

## Estimation of thermodynamic and spatial properties of organic compounds for quantitative molecular design

A. J. Hopfinger<sup>(a,b)</sup>, R. A. Pearlstein<sup>(b)</sup>, M. Mabilia<sup>(b)</sup>, S. K. Tripathy<sup>(c)</sup>

Departments of Chemistry<sup>(a)</sup> and Medicinal Chemistry<sup>(b)</sup>, University of Illinois at Chicago, Box 6998, Chicago, IL 60680

and Department of Chemistry, University of Lowell, Lowell, MA 01854<sup>(c)</sup>

**Abstract** - The ability to be able to estimate relevant molecular properties of a chemical system is critical to the successful use of computer-aided molecular modeling and design (CAMMAD). Schemes are needed in applying fundamental computational tools to derive molecular properties characteristic of the chemical system under investigation. Three such schemes are presented here to quantitatively characterize molecular shape, estimate polymerization stability, and compute the lattice geometries of crystalline polymers. The molecular shape descriptors are shown to be useful in establishing a quantitative structure-activity relationship (QSAR) for enzyme inhibition. The measure of polymerization stability is shown to correlate with the enthalpy of polymerization for some vinyl polymers. The calculated lattice geometries for a series of polymers are in good agreement with observed crystal structures. These three examples of molecular property estimation demonstrate ways in which the methods of CAMMAD can be applied to complex molecular systems.

### INTRODUCTION

The ability to reliably estimate a wide variety of molecular properties is critical to successful applications of computer-aided molecular modeling and design (CAMMAD). Most of the applications to date have been in the pharmaceutical sciences with the goal of "drug design" (ref. 1). Correspondingly, the types of molecular properties computed have been those traditionally advanced as being responsible for illicit biological action (ref. 2).

CAMMAD is undergoing expanded use in the chemical sciences. Hence, it is appropriate to begin to develop capabilities to estimate properties not necessarily employed in drug design, and, in some cases, to expand the representation of molecular properties found useful in pharmaceutical research. The purpose of this paper is to discuss three of our recent efforts in molecular property estimation and representation. In particular, molecular shape properties, measures of polymer stability, and prediction of polymer crystal structures will be discussed.

### MOLECULAR SHAPE PROPERTIES

Molecular shape is a property which spans all applications in the chemical sciences. The ability of a molecule to "fit", as represented by molecular shape, is virtually a universal requirement for the desired macroscopic properties of the molecule.

Unfortunately, the term "molecular shape" is a catch-all term for a collection of molecular sub-features dependent upon the spatial locations of atoms composing the molecule. Our approach has been to devise quantitative representations of these sub-features. The emphasis on quantitation is due to the need of being able to meaningful compare shape to other molecular properties (ref. 3). That is, while molecular shape is often a critical component to the useful macroscopic properties of a molecule, it may not be the only important feature. Some way of weighing the relative contributions of the relevant molecular properties is needed in order to fully characterize the relationship between molecular and macroscopic properties of a molecule. Ideally, one would like to construct a quantitative structure-property relationship which might be considered a generalization of the quantitative structure-activity relationship (QSAR) in drug design (ref. 3).

Our most recent approach to the quantitative characterization of molecular shape has been to compute the potential energy fields about a molecule in some conformation (ref. 4). Different potential energy fields can be generated by using different probe species to evaluate field potential. While any probe can be used, those which are spherically symmetric offer the advantage of not having to consider the rotational degrees of freedom of the probe about its internal coordinate axes.

Some useful molecular shape subfeatures that can be determined from the potential field of an isolated molecule are:

1. The spatial extrema, maxima, minima, and saddle points, in the field potential.
2. The change in the spatial extrema of the field as a function of change in conformation of the isolated molecule.
3. The radial distribution function of integrated field potential using the center of mass of the isolated molecule as the origin.
4. The total spatially integrated potential field about the molecule.

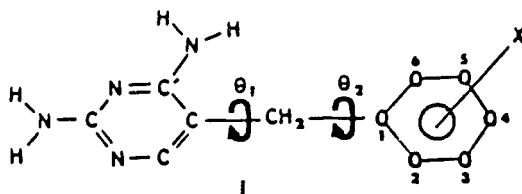
The first two sets of spatial subfeatures can be used to map out specific sites of attraction and repulsion for the molecule being studied with other compounds containing the probe species. The latter two sets of spatial subfeatures reflect the net interaction potentials of a molecule with different chemical units (the probes).

More often than not, one is interested in comparing the shapes of two or more molecules with the goal of understanding the differences in the macroscopic properties between/among the compounds.

