SRC TR 90-43

Estimation of Properties from Conjugate Forms of Molecular Structures: The ABC Approach

By

M.L. Mavrovouniotis

Estimation of Properties from Conjugate Forms of Molecular Structures: The ABC Approach

Michael L. Mavrovouniotis
Systems Research Center and Chemical Engineering
A.V.Williams Building (115), Room 2149
University of Maryland
College Park, MD 20742

Abstract

The ABC technique for estimating properties of organic compounds from their molecular structures is presented, based on the contributions of Atoms and Bonds to the properties of Conjugates. Conjugates are alternative formal arrangements of valence electrons; a real chemical compound is a hybrid of conjugates. In ABC all conjugates of a molecule are generated, and thermodynamic and quantum-mechanical properties are assigned to each conjugate, by summing contributions from atoms and bonds of the conjugate; the properties of the compound are then obtained as combinations of the properties of all the conjugates. In a simple application, ABC estimates the heat of formation of alkanes within 2.8kJ/mol, compared to 4.7kJ/mol for group contributions. The ultimate goal is the estimation of fractional charges on individual atoms of a compound and electron densities of bonds, because these are related to intermolecular interactions and chemical properties.

1. INTRODUCTION AND BACKGROUND

Property Estimation

In the modelling of a chemical engineering system, a variety of physical and chemical properties of pure compounds and mixtures, under specified conditions, must be used for accurate design and analysis of the system. All too often, it is not possible to find experimental values for the necessary parameters in the literature, neither is it practical or efficient to initiate the appropriate experiments to determine them as the need arises. While the need for accurate data on properties is increasing, the procedures of determining such data experimentally remain time-consuming and expensive [Reid *et al.*, 1987, Cox and Pilcher, 1970].

It is worth remembering that for every system for which detailed modelling and design is undertaken, several other systems were evaluated or designed quickly and approximately, only to be rejected based on the preliminary results. While for many instances of detailed system design it might be worthwhile to pursue experimental measurements, these can rarely be justified when several candidate systems are being roughly examined in order to select the most promising candidates for further study. The number of alternatives examined in this context could be virtually unlimited. For example, if one wishes to select a material that is most suitable for a very specific process (e.g., selecting a refrigerant or a solvent), the number of possible compounds and the mixtures they could form make experimental investigation especially difficult.

When the experimental determination of properties is impossible or impractical, the properties must be estimated from information on the molecular structures of the compounds involved.

It has been observed that in the physics of atoms, molecules, and solids, the fundamental laws have been determined (with the Schrödinger equation being a central fundamental principle), and we are exploring the deductions from them [Slater, 1955]. Although this is correct, the distance from the fundamental laws of mathematical physics to chemical engineering properties continues to be a major obstacle decades after these laws were fully established; relating the molecular structure of compounds to physical, thermodynamic, and chemical properties remains an extremely difficult problem.

In fact, the quantum mechanical principles underlying the structure and behavior of molecules have been so difficult to apply to property estimation that in many of the most commonly used estimation methods, these principles play a peripheral role, if any; for example, a variety of successful methods based on additive contributions of groups do not use these principles to any significant extent.

Thermodynamic principles, which cannot be used to estimate any single property directly from the molecular structure, have been quite successful in developing rigorous as well as empirical interrelationships among properties [Reid *et al.*, 1987]. In effect these principles reduce to a few basic properties the need for direct estimation from molecular structure — because other properties can be derived indirectly, from equations relating them to the basic properties.

Focusing solely on the derivation of properties from molecular structure, rather than interrelationships among macroscopic properties, let us briefly discuss and contrast quantum-mechanical investigations to group-contribution methods.

Quantum mechanical techniques

Quantum mechanical *ab initio* calculations are based on computing – or approximating analytically – the wavefunction from first principles, namely from the Schrödinger equation [Adamson, 1979], without resorting to any kind of regression with experimental data. Through quantum mechanical calculations, many properties of the compound as a whole, or individual atoms and bonds within the compound, can be derived from the wavefunction [Pauling and Wilson, 1963].

Solving the Schrödinger equation accurately is extremely difficult, even for fairly simple systems. The necessary procedures are extremely intensive in (super-)computer time; the complexity of the problem also necessitates tremendous manual effort for the creation of the necessary computer programs [Löwdin, 1989], which are usually specialized and tailored to particular compounds, systems, or features of the wavefunction [for example: Marconi, 1989, Ishida, 1989]. These intensive computations which are specially tailored to particular compounds can be likened to experimental measurements and are equally impractical for systematic and generalized estimation of properties.

Although it is possible to make drastic simplifying assumptions and resort to rough, approximate methods of solution, the results so obtained are suitable only for qualitative comparisons. For rough quantum mechanical approximations, discrepancies of the order of more than 20 kJ/mol between calculations and experimental data are common [Adamson, 1979]; such errors are clearly unacceptable for chemical engineering applications. Thus, from the engineering viewpoint, the quantum-mechanical estimation of properties of complex molecules remains beyond reach.

An alternative is presented by certain methods of non-quantum molecular mechanics which examine the details of the force field (instead of the wavefunction) around a molecular structure [for example: Allinger *et al.*, 1989, Lii and Allinger, 1989 - I and II] . This complex field is determined by the structure of the compound, and it involves a large number of parameters which are determined through a comparison with experimental data. While the use of experimental data permits accurate estimations, the computational complexity involved remains very high. The relationship between the force field and the more fundamental concepts of organic and quantum chemistry is very complex; as a consequence, it is difficult to verify the physical significance of the force field and its parameters, or to derive any qualitative predictions that would be independent of the actual values of the parameters.

Both force fields and wavefunctions entail computations that are essentially too detailed for most common physical and chemical properties in homogeneous systems. For most properties of interest in chemical engineering, the intricate shape of the wavefunction or the force field is not important; only quantification of some rough characteristics is necessary. Thus, after all the computational burden to derive a wavefunction, one ends up computing fairly simple properties, by integrating expressions involving the wavefunction over the whole space of coordinates, thus discarding most of the hard-earned information contained in the detailed shape of the original function.

It would be clearly an advantage if one could derive a coarse or discretized description of the system, such that the properties of interest can be estimating without resorting to the full detailed wavefunction or force field.

Group-Contribution Methods

Using much simpler procedures are several diverse methods for the estimation of physical properties of organic compounds, based on the contributions of groups [Joback and Reid, 1987, Benson, 1968, Domalski and Hearing, 1988, Mavrovouniotis *et al.*, 1988, Reid *et al.*, 1977, Reid *et al.*, 1987; the last reference includes several methods, for a variety of properties].

In such methods, the molecular structure of the substance is broken down into substructures representing functional chemical groups. Each property of the material is then estimated as a mathematical combination of the *contributions* of the groups (with a few corrections or additional factors, which refer to special sets of groups or to the molecular structure as a whole). A group-contribution method is developed by utilizing experimental data for a large number of compounds in a regression scheme to estimate the contributions of groups (and any other corrections utilized).

Because they can give quick estimates without requiring substantial computational resources, group-contribution methods are currently the most widely employed methods for the estimation of thermodynamic and physical properties of organic compounds (including properties of mixtures [Prausnitz *et al.*, 1986]). However, they are based on a very simplified view of chemical molecular structures, largely ignoring concepts of organic chemistry and quantum mechanics (for example, they cannot represent delocalized bonds and resonance). As a consequence, the scope of group-contribution methods is restricted by significant difficulties, which we discuss below.

The fundamental assumption of group additivity for ideal-gas properties is that groups do not interact with each other (except where group interactions are accounted for by special corrections), i.e., once the compound is decomposed into groups, the particular way in which groups are connected does not matter. In many cases, this assumption leads to isomers with identical estimated properties. For example, 2-methylpentane and 3-methylpentane are predicted to have identical properties in virtually all group-contribution methods; in reality, their heats of formation differ by 2.7 kJ/mol [Reid *et al.*, 1987].

