

Systematic Signatures for Organic Reactions

James B. Hendrickson

Department of Chemistry, Brandeis University, Waltham, Massachusetts 02454-9110

Received February 2, 2010

A signature for any reaction is defined by just the net change in bonding of the reacting atoms in the conversion of reactant to product structures. This reaction signature is both unique and definitive for any reaction and consists of a simple linear string of letters suitable to index every reaction in a reaction database for computer access. This allows daily entry of new reactions to be easily incorporated and later retrieved with all related reactions from the reaction database.

INTRODUCTION

For over a century organic reactions have been classified verbally, either gratuitously by nominal inventors' names or by mechanism, bases which are neither general nor rigorous. Furthermore, various usages of "reaction" include multiple mechanistic or successive reaction steps in a single description. A unifying basis for reaction notation is defined here, based only on the overall net change in bonding of the atoms of the reactants as they pass to those of the products, *irrespective of the reagents used or the mechanistic or operational steps involved*.

With the advent of computer databases it has become imperative to index descriptions by a linear string of symbols, as in the SMILES or InChI strings for organic structures. In the same fashion, we define here a unique, definitive signature for any reaction such that all versions of a reaction will be indexed and retrieved together in a reaction database.

Reaction classification was recently reviewed¹ and was there summarized: "While compound structures have long been rigorously classified in universal systems, such as Beilstein and IUPAC nomenclature, a parallel classification of reactions has lagged far behind". No significant publications seem to have been added more recently, but the RInChI project, still at an early stage, aims "to create a unique data string to describe a reaction".² The signatures described here have been developed from the ideas in several earlier publications.^{1,3–6}

DERIVATION OF REACTION SIGNATURES

The basic fact of an organic reaction is the breaking and making of covalent bonds. The simplest definition of any reaction is just a list of those atoms in the reactant structures that change their bonding during the reaction to form the product structures. Since most of the atoms do not change any bonds in a reaction, only the atoms that change their bonding need be specified to define the reaction change. The ensemble of atoms that exchange their bonds is the reaction center, commonly no more than half a dozen reacting atoms.

The reaction center is defined first by mapping all the atoms in the reactant molecules with the resultant atoms in the product molecules. In this mapping each atom retains its identifying number as it passes from the reactant to the

product state. This mapping assures that the change in bonding of each reaction center atom, from reactant to product, may be clearly identified. The reactant and product structures here are just stable molecules, entering into and being isolated from the reaction. They are neither radicals nor high-energy species or intermediates en route.

The focus in this definition of a reaction is not the physical description of the several steps in the reaction mechanism. A "reaction" is defined here solely by the overall net change in bonding of the reacting atoms between the stable reactant and product structures. In fact, the actual mechanism of the reaction may often include bonding changes which later cancel, reverting to the original bonds, as in the intermediacy of keto–enol tautomerism or the last changes for hydrolytic workup.

This mapping of the atoms in the reactant to product structures identifies not only the nature of the changing atoms in the reaction center but also the sequence of breaking and making the bonds between them.

The reaction center is characterized by the atoms that exchange bonds. Each reaction center atom will have one or more bonds broken and also the same number of bonds made. The bonds broken are the bonds held by those atoms in the reactant(s), while the bonds made are the bonds to those same atoms that are found in the product(s). Each reaction center atom will exhibit the same number of reactant bonds broken from it as there are product bonds made to it.

The reaction center atoms are naturally linked in a sequence to one another, since any reactant atom that breaks a bond to one neighbor atom must then make a bond to another neighbor atom in forming the product. Since this holds for each reacting atom, a natural progression of the reacting atoms is defined as the alternating breaking and making of the bonds between them.

This ordered progression of the reacting atoms is called the reactive strand and constitutes a sequence of the atoms that exchange bonds in the reaction center. The development of this reactive strand assumes that every bonding change in the reaction is the exchange of a breaking for a making bond at each atom, and so the number of bonds broken is the same as the number of bonds made. This reactive strand, just of the atoms that exchange bonds, ignores reactions of unshared electron pairs which, therefore, must also be

converted into bonds, as discussed below. This sequence of the reacting atoms and their bond exchanges is characteristic of any particular reaction and serves to define it.

By way of illustration, as a reaction proceeds, any reaction center atom, A_1 , will break its bond to an atom A_2 , which in turn makes a bond to atom A_3 , which must then break its bond to atom A_4 , and so forth until the last changing atom A_n makes a bond back to the first atom A_1 . This first atom, A_1 , which began by breaking its bond to A_2 , is now the only atom not having had a full bond exchange and is thus the last reactive atom needing completion by forming a new bond, i.e., to the last atom A_n .

This creates an ordered list of atoms that is also a list of the bonds between them, alternately breaking and making during the reaction. Such a list must include all changing bonds in the reaction center, with the first bond breaking and the last bond making a new bond between the last and first of the listed atoms in the reaction center, as in $[A_1 - A_2 + A_3 - A_4, \dots, A_n +] A_1 - A_2, \dots$, now repeating through the same bracketed sequence again on another set of the same reaction center atoms in the next reactant molecules.

This sequence is the reactive strand, which passes through each of the reaction center atoms $A_1 - A_n$ and closes by making the last bond from A_n back to A_1 . Therefore, the sequence defines an ordered cycle of all reacting atoms with alternating bonds between them, breaking and making successively around the cycle, and with the end returning to make a bond to the first atom, A_1 . This recognition of a cycle of reacting atoms contains the required linear string for a reaction signature, as the list of all reacting atoms alternately breaking and making the bonds between them, hence a signature list $[A_1A_2A_3A_4, \dots, A_n]$. The cycle is thus recorded as a linear list of all reacting atoms A_1 to A_n in sequence within brackets.

In the signature list, the alternately breaking and making of the bonds between each atom are assumed and need not be specified. The bond between the first two listed atoms is understood as broken, to start the sequence. The last entered atom in the signature list (A_n) is also set to make a new bond to the first atom of the list (A_1), so as to repeat the signature sequence again. This signature list will have the same number of atoms as changing bonds between them, and of course, an equal number of bonds were broken and made.

To include all changing atoms, such a list can start with any chosen atom in the cycle breaking its bond to a neighbor and end with the atom just before the chosen one now making a bond to it, then to start around the cycle again. Hence any atom in the cycle can be the first one listed.

Useful reaction indexing, however, requires a unique list convention for the atoms in the reaction signature list. In most reactions, the first step breaks a bond to hydrogen, implying acid–base catalysis. Since this suggests that the first atom in an ordered list will be H, hence the order of preference of subsequent atoms should be one by the number of bonds each atom has in its normal, neutral state, taking top-row atoms first. This leads to the ordered list convention: $H > X > O > S > N > P > C$.

