Comprehensive System for Classification and Nomenclature of Organic Reactions

James B. Hendrickson

Department of Chemistry, Brandeis University, Waltham, Massachusetts 02254-9110

Previous authors' ideas on reaction classification have been incorporated into a unified and comprehensive system to record the bonding change in any reaction. The system offers a "reaction logo" as a single drawing of the bonding change at the reaction center. It also provides a smooth hierarchy of increasing detail from the two monocyclic basic forms (homovalent and ambivalent), through their subclasses of unchanging shell σ-bonds on the reaction center, to the full detail of the individual atom types in the reaction center. A simple linear notation incorporates all of this information in a form such as $-\mathbf{3_L}$[OCCCOH], for the reaction center of the aldol reaction.

INTRODUCTION

It is surprising in a mature science like organic chemistry to have so little organized nomenclature for organic reactions, the central tools of the trade. Despite the systematic nomenclature system for compounds established a century ago by IUPAC, the literature on systematic classification for reactions is very sparse.[1−7] This article combines and develops the ideas of previous authors into a unified, comprehensive system for classifying reactions. The system provides a standard representation of any reaction in a pictorial format as well as a linear notation for nomenclature.

DERIVATION OF THE SYSTEM

At a basic level a reaction is described by the net bonding change between reactant and product, over just those atoms which change their bonding, defined as the *reaction center*. Figure 1A shows a bond moved from one atom pair (AB) to an adjacent one (BC), with atom A losing a bond and atom C gaining one. Since the overall change from reactant to product must retain constant bonding of the atoms, there must be a fourth, carrier atom (D) to carry bonding electrons back to atom A. This results in a cyclic, four-atom format for a single reorganization of two bonds, as in Figure 1B. Reorganization of three bonds affords an analogous six-atom cycle; higher, even-numbered cycles are possible but very rare.

This basic reaction exchanges just one bond at each atom in the cycle and so is defined as a *unit reaction*.[8] Almost all reactions are described by these monocycles. Of these some 80% are single unit reactions, the rest usually being composites of successive unit reactions in sequence, rarely more than two.[9]

Reactions are normally described with two structures, as reactant → product in Figure 1B. However, the essential information of the reaction center can be condensed into a single drawing which is more compact and economical[4,6] and has been called an "imaginary transition state".[6] The simplest single drawing incorporates the reactant bonds, which break, as plain lines; and the product bonds, which are formed, as dotted lines. This is defined here as the *reaction logo*, exemplified in Figure 1 with the structure in
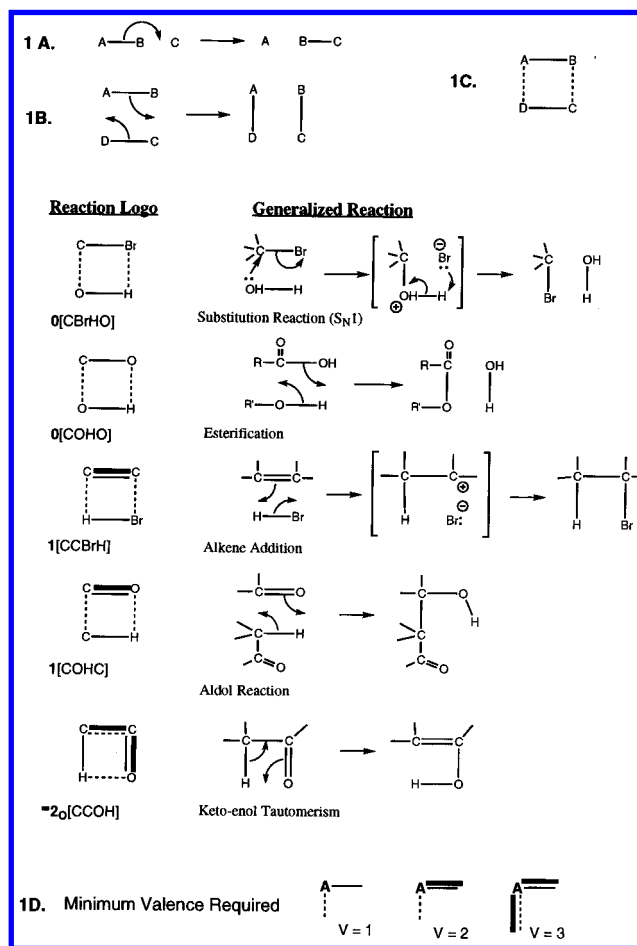

Figure 1. Reaction logos to represent reaction change.

Figure 1C for the general reaction of Figure 1B. The reaction logo fully describes the reaction in terms of the bonding changes in the reaction center. The arrows in Figure 1B may be placed inside the logo but are unnecessary for depicting the change in the reaction.

Reaction 1C is a metathesis or substitution, the only bonds among the cycle atoms being those which change. However, reaction center atoms may be further linked by unchanging bonds, hence double bonds in either reactant or product. These unchanging bonds are called *shell bonds*[2] and are the skeletal σ-bonds in a double bond; the changing bonds are the π-bonds. These shell bonds are distinguished here with

COMPREHENSIVE SYSTEM FOR CLASSIFYING REACTIONS

J. Chem. Inf. Comput. Sci., Vol. 37, No. 5, 1997 **853**
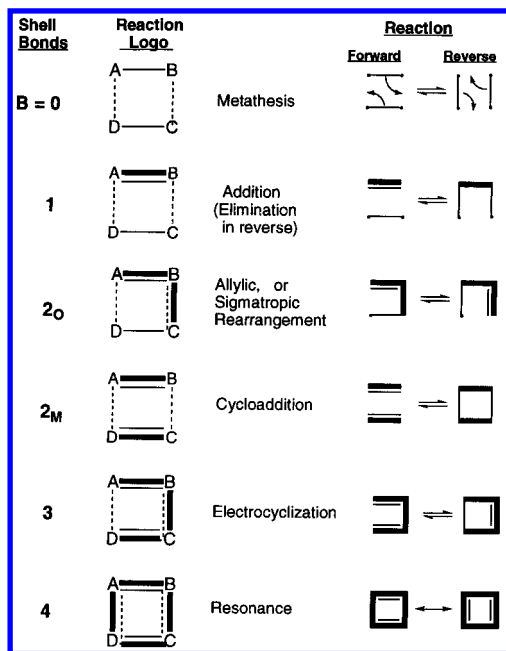

Figure 2. Shell bond subclass logos for 4-cycle reactions.

boldface lines. This is, however, not obligatory since any reactant single bond shown between reaction center atoms must be a breaking bond, while with any double bond only one bond (the π-bond) is broken, the other bond being a shell σ-bond which remains unchanged.

