Organic and Biological Chemistry

A Systematic Characterization of Structures and Reactions for Use in Organic Synthesis

James B. Hendrickson

Contribution from Brandeis University, Department of Chemistry, Waltham, Massachusetts 02154. *Received December 28, 1970*

Abstract: A system is offered for mathematical presentation of the structural and functional features of organic molecules and their interconversion reactions. The system offers a logical structure for reaction classification which can be used either to catalog synthetically useful reactions (presently known or unknown) or to establish all possible synthetic routes to given molecular sites. It also provides for a simple linear notation both for structures and their interconversions, hence for synthetic sequences. The system is oriented to use in the design of organic syntheses but is presumably applicable to other problems requiring a systematic view of organic structure and reactions.

In order to put organic synthesis design into a more systematic format it is necessary to orient the vast collection of organic functionalities and their reactions into a more manageable framework. Such a framework must be organized around the needs of synthesis, *i.e.*, the nature of structure and the net change in structure during chemical reaction, rather than around mechanistic understanding of reaction pathways, which is our usual current organization.

A system must ultimately derive directly and clearly from structural fundamentals both so that its tenets are rooted in fundamentals and so that its generality allows inclusion of all new reactions without breakdown of the system, but it must do this without loss of significant information in the process of simplification necessary to its construction. Furthermore, a system should be general so that it defines all possible reactions or meaningful interconversions, including those not presently feasible which might then conceivably be invented to swell the armory of synthetic methods. Finally, the central information available to the synthetic chemist when he approaches a synthesis is the structure of his target molecule, *i.e.*, a twofold recognition of *structure*—the carbon skeleton—and *functionality*—the nature and placement of functional groups on that skeleton. Any meaningful system must reproduce the important elements of this dual description and imply all the possible synthetic routes so that when some are selected as preferable, one knows that all have been at least tacitly examined and that no good route has escaped attention.

A logical framework aimed at fulfilling these criteria is developed here and its application to systematic synthesis design implied; more detailed development of these applications will appear later. The present ideas are offered as a logical basis for systematizing synthesis design; accordingly they should be equally applicable for use by hand or by computer, but neither application is primary at this stage. It is hoped that they may also have future value in areas other than synthesis, such as in codifying structure elucidation, or simply organizing organic structure and reactions for teaching purposes.

We may consider any organic structure as a set of carbon sites, each of which has a characteristic quality both of functionality and of carbon skeleton. Any reaction is then a change in these characteristics at one or more carbon sites on the skeleton. A carbon site may be defined by its four attachments: R = σ bond to carbon,[1] Π = π bond to carbon, H = bond to H or other less electronegative atom,[2] Z = bond to more electronegative atom (heteroatoms, Z = N, O, X, P, S).[3]

The sum of these attachments to any carbon site is four, so that the *numbers* of each kind may also be defined: σ = number of σ bonds to carbon (number of attached carbons), π = number of π bonds to carbon, h = number of bonds to H (or other less electronegative atoms), z = number of bonds to heteroatoms (N, O, X, P, S).

Except for π, each of these values may range from 0 to 4; but the range of π is only 0–2. The π bond to carbon is also anomalous in that it may be regarded as a bond to carbon, structurally, or as a functionality like Z (to which it is related by elimination of HZ). It is therefore convenient to define the *functionality* at a carbon site as the sum of functional heteroatom attachments (Z) and carbon–carbon π bonds (Π): $f = \pi + z$, the number of functional bonds (to heteroatoms or π bonded to carbon).

Hence, for any single carbon site, $\sigma + \pi + h + z = \sigma + f + h = 4$.

The important data about a carbon site are its skeletal and functional characteristics, *i.e.*, the number of its links to other carbons or to reactive functions, and this may be expressed as the *character* (c) of the site: $c = 10\sigma + f$, a two-digit number in which the first

---

(1) The letter R is chosen following the common practice in denoting generalized alkyl substituency; the σ bond is specified so that each R implies a single carbon link to the carbon site described, whether it be by a single, double, or triple bond, each of which has only one σ bond.

(2) The letter H usually refers to hydrogen but may also be an unshared pair of electrons (the conjugate base or carbanion) or a link to a metal or other less electronegative atom such as boron.

(3) Both σ and π bonds to electronegative atoms are counted in Z and not distinguished; *cf.* the carbonyl group represents two Z attachments to the carbon site. Carbonium ions are also represented as Z, implying loss of an electronegative atom attachment with its bonding electron pair.