However, the reactive strand already specifies the sequence of atoms for the signature list, and the first atom in most reactions will be chosen as H, to accord with this convention. Therefore, the common reaction signature will begin as a

list $[H - A_2 + A_3 + \dots]$, etc. with the last atom in the list making a bond to H to repeat the list on through the cycle again. When there are more hydrogens in the cycle, the one chosen first is that with its next atoms A_2, A_3 , etc., in the lowest convention order when there is a choice. Similarly in those few reactions with no hydrogen, the first atom in the signature list will be the lowest in the convention order that breaks a bond. This is characteristic of truly cyclic reactions, most of them pericyclic reactions, as noted below.

While the skeletal (non-hydrogen) atoms are all numbered/specified by automapping reactant atoms with corresponding product atoms, the hydrogens are not so specified but instead are just taken as available from or returned to a general hydrogen pool for acid–base catalysis.

Charges on the atoms are unnecessary and ignored in reaction signatures. Electron pairs that mechanistically engage in the reaction are treated as their conjugate acids. This is simply a matter of replacing any unshared pairs ($A:$) with bonds to hydrogen ($A-H$), as in N: by $N-H$, P: by $P-H$, or C: in carbanions or isocyanides as $C-H$, when it is to exchange with another bond in the reaction. Reactions of $C-H$ ignore the chemistry used for activation of $C-H$; rather this activation is subsumed in the overall signature.

Nitrogen (or phosphorus) with no charge in reactant or product is simply N (or P) in the signature. If either reactant or product is quaternary, as R_4N^+ , the other is treated as $H-N^+$ for the H needed to exchange in the quaternizing reaction; since both reactant and product are understood as N^+ , the $+$ sign is still ignored in the signature.

The reaction signature uniquely describes any reaction. While the reaction cycle is an endlessly repeating list of the atoms that are exchanging their bonds, the bracketed signature list is the unique segment of that cycle containing all bond exchanges once, without repeating. The signature is a list of all the changing atoms ordered by the sequence of the breaking and making of their bonds, starting from the first atom breaking its bond to the second, in the reactive strand. It records only the overall bonding changes ($- + - + -$ etc.) from reactants to products. The sequence of reacting atoms in this signature list is not necessarily the same as the reaction mechanism, which may include steps that cancel out in the overall bonding change of reactants to products.

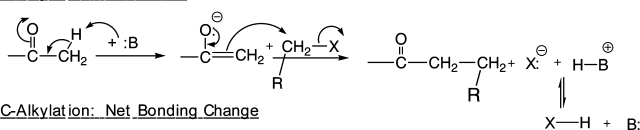
SIMPLE UNIT REACTIONS

The simplest and most common reactions are unit reactions, defined³ as only a single exchange of bonds at each reacting atom. We have shown⁷ that some 80% of reactions in a major reaction database are unit reactions and that about 15% more are just two successive unit reactions, described as composite reactions.

Common unit reactions may be exemplified by C-alkylations, shown in Figure 1.

The mechanism of the reaction has several steps and includes bonding changes between keto–enol forms that disappear in the overall net change from reactants to products. The simplest rendering of the overall net change is just a four-cycle of bond exchanges, listed as $[H - C + C - X +] H - C + \dots$, repeating the cycle. Thus, hydrogen is lost to the hydrogen pool from carbon, and this first carbon makes a bond to another carbon, which in turn loses its bond to halogen (X). The released halogen atom now takes a

C-Alkylation: Mechanism



C-Alkylation: Net Bonding Change

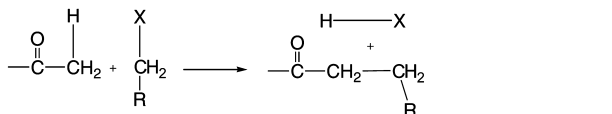
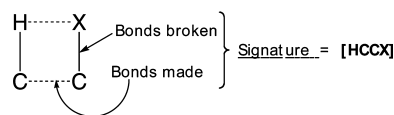
Atom Diagram:
(simplest net bonding change)

Figure 1. C-alkylations: mechanism and signature.

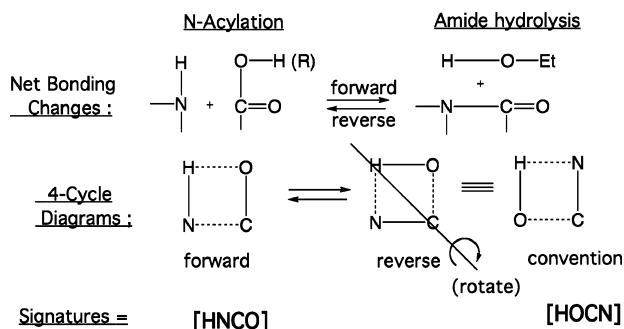


Figure 2. Reverse reactions as atom diagrams.

hydrogen from the pool to recommence this reactive strand, a sequence listed as the signature [HCCX].

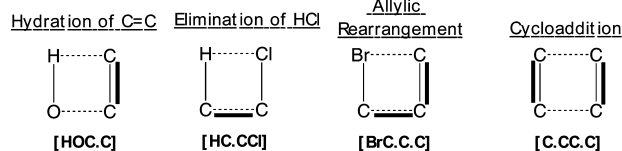
This is represented in Figure 1 by an atom diagram. This atom diagram illustrates in a single structure the changing bonds of the reaction center atoms with alternate solid lines for bonds broken in the reactant and dotted lines for bonds made in the product.³ This atom diagram is read as the signature [HCCX]; this records the reactive strand as the counter-clockwise sequence in the atom diagram of the four single-bond exchanges starting from H – C through to the cycle closure at X + H, so as to make the last bond to form HX. Implicitly this continues through many such identical cycles using the same atom sequence again.

By convention for these atom diagrams of reaction signatures, H (or lowest atom A₁) is placed at the upper left, and the cycle is read counter clockwise, so that the first bond broken is the left-hand, solid, bond of the cycle, and the last bond made is the top one, dotted, returning to the initial atom, H or A₁, so as to repeat the cycle.

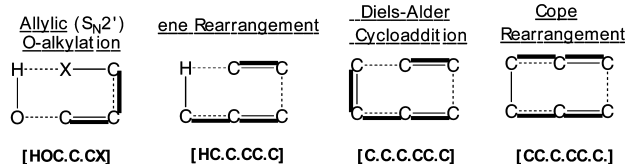
Analogous reactions are O-alkylations, the reacting atoms being O–H + C–X → C–O + H–X, and may be displayed as a similar four-cycle diagram with a signature of [HOCX]. The same signature expresses O-acylation, while N-alkylation or -acylation will be expressed as [HNCX] for a halogen leaving group or [HNCO] displacing an oxygen atom to bond with the starting H, as H₂O or ROH.