Examples of several common reactions are shown in Figure 1 with both the reaction logo and the traditional depiction; each example is also labeled with the linear notation described below. The examples emphasize that only the bonds which change and only the atoms to which they are attached are recorded in the reaction logo.

The reaction logo representation is intended only to describe the net change from reactant to product. It does not describe the mechanism of the reaction, despite the possibility of adding arrows to the logo. Since radical reactions usually begin and end with fully bonded molecules, the notation serves equally well for these. This is apparent in Figure 1 with the addition of HBr to an alkene, which can occur by either ionic or radical mechanisms.

The classification of reactions must proceed in a hierarchy with levels of subclasses defined for increasing detail. This taxonomy of reactions may be seen in three levels: the primary generalizations as *n*-cycles for *n* = 4, 6, 8, ..., then the parallel subclasses of shell bonds at the next level, and finally at a third level of detail the subclasses which define the atom types possible for the *n* atoms in the cycle.[2,10] The 4-cycle has six shell bond subclasses, the 6-cycle has 13, and the 8-cycle has 30. The considerably greater number of atom type permutations has been organized and enumerated.[2,6] The examples in Figure 1 specify both shell bond and atom type subclasses.

There are six possible combinations of shell bonds on the periphery of the 4-cycle. These are the six subclasses summarized in Figure 2, and each is labeled by the number of shell bonds (**B**) shown at the left. The basic form is the metathesis reaction with no shell bonds (**B = 0**). There are two orientations for the two-bond shell, one with the shell bonds adjacent or vicinal, the other with them separated. The former shell is labeled ortho (**O**) and the latter *meta* (**M**), i.e., vicinal or separated like substituents on benzene and

consistent with the two-bond shell labels in the 6-cycle below. The chemical type of each subclass in Figure 2 is shown at the right along with the full reaction change in the forward and reverse directions.

The convention for shell bond placement shown here results in the structure with the fewest π-bonds appearing at the right, and these are usually the more stable in any pair of structures. Hence, as written, the reaction to the right will generally be the more favored one thermodynamically, and so is taken as the forward direction of the logo. The reaction in the reverse direction (right to left in Figure 2) is simply designated by placing a minus sign before the same shell bond label, as in **1** for addition and **−1** for elimination, or **2$_M$** for [2+2]-cycloaddition and **−2$_M$** for its cycloreversion.

It may be noted that the full-shell representation (**B = 4**) is not a reaction since no σ-bonds are made or broken. It just shows the generation of one resonance form from another, reflecting the common use of the arrow convention for generating either reaction products or resonance forms.

In order to standardize an unambiguous presentation of reactions, it is important to establish certain uniform conventions. The symmetry of the square cycle has a 4-fold axis orthogonal to the plane and four 2-fold axes in plane. Unique placement of the shells on the 4-fold axis is realized by placing the first shell bond horizontal at the top of the cycle and the others clockwise from it and as close to it as possible. The moving bonds are also oriented with a breaking reactant bond (plain line) horizontal at top, which determines the forward direction of the reaction as depicted.

In this convention the reaction logo describes the forward direction, which is defined with the A−B bond, at the top, as a plain line; i.e., A−B is broken in the forward reaction as illustrated in Figure 2. This allows easy definition of either reactant or product from the logo, the reactant with the plain, solid bonds and the product with the dotted bonds.

For nomenclature purposes and presentation in text form the reaction logos in Figure 2 may also be simply described with a linear notation or text label. This consists of two parts: the shell bond subclass first and the atom type subclass next as an atom list. The shell bond subclass is identified by the shell bond number at left in Figure 2. This is followed by a bracketed list of the reaction center atoms in clockwise order from top left (A, B, C, D in Figure 2). The text labels for the reactions in Figure 1 are appended below their logos.

The convention for the reaction logo orients the shell bonds clockwise from the top to create an unambiguous orientation of the unchanging shell with respect to the orthogonal 4-fold axis. Ambiguity in the atom type order can still arise from inversions around the 2-fold axes in the plane. A unique orientation here can arise from the valence demands of the shell, since only higher-valent atoms can appear first in the atom order.

Where an ambiguity exists, the convention requires that the atom order in the logo and bracketed atom list is prioritized by valence; i.e., C > N > O > X, H. This will commonly result in carbon at top left and first in the list. The required minimum valences for the logo sites are shown at the bottom of Figure 1. The atoms in the HBr addition are unambiguously ordered as [CCBrH]. A reaction in the reverse direction from the convention is simply indicated with a minus sign. Hence, the keto → enol conversion is
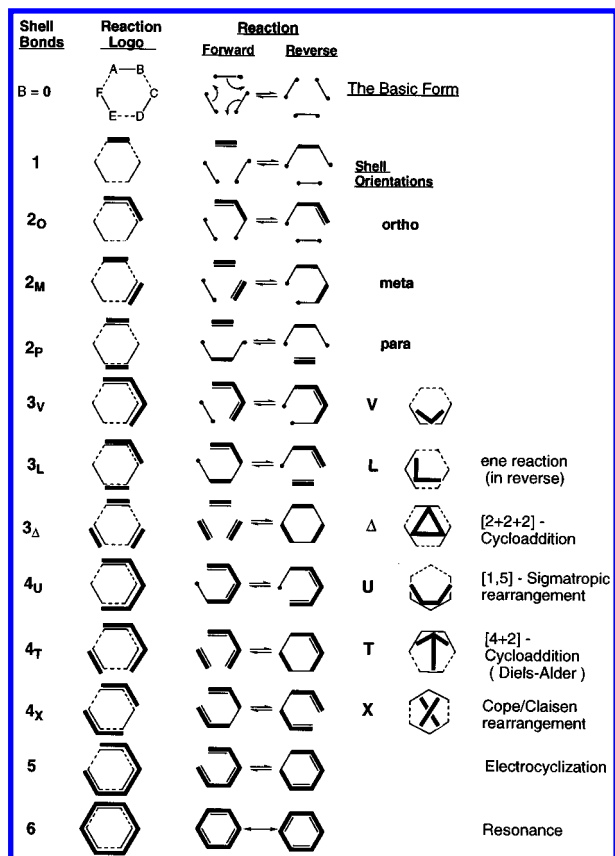
Figure 3. Shell bond subclass logos for 6-cycle reactions.