Many methods restrict themselves to those atoms and groups that behave well with respect to group additivity. For example, compounds that contain phosphorous present more difficulties than compounds containing only carbon, hydrogen, oxygen, sulphur, and nitrogen. Phosphorous compounds are thus omitted by most methods [e.g., Joback and Reid, 1987].

Compounds that are affected significantly by resonance, i.e., they are hybrids of a number of conjugate forms, are not handled systematically by any method. Some methods do not address resonance at all, even for the case of aromatic compounds for which resonance is crucial [e.g., Joback and Reid, 1987]. Other methods introduce special corrections for some instances of resonance. For example, they use distinct groups for aromatic carbons or bonds [Benson, 1968, Benson *et al.*, 1969], or an additional contribution for conjugated dienes, in which two carbon-carbon double bounds are separated by one single bond.

There are no group contribution methods for the properties of charged molecules or other unstable intermediates of reactions. There are two reasons for the absence of such methods. First, the data on properties of such intermediates, necessary for the estimation of contributions, are quite sparse. Second, the properties of intermediates are affected strongly by conjugation. The ability of an ion to *disperse* the charge (which is formally placed on one atom) to several atoms is a major factor affecting the stability of the ion. The conjugation that leads to the delocalization of charge is not taken into account by group contributions. In the case of ions, if one attempts to use special corrections for conjugation, the number of parameters would be too high and there are not sufficient accurate data on ions to estimate these parameters.

Group-contribution methods estimate only properties of the whole molecule. They do not provide any information on the characteristics of parts of the molecule, such as fractional charges on atoms, or (σ- and π-) strengths of particular bonds. This prohibits any reasoning on chemical aspects of the molecule, such as the susceptibility to nucleophilic or electrophilic attack, for different parts of the molecule.

Finally, it is usually not possible to derive any significant *a priori* qualitative predictions that would not require the actual values of the contributions determined by the regression. In other words, there are no specific qualitative trends built into group-contribution methods.

Scope of the proposed approach

The ultimate goal of the research discussed in this paper is the development of an accurate method for the estimation of diverse physical and chemical properties of a variety of organic compounds from their molecular structure. The proposed new method rests on quantum-mechanical and organic chemical principles and estimates properties from the contributions of Atoms and Bonds to the properties of Conjugates (ABC). Sections 2, 3, and 4 of this paper present the foundation of the approach on conjugation of organic molecular structures; Section 5 provides a complementary, quantum-chemical view.

The ABC approach is applicable not only to stable compounds but also to unstable reaction intermediates (such as carbonium ions) and to properties of parts of a molecular structure (such as fractional charges on atoms and strengths of bonds). We have started with the estimation of heats of formation and plan to proceed with the estimation of fractional charges and bond densities; these localized properties are the key to other physical and chemical properties [Dewar, 1969, Fliszar, 1983].

2. CONJUGATION AND HYPERCONJUGATION

In organic chemistry, we often analyze and qualitatively compare the physical properties and stability of similar compounds, based on their molecular structures. We also reason about chemical properties, reaction selectivities, and relative stabilities of reaction intermediates. The analysis is based partly on basic principles and partly on empirical rules, the latter being often high-level abstracted manifestations of the former.

We examine some of these principles below, asserting that, although they are largely ignored by group contribution methods, they could in fact serve as the basis for quantitative property estimation.

Conjugation

Conjugation or *resonance* is a **delocalization of π-electrons** that changes the character of individual bonds, and affects the stability and properties of a compound. In conjugation, a compound is viewed as a *hybrid* of a number of *conjugates*. Each conjugate can be viewed as an arrangement of atoms which are connected, in pairs, by single, double, or triple bonds; each electron pair belongs to a specific bond that connects a specific pair of atoms; each charge present on a conjugate must be integer, because it is due to the deficiency or surplus of an integer number of electrons. The hybrid, however, cannot be represented in this way, because its electronic structure involves delocalized electrons, which bond different pairs of atoms in different conjugates. If we want to view the compound (hybrid) without referring to the original conjugates, we must represent its bonds as being *partially double* and partially single (more precisely, there are delocalized bonds distributed over more than two atoms); we may also need to represent fractional charges.

Group contribution methods do not take into account the conjugates, and do not view the compound as a hybrid (except in the special case of aromatic compounds). Instead, they work with a single form of the molecular structure — essentially the most prevalent of the conjugates. The basic premise of the ABC approach is that the properties of the hybrid compound can be estimated by examining the properties of each conjugate form of the compound. To explain this idea in the context of a simple example, consider the two isomeric dienes below:

$CH_3—CH=CH—CH=CH—CH_2—CH_3$ (2,4-heptadiene)

$CH_2=CH—CH_2—CH_2—CH=CH—CH_3$ (2,5-heptadiene)

Any of the carbon-carbon double bonds in these compounds can be depicted as having three conjugates:

$>C^{1+}—C^{1-}< \quad \leftrightarrow \quad >C=C< \quad \leftrightarrow \quad >C^{1-}—C^{1+}<$

These three conjugates account for the fact that the electrons of the π-bond have considerable mobility (compared to the electrons of the σ-bond). However, the proximity of the two double bonds in 2,4-heptadiene allows two additional conjugates, involving both double bonds:

$CH_3—CH^{1+}—CH=CH—CH^{1-}—CH_2—CH_3$ ↔ [I]

$CH_3—CH=CH—CH=CH—CH_2—CH_3$ ↔ [II]

$CH_3—CH^{1-}—CH=CH—CH^{1+}—CH_2—CH_3$ [III]

These additional conjugates make 2,4-heptadiene more stable than 2,5-heptadiene, by about 13 kJ/mol [Morrison and Boyd, 1973]. The conjugates also affect the properties of individual bonds. In this case, the $^3C—^4C$ bond has a partial double bond character, because it appears as a double bond in the forms [I] and [III]; naturally, its single bond character is more prevalent, because it occurs in the most stable structure [II]. Similarly the $^2C—^3C$ and $^4C—^5C$ bonds are weaker than a normal double bond [Morrison and Boyd, 1973].

What is of most importance in property estimation is that conjugation explains interactions among atoms that are separated by several bonds. In general, group contribution methods cannot account for such interactions in any systematic way, and can only introduce special corrections for a few common conjugating systems. On the other hand, most non-nearest neighbor interactions encountered in chemical structures do not involve delocalization of π-electrons. Thus, the significance of conjugation and any applications it may have in quantitative property-estimation, would appear to be limited to a few relatively unusual cases, such as conjugated dienes.

Hyperconjugation

While conjugation with π-electrons has the most pronounced effect, so-called *hyperconjugation* with σ-electrons is also possible [Dewar, 1962]. In the simplest case, even one single bond can be viewed as a hybrid of three forms — one purely covalent and two purely ionic conjugates:

$C^{1+} \cdots H^{1-}$ ↔ $C—H$ ↔ $C^{1-} \cdots H^{1+}$

The prevalent conjugate is the covalent one. However, the presence of the ionic conjugates implies that the single bond is polar and can be viewed as partly ionic. Consequently, in the hybrid of the forms, the carbon and the hydrogen have fractional charges. The sign of the charge on the hydrogen is negative if the first conjugate is more stable than the third; this is generally the case (barring the effect of other substituents on the carbon) if we assume that hydrogen is more electronegative than carbon. Although that is the currently accepted view [Fliszar, 1983], the definitions of electronegativity and partial charges are still under debate and some researchers actually assign a *positive* charge on the hydrogen of methane, i.e., they take hydrogen to be less electronegative than carbon [Mullay, 1986]. Regardless of the relative electronegativity of carbon and hydrogen, the presence of additional conjugates stabilizes the hybrid. The *overall bond* is *more stable* because of the contributions of the ionic forms; for the same reason, however, the *covalent character* of the bond is actually *weakened.*

Hyperconjugation can extent over several bonds and involve single and double bonds [Dewar, 1962]. To generate all the conjugates, we need only construct or dissociate *one bond at a time*, allowing conjugation to extend gradually over several atoms and bonds, as a large number of conjugates is formed. A few representative conjugates of the compound $HO—CH_2—CH=CH—CH=O$, listed in Table I, show how conjugates are formed from each other. Successive conjugates in each of the three series in the table differ only in one bond.

