

- Relevant publications are H. R. Schenk and F. Wegmüller, "Substructure Search by Means of the Chemical Abstracts Service Registry II System", *J. Chem. Inf. Comput. Sci.*, **16**, 153-61 (1976), for the BASIC system; and J. Page, R. Theisen, and F. Kuhl, "The Walter Reed Army Institute of Research Chemical Information System", in "Retrieval of Medicinal Chemical Information", ACS Symposium Series, American Chemical Society, Washington, D.C., 1978.
- (8) M. Milne, D. Lefkowitz, H. Hill, and R. Powers, "Search of CA Registry (1.25 Million Compounds) with the Topological Screens System", *J. Chem. Doc.*, **12**, 183-189 (1972).
 - (9) P. G. Dittmar, R. E. Stobaugh, and C. E. Watson, "The Chemical Abstracts Service Chemical Registry System. I. General Design", *J. Chem. Inf. Comput. Sci.*, **16**, 111-121 (1976).
 - (10) A. Feldman and L. Hodes, "An Efficient Design for Chemical Structure Searching. I. The Screens", *J. Chem. Inf. Comput. Sci.*, **15**, 147-52 (1975).
 - (11) L. Hodes, "Selection of Descriptors According to Discrimination and Redundancy. Application to Chemical Structure Searching", *J. Chem. Inf. Comput. Sci.*, **16**, 88-93 (1976).
 - (12) A. Feldman, "An Efficient Design for Chemical Structure Searching. III. The Coding of Resonating and Tautomeric Forms", *J. Chem. Inf. Comput. Sci.*, **17**, 220-223 (1977).
 - (13) The WRAIR system uses the SSS screens for full structure search. See L. Hodes and A. Feldman, "An Efficient Design for Chemical Structure Searching. II. The File Organization", *J. Chem. Inf. Comput. Sci.*, **18**, 96-101 (1978).
 - (14) L. Hodes, "A Square Root Algorithm for Inclusive Matching. Application to Chemical Structure Searching", Proceedings of the Conference on Computer Graphics, Pattern Recognition, and Data Structures, IEEE Computer Society, 1975.
 - (15) This file contained about 17 000 entries, representing a random sample of the quarter-million entry WRAIR file.

A Systematic Organization of Synthetic Reactions^{†,1}

JAMES B. HENDRICKSON

Edison Chemistry Laboratories, Brandeis University, Waltham, Massachusetts 02154

Received September 12, 1978

A base of description of organic structure first defines the kind and number of four important kinds of attachments to a single carbon. From this may be derived two interconnected descriptions of a reaction. Linear digital descriptions of the strand of carbons involved in the reaction may be listed for substrate and for product. Also the change in the reaction may be denoted by the kind of attachment made and that broken for each carbon. A single reaction step is defined as one with only a unit exchange in attachment at any involved carbon. This system allows a systematic organization of all possible synthetic reactions in a simple but rigorous format. The basic parent reactions are developed first, then a set of ways to modify these to more complex variants. The reactions can be presented in a compact graphical form, and the system can be utilized to describe and find all possible pathways for multistep reaction sequences between specified generalized substrates and products.

The reactions of organic chemistry constitute an enormous quantity of information, their number reflecting the fineness of description. There are also many possible ways to organize this collection for presentation, depending on the focus of interest, and no single mode dominates present practice. The system developed here is aimed at the needs of the synthetic chemist and consists of a set of nested categories, the logic of which assures that all possible reactions are included. Hence any possible reaction has a place in the system fixed by its logic. This approach assures that one can find any wanted reaction and also can see what reaction possibilities are yet unknown in practice and waiting to be invented.

Chemists commonly express generalized reactions as partial structures of substrate and product bearing only those atoms and bonds which change in the reaction. With several adjacent carbons involved, there can be a great number of possible variants. The present system offers a concise logical notation of these part structures in order to generate and organize all possible variants in a structured way. The basis for the notation is a numerical, or digital, one easily adapted to a rigorous system, the completeness of which is then simply defined by mathematical combinations, unbiased as to current chemical practicality.^{1b}

The system describes each involved carbon separately in terms of four kinds of attachment basic to synthetic operations. Any reaction or reaction sequence is then the net change in these attachments at each involved carbon, between the

substrate and product partial structures. The kinds of attachments defined for any carbon are: H for hydrogens, R for single (σ) bonds to carbon, Π for double (π) bonds to carbon, and Z for bonds (σ or π) to electronegative heteroatoms.

The numbers of each kind of attachment are, respectively, h , σ , π , and z , limited by valency to $h + \sigma + \pi + z = 4$. The "skeletal level" of any carbon is then its value of σ ($\sigma = 1$, primary; $\sigma = 2$, secondary, etc.), and the "functional level" (f) is the sum of attachments to heteroatoms (z) and Π bonds to other carbons (π , limited to 2), symbolized as $f = z + \pi$. The distinction in kind of functionality is denoted by placing one or two overbars over the value of f to indicate $\pi = 1$ or 2, respectively (examples in Figure 1). Metals or other electropositive attachments are included in H, i.e., as the conjugate acid. Hence any carbon is characterized by two digits (each of 0-4), the skeletal value, σ , and the functionality value, f , such that $\sigma + f = 4 - h$. The oxidation state of any carbon is then simply given by $x = z - h$, with values of $-4 \leq x \leq 4$. The functional oxidation state, $x' = 2f - \pi$, is useful in describing reactions and is discussed below. This digital mode of description is based on fundamentals and is very easy to assimilate. For cataloging purposes and for most general use, the system does not require the use of a computer. Used by hand, this notation system can drastically simplify the chemist's consideration of synthetic problems.

