Measuring Splitting and Overlapping in Molecular Systems. A Fuzzy Set Estimation of Reaction Hazards

Lech Schulz[†]
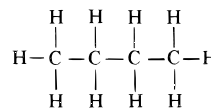
Quantum Chemistry Group, Department of Chemistry, A. Mickiewicz University, Grunwaldzka 6,
60-780 Poznań, Poland

An algorithm for assessments of fuzzy molecular structural characteristics has been presented, and its chemical relevance is approved for numerous evidences of reaction mechanisms. Empirical rules pervading the results of chemical synthesis are continuously the source of information efficiently used by experimental chemists and technologists. Approaching them theoretically to broaden areas of applicability and to make them more precise for better prediction is of great importance to reach progress in the chemical information recognition and to steer and control technological processes. Fuzzy sets called splitting and overlapping have been applied for assessments of the reaction hazards. The mathematically grounded properties turned out to underlie well-known empirical rules for preliminary estimations of organic reaction tendencies. Informative quantitative data shown are revealing new ways of understanding basic chemical reaction mechanisms.

INTRODUCTION

The interest in recognizing the principles of chemical structures is entirely justified as the cognitive and technologically applied problem. One of the basic concepts in chemistry is the notion of the elementary unit called the molecule, which is difficult to be overappreciated in the recognition and diversification of chemical substances. These objects and their pertinent features are the goal of investigating them from the viewpoint of mathematical discrete chemistry approaches. The relatively new idea of self-organization has widely been used in a similar concept in the foundations of chemistry. This is done by the basic stoichiometry regularities and, for example, the phenomenon of saturation. It could be noted that modern chemistry started to develop in the 19th century from the idea of assembling atomic species and giving elementary quantitative characteristics in the form of empirical rules. The result of such self-organization (self-assembly) distinguishes an elementary unit first described by the molecular formula and by the so-called structural formula. Since different compounds can have the same molecular formula, the result of distinguishing a molecular species must be attributed to the self-assembly structure, i.e., by distinguishing some relationships between nuclei in the molecules. This is, as usual, performed on the grounds of the traditional theory of bonding based on the electronic theories or on the quantum mechanical concepts which overbuild the chemistry starting from the microscopic level of matter constitution.[1,2] The theory of multiple bonding (i.e., multiple links) is not necessary to explain the self-assembly phenomenon in the chemical world, and there exist general mathematical entries free of redundant interpretations but not rejecting the core of quantum chemical and experimental results.[3]
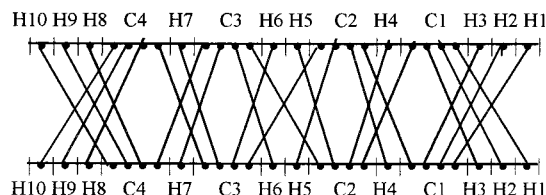
† Mailing address: Institute of Bioorganic Chemistry, Noskowskiego 12/14, 61-704 Poznań, Poland. Phone: 48-61 852 85 03. Fax: 48-61 865 87 01. E-mail: lsch@artemida.amu.edu.pl.

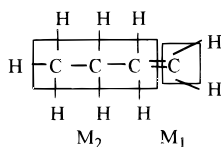**Figure 1.** A common iconic language insert in chemistry.

Consider a chemical space[3,4] prespecified by a molecular formula. For example, in the case of butane, it can be pictured as the well-known chemical icon[5,6] (Figure 1). Such a picture represents some entities having full mathematical status as a graph or as a specific mathematical structure elicited.[3] The specification and modification has been made on the graph realm[7] in the hope of finding information in addition to the basic chemical assumptions included. To approach it, a little more abstract situation is useful to consider. Let StrA denote all the structures created over the set A. The concept of self-organization given in ref 2 (p 168) is close to a mathematical simple invention to make self-assembly by discrimination between some structures of all structures which can be induced on the given set A (denoted StrA). Even finiteness of A does not imply finiteness of a variety of structures of the class StrA. The better the findings in the "bag" called StrA, the richer the chemical information revealed. It has been shown in refs 4 and 8 that the determination of the so-called self-assembly chemical SAC spaces implies also the choice of other, alternative systems which have interesting peculiarities regarding both mathematical and interpretation aspects. For example, owing to it, there is no necessity to consider the connecting lines between atomic species as multiple linking bonds. Moreover, the structure can equivalently be transformed into a permutation which is expressing a one-to-one correspondence, only between the different nuclei. This can be exhibited as the so-called compensating system[8] (Figure 2).

Thus, the important information encoded in the "topological" structure of chemical spaces is unveiled by considering the equivalent permutation, which clearly shows the specific assignments between atoms rather than bonding conceived

SPLITTING AND OVERLAPPING IN MOLECULAR SYSTEMS

J. Chem. Inf. Comput. Sci., Vol. 40, No. 4, 2000 **1019**

Figure 2. Compensating function underlying the icon in Figure 1.



Figure 3. Discriminated fragments of 1-butene (denoted M_1 and M_2).

in the classical sense. It should moreover be stressed that the terms "bond", and "double bond" used here do not bear any meaning, which includes the traditional understanding of multiplicity of bonds using, e.g., Lewis pairs or orbital theory. Stressing the concept of self-assembly in molecular structure, the point of gravity is moved from the question, "What is the matter constituting molecular arrangement?" onto the problem, "how is it made?". The mathematical sense introduced bears limitations only by general chemical assumptions coming from stoichiometric and saturation phenomena. Besides, due to some findings, no interaction of particles is necessary to distinguish the general structural form consisting of certain mathematical correspondences, though the results show the self-assembly design to be well correlated with the intramolecular relationships essentially affecting chemical reactivity. Since such SAC[3,4,8] spaces are easily transformed into multigraphs, they also turn out to be underlain by some partial one-to-one correspondence easily symmetrized to the respective permutations (cf. Figures 1 and 2). Hence, the scope of potential fuzzy set applied approaches presented in this paper goes far beyond the chemical examples, though they are extremely stimulating in the research. The mathematical background does include interesting entries into category theory and foundations of mathematics, a detailed presentation of was left to the mathematically oriented presentations in contrast to the below-stated descriptive way suitable for experimentally and technologically practicing chemists.

ELICITING CHEMICAL STRUCTURE INFORMATION ON SPLITTING AND OVERLAPPING IN MOLECULAR SYSTEMS

A collection of hydrocarbons will be considered to make analyses of the mechanisms of their reactions with the use of overlapping and splitting introduced. Computational algorithms will be illustrated by the case of SAC space of 1-butene with suppressed hydrogen atoms. The fragments under consideration can be pictured as shown in Figure 3.

For the corresponding compensating system (with suppressed hydrogen) we have function κ illustrated as shown in Figure 4. Chemical structure approached in the way pictured in Figure 4 has some sets distinguished as seen at first glance. To approach it, only the well-known basic set theoretic operations will be used in further considerations,
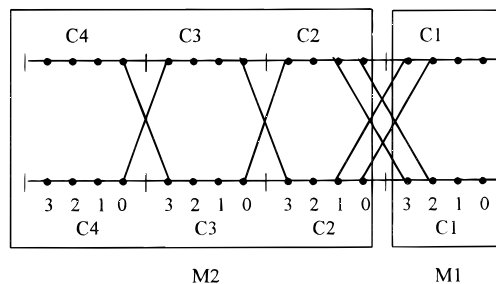
Figure 4. Compensating function to be analyzed.
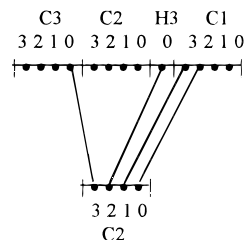
Figure 5. A selected partial correspondence received of Figure 1.

i.e., the union of sets, say a and b (denoted a ∪ b) which includes elements of a or b and their intersection (denoted a ∩ b) consisting of common elements of both sets. $\{x\}$ denotes a set having only one element, x, $\{x,y\}$ is a set having two elements assuming $x \neq y$, $\{x,y,z\}$ is a set having three elements, etc. $\{x\}$ can of course also be only one element of a set $\{\{x\}\}$. $|a|$ is sometimes called the cardinality of a, which means the number of its elements.

Let us consider a particular carbon atom, say C^2 (=C2; Figure 5). The correspondences shown in the figure clearly distinguish some subsets on the carbon C^2, i.e., the subset which is induced by carbon C^1 consisting of elements $\{(C^2,0),(C^2,1)\}$, the subset which is induced by hydrogen H^3 consisting of one element $\{(C^2,2)\}$, and the subset $\{(C^2,3)\}$ which is induced by carbon C^3. Thus, a family of subsets has been distinguished on C^2 which equals

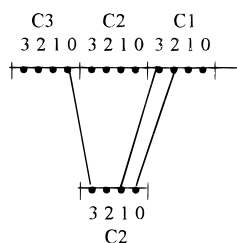$$\{\{C^{20},C^{21}\},\{C^{22}\},\{C^{23}\}\} \tag{1}$$

where elements such as $(C^2,0)$ have been denoted as C^{20} for short. The family determined consists of the sets which mutually have no elements in common (they are disjoint), and their union covers C^2. A partition has been induced in this way on the carbon by correspondences shown in Figure 5.