This can be done using potential fields if one recognizes that molecular shape is a relative measure. One molecule in one conformation must be selected as the reference to which all other compounds are compared. The reference molecule provides a common base for making molecular shape comparisons.

Any molecule being studied can be superimposed upon the reference molecule and their difference in field potential at any point in space computed. Differences in field potential between the two molecules can be spatially integrated in one-, two- or three-dimensions. This permits a quantitative estimation as to how similar or different the shapes are of molecules as a function of field probe, conformation, and spatial superposition.

We have applied this formalism to a set of substituted 2,4-diamino-5-benzylpyrimidines (I) whose *in vitro* inhibition potencies of dihydrofolate reductase had been measured (ref. 4).



Molecular mechanics potentials were used to estimate the molecular potential energy fields. Hence, the molecular potential energy field,  $P_u(R, \theta, \phi)$ , at any point  $(R, \theta, \phi)$  for molecule  $u$  is given by eq. 1. In eq. 1  $a(T)_i$  and  $b(T)_i$  are the attractive and repulsive potential energy coefficients, respectively, of atom  $i$  of molecule  $u$  interacting with the test probe,  $T$ , treated as a single force center. The  $a(T)_i$  and  $b(T)_i$  can be taken from any of a number of different reported potential energy sets. Those of Hopfinger (ref. 5) were used in this work.  $Q_i$  and  $Q(T)$  are, respectively, the charge densities of the  $i$ th atom and the test probe.  $\epsilon(r_i)$  is the dielectric term. In the current study,  $\epsilon(r_i) = 3.5$ . Lastly,  $n$  is the number of atoms in  $u$ , and  $r_i$  is the distance between atom  $i$  and the test probe. The coordinate system is spherical and, by convention, located at the center of mass of the reference compound when pairs of fields are being compared.

$$P_u(R, \theta, \phi) = \sum_{i=1}^n \left[ \frac{a(T)_i}{r_i^6} + \frac{b(T)_i}{r_i^{12}} + \frac{Q_i Q(T)}{\epsilon(r_i) r_i} \right] \quad (1)$$

Figure 1 illustrates the geometry used to compute  $P_u(r, \theta, \phi)$  when  $u$  is the compound having  $X = H$ , and it is in the postulated "active" conformation (ref. 4). The test probe is  $CH_3$  which is treated as a sphere in eqn. (1). Three test probes,  $CH_3$ ,  $H^+$  and  $O^-$  were used in the study.

The total difference in field potential,  $\Delta P_{\alpha, \beta}$  for any pair of molecules  $\alpha$  (reference) and  $\beta$  can be expressed as:

$$\Delta P_{\alpha, \beta} = \int_R \int_{\theta} \int_{\phi} |P_{\alpha}(R, \theta, \phi) - P_{\beta}(R, \theta, \phi)| R^2 \sin \theta \, d\theta \, d\phi \, dR \quad (2)$$

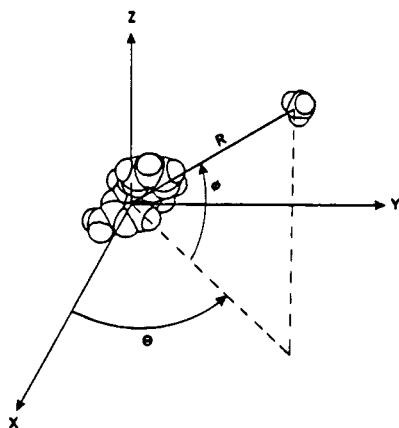


Fig. 1. The geometry used to compute  $P_U(R, \theta, \phi)$

TABLE 1. Measured enthalpies of polymerization,  $H_p$ , and the various calculated energies, all in kcal/mole

Polymer	$\Delta H_p$ (mea.)	$\Delta E_p$ Eq. (4)	$\Delta H_p$ (cal.) Eq. (5)
Poly(styrene)	-16.1 to -17.0	-5.8	-18.6
Poly(vinyl acetate)	-20.1 to -21.3	-11.21	-23.5
Poly(methyl acrylamide)	$-8.4 \pm 0.2$	5.54	-8.4
Poly(acrylamide)	$-13.8 \pm 0.3$	1.21	-12.3
Poly(4-vinyl pyridine)	$-18.7 \pm 0.3$	-5.23	-18.1
Poly(vinylpropionate)	$-20.6 \pm 0.2$	-9.56	-22.5
Poly(acrylonitrile)	$-17.2 \pm 0.5$	-8.33	-21.4
Poly(vinyl chloride)	-22.9 to -26.0	-13.70	-25.7
Poly(oxymethylene)	-12.17 to -17.14	-3.07	-16.2
Poly( $\alpha$ -methyl styrene)	-8.6	3.57	-10.2