CHARACTERIZATION OF REACTIONS

In any synthetic reaction only a few of the involved carbons change their attachments, and these carbons are always

[†] Presented in the symposium, "Retrieval, Analysis, and Indexing of Chemical Reactions", 176th National Meeting of the American Chemical Society, Miami Beach, Fla., Sept 12, 1978.

Kind	Number	
H	h	
R	σ	Skeletal level
Π	π	Functional level
Z	z	
$\Sigma=4$		$\sigma+f=4-h$
		$f=z+\pi$
		Oxidation state, $x=z-h$
		Functional oxidation state, $x'=2f-\pi$

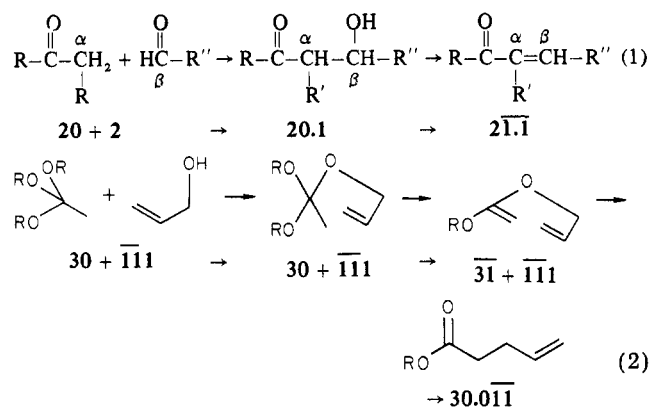
SKELETAL	FUNCTIONAL*	
$\sigma=0$	$f=0$	saturated hydrocarbon
1 Primary	1	$>C-Z$ (ROH, RCl, RNH ₂ , etc.) $\bar{1}$ $>C=C$
2 Secondary	2	$>C=O$ (RCHO, R ₂ CO + derivs.) $\bar{2}$ $Z-C=C$
3 Tertiary	3	$-COOR, CN$, etc. (acid derivs.) $\bar{3}$ $-C\equiv C$
4 Quaternary	4	$CO_2, ClCOOR$, etc. (CO ₂ derivs.) $\bar{4}$ $Z-C\equiv C$

(* unconnected bonds = R, H)

Figure 1. Numerical characterization of structure.

adjacent, generally as a linear string or strand of involved carbons. The structures (or partial structures) of the several successive molecules in a sequence of reactions may then be characterized by a list of f values for this reacting strand of carbons, known as the " f -list". The f -list, and its change from one molecule to the next, contains the functionality information, at one digit per carbon. The skeletal information is implicit in that the carbons with unseparated f values are understood to be bonded together, while those separated by a plus sign are not. It is convenient to show the location of a carbon-carbon bond formed in a construction reaction in the sequence by placing a dot between their f values. The overbars which indicate the Π values are joined across two numbers to show that the Π bond is common to those two carbons. With these conventions the f -lists fully characterize any single reaction or any reaction sequence. The aldol reaction (eq 1) and Claisen rearrangement (eq 2) illustrate the use of f -lists to characterize reactions by their net structural change from substrate to product.

While reactions can be characterized by the involved part structures of substrate and product, they can also be described by a name or label, which describes the nature of the change taking place. A reaction is definable by the net structural change in the four kinds of attachments at each of the involved carbons. In common practice the word reaction is not rigorously defined. In the present system a single *reaction step* is specifically defined as a *single unit exchange* of one kind of attachment (A) for another at each carbon for which these exchanges are linked. A gain or loss of H or Z at any carbon



involves only that carbon, but a gain or loss of R or Π requires a link to an adjacent carbon at which the same gain or loss occurs.

We can begin by considering only one carbon at a time. A *reaction symbol* describes a single reaction step at one carbon as a two-letter designation: first the attachment bond made (+A), then the attachment bond broken (-A). A halide reduction then is HZ and a double bond epoxidation is Z Π at each of the two carbons linked by the double bond. The four attachment letters (H, R, Π , Z) allow $4 \times 4 = 16$ such two-letter combinations as reaction symbols describing all possible single reaction steps at one carbon.^{1b} These 16 reaction symbols are shown in Figure 2. The change in carbon α in eq 1 is RH, that at β is RZ, for the first reaction, and ΠH and ΠZ , respectively, in the second.

Thus one step may involve only one carbon or several adjacent linked carbons if their attachment changes include $\pm R$ or $\pm \Pi$. Actual transformations which are commonly called reactions may contain several reaction steps by this definition, and each step may incorporate more than one carbon. However, most ordinary "reactions" are identical with the reaction steps specifically defined here. A single reaction step may be denoted by a string of the two-letter symbols for the linked carbons.² The steps in the aldol and Claisen reactions of eq 1 and 2 are shown for illustration; constructions are implicit from +R on each of two linked carbons.

aldol (1) (RH·RZ) ($\Pi H \cdot \Pi Z$)

Claisen (2) (ZZ) ($\Pi Z \cdot \Pi H$) (Z $\Pi \cdot R\Pi \cdot R\Pi \cdot \Pi Z$)²

The changes in f -lists for the substrate and product are directly related to these reaction symbols for the converting reaction step, and each may be derived from the other. The

Reaction Symbols: 1st letter = bond made; 2nd letter = bond broken

At any one carbon:	Refunctionalization ($\Delta\sigma=0$)			Construction (+ $\Delta\sigma$)	Fragmentation (- $\Delta\sigma$)	
	Substitution (s=1)*	Addition (- $\Delta\pi$)	Elimination (+ $\Delta\pi$)			Span*
Oxidative (+ Δx)	ZH	Z Π	Π H	RH		ZR s=2
Isohypsic ($\Delta x=0$)	HH ZZ		$\Pi\Pi$	R Π	RR	Π R s>3
Reductive (- Δx)	HZ	H Π	Π Z	RZ		HR s=2

(Reverse reactions have reverse symbols)

* Four reactions with spans of one require no other involved carbons;
12 reactions require other involved carbons and must have spans
greater than one.

