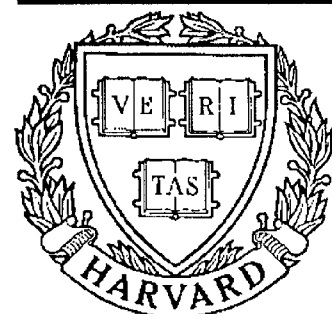


# TECHNICAL RESEARCH REPORT



S Y S T E M S  
R E S E A R C H  
C E N T E R



*Supported by the  
National Science Foundation  
Engineering Research Center  
Program (NSFD CD 8803012),  
Industry and the University*

## Symbolic Computing in the Prediction of Properties of Organic Compounds

*by M.L. Mavrouniotis*

## **SYMBOLIC COMPUTING IN THE PREDICTION OF PROPERTIES OF ORGANIC COMPOUNDS**

Michael L. Mavrovouniotis  
Chemical Engineering and Systems Research Center  
University of Maryland  
College Park, MD 20742

### **1. SUMMARY**

A completely new approach is under development for estimating properties of chemical compounds from their molecular structure. The approach is based on viewing each compound as a hybrid of several different molecular structures. The molecular structures are generated and manipulated by symbolic computation methods. The properties to be estimated include not only physical and thermodynamic properties, but also chemical properties of compounds which, currently, cannot be predicted by any method.

### **2. SIGNIFICANCE**

Knowing only the chemical structure of a compound, will allow one to use the proposed new method to estimate a variety of physical and chemical properties of the compound. Accurate values for the properties of compounds are extremely important in the design and modelling of any chemical processing system. The proposed method will allow prediction of certain properties which previously could not be estimated at all. These chemical properties include the stability and degradation of compounds with severe environmental impact. For example, the effect of chloro-fluoro-carbons (CFCs) on the ozone layer is determined by two factors: First, the stability of these compounds as they travel towards the ozone layer. Second, the activity of the free radicals which are created by the CFCs and destroy the ozone. Both of these properties could be predicted by the proposed approach, enabling the identification of the CFCs that are most harmful environmentally. Another application of the method would entail the prediction of cracking patterns, which determine the way in which certain fractions of oil can be converted into valuable fuels. This predictive ability will allow more efficient production of fuels and reduction of those byproducts that represent both an economic loss and an environmental burden.

### 3. BACKGROUND: GROUP-CONTRIBUTION METHODS

In order to design and model systems that process a particular set of substances, one needs the values of various physicochemical and thermodynamic properties of the substances. If the properties of interest have not been determined experimentally, one must estimate them from the known molecular structure of the compound. The molecular structure describes the atoms the make up a molecule of the compound, and the way these atoms are interconnected through bonds.

Several diverse methods for the estimation of physical properties of organic substances are based on the contributions of groups. In such methods, the molecular structure of the substance is broken down into substructures representing functional chemical groups; usually, a group consists of 1 to 5 atoms. Each property of the material is then estimated as a mathematical combination of the contributions of the groups and various correction factors. A group-contribution method is developed by utilizing experimental data for a large number of substances in a linear regression scheme to estimate the contributions of groups. In other words, data on a particular property for a number of substances are used to estimate the contributions; in turn, these contributions can be used in the prediction of the property for other substances.

Although traditional group-contribution methods are currently the most widely employed methods for the estimation of physical properties of organic substances, they are deficient in many respects:

- The applicability of each method is limited. Large and important classes of substances are often excluded; many methods exclude particular atoms (e.g., halogens, sulphur, phosphorous) or other classes of substances (e.g., esters, aromatics, heterocyclic compounds).
- Frequently, a compound cannot be properly described by a single molecular structure. Instead, the compound must be described as a *hybrid* of many molecular structures, called *conjugates*. Although this hybridization (also called *resonance*) among conjugates has a profound effect on a substance's properties, it is not taken into account by group-contribution methods.

- The properties of electrically charged molecules, free radicals, or other unstable intermediates of reactions, cannot be estimated.
- Only properties of the whole molecule can be estimated, i.e., no conclusions can be reached on the characteristics of parts of the molecule. Properties such as fractional electrical charges on atoms, bond strengths, etc., are essential for the prediction of chemical properties.

In order to reduce the number of parameters required, some methods resort to the use of a set of groups so small that important distinctions in functional groups are not retained (for example, some methods do not use special groups for aromatic carbons and have inherently high inaccuracy for all unsaturated cyclic compounds). Other methods allow a large number of groups and parameters, achieving good accuracy for substances very similar to those employed in the regression, but very poor accuracy for substances that deviate significantly.