Equation (2) was used to analyze the structure-activity data for 23 benzylpyrimidines. Each compound was allowed to be a candidate as the shape reference. The optimum multi-dimensional linear regression equation (QSAR) found is:

$$\log(1/C) = -2.34(0.67)[F] + 0.29(0.09)[F^2] + 0.37(0.05)[\pi_3 + \pi_4] + 9.39 \quad (3)$$

$N = 22 \quad R = 0.961 \quad S = 0.105$

where  $C$  is the concentration of compound necessary for 50% enzyme inhibition,  $F = \log(\Delta P_{\alpha, \beta})$ ,  $[\pi_3 + \pi_4]$  is the sum of lipophilicity fragment values (ref. 6) for the *meta* and *para* substituents,  $N$  the number of compounds analyzed (the reference compound must be deleted from the data set),  $R$  is the correlation coefficient, and  $S$  the standard deviation of fit. The reference compound for eq. (3) is  $X = 4\text{-NHCOCH}_3$  although most compounds in the data base could be used to generate significant QSARs. The probe for eq. (3) is  $H^+$ . The correlation coefficient of eq. (3) indicates about 92% of the variance in the data is accounted for by  $F$  and  $[\pi_3 + \pi_4]$ . This suggests that the shape of the molecule and the relative hydrophobicity of the *meta* and *para* substituents are responsible for inhibition potency.

## POLYMERIZATION STABILITY

Polymerization stability can be divided into three computational components. These components consist of estimating electronic energy, conformational energy, and molecular volume (ref. 7). Each of these three properties must be determined for both the reactant and the monomer so that the requisite differences in properties for the *pre* and *post* polymerized states can be calculated. Recently, we have developed a model to compute the internal energy of polymerization,  $\Delta E_p$ , and to show that, for a series of vinyl polymers, the work term (volume change) can be neglected (ref. 8). Hence,  $\Delta E_p$  directly correlates with the enthalpy of polymerization,  $\Delta H_p$ . A critical assumption made in this study, which appears valid, is that electronic energies computed from a semi-empirical molecular orbital method can be combined with conformational energies estimated using molecular mechanics. Our model predicts that  $\Delta E_p$  can be computed from,

$$\Delta E_p = \epsilon_3 - (\epsilon_2 + \epsilon_R) + E(1) + E(2) + 2(E(3) + E(4)) \quad (4)$$

where  $\epsilon_3$ ,  $\epsilon_2$  and  $\epsilon_R$  are the total electronic energies of a trimer, dimer, and reactant, respectively.  $E(i)$  is the conformational energy of the  $i$ th nearest-monomer neighbor interaction. The results of applying eq. (4) to a series of vinyl polymers are given in the third column of Table 1. The measured  $\Delta H_p$  are given in the second column of Table 1 (ref. 8). A linear regression fit between measured  $\Delta H_p$  and calculated  $\Delta E_p$  gives

$$\Delta H_p = 0.89 \Delta E_p - 13.39 \quad (5)$$

$$N = 10 \quad R = 0.96 \quad S = 1.30$$

The last column of Table 1 gives the  $\Delta H_p$  computed using eq. (5) and the calculated  $\Delta E_p$ .

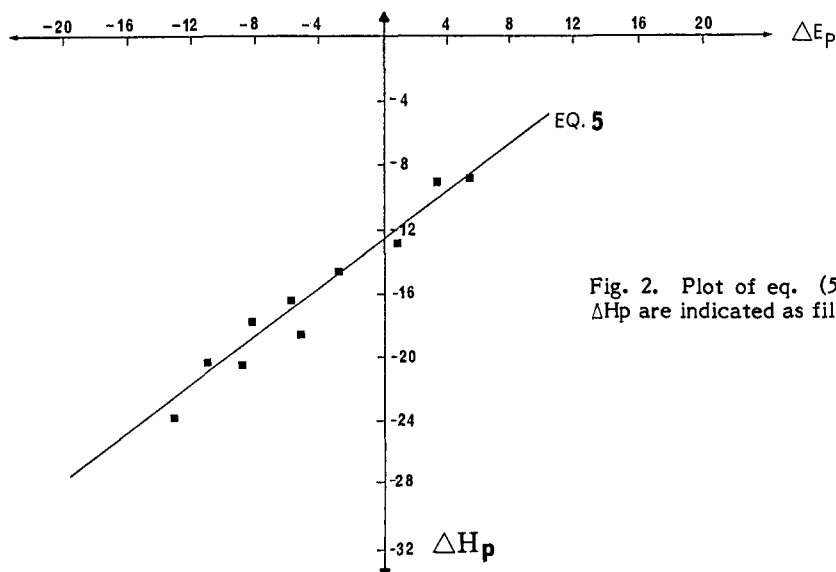
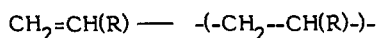