Reaction relations: $\Delta h + \Delta\sigma + \Delta\pi + \Delta z = 0$

(at any one carbon $\Delta h = -\Delta\sigma - \Delta\pi$)

or summed over $\Delta z = \Delta f - \Delta\pi$

all involved $\Delta x = \Delta z - \Delta h$

carbons) $2\Delta f = \Delta x + \Delta\pi - \Delta\sigma$

$\Delta x' = 2\Delta f - \Delta\pi$

$\Delta x = \Delta x' + \Delta\sigma$

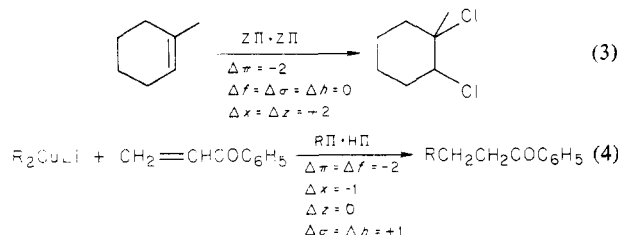
Figure 2. Types of reactions: reaction symbols and net structural change.

relations at the bottom of Figure 2 allow all the changes in numerical values to be derived from the f -list changes and implicit constructions, since these indicate Δf , $\Delta\pi$, and $\Delta\sigma$, which in turn determine Δh , Δz , and Δx at each carbon. These changes may also be derived directly from the reaction symbols, as spelled out in Table I of ref 1b or easily deduced. In the first reaction of eq 1 for the α carbon: $\Delta\sigma = \Delta x = +1$, $\Delta f = \Delta x = \Delta\pi = 0$, $\Delta h = -1$; and for the β carbon: $\Delta\sigma = +1$, $\Delta h = \Delta\pi = 0$, $\Delta f = \Delta z = \Delta x = -1$.

All reaction steps are characterized by three independent variables: $\Delta\sigma$, $\Delta\pi$, and $\Delta x = \Delta z - \Delta h$, which refer, respectively, to construction-fragmentation³ ($\pm R$), elimination-addition ($\pm\Pi$), and oxidation-reduction ($+Z$, $-H$ or $+H$, $-Z$), paired as reversals of each other with reverse two-letter symbols. They may be plotted on a three-dimensional coordinate system of $\Delta\sigma$, $\Delta\pi$, Δx .⁵ More simply the 16 reactions are grouped by symbol and kind in Figure 2. Synthetically the most important reactions are the *construction* reactions ($+R$) which create C-C σ bonds. If these construction steps are collected separately, as well as their reversals, the fragmentation reactions ($-R$),³ then the remaining reaction steps are those which occur with no alteration of carbon skeleton ($\Delta\sigma = 0$), denoted as *refunctionalization* reactions. The first step of eq 1 is a construction ($\alpha\beta = RZ\cdot RH$), the second a refunctionalization ($\alpha\beta = \Pi Z\cdot \Pi H$). Thus, with $\Delta\sigma$ fixed by reaction type, only the two variables ($\Delta\pi$ and Δx) are needed to define reactions of each type.

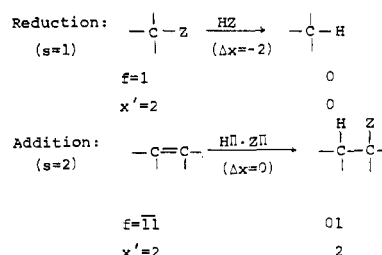
Some reactions (four) occur at only one carbon (Figure 2); the others require functionality changes at two or more adjacent (linked) carbons. The number of adjacent carbons involved in a reaction step is called the *span*, s . This is the number of involved (changing) f values in the f -list or the number of symbols in the reaction symbol string describing the reaction. The simple substitution steps at only one carbon ($s = 1$) are the four in Figure 2 with only $\pm H$ and/or $\pm Z$. Those with $\pm R$ or $\pm\Pi$ necessarily demand changes in attachment at two carbons ($s = 2$) and those with both $\pm R$ and $\pm\Pi$ at three carbons ($s = 3$). However, construction reactions may be characterized as two independent *half-reactions*.⁶ Each half-reaction, RH or RZ , will then require a *half-span* (s') of only $s' = 1$ each, with an overall full construction span, $s = s'_1 + s'_2 = 2$. The $R\Pi$ constructions require a half-span of $s' = 2$. The fragmentations, as reversals of constructions, are analogously described in terms of half-reactions.

Each involved carbon in the reacting span is separately denoted, e.g., $H\Pi\cdot H\Pi$ for hydrogenation of a double bond, $Z\Pi\cdot Z\Pi$ for its halogenation or epoxidation (eq 3), and $R\Pi\cdot H\Pi$ for an addition-construction to a double bond ($C=C \rightarrow R-C-C-H$) as in eq 4, for which only this half-reaction is characterized (the organometallic half-reaction in eq 4, is RH). These symbols show the changes in attachment at each carbon and further imply the overall changes in the numerical values, $\Delta\sigma$, Δf , $\Delta\pi$, etc., summed over all the involved carbons, as shown for the examples in eq 3 and 4. Their interrelations



are summarized in Figure 2, and further examples are collected in Figure 3, some with generalized heteroatoms and some with specific functional groups. It should be noted that for each reaction step the traditional partial structure representation is itself essentially a rigorous, mathematical symbolism and

