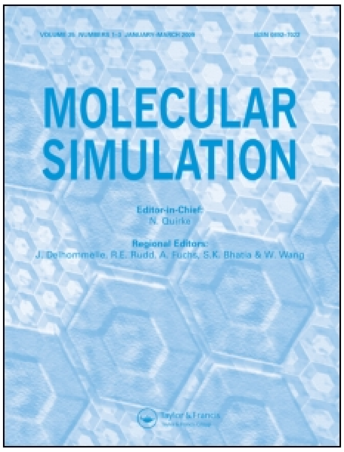

Molecular Simulation
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713644482

Computational Identification of Conjugate Paths for Estimation of Properties of Organic Compounds

Suzanne E. Prickett[a]; Leonidas Constantinou[a]; Michael L. Mavrovouniotis[a]
[a] Department of Chemical Engineering and Institute for Systems Research University of Maryland College Park, MD

To cite this Article Prickett, Suzanne E. , Constantinou, Leonidas and Mavrovouniotis, Michael L.(1993) 'Computational Identification of Conjugate Paths for Estimation of Properties of Organic Compounds', Molecular Simulation, 11: 2, 205 — 228

To link to this Article: DOI: 10.1080/08927029308022509

URL: http://dx.doi.org/10.1080/08927029308022509

Molecular Simulation, 1993, Vol. 11(2-4), pp. 205-228
Reprints available directly from the publisher
Photocopying permitted by license only

COMPUTATIONAL IDENTIFICATION OF CONJUGATE PATHS FOR ESTIMATION OF PROPERTIES OF ORGANIC COMPOUNDS

SUZANNE E. PRICKETT, LEONIDAS CONSTANTINOU, and MICHAEL L. MAVROVOUNIOTIS*

Department of Chemical Engineering and Institute for Systems Research
University of Maryland College Park, MD 20742

(Received December 1992, accepted March 1993)

We have developed and implemented a computer-based method to generate and enumerate all conjugate forms of acyclic organic compounds. These algorithms are used in a new approach for the estimation of thermodynamic and physical properties of acyclic organic compounds from their molecular structure. The approach is based on the contributions of *A*toms and *B*onds in the properties of *C*onjugate forms (ABC) of a compound and has produced more accurate results than group-contribution methods. Generating all conjugate forms of the molecule whose properties we wish to estimate is necessary in the application of this technique. The use of symbolic computing environments allows the flexible representation and manipulation of molecular structures. Atoms, bonds, molecules, and other entities are represented as interconnected objects. The generation, comparison, and analysis of conjugates are carried out through computer-based manipulation of the objects and their interconnections.

KEY WORDS: Property estimation, conjugate form, molecular structure, object-oriented programming, conjugation

I THE CONCEPT OF CONJUGATION

Physical and thermodynamic properties of organic compounds are necessary in the design of products and processes. When experimental data is unavailable, these properties must be estimated. Often, properties are estimated from additive group contributions [e.g., 1, 2, 3, and 4]. A new approach is based on the contributions of Atoms and Bonds in the properties of Conjugate forms (ABC) of molecular structures [5]. A conjugate form is an alternative formal arrangement of electrons in a molecule. In ABC, all compounds are viewed as hybrids of conjugates. To estimate the properties of a compound, one must determine and combine formal properties of its conjugate forms. The properties of each conjugate are estimated from contributions of individual atoms and bonds.

A conjugate of a molecular structure is an idealized arrangement of atoms connected by bonds. Each electron pair belongs either to a specific atom or to a specific bond, which connects two atoms. Therefore the order of all bonds must be an integer and any charge on an atom must also be an integer. A real compound is viewed as a hybrid of all its conjugates. A hybrid compound has bonds of fractional

---

[*] To whom all correspondence should be addressed.

order and atoms with fractional charges. The dominant conjugate of a compound is defined as the arrangement of electron pairs that results in the maximum number of bonds. In general, this is the standard written representation of a molecule. A recessive conjugate is any other arrangement of electron pairs.

It has long been recognized that certain properties of a compound are affected by conjugation. Consider the two isomers:

$$\mathrm{CH_3{-}CH{=}CH{-}CH{=}CH{-}CH_2{-}CH_3} \quad \text{(2,4-heptadiene)}$$

$$\mathrm{CH_3{-}CH{=}CH{-}CH_2{-}CH{=}CH{-}CH_3} \quad \text{(2,5-heptadiene)}$$

The first compound gives rise to conjugates that depend on the proximity of its double bonds.

$$\mathrm{CH_3{-}CH^{1+}{-}CH{=}CH{-}CH^{1-}{-}CH_2{-}CH_3}$$

$$\mathrm{CH_3{-}CH^{1-}{-}CH{=}CH{-}CH^{1+}{-}CH_2{-}CH_3}.$$

These additional conjugates result in 2,4-heptadiene having a lower enthalpy than 2,5-heptadiene by about 13 kJ/mol [6].

In the ensuing discussion, the term "energy" will be used for the enthalpy of formation. The energy of a compound, E, is determined by combining the energy of each conjugate. A suitable function is

$$E = -A \ln \left[\Sigma_i \exp(-E_i/A)\right] \qquad (1)$$

where A is any positive parameter which has molar energy units (kJ/mol) [5]. For stable compounds, equation (1) may be written as [5]

$$E = E_0 - \sum_{i \neq 0} N_i e_i \qquad (2)$$

where $e_i = A \exp\left[A^{-1}(E_0 - E_i)\right]$

E_0 = the energy of the dominant conjugate

and N_i = the number of conjugates with energy E_i.

In general, the energy of the dominant conjugate, E_0, is determined by the contribution of atoms and bonds. For acyclic hydrocarbons, for example

$$E_0 = x(\mathrm{C{-}C})n(\mathrm{C{-}C}) + x(\mathrm{C{=}C})n(\mathrm{C{=}C}) + x(\mathrm{C{\equiv}C})n(\mathrm{C{\equiv}C}) + x(\mathrm{C{-}H})n(\mathrm{C{-}H})$$

where $n(\cdot)$ represents the number of bonds and $x(\cdot)$ represents the contributions [7]. Only independent, adjustable parameters are considered; thus, the contributions of the individual atoms are not included, because in acyclic hydrocarbons the number of carbons and hydrogens is dependent on the number of bonds.

Conjugate operators

A recessive conjugate may be obtained from the dominant form by a rearrangement of electron pairs. A conjugate operator defines a particular pattern of electron rearrangement. When applied to a dominant conjugate, an operator yields an entire

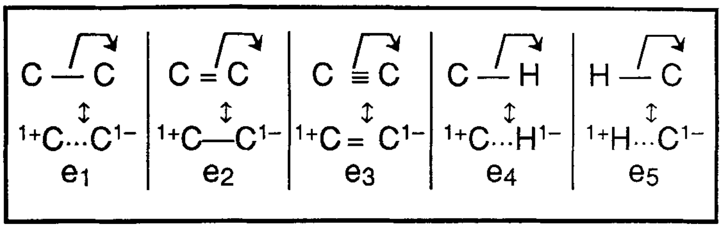

Figure 1 The fundamental conjugate operators for hydrocarbons.

class of recessive conjugates. Acyclic hydrocarbons have five one-bond operators shown in Figure 1. These are called fundamental operators. As isomers have the same number of carbon-carbon and carbon-hydrogen bonds, fundamental operators cannot account for the lower energies of branched alkanes. It is necessary to consider conjugate operators of greater bond length.

There are twenty-five three-bond operators for acyclic hydrocarbons. An example of a conjugate operator is shown in form (3).

$$\mathrm{C{-}C{-}C{-}H} \leftrightarrow \mathrm{C^{+}\cdots C{=}C\cdots H^{-}} \qquad (3)$$

The CCCH conjugate operator can be applied to each distinct subchain of the form C—C—C—H in a molecule. Each conjugate operator changes the energy of the dominant conjugate by a constant amount. Thus the operator can be related to $(E_0 - E_i)$ and therefore e_i in Equation (2). The parameter e_i determines the energy contributions of recessive conjugates.