Figure 4. Examples of 6-cycle reactions: logos and mechanisms.

shown as -2_O[CCOH], which has the reverse direction for the 2_O shell convention with the valence priority. The alternative 2_O[OCCH] would represent keto → enol as the forward direction, but the atom priority is incorrect.

6-CYCLE REACTIONS

The formulation above for the 4-cycle is paralleled completely in the 6-cycles, with six atoms and three changing bonds in the reaction center ring.[1,2] Here there are 13 shell bond subclasses which are shown in the same convention in Figure 3, with both the first shell bond and a breaking reactant bond horizontal at the top.

The shell bond subclasses are characterized as before by the number of shell bonds and their orientation on the cycle; the shells with two to four shell bonds each have three different orientations. In the case of two shell bonds, these were labeled ortho (**O**), meta (**M**), and para (**P**) as with the orientations of two substituents on the six-atom benzene ring. The larger shells were then labeled with letters whose form shows the orientation of the shell bonds.[2] The shell labels themselves are shown down the left side of the set of reaction logos in Figure 3.

For each orientation of three to four shell bonds, joining the midpoints of the shell bonds creates the form of the labeling letter, as indicated down the right side of Figure 3. The shells are rotated there to show the labeling letteer upright, but the particular orientation of the shells in the reaction logos themselves is dictated by the convention above, i.e., placing the first shell bond horizontal at the top and the others appearing clockwise from it as early as possible. Thus the order shown places the vicinal, or connected, shell bonds first in any set, clockwise from the top bond.
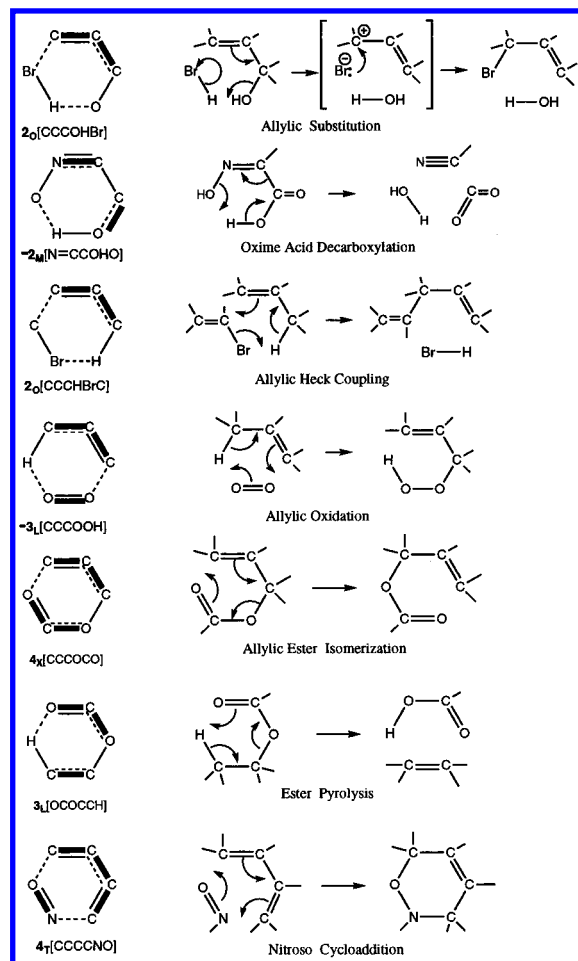
Since the 6-cycle reactions are often thermal pericyclics, characteristic pericyclic reaction names are listed at the right in Figure 3, but the logos apply equally to reactions which may not be concerted pericyclics. The convention for shell bond placement here again results in the structure with the fewest π-bonds appearing at the right. Hence, as written, the forward reaction (to the right) will again usually be the more favored one thermodynamically; the only exception here is the ene reaction (**3**). As with the 4-cycle the reverse reaction (to the left) is distinguished with a minus sign on the same label, as in 4_T for Diels—Alder cycloaddition and -4_T for its cycloreversion, or -3_L for the ene reaction in its favored direction.

Figure 4 collects examples of 6-cycle reactions, many of them pericyclic, with their characteristic reaction logos. In the traditional depiction with separate reactant and product structures the arrows may be placed on either clockwise or counterclockwise. This makes no difference to the net overall bonding change shown in the reaction, but in a rough mechanistic sense the arrow direction is often more acceptable in the direction which breaks bonds in the direction of their natural polarity. For this reason the arrow direction is reversed from the Heck reaction to the three allylic cases. The allylic oxidation is also seen as proceeding in the conventional reverse direction, with the dotted bond at the top of the reaction logo and the -3_L shell designation. Ester pyrolysis is in the conventional forward direction for the 3_L shell.

The allylic nature of a class of reactions is recognized in several cases in Figure 4 by a minimum shell of 2_O and CCC
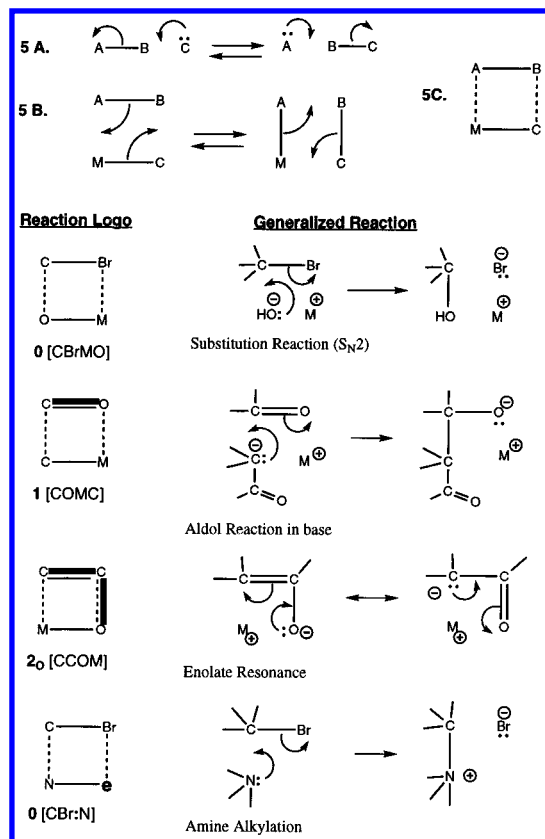
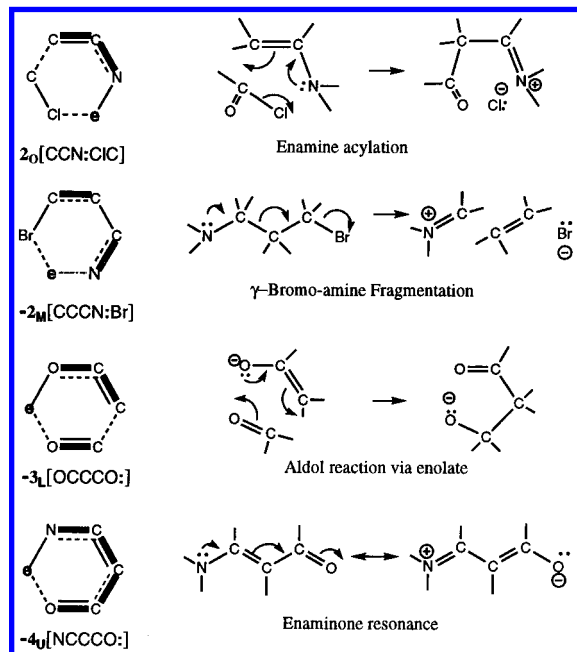**Figure 5.** Equivalent formulation for reactions with unshared electron pairs.