1. Refunctionalization Reactions ($\Delta x = \Delta x'$)



2. Construction Half-reactions ($\Delta x = \Delta x' + 1$)

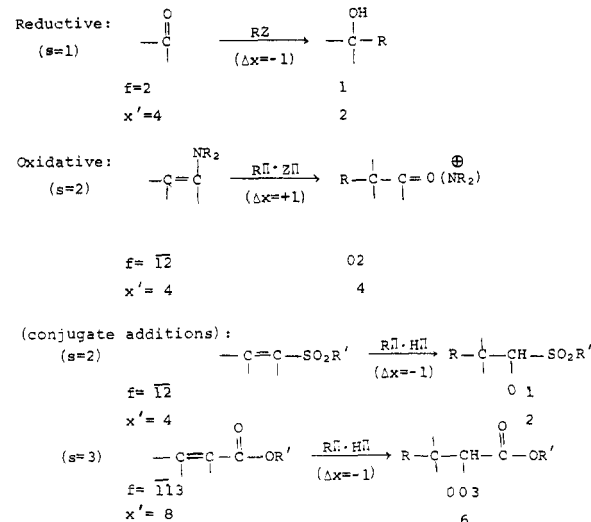


Figure 3. Examples of characterized reactions.

is translated directly into the simpler f -list description shown below it.

The overall oxidation state change in any reaction is Δx , the sum of ($x_{\text{PROD}} - x_{\text{SUB}}$) for all involved carbons, and is related to the summed changes in f , π , and σ as outlined in Figure 2. With partial structures as in Figure 3 it is implicit that the unspecified bonds are not functional, hence are attached to H or R. Although variations there in attachments of H or R do not change the reaction or its f -list description or its overall oxidation state change (Δx), they do change the oxidation states of the individual compounds, substrate, and product. Hence it is convenient to define a "*functional oxidation state*", x' , which relates only to the reactive functionality of a structure, or partial structure, and also gives Δx for their reactions. With this definition the partial structures are freed from consideration of the unspecified bonds and their various possible contributions to the oxidation state. The functional oxidation state is quickly calculated from the f -list with its associated overbars as π values, and examples are shown in Figure 3.

$$x' = 2\sum f - \sum \pi \text{ for a structure or part structure}$$

Hence $\Delta x = \Delta x' + \Delta\sigma$:

$$\left. \begin{array}{l} \Delta x = \Delta x' \text{ for refunctionalization reactions} \\ \Delta x = \Delta x' + 1 \text{ for construction half-reactions} \\ \Delta x = \Delta x' - 1 \text{ for fragmentation half-reactions} \end{array} \right\} \Delta x' = \text{PROD} - \text{SUB}$$

These considerations allow a master chart of all reactions steps to be constructed, plotting reaction families against Δx and $\Delta\pi$, as in Figure 4. In effect the three-dimensional graph is now reduced to two dimensions by separating refunctionalizations from constructions and fragmentations and so removing the $\Delta\sigma$ dimension from the chart. The refunctionalizations are all conveniently named as shown to indicate their changes in Δx and $\Delta\pi$ and these names fall very well into accepted usages as oxidative and reductive eliminations,


















REDUCTIVE				ISOHYPSIC	OXIDATIVE				SPAN
ΔX $\Delta \pi$	-2 REFUNCTIONALIZATION	-1 CONSTRUCTION (Half-reactions) FRAGMENTATION		0 REFUNCTIONALIZATION	+1 FRAGMENTATION (Half-reactions) CONSTRUCTION		+2 REFUNCTIONALIZATION		
+2	Reductive Elimination [RE] ($\Pi Z \cdot \Pi Z$)  $f: \Pi \rightarrow \Pi$ ($x': 4 \rightarrow 2$) 9(2) $\Delta f = 0$ 148(22)	($\Pi R \cdot \Pi Z$)  $f: O \rightarrow \Pi$ ($x': 2 \rightarrow 2$) 13 $\Delta f = +1$ 280		Elimination [E] ($\Pi H \cdot \Pi Z$)  $f: O \rightarrow \Pi$ ($x': 2 \rightarrow 2$) 13(2) $\Delta f = +1$ 280(38)	($\Pi R \cdot \Pi H$)  $f: O \rightarrow \Pi$ ($x': 0 \rightarrow 2$) 13 $\Delta f = +2$ 280		Oxidative Elimination [XE] ($\Pi H \cdot \Pi H$)  $f: O \rightarrow \Pi$ ($x': 0 \rightarrow 2$) 9(2) $\Delta f = +2$ 148(22)	Eliminations (S=2)	
0	Reduction [R] (HZ)  $f: I \rightarrow 0$ ($x': 2 \rightarrow 0$) 4(1) 75(11) $\Delta f = -1$	(RZ)  $f: I \rightarrow 0$ ($x': 2 \rightarrow 0$) 4 75 $\Delta f = -1$	(HR)  $f: I \rightarrow 0$ ($x': 0 \rightarrow 0$) 4 75 $\Delta f = 0$	Substitution [S] (ZZ)  $f: I \rightarrow I$ ($x': 2 \rightarrow 2$) 4(1) 75(11) $\Delta f = 0$	(ZR)  $f: 0 \rightarrow I$ ($x': 0 \rightarrow 2$) 4 75 $\Delta f = +1$	(RH)  $f: 0 \rightarrow 0$ ($x': 0 \rightarrow 0$) 4 75 $\Delta f = 0$	Oxidation [X] (ZH)  $f: 0 \rightarrow I$ ($x': 0 \rightarrow 2$) 4(1) 75(11) $\Delta f = +1$	Substitution (S=1)	
-2	Reductive Addition [RA] ($\Pi \Pi \cdot \Pi \Pi$)  $f: \Pi \rightarrow 0$ ($x': 2 \rightarrow 0$) 9(2) 148(22) $\Delta f = -2$	($\Pi \Pi \cdot \Pi \Pi$)  $f: \Pi \rightarrow 0$ ($x': 2 \rightarrow 0$) 13 280 $\Delta f = -2$		Addition [A] ($\Pi \Pi \cdot \Pi \Pi$)  $f: \Pi \rightarrow O$ ($x': 2 \rightarrow 2$) 13(2) 280(38) $\Delta f = -1$	($\Pi \Pi \cdot \Pi \Pi$)  $f: \Pi \rightarrow O$ ($x': 2 \rightarrow 2$) 13 280 $\Delta f = -1$		Oxidative Addition [XA] ($\Pi \Pi \cdot \Pi \Pi$)  $f: \Pi \rightarrow I$ ($x': 2 \rightarrow 4$) 2(9) 148(22) $\Delta f = 0$	Additions (S=2)	
$\Delta \sigma$	0	+1	-1	0	-1	+1	0		
$\Delta x'$	-2	-2	0	0	+2	0	+2		