The energy contribution of three-bond conjugate operators may be obtained by matching the fundamental operators against the conjugate operators. A fundamental operator which forms a bond may be considered the inverse of the one that breaks a bond. If e_k is used to denote the contribution of each fundamental operator and the operator itself, the conjugate operator CCCH may be decomposed as follows:

$$\mathrm{C{-}C{-}C{-}H} \xrightarrow{e_1} \mathrm{{}^{+}C\cdots{}^{-}C{-}C{-}H} \xrightarrow{e_4} \mathrm{{}^{+}C\cdots{}^{-}C{-}{}^{+}C\cdots{}^{-}H} \xrightarrow{e_2^{-1}} \mathrm{{}^{+}C\cdots C{=}C\cdots{}^{-}H}$$

The symbols e_1 and e_4 represent the application of the corresponding fundamental operators while e_2^{-1} represents the application of the fundamental operator e_2 in the reverse direction. To determine the effect of the operator CCCH on the energy of the dominant conjugate, the energy contributions from operators e_1 and e_4 are applied while the energy contribution of e_2 is applied in the reverse. It can be shown that multiplication of the energy parameters (e_k and e_k^{-1}) is a suitable way to combine the fundamental e_k's to obtain the e_i of the conjugate operator [7]. Thus $e(\mathrm{CCCH}) = e_1 e_2^{-1} e_4$. The technique of obtaining the energy contribution of a composite operator from fundamental operators may be extended to include longer operators. Thus increasing the length of the conjugate operators does not increase the number of adjustable parameters, e_i, in Equation (2).

One may create for each conjugate operator an array where each position corresponds to a fundamental operator, $e_k (1 \leq k \leq 5)$. Each time a fundamental operator is applied to a conjugate operator, one is added to that position. If the inverse of an operator is applied, one is subtracted from that position. The result is an array from which the energy parameters of conjugate operators may be expressed as a function of the energy parameters of the fundamental operators. This array is called the energy array, X_i, of the operator.

The above discussion was restricted to acyclic hydrocarbons. However the principles may be generalized to include organic compounds, cyclic and acyclic, that contain hetero-atoms. The procedure presented here was developed to manipulate acyclic compounds only. It could, however, be modified to handle cyclic molecules.

This paper describes a set of algorithms which generate and enumerate the conjugate forms of acyclic organic compounds to be used in the ABC approach for estimating the properties of compounds. We begin by discussing the data structures which are used to represent and manipulate molecules and operators. The rules for determining whether or not a chain of atoms and bonds represents a legitimate operator are then developed. Based on these restrictions, algorithms which generate the sets of conjugate and fundamental operators are presented. The algorithms which transform the input representations of molecules, calculate N_i, and determine the energy arrays of conjugate operators are also discussed. Finally, we present the results of applying the ABC technique of estimating properties of compounds using the output of these algorithms.

II DATA STRUCTURES

There are many approaches to the computer representation and transformation of molecular structures targeted to different applications [e.g., 8, 9, 10, and 11]. In this paper, we focus on the representation that suits our problem - the conjugation of molecular structures.

The data structures and algorithms presented here are based on Object-Oriented Programming. An object is a data structure that defines a component or building-block of a complex system. Each object has a set of properties called attributes. A slot is defined as the space in an object that is allocated to a particular attribute of the object. The objects are arranged into classes based on similarities in their attributes. The term "object" may refer to a class (effectively, a group of similar items) or to a specific instance of an item [12].

The set of objects chosen for our application is the element, atom, bond, molecule, and conjugate operator. A set of important attributes that play a central role in the generation and manipulation of the conjugates is given in Tables 1 through 4 for the objects which represent elements, bonds, atoms, and molecules. The attributes of an element are its valence capacity, symbol, inner shell charge,

Table 1 The attributes of the computational representation of an element, with carbon as an example.

Attributes of an Element	*Carbon*
Name	Carbon
Symbol	c
Valence Capacity	4
Inner Shell Charge	2
Atomic Number	6

Table 2 The attributes of the computational representation of an atom and examples of a primitive and complex atom. ATOM-1 is a complex atom because it is connected to the atoms and bonds listed in its atom-neighbors and bond-neighbors slots. These slots for ATOM-26 are empty, therefore it is a primitive atom.

Attributes of an Atom	*Primitive Positively Charged Oxygen*	*Complex Neutral Carbon*
Name	ATOM-26	ATOM-1
Atom Neighbors	( )	(ATOM-2 ATOM-3 ATOM-4 ATOM-5)
Bond Neighbors	( )	(BOND-1 BOND-2 BOND-3 BOND-4)
Type Element	O	C
Shell Electrons	5	4
Maximum Bonds	3	4
Charge	1	0

Table 3 The attributes of the computational representation of a bond and examples of a primitive and complex bond. BOND-1 is complex because its atom-neighbors slot has been assigned. BOND-24 is primitive because its atom-neighbors slot is empty.

Attributes of a Bond	*Primitive Double Bond*	*Complex Single Bond*
Name	BOND-24	BOND-1
Atom Neighbors	( )	(ATOM-1 ATOM-2)
Order	2	1

Table 4 The attributes of the computational representation of a molecule and an example. The atoms and bonds in the slots must be complex (i.e., connected).

Attributes of a Molecule	*Methane*
List of Bonds	(BOND-1 BOND-2 BOND-3 BOND-4)
List of Atoms	(ATOM-1 ATOM-2 ATOM-3 ATOM-4 ATOM-5)

and atomic number. This information is contained in the program. Currently carbon, hydrogen, oxygen, nitrogen, sulfur, and phosphorous are included. Defining elements as an object class obviates the necessity of storing invariant data for every atom created.

The attributes of an atom are its maximum number of permissible bonds, atom-neighbors, bond-neighbors, element type, shell electrons, and charge. When an instance of an atom is initially constructed, the type of element and charge is given. The number of shell electrons and permissible bonds are then calculated. This constitutes a primitive atom if it is not part of a molecule (i.e., has no atom-neighbors or bond-neighbors). If the atom is a portion of a molecule the atom-neighbors and bond-neighbors slots are assigned. The atom is then called a complex atom. These two attributes are fundamental in forming molecules. A bond has the attributes of atom-neighbors and bond order. To construct a bond, clearly its order must be specified. This again constitutes a primitive bond in the same sense as a primitive atom described above. If the bond is part of a molecule, a list of its neighboring atoms is placed in the atom-neighbors slot forming a complex bond. The attributes of a molecule are its list of atoms and list of bonds. The atoms and

Table 5 The attributes of the computational representation of a conjugate operator and an example. The chain is a list of primitive atoms and bonds. The chain length is the number of bonds in the operator.

Attributes of a Conjugate Operator	*Example*
Name	OPERATOR-23
Chain	(C—O—C+ = C—N—H)
Operator Type	case-1
Chain Length	5
Total Charge	1
Type Atoms	(H N C+ O C)

bonds of a molecule are all complex (i.e., they are interconnected). Thus all topological information concerning the molecule may be obtained from its list of atoms or list of bonds.

Table 5 lists the attributes of a conjugate operator. The chain length of a conjugate operator is the number of bonds it contains. The chain type is either case-1 or case-2 as defined in the discussion of legitimate operators below. The chain of a conjugate operator is a list of alternating atoms and bonds. An additional data structure is the fundamental operator. It consists of a list of the atom-bond-atom triple comprising the operator and the operator's array position, e_k. Fundamental operators and the chain of conjugate operators are built from primitive atoms and bonds.

III GENERATION AND ENUMERATION OF CONJUGATE FORMS

An overall procedure consisting of five algorithms has been developed to generate and analyze the necessary conjugates. The interrelationship between the algorithms and data structures is shown in Figure 2. The procedure first constructs all fundamental operators from a given set of atoms. All conjugate operators containing this set of atoms up to a maximum number of bonds are then generated. The energy-array, X_i, of each operator is then determined by comparing the conjugate operator to the set of fundamental operators. Thus, the energy contribution for each conjugate operator is determined as a function of the energy contributions of the fundamental operators. The preceding steps (corresponding to steps 1, 2, and 4 in Figure 2) deal solely with conjugate operators. Steps 3 and 5 in Figure 2 are directly related to the structure of the molecules. In these steps an input list representation of the molecule is converted into a form suitable for computer manipulation. Then the set of conjugate operators is applied to the dominant conjugate (i.e., the input molecule) to determine the number of occurrences for each operator (N_i).

The final output from these algorithms is the set of fundamental operators and their position in the energy array, the set of conjugate operators and their energy arrays, and the dominant conjugate and the number of times each conjugate operator (if different from zero) may be applied to it. This allows regression to be carried out, to determine the contributions of fundamental operators; or, if these adjustable parameters are known, the estimation of the energy of compounds from their structures.