In this example, the presence and stability of conjugates are affected not only by the polarity of individual bonds but also by interactions among groups that are separated by several bonds. Note that Table I shows only a small sample of the compound's conjugates; for example, of the six hydrogens contained in this compound, only two were introduced in the conjugation. For the rest of this paper, we will often omit the prefix "hyper-" when referring to conjugates, with the understanding that conjugation can entail σ- as well as π- electrons.

Conjugation in empirical rules

Several empirical rules regarding the order of stability within a set of compounds can be explained by the number of conjugates or conjugates that participate in the stabilization of the compounds. We consider some of these empirical rules below.

(a) Branched alkanes are more stable than their straight-chain isomers [Morrison and Boyd, 1973]. For example, the alkanes with molecular formula C_5H_{12}, have the heats of formation shown below [Reid *et al.*, 1987]:

n-pentane: –146.5 kJ/mol

2-methylbutane: –154.6

2,2-dimethylpropane: –166.1

It is worth noting that steric interactions among the side chains destabilize the branched alkanes and would predict an order of stability that is the reverse of the observed order. Although steric hindrance does play a role, it is clearly counteracted by a stronger effect.

Table I. A few representative conjugates of the compound $\mathrm{HO{-}CH_2{-}CH{=}CH{-}CH{=}O}$; successive conjugates in each of the three series in the table differ only in one bond.

First series:	(a.1)	$\mathrm{HO{-}CH_2{-}CH{=}CH{-}CH{=}O}$	(dominant conjugate)
	(a.2)	$\mathrm{HO{-}CH_2{-}CH{=}CH{-}CH^{1+}{-}O^{1-}}$	
	(a.3)	$\mathrm{HO{-}CH_2{-}CH^{1+}{-}CH^{1-}{-}CH^{1+}{-}O^{1-}}$	
	(a.4)	$\mathrm{HO{-}CH_2{-}CH^{1+}{-}CH{=}CH{-}O^{1-}}$	
	(a.5)	$\mathrm{HO^{1+}\cdots CH_2{}^{1-}{-}CH^{1+}{-}CH{=}CH{-}O^{1-}}$	
	(a.6)	$\mathrm{HO^{1+}\cdots CH_2{=}CH{-}CH{=}CH{-}O^{1-}}$	
Second series:	(b.1)	$\mathrm{H{-}O{-}CH_2{-}CH{=}CH{-}CH{=}O}$	(dominant conjugate)
	(b.2)	$\mathrm{H{-}O{-}CH_2{}^{1+}\cdots CH^{1-}{=}CH{-}CH{=}O}$	
	(b.3)	$\mathrm{H{-}O^{1+}{=}CH_2\cdots CH^{1-}{=}CH{-}CH{=}O}$	
	(b.4)	$\mathrm{H^{1+}\cdots O{=}CH_2\cdots CH^{1-}{=}CH{-}CH{=}O}$	
Third series:	(c.1)	$\mathrm{HO{-}CH_2{-}CH{=}CH{-}CO{-}H}$	(dominant conjugate)
	(c.2)	$\mathrm{HO{-}CH_2{-}CH{=}CH{-}CO^{1-}\cdots H^{1+}}$	
	(c.3)	$\mathrm{HO{-}CH_2{-}CH^{1-}{-}CH^{1+}{-}CO^{1-}\cdots H^{1+}}$	
	(c.4)	$\mathrm{HO{-}CH_2{-}CH^{1-}{-}CH{=}CO\cdots H^{1+}}$	
	(c.5)	$\mathrm{HO^{1-}\cdots CH_2{}^{1+}{-}CH^{1-}{-}CH{=}CO\cdots H^{1+}}$	
	(c.6)	$\mathrm{HO^{1-}\cdots CH_2{=}CH{-}CH{=}CO\cdots H^{1+}}$	

Table II. Types conjugates relevant in the relative stability of isomeric hydrocarbons

$\mathrm{H{-}C{-}C{-}C}$	$\leftrightarrow$	$\mathrm{H^{1+}\ldots C = C\ldots C^{1-}}$	[I]
$\mathrm{H{-}C{-}C{-}C}$	$\leftrightarrow$	$\mathrm{H^{1-}\ldots C = C\ldots C^{1+}}$	[II]
$\mathrm{C{-}C{-}C{-}C}$	$\leftrightarrow$	$\mathrm{C^{1+}\ldots C = C\ldots C^{1-}}$	[III]
$\mathrm{H{-}C{-}C{-}H}$	$\leftrightarrow$	$\mathrm{H^{1+}\ldots C = C\ldots H^{1-}}$	[IV]

What are the important conjugates that could relate to the relative stability of these isomers? One would expect conjugates that involve only one single bond should play an important role in the stability of any individual compound. However, the number of carbon-carbon bonds and the number of carbon-hydrogen bonds do not differ among these isomers; thus, each compound has the same number of one-bond conjugates, and these cannot be used to discriminate among the isomers. The next set of candidates would involve three bonds and is shown in Table II.

If carbon and hydrogen have different electronegativities, the conjugates of type [I] or [II] (Table II) should be the most relevant. Specifically, if the carbon is more electronegative than hydrogen, then conjugate [I] which assigns a negative charge to the carbon is more stable. If carbon is less electronegative than hydrogen then conjugate [II] is more important. Although the latter is generally accepted to be the case [Fliszar, 1983], some methods assign the a positive charge on the hydrogen [Mullay, 1986]. We do not need to make a commitment here, because the number of conjugates of the two types is the same (for any given neutral compound) because both types involve the same chain of atoms. We call these HCCC-conjugates. We can count how many distinct conjugates of this type each of the three isomers has. Each distinct subchain of one hydrogen followed by three carbons, H-C-C-C, gives a distinct conjugate, even if the hydrogen or some of the carbons are also participating in other such subchains. We find 14 HCCC-conjugates for n-pentane, 20 for 2-methylbutane, and 36 for 2,2-dimethylpropane. In full accordance with the provided heats of formation, the higher number of conjugates has a stabilizing effect.

Although group contributions would provide the same order of stability, it is important to note that, with conjugation we could reach the correct qualitative conclusion without using quantitative parameters (i.e., before fitting with experimental data). Furthermore, we can easily point out an example in which group contributions would not be able to distinguish among the isomers. Specifically, consider the three isomeric methyl-heptanes [Reid *et al.*, 1987], for which the heats of formation and numbers of HCCC-conjugates are shown in Table III.

Even sophisticated methods [Benson, 1968] would predict the same heat of formation for the three methyl-heptanes. However, inspection of the number of HCCC-conjugates (Table III) shows that, in agreement with the data, 2-methylheptane is more stable and the other two are roughly equivalent. Note that 2-methylheptane has 2 conjugates more than 3-methylheptane and is roughly 3 kJ/mol more stable; 3-methylheptane has 4 conjugates more than n-octane and is, almost proportionately, 7 kJ/mol more stable.

Table III. Analysis of the effect of HCCC-conjugates on the heats of formation of three isomeric methyl-heptanes.

Compound	Heat of Formation [Reid *et al.*, 1987] kJ/mol	Number of HCCC-conjugates
2-methylheptane	–215.6	32
3-methylheptane	–212.8	30
4-methylheptane	–212.2	30
n-octane *(for comparison)*	–208.6	26

It should be mentioned here that counting the number of HCCC-conjugates takes into account the interactions or proximity effects among groups. Specifically, the number of HCCC-conjugates is higher if there are highly branched carbons connected to primary carbons; it tends to be lower when the branched carbons are close to each other. Naturally, it is lowest for carbon skeletons that are not very branched. Consider 2-methylheptane and 3-methylheptane (Table III) which consist of the same groups. In 2-methylheptane the tertiary carbon is connected to two primary carbons (and one secondary carbon) gives the compound 32 HCCC-conjugates; 3-methylheptane, on the other hand, has only 30 conjugates because its tertiary carbon is connected to only one primary carbon (and two secondary carbons). Thus, in this case, the number of conjugates can attributed to interactions among neighboring groups.