The goal of the proposed research is the development of a new framework for the accurate prediction of diverse thermodynamic and chemical properties of a variety of substances from their molecular structure. The generation and analysis of conjugate forms lies in the foundation of the framework.

#### 4. APPROACH

The proposed approach, called ABC, is based on the contributions of Atoms and Bonds to the properties of Conjugates. The basic premise of the method is that properties can be estimated more accurately if we view each compound as a hybrid of a large number of conjugate molecular structures.

As discussed earlier, group-contribution methods generally do not take conjugation (resonance) into account. A few methods introduce corrections for some common strongly-conjugating hybrid structures (such as the benzene aromatic ring), but no method tries to take into account all conjugates systematically right from the start. The ABC approach makes conjugation an important integral part of property-estimation. *All compounds can be represented as hybrids of conjugates*; most conjugates may be unstable but they can nonetheless affect the properties of the hybrid.

In the proposed approach, resonance is viewed in a rather broad sense. Any single bond, for example, is viewed as a hybrid of three forms: (a)  $X—Y$ ; (b)  ${}^1X\cdots Y^{1-}$ ; (c)  ${}^1X\cdots Y^{1+}$ ; where the symbol  $\cdots$  denotes a deactivated single bond. Additional hybrids are formed between surrounding bonds and the negative and positive electrical charges formed by this bond. If atom Y is a strong electron-attractor and atom X is not (which in the context of a particular structure may depend on stabilization of the electrical charge by other conjugates), conjugate (b) will be more stable than conjugate (c). Consequently, in the hybrid of the form, Y will have a fractional negative charge, X will have a fractional positive charge, and the bond will be slightly weakened because of the contributions of forms (b) and (c) which do not have a single bond.

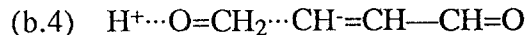
Constructing or dissociating *one bond at a time*, resonance (hybridization) can extend over several atoms and bonds and form a large number of conjugates. A few representative conjugates of the compound  $HO—CH_2—CH=CH—CH=O$  show how conjugates are formed from each other. Two successive conjugates in the list differ only in one bond:

First series of conjugates

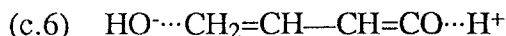
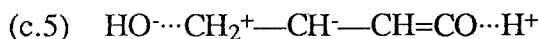
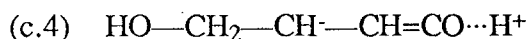
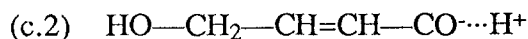
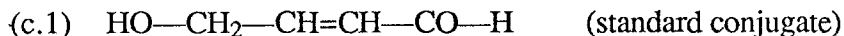
- (a.1)  $HO—CH_2—CH=CH—CH=O$  (standard conjugate)
- (a.2)  $HO—CH_2—CH=CH—CH^+—O^-$
- (a.3)  $HO—CH_2—CH^+—CH^-—CH^+—O^-$
- (a.4)  $HO—CH_2—CH^+—CH=CH—O^-$
- (a.5)  $HO^+\cdots CH_2^-—CH^+—CH=CH—O^-$
- (a.6)  $HO^+\cdots CH_2=CH—CH=CH—O^-$

Second series of conjugates

- (b.1)  $H—O—CH_2—CH=CH—CH=O$  (standard conjugate)
- (b.2)  $H—O—CH_2^+\cdots CH^-—CH=CH—CH=O$
- (b.3)  $H—O^+=CH_2\cdots CH^-—CH=CH—CH=O$



Third series of conjugates



The above conjugates are only a small sample; for example, of the six hydrogens only two were introduced in the conjugation. It should be emphasized that the linear paths implied in the three lists above would be intricately interconnected if all possible conjugates were present. As the examples demonstrate, the use of conjugates indirectly takes into account the fact that the orbital of any bond may be closer to one of its two atoms. The role of conjugates is even more important when there are interactions among groups that are separated by several bonds.

The stated basic premise of the method is that since these conjugates affect the structure of the compound, they should be explicitly taken into account during the estimation of properties. Even in the rather simple example of the conjugation of a single bond, it becomes apparent that the analysis of properties of conjugates provides information about molecular fragments, such as information on bond strengths and fractional electrical charges on atoms.