The same correspondences select some elements on C^1, H^3, and C^3; i.e., they distinguish set $\{C^{12},C^{13},H^{30},C^{30}\}$. Moreover, atomic species being related by the correspondences pictured induce a new partition of this set in a self-understood way:

$$\{\{C^{12},C^{13}\},\{H^{30}\},\{C^{30}\}\} \tag{2}$$

Thus, two different partitions have been obtained: one (1) induced *on* C^2, and the second one (2) induced *by* C^2-related species. Elements of these partitions are called constituents.

The same can be exercised for the butene structure with suppressed hydrogen. The fragment with the C^2 carbon and corresponding species is as shown in Figure 6.

C3   C2   C1
3 2 1 0 3 2 1 0 3 2 1 0

3 2 1 0
C2

Figure 6. A selected partial correspondence of Figure 4.

The partition induced *on* C^2 is in particular the following:

$$\{\{C^{20},C^{21}\},\{C^{22}\},\{C^{23}\}\} \tag{3}$$

Nonrelated elements such as C^{22} are included in one constituent consisting of all nonrelated elements for a fixed species such as C^2.

The partition induced *by* C^2 is

$$\{\{C^{30}\},\{C^{12},C^{13}\}\} \tag{4}$$

This is, of course, a partition of a set induced by C^2-related species, i.e., $\{C^{12},C^{13},C^{30}\}$.

Let's come back to the whole structure of butene with suppressed hydrogen (Figure 4) and consider partitions emerging for this structure. Set A, which is to be partitioned in various ways, does consist of elements C^{ij} for $i = 1, ..., 4$ and $j = 0, ..., 3$. The elements belonging to fragment M_1 with added C^{20},C^{21} are denoted as AI and to fragment M_2 with added C^{12},C^{13} as AII.

Basic partitions PI and PII do consist of the following sets:

fragment M_1 (PI): $\{C^{10},C^{11},C^{12},C^{13}\}, \{C^{20},C^{21}\}$

fragment M_2 (PII):

$$\{C^{20},C^{21},C^{22},C^{23}\}, \{C^{30},C^{31},C^{32},C^{33}\},$$
$$\{C^{40},C^{41},C^{42},C^{43}\}, \{C^{12},C^{13}\}$$

The sets distinguished on the cut neighboring species have been added as denoted on the last positions of the list.

Sets κPI and κPII induced *by* particular species of PI and PII and κ are of the form

fragment M_1 (κPI): $\{C^{20},C^{21}\}, \{C^{12},C^{13}\}$

fragment M_2 (κPII):
$$\{C^{12},C^{13},C^{30}\}, \{C^{23},C^{40}\}, \{C^{33}\}, \{C^{20},C^{21}\}$$

It is useful to discriminate between sets induced which include elements C^{ij} of different carbons and those consisting of elements ascribed to the same carbon. Function κ is then "ramified" or "nonramified" under different C^i species as shown in the illustrating figures. For example, fragment M_2 does admit a new breakout into κ^*PII and $\kappa^{\bullet}$PII families:

$$\kappa^*\text{PII:}\quad \{C^{12},C^{13},C^{30}\}, \{C^{23},C^{40}\}$$

$$\kappa^{\bullet}\text{PII:}\quad \{C^{33}\}, \{C^{20},C^{21}\}$$

κ^*PI is an empty class, and $\kappa^{\bullet}$PI includes only sets: $\{C^{20},C^{21}\}$ and $\{C^{12},C^{13}\}$.

Since the cut neighboring induced sets such as $\{C^{12},C^{13}\}$ and $\{C^{20},C^{21}\}$ permanently appear in the structures used in this work, they are often included in numerical parts of the

formulas (7), (9), (12), and (14). To indicate exactly, $\kappa^{\bullet}$Pi families are distinguished for carbons not neighboring the cut and P′ denotes all sets of basic partitions PI and PII except of, $\{C^{20},C^{21}\}$ and $\{C^{12},C^{13}\}$. For example, the case considered assumes $\kappa^{\bullet}$PII consisting of only set $\{C^{33}\}$ and $\kappa^{\bullet}$PI to be empty if applied in the formulas just mentioned.

Continuing the example, we have two families of sets determined for particular fragments of hydrogen-suppressed structure:

$$\text{FI} \stackrel{\text{df}}{=} \text{PI} \cup \kappa\text{PI} \tag{5}$$

$$\text{FII} \stackrel{\text{df}}{=} \text{PII} \cup \kappa\text{PII}$$

Constituents induced in "*on*" and "*by*" modes are the following:

fragment M_1:   @FI $= \{\{C^{10},C^{11}\},\{C^{12},C^{13}\},\{C^{20},C^{21}\}\}$

fragment M_2:

$$@\text{FII} = \{\{C^{12},C^{13}\},\{C^{20},C^{21}\},\{C^{22}\},\{C^{23}\},\{C^{30}\},$$
$$\{C^{31},C^{32}\},\{C^{33}\},\{C^{40}\},\{C^{41},C^{42},C^{43}\}\}$$

It is easily seen that each of the sets of FI and FII can be fully covered by a number of mutually disjoint constituents. If a belongs to

$$\text{P} \stackrel{\text{df}}{=} \text{PI} \cup \text{PII}$$

then P*i*a ($i =$ I or II) or Pa denotes the partition of a by constituents. If a′ belongs to

$$\kappa^*\text{P} \stackrel{\text{df}}{=} \kappa^*\text{PI} \cup \kappa^*\text{PII}$$

then κ^*Pa′ or κ^*P*i*a′ (for $i =$ I, II) denotes the family of all constituents included in a. The same is valid for $\kappa^{\bullet}$Pa′ consisting of constituents covering a′ being the member of

$$\kappa^{\bullet}\text{P} \stackrel{\text{df}}{=} \kappa^{\bullet}\text{PI} \cup \kappa^{\bullet}\text{PII}$$

These denotations are used in the formulas stated below to calculate useful structural characteristics such as overlapping and splitting.

All the partitions indicated have been induced due to some one-to-one correspondences inherent in the chemical structure. Having the possibility to compare, e.g., F*i* and @F*i* (for $i =$ I, II) families certain molecular characteristics that are of chemical significance can be obtained. Moreover, it turns out to be useful to compare, e.g., family {A*i*} consisting of only the set indicated and F*i* ($i =$ I, II). Certain similarity−dissimilarity relationships of these entities can even be calculated. To carry out this facility, the set families under consideration are to be transformed into relations. It is especially easy to perform in the case of constituents. For example, constituent $\{C^{10},C^{11}\}$ can be changed into a relation consisting of all pairs ($C^{ij}, C^{ij'}$) where $i = 1, j = 0, 1$, and $j' = 0, 1$. There are exactly 2^2 such pairs. The number of elements in all relations appearing after conversions of the sets of @FI is $|f@\text{FI}| = 12$. Similarly for constituents of @FII, the cardinality of the corresponding relations is $|f@\text{FII}| = 26$.

Another transformation into relations is used for sets of families FI and FII when the number of constituents included

SPLITTING AND OVERLAPPING IN MOLECULAR SYSTEMS

J. Chem. Inf. Comput. Sci., Vol. 40, No. 4, 2000 **1021**

in the set to be converted is more than 1. In particular, set

$$a \stackrel{df}{=} \{C^{20}, C^{21}, C^{22}, C^{23}\}$$

is an element of FII, and it is covered by three constituents: $\{C^{20}, C^{21}\}$, $\{C^{22}\}$, and $\{C^{23}\}$. The total relation for set a consists of all pairs $(C^{2j}, C^{2j'})$ for $j, j' = 0, ..., 3$. There are 4^2 such pairs. This relation is however reduced by subtracting all pairs arising from transformations of the included constituents. Thus, the pairs (C^{23}, C^{23}) and (C^{20}, C^{21}) are to be removed, but pair (C^{21}, C^{22}) remains in the relation transformed. The number of elements in such a reduced relation of set a is $4^2 - 2^2 - 1^2 - 1^2$. All sets of FI and FII which consist only of one constituent are transformed into relations as these constituents. Thus, the number of its elements (cardinality) is the square of the number of elements of the set transformed. The cardinality of the whole relation for FI is thus the following: $|fFI| = (4^2 - 2^2 - 2^2) + 2^2 + 2^2 = 16$. The second case is also easy to calculate: $|fFII| = 41$. Consider the cardinality of the common part of the relations fFII and f@FII, i.e., the so-called factor of similarity FS(fFII,f@FII). Since the pairs belonging to the same constituent of one of at least two constituents included in a "many-constituent" set of fFII are removed as mentioned above, the relations on such sets cannot be members of the common part just considered. Therefore, only constituents of FII yield required relations. Hence, the factor of similarity is easy to estimate: FS(fFII,f@FII) $= 2^2 + 2^2 + 1^2$. It is easy to check that FS(fFI,f@FI) $= 8$. Using the general theory of resemblance and set spectra,[9] the factors of dissimilarity should be yet determined on the known values just derived.

$$FD(fFI, f@FI) = |fFI| + |f@FI| - 2FS(fFI, f@FI) \quad (6)$$

$$FD(fFII, f@FII) = |fFII| + |f@FII| - 2FS(fFII, f@FII)$$

Moreover, as prompted by the general theory[9] the sum of two factors of similarity, FS(fFI,f@FI) + FS(fFII,f@FII), is also a factor of similarity, and the summation of factors of dissimilarity, FD(fFI,f@FI) + FD(fFII,f@FII), results in a factor of dissimilarity too.