Fig. 2. Plot of eq. (5) and the measured  $\Delta H_p$  are indicated as filled squares.

The slope of eq. (5), 0.89, is sufficiently close to 1 ( $45^\circ$ ) to suggest that any errors and/or shortcomings inherent to this formalism in computing polymerization stability are relatively constant and uniform across the set of vinyl polymers investigated. On the other hand, the intercept in eq. (5) of -13.39 kcal/mole is both problematic and intriguing.

The intercept of -13.39 kcal/mole must be account for by some property common to all ten polymers. We believe that this common property is the breaking of the carbon-carbon double bond upon polymerization,



The change in electronic energy associated with this process is incorrectly computed using the MNDO molecular orbital method (ref. 9). Partial support of this ascertainment has been realized by computing  $\epsilon_3 - (\epsilon_2 + \epsilon_R)$  for poly(styrene) and poly( $\alpha$ -methylstyrene) using the CNDO method (ref. 10). The quantity  $\epsilon_3 - (\epsilon_2 + \epsilon_R)$  goes from 0.89 (MNDO) to -6.11 (CNDO) kcal/mole for poly(styrene), and 10.87 (MNDO) to -1.93 (CNDO) kcal/mole for poly( $\alpha$ -methylstyrene). Clearly, the value of  $\epsilon_3 - (\epsilon_2 + \epsilon_R)$  is method dependent. Nevertheless, MNDO is a reliable method for computing electronic structure contributions to differences in  $\Delta H_p$  between congeneric polymers. Moreover, it is only the difference in energy between states, and not the absolute energy of a state, that can currently be accurately estimated by computational methods.

## PREDICTING POLYMER CRYSTAL STRUCTURES

The arrangement of atoms in a polymer chain, having a conformation that can be described as a rational helix, is equivalent to a set of periodic row atoms parallel to the chain (helix) axis. The spacing along each row is determined by the molecular structure of the monomer repeat unit and the helix type. Thus, the crystal lattice can be considered as the collective set of superimposed row atom lattices. This construction permits the interaction of an infinite row of atoms with a reference atom to be represented in a series sum. This series sum, in turn, can be converted into another series that is rapidly convergent by expanding the original series in its Fourier representation. Thus, the time-consuming pairwise atomic interaction summing process can be reduced to the evaluation of the first few terms of a rapidly convergent series, satisfying desired truncation criteria. Consequently, the interchain lattice energy calculation is reduced to evaluating the associated interaction "series functionals" for the set of atoms of a single helical repeat unit with the unique set of atom rows comprising the lattice (ref. 11).

When pairwise atomic potentials are used to represent Lennard-Jones dispersive-repulsive and Coulomb-type electrostatic interactions only  $n = 1, 6,$  and  $12$  need to be considered in dealing with series sums of the type

$$S^n(r, \rho) = \sum_1 \frac{1}{[\rho^2 + (r - r_1)^2]^{n/2}} \quad (6)$$

TABLE 2. Polymer crystal structures from lattice energy calculations

Polymer	Calculated lattice (A)					Observed lattice				
	Symmetry	a	b	c	Helix	Symmetry	a	b	c	Helix
Polyethylene	Ortho	7.2	4.9	2.5	2/1	Ortho	7.42	4.95	2.55	2/1
	Mono	7.9	2.5	4.6	2/1	Mono	8.09	2.53	4.79	2/1
<u>it</u> -Polystyrene	Trig	21.6	-	6.7	3/1	Trig	21.90	-	6.65	3/1
<u>it</u> -Poly-p-fluorostyrene	Ortho	17.7	11.9	8.3	4/1	Ortho	17.6	12.1	8.25	4/1
Polyoxymethylene	Trig	4.4	-	17.5	9/5	Trig	4.47	-	17.39	9/5
Poly(ethylene oxide)	Mono	7.9	12.9	19.6	7/2	Mono	8.05	13.04	19.48	7/2
<u>it</u> -Poly(vinyl chloride)	Ortho	10.2	5.5	5.2	2/1	Ortho	10.6	5.4	5/1	2/1
<u>it</u> -Poly(vinyl fluoride)	Ortho	8.3	5.1	2.5	2/1	Ortho	8.57	4.95	2.52	2/1
<u>it</u> -Poly(vinyl methyl ether)	Trig	15.9	-	6.6	3/1	Trig	16.25	-	6.50	3/1
Poly(acetylene)	(ct) Ortho	7.9	4.5	4.5	2/1	(ct) Ortho	7.61	4.39	4.47	2/1
	(t) Ortho	4.2	7.6	2.5	2/1	(t) Ortho	4.08	7.41	2.48	2/1