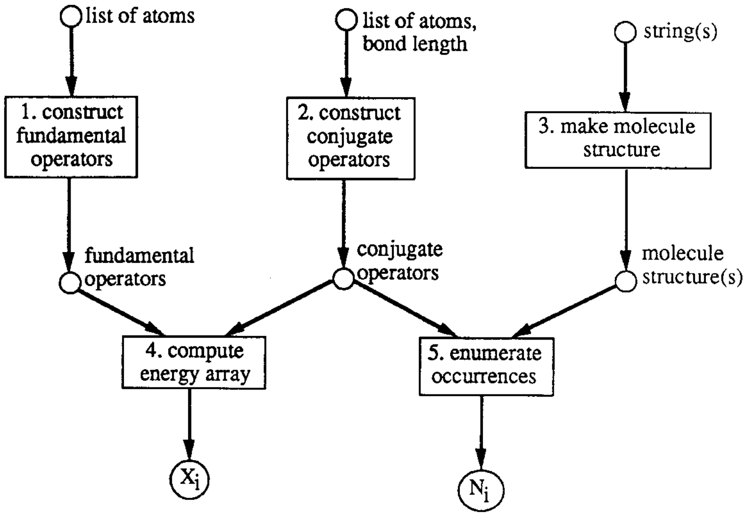

Figure 2 The interrelationship between the algorithms and the data structures.

The algorithms can manipulate acyclic molecules containing carbon, hydrogen, nitrogen, oxygen, sulfur, and/or phosphorous. Additional elements, with a maximum bond capacity of four, may be added to the program. The atoms may be neutral or carry a single (positive or negative) charge. Thus, compounds as well as ions can be manipulated. An overview of these algorithms appears below. A detailed description of the algorithms can be found elsewhere [13].

Legitimate conjugate operators

The generation of conjugate operators is an integral part of this program. The maximum number of bonds, types of atoms, and whether a specific type of atom may contain a single charge (positive and/or negative) must be specified. Charged atoms are generally treated as distinct from their neutral counterpart.

There are constraints that determine whether an operator may be legitimately applied to a dominant conjugate. Table 6 lists the types of legitimate conjugate operators. The restrictions in the generation of operators are based on considerations of the valence capacities and valence shell electrons of the individual atoms. The range of permissible bond orders depends on the pair of atoms. Carbon-carbon bonds can have orders between zero and three, but carbon-hydrogen bonds are limited to zero and one.

The application of an operator represents a movement of electron pairs in the dominant conjugate. This may also be viewed as alternately decreasing and increasing the orders of the bonds by one. The number of bonds and types of atoms involved in the electron rearrangement depend upon the origin of the electron pair. Electron pairs may originate from the first bond or the first atom of an operator as seen in Figure 3. Case-1 operators are defined as operators in which the first bond donates the electron pair. If the bonds are numbered from left to right, applying a case-1 operator will cause the odd bonds to decrease in order by one and the even

Table 6 Legitimate conjugate operators based on the valence capacities and valence shell electrons of the atoms.

Operator type:	*case-1*	*case-1*	*case-2*	*case-2*
electron pair from:	first bond	first bond	first atom	first atom
first atom:	any	any	e-donor	e-donor
odd bonds:	decrease	decrease	increase	increase
bond length:	odd	even	even	odd
last atom:	any	e-acceptor	any	e-acceptor

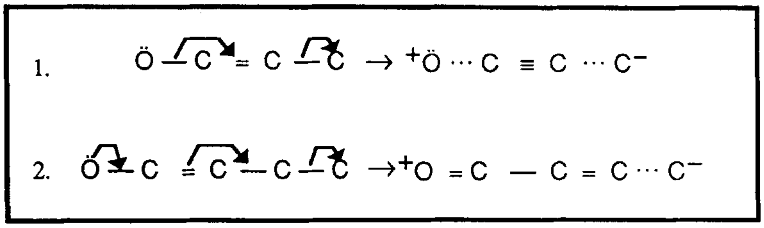

Figure 3 Examples of conjugate operators. The electron pair may originate from the first bond of an operator, called a case-1 operator, or from the first atom, called a case-2 operator.

bonds to increase in order by one. When the electron pair originates from the first atom we have a case-2 operator and the odd bonds increase in order by one while even bonds decrease by one. Clearly, the first atom of a case-2 operator must be able to donate an electron pair. Therefore, the atom must have valence-shell electrons which do not participate in bonds. Oxygen, nitrogen and negatively charged atoms (including C$^-$ and H$^-$), for example, have such extra valence-shell electrons which they can donate.

When case-1 operators of even bond length are applied to a compound, the movement of electrons will result in the last bond increasing in order by one. Therefore the last atom must be able to accept an electron. Thus it must have vacancies in its valence shell, which usually means it is positively charged. Similarly, case-2 operators of odd bond length must also terminate with an atom capable of accepting an electron. This can easily be seen by considering Figure 4. Applying form C—C—C on a dominant conjugate results in a negatively charged carbon with five bonds. Therefore C—C—C is *not* a legitimate operator. However applying C—C—C$^+$ on a dominant conjugate results in a neutral carbon with four bonds. Thus, C—C—C$^+$ is a legitimate case-1 conjugate operator. The results of applying forms Ö—C—C—C and Ö—C—C—C$^+$ when the electron pair is donated by the oxygen atom are analogous. Thus Ö—C—C—C$^+$ is a legitimate case-2 operator whereas Ö—C—C—C is not.

There is no restriction placed on the last atom of case-1 operators of odd bond length and case-2 operators of even bond length. The movement of electrons in these operators will cause the charge of the last atom and the number of bonds associated with the last atom to decrease by one, as shown below in form (4). Therefore, if the last atom in the operator already carries a single negative charge, the recessive conjugate which is generated will contain an atom with a negative charge of two. This can be seen from form (5) below.

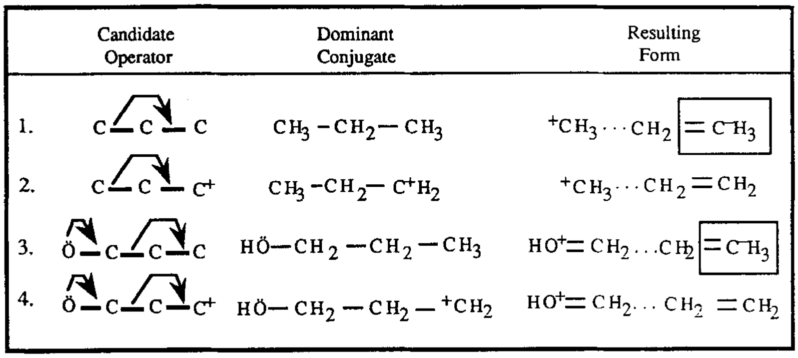

Figure 4 The effect of candidate operators on dominant conjugates. Applying candidate operators 1 and 3 on a dominant conjugate result in a negatively charged carbon with five bonds. Therefore these are *not* legitimate conjugate operators. Applying candidate operators 2 and 4, however, result in a neutral carbon with four bonds. Therefore these are legitimate conjugate operators.

$$C-C-O-C \leftrightarrow C^+ \cdots C=O \cdots C^- \qquad (4)$$

$$C-C-O-C^- \leftrightarrow C^+ \cdots C=O \cdots C^{2-} \qquad (5)$$

The operators which terminate with a negatively charged atom are generated, but the recessive conjugates thus formed will have a negligible influence on the energy of the hybrid compound.

To further reduce the number of operators generated, two external restraints are imposed. The sequence X^a BOND X^b, where a and b are 1, 0, or -1, is allowed only if X is a carbon. Thus, $N^+-N-N-N$ is not generated whereas $N^+-O-N-O$ is. This is one of the few instances where charged atoms are not distinguished from the neutral atom. Also, only one charged atom per operator is allowed, (e.g., C^+-C-C^--C is not generated). In imposing these restraints, the number of conjugates is greatly reduced while only a minimal amount of generality is sacrificed.

Construction of conjugate operators

This algorithm (corresponding to step 2 in Figure 2) generates all conjugate operators from a given set of atoms up to a maximum number of bonds. In this algorithm charged atoms are treated as distinct from their neutral counterpart. Thus the term "atom" refers *either* to X *or* X^+ *or* X^-. The general form of this algorithm is outlined below.

i. For each atom determine if the atom can donate electrons. If yes, construct two operators. Place copies of the atom in the chain of each operator. Label one operator "case-1" and the other as "case-2." Otherwise, construct one operator. Place a copy of the atom in the operator's chain. Label the operator as "case-1."

ii. Determine the set of bonds which can be connected to the last atom in the chain of the operator, based on the bond position (i.e., will applying this operator cause the bond to increase or decrease in order by one) and the atom's electron pairs which are not participating in bonds. If the set is empty, stop. Otherwise add each bond to the chain of a copy of the operator.

iii. Determine the set of atoms which can be connected to the last bond in the operator's chain based on the bond order and the externally imposed restrictions discussed previously. If the set is empty, stop. Otherwise add each atom to the chain of a copy of the operator.

iv. Determine if the chain is a legitimate operator based on the criteria summarized in Table 6. If yes, copy the operator and place it in the set of conjugate operators.

v. Determine if the chain has the maximum allowable number of bonds. If yes, stop. Otherwise go to step ii.