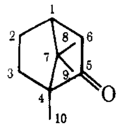
digit (σ) shows the skeletal nature (σ = 1, primary; 2, secondary; 3, tertiary; 4, quaternary) and the second (f) shows the functionality level (f = 1, alcohol, ether, halide, olefin; 2, ketone, aldehyde, acetylene; 3, carboxylic acid family, nitrile; 4, derivative of carbon dioxide).[4]

In arriving at this simple characterization of individual carbon sites by $c = 10\sigma + f$, the major condensation of information lies in the functionality term, f, which includes both π bonds to carbon as well as the various possible heteroatom attachments (Z = N, O, X, S, P, etc.). In practice this represents our recognition that there exist functional families which are easily interconverted for synthetic purposes. Thus simple halides, olefins, and alcohols (f = 1) are interconvertible, as are ketones, ketals, imines, *gem*-dichloro compounds, etc., at f = 2, or the nitrile–carboxylic acid–ester–amide family (f = 3). Thus the generalization leads to an advantage as well as a simplification for synthetic reasoning. The same can be said for the recognition of similarity conferred on tautomers, such as the carbonyl carbon of ketones and their enols, enamines, or enol ethers, or the possible tautomers of complex functions as in citrinin (below); in these cases the system shows each carbon to have the same functional character in each tautomer (same f value) and hence the same synthetic origin. Thus carbon-10 of citrinin is revealed as an aldehyde (c = 12), and its synthesis reflects this. When it is useful to distinguish π from z in the f value part of the character number, c, this may be done by placing one or two bars over the c value to indicate one or two π bonds to carbon. Carbon-10 of citrinin is then $c = \overline{12}$, carbon-6 is 22, and chloroacetylene is $\overline{\overline{12}}$–$\overline{\overline{13}}$.

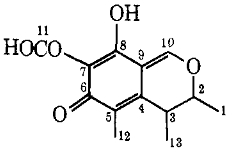
Furthermore, the oxidation state (x) of any single carbon site is easily calculated:[6] $x = z - h = 2z + \sigma + \pi - 4$.

For the overall structure of any molecule, $C_nH_mZ_p$, the number of carbon–carbon links (σ bonds) is $0.5\cdot\Sigma_n\sigma_i$, which equals $(n - 1)$ for acyclic molecules. The number of carbocyclic rings is accordingly $(1 - n + 0.5\cdot\Sigma_n\sigma_i)$. The overall oxidation state of the molecule (specifically of the carbon atoms in it) is given by: $X = \Sigma_n x_i = \Sigma_p v_z - 2b_z - m = \Sigma_n(z_i - h_i)$, where v_z = normal valence of each heteroatom (= 8-periodic table row) and b_z = number of bonds between heteroatoms.

If a given structure of n carbons is numbered, a simple connectivity list may be written to express the carbon skeleton and its functionality. This is a list of site numbers (carbon labels, $i = 1 \rightarrow n$), each with its character ($c = 10\sigma + f$) as a superscript. Each site number is linked by a dash to the adjacent numbers if they are bonded, or separated by a slash if they are not. Sites other than sequential ones which are attached to a given site are then listed in parentheses after it. Such a list has full connectivity information, allowing the skeleton to be reassembled from the list and its functional sites marked.[8] Such lists also have, of course, a linear presentation for computer input. Examples are shown for camphor and citrinin, numbered arbitrarily:



camphor



citrinin

camphor: $1^{30}(6,7)-2^{20}-3^{20}-4^{40}(7,10)-5^{22}-6^{20}(1)/7^{40}(1,4,9)-8^{10}/9^{10}(7)/10^{10}(4)$

citrinin: $1^{10}-2^{21}-3^{30}(13)-4^{31}(9)-5^{31}(12)-6^{22}-7^{31}(11)-8^{22}-9^{31}(4)-10^{12}/11^{13}(7)/12^{10}(5)/13^{10}(3)$

Organic Reactions. With the four defined kinds of attachment to carbon any desired reaction *at that site* may be described as the replacement of one kind by another. A reaction at one carbon site may then be symbolized by a pair of the capital letters (R, Π, H, Z), the first indicating the attachment *formed* in the reaction, the second the attachment *removed*. The first letter is registered as + (group added: +R = creation of σ bond to carbon; +Z = formation of heteroatom bond), the second as − (group lost: −Π = loss of olefin, *i.e.*, addition; −H = loss of hydrogen). Thus HZ is a replacement of heteroatom by hydrogen, *i.e.*, a reduction of halide, ketone, etc., and ZH is the reverse oxidation. HΠ indicates saturation at one carbon of an olefin bond by hydrogen, either by hydrogenation or by addition of HX, H_2O, etc., in which the carbon described receives the proton (the other carbon is then HΠ or ZΠ, respectively). Similarly RH implies replacement of hydrogen by carbon, as at the nucleophilic, carbanion site of an aldol reaction or alkylation, while RZ would commonly denote the other, electrophilic carbon site in such a carbon–carbon bond formation. The four kinds of attachments (R, Π, H, Z) thus lead to seven major reaction descriptions, grouped in Figure 1.[9]

The terms *construction* and *cleavage* have been adopted to describe carbon–carbon σ-bond formation (+R) and cleavage (−R), respectively; generic terms for these reaction types are essential for discussions of synthesis.[10] The 16-line graph in Figure 1 interrelates the major reactions, showing the 16 possible combina-

---

(4) All the functional groups are grouped into four classes according to the number of heteroatom bonds to carbon in section 3-1 of ref 5; these groups are the f numbers here.

(5) J. B. Hendrickson, D. C. Cram, and G. S. Hammond, "Organic Chemistry," 3rd ed, McGraw-Hill, New York, N. Y., 1970.