The procedures for using other than butene alkenes (acyclic on the cut) on the grounds of the pattern presented are clear. Some general formulas have even been derived to ease the calculations:

$$FS(fFi, f@Fi) = 8 + \sum_{a' \in \kappa^* Pi} |a'|^2 \quad (7)$$

$$FD(fFi, f@Fi) = 16n_i - 4 +$$
$$\sum_{a' \in \kappa^* Pi} |a'|^2 - \sum_{a' \in \kappa^* Pi} |a'|^2 - \sum_{a' \in \kappa^* Pi} \sum_{b' \in \kappa^* Pia'} |b'|^2$$

where n_i denotes the number of carbons in the fragment for $i = $ I, II.

Overlapping in chemical molecules can thus be determined

$$o \stackrel{df}{=}$$
$$\frac{FD(fFI, f@FI) + FD(fFII, f@FII)}{FD(fFI, f@FI) + FD(fFII, f@FII) + FS(fFI, f@FI) + FS(fFII, f@FII)} \quad (8)$$

and in particular, for hydrocarbons under a butene pattern, we have

$$o = \frac{16n - 8 + \sum_{a' \in \kappa^* P} |a'|^2 - \sum_{a' \in \kappa^* P} \sum_{b' \in \kappa^* Pa'} |b'|^2 - \sum_{a' \in \kappa^\bullet P} |a'|^2}{16n + 8 + \sum_{a' \in \kappa^* P} |a'|^2 - \sum_{a' \in \kappa^* P} \sum_{b' \in \kappa^* Pa'} |b'|^2} \quad (9)$$

This is another example of fuzzy set usage applied already in ref 10 and presented more generally in ref 8.

Yet one important quantitative characteristic called splitting[8] can be introduced. For this purpose the transformations of the sets AI and AII in the relations are necessary. The procedure is as above-applied. First, the set of all possible pairs $(C^{ij}, C^{i'j'})$ ($i, i' = 1$ and $j, j' = 0, ..., 3$ for AI, and i, $i' = 2, 3, 4$ and $j, j' = 0, ..., 3$ for AII) is distinguished. Then the pairs $(C^{ij}, C^{i'j'})$ such that C^{ij} and $C^{i'j'}$ belong to the same constituent are removed. The numbers of elements in the relations f{AI} and f{AII} determined in this way for the butene case are as follows:

$$|f\{AI\}| = (4 + 2)^2 - (2^2 + 2^2 + 2^2)$$

$$|f\{AII\}| = (12 + 2)^2 - |f@FII|$$

We are now asking about the factors of similarity FS(f{AI},fFI) and FS(f{AII},fFII). The task is to find and count all the common elements of the relations involved. One can easily infer from the nature of considered relations that such a factor should include pairs for, e.g., relation fFII, which do not belong to the common elements of fFII and f@FII; i.e., they are not elements of intersection fFII $\cap$ f@FII $= $ FS(fFII,f@FII).

Hence, the procedure leads to the relationships valid for the kind of cases considered here

$$FS(f\{AI\}, fFI) = |fFI| - FS(fFI, f@FI) \quad (10)$$

$$FS(f\{AII\}, fFII) = |fFII| - FS(fFII, f@FII)$$

and, respectively, to the factors of dissimilarity

$$FD(f\{Ai\}, fFi) = |f\{Ai\}| + |fFi| - 2FS(f\{Ai\}, fFi)$$
$$\text{for} \quad i = \text{I, II} \quad (11)$$

For the compounds under question

$$FS(f\{Ai\}, fFi) = 4 + 16n_i -$$
$$\sum_{a \in Pi} \sum_{b \in Pia} |b|^2 + \sum_{a' \in \kappa^* Pi} |a'|^2 - \sum_{a' \in \kappa^* Pi} \sum_{b' \in \kappa^* Pia'} |b'|^2$$

$$FD(f\{Ai\}, fFi) = (2 + 4n_i)^2 - 16n_i + 4 -$$
$$\sum_{a' \in \kappa^* Pi} |a'|^2 + \sum_{a' \in \kappa^* Pi} \sum_{b' \in \kappa^* Pia'} |b'|^2 + \sum_{a' \in \kappa^\bullet Pi} |a'|^2 \quad (12)$$

Using the summation principle for factors mentioned above, the splitting is defined by the following formula:

$$s \stackrel{df}{=}$$
$$\frac{FD(f\{AI\}, fFI) + FD(f\{AII\}, fFII)}{FD(f\{AI\}, fFI) + FD(f\{AII\}, fFII) + FS(f\{AI\}, fFI) + FS(f\{AII\}, fFII)} \quad (13)$$

$$s = \frac{(2 + 4n_\text{I})^2 + (2 + 4n_\text{II})^2 - 16n + 8 - \sum\limits_{a' \in \kappa^*P} |a'|^2 + \sum\limits_{a' \in \kappa^*P} \sum\limits_{b' \in \kappa^*Pa'} |b'|^2 + \sum\limits_{a' \in \kappa^\bullet P} |a'|^2}{(2 + 4n_\text{I})^2 + (2 + 4n_\text{II})^2 + 8 - \sum\limits_{a \in P'} \sum\limits_{b \in Pa} |b|^2 + \sum\limits_{a' \in \kappa^\bullet P} |a'|^2} \tag{14}$$

For the particular case of hydrocarbons illustrated by the butene example, splitting can be expressed as
where n_I and n_II are the numbers of carbon species in fragments I and II and $n = n_\text{I} + n_\text{II}$.

Do the algorithms pursued above really afford chemical information? This is subject for testing in the next sections.

FUZZY SET CONTROL OF ORGANIC REACTIONS

If a hydrocarbon has substituted one hydrogen, e.g., by a halide atom X (at a carbon α position), then removing X yields, as usual, an olefin. Depending on the surrounding carbons (at the β position) and their hydrogens, alternative alkenes can arise. In this paper, both α and β carbons are exclusively regarded on the acyclic parts of hydrocarbons. The well-known empirical Saytzeff rule[11] states that the double bond goes mainly toward the most highly alkyl substituted carbon (cf. ref 12, p 998). The opposite direction is called Hofmann orientation. Some coefficient is ascribed to each β-carbon with hydrogen (denoted β-C) in this way, i.e.

$$\text{Coeff}(\beta\text{-C}) \overset{\text{df}}{=} \text{number of substituted alkyl groups} \tag{15}$$

A first glance at this rule reveals that the contribution of each alkyl at the β-carbon is the same and equals 1. Thus, the coefficient (15) is a summation of roughly counted contributions of alkyl substituents independently of their size and structure. Some more sophisticated approaches have been used in this paper to take into account the nature of the alkyl substituents. The contributions of alkyls or other hydrocarbon substituents were evaluated in such a way to grasp differences between features of participating fragments.

The regularity considered, as many other ones in chemistry, uses the directly observational local properties coming from iconic chemical languages. Writing it more formally, one can note that *the yield toward β-C is less than the yield toward β-C′ if and only if Coeff(β-C) < Coeff(β-C′)* (**1**). Additionally, it is of interest to know whether the predominant yield of olefin in E1 eliminations[12] depends only on the local arrangement of alkyl branching or the whole self-assembly composition is involved? To answer this question, some general characteristics of SAC spaces have been found to show the contribution of all the structure features to the considered chemical reactions.

For this purpose, substituted hydrocarbons (acyclic) from C_4 to C_7 have been investigated to obtain their overlapping and splitting (Table 1) according to the example shown for 1-butene. As for the hydrocarbons C_4–C_7, the ratio o/s is greater for the Saytzeff orientation than for the Hofmann one. Besides, the greater value of o/s is reasonably observed in the direction of β-carbons with equal numbers of alkyl substituents but of different types (cf., e.g., compounds 7.5 and 7.12 in Table 1). It is supposed that the ability to create olefin in all the directions toward β-carbons can be assessed by the summation of o/s ratios assigned to each β-carbon

instead of the summation of alkyl substituent contributions as in (15). A conjecture has been framed that this summation called the *S*-coefficient (ScC) is capable of estimating how the related system will behave, e.g., in the E1 solvolysis reactions. This is expressed below in statements *A* and *B*. Such an assessment is very useful since the quantitative characteristic considered remains in relation to many phenomena, the meaning (and the term "*S*-coefficient") of which is discussed in the next section. For example, the stability order of alkyl-substituted free radicals is the same as for carbocations in certain cases (cf. ref 12, p 751). So, when the *S*-coefficient is involved in a reasoning, then it is not automatically assumed to be assigned to a fixed reaction mechanism, e.g., to a fully acting carbocation. Moreover, each carbon has its ascribed *S*-coefficient consisting of the summation of o/s ratios just mentioned. Interesting relevance of these parameters in the course of organic reactions will be shown.