Construction of fundamental operators

This algorithm corresponds to step 1 in Figure 2. The generation of the fundamental operators (operators with only one bond) follows the same restrictions outlined above for the conjugate operators. Thus the algorithm which generates the fundamental operators is very similar to the one which generates the conjugate operators. Case-1 and case-2 fundamental operators are generated and stored separately. The fundamental operators are assigned an index, k. This index corresponds to the fundamental operator's position in the energy-array. It is used in the algorithm to compute the conjugate operators' energy-arrays.

Transformation of input molecule

The input list representations of the molecules are transformed into graphs suitable for computer manipulation by this algorithm, step 3 in Figure 2. The graph of a molecule contains information on the connectivity of the atoms and bonds. These structures do not contain any information on the stereochemistry of the molecule because such information is not used in the ABC approach.

This algorithm receives a list of symbols and determines whether each symbol represents an atom, bond, or charge. The appropriate objects, atoms and bonds shown in Tables 2 and 3, are then created and/or modified. Side chains are represented as nested lists. Hydrogen atoms and single bonds can be inferred by the algorithm. For example, 3-methenyl-1,4-pentadiene is represented as (c db c c (db c) c db c), where "db" is the symbol for a double bond. While it is assumed that the input representation is an actual chemical structure, the program will recognize and signal an error if an atom is assigned too many bonds based on the element's valence capacity and valence shell electrons. The program will not, however, signal a warning for other improbable structures (e.g., H_2N-NH-S-O-NH-NH_2). Clearly, this algorithm is highly dependent on the requirements of the input representation of the molecule.

Determination of the energy array, X_i

This algorithm, step 4 in Figure 2, receives the set of fundamental operators and the set of conjugate operators. The energy array of a conjugate operator is determined by comparing it to the fundamental operators. This operation was described in Section I for hydro-carbons. Determining the energy array of a conjugate operator that contains heteroatoms is described below.

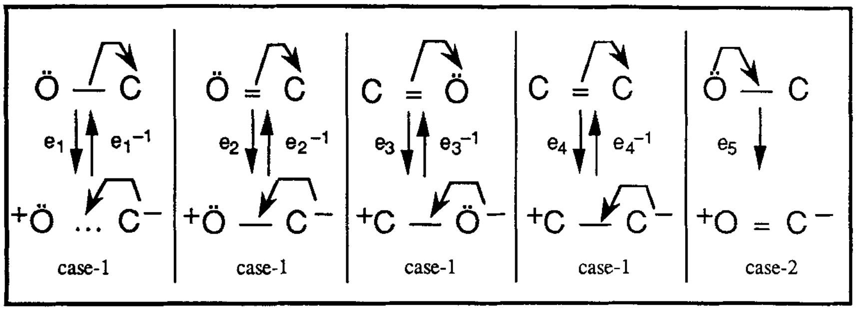

Figure 5 Some fundamental operators containing carbon and oxygen. The case-1 operators may be applied to conjugate operators either to break a bond (as e_k), or to form a bond, (as e_k^{-1}). Case-2 fundamental operators are only applied to conjugate operators to form a bond (as e_k).

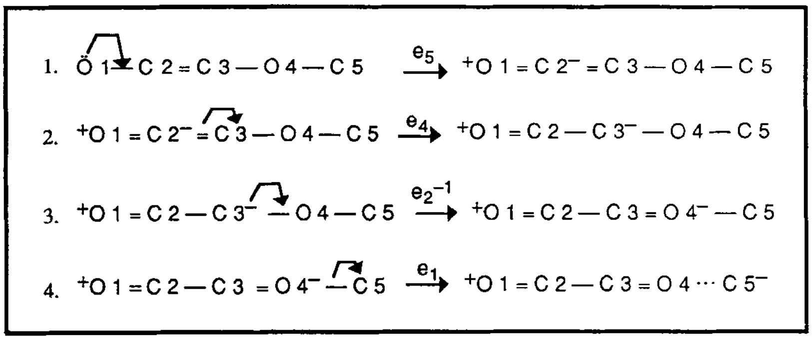

Figure 6 The effect of applying fundamental operators on the case-2 operator Ö—C=C—O—C. The resulting energy array is {1 −1 0 1 −1} where the positions correspond to $\{e_1 e_2 e_3 e_4 e_5\}$ shown in Figure 5.

Case-2 fundamental operators are only applied to the first atom-bond-atom triple of case-2 operators and always result in the order of the bond being increased. Case-1 fundamental operators are applied the remainder of the time. Several fundamental operators containing oxygen and hydrogen are shown in Figure 5. Consider the case-2 conjugate operator Ö—C=C—O—C. Applying the fundamental operators to this conjugate operator is shown in Figure 6. Because Ö—C=C—O—C is a case-2 operator, the set of case-2 fundamental operators is compared to the first atom-bond-atom triple (Ö1—C2 in Figure 6) of the conjugate operator. When a match is found, one is subtracted from position *k* of the case-2 fundamental operator in the energy array. The remainder of the conjugate operator is then compared to the set of case-1 fundamental operators.

Note that the inverse of a fundamental operator has the atoms in the reverse position with the bond order increased by one. In line 3 of Figure 6 the electron pair that increases the bond order originates from the carbon. As can be seen from Figure 5, this corresponds to the inverse of applying e_2 *not* e_3. Therefore, to transform (C—O) into (C=O), e_2^{-1} must be applied. The limited example shown in Figure 6 would result in the array {1 −1 0 1 −1} where the positions correspond to $\{e_1 e_2 e_3 e_4 e_5\}$. Determining the energy array of a case-1 conjugate operator is similar except that it is only compared to case-1 fundamental operators.

The algorithm can be summarized as follows:

i. If the conjugate operator is a case-1 operator, go to step ii. Otherwise, compare the first atom-bond-atom triple to the case-2 fundamental operators. When a match is found, subtract one from position k in the energy array. Remove the first atom and bond from the conjugate operator.
ii. For each case-1 fundamental operator, determine if it matches the next atom-bond-atom set of the conjugate operator. If yes, continue with step iii. Otherwise go to step iv.
iii. If the bond is in a "breaking" position, subtract one from position k of the energy array. If the bond is in a "forming" position, add one to position k of the fundamental operator's *inverse* in the energy array.
iv. Remove the first atom and bond from the conjugate operator. For each case-1 fundamental operator, determine if it matches the next atom-bond-atom set of the conjugate operator. If yes, go to step iii. Otherwise repeat step iv until there are no more bonds in the conjugate operator.

Enumeration of occurrences

This algorithm corresponds to step 5 of Figure 2 and determines the number of times a conjugate operator may be applied to a compound, N_i, by comparing the conjugate operators to the molecular structure. In this way, all recessive conjugates are enumerated. When counting the number of times a specific operator may be applied to a compound, symmetry is not taken into account. Consider the dominant conjugate of butane and the conjugate operator (C—C). The movement of electrons in the operator will yield the form ($C^+ \cdots C^-$). The number of recessive conjugates of this type in butane (C1—C2—C3—C4) is six rather than three. The conjugate ($C1^+ \cdots C2^-$—C3—C4) is clearly different than ($C1^- \cdots C2^+$—C3—C4). Therefore, both C1—C2 and C2—C1 are counted as occurrences.

The atoms and bonds of the molecule are unique and connected objects (i.e., complex bonds and atoms as shown in Tables 2 and 3), whereas those of the conjugate operator are not. When the first matching atom is found, a list is created and the atom of the *molecule* is placed in this list. All subsequent matching atoms and bonds of the molecule are also placed in the list. The list is checked when a possible match is found to ensure that the atom or bond of the molecule has not already been visited. The algorithm for comparing one conjugate operator to a molecule is as follows:

i. For each atom in the molecule, compare it to the first atom in the operator. If there is a match, place the atom in an empty list. Otherwise stop.
ii. For each bond in the molecule that is connected to the last atom in the list, determine if it is the same order as the next bond of the operator and not already in the list. If yes, remove the bond from the operator and place the bond from the molecule in a copy of the list. Otherwise stop.
iii. For each atom in the molecule that is connected to the last bond in the list, determine if it is the same element and charge as the next atom in the operator and not already in the list. If yes, remove the atom from the operator and place the atom from the molecule in a copy of the list. Otherwise stop.
iv. Determine if there are any more atoms or bonds in the operator. If yes, go to step ii. Otherwise increase N_i by one and stop.