(6) Oxidation state in organic molecules is only viably described by a separate number (x) for each carbon, the sum being the state of the whole molecule, as shown by Ferguson[7] and recast in ref 5, section 19-1. Oxidation states so computed are generally useful in the balancing of oxidation–reduction equations with inorganic reagents. The generality, f or F, is not useful here, since *all* bonds to carbon (R + Π) count alike in computing oxidation state, *i.e.*, Π is here grouped with R instead of Z. For saturated carbon (π = 0), $x = 2f + \sigma - 4$. For olefinic carbon (π = 1), $x = 2f + \sigma - 5$. For acetylenic carbon (π = 2), $x = 2f + \sigma - 6$.

(7) L. N. Ferguson, *J. Chem. Educ.*, 23, 550 (1946).

(8) The σ value equals the number of adjacent dashes to the site plus the number of sites in parentheses. Other connectivity lists and tables have been advanced, usually in connection with chemical documentation, and are discussed in the pioneering work of Corey and Wipke on synthesis analysis; E. J. Corey and W. T. Wipke, *Science*, 166, 178 (1969).

(9) These descriptions contain the assumption that a reaction may be described by a change in only one carbon attachment, and indeed that is by far the most usual circumstance. However, it may be convenient to symbolize reactions changing two attachments, even though these can usually be described sequentially in the standard symbols. In changing two attachments, the two formed may be listed followed by the two removed in a four-letter symbol, as in hydride reduction of carboxyl to alcohol, HHZZ (or H^2Z^2), which could also be written sequentially as HZ + HZ. In changes involving carbene intermediates the four-letter symbol is more obviously necessary as in the change on cycloaddition of dichlorocarbene from $CHCl_3$ or of diazo compounds to olefins: R^2HZ or R^2Z^2 for the added carbon, respectively.

(10) Construction reactions all have R first; in cleavage reactions R is last. In construction reactions: RH changes usually denote a carbon site acting as a nominal carbanion; RZ changes usually denote a carbon site acting as a nominal carbonium ion.

Table I. Types of Reaction at One Carbon Site

	Symbol	$\Delta\sigma$	Δz	$\Delta\pi$	Δh	Δc	Δx
I. Substitution							
Proton exchange	HH	0	0	0	0	0	0
Carbon interchange	RR	0	0	0	0	0	0
π Rearrangement	ΠΠ	0	0	0	0	0	0
Nucleophilic substitution	ZZ	0	0	0	0	0	0
II. Oxidation–reduction							
Oxidation	ZH	0	+1	0	−1	+1	+2
Reduction	HZ	0	−1	0	+1	−1	−2
III. Construction–cleavage							
Oxidative construction	RH	+1	0	0	−1	+10	+1
Reductive cleavage	HR	−1	0	0	+1	−10	−1
Reductive construction	RZ	+1	−1	0	0	+9	−1
Oxidative cleavage	ZR	−1	+1	0	0	−9	+1
Constructive addition	RΠ	+1	0	−1	0	+9	0
Fragmentation	ΠR	−1	0	+1	0	−9	0
IV. Elimination–addition							
Oxidative elimination	ΠH	0	0	+1	−1	+1	+1
Reductive addition	HΠ	0	0	−1	+1	−1	−1
Reductive elimination	ΠZ	0	−1	+1	0	0	−1
Oxidative addition	ZΠ	0	+1	−1	0	0	+1

Table II. Possible Reaction Types at Two Carbon Sites Only

	Construction (+R)	Cleavage (−R)	Δx
Oxidative (−H or +Z)	RH·RH	ZR·ZR	+2
Isohypsic	RH·RZ	ZR·HR	0
Reductive (+H or −Z)	RZ·RZ	HR·HR	−2
	Elimination (+Π)	Addition (−Π)	
Oxidative (−H or +Z)	ΠH·ΠH	ZΠ·ZΠ	+2
Isohypsic	ΠH·ΠZ	ZΠ·HΠ	0
Reductive (+H or −Z)	ΠZ·ΠZ	HΠ·ΠH	−2

tions which may occur at one site and the associated change in its character. The 16 combinations and character changes are listed in Table I. Each of the 16 possible site changes is shown as a reaction type with its associated changes in site characteristics and oxidation state (Δx); reactions labeled oxidative or reductive show a corresponding oxidation state change. Among substitution reactions, HH is usually trivial, RR is usually rearrangement (migrating carbon), and ZZ is labeled nucleophilic since this is the most usual case but ZZ is not necessarily always nucleophilic. The other reactions are bracketed into forward–reverse pairs.

Description of the total change (at all carbon sites) in a reaction is readily obtained by listing the reaction changes at each involved carbon site sequentially, *e.g.*, RH·RZ denoting an aldol-type reaction, HΠ·ZΠ an addition of HBr to an olefin. The twelve combinations representing the possible reactions involving just two carbon sites are shown in Table II; the upper group (±R) making and breaking σ bonds between carbons, the lower group (±Π) involving only π bonds. The term *isohypsic* ("equal level," from the Greek *hypsos*, level) is introduced to describe reactions that occur with no change in oxidation state ($\Delta x = 0$). The main ionic reactions for carbon–carbon bond formation are isohypsic constructions, *e.g.*, enolate alkylation and acylation, organometallic additions, etc.; and ordinary acid- or base-catalyzed eliminations or additions are isohypsic also, *e.g.*, dehydrohalogenation (ΠH·ΠZ).

