504	139.0°	140.9°	0.144	0.108
2,4-Dimethyl-2-pentanol	3.416	133.0°	137.5°	0.117	0.135
2,4-Dimethyl-3-pentanol	3.552	138.8°	149.3°	0.0607	0.0704
2,2-Dimethyl-3-pentanol	3.481	136.0°	146.5°	0.0711	0.0843
3-Heptanol	3.807	156.8°	159.1°	0.041	0.0370
4-Heptanol	3.807	155.0°	159.1°	0.0409	0.0370
1-Octanol	4.414	195.2°	193.4°	4.51×10^{-3}	4.76×10^{-3}
2,2,3-Trimethyl-3-pentanol	3.811	152.5°	152.7°	0.0533	0.0500
2-Octanol	4.269	179.8°	176.8°	8.6×10^{-3}	1.15×10^{-2}
2-Ethylhexanol	4.345	184.6°	190.8°	6.76×10^{-3}	5.67×10^{-3}
1-Nonanol	4.914	213.1°	212.6°	1.0×10^{-3}	1.34×10^{-3}
2-Nonanol	4.769	198.5°	196.0°	1.8×10^{-3}	3.25×10^{-3}
3-Nonanol	4.807	194.7°	197.5°	2.2×10^{-3}	2.95×10^{-3}
4-Nonanol	4.807	193.0°	197.5°	2.6×10^{-3}	2.95×10^{-3}
5-Nonanol	4.807	195.1°	197.5°	3.2×10^{-3}	2.95×10^{-3}
2,6-Dimethyl-3-heptanol	4.517	178.0°	186.3°	3.1×10^{-3}	6.14×10^{-3}
3,5-Dimethyl-4-heptanol	4.628	187.0°	190.6°	5.0×10^{-3}	4.64×10^{-3}
1,1-Diethylpentanol	4.683	192.0°	186.2°	3.8×10^{-3}	5.47×10^{-3}
7-Methyloctanol	4.769	206.0°	207.1°	3.2×10^{-3}	1.94×10^{-3}
3,5,5-Trimethylhexanol	4.454	193.0°	195.0°	3.12×10^{-3}	4.30×10^{-3}
1-Decanol	5.414	—	—	2.0×10^{-4}	6.35×10^{-5}
1-Dodecanol	6.414	230.2°	231.9°	2.3×10^{-5}	3.04×10^{-5}
Butane	1.914	—	—	2.34×10^{-3}	1.74×10^{-3}
2-Methylpropane	1.732	—	—	2.83×10^{-3}	2.75×10^{-3}
1-Pentane	2.414	—	—	5.37×10^{-4}	4.91×10^{-4}
2-Methylbutane	2.269	—	—	6.61×10^{-4}	7.06×10^{-4}
3-Methylbutane	2.307	—	—	1.48×10^{-4}	1.82×10^{-4}
Neopentane	2.000	—	—	7.48×10^{-4}	1.40×10^{-3}
2,2-Dimethylbutane	2.561	—	—	2.14×10^{-4}	3.38×10^{-4}
2,4-Dimethylpentane	3.125	—	—	4.07×10^{-5}	8.13×10^{-5}
2,2,4-Trimethylpentane	3.417	—	—	7.48×10^{-5}	3.91×10^{-5}
2,2,5-Trimethylhexane	3.917	—	—	8.95×10^{-6}	1.10×10^{-5}
Cyclohexane ^d	3.000	—	—	6.61×10^{-4}	3.95×10^{-4}
Methylcyclohexane ^d	3.393	—	—	1.41×10^{-4}	1.46×10^{-4}
1,2-Dimethylcyclohexane ^d	3.804	—	—	5.38×10^{-5}	5.17×10^{-5}
Cycloheptane ^d	3.500	—	—	3.05×10^{-4}	1.12×10^{-4}
Cyclooctane ^d	4.000	—	—	7.05×10^{-5}	3.17×10^{-5}
Hexane	2.914	—	—	1.11×10^{-4}	1.39×10^{-4}
Heptane	3.414	—	—	2.93×10^{-5}	3.95×10^{-5}
Octane	3.914	—	—	5.79×10^{-6}	1.11×10^{-5}
Cyclopentanol	2.893	140.85°	124.0°	—	—
Cycloheptanol	3.893	185.0°	162.4°	—	—
1-Ethylcyclohexanol	4.269	166.0°	170.3°	—	—
2-Ethylcyclohexanol	4.341	181.0°	179.6°	—	—
1-Methylcyclohexanol	3.708	155.0°	148.7°	—	—
2-Methylcyclohexanol	3.803	165.0°	158.9°	—	—
3-Methylcyclohexanol	3.803	174.5°	158.9°	—	—
4-Methylcyclohexanol	3.803	173.5°	158.9°	—	—
1,3,5-Trimethylcyclohexanol	4.494	181.0°	178.9°	—	—
Ethanol	1.000	78.5°	78.3°	—	—