First, a temptation arises to generalize the Saytzeff rule to the form *the predominant yield of olefin in E1 eliminations is toward the β-carbon having the greater S-coefficient* (**2**). Let us note that for all hydrocarbons C_4–C_7 having alternative β-carbons, the ratio of overlapping and splitting (o/s) indicates the Saytzeff orientation by its greater value and complies with the common organic chemist conceiving of the effect of carbon chain influence in the case of equally substituted β-carbons. However, compound $Me(CH_2)_3$-$CHXCHMe_2$ (8.2, Table 2) has an o/s ratio greater toward the secondary β-carbon compared with the tertiary β-carbon. Some results for C_8 hydrocarbons are given in Table 2. It is clearly seen that the *S*-coefficient distribution is reasonably obeyed in most of these cases too, even for the complex hierarchy of primary, secondary, and tertiary β-carbons under a fixed α position in one compound.

There are even more sophisticated features of the data given in Table 1. If an alkane is substituted in position 2 and has a methyl group at this position, then the summation of primary carbon *S*-coefficients is always less than the *S*-coefficient of the third β-carbon, so it easy to check that within the compounds of Table 1 (5.4, 6.5, 6.6, 7.7, 7.8, 7.12) the following inequality holds:

$$\text{Sc}\beta\text{-}p\text{-C} + \text{Sc}\beta\text{-C}^1 < \text{Sc}\beta\text{-C}^3 \tag{16}$$

where p denotes primary carbon.

This rule is still valid for compound 8.1, but it clearly fails in the case of 8.5 since

$$\text{Sc}\beta\text{-}p\text{-C} + \text{Sc}\beta\text{-C}^1 = 1.94346 > 1.01247 = \text{Sc}\beta\text{-C}^3 \tag{17}$$

for compound $Me_3CCH_2CMe_2X$ (Table 2, 8.5).

Though the "double bond" should go mainly toward the most highly alkyl-substituted β-carbon, which coincides with the hierarchy of *S*-coefficients, a perturbation in other regularities gives rise to suspect changes in the system

Table 1. Splitting and Overlapping of Alkenes (C_4-C_7)

hydrocarbon derivative	n-ene	o	s	o/s	β-C	α-C
4.1. MeCH$_2$CHMeX	1	0.78205	0.81043	0.96499	0.96499	1.96463
	2	0.77500	0.77528	0.99964	1.96463	
5.1. Me(CH$_2$)$_2$CHMeX	1	0.82292	0.84384	0.97520	0.97520	1.98805
	2	0.81633	0.80597	1.01285	2.02570	
5.2. MeCH$_2$CHXCH$_2$Me	2	0.81633	0.80597	1.01285	1.98805	2.0257
	3	0.81633	0.80597	1.01285	1.98805	
5.3. Me$_2$CHCHMeX	1	0.81633	0.84337	0.96793	0.96793	1.98031
	2	0.81373	0.80377	1.01238	2.95914	
5.4. MeCH$_2$CMe$_2$X	p	0.82000	0.84242	0.97338	0.97338	2.95914
	1	0.82000	0.84242	0.97338	0.97338	
	2	0.81373	0.80377	1.01238	1.98031	
6.1. Me(CH$_2$)$_3$CHMeX	1	0.85088	0.86858	0.97962	0.97962	1.99030
	2	0.84483	0.83590	1.01068	2.03942	
6.2. Me(CH$_2$)$_2$CHXCHMe	2	0.84483	0.83590	1.01068	1.99030	2.03942
	3	0.84483	0.82123	1.02874	2.03942	
6.3. Me$_2$CHCH$_2$CHMeX	1	0.84483	0.86831	0.97295	0.97295	1.97715
	2	0.83898	0.83548	1.00420	2.03098	
6.4. Me$_2$CHCHXCH$_2$Me	2	0.83898	0.83548	1.00420	1.97715	2.03098
	3	0.84167	0.81972	1.02678	2.97996	
6.5. Me(CH$_2$)$_2$CMe$_2$X	p	0.84746	0.86777	0.97659	0.97659	2.97996
	1	0.84746	0.86777	0.97659	0.97659	
	2	0.84167	0.81972	1.02678	2.03098	
6.6. Me$_2$CHCMe$_2$X	p	0.84167	0.86749	0.97023	0.97023	2.96555
	1	0.84167	0.86749	0.97023	0.97023	
	2	0.83871	0.81818	1.02509	2.96555	
6.7. MeCH$_2$CHMeCHMeX	1	0.84483	0.86831	0.97295	0.97295	1.98139
	2	0.84167	0.83463	1.00844	2.99347	
6.8. MeCH$_2$CMeXCH$_2$Me	p	0.84746	0.86777	0.97659	0.97659	2.99347
	2	0.84167	0.83463	1.00844	1.98139	
	3	0.84167	0.83463	1.00844	1.98139	
7.1. Me(CH$_2$)$_4$CHMeX	1	0.87121	0.88707	0.98212	0.98212	1.98837
	2	0.86567	0.86029	1.00625	2.03477	
7.2. Me(CH$_2$)$_3$CHXCH$_2$Me	2	0.86567	0.86029	1.00625	1.98837	2.03477
	3	0.86567	0.84167	1.02852	2.05704	
7.3. Me(CH$_2$)$_2$CHX(CH$_2$)$_2$Me	3	0.86567	0.84167	1.02852	2.03477	2.05704
	4	0.86567	0.84167	1.02852	2.03477	
7.4. Me$_2$CH(CH$_2$)$_2$CHXMe	1	0.86567	0.88690	0.97606	0.97606	1.97636
	2	0.86029	0.86004	1.00030	2.02283	
7.5. Me$_2$CHCH$_2$CHXCH$_2$Me	2	0.86029	0.86004	1.00030	1.97636	2.02283
	3	0.86029	0.84134	1.02253	2.04828	
7.6. Me(CH$_2$)$_2$CHXCHMe$_2$	2	0.86232	0.84067	1.02575	2.98307	2.04828
	3	0.86029	0.84134	1.02253	2.02283	
7.7. Me(CH$_2$)$_3$CMe$_2$X	p	0.86765	0.88657	0.97866	0.97866	2.98307
	1	0.86765	0.88657	0.97866	0.97866	
	2	0.86232	0.84067	1.02575	2.04828	
7.8. Me$_2$CHCH$_2$CMe$_2$X	p	0.86232	0.88640	0.97284	0.97284	2.96568
	1	0.86232	0.88640	0.97284	0.97284	
	2	0.85714	0.84034	1.02000	2.04000	
7.9. Me$_2$CHCHXCHMe$_2$	2	0.85714	0.84034	1.02000	2.96568	2.04000
	3	0.85714	0.84034	1.02000	2.96568	
7.10. Me$_2$CHCHMeCHMeX	1	0.86029	0.88674	0.97018	0.97018	1.96772
	2	0.85714	0.85926	0.99754	2.99359	
7.11. Me$_2$CHCMeXCH$_2$Me	p	0.86232	0.88640	0.97284	0.97284	2.99359
	2	0.85714	0.85926	0.99754	1.96772	
	3	0.85915	0.83966	1.02321	2.96889	
7.12. MeCH$_2$CHMeCMe$_2$X	p	0.86232	0.88640	0.97284	0.97284	2.96889
	1	0.86232	0.88640	0.97284	0.97284	
	2	0.85915	0.83966	1.02321	2.99359	
7.13. Me$_3$CCH$_2$CHMeX	1	0.86232	0.88640	0.97284	0.97284	1.97038
	2	0.85714	0.85926	0.99754	0.99754	
7.14. MeCH$_2$CH(CH$_2$Me)CHMeX	1	0.86567	0.88690	0.97606	0.97606	1.97932
	2	0.86232	0.85952	1.00326	3.00978	
7.15. MeCH$_2$C(CH$_2$Me)XCH$_2$Me	s	0.86232	0.85952	1.00326	1.97932	3.00978
	2	0.86232	0.85952	1.00326	1.97932	
	3	0.86232	0.85952	1.00326	1.97932	
7.16. MeCH$_2$CHMeCH$_2$CHMeX	1	0.86567	0.88690	0.97606	0.97606	1.97636
	2	0.86029	0.86004	1.00030	2.02605	
7.17. MeCH$_2$CHMeCHXCH$_2$Me	2	0.86029	0.86004	1.00030	1.97636	2.02605
	3	0.86232	0.84067	1.02575	3.00767	
7.18. Me(CH$_2$)$_2$CMeXCH$_2$Me	p	0.86765	0.88657	0.97866	0.97866	3.00767
	2	0.86232	0.85952	1.00326	1.97932	
	3	0.86232	0.84067	1.02575	2.02605	
7.19. Me(CH$_2$)$_2$CHMeCHMeX	1	0.86567	0.88690	0.97606	0.97606	1.97932
	2	0.86232	0.85952	1.00326	3.00767	