The corresponding reverse reaction list will be the same list with the implicit break/make signs (solid and dotted lines) reversed, as shown in Figure 2 but altering the diagram for convention with the breaking H atom at top left on the atom diagram (shown by rotation of the four-cycle square) and first in the signature list. The reverse reaction of N-acylation will then be amide hydrolysis, [HOCN]. Esterification or ester hydrolysis is [HOCO] in either direction.

4-CYCLE FORMS



6-CYCLE FORMS



equivalent 6-cycle forms

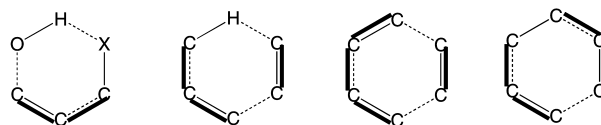


Figure 3. Unchanging skeletal bonds.

The use of H–N⁺ (or H–P⁺) to act for their unshared electron pairs is illustrated by quaternizing tertiary amines in the four-cycle signature, [HNCX]. The deoxygenation of amine oxides by phosphines is also understood as R₃P⁺–H + HO–N⁺R₃ → R₃P⁺–OH + H–N⁺R₃ and has the signature [HPON] with no charges necessary for N and P and with the O–H just carried unchanged from N to P. This is just an internal oxidation of phosphorus and reduction of nitrogen and so not a redox reaction with an added redox reagent, as discussed in the redox section below.

The break/make alternation in signature cycles requires that all simple cycles will be even numbered, four-cycles being by far the most common simple reactions and six-cycles less so. The rarity of higher cycles for unit reactions implies the difficulty of more atoms coming together for a single-bond exchange at each atom in the reaction center.

The generalized four-cycle signature, [HABC], allows a breakdown by chemical function. If atoms AB are carbon and/or nitrogen, then the reaction is a skeletal construction, while if BC are carbon and/or nitrogen, the reaction is a skeletal fragmentation. Functional group exchanges on carbon are the group with atom B = carbon, the leaving group as atom C.

REACTIONS WITH UNCHANGING BONDS

The bonds broken and made above are implicitly σ-bonds. When there are also unchanging bonds linking the atoms in the cycle, the reaction change involved is that of a π-bond with the σ-bond unchanged. The unchanging σ-bonds are called skeletal bonds as they link skeletal (non-H) atoms in the reaction center. The skeletal bonds are indicated in the bracketed signature list by placing a period between any two skeletal atoms that are doubly bonded. These skeletal bonds are distinguished in the atom diagrams as boldface bonds added outside the normal solid and dotted bond lines of the cycle, as exemplified in Figure 3.

The basic reactions with one skeletal bond are additions and eliminations to double bonds, illustrated in Figure 3 with hydration of the C=C π-bond as [HOC.C], and elimination of HCl to form it as [HC.CCl]. Allylic groups will have two adjacent skeletal bonds, as in the rearrangement of allylic

bromides. Photolytic (2 + 2)-cycloadditions of a pair of C=C π -bonds are just [C.CC.C] or with nitrogen the thermal cycloadditions of isocyanates (R–N=C=O) to alkenes as [N.CC.C].

The six-cycle reactions almost always involve π -bonds, with their skeletal bonds marked with periods in the signature, as in the allylic (S_N2') alkylation of an alcohol with an allylic halide [HOC.C.CX], or N-allylation, [HNC.C.CX], both with two skeletal bonds in the six-cycle. The alternation of π -bond breaking and making is implicit in these six-cycle signatures, recognized here as [H–O+C–C–C–X+] ... with the first C.C breaking the π -bond and the second C.C π -bond being formed in the reactive strand.

These allylations are the six-cycle extensions of the simple four-cycle alkylations, [HOCX], [HNCX] in Figure 2. The six-cycle ene reaction is [HC.C.CC.C], with three skeletal bonds. Thus the signature of any simple unit reaction is readily understood, or created by hand, with the aid of a normal atom diagram of the reaction cycle like those in Figures 1–3. For convenience, the six-cycle atom diagrams most simply just extend the common four-cycle form horizontally.

Pericyclic reactions were the model for these atom diagrams to represent cyclic forms^{3,4} and were commonly represented as hexagonal, shown in Figure 3 as alternative equivalent atom diagrams, with the first breaking atom placed at the top position and its breaking bond down to the left. The signature is the same from either form. The pericyclic reactions without hydrogen include the Diels–Alder reaction, which has only carbons and four skeletal bonds; its reaction signature is written as [C.C.C.CC.C], while that of the all-carbon Cope rearrangement [C.C.CC.C.C] is distinguished from it by the difference in the order of the skeletal bonds. The pericyclic Claisen rearrangement is analogous but the signature starts with the lower-order oxygen, [O.C.CC.C.C].

SKELETAL REARRANGEMENTS

Rearrangements are named as such in over 40 of the 250 entries in ref 11. Rearrangements have commonly been used in synthesis for ring expansion or contraction, as formalized in ref 8. They may be conveniently divided into two classes: pericyclic reactions and 1,2 shifts of a migrating atom. The pericyclic reactions are unit reactions, either four- or six-cycle, commonly the latter since they require 1–4 π -bonds, mostly in the form of the Cope rearrangement in Figure 3, with heteroatoms in place of carbon, as discussed for the Claisen rearrangement above.

Common skeletal rearrangements are usually 1,2 shifts of a migrating atom across an unchanging skeletal bond. However, in unit reaction cycles this skeletal bond lies across, not within, the cycle and is called a cross-bond.³ Accordingly, it is an unchanged bond and is not involved in the alternating bond changes of the cycle itself. This cross-bond appears as a boldface bond across the cycle in the atom diagrams of Figure 4; in the signature, the presence of the cross-bond is indicated by labeling its two bonded atoms in the signature with an asterisk.