Figure 6. 6-Cycle reaction examples with unshared electron pairs.

the bracketed list. In this way the cycloaddition of three acetylenes to benzene is $3_4[C≡CC≡CC≡C]$, and the Diels–Alder addition of a diene to a nitrile would be 4_T-[CCCCC≡N]. The second example in Figure 4 also exhibits the triple bond, as do Figures 7E and 11B.

REACTIONS WITH UNSHARED ELECTRONS

The model reaction in Figures 1–3 consists of a cycle of changing bonds on an even number of atoms. There are also reactions in which an unshared electron pair is changed into a bond, the simplest model being the linear three-atom conversion in Figure 5A, analogous to that in Figure 1A with an extra electron pair. Like Figure 1A, it does not maintain constant valence across the three atoms, but it can if expanded to a four-atom cycle.

If the unshared pair is a negative ion, it is easy to add its counterion as a metal ion (M^+) bonded to it ionically, as in Figure 5B. This maintains constant valence, in a four-atom cycle with two changing bonds and so is fully analogous to the previous treatment (cf., Figure 1B). This is described with the corresponding reaction logo of Figure 5C reorganizing two bonds on four atoms, the same as Figure 1C.

In another analogy the unshared pair in Figure 5A is a conjugate base, so the same reaction on its conjugate acid would show H instead of M bonded to the electron pair. The first two reaction examples in Figure 5 correspond to their conjugate acids in Figure 1, for substitutions and aldol reactions. Their logos differ only in substituting M for H.

In the full generalization of these reactions the unshared electron pair can itself by just annotated as a bond, with the end-of-bond notation **e** in the reaction logo, whereby it is seen as a bond or as an orbital with no atom nucleus as its other end. Thus, the reaction is represented as cyclic with an unshared pair passed around the ring. This is exemplified with the reaction logo for amine alkylation in Figure 5. In any of the cases in Figure 5 it is equivalent to substitute **e** for M in the logo, just as common usage leaves out the metal ion in writing mechanisms like the S_N2 substitution (cf., Figure 5A). Similarly, the conjugate base form of any reaction results from substituting **e** for H in a reaction logo.

In the bracketed atom list for the linear notation the unshared pair **e** is designated with a colon in the proper place in the clockwise atom list. This avoids confusion with atoms which have an **e** in the atom symbol (cf., Se). In this way both families of reactions have a single unified mode for drawing and for the linear notation.

Like H and M, the unshared pair **e** can only appear at the end of a single bond (valence of 1; see the bottom of Figure 1). Hence, it can only occupy an atom position with no shell bonds since a shell bond implies a double bond in either reactant or product. Therefore, some of the shell bond forms of Figures 2 and 3 will not be viable for these reactions. The valence priority and shell bond convention—with shell bonds first in the clockwise order—dictate that the colon for the unshared pair in these odd-atom homovalent reactions will appear at or near the end of the bracketed atom list.

Thus, in Figure 5 the notation for the enolate resonance is $2_O[CCO:]$, the conjugate base form of keto–enol tautomerism in Figure 1. The aldol reaction in base is just the conjugate base form of the aldol reaction in Figure 1 with ":" instead of H. Common examples of five-atom reactions with unshared pairs are illustrated as 6-cycle reaction logos
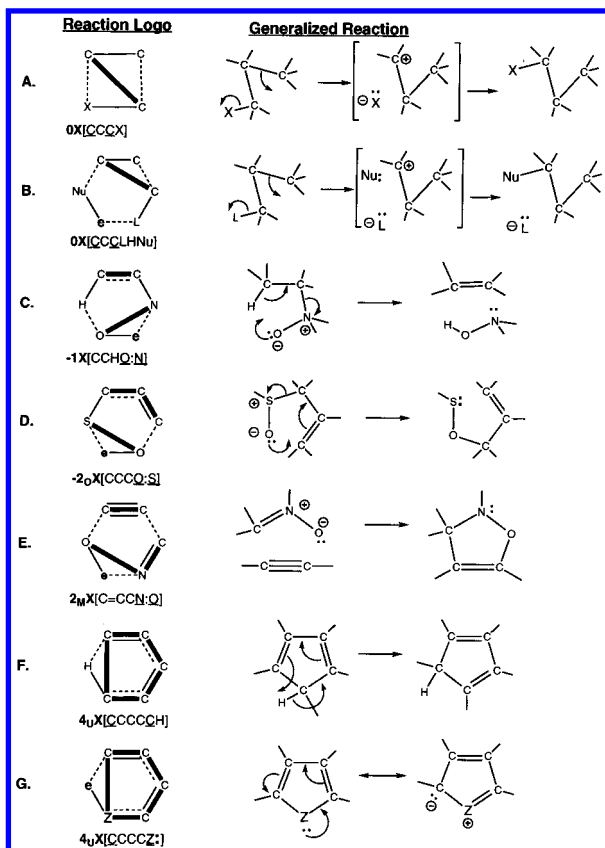
as the first atoms in the bracketed list. A whole family of Claisen/Cope reactions is modeled in the 4_X shell just by changing the atom types in the last three atoms in the list. The many variants of [4 + 2] cycloaddition also arise by varying the atoms in the 4_T shell.

The presence of triple bonds (in either the reactant or product) implies an unchanging π-bond as well as the σ-shell bond. To avoid excessive expansion of nomenclature, the same shell label can be used by just designating the extra π-bond with an = mark between the triply bonded atoms in

Figure 7. Formulation of reaction cycles with unchanging cross-bonds.