Table 7 Comparison of the ABC technique with the group contribution method for acyclic hydrocarbons.

			Group Contributions			*ABC*		
	Property	*Data*	*Parameters*	*Std. Deviation*	*Average Absolute Error*	*Parameters*	*Std. Deviation*	*Average Absolute Error*
ΔH_f	Heat of Formation	117	9	5.0 kJ/mol	3.9 kJ/mol	8	4.2 kJ/mol	3.4 kJ/mol
S^0	Standard Entropy at 298 K	90	9	3.5 J/mol	2.7 J/mol	9	2.8 J/mol	2.4 J/mol
V_c	Critical Volume	47	9	0.009 l/mol	0.007 l/mol	9	0.007 l/mol	0.006 l/mol
T_c	Critical Temperature	41	10	5.1 K	3.5 K	10	2.4 K	1.9 K
P_c	Critical Pressure	49	9	0.86 atm	0.65 atm	9	0.50 atm	0.31 atm
T_b^0	Normal Boiling Point	67	9	4.2 K	3.1 K	5	3.6 K	2.7 K
ΔH_v^0	Standard Heat of Vaporization at 298 K	44	9	0.46 kJ/mol	0.38 kJ/mol	6	0.39 kJ/mol	0.33 kJ/mol

IV ESTIMATION OF PROPERTIES OF ACYCLIC HYDROCARBONS

Here, we give an example of a property-estimation scheme which uses the algorithms described above. In this example, we used conjugate operators of length ≤ 3, to estimate the heat of formation of acyclic hydrocarbons. A regression was carried out using experimental data for the heat of formation of 117 acyclic hydrocarbons [14]. The standard deviation of the regression was 4.2 kJ/mol. In the regression analysis it was determined that the contribution from the fundamental operator (H—C) is insignificant, as expected. Thus, ABC uses only eight adjustable parameters; the four bond contributions and the four contributions from the remaining fundamental operators. The values derived for the adjustable parameters were:

$$e_1 = 2.035; \quad e_2 = 14.205; \quad e_3 = 28.848; \quad e_4 = 3.530; \quad (\text{in kJ/mol})$$
$$x_d(\mathrm{C{-}C}) = 20.574; \quad x_d(\mathrm{C{=}C}) = 144.045; \quad x_d(\mathrm{C{\equiv}C}) = 317.154;$$
$$x_d(\mathrm{C{-}H}) = -13.692; \quad (\text{in kJ/mol}).$$

For comparison, a group contribution fit on the same set of compounds, using a total of nine groups yielded the following contributions:

$$k(-\mathrm{CH_3}) = -42.781; \quad k(>\mathrm{CH_2}) = -20.569; \quad k(>\mathrm{CH}-) = -2.340;$$
$$k(>\mathrm{C}<) = 15.793; \quad k(=\mathrm{CH_2}) = 21.425; \quad k(=\mathrm{CH}-) = 37.086; \quad k(=\mathrm{C}<) = 48.090;$$
$$k(\equiv\mathrm{CH}) = 112.606; \quad k(\equiv\mathrm{C}-) = 115.901; \quad (\text{in kJ/mol}).$$

The standard deviation of the group-contribution regression, using the same experimental data, was 5.0 kJ/mol [7]. Even though ABC uses fewer adjustable

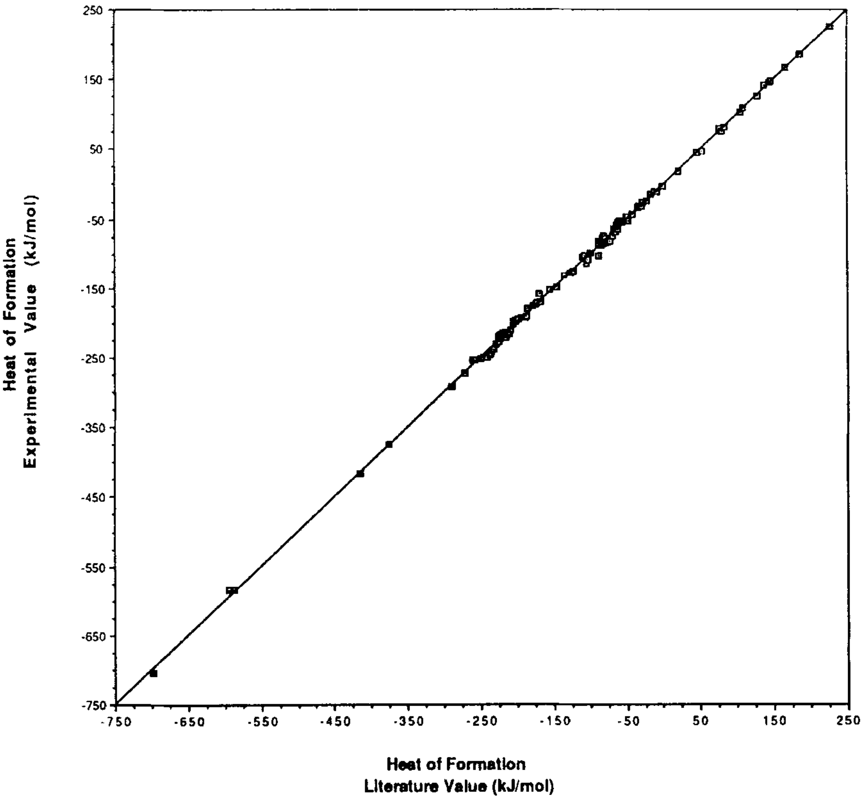

Figure 7 Experimental versus literature value of the heat of formation, ΔH_f.

parameters (8 parameters vs. 9 for group contributions), it performs better than group contributions (i.e., gives a lower standard deviation) in this example.

The product of the fundamental operators contribution, e_i, has been viewed above as an *energy*-parameter. However, the values of e_i are also used as weighting factors for the estimation of standard entropy of formation, normal boiling point, heat of vaporization, critical pressure, critical temperature, and critical volume of acyclic hydrocarbons [15]. Table 7 gives, for each property, a comparison between ABC and the group-contribution framework. In all cases, ABC exhibits a lower standard deviation and average absolute error. Compared to the group-contribution scheme, ABC often provides dramatic improvements either in the standard deviation (for T_c and P_c - critical temperature and pressure) or in the number of adjustable parameters (for T_b^0 and ΔH_v^0 - normal boiling point and standard vaporization). Figures 7, 9, 11, 12, 14, 16, and 17 show the correlation between the estimated values and experimental values found in the literature for each property showing all data points used in the regression. Figures 8, 10, 13, and 15 are enlargements of a portion of the estimated values vs. the literature values of the heat of formation, standard entropy, critical temperature, and critical pressure, respectively. There is no correlation between the size of a molecule (i.e.,

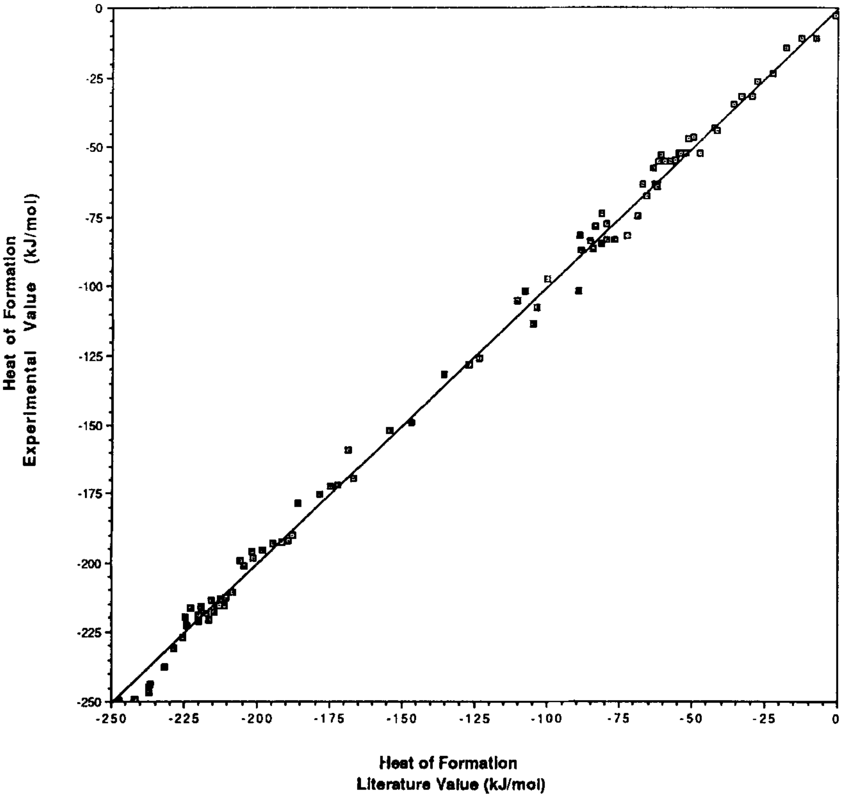

Figure 8 Experimental versus literature value of the heat of formation, ΔH_f, showing 91 data points. This figure is an enlargement of part of Figure 7.

number of bonds) and the accuracy of the value estimated using the ABC method. A detailed description of the regression and analysis of the results may be found elsewhere [15, 16]. The correlation between N_i, e_i, and the properties of acyclic compounds containing hetero-atoms is currently being analyzed [16].