(b) Within a set of isomeric alkenes, the stability increases with the number of side chains attached on the double-bonded carbons [Morrison and Boyd, 1973]. Another common way in which this rule is phrased is that more substituted alkenes are more stable.

Consider the set of isomeric pentenes given below with their heats of formation [Reid *et al.*, 1987], and the number of side chains on double-bonded carbons, shown in Table IV. The empirical rule provides the correct order of stability. The only difficulty is that 2-methyl-1-butene cannot be distinguished from 2-pentene with this rule.

For the reasons mentioned in example (a), conjugates within each individual bond cannot be used to distinguish among isomers. Thus, we examine again conjugates that involve three bonds. The HCCC-conjugates discussed for alkanes are relevant for this example, but they take a secondary role because of another, more important class, which we call HCC=C-conjugates:

$$\mathrm{H—C—C=C} \quad \leftrightarrow \quad \mathrm{H^{1+}\ldots C=C—C^{1-}}$$

We know that a conjugate is more important if it is more similar to the dominant conjugate of the compound, structurally and energetically. An HCC=C-conjugate differs from the dominant conjugate in the separation of charges and one broken single bond. On the other hand in an HCCC-conjugate, there is a second broken single bond, and a double bond is created. Because two single bonds are more stable than one double bond [Mortimer, 1962], an HCC=C-conjugate is more stable, and thus exercises more influence, than an HCCC-conjugate of the same compound. Permitting a carbon to have two double bonds in HCC=C- conjugates, we obtain the number of conjugates of each kind for the compounds we are comparing here (Table V). The HCC=C-conjugates capture the same ordering as the original empirical rule, but with the aid of HCCC-conjugates we can also distinguish 2-methyl-1-butene from 2-pentene. However, the cis and trans versions of 2-pentene remain indistinguishable.

Table IV. Analysis of the effect of the number of side-chains on the heats of formation of isomeric pentenes.

ΔH_f [Reid *et al.*, 1987] kJ/mol	compound	number of side chains
–20.93	1-pentene	1
–28.09	2-pentene, cis	2
–31.78	2-pentene, trans	2
–36.34	2-methyl-1-butene	2
–42.58	2-methyl-2-butene	3

Table V. Analysis of the effect of the number of conjugates on the heats of formation of isomeric pentenes.

ΔH_f [Reid *et al.*, 1987] kJ/mol	compound	HCC=C-conjugates	HCCC-conjugates
–20.93	1-pentene	2	8
–28.09	2-pentene, cis	5	3
–31.78	2-pentene, trans	5	3
–36.34	2-methyl-1-butene	5	7
–42.58	2-methyl-2-butene	9	6

This example points out some difficulties. First, when we start dealing with more kinds of conjugates it becomes increasingly more difficult to draw qualitative conclusions; an automated, quantified approach would be of great value. Second, conjugation does not give us any information about steric effects, such as those arising in cis and trans isomers. While the ABC framework should resolve the first issue, the second is a limitation we must accept as inherent in this type of analysis. Naturally, one could introduce steric corrections as additional factors after the analysis of the electronic effects of conjugation.

Conjugation in unstable intermediates

In unstable intermediates, conjugation plays an even more important role, because the various conjugates tend to be closer energetically to the dominant conjugate. Consider carbonium ions, for which a qualitatively rule states that order of stability (from most stable to least stable) is:

tertiary ions > secondary ions > primary ions

In order to explain the rule we would consider conjugates that disperse the positive charge to other atoms:

$$\text{H—C—C}^{1+} \quad \leftrightarrow \quad \text{H}^{1+}\ldots\text{C} = \text{C}$$

$$\text{C—C—C}^{1+} \quad \leftrightarrow \quad \text{C}^{1+}\ldots\text{C} = \text{C}$$

The number of conjugates of this form (which we call HCC^+-conjugates and CCC^+-conjugates) is equal to the number of hydrogens and carbons (respectively) that are in α-position to the charged carbon. It is easy to see that this leads to the known order of stability. An example is shown in Table VI, where these charge-dispersing conjugates explain the order of stability of the carbonium ions. The table shows that charge-dispersing conjugates explain the order of stability of the carbonium ions.

We stated that conjugation is more important in such unstable intermediates than in stable compounds. This assertion is based on comparing the conjugates (such as $H^{1+}\ldots C = C$) to the dominant conjugate (H—C—C^{1+}). The energy difference between the two is due to two factors. First, movement of a charge (rather than the *separation* of charges we had in HCCC-conjugates for neutral compounds). Second, creation of a double bond replacing the two single bonds (rather than net loss of one bond which was the case in HCCC-conjugates). For each of these factors, the energy difference between the conjugates is drastically smaller in the HCC^+-conjugates for the ion than in the HCCC-conjugates. Hence, the conjugation has a greater effect on the stability of the ion.

Table VI. Analysis of the effect of the number of conjugates on the stability of isomeric carbonium ions.

formula	order	**HCC^+-conjugates**	**CCC^+-conjugates**
$CH_2^{1+}\text{—}CH_2\text{—}CH(CH_3)_2$	primary	2	1
$CH_3\text{—}CH^{1+}\text{—}CH(CH_3)_2$	secondary	4	2
$CH_3\text{—}CH_2\text{—}C^{1+}(CH_3)_2$	tertiary	8	1

3. CONJUGATION IN QUANTITATIVE PROPERTY-ESTIMATION

Conjugation is a manifestation of quite complex underlying quantum–mechanical phenomena. However, it is simple to use in a variety of qualitative comparisons. The basic tenet of the ABC property-estimation approach is that conjugation can be used in a quantitative framework: We can examine, separately, formal properties of the conjugates, and then combine them mathematically to estimate the properties of the (hybrid) compound.

We confine the discussion to a simple yet extremely useful thermodynamic property, the *ideal-gas heat of formation.* In the ideal gas state, there are no intermolecular interactions. We further assume that the electron clouds of the different bonds, and their location determine the energy of the system, and intramolecular interactions do not occur between atoms that are not bonded (i.e., molecules do not fold back to permit interactions among atoms that are, formally, several bonds away from each other).

As discussed earlier, group-contribution methods generally do not take conjugation (resonance) into account. A few methods introduce corrections for common strongly-conjugating hybrid structures (such as the benzene aromatic ring), but no method incorporates conjugation in a systematic way. The ABC approach proposed here makes conjugation central to property-estimation. In ABC, *all compounds are represented as hybrids of conjugates.*

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 capture the properties of a variety of molecular structures. It should be remembered that ultimately it is the properties of the hybrid that matter, and not the properties of the conjugates. Although a particular interaction among atoms (leading to certain fractional electrical charges, some bonds weakened, other bonds strengthened, etc.) may be neglected by one conjugate, it undoubtedly appears in another conjugate and is ultimately reflected in the results for the hybrid. Thus, in ABC, the intramolecular interactions among groups are not captured through a large variety of groups (which would require a large number of parameters), but through a variety of conjugates, whose properties can in fact be captured with just a small number of contributions from atoms and bonds. This is evident in the way we used conjugation to explain empirical rules on the stability of compounds in a previous section; we assumed that two conjugates are equivalent if they contain the same bonds, but *by using different types of conjugates* we effectively discriminated among the compounds.

The ABC framework entails fewer parameters than group contribution methods, and it is conceptually simple, in the sense that the generation of conjugates entails very simple and well-defined steps. However, ABC requires a large number of (simple) steps to carry out any given estimation and is computationally more intensive than group contribution techniques. Thus, ABC will require computational generation of the conjugates, except for very restricted classes of compounds. The quantitative example in a subsequent section illustrates these ideas.

To summarize the ABC 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 type of approach has not been investigated before.