Figure 8. Comparison of reaction logos for homovalent and ambivalent reactions.

in Figure 6. The generation of resonance forms must have a full shell and is exemplified in the last example (-4_U) with the normal linear resonance on five atoms.

CROSS-BONDS

The only unchanging bonds so far recognized have been the shell bonds between *adjacent* atoms on the reaction center cycle. Separate recognition is accorded bonds *across* the cycle,[3] which are called **cross-bonds**.[2] When a cross-bond is present within a given shell, the shell label is used as before and an X is added for the cross-bond. The position of the cross-bond is marked in the linear notation by underlining the symbols for the atoms at the two ends of the cross-bond. Examples are collected in Figure 7.

Cross-bonds are common in rearrangements. The simplest instance is the 4-cycle with a cross-bond but no shell bonds, as shown in Figure 7A for a simple 1,2-shift. The 6-cycle in Figure 7B shows the same reaction with a different nucleophile and leaving groups. They are also common in the odd-atom homovalent reactions, often found across electron pair, **e**, hence linking the atoms which pass the unshared pair between themselves. The logos for the thermal N-oxide elimination (Figure 7C) and for allylic sulfoxide to sulfenate (Figure 7D) illustrate this. The dipolar cyclo-addition of nitrones to alkynes in Figure 7E is the same reaction on a cross-bonded 2_M shell.

A cross-bond on a 4_U shell would represent the 1,5-sigmatropic rearrangements on a cyclopentadiene with H (or C) migrating in Figure 7F. The same form with **e** in place of H denotes the resonance of the cyclopentadiene anion. The resonance generation for furan, pyrrol, and thiophene
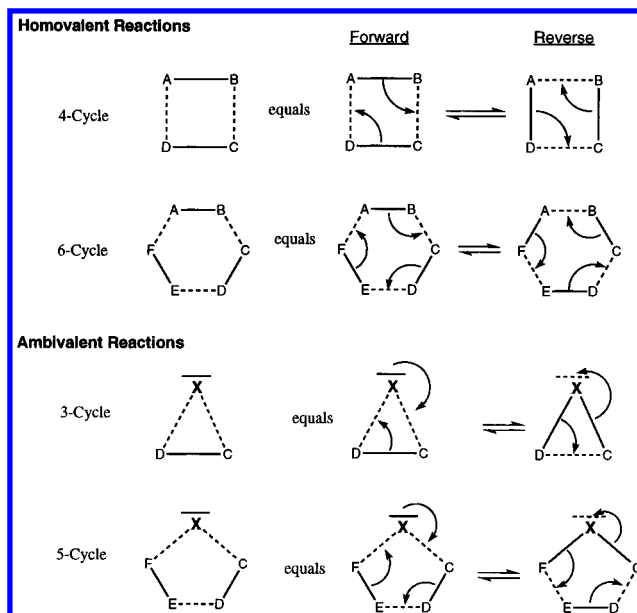
is then just the same with Z = O, N, S replacing the carbon bearing **e**, as shown in Figure 7G.

AMBIVALENT REACTIONS

Thus far, the reactions have involved only atoms which do not change valence. Unit reactions in which no atom changes its valence have been defined as *homovalent*, while those in which one or more atoms change their valence via redox processes are *ambivalent*.[3] The two kinds of reactions considered above are homovalent and share a single reaction logo form. The first kind involves a cycle of moving bonds with an even number of atoms, without changing any atom valence. The second kind has an odd number of atoms plus an unshared pair of electrons passed around the ring from the first atom to the last. In this second kind there is one unshared pair in both reactant and product. Both kinds, however, share the same notation.

In ambivalent reactions an ambivalent (redox) atom with one unshared pair passes that pair into an odd-atom cycle to form a product with two bonds and no unshared pair on that atom. The two families are contrasted in Figure 8. The ambivalent reactions are generalized with cycles of three and five atoms in which the ambivalent atom is designated with **X**. The total number of electron pairs remains constant but changes form from an unshared pair to a bond pair. The ambivalent 3-cycles and the homovalent 4-cycles each have two changing electron pairs/bonds; the 5-cycle ambivalent and 6-cycle homovalent each have three pairs.

The unit reaction definition of only a single exchange of one bond (or electron pair) for another is expanded to include ambivalent reactions with this unit exchange at atom **X** (only) of one unshared pair for two bonds; the ambivalent atom **X** is oxidized to accept two bonds in exchange for its unshared pair.

With carbon the ambivalent atoms will be carbenes with a valence change of 2 ↔ 4. These are usually produced as reactive intermediates from prior reactants but also include CO or isocyanides as stable reactant/product entities. With phosphorus and sulfur, stable entities of different valence
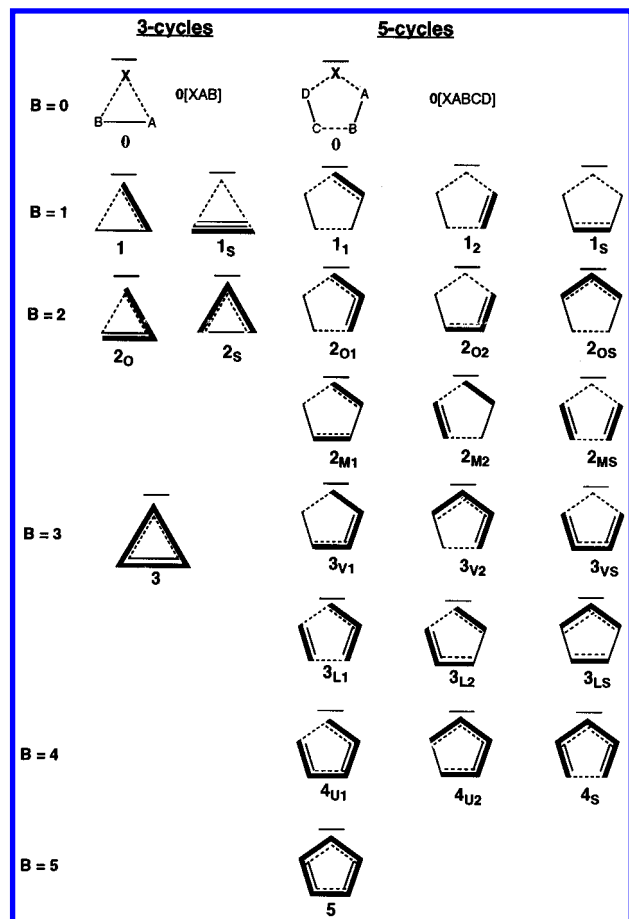
Figure 9. Shell bond subclass logos for 3- and 5-cycle ambivalent reactions.