The basic form of common 1,2 rearrangements (Figure 4) involves a carbon atom (C_M) migrating across two skeletally bonded atoms, A and B, with an outgoing atom (o) breaking its bond to A and an incoming atom (i) forming a bond to B. The three atoms, A, B, and C_M , will appear

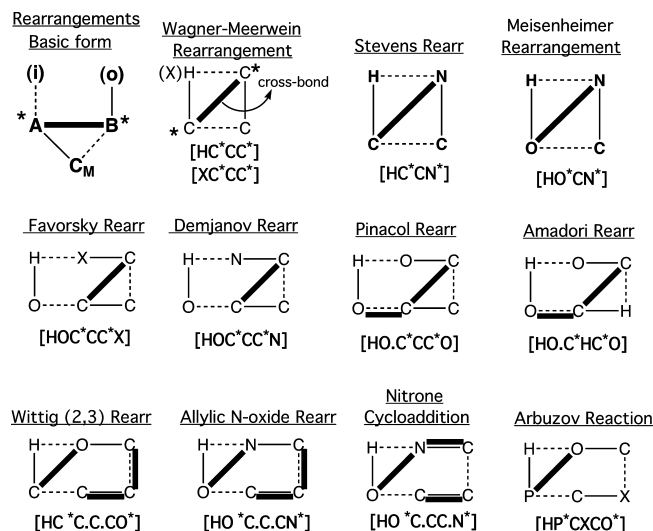


Figure 4. Skeletal 1,2 rearrangements.

together in the cycle as... (o)–A + C_M –B + (i), ... This is seen in the Wagner–Meerwein rearrangement, a four-cycle with the same atom (H or X) incoming and outgoing, and the more common six-cycles below, Favorsky, Demjanov, and Pinacol rearrangements, with different incoming and outgoing atoms.

The form of four-cycle rearrangements is such that either of the atoms above and below the cross-bond may be seen as the migrating atom, i.e., the Stevens and Meisenheimer rearrangements, but the migrating skeletal atom is usually carbon, as also in the Wittig rearrangement [HC*CO*]. In all the signatures of these rearrangements, the cross-bond is designated by a pair of asterisks in Figure 4. The migrating atom appears isolated between the two atoms with asterisks in the signature.

The process outlined for signature creation affords a new general basis for defining rearrangements. The common feature of all 1,2 rearrangements is the triangle in the atom diagram, and it is this common feature—a three-membered semicycle in the reactive strand—that defines a 1,2 rearrangement. In the basic form in Figure 4, atoms A and B may be any skeletal atom, but a migrating skeletal atom is usually carbon. Rearrangements of other skeletal atoms (O and S), not named in ref 11, all have the same form, and their atom diagrams and signatures are analogously created and understood like those above.

However, there is a second category of 1,2 rearrangements having hydrogen as the only migrating atom. This is common in reactions that pass an unshared pair of electrons from one (–) atom to its (+) neighbor, where they share an unchanging bond (cf., $N^+–O^-$), as in “dipolar” cycloadditions and rearrangements. As described above, the unshared pair is treated as a hydrogen and so migrates across a cross-bond, A–B.

The cross-bond here isolates the migrating hydrogen, on the right side of the six-cycle atom diagram, as in the Amadori rearrangement or, more commonly, at the left side because the convention places hydrogen at the upper left, as illustrated by the Wittig (2,3) and allylic N-oxide rearrangements. Many dipolar reactions, not commonly named rearrangements, show this migrating hydrogen, as in nitron cycloaddition, the Arbuzov reaction and Cope elimination [HO*HC.C*N]. In all these H rearrangements, the cross-

bond asterisks in the signature list are more clearly seen as isolating the migrating hydrogen by taking account of the repeating cycle, in which the last starred atom is understood as forming its next bond to H, [HO*C.C.CC*]HO*C.C. ..., as in the signature of the Wittig (2,3) rearrangement.

Most rearrangements are unit reactions, almost none larger than six-cycle. There are some more among the six-cycle reactions with one doubly exchanging atom, discussed below with Figure 8, but few are found with larger cycles.

REAGENTS AND REDOX REACTIONS

Reagents are not considered unless they incorporate their skeletal atoms in the product, in which case they are counted as reactants. Thus with the use of organometallic reagents as catalysts, the reagent is not included, since only the net bonding change from reactants to products appears in the signature and so ignores the catalyst.

The major use of reagents is in oxidation and reduction reactions, in which the overall net bond change between reactants and products is always just the addition or removal of hydrogen atoms and does not need to record the reagent atoms. In the case of simple reductions, the overall net bond change on the reactant is always just +2H, equivalent to hydrogenation, i.e., reaction with H–H. Similarly, the overall change in the product for oxidations is just –2H, equivalent to dehydrogenation (loss of H–H). In each case, multiple redox reactions will be $\pm 2n\text{H}$. With this focus, there is no need to incorporate the molecular changes in the redox reagents since no atoms other than hydrogen from these reagents are found altered in the overall reaction change of reactant to product.

In many examples, however, the net bond change for redox reactions also includes adding X or O atoms to the reactant or removing them from the product. When they appear in redox reactions, halogen atoms can pass between oxidation states⁹ of 1 and 0, oxygen atoms between 2 and 0. These atoms can be used to substitute for the redox reagents actually used that introduce or remove oxygen and/or halogen in any redox reaction.

In order to ignore the actual reagent, any such redox reaction is therefore duplicated by gain or loss of 2HX or H₂O and by appending the missing $\pm 2\text{H}$ to the change in the reaction center atoms. Thus halogenation (oxidation)⁹ is simply +2HX–2H without having to add X₂ to the reactant, and reductive removal of halogen is the reverse –2HX+2H. Similarly, oxidatively adding oxygen to the reactant will be +H₂O–2H and removal is –H₂O+2H. In this way, no halogens (X₂) or hypohalous acids (HOX) or peroxides (or ozone) are needed as reagents in this process, and so the only molecules that need to be added to reactants and products in the signatures of redox reactions are 2nHX and/or nH₂O along with $\pm 2n\text{H}$.

This process encompasses all cases of the same net change no matter what reagent is used and so coalesces all redox reactions with the same net change into one place in the database, to include all examples with different exterior reagents. However, there are more oxidation levels in nitrogen compounds,⁸ and these must be entered directly as reactant and/or product to accommodate their redox reactions.

In creating signatures for any reactions, the changing atoms in the reactants are summed and compared with the sum in

the products to establish the empirical formula change in the reaction, and these sums must be the same. If not, the missing X, O, and N atoms for any reaction are then added to either the reactant or product sum (cf., hydrolysis or dehydration, respectively) as necessary, by adding the elemental hydrides, i.e., HX, H₂O, and NH₃, to the resultant empirical formula change such that all the involved skeletal atoms are the same on each side of the reaction. In the same way, loss of elemental N₂ requires it be added to the product sum. However, all carbons must be seen in the atom sums for both reactant and product.