The energy combination function

Let E_i be the energy (heat of formation) of conjugate i, obtained by summing contributions from atoms and bonds. Let F be the function we use to combine the heats of formation of the conjugates:

$$E = F(E_1, E_2, ..., E_{n-1}, E_n) \qquad [1]$$

While the function F cannot be fully determined from the qualitative use of conjugates in organic chemistry, certain properties of the function can be postulated. We mention here only a few basic properties. First, the hybrid must be more stable than each individual conjugate, and it must be stabilized as each conjugate is stabilized:

$$F(E_1, ..., E_n)<E_i \quad \text{and} \quad \partial F/\partial E_i \geq 0, \text{ for all i} \qquad [2]$$

Second, if a conjugate is extremely unstable then it should not affect the hybrid at all; if a conjugate is extremely stable it alone should determine the energy of the hybrid:

$$\partial F/\partial E_i \rightarrow 0, \text{ as } E_i \rightarrow +\infty \qquad [3]$$

$$\partial F/\partial E_i \rightarrow 1, \text{ as } E_i \rightarrow -\infty \qquad [4]$$

Third, the function should be invariant with respect to the choice of reference state. Changing the reference state adds a constant K to each conjugate. The hybrid should be shifted by the same constant K:

$$F(E_1-K, E_2-K, ..., E_{n-1}-K, E_n-K) = F(E_1, E_2, ..., E_{n-1}, E_n)-K \qquad [5]$$

It might at first appear that the statistical-mechanical Boltzmann averaging formula:

$$F_B(E_1, ..., E_n)=[\Sigma_i E_i \exp(-E_i/RT)]/\Sigma_i \exp(-E_i/RT) \qquad [6]$$

is a suitable combination function, where R is the ideal-gas constant and T the temperature. However, this function fails an important requirement we imposed here. Specifically, since it averages the E_is it will produce an overall energy E that is *not* lower than all the E_is. From the physical viewpoint, note that conjugates are *not* possible states of the compound, and the compound is *not* an equilibrium mixture of conjugates. The compound has an electronic arrangement which is an interpolation of the electronic arrangements of the conjugates, ***but more stable than each conjugate*** (which is precisely why there is hybridization). For example, if there are only two conjugates with equal energies, the compound should be more stable than either conjugate, while Equation [6] would incorrectly predict that the energy of the compound is the same as that of either conjugate. The fact that conjugation causes stabilization is apparent in the quantum-chemical treatment in Section 5.

One possible function that satisfies these requirements is:

$$E = -A \ln [\Sigma_i \exp(-E_i/A)] \qquad [7]$$

where A is any *positive* constant. ***For compactness, we use the notation*** Σ_i ***to denote a summation of terms; we similarly use*** Π_i ***to denote a product.***. The relationship can be also written in other forms:

$$\exp(-E/A) = \Sigma_i \exp(-E_i/A) \Leftrightarrow \Sigma_i \exp[(E-E_i)/A] = 1 \qquad [8]$$

$$E = E_1 - A \ln \{\Sigma_i \exp[(E_1-E_i)/A]\} \qquad [9]$$

This entirely empirical function yields satisfactory results for the simple studies which are discussed below, but it may be inadequate for more complex cases. We show in Section 5 that better means of combining energies of the conjugates can be developed through quantum mechanical analysis.

We mentioned in Section 2 that a conjugate is more important if it is more similar to the dominant conjugate of the compound, structurally and energetically. Only energetic similarity is actually used in the proposed combination function. The quantum mechanical analysis (Section 5), however, takes into account structural similarity.

An example: Heats of Formation of Alkanes

To demonstrate that it is indeed feasible to use conjugates in the quantitative estimation of properties, we develop here a simple scheme for estimating the heat of formation of alkanes.

The most important conjugates, apart from the dominant one for each compound, are those involving a single bond (CC-conjugates and HC-conjugates in our terminology) and the HCCC-conjugates defined in an earlier section. Ordinarily, we would take the heat of formation of each conjugate to be the summation of the contributions of its atoms and bonds (or charges on atoms). In this case, however, there are not enough independent types of conjugates to permit the estimation of contributions from the atoms and bonds. To lump these contributions, we consider the heat of formation of the dominant conjugate to be a linear function in n, the number of carbons:

$$\Delta H_f \text{(dominant conjugate)} = b\,n + d \qquad [10]$$

Remembering the molecular formula of alkanes, C_nH_{2n+2}, b represents one carbon-carbon bond, two carbon-hydrogen bonds, one carbon atom, and two hydrogen atoms; d represents two carbon-hydrogen bonds, and two hydrogen atoms. The heat of formation for each of the other types of conjugates differs from the dominant conjugate by a constant term independent of the number of carbons. For the HCCC-conjugate:

$$\Delta H_f \text{(HCCC-conjugate)} = b\,n + d + h \qquad [11]$$

where h represents the contributions of a carbon-carbon double bond, plus a negative charge on a carbon, plus a positive charge on a hydrogen, minus two carbon-carbon single bonds, minus one carbon-hydrogen single bond. As previously discussed, the energy of each conjugate depends only on the numbers of bonds and not on more detailed structural features; the different types of conjugates should nonetheless permit accurate estimation of the heat of formation of the compounds.

We can express similarly the heats of HC-conjugates and CC-conjugates (the number of conjugates of these types is, respectively, 2n+2 and 2n-2). For alkanes, however, these conjugates introduce algebraic and numerical degeneracies: First, the combination function we use (Equation [9]), renders the parameters of these two conjugates mathematically inseparable — they are algebraically interdependent. Second, the effect of the number of carbons, n, on the energy is exerted primarily through the linear

Table VII. Estimation of heats of formation of saturated hydrocarbons

Number and name of the compound in [Reid *et al.*, 1987]		Number of Carbons	Number of HCCC Conjugates	Heat of Formation [Reid *et al.*, 1987]	Error via ABC	Error via group contributions	Product of errors
168	ethane	2	0	–84.7	3.7	–5.7	–21.1
206	propane	3	6	–103.9	–2.2	–6.4	14.2
258	n-butane	4	10	–126.2	–1.4	–3.9	5.5
259	isobutane	4	18	–134.6	–1.4	1.1	–1.6
311	n-pentane	5	14	–146.5	–2.1	–3.5	7.3
312	2-methyl butane	5	20	–154.6	0.2	1.3	0.2
313	2,2-dimethylpropane	5	36	–166.1	–1.0	7.5	–7.1
379	n-hexane	6	18	–167.3	–1.9	–2.6	4.8
380	2-methyl pentane	6	24	–174.4	–0.1	1.5	–0.2
381	3-methyl pentane	6	22	–171.7	–1.1	–1.5	1.7
382	2,2-dimethyl butane	6	36	–185.7	2.1	7.2	14.8
383	2,3-dimethyl butane	6	28	–177.9	0.1	1.4	0.1
429	n-heptane	7	22	–187.9	–1.5	–1.8	2.8
430	2-methylhexane	7	28	–195.1	0.7	2.0	1.4
431	3-methylhexane	7	26	–192.4	–0.4	–0.7	0.3
432	2,2-dimethylpentane	7	40	–206.3	3.4	8.0	27.0
433	2,3-dimethylpentane	7	30	–199.4	3.5	3.0	10.5
434	2,4-dimethylpentane	7	34	–202.1	3.2	5.7	18.5
435	3,3-dimethylpentane	7	36	–201.7	1.5	5.3	7.7
436	3-ethylpentane	7	24	–189.8	–1.3	–3.3	4.4
437	2,2,3-trimethylbutane	7	42	–204.9	0.7	3.2	2.3
483	n-octane	8	26	–208.6	–0.8	–1.0	0.8
484	2-methylheptane	8	32	–215.6	1.6	2.7	4.2
485	3-methylheptane	8	30	–212.8	0.3	–0.1	0.0
486	4-methylheptane	8	30	–212.2	–0.3	–0.7	0.2
487	2,2-dimethylhexane	8	44	–224.9	2.9	6.7	19.4
488	2,3-dimethylhexane	8	34	–214.1	–1.4	–2.1	2.9
489	2,4-dimethylhexane	8	36	–219.6	2.8	3.4	9.2
490	2,5-dimethylhexane	8	38	–222.8	4.6	6.6	30.2
491	3,3-dimethylhexane	8	40	–220.3	0.8	2.1	1.7
492	3,4-dimethylhexane	8	32	–213.1	–0.9	–3.1	2.9
493	3-ethylhexane	8	28	–211	0.0	–1.9	0.0
494	2,2,3-trimethylpentane	8	44	–220.3	–1.7	–1.2	2.1
495	2,2,4-trimethylpentane	8	50	–224.3	–1.2	2.8	–3.4
496	2,3,3-trimethylpentane	8	42	–216.6	–4.2	–4.9	20.6
497	2,3,4-trimethylpentane	8	38	–217.6	–0.6	–2.0	1.2
498	2-methyl-3-ethylpentane	8	32	–211.3	–2.7	–4.9	13.4
499	3-methyl-3-ethylpentane	8	36	–215.1	–1.7	–3.1	5.4
529	n-nonane	9	30	–229.2	0.1	–0.3	0.0
530	2-methyloctane	9	30	–229.2	0.1	–3.6	–0.2
531	2,2-dimethylheptane	9	48	–247	6.0	8.9	53.8
532	2,2,3-trimethylhexane	9	48	–241.4	0.4	0.0	0.0
533	2,2,4-trimethylhexane	9	50	–243.4	1.3	2.0	2.6
534	2,2,5-trimethylhexane	9	54	–254.2	9.9	12.8	126.7
535	3,3-diethylpentane	9	36	–232.1	–1.3	–6.0	8.0
536	2,2,3,3-tetramethylpentane	9	54	–237.4	–6.9	–9.3	64.0
537	2,2,3,4-tetramethylpentane	9	47	–237.1	–3.3	–7.6	25.1
538	2,2,4,4-tetramethylpentane	9	66	–242.1	–8.2	–4.6	37.4
539	2,3,3,4-tetramethylpentane	9	42	–236.4	–1.0	–8.3	8.1
567	n-decane	10	34	–249.8	1.1	0.5	0.6
568	3,3,5-trimethylheptane	10	50	–258.7	0.0	–2.5	0.1