Table 2. Splitting and Overlapping of Certain Eight-Carbon Alkenes

hydrocarbon derivative	n-ene	o	s	o/s	β-C	α-C
					S-coefficient	
8.1. $Me(CH_2)_4CMe_2X$	p	0.88312	0.90090	0.98026	0.98026	2.98105
	1	0.88312	0.90090	0.98026	0.98026	
	2	0.87821	0.86054	1.02053	2.04723	
8.2. $Me(CH_2)_3CHXCHMe_2$	2	0.87821	0.86054	1.02053	2.98105	2.04723
	3	0.87662	0.85358	1.02670	2.04487	
8.3. $Me(CH_2)_3CMeX\ CH_2Me$	p	0.88312	0.90090	0.98026	0.98026	3.00885
	2	0.87821	0.87895	0.99915	1.97746	
	3	0.87821	0.85309	1.02944	2.04761	
8.4. $Me(CH_2)_2CMeXCHMe_2$	p	0.87821	0.90079	0.97493	0.97493	3.00775
	2	0.87500	0.85987	1.01759	2.96745	
	3	0.87342	0.86032	1.01523	2.00690	
8.5. $Me_3CCH_2CMe_2X$	p	0.87500	0.90045	0.97173	0.97173	2.95593
	1	0.87500	0.90045	0.97173	0.97173	
	2	0.87037	0.85965	1.01247	1.01247	

Table 3. S-Coefficients and Parameters of Solvolysis (in n-Butyl Cellosolve)[a]

hydrocarbon derivative	rate of reaction (h^{-1})	S-coefficient of α-C	experimental p/Z olefin ratio yields	p-olefin part of the S-coefficient of α-carbon
1. $MeCH_2CMe_2Br$	0.387	2.95914	0.27	0.65788
2. $MeCH_2CMeBrCH_2Me$	0.410	2.99347	0.036	0.32624
3. Me_2CHCMe_2Br	0.197	2.96555	0.22	0.65433
4. $Me(CH_2)_2CMe_2Br$	0.297	2.97996	0.41	0.65544
5. $Me_2CHCH_2CMe_2Br$	0.697	2.96568	0.7	0.65607
6. $Me_3CCH_2CMe_2Br$	4.71	2.95593	4.26	0.65748

[a] Experimental data: (a) rates of solvolysis, compounds 2 and 3, see ref 13, p 3608, compounds 1, 4, 5, and 6, see ref 15, p 3611. (b) p/Z ratios of olefin yield: compounds 2 and 3 see ref 13, p 3609, compounds 1, 4, 5, and 6 see ref 15, p 3611.

behavior. Really, the yield ratios of Hofmann and Saytzeff products are inverted as well as the inequality (16) compared with (17). Reference 13 brings the result of an E1 elimination, giving 4.7 times more of the Hofmann product than the Saytzeff one. This fact is the reason for a more thorough insight into the experimental results as well as the formulation of the yet more adequate Saytzeff rule, which is discussed in the last section. The classic explanation of E1 eliminations employs the carbocation concept to understand the reaction mechanism.[12] This is tentatively taken into account to formulate and test the following statement: (A) For two hydrocarbons (C) and (C′) with leaving groups at α carbons C and C′, respectively, if S-coefficients ScC and ScC′ are related by the inequality ScC < ScC′, then the rates of E1 solvolysis are in the opposite relationship, i.e., rate(C) > rate(C′).

Another statement to be tested concerns the 1-olefin and 2-olefin yields for respective hydrocarbons. The quantitative criteria have been established for Saytzeff and Hofmann orientation in Table 1. However, the ratio of 1/2 for olefin yields is a more subtle characteristic functioning within the predominant orientation mentioned. Olefin formation can have one or more alternative orientations toward primary carbons. It is supposed that the primary olefin contribution can be assessed by the respective β-carbon S-coefficients summarized. For example, compound 5.4 (Table 1) has an S-coefficient of 2.95914, and the sum of Hofmann-oriented S-coefficients is 1.94676. Thus, the p-olefin part of the total S-coefficient (denoted p-Sc(C)) equals 0.65788. The letter p is used to indicate that the generally double binding orientation can go to a primary carbon which is not necessarily on position 1. Other orientations of double binding directed to the higher order alkyl-substituted β-carbons are Saytzeff (Z) orientations.

Table 4. Fuzzy Control of Solvolysis (Table 3)[a]

	1	2	3	4	5	6
1	0	+	+	−	−	−
2	−	0	+	+	+	+
3	+	−	0	+	+	+
4	+	−	−	0	+	+
5	−	+	−	+	0	+
6	+	+	+	+	+	0

[a] Statement A, lower diagonal; statement B, upper diagonal.

The following statement will be checked: (B) For two substituted hydrocarbons (C) and (C′) where C and C′ are α carbons (like in Tables 1 and 2), if p-olefin parts of the S-coefficients are related by the inequality p-Sc(C) < p-Sc(C′), then the p/Z ratios of olefin yields in E1 solvolysis remain in the same relationship; i.e., the ratio p/Z for hydrocarbon (C) is less than the ratio p/Z for hydrocarbon (C′).

How can such theses as A and B underlie fuzzy set evaluations?[14] If we have a collection of compounds, say **X**, then the ratio of compounds fulfilling the statement in question is easy to assess. If there is a family of such collections, then the ratios determined can serve as grades of membership. Finally, a fuzzy set is always involved in this way, and depending on its evaluations, one can measure how far a condition like A or B is influential in the chemical process. When the fulfillment for a collection **X** under a fixed statement is 100% (i.e., the ratio considered equals 1), then it is said that **X** belongs to the core of the fuzzy set.[14] A case of this kind has been found for certain allene derivatives as presented in Table 14. Below-stated examples give practical presentation of the needed assessments for computational and experimental results exposed for concrete cases of bromide derivatives in Table 3.

Table 5. *S*-Coefficient and Parameters of Solvolysis (in Anhydrous Acetic Acid)[a]

hydrocarbon derivative	rates of E1 solvolysis (h^{-1})	*S*-coefficient of α-C	experimental p/Z olefin ratio yields	p-olefin part of the *S*-coefficient of α-C
1. MeCH$_2$CHMe(OBs)	0.464	1.96463	0.11	0.49118
2. Me(CH$_2$)$_2$CHMe(OBs)	0.443	1.98805	0.19	0.49053
3. Me$_2$CHCH$_2$CHMe(OBs)	0.407	1.97715	0.25	0.49210
4. Me$_3$CCH$_2$CHMe(OBs)	0.950	1.97038	0.32	0.49373

[a] Experimental data, cf ref 16, p 3614.

Table 6. Fuzzy Control of Solvolysis (Table 5)[a]

	1	2	3	4
1	0	−	+	+
2	+	0	+	+
3	+	−	0	+
4	−	+	+	0

[a] Statement *A*, lower diagonal; statement *B*, upper diagonal.

Table 7. *S*-coefficient and Parameters of Solvolysis (in 80% ethanol)[a]

hydrocarbon derivative	rate of reaction (h^{-1})	*S*-coefficient of α-C
1. Me$_3$CCl	0.033	2.89895
2. MeCH$_2$CMe$_2$Cl	0.055	2.95914
3. Me$_2$CHCMe$_2$Cl	0.029	2.96555
4. Me$_3$CCMe$_2$Cl	0.040	1.93965
5. MeCH$_2$CMe$_2$CMe$_2$Cl	0.187	1.94346

[a] Experimental data, cf. ref 16, p 3627.

Table 8. Fuzzy Control of Solvolysis (Table 7)[a]

	1	2	3	4	5
1	0				
2	−	0			
3	+	+	0		
4	+	−	+	0	
5	+	+	+	−	0

[a] Statement *A*.

Table 9. Rates of Halogen Eliminations

substrate	temp (°C)	rate × 10^5 of elimination reaction	*S*-coefficient of α-C
1. MeCH$_2$Br	55	1.6	0.80000
2. Me$_2$CHBr	25	0.237	1.86488
3. Me$_3$CBr	25	4.17	2.89896
4. n-PrBr	55	5.3	0.93244
5. Me$_2$CHCH$_2$Br	55	8.5	0.96632

Table 10. Fuzzy Control of Eliminations (Table 9)[a]

	1	2	3	4	5
1	0				
2	+	0			
3	−	−	0		
4	−	+	+	0	
5	−	+	+	−	0

[a] Statement *A*.

Let us indicate by a plus sign that condition *A* is fulfilled for any pair of different compounds of Table 2 and by a minus sign that *A* does not hold (Table 4). The same has been done for p/Z ratios of olefin yields to check condition *B*. Thus, the results of this estimation are as follows.