In this procedure the elemental hydride components, HX, H₂O, NH₃, or N₂, need not have been recognized in advance as essential parts of the reaction but are just discovered by their absence in the empirical formula change. Hence, they will not be omitted inadvertently in signature generation. Finally, for redox reactions only hydrogens are missing from the empirical atom sums, and the difference of $\pm 2n\text{H}$ between the reactant and product sums for any redox reaction is also added to the empirical formula change for generation of the signature.

Reduction reactions (+2H) are now just a sequence with hydrogen first by convention (H – H + A₁ – A₂ +, ...) and so appear in the signature as [HH., ...], which implies initial breaking of H₂ as a reactant for net hydrogenation. With oxidations, as –2H, the sequence will be (H – A₁ + A₂ –, ..., – H +) H – A₁, ..., in which the last H is lost and then adds to the first H in the repeating sequence to form H–H (H₂). This will appear in the signature as [H.H] implying dehydrogenation, with the formation of H₂ as product without regard to the actual reagent used. The simple reduction of ketones to alcohols is [HHO.C] and the reverse, oxidation of alcohols, is [HO.CH], with the ordering convention placing oxygen first before carbon in the signature. The formation of bromohydrin is recognized as the oxidation [HBrC.CO]. Cases of redox reactions by valence change of ambivalent atoms are discussed in the next section.

AMBIVALENT ATOM REACTIONS

Ambivalent atoms¹⁰ are those that can change valence in a reaction, e.g. P, S, and C (in carbon monoxide or isocyanides, etc.). The valence change for phosphorus between phosphines (3) and phosphine-oxides (5) is already covered with the use of P–H for P, outlined above, as is the use of C–H for C: h in isocyanides. This leaves sulfur with three valence or bonding states (2, 4, and 6), and carbon monoxide to ketones, as the major ambivalent situations to consider here.

No valence changes occur in reactions of common atoms (H, X, O, N, and C), and so the signature has an even number of atoms and exchanging bonds, with the bond changes alternating between the atoms, A–B+C–D+E, ... In contrast, signatures with one atom changing valence have an odd number of atoms and bond changes. An ambivalent atom will either make or break two successive bonds in one unit reaction. Therefore, the bond changes on each side of the valence change atom in the sequence of the signature are the same, i.e., ++ or --. Therefore, in all these cases the total of + changes is one more in number than the total of – changes, or vice versa. This results in an odd number of atoms in the signature list as well as an odd number of bond changes.

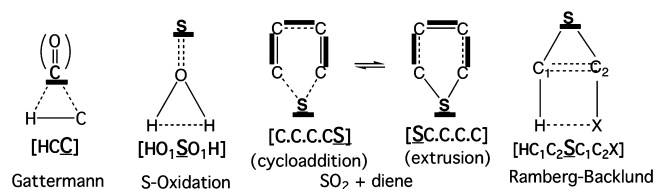


Figure 5. Ambivalent reactions.

Reaction signatures that include these atoms must employ a convention to indicate whether or not a change in valence occurs as part of the reaction. An ambivalent atom that changes valence in the reaction is indicated by underlining the atom symbol in the signature. Several examples are illustrated in Figure 5 with the underlined ambivalent atoms enlarged for emphasis.

The simplest ambivalent reaction is the Gattermann formylation, i.e., $\text{ArH} + :\text{CO}$ to ArCHO , with the three-atom signature, $[\text{HCC}]$. The oxidation of sulfur, as sulfide to sulfoxide to sulfone and also the several levels of oxidation of sulfur acids, all have the simple form $[\text{HO}_1\text{SO}_1\text{H}]$, as shown. The cycloaddition of SO_2 to a diene as $[\text{C.C.C.CS}]$ and its reverse, the extrusion of SO_2 from a cyclic sulfone, $[\text{SC.C.C.C}]$, are also diagrammed in Figure 5. Their five-atom signatures are reversed but reordered to accord with convention. The Ramberg–Backlund reaction forms a double bond by extrusion of SO_2 and has a seven-atom signature, $[\text{HC}_1\text{C}_2\text{SC}_1\text{C}_2\text{X}]$.

In general, ambivalent reactions are internal redox reactions: reduction of the ambivalent atom accompanied by oxidation in the product, or vice versa. When the ambivalent atom is incorporated in the isolated product, as in the examples of Figure 5, there is no net change in oxidation state or in hydrogens. However, if the ambivalent atom is just a part of an external reagent and not incorporated in the net change of reactant to product, it has no place in the signature, and its net effect is replaced by the redox ($\pm 2n\text{H}$) process described above.

SUMMARY OF REACTION SIGNATURES

In order to establish the validity and accuracy of the signature concept, the signatures of all the 250 named reactions in an excellent compendium¹¹ were created following the procedure of Figure 1. These are all listed in order in the Supporting Information, but a summary is presented in Figure 6, grouped by unit reactions (A) and those with multiple bond exchanges, i.e., section (B), with only one exchanging atom and section (C) with more.

The total of 250 named reactions is expanded to 334 in this summary since a number of these reactions have multiple examples in the compendium. Unit reactions (section A), having only one bond exchange per atom in the reaction center, are the most common, and their signatures are easily read and understood, related to simple atom diagrams as in Figures 1–4 for familiar chemical recognition.

Many common unit reactions have no “name” in common use and so are not included in this compendium, i.e., such general reactions as simple substitutions, additions, hydrolysis, etc. Some of these, like hydrolytic workup, have signatures incorporated into the named reactions, others are simple and easily coded with signatures as in Figure 1, where

A. Unit Reactions	Total	Redox
4-cycle =	102	21
6-cycle =	73	19
8-cycle =	4	0
Ambivalent =	12	2
	191	
B. Double Exchange Reactions		
6-cycle =	17	9
8-cycle =	7	2
	24	
C. Multiple Exchange Reactions		
8-cycle =	43	9
10-cycle =	22	2
12-cycle =	26	5
14-cycle =	12	3
16-cycle =	9	1
> 16-cycle =	7	2
	119	
TOTALS	334	73

Figure 6. Summary of reaction signatures.

the signature $[\text{HCCX}]$ is the same for several named alkylation reactions.

In fact the 102 four-cycle reactions in Figure 6, that are listed by name, contain only 50 different reaction signatures and $[\text{HCCX}]$ account for 10 of those listed. However, the given signature is always unique in the sense that it is defined just by the reacting atoms that change their bonding in the course of the conversion of the reactants into the products of the reaction. Therefore, it is no surprise that 10 net C-alkylations appear in the compendium under the names of different chemists who carried them out under different conditions or on limited, defined molecule types. Thus the Friedel–Crafts reaction is $[\text{HCCX}]$ under acid conditions as is also the basic $\text{S}_\text{N}2$ alkylation by halides.