IV CONCLUDING REMARKS

We have developed the algorithms necessary to generate, compare, and analyze the recessive conjugates of acyclic organic compounds. The use of symbolic computing environments allows for the flexible representation and manipulation of the objects used in the algorithms. The algorithms may be easily modified to include additional hetero-atoms and remove and/or add external restrictions on the generation of conjugate operators. These algorithms play an integral role in the application of the ABC technique for estimating physical and thermodynamic properties of compounds. This approach leads to an accurate additive property estimation technique based exclusively on the molecular structure of a compound.

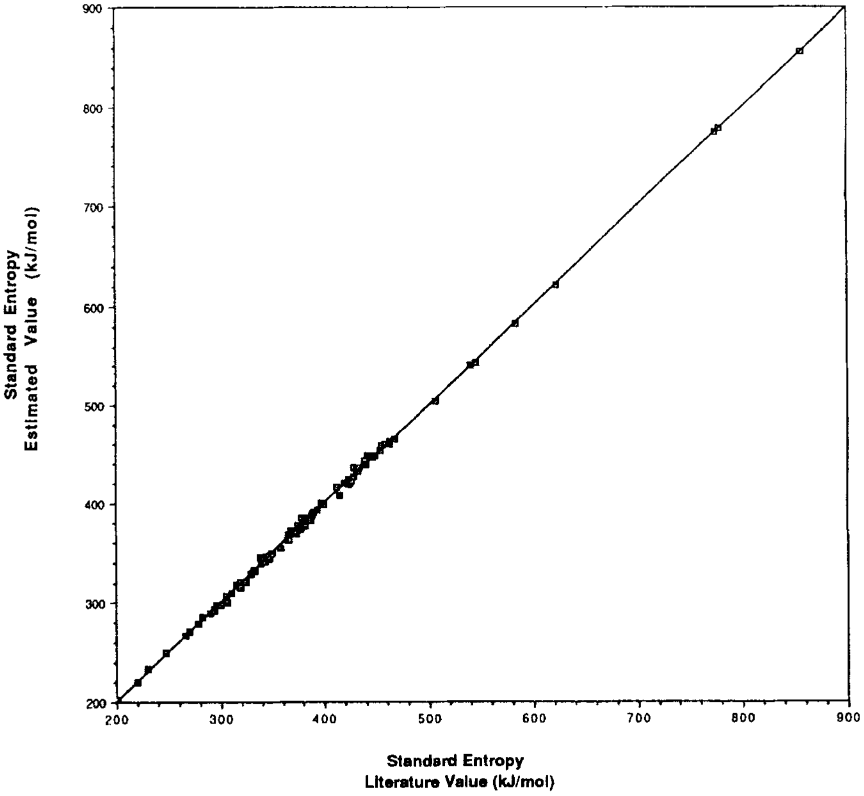

Figure 9 Experimental versus literature value of the standard entropy, S^0.

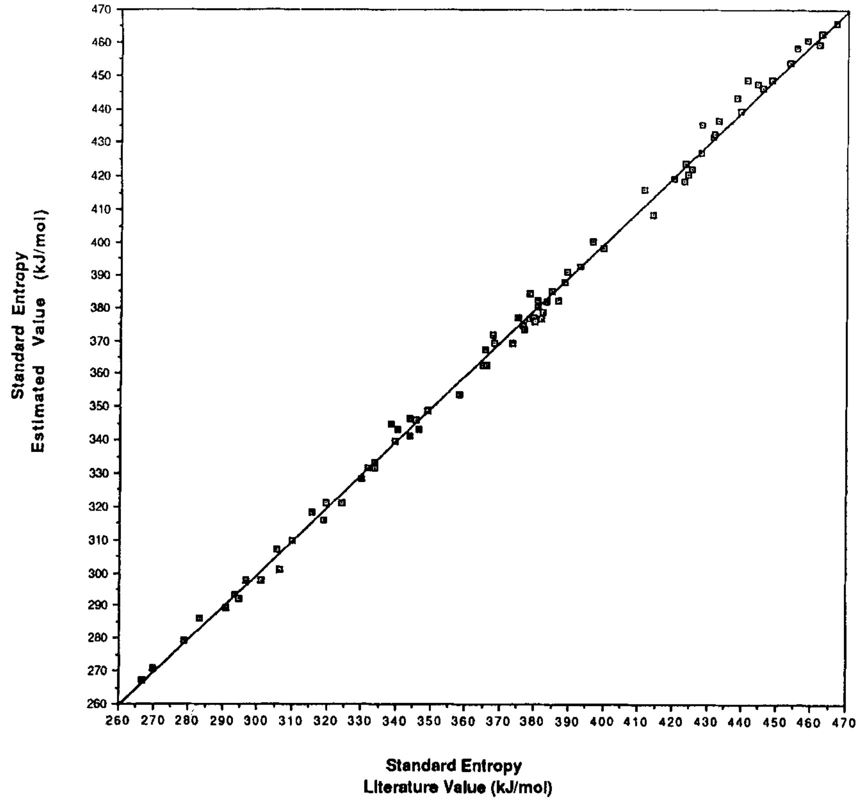

Figure 10 Experimental versus literature value of the standard entropy, S^0, showing 78 data points. This figure is an enlargement of part of Figure 9.

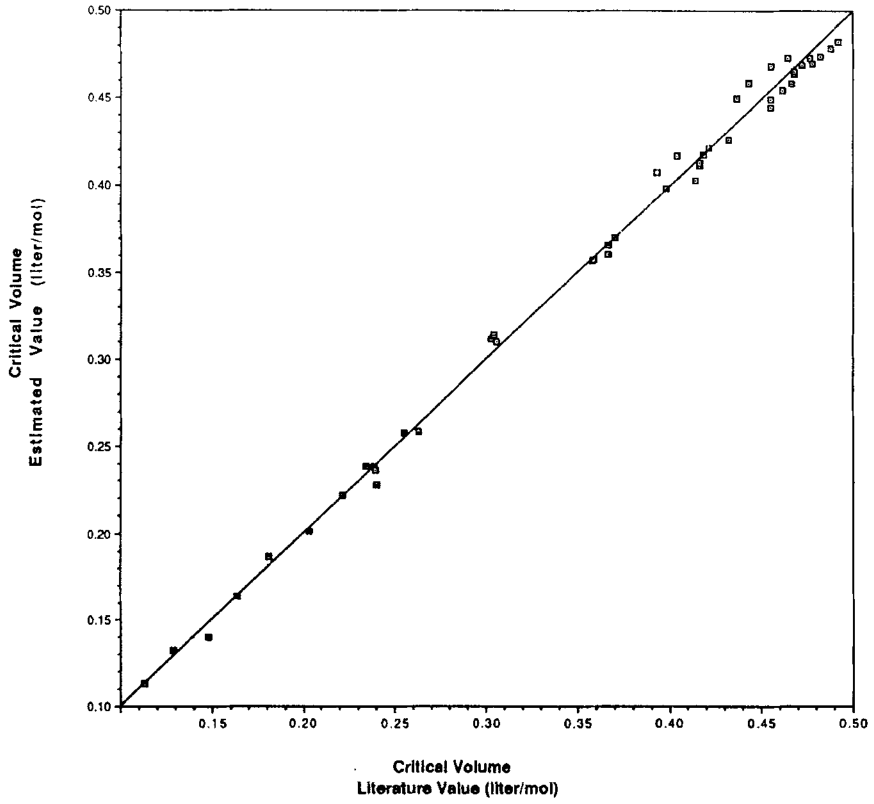

Figure 11 Experimental versus literature value of the critical volume, V_c.