functions [10] and [11], and it is very difficult to discern, numerically, the secondary effect that n exerts through the number of HC- and CC-conjugates, especially in the presence of significant errors in our data. Because of these degeneracies, we must neglect the HC-conjugates and CC-conjugates.

If the number of HCCC-conjugates is N, the heat of formation of an alkane is given by Equation [9], into which we can substitute [10] and [11] to obtain:

$$E = b\,n + d - A \ln\left[1 + N \exp(-h/A)\right] \qquad [12]$$

where n and N are known for each compound, and b, d, h, and A are unknown constants, independent of the compound. We use the heats of formation of 51 alkanes with 2 to 10 carbons from [Reid *et al.*, 1987] to perform a regression (Table VII) and estimate these parameters. We obtain the values:

$$b = -16.6,\ d = -47.8,\ h = 146.2,\ A = 43.9 \ \text{(all in kJ/mol)} \qquad [13]$$

One should be cautious not to confuse the contributions of bonds with bond-dissociation energies. The contributions calculated here are devoid of delocalization and resonance effects because they refer to isolated conjugates. Bond dissociation energies on the other hand refer to real chemical species (including free radicals) and therefore contain resonance effects.

The standard deviation of the regression **±2.8 kJ/mol** (or about 0.67 kcal/mol); the average of the absolute values of the errors is 1.9 kJ/mol. The results of both fits for all the compounds are in Table VII. For comparison, we performed a group-contribution fit using four groups; one often uses a constant "origin" term in addition to the group contributions, but, in this case, such a term would not be linearly independent from the other contributions. The residuals for the group-contribution approach are also listed in Table VII. We obtained contributions (in kJ/mol): –45.21 for CH_3–; –19.86 for –CH_2–; 2.15 for >CH–; and 22.22 for >C<. In the group contribution fit, the standard deviation was ±4.7 kJ/mol and the mean of the absolute values of the errors 3.7 kJ/mol. Finally, to assess the correlation between the two fits, we list the products of the errors. The correlation factor between the errors (not the estimates) of the two methods was 0.9.

We consider the performance of the ABC approach excellent in this case: Using the same number of parameters (four), it achieved a standard deviation of 2.8 kJ/mol, compared to 4.7 kJ/mol for the group contribution approach. Several important factors should be kept in mind in evaluating the results. First, we essentially used a first-order ABC approach by focusing only on the most stable conjugates; once the technique is computerized, more conjugates are taken into account, improving the performance.

Second, we show in Section 5 that when ABC is established in a quantum-mechanical context, not only the energy but also the *overlap* of the conjugates must be taken into account; we did not take into account these quantum mechanical considerations in this example.

Third, the data probably contain random errors comparable to our estimation errors. Specifically, the standard deviation in heats of formation of alkanes from experimental measurements are 1 to 2 kJ/mol [Cox and Pilcher, 1970]. Furthermore, some data points appear especially suspect. For example, 2,2,5-trimethylhexane appears inconsistent with 2,2,4-trimethylhexane and 2,2,3-trimethylhexane; specifically, the difference (in the data) of 10.8 kJ/mol between 2,2,5-trimethylhexane and 2,2,4-trimethylhexane is not justified. The discrepancy led to the largest error for ABC (9.9 kJ/mol), as well as the group-contribution technique (12.8 kJ/mol). If we eliminate this data point from consideration, the standard deviation is reduced by 0.3 kJ/mol (to 2.5 kJ/mol for ABC and 4.4 kJ/mol for the group-contribution fit).

Throughout this section, we are comparing ABC to a group-contribution fit that uses the same number of parameters and the same data points. If we use contributions taken from the literature, the fit will deteriorate, because such contributions were adjusted to accommodate more diverse compounds and not just alkanes. In other respects, however, some group-contribution techniques lean towards unwarranted and hidden specialization. Notably, the methods [Benson, 1968] and [Joback and Reid, 1987] are strongly biased in favor of straight-chain alkanes and often yield very large errors (in excess of 20kJ/mol) for branched alkanes.

Estimation of other Properties

We claimed that the conjugates provide information on the character of bonds and fractional charges on atoms. How could such properties be estimated? We accept that these simple properties of the compounds lie *between* the properties of the conjugates, and they are closer to the properties of the most stable conjugates. For example, for most neutral organic compounds, the conjugate that contains no charge separations is the most stable, and the charges on most atoms are low. The effect of unstable conjugates that contain a positive charge on an atom is to shift the charge in the positive direction by a small amount; the more stable these charged conjugates become, the higher the charge of the atom is.

Let P_i be the property in question for conjugate i, and P the same property for the hybrid compound. A simple combination function with the correct qualitative behavior can be formulated, similar to Equations [7] to [9]:

$$P \exp(-E/A) = \Sigma_i P_i \exp(-E_i/A) \qquad [14]$$

This is just a simple weighting formula which uses $\exp(-E_i/A)$ as the weighting coefficients and normalizes the result by $\exp(-E/A)$. The equation can be solved for P to yield a form similar to [8]:

$$P = \Sigma_i P_i \exp[(E-E_i)/A] \qquad [15]$$

One can easily verify that this formula has attractive qualitative properties, similar to the ones of the energy combination function (Equations [2] to [9]).

It is worth noting that, for the estimation of partial charges and bond densities no additional parameters are required beyond those used in the calculation of the energy. The only new variables appearing in [15] are the P_is. These are known by simple inspection of the conjugates. If, for example, P is the fractional charge on an atom, then P_i is 0 in all the conjugates in which the atom is neutral, +1 in all conjugates in which the atoms has a positive charge, etc.