Because the generated conjugates contain electrically charged atoms, missing bonds, etc., a very large number of groups would normally be required for the estimation of the properties of conjugates. In the ABC framework, however, the properties of each conjugate are estimated from the *contributions of individual atoms and individual bonds*, rather than larger functional groups. One may doubt whether a small number of simple atoms and bonds can still capture the rich variety of conjugate molecular structures. It

should be remembered, however, that ultimately it is the properties of the hybrid that matter, and not the properties of the conjugates. If a particular interaction among atoms (leading to certain fractional electrical charges, some bonds weakened, other bonds strengthened, etc.) is neglected by one conjugate, it will, in all likelihood, be captured by another conjugate and will be ultimately reflected in the results for the hybrid. Thus, in ABC, the intramolecular interactions among groups are not captured through a rich variety of groups (which would require a large number of parameters), but through a rich variety of conjugates, whose properties can in fact be captured with just a small number of contributions from atoms and bonds.

To summarize the proposed framework, the properties of each conjugate must be estimated independently, through contributions of atoms and bonds; the properties of the hybrid must be subsequently derived through a mathematical combination of the properties of conjugates. This novel approach has not been ever investigated before.

While the use of contributions from both atoms and bonds in the estimation of properties can be carried out manually in a straightforward fashion, it is not possible to generate manually the vast set of all possible conjugates of any given molecular structure. It would also be quite difficult to develop and implement suitable algorithms within traditional computing environments and programming styles. With the advent of symbolic computation, the generation of conjugates becomes pragmatically feasible. Atoms, bonds, molecules, electron pairs, and other entities can be represented as interconnected objects within an Object-Oriented Programming (OOP) computation paradigm. The generation, comparison, and analysis of conjugates can be carried out through computer-based manipulation of the objects and their interconnections. One only needs to identify, formalize, and encode the operators which produce one conjugate from another, the canonical forms for the comparison of conjugates (to avoid duplicates), and the rules for eliminating excessively unstable conjugates.

## 5. APPLICATIONS

The proposed new method will permit the accurate estimation of a variety of physical properties of the compound. The proposed method will also allow prediction of *chemical* properties, such as sites for nucleophilic or electrophilic attack by other reagents, weak bonds that could be cracked at high temperatures, and stability of unstable reaction-intermediates (such as carbonium ions or free radicals) which affect the kinetics of chemical

reactions. Values for the properties of compounds are extremely important in the design and modelling of any chemical processing system.

One application of these chemical properties is the prediction of the effect of chloro-fluorocarbons (CFCs) on the ozone layer. The stability of these compounds as they travel through the atmosphere determines whether they will reach the ozone layer or will instead degrade like most other pollutants. Should a CFC compound reach the ozone layer, its destructive effect will depend on the nature and properties of the free radicals formed by the compound. The proposed approach could predict all of these properties of the CFCs and the free radicals, enabling the identification of the CFCs that are most harmful environmentally.

Another application of the ABC method would be the prediction of the ions a given hydrocarbons will form under high-temperature cracking. The method can accomplish this by estimating the strength of each bond of the hydrocarbon, since the weakest bonds are the ones likely to crack. Furthermore, the properties of the created ions, and the way they would react with the electrically neutral molecules, could also be predicted, because it depends on the distribution of partial electrical charges on different atoms of the compound. These predictions allow more efficient production of valuable fuels (from oil fractions of lesser value) and reduction of those byproducts that represent both an economic loss and an environmental burden.

## 6. RESEARCH PLAN

The initial focus is a proof-of-concept case study (already under way), involving the heat of formation of branched alkanes. Early results for selected isomeric alkanes show the method to be superior to any other method for the estimation of properties from molecular structure. For example, the ABC method has predicted the heat of formation of the five isomers of  $C_6H_{14}$  with an accuracy of  $\pm 0.1$  kJ/mol, while group-contribution methods are limited to an accuracy of  $\pm 4$  kJ/mol. The next step will involve the prediction of heat of formation for any organic compound. This will require the development and implementation of sophisticated symbolic computation algorithms for the generation of conjugate molecular structures, which can prune the least stable structures to reduce the complexity of the search. Having proven the feasibility of the method and having these computational tools, we will proceed to tackle the estimation of the Gibbs energy of formation, liquid molar volume, vapor pressure, other thermodynamic properties, and



chemical properties. Along the way, the method will be tested on the applications discussed in the previous section.

The work performed to date involved demonstration of the basic principles of the approach as applied to the comparison of the heat of formation of small sets of hydrocarbons. The next short-term steps, already under way, include:

- Detailed calculations aimed at a rigorous demonstration of the validity of the method, for one particular physical property, specifically the heat of formation, of a variety of compounds.
- Preparation of a journal paper presenting these results.
- Submission of a proposal to a funding agency, requesting substantial funding for the full development of the method over the next 4-5 years.