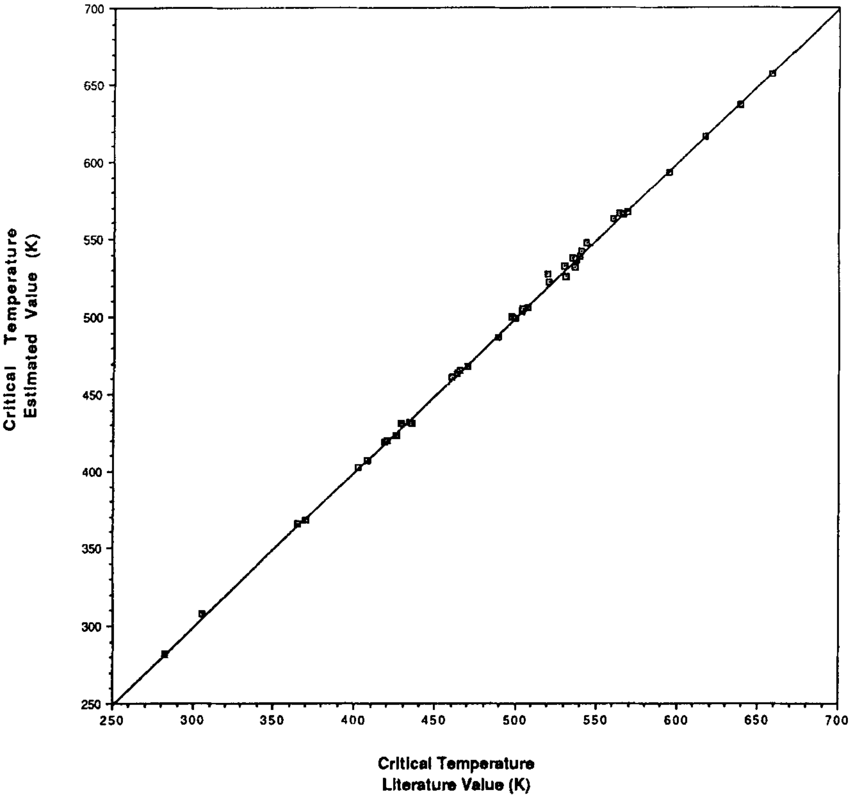

Figure 12 Experimental versus literature value of the critical temperature, T_c.

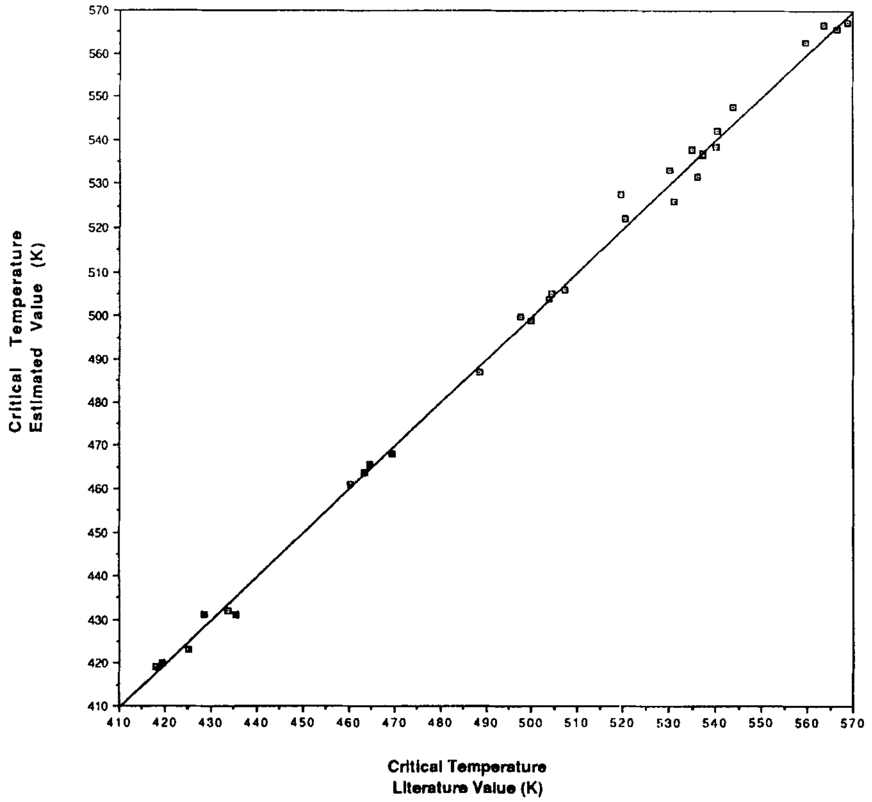

Figure 13 Experimental versus literature value of the critical temperature, T_c, showing 31 data points. This figure is an enlargement of part of Figure 12.

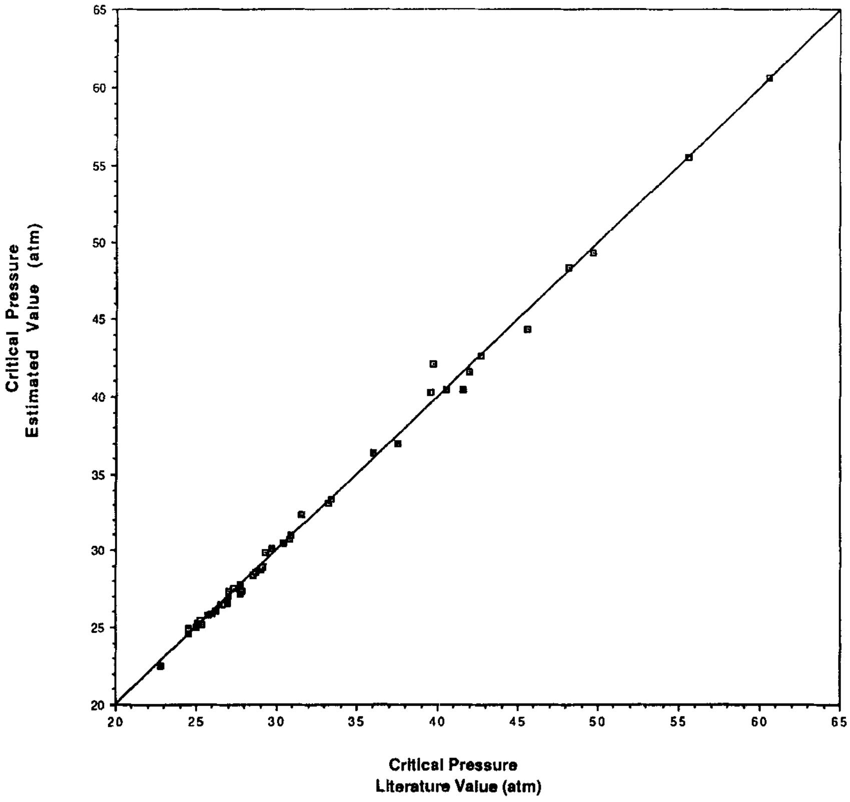

Figure 14 Experimental versus literature value of the critical pressure, P_c.