Figure 10. Ambivalent reactions: logos and mechanisms.

are more common, as with phosphites to phosphates (3 ↔ 5) or sulfides to sulfoxides (2 ↔ 4).

In Figure 8 the atom **X** may be seen as coalescing the two atoms A−B in the homovalent family and that A−B bond becomes the unshared pair on **X**. This is conveniently placed at the top on the vertical axis like the A−B bond in the homovalent forms of Figures 2 and 3. The ambivalent electron pair is drawn on it as a line like a bond, so as to distinguish its function from **e** or : in the homovalent odd-atom reactions above. The forward direction of reaction is defined, consistent with the previous convention, as having the ambivalent electron pair present in the reactant and missing in the product, as shown in Figure 8.

As in the homovalent reaction form the next clockwise bond (B−C/ homo and **X**−C/ ambi) will be dotted, i.e., formed in the forward direction. In the reverse reaction this bond is a plain line and the line above **X** is dotted, indicating that the unshared pair on **X** is formed in the reverse reaction. This is also illustrated with added arrows in Figure 8.

The shell bond nomenclature here is more troublesome since the symmetry is less than in Figures 2 and 3. While there are still six shell forms for the 3-cycle, they are different from the six homovalent ones on the analogous 4-cycles, and there are now 20 for the 5-cycle which need to be labeled, instead of 13 for the 6-cycle.

The possible shells for the 3- and 5-cycles are summarized in Figure 9 with labels adapted from the homovalent ones, and adding the **s**-subscript to designate orientations which are symmetrical on the vertical axis. Some labels from the 6-cycle (2_P, 3_Δ, 4_X, 4_T) are not transferable at all to the
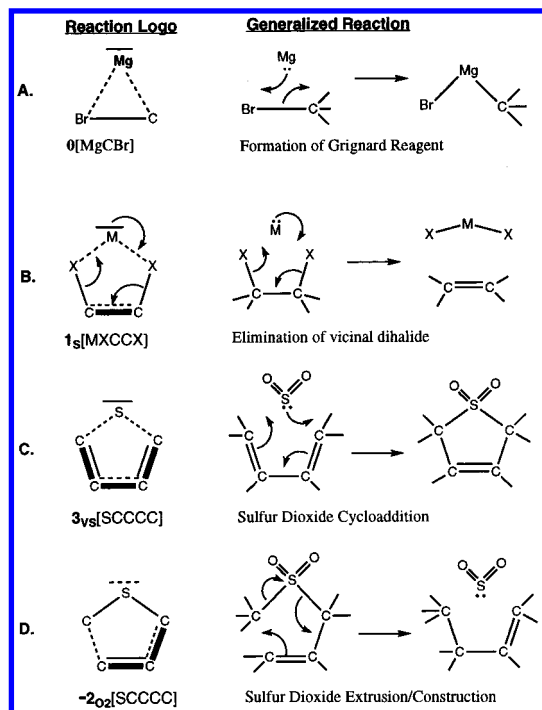
smaller 5-cycle, but there are still three variants for each six other labels because of the lower symmetry. The labeling chosen here for the 5-cycle simply numbers the two unsymmetrical shells for each of the 2_O, 2_M, 3_V, 3_L, and 4_U shells, with number 1 indicating that the shell bond sequence begins at atom **X**. The symmetrical ones are just labeled with **s**. More elaborate labeling would be hard to justify in view of the rarity of these ambivalent reactions.

The linear notation can have the same form as before with the specification that the ambivalent atom **X** is placed first in the clockwise atom list. In this way no indicator for the unshared pair is necessary and the atom list will have an odd number of items, hence distinguishing it from homovalent reactions. Samples of ambivalent reactions are illustrated in Figure 10, the simplest case being the 3-cycle formation of a Grignard reagent as **0**[MgCBr] in 10A; carbene insertion into a CH-bond would be analogous, annotated **0**[CCH].

As 5-cycles the metal reduction of a vicinal dihalide is shown in Figure 10B with a symmetrical one-bond shell and the cycloaddition of SO_2 to a diene in Figure 10C has a symmetrical three-bond shell as 3_{VS}[SCCCC]. The phosphine addition to α-diektones is just an atom type variant of the latter with the same shell form, labeled 3_{VS}[POCCO]. The reverse reaction in Figure 10D, with sulfur dioxide extrusion, provides for C−C construction as well.[11]

A common ambivalent atom is carbene. The reaction logo for dichlorocarbene forming from chloroform is shown as reaction 1 in Figure 11A, followed by its addition to an alkene, shown as reaction 2 next to it. As with homovalent reactions, there may also be ambivalent cases with a moving electron pair which is retained between reactant and product as well as the ambivalent atom pair which becomes a bond in the product. The designation of the former by **e** in the reaction logo and : in the atom list is distinguished from the plain and dotted bond representation for the unshared pair on the ambivalent atom **X**. This is illustrated with carbene formation from diazomethane in Figure 11B, which
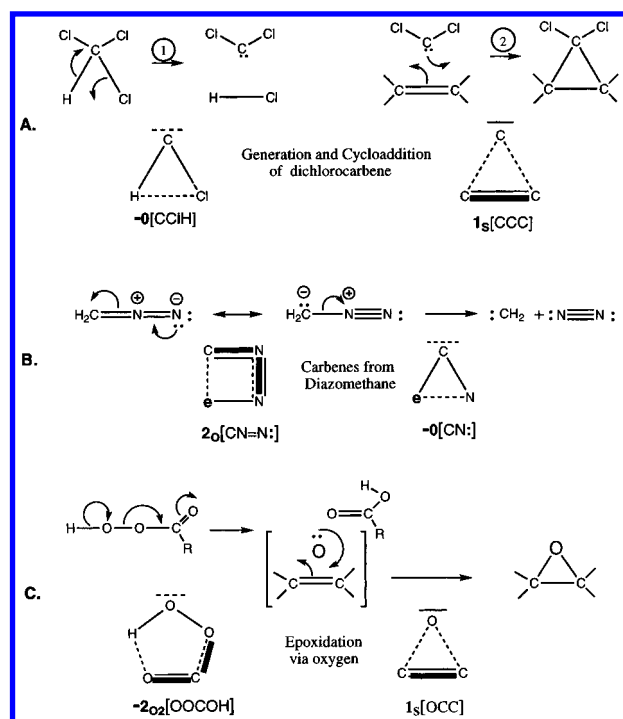
Figure 11. Composite reactions as successive unit reactions.

shows first the homovalent generation of one resonance form of diazomethane from the other and then the ambivalent loss of nitrogen to give the carbene.