As noted above, skeletal restrictions on the reactive strand limit the size of any unit reaction center, four-cycle reactions being the most common, while six-cycle examples are mostly rigidified by skeletal bonds. Of the few eight-cycle unit reactions, the dienone-phenol rearrangement $[\text{HC}^*\text{CC}^*.\text{C.C.C.O}]$ is skeletally rigid, incorporating the skeletal bonds of its concurrent aromatization. Larger cycles appearing in Figure 6 do not necessarily have more atoms in the reaction center since, in reactions with multiple exchanges on some of the same atoms, these atoms appear more than once in the signature.

There are 64 combinations possible for the common four-cycle unit signatures, generalized as $[\text{HABC}]$, with the atoms A, B, and C chosen from X, O, N, and C. Removing those which make or break weak bonds of O and X with each other, there remain only 40 possible chemically viable combinations, and all are represented by examples in the compendium. There are also 32 possible four-cycle redox reactions, 16 reductions as $[\text{HHBC}]$ and 16 oxidations as $[\text{HABH}]$. Of these possible signatures, only 9 reductions and 14 oxidations are present in the compendium, implying there are some reactions left to be invented (and named).

With one skeletal bond the atom B requires a valence of 3 and so cannot be oxygen, so the possible combinations are only 24 elimination reaction signatures, $[\text{HA.BC}]$, and the 24 reverse, addition reaction signatures, $[\text{HAB.C}]$. With

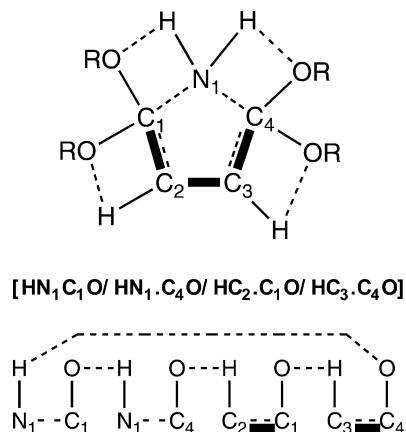


Figure 7. Paal–Knorr pyrrole synthesis; extended atom diagram.

two skeletal bonds, the 18 possible signatures [H.A.B.C] are all variants of keto–enol tautomerizations.

Many longer reaction signatures are just composites of one or more successive unit reactions. The simple aldol reaction, [HCCO], may be followed mechanistically by dehydration, the overall change being then [HCCO/HCCO], creating the two bonds of the double bond in two successive unit reactions. This, an eight-cycle reaction in section C of Figure 6, is in effect two successive four-cycle unit reactions over the same C and O atoms twice and so is an example of a multiple exchange reaction in the next section. Another eight-cycle composite reaction is the Clemmensen or Wolff–Kishner reduction of C=O to CH₂ by two successive identical unit reductions, [HHOC/HHOC]. In each case the successive unit reactions, divided by the slash, if isolated or taken separately, are just simple four-cycle unit reaction signatures taken sequentially and referred to as quartets within the overall signature list.

Many long signatures, rarely longer than 20 atoms, are actually composed this same way of unit reaction quartets (with some intervening unit sextets), the whole understood (and named) as a single reaction with the reactive strand incorporating succeeding quartets over just a few of the same atoms multiply exchanging in the reaction center. When the overall signatures of section C in Figure 6 are divided with slashes into such quartet and sextet units, it is no surprise that 70 of the 119 are just composed of such units all starting with hydrogen.

The Paal–Knorr synthesis of pyrroles from amines and 1,4-diacetals in Figure 7 affords a clear atom diagram of the sequential bond exchanges in each quartet. However, the quartets themselves could be ordered in several ways in the signature since each begins with H, taken from a general hydrogen pool and not necessarily from the nearest quartet. The unique order of the quartets in the whole signature is fixed by the ordering convention. The whole signature is spelled out in an extended form in Figure 7 to see all of the four quartets sequentially as a long reactive strand. The atoms are numbered here not just in sequential order but also to identify the 1,4-diacetal carbons.

Another example is the oxidative synthesis of pyridines from 1,5-diketones and ammonia, as [HNCO/HNCO/HNCO/HC.CH/HC.CO] with 20 bond changes in five quartets. All the five quartets begin with a loss of hydrogen, the fourth quartet being oxidative with a loss of H–H.

Creating a long signature by hand starts with assembling all reactant bonds that are broken in the reaction, then all product bonds that are newly formed. In a unit reaction, the two totals will be equal, and their sum is the length of the signature cycle list.

MULTIPLE EXCHANGE REACTIONS

One Atom Exchanging Two Bonds. As we have seen in the foregoing discussion, a complete cycle is a sequence of single exchanges of bonds at each reacting atom in the reactive strand that returns from the last to the first atom at the end to repeat the strand. The order of atoms in the signature list is the sequence of the alternating exchanges of the bonds between them in the reactive strand. With simple unit reactions, the reacting atoms lie on a simple cycle, and each atom has just a single bond exchange.

When one atom undergoes two bond exchanges, the single cycle is maintained but it crosses over itself at that atom. If we envision this simple atom cycle as a string loop or necklace, when twisted once it becomes a figure eight, two semicycles with one atom at the crossover juncture. Following the whole cycle from the start, we will now encounter the crossover atom twice, defining it as one atom with two bond exchanges at different places in the reactive strand. The crossover atom will also appear twice in the signature list, once for each single bond exchange it undergoes en route.

The general convention to identify atoms that exhibit more than one exchange in any sequence is to use subscripts. These subscripted atoms will appear more than once in the sequence list, and the subscripts distinguish them from like atoms that have only one exchange and so only appear once in the sequence list. In this section with only one crossover atom, it will be identified with subscript 1.

In multiple exchange reactions, each recurring atom has its own identifying subscript. The number of appearances of any such recurring atom in the signature list will equal the number of single exchanges it undergoes. While the recurring atoms will already have been numbered in mapping the reaction center, it is clearer if they are renumbered for the signature list by numbering the first recurring atom of each kind to appear in the list with subscript 1.

The length of the signature list is the total number of bond exchanges in the reaction. Only in a unit reaction is the number of atoms in the signature list equal to the total number of different atoms in the reaction center; in cases of multiple exchange, those atoms appear multiple times in the list, once for each subscripted single exchange.