Clearly, reactions involving more than two carbon sites are easily symbolized, listing the symbols for each carbon in the order the carbons assume in the reaction. Several examples are appended: Michael addition, RH·RΠ·HΠ;[11] Wagner–Meerwein re-

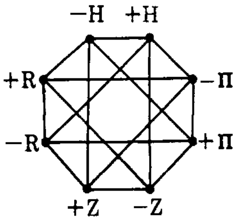
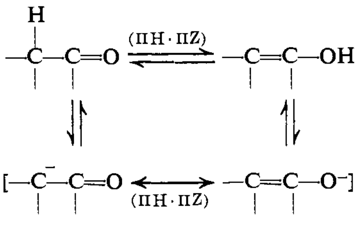


Figure 1. Major reaction groups and interrelations: substitution (no character change), oxidation (−H or +Z), reduction (+H or −Z), elimination (+Π), addition (−Π), construction (+R), and cleavage (−R).

---

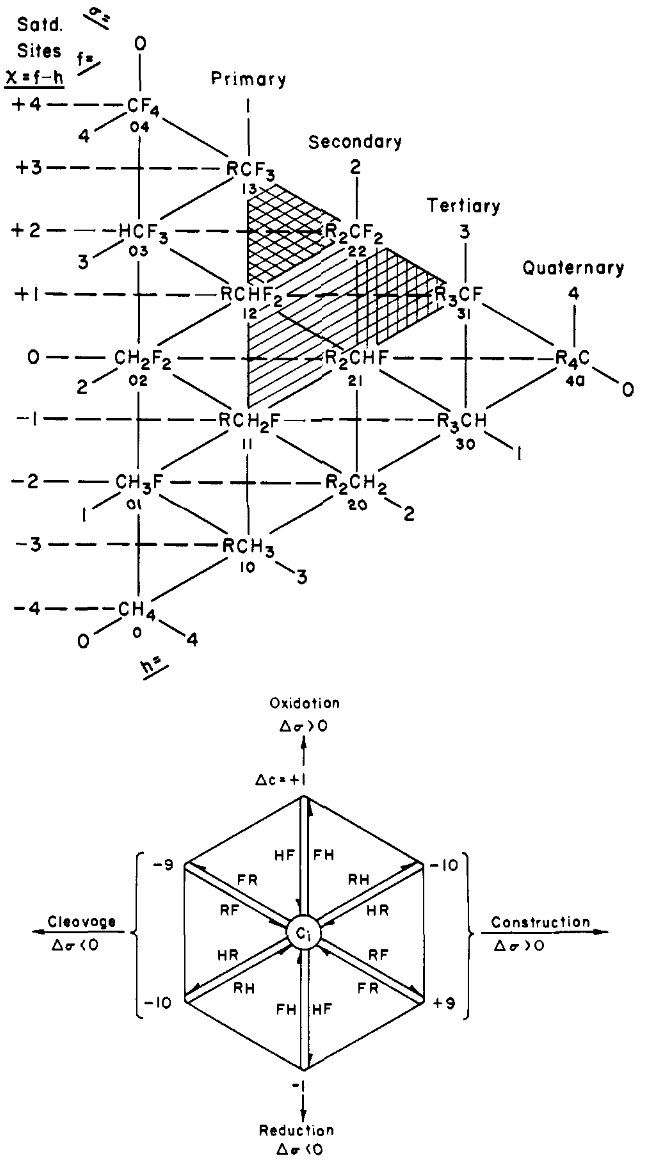
[11] The system, while basically intended for overall structural change in any reaction, is equally suitable for description of reaction intermediates and also of resonance. The notation treats the formal electron shifts between canonical resonance forms as reactions (just as the curved arrow convention does), parallel to the corresponding tautomerizations. The intermediates in the Michael addition may be char-



arrangement, ZR·RR·RZ; allylic rearrangement (or SN2′), ZΠ·ΠΠ·ΠZ; enamine acylation, ZΠ·RΠ·RZ; common fragmentation, ΠZ·ΠR·ZR; [2 + 2]cycloaddition, RΠ·RΠ·RΠ·RΠ. Several features of the symbols may be observed. Reversing the order of symbols causes no change in sense but reversing the two letters in every symbol creates the reverse reaction. Any site with +R or +Π must be adjacent to another with +R or +Π, and the same pairing must also occur with −R and −Π. Thus any reaction involving ±R or ±Π must have two or more sites while reactions with only ±H and ±Z occur at only a single site. Reactions involving two or more sites must always terminate in ±H and/or ±Z unless the involved carbons are cyclic, as in cycloaddition. The central (nonterminal) sites exhibit only ±R, ±Π symbols. The oxidation state change in the reaction is quickly discerned by removing all ±R and ±Π from the symbol list so as to leave only ±H and ±Z and obtain Δx by adding $\Delta x = +1$ for +Z or −H and $\Delta x = -1$ for −Z or +H.