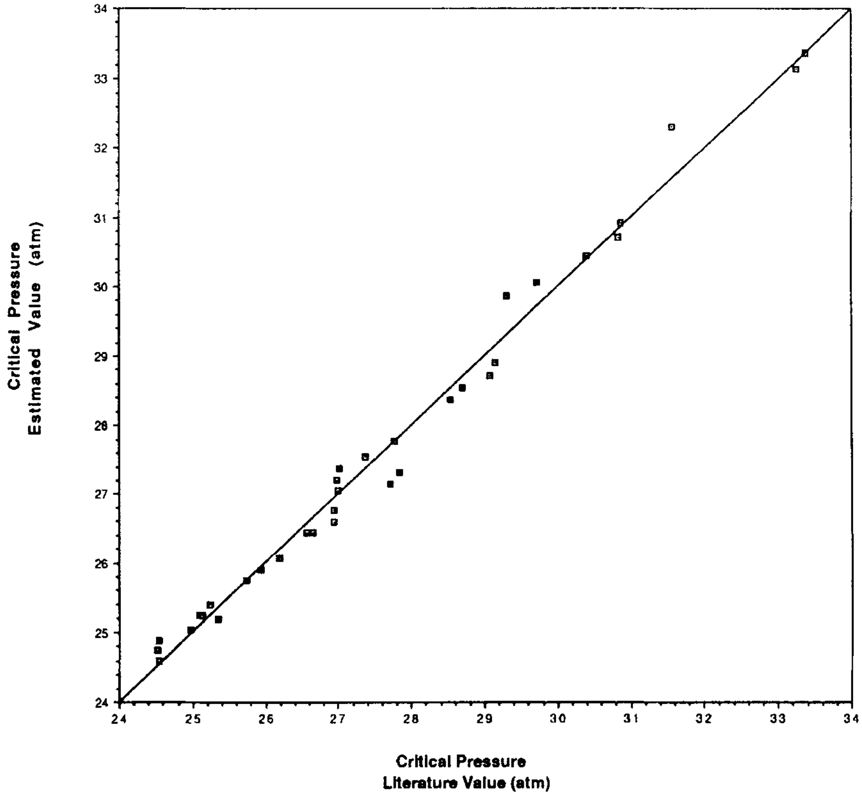

Figure 15 Experimental versus literature value of the critical pressure, P_c, showing 34 data points. This figure is an enlargement of part of Figure 14.

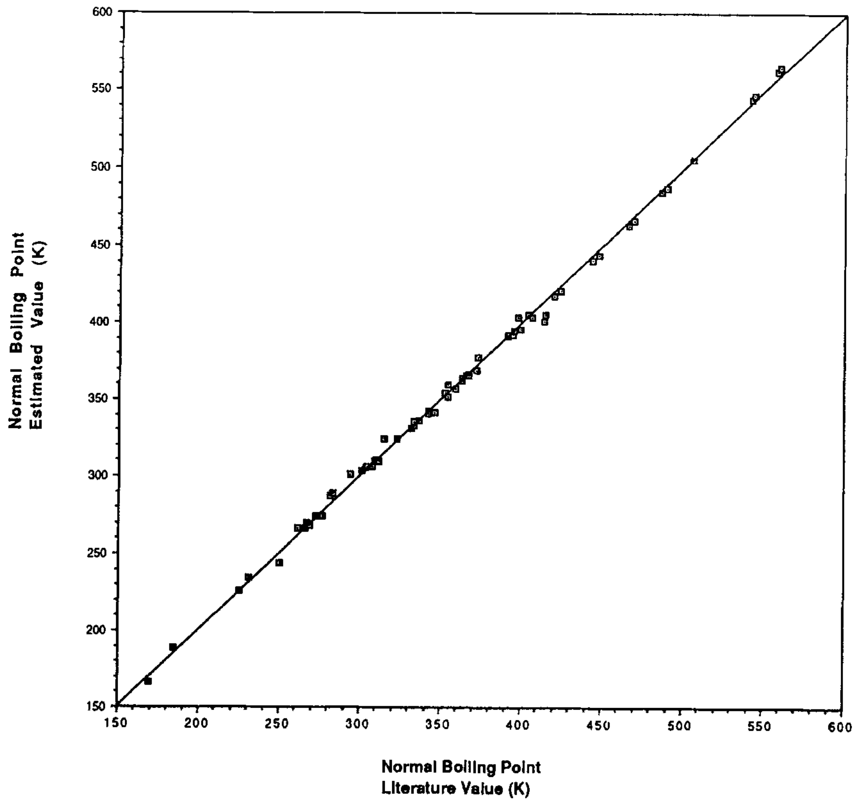

Figure 16 Experimental versus literature value of the normal boiling point, T_b^0.

References

[1] Joback, K.G.; Reid, R.C. "Estimation of Pure-Component Properties from Group Contributions". *Chemical Engineering Communications*, **57**, 233-243, 1987.

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

[3] Domalski, E.S.; Hearing, E.D. "Estimation of the Thermodynamic Properties of Hydrocarbons at 298.15 K". *Journal of Physical and Chemical Reference Data*, **14**, 1637-1678, 1988.

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

[5] Mavrovouniotis, M.L. "Estimation of Properties from Conjugate Forms of Molecular Structures: The ABC Approach". *Industrial and Engineering Chemistry Research*, **29**, 1943-1953, 1990.

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

[7] Mavrovouniotis, M.L.; Prickett, S.E.; Constantinou, L. "Object-Oriented Programming in the Estimation of Properties from Molecular Structures". *Computers and Chemical Engineering*, **16 Supplement**, S353-S360, 1992.

[8] Meyer, E. "Computer Representation and Handling of Structures: Retrospect and Prospects". *Journal of Chemical Information and Computer Science*, **31**, 68-75, 1991.

[9] Fisanick, W. "The Chemical Abstracts Service Generic Chemical (Markush) Structure Storage and Retrieval Capability. 1. Basic Concepts". *Journal of Chemical Information and Computer Science*, **30**, 145-154, 1990.

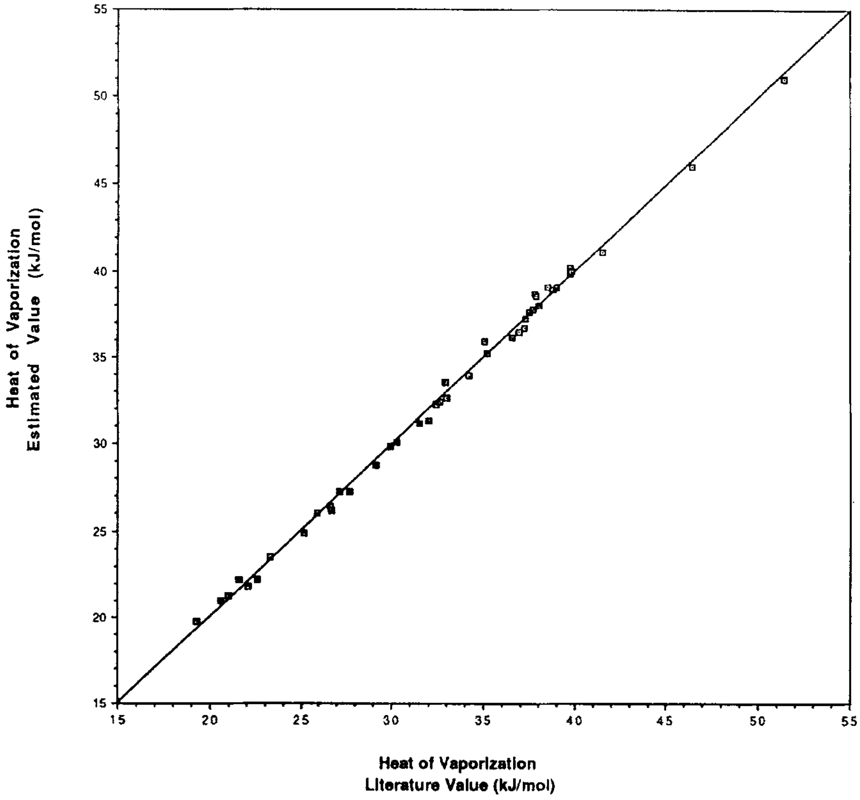

Figure 17 Experimental versus literature value of the standard heat of vaporization, ΔH_v^0.

[10] Blower, P.E.; Whitlock, H.W. "An Application of Artificial Intelligence to Organic Synthesis". *Journal of the American Chemical Society*, **98**, 1499-1510, 1976.

[11] Gasteiger, J.; Ihlenfeldt, W.D.; Röse, P. "A collection of computer methods for synthesis design and reaction prediction". *Recueil des Travaux Chimiques des Pays-Bas*, **111**, 270-290, 1992.

[12] Forsythe, R.G.; Prickett, S.E.; Mavrovouniotis, M.L. *An Introduction to Object-Oriented Programming in Process Engineering*; CACHE Corp.: Austin, Texas, 1993.

[13] Prickett, S.E. "Generation and Enumeration of Conjugation Chains in Acyclic Compounds". *M.S. Thesis, Dept. of Chemical Engineering, University of Maryland*, December, 1992.

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

[15] Constantinou, L.; Prickett, S.E.; Mavrovouniotis, M.L. "Estimation of Thermodynamic and Physical Properties of Acyclic Hydrocarbons using the ABC Approach and Conjugation Operators". Accepted for publication in *Journal of Industrial and Engineering Chemistry Research*, March, 1993.

[16] Constantinou, L. "Estimation of Properties of Acyclic Organic Compounds through Conjugation". *Ph. D. Dissertation, Dept. of Chemical Engineering, University of Maryland*, May, 1993.