Section B of the summary in Figure 6 collects those reactions with only one atom doubly exchanging. The simplest examples are cycloadditions of single atoms to π -bonds, adding to form three-membered rings, in which the single adding atom forms two bonds, and so must lose two bonds, in the full cycle. This creates an atom diagram with two semicycles (3,3) linked at the crossover atom but still overall a six-cycle of six alternately changing bonds on five atoms. Only the subscripted atom at the crossover exchanges its bonds twice and so is listed twice for a six-atom signature, as shown in Figure 8.

The addition of dihalocarbene (from CHX₃) to a double bond is [HC₁C.CC₁X] with loss of HX, a full six-cycle

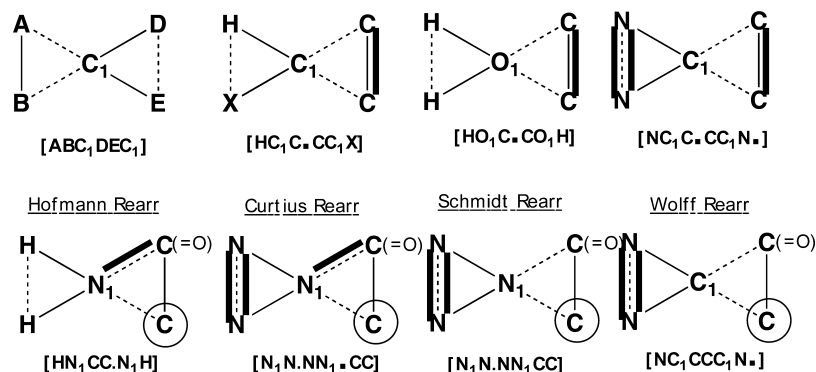


Figure 8. Double exchange at one atom in six-cycle reactions.

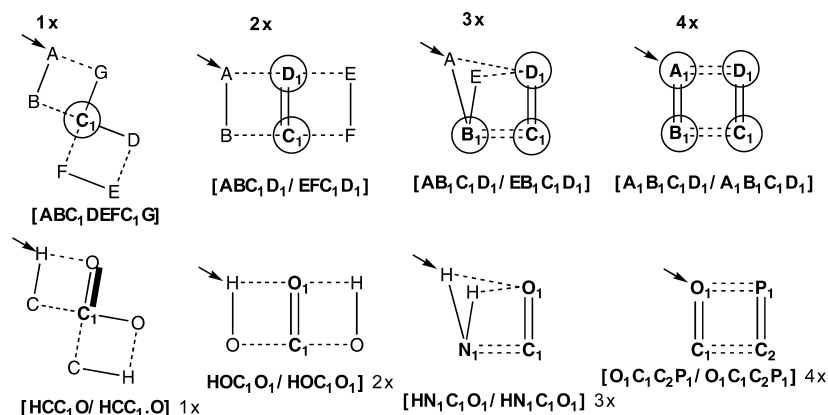


Figure 9. Major forms of eight-cycle reactions.

sequence crossing over at C₁. This carbon forming two bonds is subscripted (C₁) to distinguish it from the two different double-bond carbons (C) with only one exchange each. A similar example is that of epoxidation, seen as the oxidation of water to oxygen and its addition to a double bond, [HO₁C.CO₁H]. The same oxygen atom (O₁) appears twice as it twice breaks a bond to H and twice makes a bond to carbon in the course of the cycle.

The same sequence list without the skeletal bond between carbons is the signature for the oxidative insertion of oxygen in the Baeyer–Villiger reaction, [HO₁CCO₁H]. The signature is also distinguished from the unit hydroxylation of the double bond, [HOC.CO₁H], in which two separate, unsubscripted oxygen atoms are added to a π -bond in a simple six-cycle signature with each oxygen having only a single exchange.

Nitrogen groups with internal charges like azides and diazo compounds are handled via their less stable but uncharged cyclic tautomer, as seen in the fourth example in Figure 8 for diazomethane addition to form a cyclopropane. The neutral cyclic tautomer for CH₂N₂ is taken as initially forming from the more stable charged form, H₂C=N⁺=N⁻, and so eliminates the need to deal with charged species in creating signatures. The same usage is applied to azides, R–N=N⁺=N⁻ on reactions like the Curtius and Wolff rearrangements that form N₂.

As shown in Section B of the summary (Figure 6), double exchange at only one subscripted atom defines two semi-cycles on the atom diagram. The six-cycle reactions can have only (3,3)-semicycles, while the 8-cycle reactions can have (4,4)- or (3,5)-semicycles.

Rearrangements are characterized above as defined by semicycles with only three atoms (cf., Figure 4), and so this

crossover mode in Figure 8 is a model for rearrangements. In the Hofmann and Curtius rearrangements of acid to isocyanate, the unchanging bond defining the rearrangement is not a cross-bond but exists in the cycle itself; the migrating carbon is circled, and the original acid carbon is identified for clarity with (=O). The third bond of the triangle need not even be an unchanging bond but only a bond that is made in the cycle and allows rearrangement to occur, as in the Schmidt and Wolff rearrangements.

Several Atoms Exchanging More Than Once. Reactions with more than one atom having more than one bond exchange now become the norm with eight-cycles and above, enumerated in section C of the summary in Figure 6. Multiple exchanges are implicit in many common reactions, especially those adding or releasing water or ammonia, which may have double exchanges at oxygen and double or triple exchanges at nitrogen. In rare cases carbon atoms can have as many as four exchanges of all its bonds in one reaction sequence.

The common atom diagrams for eight-cycle reactions feature two successive four-cycles linked, in effect a composite reaction of two successive unit reaction quartets. Each quartet can have 1–4 atoms doubly exchanging, the four basic types illustrated in Figure 9 as composite eight-cycles with the double-exchange atoms circled. The four groups are labeled by 1x–4x for the number of doubled atoms, i.e., those with two bond exchanges each. These four forms account for all but 10 of the 43 eight-cycle reactions in the compendium.

The first group, 1x, have just one single atom exchanging twice and so are the eight-cycle equivalents of the six-cycle crossover reactions of Figure 8. The 4x group has all atoms with two (or three) bonds and so cannot include H. The

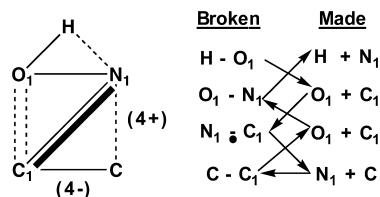


Figure 10. Derivation of a signature: beckmann rearrangement.

1x–4x designations are usefully appended to the bracketed signature lists to quickly identify their atom diagram form.