If we coalesce Π and Z into one grouping of *functionality*, F ($f = \pi + z$), it is possible to display all the possible (15) single carbon sites in a planar, triangular graph with σ, f, and h axes (and $\sigma + f + h = 4$), as shown in Chart I.[12] Each point represents a carbon type with its character, $c = 10\sigma + f$, shown below it. If the sites are saturated (no C–C multiple bonds) the oxidation state (x) is shown on the left.[6] The parallels indicate constancy or level of σ, f, or h, *i.e.*, the vertical lines represent groups of constant σ: primary ($\sigma = 1$), secondary ($\sigma = 2$), etc. and the f lines represent functionality level (f = 0, 1, 2, 3, 4). Thus, for $f = 3$, c = 03 is HCN, HCOOR, or $CHCl_3$ while c = 13 is RCN, RCOOR, RCX_3, while ketones and aldehydes and their derivatives are at the next lower functionality level ($f = 2$), being characterized as c = 22 and 12, respectively. Skeletal level families (constant σ) may be exemplified by $\sigma = 2$ (the secondary carbon site): ketone (c = 22), secondary alcohol (c = 21), and methylene (c = 20). The $\sigma = 0$ family (left-most column) groups one-carbon compounds, not bonded (by carbon) to a larger carbon skeleton. Carbon sites with π = 1 or 2, *i.e.*, alkenes or alkynes, can only appear at sites with $f \geqslant 1$[13] and $1 \geqslant \sigma \leqslant 3$ and this structural limitation is indicated for alkene, alkyne, and aromatic carbons by shadings on the triangle in Chart I.

Chart I. Carbon Site Characterizations and Interconversions[a]

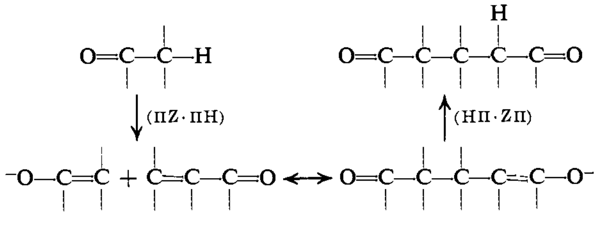


[a] Character, c, shown beneath each structural type. Shaded areas: ▨, possible double bond sites (C=C) (6); ▦, possible triple bond sites (C≡C) (3); ||||, possible aromatic sites (3).

There are 30 links on the triangle representing 30 forward reactions—FH, RF, RH (ten each)—and 30 reverse reactions—HF, FR, HR (ten each)—and ten points of $f \geq 1$ for substitution reactions, FF ($\Delta c = 0$). This is a total of 70 possible interconversion modes for single carbon sites, grouped into seven *types* of ten examples each in Table III; each type is characterized by one characteristic (σ, h, f, c) which does not change in the reaction. These types are each conveniently subdivided into two parallel kinds of *classes* according to the change in either skeletal level (σ classes) or functionality level (f classes). There are four possible classes for each of the (seven) reaction types; the f classes for each are shown in Table III and the alternate

---

acterized as involving five carbon sites if the resonance is invoked; the sum of the intermediate reactions yields the overall reaction statement in the text.



(ZΠ·RΠ·RΠ·ΠΠ·ΠZ)

Sum: ΠZ·ΠH
ZΠ·RΠ·RΠ·ΠΠ·ΠZ
HΠ·ZΠ
RH·RΠ·HΠ

(12) A graph of all four parameters requires a more difficult three-dimensional presentation, such as the present triangle for the σ, h, and z characteristics with the two levels of possible π bonding as an axis perpendicular to the triangle. The π characteristic is also structurally limited to sites with σ = 1, 2, 3 only. An expanded system distinguishing the various functional kinds of Z would also have a three-dimensional presentation, with a perpendicular axis for substitution reactions, ZZ, presumably on a scale noting the particular atom (N, O, X, S, P, etc.) attached to the functionalized carbon and condensed in our notation to Z.

(13) A site with $f = 2$, for example, could be an alkyne, vinyl halide, enol ether, or ketone site at the same functionality level, since $f = \pi + z$.