(continued)

Table I—Continued

Compound	χ	Tb_{obs}^a	Tb_{calc}^b	S_{obs}^a , molal	S_{calc}^c , molal
2-Propanol	1.732	82.4°	79.4°	—	—
1-Propanol	1.914	97.4°	97.5°	—	—

^a Solubility (S), expressed as molality, and boiling points (Tb), expressed in degrees centigrade, were taken from Refs. 5 and 6 and "Handbook of Chemistry and Physics," 51st ed., Chemical Rubber Co., Cleveland, Ohio, 1971, and the references cited therein. ^b Calculated boiling points are based on Eq. A3. ^c Calculated solubilities are based on Eq. A2. ^d For these cyclic compounds, the connectivity index is modified as described in Part I (4): a value of 0.500 is subtracted from the computed connectivity index for the solubility study only.

connectivity index with the water solubility of aliphatic alcohols and hydrocarbons. Furthermore, since boiling points tend to show a variation with structure similar to that of solubilities, the relationship between the connectivity index and the boiling point was also developed.

EXPERIMENTAL

Formulation of Connectivity Index—The connectivity index (χ) (4) is obtained from the molecular skeleton, which is also called the skeletal graph or, more simply, the graph. Hydrogen atoms are suppressed. The degree of connectivity, δ_i , is assigned to each vertex of the skeletal graph. The value of $\delta_i = 1, 2, 3,$ or 4 , the number of connections at the i th vertex. Then the connectivity index is the sum of connectivities, C_k , of each connection (or edge) of the graph between vertexes i and j :

$$\chi = \sum_{k=1}^N C_k = \sum_k \frac{1}{(\sqrt{\delta_i \delta_j})_k} \quad (\text{Eq. 1})$$

where N is the number of edges and χ is then simply a sum of terms, each of which is a connectivity contribution for each bond in the molecular skeleton. Illustrations of the calculation of the connectivity index were given in Part I (4). Only simple arithmetic is required: the summation of reciprocal square roots.

Correlation of Water Solubility with Connectivity Index—The correlation of the connectivity index with the water solubility of the hydrocarbons (5, 6) and alcohols (6) was determined. The regression equations, correlation coefficients, and standard errors are as follows (Table I).

For alcohols:

$$\ln S = 6.702 - 2.666\chi \quad (\text{Eq. 2})$$

$$r = 0.978 \quad s = 0.455 \quad n = 51$$

For hydrocarbons:

$$\ln S = -1.505 - 2.533\chi \quad (\text{Eq. 3})$$

$$r = 0.958 \quad s = 0.511 \quad n = 18$$

where S is solubility expressed as molality (see Appendix).

The correlations are of a general high quality (Fig. 1). It is useful to consider the predictive power of this relationship based on connectivity. Table II presents measured and predicted solubilities for three solid alcohols. The solubilities are for the supercooled liquid (6).

The obvious utility of the connectivity index in correlating and predicting solubility is the extraordinary simplicity of its calculation by simple arithmetic from the integers 1, 2, 3, and 4, representing the degrees of connectedness in the molecular skeleton.

Correlation of Boiling Point with Connectivity Index—The correlation between the connectivity index and the boiling points of two sets of alcohols was determined. The regression equations, correlation coefficients, and standard errors are as follows.

For 51 alcohols (from Ref. 6):

$$\text{bp} = 10.25 - 39.22\chi \quad (\text{Eq. 4})$$

$$r = 0.964 \quad s = 8.68 \quad n = 51$$

For 63 alcohols (from Ref. 6 plus the last 12 compounds in Table I):

$$\text{bp} = 11.60 - 39.13\chi \quad (\text{Eq. 5})$$

$$r = 0.970 \quad s = 9.35 \quad n = 63$$

It can be clearly seen from Fig. 2 that the use of the connectivity index is excellent both by virtue of its great simplicity and its superior correlation in the case of boiling points. Furthermore, this study included 10 cyclic alcohols in addition to straight-chain and variously branched alcohols.

An empirical improvement to the correlation may be obtained if the degree of connectivity of the hydroxyl group is allowed to vary and its value is determined by multiple regression analysis of alcohol solubility (see Appendix).