Figure 4. Master chart of reaction families.

additions, and substitutions. No confusion results from dropping the adjective isohypsic (= no oxidation state change)^{1b} from those with $\Delta x = 0$ nor from calling HZ, ZH, and ZZ reaction steps ($\Delta \pi = 0$) simply reductions, oxidations, and substitutions, respectively. The abbreviations of the names in brackets provide simple category designations. Similar naming of construction and fragmentation half-reactions is more unwieldy, requiring three words to avoid ambiguity. Their categories may, however, be easily designated by the construction half-reaction symbols (RH, RZ, RII) and their reverse for fragmentations (HR, ZR, IIR).

Any organization mode for all possible organic reactions must necessarily appear to be rather complex, but the categories of Figure 4 do cleanly fulfill this intention from a fundamental basis while still cleaving closely to common usage. In each category is shown the reaction symbol, the actual partial structure depicting in familiar terms the net reaction, and under it the f -list change describing it. The partial structure itself is a mathematical symbolism and the f -list change simply summarizes it in a more condensed, digital form. The change in the functional oxidation state (x') is included in parentheses for each reaction, and Δf is also annotated for each one.

Such a chart must incorporate all possible single reactions since every permutation of possible changes in the four attachment types is included. However, Figure 4 lacks three of the 16 two-letter reaction symbols: HH, $\Pi \Pi$, RR; of the four substitutions with no net change in Δx , $\Delta \pi$, $\Delta \sigma$, and Δf only ZZ is included as substitution. The HH reaction is essentially the simple acid-base reaction⁷ and the RR designation refers to a half-reaction (of $s = 1$), necessarily coupled to two other half-reactions, one construction and one fragmentation, as in RZ·RR·ZR, the three-carbon label of a skeletal (Wagner-Meerwein) rearrangement.⁸ Both of these can be added to the central substitution category of Figure 4 if desired. However, the $\Pi \Pi$ reactions necessarily require three or more adjacent linked carbons ($s \geq 3$). These reactions

will be seen later to be incorporated as vinylogous variations of the basic categories shown in Figure 4. From the 16 basic two-letter reaction symbols, then, three substitutions (HH, RR, $\Pi \Pi$) are not separately placed on the chart, but the addition-elimination reactions require two symbols, thus expanding the number of reaction categories in Figure 4 to 17 (i.e., $16 - 3 + 4 = 17$).

These 17 categories represent all the basic reaction steps of refunctionalization and half-reactions of construction/fragmentation. The categories derive from all possible single attachment exchanges (reaction symbols) at each involved carbon. These in turn directly generate the f -lists for substrate and product and also the derived values for s , x' , Δx , Δf , $\Delta \pi$, $\Delta \sigma$, etc. The partial structures shown are simply a parallel description of each reaction and so are also derived from the same reaction symbols.⁹ All the reaction steps have spans (or half-spans for half-reactions) of one, except the $\pm \Pi$ categories (elimination and addition), which exhibit $s = 2$. Excepting the substitution category in the center block, the 16 remaining categories are all paired as reverse reactions, each reaction of a pair centrosymmetrically placed across the center block in the master chart (Figure 4), e.g., reductive addition [RA], the reverse of oxidative elimination [XE]. The symbols read backwards in reverse reactions, e.g., $\Pi \Pi \cdot \Pi \Pi$ vs. $\Pi \Pi \cdot \Pi \Pi$.

MORE COMPLEX REACTIONS

There are, of course, many more—and more complex—synthetic reactions than just the 17 shown in Figure 4. These, however, are the 17 *parent reaction* steps with the minimum functionality required by the reaction definitions. The rest are just more complex variants of these, generated by applying two kinds of modifications to the 17 parent reactions: (1) extended span; and/or (2) excess functionality within the span of the parent, or extended parent, reaction step. Each category will consist of the parent reaction step and a family of these related modified reaction steps. Each modification may be

(c) Vinylogous extensions are denoted with a prime on the label of the parent category as in Figure 6. Multiple primes can then be used for higher vinylogs, i.e., A'' for 1,6 addition.

All three kinds of span extension exhibit in both substrate and product an increase in x' of two, i.e., $\Delta x' = +2$ for every extended parent, although for higher vinylogs, $\Delta x' = +2n$ (n = number of added vinylogous double bonds). With excess functionality in the span, the substrate and product values of x' will increase further (see below).

2. Excess Functionality. Given a parent reaction step or an extended parent at the minimum necessary functionality, it is an easy matter to generate all the variants made possible by adding extra functionality within the span. The unspecified bonds on the partial structure in each category (and their extended modifications) are assumed to be skeletal (R or H). If they are replaced instead by functional groups (Z or Π) the f -list values increase for both the substrate and product but the net structural change for the step remains the same (Δf , $\Delta \sigma$, Δx , etc.) and so does the span. The number of such possible combinations is clearly defined by the partial structure and the f -list, and these are readily and systematically generated by successive additions of Z or Π to these unspecified bonds in the parent. The number of unspecified bonds on any parent partial structure is $u = s + 2$. This is also true for the extended span modifications except for the contiguous π bonds, which have $u = s + 1$ unspecified bonds available for adding extra functional groups.