Since condition *A* is fulfilled in 60% of the cases considered in Table 3, it can be said the condition *A* controls the rates of E1 solvolysis reactions (of Table 3) in the degree 0.6. Thus, the fuzzy situation has been determined for an organic reaction first, by the use of fuzzy set pseudometrics (8) and (13), and for the second by checking conditions *A* and *B* fulfillment in the series of experiments.

Another example of experimental data analysis (in the fuzzy mode) has been shown for alkyl brosylates in Table 5. The corresponding complement exhibiting the cases fulfilling conditions *A* and *B* is presented in Table 6. Fuzzy control of the rates of solvolysis reactions is in degree 0.67 in this case, and p/Z ratios of olefin yields are subordinate to statement *B* with the degree 0.83.

Yet one experiment has been tested as suitable for the fuzzy estimation of condition *A* involvement for alkyl chlorides (Table 7). Associated Table 8 shows the degree of fuzzy control of the reactions considered equal to 0.7 under condition *A*.

The features of compounds which influence organic reactions are undoubtedly fuzzy in this sense that any particular property (as considered above) does contribute as usual to the mechanism of chemical reaction (in a collection of compounds) only to some extent. Pertaining grades of such contributions are high (0.6−0.83) taking into account the complexity of chemical processing considered. In particular, compound 6 (Table 3) reveals a sharp increase of solvolysis rate as well as p/Z ratio of the olefin yield. The increases are signalized by the lowest value of the *S*-coefficient (for α-carbon) and one of the greatest values of the p-olefin part in this coefficient. However, the scale of increase cannot be explained in this way since the respective coefficients are not distinguished so sharply.

Nevertheless, another glance at the tables unveils spectacular results. For example, compound 2 (Table 3) having an unusual decrease of p-olefin ratio in *S*-coefficient is accompanied by the great diminishing p/Z ratio of olefin yields. It is easy to find more examples of this kind.

Do the fuzzy set control indicators determined for particular properties described in the tables give rise to concludsions about system processing? Consider once more elimination rates for a number of hydrocarbon bromine derivatives (Table 9). It is of interest to estimate the fuzzy set value for the rates given (Table 10).

The grade of membership to fulfill statement A is 0.5 for the data calculated. Since with rising temperature, e.g., for 2, an increase of the reaction rate can be expected, the real fuzzy set value for this table should not exceed 0.4, much less than received above. It is often the case to encounter new diversifying phenomena in such circumstances. Indeed, the reactions just involved undergo an E2 mechanism (cf. ref 12, p 1004) of elimination as opposed to the fuzzy control tables previously explored.

GENERAL NOTES AND DISCUSSION

As for practical chemical recognition, the analyses performed show that neither classic formulations of the Saytzeff rule nor generalizations made by the use of the overlapping and splitting parameters embrace the cases considered above. However, such a general statement is possible to introduce it within the investigated examples, giving a more informative law than in the classic formulations. It should be noted that β-carbons (which are maximally three in a fixed case) can belong to the isomorphic alkyl substituents (e.g., primary methyl groups), or they represent alkyls of a different type (such as 5.4 or 7.11 in Table 1). β-Carbons which occur in the isomorphic alkyls (or other hydrocarbon substituents) are called twin β-carbons. Every β-carbon does or does not have a twin β-carbon which is respectively characterized by the number τ equal to 1 or 0. Let us consider an α-carbon. Then a characteristic value can be assigned to each β-carbon C. It is of the form

$$2^{\tau} Sc\beta\text{-}C \qquad (18)$$

This is principally the same symbol as used for the sum in (16) (when $\tau = 1$) since the formulas contain there additionally only the concrete position of the distinguished β-carbon. The Saytzeff rule for mechanisms of its applicability is easy to formulate in a new way using the characteristic value.

If C and C′ are β-carbons in a hydrocarbon, and τ and τ' are their twin numbers, then *the yield toward C is less than the yield toward C′ if and only if* $2^{\tau} Sc\beta\text{-}C < 2^{\tau'} Sc\beta\text{-}C'$ (**3**). One can easily check that such a formulated generalized rule covers all the cases taken into account in this work including the case of 8.5 from Table 2. The result suggests that the enormously large yield of 1-alkene is due to shifting of the electron density to the respective terminal methyl groups during the E1 elimination, which can be illustrated by an analogy to the "double wire" electric current intensification under constant voltage. As for the understanding of the *S*-coefficient (Sc) in the organic chemistry manner, it cannot be considered as a quantitative assessment of the stability of the whole carbocation (cf. ref 12, p 168) since, for example, *tert*-pentyl has a greater Sc value (Table 1, compound 5.4) than *tert*-butyl (Table 7, compound 1), which is claimed as the most stable in the reference mentioned. Other understandings of stabilities (e.g., the stability of electric charges at a carbocation) have been left in terms of probable but not employed here concepts. Taking into account the nature of the *S*-coefficient and the results of fuzzy set analyses involving experimental data, the evaluations performed do express rather the susceptibility of a carbon for its valence electron shifting toward the vicinity of interacting chemical species (particle). It is supposed that Sc values give the best correspondence to the reaction phenomena when a double bond is to be removed (as in electrophilic addition) or created (as in elimination reactions). Since only the "layout" of the carbon skeleton was involved in the calculations, it seems the role of the intramolecular system of carbon–carbon relationships can be greater than it is commonly thought regarding the phenomena considered. There is a surprising impression that no steric, ionic pair, hyperconjugation, or bulkiness explanations were necessary to reach an insight into the mechanisms of reaction hazards.

Table 11. ^{13}C Chemical-Shift Values and *S*-Coefficients (ppm) for Some Carbocations (cf. Ref 12, p 173)

ion	chemical shift	temp (°C)	*S*-coefficient
1. Et$_2$MeC$^+$	−139.4	−20	2.99347
2. Me$_2$(cyclopropyl)C$^+$	−86.8	−60	3.13436
3. Me$_2$CH$^+$	−125.0	−20	1.86488
4. Me$_2$EtC$^+$	−139.2	−60	2.95914
5. Me$_3$C$^+$	−135.4	−20	2.89896
6. Ph$_2$CH$^+$	−5.6	−60	0
7. Ph$_3$C$^+$	−18.1	−60	0
8. PhMe$_2$C$^+$	−61.1	−60	2.04518
9. PhMeCH$^+$	−40		1.03563

The relevance of *S*-coefficients to the experimentally found NMR chemical shifts for some carbocations can be tested. As shown in Table 11 alkyl-substituted hydrocarbons have a well-preserved hierarchy of *S*-coefficients related to negative values of chemical shifts; i.e., the greater the Sc value, the smaller the chemical shift evaluation. There are some disturbances in the case of aromatic substituents and cyclopropyl. For example, carbocations 2 and 4 have inverted expected hierarchy. This leads to the conjecture that cyclopropyl having double bond properties can underlie conjugation, diminishing the intrinsic Sc value of α-carbon as framed in another case mentioned above.

There are some further regularities of ^{13}C chemical shifts. For example, the one of double bond paired carbons with a greater Sc value has a greater ^{13}C NMR chemical shift (in the positive scale; cf. ref 17, p 192). Generally, however, the sensitivity of ^{13}C chemical shifts with respect to the changes of Sc values is not very high as the data of ref 17 show for dienes and allenes. It seems they cannot be directly related to the considered structural characteristics.

Heterolysis of hydrocarbons into carbocation and hydrogen anion does afford yet one evidence in support of *S*-coefficient influence on the reaction results. An energetic effect of such a heterolytic reaction should decrease with the growth of electronic shift susceptibility expressed by the coefficients considered. This is fully confirmed by the data in Table 12.