MORE COMPLEX TRANSFORMATIONS

More involved reactions will be recognized by the change of more bonds than one at some of the atoms. In general these are readily broken down into a sequence of unit reactions, as indeed they usually are also in formulating their mechanisms. Thus, the Wittig reaction is a two-step conversion of C=O + C=P → C=C + P=O and may be denoted as $2_M[CPOC] + 2_M[CCPO]$. All four atoms change two bonds each.

Complex reactions can usually be annotated as composites of successive unit reactions, rarely more than two.[9] This is apparent in the treatment of the carbene formation and its immediate cycloaddition in Figure 11A. The treatment of diazomethane in Figure 11B is also easily adapted to nitrene formation from azides, followed by cycloaddition to aziridines. Similarly, the epoxidation of an alkene is a parallel cycloaddition of an electron-deficient oxygen atom. Between peracid and epoxide the oxygen atom exchanges two bonds, as carbene CCl_2 does from chloroform to cyclopropane. In each case we should write two successive unit reactions, as outlined in Figure 11A,C.

Conversions in which two unit reactions occur successively via an unstable intermediate atom **X**, like carbene, nitrene, or atomic oxygen, can be represented by a spiro bicyclic reaction logo with the **X** atom at the spiro center.[7] This may be simply constructed by merging the two successive ambivalent monocycles into one bicycle with **X** in common
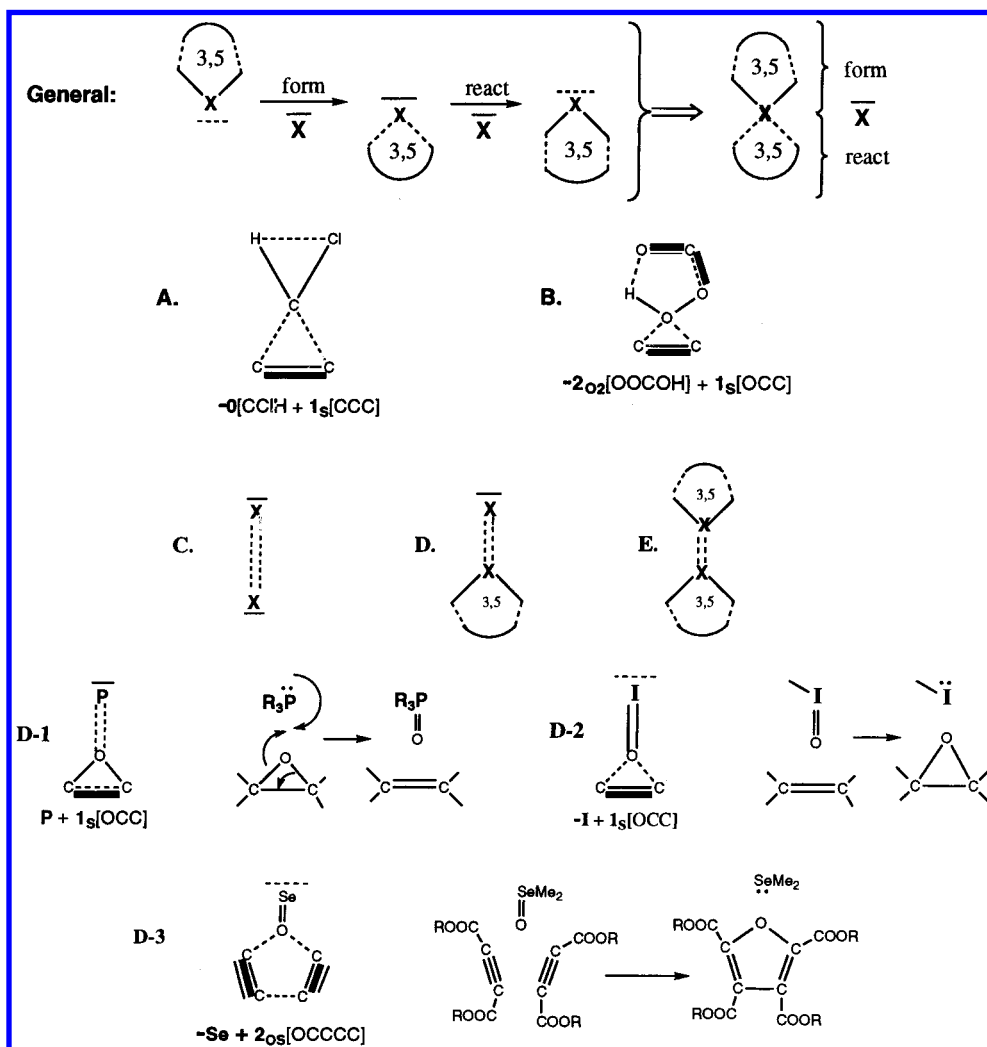

Figure 12. Bicyclic reactions logos to designate successive unit reactions.

COMPREHENSIVE SYSTEM FOR CLASSIFYING REACTIONS

J. Chem. Inf. Comput. Sci., Vol. 37, No. 5, 1997 **859**

at their juncture, as shown in Figure 12 for the general case with either 3- or 5-cycles. The merged bicyclic logo shows both the two bonds broken (dotted) and the two formed (plain) at atom **X**. This is illustrated with bicyclic logos for carbene formation and cycloaddition from Figure 11A, and for epoxidation from Figure 11C, in logos of Figure 12A,B, respectively. The text labels are just a combination of the two successive ones in Figure 11; the inverted monocycle for **X** formation retains its conventional label.

More complex composite reactions can be similarly created with more than one ambivalent atom **X**, and examples of these exist. In the simple monocyclic transitions, the number of changing electrons equals the cycle size for homovalent reactions, i.e., 4- or 6-cycles; the cycle size of ambivalent reactions with one ambivalent atom is one less than the number of electrons changing, i.e., 3- or 5-cycles with four and six moving electrons, respectively (see Figure 8). With two ambivalent atoms the cycle size will be two less than the number of moving electrons. This imples a "2-cycle" for the smallest case on four electrons, and this is generalized in Figure 12C, a 2-cycle being just a double bond.

The simplest case of two ambivalent atoms **X** in Figure 12C would be the dimerization of a carbene to an alkene. As with the bicyclic logos above for composite reactions, the ambivalent **X** for its forward reaction may be replaced by the ambivalent monocycle which forms it. Replacement of one **X** affords the bicyclic logo in Figure 12D (a 2- and a 3- or 5-cycle), and replacement of both gives the tricyclic one in Figure 12E.