These higher functional variants of the parent and extended parent reactions are all at higher oxidation states. If the simple parent has a functional oxidation state of x' for its substrate (as in Figure 4), that for the extended parent substrates will be $(x' + 2)$. Substrate x' values are derived from the reaction symbol and are shown in Figure 4 for each category. As the unspecified bonds are successively used for functional groups (Z and Π) the substrate x' values will increase to a maximum of $(x' + 2u)$ for parent reactions, and $(x' + 2u + 2)$ for extended parents, with all unspecified bonds loaded with heteroatoms.

The 17 categories are characterized by Δx and $\Delta \pi$ (Figure 4), which remains the same in all modifications within the category. The modifications may be cataloged within each category by listing the parent and the three kinds of extensions and then subcategorizing each by x' and π for the substrate. Subcategories for reductive elimination [RE] are illustrated as an example with f -lists of substrate \rightarrow product shown; only the parent span variants are included (no extended span modifications):

	$\pi = 0$	$\pi = 2$
$x' = 4$	11 \rightarrow 11	(parent)
6	12 \rightarrow 12	22 \rightarrow 22
8	13 \rightarrow 13	23 \rightarrow 23
	22 \rightarrow 22	
10	23 \rightarrow 23	33 \rightarrow 33
12	33 \rightarrow 33	

It may be noted here that, as unspecified bonds in the partial structure are preempted for adding extra functionality, there remain fewer links for attaching other skeletal carbons. Hence in the maximum functionality example the whole molecular skeleton is limited to the carbons of the f -list. With refunctionalization reactions these are synthetically trivial reactions. They are recognized as f -lists with each end at $f = 3$ (hence $\sigma = 1$, a terminal carbon), as in the foregoing reductive elimination, $33 \rightarrow 33$ which is $CX_3CX_3 \rightarrow X_2C=CX_2$ or $ROOC-COOR \rightarrow (RO)_2C=C(OR)_2$, etc. No $f = 3$ terminal carbon may appear *within* an f -list. With half-reactions, however, none are trivial since they are coupled to a

I. Refunctionalization Reactions

A.-J. Categories (10): RE, R, RA; E, S (ZZ and HH), A; XE, X, XA

1. Added functionality within span
 - (a) Ordered by π /substrate
 - (i) Ordered by x' /substrate
2. Extended span
 - (a) Added activating functional groups
 - (i) $\Delta s=1$
 - (ii) $\Delta s=2$
 ordered by π and x' /substrate
 - (b) Contiguous π -bonds
 - (i) ordered by π and x' /substrate
 - (c) Vinylogs
 - (i) $\Delta s=2$
 - (ii) $\Delta s=4$
 ordered by π and x' /substrate

II. Construction Half-reactions

A.-E. Categories (5): RZ, R Π , H Π ; RR; RH, R Π , Z Π

1. Added functionality within span as above
2. Extended span as above

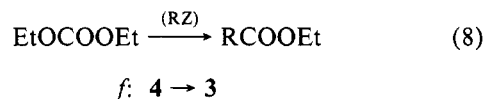
III. Fragmentation Half-reactions

A.-D. Categories (4): HR, R Π , RZ; ZR, R Π , H Π

1. Added functionality within span as above
2. Extended span as above

Figure 7. Reactions catalog: nested listing.

partner half-reaction off any skeletal size, as in carbalkoxylation with a fully functionalized single carbon reagent like chloroformate or carbonate esters (eq 8).



REACTION CATALOG

The organization developed here leads itself easily to an outline form, or nested listing of all reactions, for cataloging and indexing purposes. Such a catalog is outlined in Figure 7. The categories have been listed to group together oxidative, reductive, and isohypsic reactions, and these would be headed in an index by category names. The modification variants are subheaded within each category as shown although other orderings of the modifications are possible and might prove in practice to be better adapted to useful associations of reaction types. A catalog along similar lines, also based on this numerical system, has been appended by Hudrlik to a major reaction survey.¹⁰ This is a very helpful index which shows how such a catalog can be used to index a large reaction compendium.

Most of the practical reactions of synthesis occur among the 150 variants of increasing functionality within sp^{an} , only a few among the much larger number of extended span combinations, and even these are mostly at the lowest functionality level. In the use of an index of this type (Figure 7) the chemist would presumably first convert his reaction of interest to an f -list description and identify its category and modification (as in eq 5-8). This is a simple task and serves to clarify the nature of the reaction as well as to guide one to the correct position in the catalog. This catalog of reactions is analogous to the Beilstein catalog of compounds, and like that catalog also has a prefixed location for any reaction, whether presently known or not.

GRAPH OF MAJOR REACTIONS

Most of the common reactions by far are parent reactions with minimal added functionality. Thus they are mostly