Compound Me$_2$CHCHMeSMe$_2$$^+$ cited in ref 12 (p 999) gave 91% of Saytzeff product in E1 elimination. The *S*-coefficients of 1 and 3 β-carbons are, respectively, 0.96793 and 2.95914 (Table 1, compound 5.3). Thus, it is justified to expect minor differences in yields of Hofmann and Saytzeff products in the case of nearer values of β-carbon *S*-coefficients. Really, 1,2-diphenyl-2-X-propanes (PhMeCXCH$_2$Ph, (ref 12, p 999) have β-carbon *S*-coefficients of 1.01751 for Hofmann and 1.06621 for Saytzeff orientations with yields of approximately 50% for each product. Consequently, the orientation conditioned by the conjugation effect would emerge to be of less importance though observable as some further analyses show. It is likely that the real value of the *S*-coefficients at nonterminal substituted propane carbons is slightly reduced in the view of double conjugation with aromatic rings. Reference 12 is commonly used here to facilitate the search for confirmation when necessary.
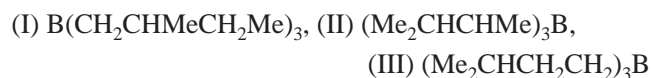
Using the just-mentioned conception of the *S*-coefficient (Sc), a conjecture on assessments of shifting can also be tested directly for reactions of other kinds such as addition to carbon–carbon multiple bonds when the Markovnikov rule is applied. A classic formulation of this rule claims that the more additive species goes to the carbon having more

Table 12. Heterolytic R−H → R$^+$ + H$^-$ Dissociation Energies $D(R^+ - H^-)$ (cf. Ref 12, p 171)

ion	S-coefficient	$D(R^+ - H^-)$ (kJ/mol)	ion	S-coefficient	$D(R^+ - H^-)$ (kJ/mol)
C_2H_5+	0.8	1158	Me_2CH^+	1.86488	1043
$H_2C=CH-CH_2^+$	0.92663	1070	Me_3C^+	2.89896	970.3

hydrogen. Another concept refers to the concept of stability.[12] The method pursued in this work does avoid ascribing physical meaning to the structural features as far as possible. Thus, the modified Markovnikov rule claims that the more additive species goes to the site with a lower value of the S-coefficient. For example, 2-methyl-4,4-dimethyl-2-pentene ($Me_3CCH=CHMe_2$ (ref 12, p 784, compound 8.5 in Table 2)) has an S-coefficient for the 2-carbon of 2.95593 and for the 3-carbon 1.01247. Since hydroboration runs easier to create a carbocation at a carbon with a greater S-coefficient (in accord with the new rule), the prevailing yield (98%) has been received for the product $Me_3CCH(BH_2)CHMe_2$. When another compound was hydroborated (i.e., i-PrCH=CHMe, ref 12, p 785), the closer values of S-coefficients (for the 2-carbon, 1.97715; for the 3-carbon, 2.03098; cf. compounds 6.3 and 6.4 in Table 1) were followed by 57% yield of hydroboron derivative with boron added at the 2-carbon and 43% at the 3-carbon. The predominant products have been obtained owing to more suitable electrophilic addition (by boron) at the carbon with a lower S-coefficient. Thus, the clasic Markovnikov rule has been extended to a more abstract and general form since neighboring carbons involved were of the same order under the count of alkyl substituents (or hydrogens). Sc values are not also ascribed to carbocation charge or structure stabilities, though their mathematical sense does not deny it.

Is it really so that the more positive species tends predominantly to the carbon having a minimal S-coefficient? Certain heated, nonterminal boranes give a mixture consisting of the following compounds (ref 12, p 1088):
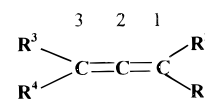
(I) $B(CH_2CHMeCH_2Me)_3$, (II) $(Me_2CHCHMe)_3B$,

(III) $(Me_2CHCH_2CH_2)_3B$

Boron-substituted carbons have the following S-coefficients:

ScI = 0.97338, ScII = 1.98031, ScIII = 0.96793
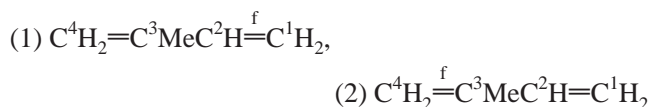(cf. compounds 5.3 and 5.4 in Table 1)

The comparisons by means of the number of hydrogens (as in Saytzeff and Markovnikov rule classic formulations) are not possible for terminal alkyl carbons in I and III. Therefore, S-coefficients turn out more useful in this case. Boron, as in the electrophilic addition, goes predominantly (when competing with hydrogen) to the carbon with a minimal Sc value in hydro, dialkylboro interchange rearrangements. The yields follow the expectations given by the inverted hierarchy of S-coefficients: I, 40%; II, 1%; III, 59%. Another example of rearrangements caused by S-coefficient distribution is exhibited by double bond migrations in electrophilic substitutions. The reaction runs in two steps: first hydrogen cation H$^+$ tends to the double bond carbon with a lower S-coefficient. Thus, a carbocation charged at the carbon with a greater Sc value appears to give rise to the double bond creation in accord with rule **3**. Double bonding is moved, as usual, from its position in this way. For example, due to this mechanism 1-decene gave 2-, 3-,

Figure 7. A scheme for substituted allene.

Table 13

isoprene	ScC^4	ScC^3	ScC^2	ScC^1
1	0.95351	2.90880		
2			0.95351	0.95351

and 4-decenes (ref 12, p 583) since S-coefficients are increased from the ends to the middle in the following order: 0.98622, 1.98373, 2.01114, 2.03991, 2.05745.

A more difficult challenge is afforded by electrophilic addition to conjugated dienes. To determine S-coefficients in this case, one double bond is fixed while double bonding is alternated at the carbon for which the required coefficient is to be found. For example, hydrohalogenation of isoprene ($CH_2=CMeCH=CH_2$) by HCl can combinatorially result in a number of products of addition. The experimental result is however highly selective, and reliable indicators would be useful to foresee the reaction products. Double bonding can be fixed (f) as follows:
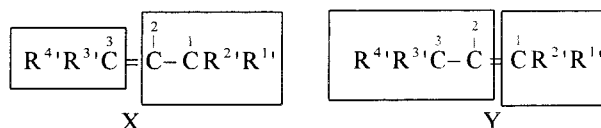
(1) $C^4H_2=C^3MeC^2H\overset{f}{=}C^1H_2$,

$$(2)\ C^4H_2\overset{f}{=}C^3MeC^2H=C^1H_2$$

Evaluations of S-coefficients for particular carbons resulted in the data shown in Table 13.

There are two products observed: $Me_2CClCH=CH_2$ and $Me_2C=CHCH_2Cl$ (ref 12, p 752). The first one is obvious as H$^+$ attack on "$C^4=C^3$" pair site is expected to give this product. The second one is supposed to arise due to the diminishing ScC^2 value in view of conjugation and following double bond rearrangement. Otherwise, $CH_2=CMeCH_2CH_2$-Cl should be observable as the result of addition HCl.

The same method has been applied for phenyl-substituted butadiene $PhC^4H=C^3HC^2H=C^1H_2$. S-coefficients have the following values: ScC^1 = 1.02438, ScC^2 = 2.06664, ScC^3 = 2.06940, ScC^4 = 1.04457. Hydrogen goes to the carbon with the smallest value of the S-coefficient at the end of acyclic hydrocarbon, giving $PhCH=CHCHClCH_3$. The last compound is reported (ref 12, p 752) to appear as only one product though in view of conjugation and reducing ScC^4 also $PhCH_2CHClCH=CH_2$ could be predicted.

Free radical addition of hydrogen bromide to acyclic olefins has anti-Markovnikov orientation (ref 12, p 758), which would indicate that Br$^•$ radical predominantly goes to the carbon with a lower S-coefficient (or equivalently: it does not go to the carbon with a greater S-coefficient). This method of hydrohalo addition for other halogens is difficult or never observed. Radical mechanisms in addition to allenes are yet more capricious. For example, certain radicals such as Br$^•$, p-$CH_2C_6H_4SO_2^•$, and $C_6H_5S^•$ reveal discrepancies in the attack on terminal and central carbons of nonsubstituted

1028 *J. Chem. Inf. Comput. Sci., Vol. 40, No. 4, 2000*

SCHULZ

Figure 8. Cuts for for the allene structure shown in Figure 7 where $R^{i'}$ denotes hydrocarbon substituents with suppressed hydrogen.

Table 14. S-Coefficients of Allene Carbons as Shown in Figure 7

	R^1	R^2	R^3	R^4	ScC^1	ScC^2	ScC^3
1	H	H	H	H	0.92663	0.93244	0.92663
2	H	H	H	methyl	1.01032	0.98003	1.93274
3	H	H	H	ethyl	1.00303	0.99109	1.99483
4	H	H	H	*n*-propyl	0.99830	0.99271	2.00914
5	H	H	H	isopropyl	0.99180	0.98614	2.00225
6	H	H	H	*n*-butyl	0.99550	0.99236	2.00568
7	H	H	H	isobutyl	0.98959	0.98636	1.99458
8	H	H	H	*tert*-butyl	0.98636	0.98334	0.98547
9	H	H	methyl	methyl	1.00111	0.98693	2.90880
10	methyl	H	H	methyl	2.04524	1.01285	2.04524
11	methyl	H	methyl	methyl	2.02317	1.01495	3.03542
12	methyl	methyl	methyl	methyl	3.01409	1.0200	3.01409
13	H	H	*tert*-butyl	methyl	0.98194	0.97896	1.93662
14	H	H	H	phenyl	1.03335	1.03525	1.03590

and substituted allenes (ref 18, p 3882, and ref 19), while $F_3C^{\bullet}$, $Cl_3C^{\bullet}$, $H_3C^{\bullet}$ go mainly to the ends independently of allene substitution by methyl groups. Therefore, to grasp *S*-coefficient influence in this case, only those radicals can be taken into account which recognize the changes in substituted and nonsubstituted allenes as to predominant attack on central carbon 2 or terminal 1 or 3 as pictured in Figure 7.

Overlapping and splitting for the middle carbon is taken as an average for both orientations to which double bond goes. This is another situation different from that for alkenes. That is, we have the cuts shown in Figure 8.