CONCLUSIONS

A simple connectivity index, easily calculated by arithmetic, based on the degree of connectedness at each vertex in a molecular graph which depicts the molecular skeleton, gives highly significant correlations with the water solubility of hydrocarbons and alcohols as well as with boiling points of alcohols. These correlations

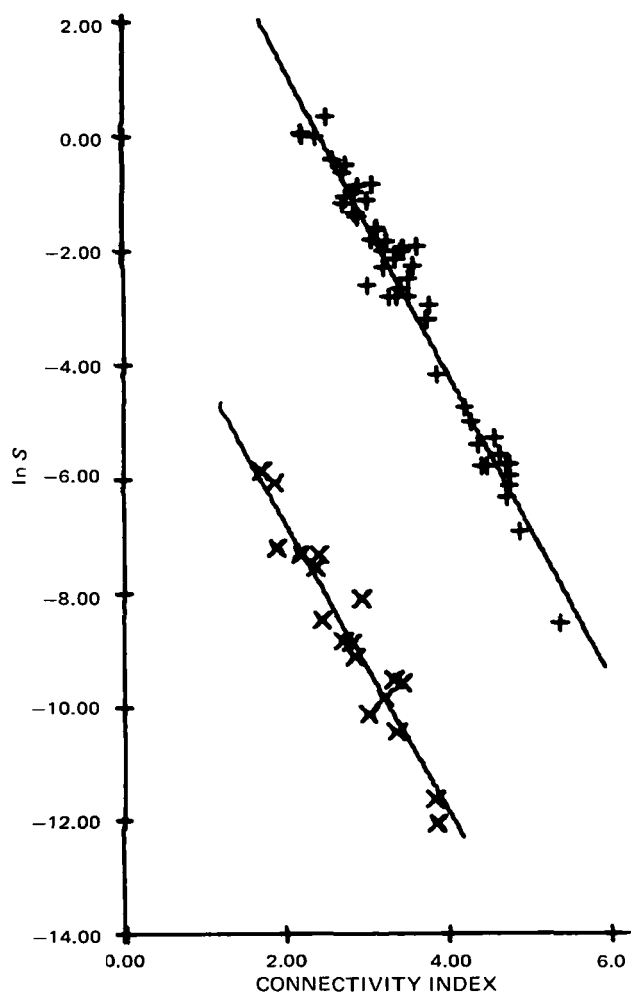


Figure 1—Computer-drawn plot of the natural logarithm of solubility versus the connectivity index for 51 alcohols (+) and 18 hydrocarbons (X) with the corresponding regression lines.

Table II—Observed and Predicted Solubilities for Three Solid Alcohols

Compound	S_{obs} , molal	S_{calc} , molal	S_{calc}^a , molal	χ	c_{OH}
1-Tetradecanol	2.84×10^{-6}	2.12×10^{-6}	2.43×10^{-6}	7.414	0.707
1-Pentadecanol	1.02×10^{-6}	0.56×10^{-6}	0.69×10^{-6}	7.914	0.707
1-Hexadecanol	4.55×10^{-7}	1.47×10^{-7}	1.94×10^{-7}	8.414	0.707

^a Based on Eq. A2.

are superior to those based on well-founded theory which relates these properties to the surface area of the cavity required to accommodate the solute in water.

The great appeal of the connectivity index lies both in the utter simplicity of its computation and in its fundamental relationship to molecular connectivity. At the heart of chemistry is the concept that an important characteristic of a molecule is the manner in which its substituent parts are connected. The evidence of the first two papers in this series strongly suggests that the connectivity index expresses the fundamental connectedness of organic molecules.

The ability to correlate water solubility with the connectivity index strongly suggests a possible correlation of the connectivity index with the partition coefficients studied by Hansch *et al.* (7). These correlations, as well as those with related biological activity, are presented in Part III.