The bicyclic (12D) is exemplified in the use of phosphines or phosphites to deoxygenate epoxides in Figure 12D-1, and the reverse reaction is seen in epoxidation by iodosobenzene, in Figure 12D-2. A 5-cycle analogue of Figure 12D for furan synthesis in Figure 12D-3 was derived and successfully demonstrated by Zefirov[5,12] using a similar analysis but without the distinction of shell bonds. The shell bond subclasses are shown in the linear notation below the examples in Figure 12. These subclasses allow not only designation of the ambivalent monocycles which replace **X** but also a systematic search among other shell bonds for new reaction analogues.

The replacement of both ambivalent atoms with mono-cycles is generalized in Figure 12E. If the two atoms **X** are connected by a shell bond, the form becomes a synthesis of triple bonds. The simplest generation of two adjacent carbenes can occur by loss of nitrogen from two diazo groups, which are presumably formed in the oxidation of α-diketone bishydrazones to acetylenes.[13] This is formulated as Figure 13A with an attached mechanism.

The form of Figure 12E is also characteristic of a number of "coarctate" reactions discovered by Herges.[14] He demonstrates that these reactions probably take place in a single, concerted step, using a single, Möbius transition state which allows at a crossover atom the making and breaking of two bonds which are orthogonal to each other. He also shows that the coarctate model serves as a basis to invent new reactions for forming triple bonds. One of his typical coarctate reactions is shown in Figure 13B.

The description of multiple ambivalent atoms formulated here can be extended to more than two **X** atoms; a sequence of three atoms in a row with two bond changes each is shown in Figure 13C for the phosphine/phosphite desulfurization of thionecarbonates to alkenes. The central 2-cycle in Figure
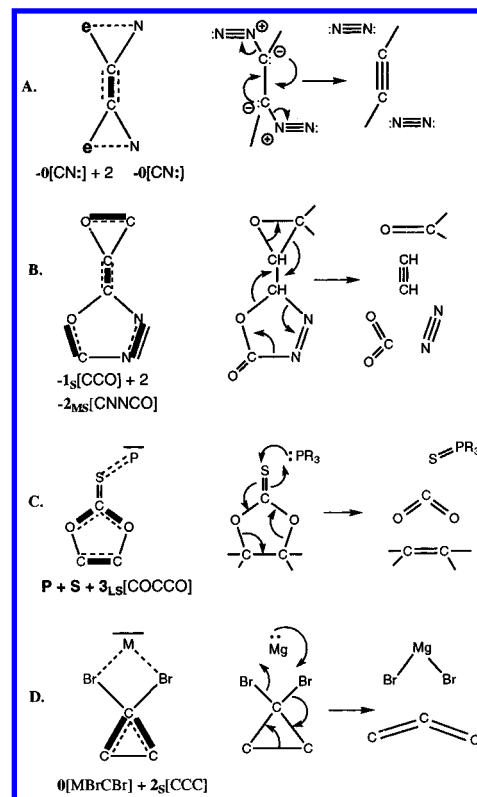

Figure 13. More complex reactions extend the reaction logo representation.

12C−E may also be expanded to the next higher 4-cycle with two ambivalent atoms and six moving electrons. This is represented in Figure 13D for the metal reduction of dibromocyclopropanes to allenes.[15]

SUMMARY

In summary a simple and rigorous representation for all common homovalent or ambivalent unit reactions is provided both as a single drawing, the reaction logo, and also a linear text label. These have been designed to be both comprehensive and simple enough to use easily. The system affords a taxonomic classification of reactions with a small set of basic forms (Figure 8) and two levels of subclasses, the shell bonds in Figures 2, 3, and 9, and then the atom types presented in a clockwise order list. These monocyclic unit reaction forms correspond to over 90% of known reactions,[14] with a horde of atom variants not yet realized.

This organization of reactions is developed rigorously from the net change in the bonds at the reaction center and as such is independent of mechanistic considerations but does not contravene them. With this orthogonal viewpoint, it is capable of suggesting new analogies across traditional lines and can inspire the invention of new reactions. The extrusion/construction in Figure 10D was derived in such a way.[11]

REFERENCES AND NOTES

(1) Balaban, A. T. *Rev. Roum. Chim.* **1967**, *13*, 875−898.
(2) Hendrickson, J. B. *Angew. Chem., Int. Ed. Engl.* **1974**, *13*, 47−76.
(3) Arens, J. G. *Recl. Trav. Chim. Pays-Bas* **1979**, *98*, (a) 155−161, (b) 395−399, (c) 471−483.
(4) Vladutz, G. In *Modern Approaches to Chemical Reaction Searching*; Willett, P., Ed.; Gower: London, 1986; pp 202−220.
(5) Zefirov, N. S. *Acc. Chem. Res.* **1987**, *20*, 237−243.

(6) Fujita, S. *J. Chem. Inf. Comput. Sci.* **1986**, *26*, 205−252; **1987**, *27*, 99−125.

(7) Hendrickson, J. B. *Recl. Trav. Chim. Pays-Bas.* **1992**, *111*, 323−334.

(8) Hendrickson, J. B. *J. Chem. Inf. Comput. Sci.* **1979**, *3*, 129−136.

(9) Hendrickson, J. B.; Miller, T. M. *J. Chem. Inf. Comput. Sci.* **1990**, *30*, 403−408.

(10) Ugi, I.; Wochner, M.; Fontain, E.; Bauer, J.; Gruber, B.; Karl, R. In *Concepts and Applications of Molecular Similarity*; Johnson, M. A., Maggiora, G. M., Eds.; Wiley: New York, 1990; pp 239−288.

(11) Hendrickson, J. B.; Bergeron, R. *Tetrahedron Lett.* **1973**, 3609−3610.

(12) Zefirov, N. S.; Tratch, S. S. *Anal. Chim. Acta* **1990**, *235*, 115−134. Zefirov, N. S. *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 994−999.

(13) Blomquist, A. T.; Burge, R. E.; Sucsy, A. C. *J. Am. Chem. Soc.* **1952**, *74*, 3636−3642.

(14) Herges, R. *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 91−102; *Angew. Chem., Int. Ed. Engl.* **1994**, *33*, 255−276.

(15) Schuster, H. F.; Coppola, G. M. *Allenes in Organic Synthesis*; Wiley: New York, 1984.