Factors of dissimilarity (6) and similarity can thus be determined for both cuts X and Y as for other hydrocarbons has been done. Moreover, overlapping (*o*) and splitting (*s*) are given for cuts X and Y:

$$oX = \frac{FD^oX}{FD^oX + FS^oX} \qquad oY = \frac{FD^oY}{FD^oY + FS^oY}$$

$$sX = \frac{FD^sX}{FD^sX + FS^sX} \qquad sY = \frac{FD^sY}{FD^sY + FS^sY} \qquad (19)$$

A useful rule is that summation of factors of similarity and dissimilarity[9] results again in new factors. This does hold also in this case, and the property leads to an overlapping and splitting for the middle carbon in substituted allenes.

$$oC^2 = \frac{FD^oX + FD^oY}{FD^oX + FD^oX + FS^oX + FS^oX}$$

$$sC^2 = \frac{FD^sX + FD^sY}{FD^sX + FD^sX + FS^sX + FS^sX} \qquad (20)$$

Hence, we get some *S*-coefficients determined in another way than previously described:

$$Sc'C^2 = oC^2/sC^2$$

The *S*-coefficient for a terminal carbon is evaluated in the usual way by determining *o*/*s* ratios for alternating double bonding toward possible orientations as for dienes. Thus, for allene $CH_2=CH=CH_2$, the particular quantities are as follows: $FD^oX = 43 = FD^oY$, $FS^oX = 17 = Fs^oY$, $FD^sX = 93 = Fd^sY$, $FS^sX = 28 = FS^sY$ and $Sc'C^2 = 0.93244$. The received parameters have been gathered in Table 14.

Let us check the statement which is an analogue of the rule already mentioned (for hydrogen bromide) that prevalent free radical attack does not go to the carbon with a greater *S*-coefficient value for neighboring (by double bond) carbons (anti-Markovnikov addition). There are however three coefficients characterizing substituted allenes as shown in Table 14. This formulation for allene hydrocarbon derivatives should sound that free radical prevalent attack (tosyl or

benzenethiyl) does not go to the greatest *S*-coefficient site. So, two positions remain to be chosen. For benzenethiyl $(C_6H_5S^{\bullet})$ and tosyl $(p\text{-}CH_3C_6H_4SO_2^{\bullet})$ radicals the yields at the terminal carbons of allene are, respectively, 75% and 100% (ref 18, p 3882). Since the greatest *S*-coefficients occupy only terminal carbons after hydrocarbon substitution, the predominant yields at them are not to be more expected. The inversion is thus attributed to the relative changes of *Sc'* values at the central carbon with respect to Sc values at the terminal allene carbons. For compounds 3, 5, 6, and 7 of Table 14, benzenethiyl reaction at the central carbon gave yields of approximately 83% (ref 20, p 2839) and 75% in the case of 8. For compound 14, the yield at the middle was 100% (ref 19, p 686). The case can involve reduced $Sc'C^2$ value in view of conjugation as supposed previously for dienes and chemical shifts (Table 12). Assuming this influence, a stronger rule can be formulated: the chief free radical attack goes to the minimal *S*-coefficient site. This is fully confirmed for tosyl iodide addition to allene derivatives, namely, for compounds 2, 9, 10, and 11 (Table 14, ref 18, p 3883) with the 100% central attack. Accompanying radicals $(H^{\bullet}, I^{\bullet})$ can thus go to one of the terminal carbons, as usual, with the lower *S*-coefficient value of the choice. The reason is not only in steric hindrances. For example, *tert*-butylallene (compound 8, Table 14) gave 38% of hydrogen addition at C^3 in contrast to 37% at the other end (ref 20, p 2839) with the accompanying fall of the Sc value at the more substituted end. The same is confirmed for the $I^{\bullet}$ radical in the case of tosyl iodide addition (ref 18, p 3883). Constructing, as previously, tables for fuzzy set control of the reactions and conditions considered, the maximal grade (=1) of membership can be achieved for distinguished allene derivatives; i.e., the selected (due to the availability of data) collection of compounds 1, 2, 3, 5, 6, 8, 9, 10, 11, 13, and 14 of Table 14 belongs to the fuzzy set core[14] when the last condition is used. The allene derivatives were approached in a slightly modified way in view of the special chemical circumstances for the central carbon. A similar approach is supposed to be required for reactions when ring carbon coefficients would be necessary. There is a permanent interest in cyclic olefins,[21] which gives rise to prospective explorations with *S*-coefficients also in this area.

It turns out the fuzzy set estimations afford an efficient theoretical method used in this paper, which is a good reason to increase the tendency for further progress in new findings of universal tools suitable for fruitful chemical applications.[22]

SPLITTING AND OVERLAPPING IN MOLECULAR SYSTEMS

J. Chem. Inf. Comput. Sci., Vol. 40, No. 4, 2000 **1029**

REFERENCES AND NOTES

(1) Haken, H. *Information and Self-Organization*; Springer: Berlin, **1988**.

(2) Cramer, F. *Chaos and Order*; VCH: Weinheim, 1993.

(3) Schulz, L. P. Decoding Molecular Processor Information. *Comput. Methods Sci. Technol.* **1999**, *5*, 75−79.

(4) Schulz, L. P. Categories in Chemistry and Chemical Computers. *J. Mol. Struct.: THEOCHEM* **1989**, *187*, 15−22.

(5) Hoffmann, R.; Laszlo, P. Representation in Chemistry. *Angew. Chem., Int. Ed.* **1991**, *30*, 1−16.

(6) Nikolić, S.; Trinajstić, N. On the concept of a chemical model. *Croat. Chem. Acta* **1997**, *70*, 777−786.

(7) Trinajstić, N. *Chemical Graph Theory*; CRC Press: Boca Raton, FL, 1992.

(8) Schulz, L. P. A Mathematical Approach to Chemical Structures−Fuzzy Set Molecular Invariants. Submitted for publication.

(9) Schulz, L. P. Set-Spectrum Method in Structure Comparison and the Concept of Topological Kinship. *Appl. Math. Comput.* **1991**, *41*, 1−21.

(10) Schulz, L. P.; Chojnacki H., Resemblance Analysis of Molecular Systems on the Grounds of DFT Evaluated Parameters. Platinum Complexes and Their Anticancer Activity. *Int. J. Quantum Chem.* **1996**, *60*, 1385−1391.

(11) Saytzeff, A. Zur Kenntnis der Reihenfolge der Anlagerung der Jodwasserstoffelemente in organischen Verbindungen. *Ann. Chem.* **1875**, *179*, 296−301.

(12) March, G. *Advanced Organic Chemistry*; Wiley: New York, 1993.

(13) Brown, H. C.; Moritani, I. Steric Effects in Elimination Reactions. I. Directive Effects in Unimolecular Eliminations; A Case of Hofmann Rule Elimination in an E1 Reaction. *J. Am. Chem. Soc.* **1955**, *77*, 3607−3610.

(14) Klir, G. J.; Yuan, B. *Fuzzy Sets and Fuzzy Logic*; Prentice Hall: New Jersey, 1995.

(15) Brown, H. C.; Nakagawa, M. Steric Effects in Elimination reactions. III. The Effect of the Steric Requirements of Alkyl Substituents upon the Extent and Direction of Unimolecular Elimination in the Solvolysis of Secondary Alkyl Brosylates. *J. Am. Chem. Soc.* **1955**, *77*, 3614−3619.

(16) Brown, H. C.; Moritani, J. Steric Effects in Elimination Reactions. V. The Importance of Steric Strains in the Reactions of Highly Branched Carbonium Ion. *J. Am. Chem. Soc.* **1955**, *77*, 3623−3628.

(17) Breitmaier, E.; Voelter, W. *Carbon-13 NMR Spectroscopy*; VCH: Weinheim, 1987.

(18) Byrd, L. R.; Caserio M. C. Orientation and Stereochemistry of Radical Addition *J. Org. Chem.* **1972**, *24*, 3881−3891.

(19) Pasto, D. J.; L'Hermine, G. Experimental and Theoretical Study of the Radical-Chain Addition of Benzenethiol to Heterosubstituted Allenes. *J. Org. Chem.* **1990**, *55*, 685−694.

(20) Pasto, D. J.; Warren, S. E.; Morrison, M. A. Radical-Chain Addition of Benzenethiol to Allenes. Analysis of Steric Effects and Reversibility. *J. Org. Chem.* **1981**, *14*, 2837−2841.

(21) Marchand, A. P.; Ganguly, B.; Shukla, R.; Krishnudu, K.; Kumar, S. V. Experimental and Theoretical Investigations into the Stereoselectivities of Peracid Promoted Epoxidations of Substituted Norbornenes and Norbornadienes. **1999**, *55*, 8313−8322.

(22) Schulz, L. P. Algebraic Criteria on Chemical Compound Saturation and Unsaturation. *Int. J. Quantum Chem.* **2000**, in press.

CI980150H