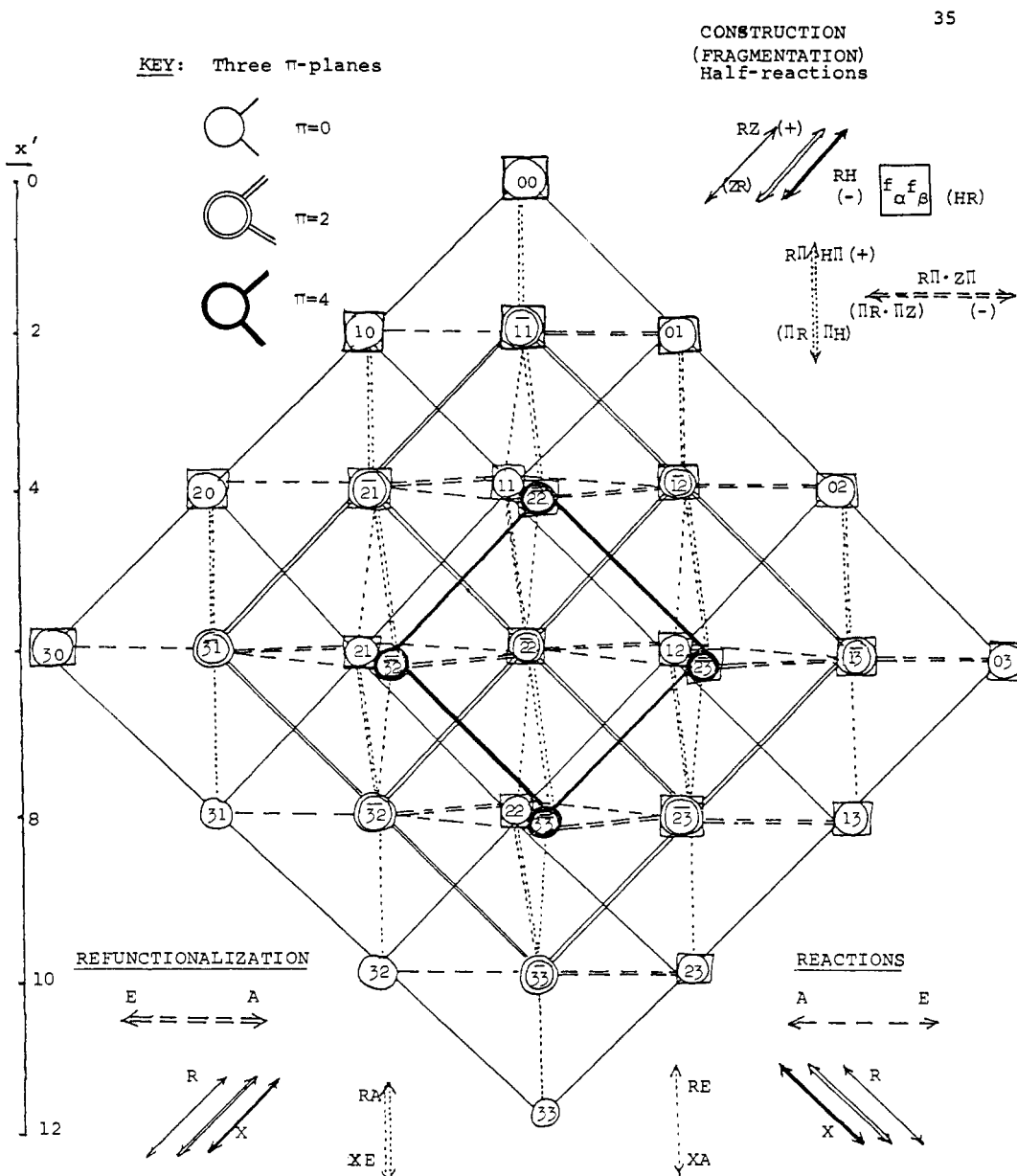


Figure 8. Interconversions graph (span, $s \leq 2$).

reactions with spans of one or two. There are exactly 29 possible f -lists with $s \leq 2$, and these are all the substrates and products of such reactions. Hence a graph of 29 points can serve as a condensed visual presentation with all of their interconversions as the connecting lines of the graph. The graph is very useful for seeing easily all synthetic routes from one partial structure (f -list) to another. Such a graph is shown in Figure 8, the 29 points indicating partial structures (as f -lists) and the lines their interconversions. Single lines are single reaction steps and paths of several successive lines represent multiple reaction sequences. Each line is a pair of reactions, forward and reverse, depending on the vector direction of the line.

The interconversions graph in Figure 8 is three dimensional, with functional oxidation state (x') plotted vertically and π values up from the plane, $\pi = 0$ the base plane and $\pi = 2$ and $\pi = 4$ understood as above it, though necessarily printed in one plane in the figure.¹¹ The double-bond plane ($\pi = 2$) is shown with double lines, the triple-bond plane ($\pi = 4$) in heavy boldface lines. Solid lines lie in one π plane ($\Delta\pi = 0$ reactions); dotted and dashed lines are $\Delta\pi = \pm 2$ reactions. Dotted lines are vertical and are oxidative (down) or reductive (up), $\Delta x = \pm 2$, while dashed lines are horizontal and isohypsic ($\Delta x =$

0) for refunctionalizations. The slope of any reaction line represents its category, reflecting Δx and $\Delta\pi$. Reactions with $\Delta x = \Delta\pi = 0$ which occur without change in the f -list (cf. ZZ reactions) remain at one point on the graph.

Construction and fragmentation half-reactions are a part of Figure 8 as well as refunctionalizations, and a key to the correspondence of slopes and categories for all is shown around the graph. The oxidation state scale of the graph is x' so that only for refunctionalizations is $\Delta x = \Delta x'$ (horizontal lines). For constructions, $\Delta x = \Delta x' + 1$ and hence oxidative construction half-reactions ($\Delta x = 1$) stay at the same x' level. RH half-reactions do not change in f -list and are represented as remaining at the same point (f -list shown in square for possible RH substrates), while oxidative RII's (i.e., RII-ZII) are horizontal (dashed) lines of $\Delta x' = 0$.

The graph is limited only to span, $s < 3$, and so all parent reaction steps are shown as well as those of extra functionality within span and extended variants with added activators ($\Delta s = 1$) for $s = 1$ parents and also their contiguous π -bond extensions. A comparable graph for $s = 3$ reaction steps can analogously be created but is much more complex and extensive and contains only a small proportion of common reactions; hence it is not presented here.