Table III. Classification of Reaction Types at Single Carbon Sites

Forward reaction types	f classes^a			Reverse reaction types
Substitution ($\Delta c = 0$)	f_{44}	$CO_2 \rightleftarrows CZ_4$		
(FF)	f_{33}	$COOH \rightleftarrows CZ_3$		Same
	f_{22}	$C{=}O \rightleftarrows CZ_2$		
	f_{11}	$COH \rightleftarrows CZ$		
Reduction ($\Delta\sigma = 0$)	f_{43}	$CO_2 \rightleftarrows HCOOH$	f_{34}	Oxidation ($\Delta\sigma = 0$)
(HF)	f_{32}	$COOH \rightleftarrows CHO$	f_{23}	(FH)
$\Delta c = -1$	f_{21}	$C{=}O \rightleftarrows CHZ$	f_{12}	$\Delta c = +1$
($\Delta f = -1$)	f_{10}	$CZ \rightleftarrows CH$	f_{01}	($\Delta f = +1$)
Reductive construction ($\Delta h = 0$)	f_{43}	$CO_2 \rightleftarrows RCOOH$	f_{34}	Oxidative cleavage ($\Delta h = 0$)
(RF)	f_{32}	$COOH \rightleftarrows RC{=}O$	f_{23}	(FR)
$\Delta c = +9$	f_{21}	$C{=}O \rightleftarrows RCZ$	f_{12}	$\Delta c = -9$
($\Delta f = -1$)	f_{10}	$CZ \rightleftarrows CR$	f_{01}	($\Delta f = -1$)
Oxidative construction ($\Delta f = 0$)	f_{33}	$HCN \rightleftarrows R{-}CN$	f_{33}	Reductive cleavage ($\Delta f = 0$)
(RH)	f_{22}	$CHO \rightleftarrows RC{=}O$	f_{22}	(HR)
$\Delta c = +10$	f_{11}	$CHZ \rightleftarrows RCZ$	f_{11}	$\Delta c = -10$
	f_{00}	$CH \rightleftarrows CR$	f_{00}	

Only f classes are listed and the sample generalizations shown exemplify common conversions (usually with oxygen groups) but any heteroatom shown may be replaced by another without change in the f class or type (*e.g.*, $f_{33} \equiv HCN$, HCOOR, etc). Unlabeled bonds are to R or H, but not Z. σ classes for each type: substitution or oxidation–reduction: σ_{33}, σ_{22}, σ_{11}, σ_{00} (σ unchanged); construction: σ_{34}, σ_{23}, σ_{12}, σ_{01}; cleavage: σ_{43}, σ_{32}, σ_{21}, σ_{10} ($\Delta\sigma = \pm 1$).

Table IV. Classification and Enumeration of Reactions at Two Carbon Sites

	Reaction type	Combinations of classes^a	Number of examples^b	Trivial examples	Significant examples
Construction	RH·RH	10	55	−10	45
(or cleavage)	RH·RF	16	100	−16	84
	RF·RF	10	55	−10	45
		36	210	−36	174
Elimination	FH·FH (≡ΠH·ΠH)	6	21	−6	15
(or addition)	FH·FF (≡ΠH·ΠZ)	9	36	−9	27
	FF·FF (≡ΠZ·ΠZ)	6	21	−6	15
		21	78	−21	57

^{a} $n(n + 1)/2$ for like reactions, n^2 for unlike; $n = 4$ for construction, 3 for elimination. ^{b} Same formulas; $n = 10$ for construction, 6 for elimination.

σ classes are listed below the table. The subscript for σ or f defining the class shows the initial and final σ or f value, respectively, for all examples in that class. Thus the σ_{33} class incorporates all reactions in which a tertiary ($\sigma = 3$) carbon site remains tertiary (*cf.* substitutions (FF), eliminations (FF), or reductions (HF) of R_3CZ or oxidations (FH) of R_3CH to R_3CZ). The σ_{23} class, necessarily construction reactions, represents specifically constructions in which a secondary ($\sigma = 2$) site becomes tertiary ($\sigma = 3$).

The number of individual examples in any given class (f or σ) varies from one to four as may be seen on the triangle. An illustration of this which provides an important conclusion for synthesis may be seen in the σ classes of the two types of construction reactions (RH; RF)

		σ_{01}	σ_{12}	σ_{23}	σ_{34}
No. of examples	RH	4	3	2	1
	RF	4	3	2	1
Total		8	6	4	2

Thus the number of ways to construct quaternary sites is less (2) than those creating tertiary sites (4), etc., and implies that an important principle in synthesis design is primary consideration of quaternary (and less, tertiary) site construction since available methods are necessarily more limited.

Focusing on functionality levels, the f_{23} class, for example, would represent reactions of ($f = 2$) → ($f = 3$) such as oxidation of aldehydes (FH, $c = 12 \rightarrow 13$), or fragmentation of ketones (FR, $c = 22 \rightarrow 13$) to carboxylic acids. These classes allow groupings of reactions by the level of change in either function or structure,[14] the important criteria of any carbon site, and they parallel many commonly accepted reaction groupings already in use, *e.g.*, interconversion of carboxylic acid derivatives (f_{33}), reduction of ketones and aldehydes (f_{23}), enolate or carbanion addition to ketones and aldehydes (f_{21} for the electrophilic or carbonyl site), etc. In the common isohypsic construction reactions the electrophilic site undergoes RF change and its four possible functionality levels are the f classes of reductive construction in Table III while the four possible functionality levels of its nucleophilic (carbanionic) partner site (RH) are shown as the f classes of oxidative construction.

When two sites are involved together in any reaction they are either ±R or ±Π, *i.e.*, forming or breaking carbon–carbon σ bond or π bond, respectively. In the first case the two sites are σ bonded in either starting material or product; in the second they are bonded in both. The reaction possibilities for any such pair of sites may now be classified and enumerated as combinations of the two single sites. For construction reactions, there are three possible types: RH·RH, RH·RF, and RF·RF (see Table II); when F = Z, these are simply two-site reactions and when F = Π, they are just the two sites forming a C–C bond in a reaction also involving other sites (*e.g.*, Michael addition, cycloaddition). The number of class com-

---

[14] In principle h classes can analogously be created to show the level of hydrogen attachment on a site, but this does not appear at present to be a useful distinction.