4. THE ROLE OF SYMBOLIC COMPUTATION

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 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. QUANTUM MECHANICAL VIEW

We provide here a quantum mechanical framework for property-estimation, essentially parallel to the empirical framework developed in the earlier sections. This brief quantum mechanical treatment shows a clearer physical basis of our property-estimation technique. Detailed analysis and examples will be presented in a future publication.

The electrons of a molecule are described by a wavefunction f in 3n coordinates where n is the number of particles (electrons and nuclei). The wavefunction is determined by the Schrödinger equation:

$$\mathbf{H}\, f = E\, f, \text{ where } \mathbf{H} = -\Sigma_i (h^2/8\pi^2 m_i)\nabla_i^2 + \Sigma_i \Sigma_j (Z_i Z_j / r_{ij}) \qquad [16]$$

with E the energy of the system, **H** the Hamiltonian operator, Z_i and m_i the charge and mass of particle i, and r_{ij} is the distance between particles i and j. The first term of **H** corresponds to kinetic energy and involves a summation over all particles, with ∇_i^2 the Laplacian with respect to the coordinates of particle i. The second term of the Hamiltonian represents the potential energy, a sum of pairwise interactions. By assuming that all non-valence electrons can be lumped with their respective nuclei and eliminating nucleus-to-nucleus interactions from the potential (Born-Oppenheimer approximation) we can reduce the number of variables and terms in the Hamiltonian. We then use the variation method which is based on the fact that, for the ground state, the wavefunction minimizes the energy:

$$E = \min_f [\, (\int f \mathbf{H} f\, dv) / (\int f f\, dv)\,] \qquad [17]$$

In the variation method, f is approximated as a linear combination of a set of basis wavefunctions:

$$f = \Sigma_i c_i f_i \qquad [18]$$

The variation method rests on picking a basis of f_is that are good approximations to f and then calculating c_is that minimize the energy. Defining $H_{ij}=\int f_i \mathbf{H} f_j \, dv$ and $S_{ij}=\int f_i f_j \, dv$, the optimization problem yields the c_is as the solution of a system of linear homogeneous equations:

$$\Sigma_i c_j(H_{ij} - S_{ij} E) = 0, \text{ for all } j \qquad [19]$$

For this system to have a non-trivial solution, the determinant of the coefficients must be zero. This yields the ground-state energy as the lowest solution of the determinantal *secular equation*:

$$\begin{vmatrix} H_{11}-S_{11}E & H_{12}-S_{12}E & \cdots\cdots & H_{1k}-S_{1k}E \\ H_{21}-S_{21}E & H_{22}-S_{22}E & \cdots\cdots & H_{2k}-S_{2k}E \\ \cdots\cdots & \cdots\cdots & \cdots\cdots & \cdots\cdots \\ H_{k1}-S_{k1}E & H_{k2}-S_{k2}E & \cdots & H_{kk}-S_{kk}E \end{vmatrix} \qquad [20]$$

We take each of the wavefunctions f_i to represent one conjugate form and define the energy E_{ii} for conjugate i as $E_{ii}=H_{ii}/S_{ii}$. Conjugates are perceived in organic chemistry to represent extreme situations which are physically unrealistic, but can form, taken together, a realistic hybrid. Much of the success of the variation method rests on adopting extreme wavefunctions f_i, which, in some sense, represent the bounds of the real overall wavefunction. The relation between conjugation and the variation method is quite apparent in the quantum-mechanical *valence-bond method* [Adamson, 1979]. From equation [17] we can show that the hybrid is more stable than each of the conjugates, i.e., $E<E_{ii}$.

The basis wavefunctions f_i are defined so that a bond that connects two atoms in a conjugate form is localized between the two atoms. Any delocalization present in the hybrid is due to disparity among several conjugates and not delocalization in any one conjugate. Furthermore, in our description of the conjugates, a bond between two atoms is a theoretical, *pure* covalent bond, devoid of any polarity or ionic character. The partially ionic character of real chemical bonds is due to ionic conjugates that exist for each bond. This choice of basis wavefunctions with extremely localized electrons leads us to decompose the wavefunction of each conjugate into distinct lobes, each corresponding to a pair of electrons. Let g_{ik} be the normalized ($\int g_{ik} g_{ik} dv = 1$) wavefunction for lobe k in conjugate i:

$$f_i = \Pi_k \, g_{ik} \qquad [21]$$

We now substitute into S_{ij}:

$$S_{ij} = \int \Pi_k (g_{ik} g_{jk}) \, dv = \Pi_k (\int g_{ik} g_{jk} dv) \qquad [22]$$

If a lobe k is in the same position in conjugates i and j, then $g_{ik}=g_{jk}$ and the respective integral in Equation [22] is equal to one. Thus, *the product in Equation [22] includes only those lobes in which conjugates i and j differ*. A simple corollary is that $S_{ii}=1$. The most valuable result of this analysis is that S_{ij} can be decomposed into a product of a small number of factors:

$$S_{ij} = \Pi_k s_{(ij)k}, \quad \text{with} \quad s_{(ij)k} = \int g_{ik} g_{jk} dv \qquad [23]$$

Each factor $s_{(ij)k}$ represents the transition of lobe i from its position in conjugate i to its position in conjugate j. The factor $s_{(ij)k}$ depends solely on the nature of the transition, and is otherwise independent of the nature of the compound, the two conjugates, the particular lobe involved, or the presence of other lobes and transitions. Since there are only a small number of distinct transitions, the decomposition introduced by Equation [23] drastically reduces the number of independent parameters that must be determined from data.

In order to decompose the integrals H_{ij}, we carry out a similar decomposition of the Hamiltonian into a summation of terms, so that there is precisely one term for each lobe. For conjugate i:

$$\mathbf{H} = \Sigma_k \mathbf{H}_{ik} \qquad [24]$$

From the definition of H_{ij} and Equation [21] we thus derive:

$$H_{ij} = S_{ij} E_{ij}, \quad E_{ij} = \Sigma_k e_{(ij)k}, \text{ and } e_{(ij)k} = \int g_{jk} \mathbf{H}_{jk} g_{jk} dv \; / \int g_{jk} g_{jk} dv \qquad [25]$$

where the interaction energy E_{ij} has been decomposed into a sum of terms $e_{(ij)k}$, each referring to the transition of a particular lobe. If the lobe is in the same position in the two conjugates (and the respective Hamiltonian term is the same), then $e_{(ij)k}$ is simply a fixed contribution, Δe_{ik}, of lobe k to the energy of conjugate i, dependent upon the nature of the lobe, but independent from other characteristics of the conjugate or other conjugates of the compound. In the special case i=j we obtain $E_{ii} = \Sigma_k \Delta e_{ik}$.

Let us summarize now how the energy of a whole compound is determined. First, we generate the conjugates of the compound. The secular equation [20], relating the conjugates to the energy of the compound, takes the form:

$$\begin{vmatrix} (E_{11}\text{-}E) & (E_{21}\text{-}E)S_{21} & \cdots\cdots & (E_{n1}\text{-}E)S_{n1} \\ (E_{21}\text{-}E)S_{21} & (E_{22}\text{-}E) & \cdots\cdots & (E_{n2}\text{-}E)S_{n2} \\ \cdots\cdots & \cdots\cdots & \cdots\cdots & \cdots\cdots \\ (E_{n1}\text{-}E)S_{n1} & (E_{n2}\text{-}E)S_{n2} & \cdots\cdots & (E_{nn}\text{-}E) \end{vmatrix} \qquad [26]$$

where $E_{ij}=E_{ji}$ and $S_{ij}=S_{ji}$, i.e., the matrix is symmetric. Each E_{ii} is the summation of contributions from the lobes of conjugate i (each lobe being either a bond or a negative charge on an atom). The contribution of a bond depends only on whether it is a σ or π bond and which atoms it connects. The contribution of a negative charge depends only on its atom. For example, the contributions of a carbon-hydrogen bond or a negative charge on a hydrogen are constants.