Each atom diagram form in Figure 9 is exemplified below with actual common reactions: 1x = double Grignard on ester; 2x = acetal formation from carbonyl; 3x = imine or oxime formation from carbonyl; and 4x = Wittig reaction. The arrow at upper left on each atom diagram indicates the first atom to start the eight-cycle sequence for the signature list, which in turn separates the two quartets with a slash in the signature notation.

The 10-cycle reactions can similarly be designated by the overlap of one quartet and one sextet as semicycles and 12-cycle reactions by two sextets. These will show their signature lists divided by slashes into quartets and sextets and similar 1x–4x forms for atoms doubly exchanging bonds en route. The hydrolysis of nitrile to acid has three hydrolysis quartets, like those in Figure 2, therefore a 12-cycle signature, [HO₁C₁N₁/HO₁C₁N₁/HOC₁N₁]. This provides for replacing of all three bonds from carbon (C₁) to nitrogen (N₁) by bonds to oxygen, all on the same carbon (C₁) but one oxygen (O₁) listed twice as forming a carbonyl, the other, nonscripted, oxygen forming the C–OH bond of the product acid.

Rearrangements are rare in these higher cycles, only a few having a three-membered semicycle in the atom diagram of the reactive strand. The Ciamician–Dennstedt rearrangement enlarging pyrroles to pyridines is a 12-cycle reaction with two overlapping 3-rings in its atom diagram. There are also reactions not named as rearrangements which show this feature; the 12-cycle Fischer indole synthesis has three semicycles of sizes 3–5, of which the 3-ring encompasses the rearrangement of C across the N–N bond.

Following the method above for creating signatures from bonds broken and made, the simple procedure for creating more complex but unique signatures by hand is outlined in Figure 10 for the signature of the Beckmann rearrangement. The first step is to assemble all the reactant bonds broken into one list, with the atoms ordered by increasing valence, and then do the same for the atoms making new bonds in the products. The two lists will be the same size and are taken in alternating order as shown to spell out the reaction signature, again with attention to the ordering convention for possible alternatives at each multiple exchanging atom.

The eight-cycle Beckmann rearrangement of oximes to amides is not an example of the common eight-cycle reaction forms in Figure 9. The Beckmann eight-cycle crosses over itself more than once, showing three atoms (O₁, C₁, and N₁) with two exchanges each, but the total of eight changing bonds is evident in the atom diagram, and the signature list derives directly from the diagram, i.e., [HO₁C₁.N₁CC₁O₁N₁], with eight bond changes (four + and four –). Here the asterisks for rearrangements are not required as the cross-bond lies directly on the eight-cycle. Surprisingly, but without the unchanging skeletal bond, the same signature character-

Signature generated to convention:

[H - O₁ + C₁ . N₁ + C - C₁ + O₁ - N₁ +]

H - O₁ + C₁ - C (rejected path: N > C)

[HO₁C₁.N₁CC₁O₁N₁]

izes the overall change in the Lossen rearrangement, i.e., [HO₁C₁N₁CC₁O₁N₁].

PRACTICAL CONSIDERATIONS

For practical use, the conversion of any defined structural change of reactants to products can be reduced to its unique signature by the methods discussed here, currently being developed into a computer program for automatic signature generation.

The actual reaction being recorded may sometimes be understood to be just one step in a synthetic sequence of several laboratory steps to a more general end. While any single reaction often includes several mechanistic steps, so also an understood “reaction” may be conceived and entered as an overall single signature of several laboratory steps, just defined as usual by the structures of its initial reactant(s) and final product(s). When just one reaction in such a sequence is separately recorded, the overall signature can also be added to its index separately to allow it to be retrieved later more generally along with other entries for the same synthetic end which utilize different intermediate steps. The simplest case of multiple reaction steps in one signature is just that of the main reaction with a hydrolytic workup added.

As illustration, replacement of carboxyl by amino can be done with Hofmann, Curtius, Lossen, etc. reactions, all with the implicit RCOOH to RCOX to start, summarized in Figure 11. The overall signature of RCOOH + NH₃ → RNH₂ + CO₂ is oxidative and simpler than the various intermediate steps of those named reactions, i.e., just [HO.CCNH] as the overall conversion of an acid and ammonia to an amine.

Similarly, several ways to make the reagents of the Wittig reaction may be recorded, but some overall conversion should also be recorded with it so that it may be retrieved with other Wittig reactions, i.e., as in a general conversion of R–CH₂X + R'–CHO → R–CH=CH–R', an overall reduction with a 10-cycle signature [HHXC₁C₂O₁/HC₁C₂O₁], forming a product double bond in two steps as with the dehydrative aldol reaction above. A number of other named cases with individual laboratory steps and overall signatures will be found in the examples listed in the Supporting Information.

The practical application of the signature generator begins with a chemist entering the reactions done on a given day into an electronic laboratory notebook directly when they are performed. For this procedure to work, it is essential that the user be given strict instructions for entering the reactant and product structures into the computer with a drawing program or by their molfiles. Only one single reaction is to be entered, without alternate products.

All compounds with skeletal bonds whose atoms appear in both reactants and products must be entered, but compounds with no skeletal bonds, i.e., HX, H₂O, and NH₃, are not entered since they are deduced from the overall empirical formula change in the mapping of total reactant(s) to

- (5) Hendrickson, J. B. Comprehensive System for Classification and Nomenclature of Organic Reactions. *J. Chem. Inf. Comput. Sci.* **1997**, 37 (5), 852–860.
- (6) Hendrickson, J. B.; Zhang, L. Duplications Among Reaction Databases. *J. Chem. Inf. Comput. Sci.* **2000**, 40 (2), 380–383.
- (7) Hendrickson, J. B.; Miller, T. M. Reaction Indexing for Reaction Databases. *J. Chem. Inf. Comput. Sci.* **1990**, 30, 403–419. *J. Org. Chem.*, **1992**, 57 (2), 988–994.
- (8) Hendrickson, J. B. Fragmentations and Rearrangements in Organic Synthesis. *J. Am. Chem. Soc.* **1986**, 108, 6748–6753.
- (9) Ferguson, L. N. A simple tutorial on oxidation states at undergraduate level. *J. Chem. Educ.* **1946**, 23, 550–552. Hendrickson, J. B. *J. Chem. Educ.* **1978**, 55, 216–220. 1985, 62, 245–249.
- (10) Arens, J. G. A formalism for the classification and design of organic reactions. *Recl. Trav. Chim. Pays-Bas* **1979**, 98, 155, 395, 471.
- (11) Kurti, L.; Czabo, B. *Strategic Applications of Named Reactions in Organic Synthesis*, Elsevier Academic Press: Amsterdam, The Netherlands, 2005.

CII000482