Table V. Numbers of Paths and Termini to Sites on Character Triangle

				Binary paths					
				Triangle[a]					
Group	Sites	No./sites	Single paths	(1)	(2)[b]	Diamond[b]	Linear	Total paths	Total termini
I	(00, 04, 40)	3	2	2	0	2	2	6	5
II	(01, 03, 10, 13, 30, 31)	6	4	2	4	4	2	12	8
III	(02, 20, 22)	3	4	2	4	6	4	16	11
IV	(11, 12, 21)	3	6	0	12	8	2	22	12
		15							

[a] (1) paths to sites reached by one triangle; (2) paths to sites reached by two triangles. [b] No./termini for these paths = half the numbers of paths.

binations is shown in Table IV as well as the number of possible examples (all possible links, for both sites combined, from the character triangle, Chart I). Three combinations of classes are synthetically trivial, namely those which combine two one-carbon units ($\sigma = 0$) into a C_2 skeleton, *i.e.*, RH·RH: $f_{33}\cdot f_{33}$ (see Table III); RH·RF: $f_{33}\cdot f_{43}$; RF·RF: $f_{43}\cdot f_{43}$ (or $\sigma_{01}\cdot\sigma_{01}$ for each type).

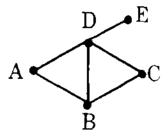
There are four possible classes (f or σ) for each type of construction (or cleavage) reaction and the number of combinations of classes or of actual examples for the two sites is $n(n+1)/2$, or n^2 if the sites differ in type. For elimination (or addition) reactions only three possible classes (f or σ) exist for each reaction type owing to the exclusion of sites with $\sigma = 0$ for olefins.

The value of Table IV arises from the fact that all organic reactions which form (or break) carbon–carbon σ bonds are represented here, either as generic classes (36, or 33 disregarding trivial combinations) or as specific examples[15] (210, or 174 without trivia). Similarly, all carbon pair situations in which π bonds are made or broken are represented and classified. A table of the 36 class combinations has not been included here but one is easily generated by combining all two-site combinations of construction f classes from Table III. This implies that for synthetic purposes we can clearly and systematically delineate all the existing or potential chemical reactions in terms relevant to synthesis and in turn observe what presently nonexistent conversions might be inverted to fill gaps in the available armory of reactions. As examples one might cite the potential oxidative coupling of aldehyde derivatives (net, RCHO + R′CHO → R-CO-CO-R′; $\Delta x = +2$), *i.e.*, $f_{22}\cdot f_{22}$, or reductive coupling of ketones with a CO_2 derivative (net, $CO_2 + R_2CO \rightarrow R_2C(OH)COOH$; $\Delta x = -2$), *i.e.*, $f_{21}\cdot f_{43}$, as potentially valuable construction reactions with little or no precedent in the present literature.

The character triangle can be used to examine all the possible routes to any given type of site in a synthesis dissection. A given carbon can be followed through a synthetic scheme to ascertain whether its overall route contains redundancies, or the total number of routes of n steps to a given site can be enumerated and applied to be sure that all synthetic avenues have been considered in particular cases. Sites on the triangle are divisible into four groups by the graphical identity of their positions: group I (00, 04, 40), first site on any exterior line (side) (corner sites); group II (01, 03, 10, 13, 30, 31), second site on any side; group III (02, 20, 22), third (middle) site on any side; group IV (11, 12, 21), central sites, not on any side. The groups are defined by concentric circles out from the triangle center in the order IV, III, II, I.

The total number of single paths to a site is its index, considered as a point on a graph, *i.e.*, the number of lines incident to that point (Table V). Following graph theory[16] a 15 × 15 adjacency matrix, A, may be written for the 15-point character triangle (Chart I), the rows and columns representing single sites (points) and the matrix elements being 1 or 0 depending on whether the sites are directly linked (adjacent) or not. The sum of any row (or column) is the index or number of single paths to the site corresponding to that row (or column). The elements in the squared matrix (A^2) represent the number of binary paths (length = 2 lines) between any two sites (points) while the sum of any row (or column) in A^2 gives the total number of binary paths leading to that site. In general the number of paths of length n (*i.e.*, a path of n lines to the point or n reactions to reach the site) to a given site is given by the sum of the row (or column) corresponding to that site in the matrix A^n. The total numbers of binary paths obtained in this way is shown in Table V.