In eq. (6)  $r_l$  is the lattice vector and  $\rho$  is the distance of the row from the reference atom.  $r$  is the projection of the field point (reference atom) onto the lattice. These sums ultimately take the forms

$$S^n(r, \rho) = \frac{1/2 (n/2 - 1/2)^{1-n}}{a\Gamma(n/2)} + \frac{2\pi^{n/2}}{a\Gamma(n/2)} \sum_h \left( \frac{|k_h|}{\rho} \right)^{1/2(n-1)} K_{1/2(n-1)}(2\pi\rho|k_h|) \exp[2\pi i k_h \cdot r] \quad (7)$$

for  $n = 6$  and  $12$ , and,

$$S^1(r, \rho) = (2/a) \sum_h K_0(2\pi\rho h/a) \exp(2\pi i h r/a) \quad (8)$$

for  $n = 1$ .  $K$  are Bessel functions of the third kind,  $a$  is the periodicity of the lattice and  $k_h$  the reciprocal lattice vector.

The lattice energies of a variety of linear polymers whose crystal structures are known (ref. 12) were minimized as a function of the lattice constants and packing symmetry. The calculated and observed  $a$ ,  $b$ , and  $c$  dimensions are reported in Table 2 along with the assumed symmetry of packing.

The observed and measured lattice dimensions are in good agreement for each polymer considered. However, in nearly every case the predicted lattice dimensions are less than the observed dimensions. This is likely due to the presence of structural and/or chemical defects in the actual polymer lattice as compared to an ideal lattice geometry used in the calculations. It is also possible that the nonbonded potential functions are "soft", that is, pair-wise interatomic contact distances are, effectively, slightly too short.

## SUMMARY

As the use of CAMMAD in the chemical sciences continues to grow it is becoming increasingly clear that computational methods and schemes must be devised to estimate molecular properties specific to the various subdisciplines of chemistry. CAMMAD needs to, in essence, learn to speak the language

(compute the molecular properties) of the various subdisciplines of chemistry. The fundamental computational tools as well as the hardware are available to achieve this goal. However, the methods, strategies and schemes required in applying basic computational tools to estimate molecular properties often still need to be determined. In this paper we report the estimation of three molecular properties - quantitation of molecular shape, polymerization stability, and crystalline polymer lattice structure. These examples may be useful in pointing the way to computing related properties.

## REFERENCES

1. A.J. Hopfinger, J. Med. Chem. **28**, 1133-1139 (1985).
2. R. Franke, Theoretical Drug Design Methods, Akademie-Verlag, Berlin (1984).
3. A.J. Hopfinger, J. Am. Chem. Soc. **102**, 7196-7206 (1980).
4. A.J. Hopfinger, J. Med. Chem. **26**, 990-996 (1983).
5. A.J. Hopfinger, Conformational Properties of Macromolecules, Academic Press, New York (1973).
6. C. Hansch and A. Leo, Substituent Constants for Correlation Analysis in Chemistry and Biology, Wiley-Interscience, New York (1979).
7. Steric and electronic effects have been recognized factors for some time; see F.S. Dainton and K.J. Ivin, Quarterly Reviews **12**, 61-82 (1958).
8. M. Mabilia, R.A. Pearlstein, M.G. Koehler, and A.J. Hopfinger, J. Macromol. Sci.-Phys. Ed., in press.
9. M.J.S. Dewar and W. Thiel, J. Am. Chem. Soc. **99**, 4899-4910 (1977).
10. J.A. Pople and D.C. Beveridge, Approximate Molecular Orbital Theory, McGraw-Hill, New York (1970).
11. S.K. Tripathy, A.J. Hopfinger, and P.L. Taylor, J. Phys. Chem., **85**, 1371-1380 (1981).
12. H. Tadokoro, Structure of Crystalline Polymers, Wiley-Interscience, New York (1979).