Each E_{ij} ($i \neq j$) is also a summation of contributions from lobes. For those lobes which do not change position (in other words for the part of a compound that, with respect to i and j, is not conjugating) the contributions are the same energy contributions encountered in E_{ii} (or E_{jj}). For lobes that are in a different position in conjugates i and j, the energy contribution depends on the nature of the transition. For example, if a pair of electrons is in a carbon-hydrogen bond for conjugate i, but moves to the hydrogen for conjugate j, there is a (constant) contribution to E_{ij} for this transition.

Each S_{ij} ($i \neq j$) is a product, with one factor for each lobe that is in a different position in the two conjugates. As with the energy contributions, the factor depends only on the nature of the transition and not on the rest of the conjugates or the rest of the structure of the compound. For lobes that are in the same position in the two conjugates the factor is omitted, because the normalization makes such factors equal to 1 (hence S_{ii} has already been set to 1 in the diagonal of the secular equation). The presence of the off-diagonal elements allows structural similarity among the conjugates to influence the energy of the compound, as mentioned in Section 2.

All the contributions of the lobes must be determined by fitting literature data for a set of compounds.

6. DISCUSSION

Systematic methods for the prediction of thermodynamic, physical, and chemical properties from the molecular structure of substances are essential for the modeling, analysis, and design of chemical processes. The objective of the work presented here is the development a new class of computer-based methods for the estimation of properties of organic compounds from their molecular structure. The ABC approach under development is based on the contributions of atoms and bonds in the properties of conjugate forms of a molecular structure. Conjugate forms, which are alternative formal arrangements of valence electrons in a molecule, have been largely ignored in property estimation for chemical engineering purposes.

In ABC, we start by generating all conjugate forms of the molecule whose properties we wish to estimate. Thermodynamic properties (or, in Section 5, quantum-mechanical properties) are assigned to each conjugate, simply by summing contributions from atoms and bonds in the particular electronic arrangement of the conjugate. In Section 5 we discuss how, for higher accuracy, pairwise quantum-mechanical relationships among conjugates are also evaluated through additive contributions. These properties of the conjugates are combined to derive the properties of the compound.

We presented in this paper the basic principles of the approach. We showed that several qualitative empirical rules governing the stability of organic compounds can be explained by conjugation. Quantifying the analysis, we developed an empirical energy combination function with the appropriate intuitive properties. Using this function, we obtained a standard deviation of ±2.8 kJ/mol for the heats of formation of alkanes, despite apparent significant errors (1 to 2 kJ/mol) in the data.

In Section 5, we provide a deeper foundation for our approach, by using the variation method of quantum mechanics and developing a preliminary scheme for calculating the energy of a compound from the energies of the conjugates, including the energies and overlaps of pairwise interactions among conjugates; we show that these quantities can be determined as summations or products of a small number of parameters describing the energies and overlaps of bonds and atomic charges.

Although our initial focus entails simple thermodynamic properties, such as the heat of formation, the fundamental basis of the approach should permit the estimation of other physical properties, fractional charges on individual atoms of the compound, and electron densities and strengths of individual bonds (either empirically or quantum-mechanically). These properties of molecular fragments will lead in turn to quantitative prediction of intermolecular interactions and chemical properties of compounds. Thus, ABC may ultimately allow accurate and general estimation of a wide variety of properties, and thus more efficient analysis and design of products and processes.

The method must be computer-based because the generation and analysis of a fairly large number of conjugates cannot be done manually. However, it does not entail the kind massive computations encountered in *ab initio* quantum mechanical analysis, because the application of ABC does not entail integration of wavefunctions. The integrals are essentially lumped into additive contributions (of bonds and atoms) which can be estimated by data regression. One non-trivial computational aspect of the method is the generation of the conjugates, which requires flexible symbolic representation and rearrangement of molecular structures. With the widespread availability of powerful symbolic computing environments, however, the requisite manipulation of molecules is greatly facilitated, and the portability of the method among different computer systems is ensured.

LITERATURE CITED

Adamson, A. W. *Physical Chemistry; 2nd edition;* Academic Press: New York, 1979.

Allinger, N. L.; Yuh, Y. H.; Lii, J.-H. "Molecular Mechanics. The MM3 Force Field for Hydrocarbons. 1". *J. of the Amer. Chem. Soc.* **1989**, 111, 8551-8566.

Benson, S. W. *Thermochemical Kinetics;* Wiley: New York, 1968.

Cox, J. D.; Pilcher, G. *Thermochemistry of Organic and Organometallic Compounds.* Academic Press: New York, 1970.

Dewar, M. J. S. *Hyperconjugation;* The Ronald Press Company: New York, 1962.

Dewar, M. J. S. *The Molecular Orbital Theory of Organic Chemistry;* McGraw-Hill: New York, 1969.

Domalski, E. S.; Hearing, E. D. "Estimation of the Thermodynamic Properties of Hydrocarbons at 298.15 K". *J. of Phys. and Chem. Ref. Data* **1988**, 14, 1637-1678.

Fliszar, S. *Charge Distributions and Chemical Effects*; Springer-Verlag: New York, 1983.

Ishida, K. "Charge Density by the Use of the Hiller-Sucher-Feinberg Identity. Perturbative Corrections and Configuration Interaction Effects". *Int. J. of Quantum Chemistry* **1989**, 36, 213-223.

Joback, K. G.; Reid, R. C. "Estimation of Pure-Component Properties from Group Contributions". *Chem. Eng. Comm.* **1987**, 57, 233-243.

Lii, J.-H.; Allinger, N. L. "Molecular Mechanics. The MM3 Force Field for Hydrocarbons. 2. Vibrational Frequencies and Thermodynamics". *J. of the Amer. Chem. Soc.* **1989 -I**, 111, 8566-8575.

Lii, J.-H.; Allinger, N. L. "Molecular Mechanics. The MM3 Force Field for Hydrocarbons. 3. The van der Waals Potentials and Crystal Data for Aliphatic and Aromatic Hydrocarbons". *J. of the Amer. Chem. Soc.* **1989 -II**, 111, 8576-8582.

Löwdin, P.-O. "On the Present Situation in Quantum Chemistry". *Int. J. of Quantum Chemistry* **1989**, 36, 191-199.

Marconi, G. "Quantum Mechanical Calculations on Vibronic Activity in the MCD Spectra of Carbonyl Compounds". *Int. J. of Quantum Chemistry* **1989**, 36, 417-425.

Mavrovouniotis, M. L.; Bayol, P.; Lam, T.-K. M.; Stephanopoulos, G.; Stephanopoulos, G. "A Group-Contribution Method for the Estimation of Equilibrium Constants of Biochemical Reactions". *Biotechnology Techniques* **1988**, 2, 23-28.

Morrison, R. T.; Boyd, R. N. *Organic Chemistry; 3rd edition;* Allyn and Bacon: Boston, 1973.

Mortimer, C. T. *Reaction Heats and Bond Strengths;* Pergamon Press: New York, 1962.

Mullay, J. "A Simple Method for Calculating Atomic Charges in Molecules". *J. Am. Chem. Soc.* **1986**, 108, 1770-1775.

Pauling, L.; Wilson, E. B., Jr. *Introduction to Quantum Mechanics with Applications to Chemistry;* Dover: New York, 1963.

Prausnitz, J.M.; Lictenthaler, R.N.; Azevedo, E.G. *Molecular Thermodynamics of Fluid-Phase Equilibria; 2nd edition;* Prentice-Hall: Englewood Cliffs, N.J., 1986.

Reid, R. C.; Prausnitz, J. M.; Poling, B. E. *The Properties of Gases and Liquids*; *4th edition;* McGraw-Hill: New York, 1987.

Reid, R. C.; Prausnitz, J. M.; Sherwood, T. K. *The Properties of Gases and Liquids*; *3rd edition;* McGraw-Hill: New York, 1977.

Slater, J. C. *Modern Physics;* McGraw-Hill: New York, 1955.

ACKNOWLEDGEMENTS

This research was supported in part by the Minta Martin Research Fund, the General Research Board, and the Engineering Research Center of the University of Maryland; and by the Systems Research Center funded by the National Science Foundation (NSFD CDR 8803012).