MULTIPLE REACTION SEQUENCES

The definition of a single reaction here, necessary for systematic organization, requires one and only one exchange of attachments per carbon. Many common synthetic reactions, however, require more, e.g., hydration of acetylenes ($22 \rightarrow 12 \rightarrow 02$), hydride reduction of acids to alcohols ($3 \rightarrow 2 \rightarrow 1$), or the Wittig reaction ($1 + 2 \rightarrow 1.1 \rightarrow 1.1$). Similarly the new synthetic methods currently appearing in the literature are very often sequences of several single reactions. It is important for synthesis to be able to catalog methods for converting one molecule to another by way of such multiple reaction sequences. The present system provides a simple basis for such a catalog, and since the system is expressed in numerical terms it is easily capable of finding all possible sequences from one molecule to another, as a mathematical exercise without the bias of current chemical practicality. This is in effect the "functional group switching" problem discussed by Whitlock¹² and reduced here to digital description.

Multiple reaction sequences can contain any kind of reaction step and may be characterized by the f -lists of the starting material(s) and product. Then all sequences of reaction steps which transform the former to the latter may be generated by adding to the several f and π values of the carbons whatever numerical changes describe each of the possible reaction categories and so finally lead to the desired transformation. Such catalogs of possible reaction sequences are currently being generated on our computer.¹³ The value for the synthetic chemist lies in knowing that all possible multistep sequences are obtainable in this way.

CONCLUSION

This system of describing and organizing organic reactions is more than just a new nomenclature. It is a rigorous mathematical form, capable of manipulation by rules which are mathematical operations. As such it can be used systematically to formulate problems and find solutions and combinations not readily accessible to the normal structural reasoning of the synthetic chemist. It has the advantage of offering both a new perspective in approaching problems and also a mathematical certainty of finding all possible answers to a defined problem.

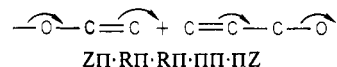
ACKNOWLEDGMENT

The author wishes to gratefully acknowledge support from the Hoffmann-La Roche Foundation and from a National

Science Foundation grant (No. CHE-77-12267).

REFERENCES AND NOTES

- (1) (a) This may be considered paper 7 in a series; paper 6: J. B. Hendrickson, *J. Am. Chem. Soc.*, **99**, 5439 (1977); (b) J. B. Hendrickson, *ibid.*, **93**, 6847 (1971); (c) J. B. Hendrickson, *J. Chem. Educ.*, **55**, 216 (1978).
- (2) These single reaction steps, made of unit exchanges of attachments, are generally directly analogous to common two-electron mechanism steps, as this comparison of the last reaction in eq (2) may illustrate:



The notation, however, carried no implication of mechanistic timing or concertedness.

- (3) The term *fragmentation*, introduced by Grob,⁴ is used to denote cleavage of a C-C σ bond, the reverse of construction.
- (4) C. A. Grob and P. W. Schiesl, *Angew. Chem., Int. Ed. Engl.*, **6**, 1 (1967).
- (5) Such a three-dimensional graph, a tetrahedron of the 24 possible (σf) descriptions of individual carbons, is presented in ref 1c, with three-coordinate axes of σ , π , x . Each of the 24 points is linked to every other to which it may be converted by a single reaction. These lines constitute a graph of all possible reactions of single carbons, and the graph therefore shows $\Delta\sigma$, $\Delta\pi$, and Δx for each reaction.
- (6) J. B. Hendrickson, *J. Am. Chem. Soc.*, **97**, 5784 (1975).
- (7) The HH reaction includes not only $\text{R}^{\cdot-} \rightleftharpoons \text{R-H}$ or $\text{R-M} \rightleftharpoons \text{R-H}$ but also carbanion silylation, $\text{R}^{\cdot-} \rightleftharpoons \text{R-SiR}'_3$ or deboration, $\text{R-Br}'_2 \rightarrow \text{R-H}$, since these are electropositive attachments, like H or metal.
- (8) Apart from skeletal rearrangements, the RR reaction is uncommon, implying displacement of a C-C σ bond by a carbon nucleophile. The opening of activated cyclopropanes by carbanions is an example.
- (9) The derivation of both the traditional partial structures and their digital descriptions from the basic reaction symbols may be carried out stepwise in this way. The number of reaction symbols (pairs of letters) is the span, i.e., the number of digits in both product and substrate f -lists and the number of linked carbons in the partial structures of each. The first and second letters of the reaction symbol at each carbon show, respectively, the parent product and substrate attachments as in Figure 2 (center table), i.e., values of f , π , x' , etc. for each carbon, and so their f -lists. From these are found the overall numerical changes from substrate to product (Δf , $\Delta\pi$, $\Delta x'$, etc.) and their relations in Figure 2 (bottom). The unspecified bonds in the partial structure equal $u = s + 2$.
- (10) P. F. Hudrlik in "Survey of Organic Syntheses", Vol. 2, C. A. Buehler and D. E. Pearson, Ed., Wiley, New York, 1977. This useful index differs from the presentation here in that Hudrlik simply equates functionality and oxidation state, but for indexing purposes this is unlikely to cause any difficulties.
- (11) The third dimension is ($f_\beta - f_\alpha$) and serves only to spread the points. The dimension with fewest values ($\pi = 0, 2, 4$) is chosen for the out-of-plane coordinate. When the graph is presented with the two main coordinates, x' and π , in the plane of the paper (i.e., Figure 8 seen from the side), there are too many points in one place for clear recognition. A version of Figure 8 distinguishing the three π planes in color is easily made and more easily read.
- (12) H. W. Whitlock, *J. Am. Chem. Soc.*, **98**, 3225 (1976).
- (13) The derivation of eq 10 and its use in computer generation of reaction sequences will be developed in a separate publication.