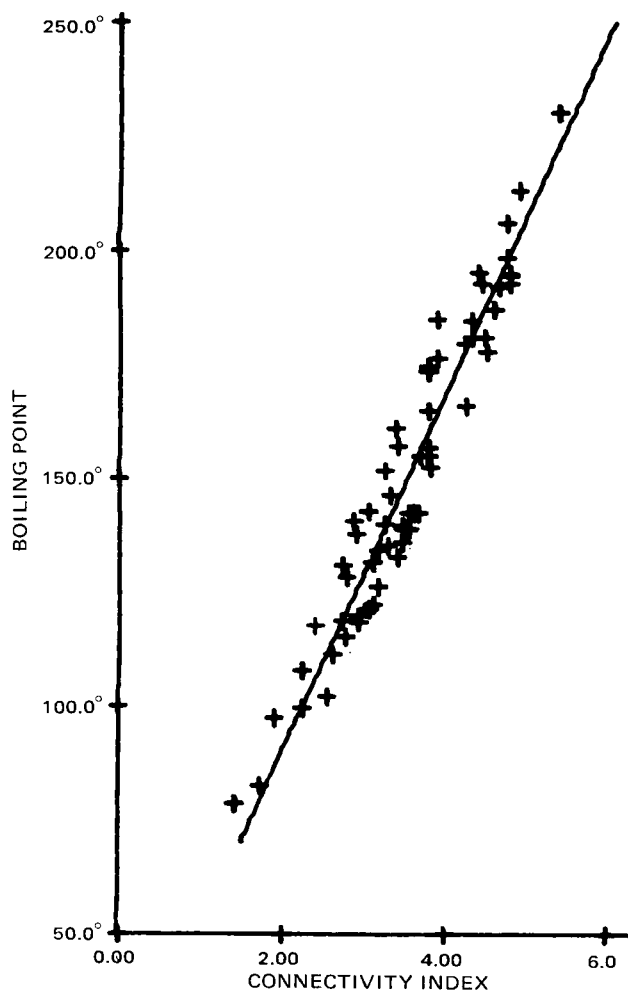


Figure 2—Computer-drawn plot of the boiling point versus the connectivity index for 63 alcohols with the corresponding regression line.

APPENDIX

The multiple regression equation in this additional study is:

$$\ln S = b_0 - b_1\chi - b_2c_{\text{OH}} - b_3Q \quad (\text{Eq. A1})$$

where b_i terms are regression coefficients, c_{OH} is the contribution of the C—OH bond graph edge to the connectivity index calculated with the OH degree of connectivity equal to 1, and Q has a value of 1 for an alcohol and of 0 for a hydrocarbon. Hence, solubility data for 51 alcohols and 18 hydrocarbons were merged for this multiple regression analysis.

The results of the analysis are:

$$\ln S = -1.516 - 2.528\chi - 3.961c_{\text{OH}} - 10.13Q \quad (\text{Eq. A2})$$

$$r = 0.994 \quad s = 0.383 \quad n = 69$$

As can be seen, this modification provides a significant improvement in correlation, with a 16% decrease in the standard error of the log of alcohol solubility from 0.455 to 0.383. The calculated solubilities (Table I) are based on the multiple regression equation (Eq. A2).

Based on this improved equation (Eq. A2), the predicted solubilities for the three solid alcohols mentioned earlier are given in Table II.

A similar approach to modification of the connectivity index was applied to the boiling points of the studied alcohols, and significant improvement was also achieved. The regression equation and associated parameters are:

$$\text{bp} = -36.23 - 38.41\chi - 2.216c_{\text{OH}} \quad (\text{Eq. A3})$$

$$r = 0.982 \quad s = 6.47 \quad n = 63$$

The calculated boiling points in Table I are based on Eq. A3. This and related modifications are currently under study and will be reported later.

REFERENCES

- (1) M. Randić, *J. Amer. Chem. Soc.*, in press.
- (2) H. Hosoyea, *Bull. Chem. Soc. Jap.*, **44**, 2332(1971).
- (3) E. A. Smolenski, *Russ. J. Phys. Chem.*, **38**, 700(1964).
- (4) L. B. Kier, L. H. Hall, W. J. Murray, and M. Randić, *J. Pharm. Sci.*, **64**, 1971(1975).
- (5) R. B. Hermann, *J. Phys. Chem.*, **76**, 2754(1972).
- (6) G. L. Amidon, S. H. Yalkowsky, and S. Leung, *J. Pharm. Sci.*, **63**, 1858(1974).
- (7) C. Hansch, J. E. Quinlan, and G. L. Lawrence, *J. Org. Chem.*, **33**, 347(1968).

ACKNOWLEDGMENTS AND ADDRESSES

Received January 30, 1975, from the Massachusetts College of Pharmacy, Boston, MA 02115

Accepted for publication March 28, 1975.

The authors acknowledge the assistance of the Eastern Nazarene College Computer Center. L. H. Hall is grateful for the support of Eastern Nazarene College during his sabbatical leave.

* On sabbatical leave from the Department of Chemistry, Eastern Nazarene College, Quincy, MA 02170

* To whom inquiries should be directed.