These totals may be understood in more detail in the following breakdown. The generalized subgraph **1** implies that binary paths may be of four kinds, depending on the angle between the two lines in the path: 60° = triangle path, *e.g.*, B → D → A; 120° = diamond path, *e.g.*, C → D → A; 180° = linear path, *e.g.*, E → D → A; 360° = trivial path, *e.g.*, A → D → A. It may be observed on the triangle (Chart I) that two sites may be linked by only one linear path (*e.g.*, 00–02), always two diamond paths (00–11), and either one (00–01) or two (01–10) triangle paths. The sums of these kinds of paths are collected in Table V, the trivial paths omitted. The trivial paths are included in and must be subtracted from the row (or column) sums from the A^2 matrix, but their number for any site is equal to its index, *i.e.*, to the number of single paths to that site (these are found as the diagonal elements in A^2), recorded as such in Table V. Trivial binary paths have chemical significance as routes involving blocking, protecting, or activating a given site and then removing the modifying group.



1

---

[15] These examples are "specific" as to character (*i.e.*, σ, h, and f), but each example is still a set of individual cases which vary in the nature of attached R or F (such as olefin, alcohol, halide, etc.).

[16] F. Harary, "Graph Theory," Addison-Wesley, Reading, Mass., 1969.

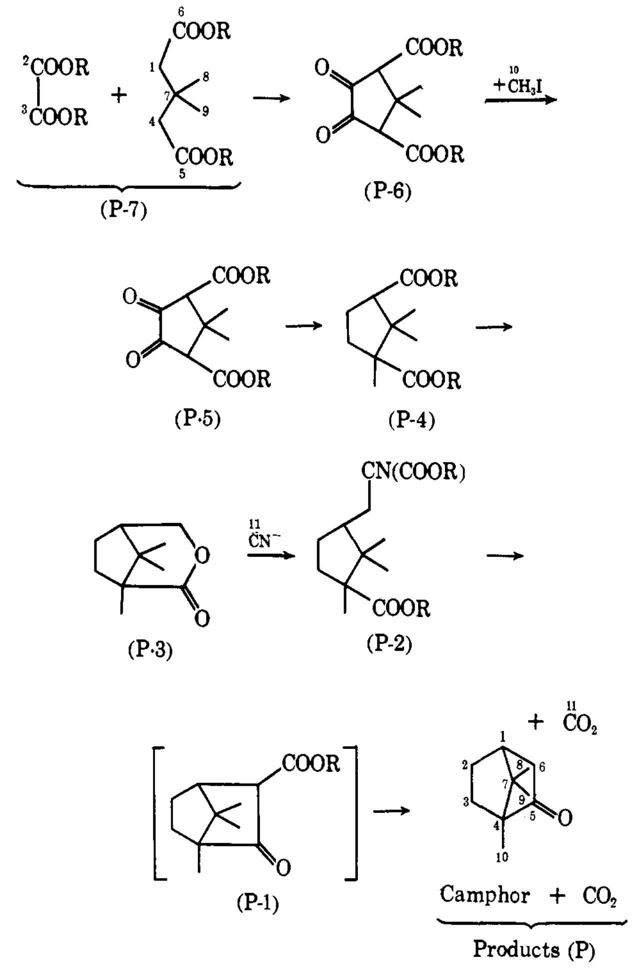


Figure 2. Total synthesis of camphor.

Products (P): $1^{30}(6,7)–2^{20}–3^{20}–4^{40}(7,10)–5^{22}–6^{20}(1)/7^{40}(1,4,9)–8^{10}/9^{10}(7)/10^{10}(5)//11^{04}$

Last reaction: HR(6)·FR(11) (reaction sites in parentheses)

(P-1) $1^{30}(6,7)–2^{20}–3^{20}–4^{40}(7,10)–5^{22}–6^{30}(1,11)/7^{40}(1,4,9)–8^{10}/9^{10}(7)/10^{10}(4)/11^{13}(6)$

Previous reactions: RF(5)·RH(6); RF(6)·RH(11); 2HF(or $(HF)^2$ or H^2F^2)(6); RH(4)·RH(10); 2HF(2), 2HF(3); RH(1)·RF(2), RF(3)·RH(4)

Starting materials: (P-7) $1^{20}(6,7)//2^{13}–3^{13}//4^{20}(7)–5^{13}/6^{13}(1)//7^{40}(1,4,9)–8^{10}/9^{10}(7)//10^{01}//11^{03}$

Following selected sites (forward):

$$1^{20} \xrightarrow[(+10)]{\text{RH}} 1^{30}$$

$$4^{20} \xrightarrow[(+10)]{\text{RH}} 4^{30} \xrightarrow[(+10)]{\text{RH}} 4^{40}$$

$$6^{13} \xrightarrow[2(-1)]{\text{2HF}} 6^{11} \xrightarrow[(+9)]{\text{RF}} 6^{20} \xrightarrow[(+10)]{\text{RH}}$$

$$6^{30} \xrightarrow[(-10)]{\text{HR}} 6^{20}$$

$$11^{03} \xrightarrow[(+10)]{\text{RH}} 11^{13} \xrightarrow[(-9)]{\text{FR}} 11^{04}$$

Figure 3. Linear representation of camphor synthesis.

The termini for paths of n steps out from a given site may be determined without visual reference to the triangle (Chart I) since the lines from that site are characterized by their slopes and the corresponding values of $\Delta c = \pm 1, \pm 9, \pm 10$, as summarized in the inset hexagon. Given a starting site, c_0, these values of Δc are allowed, depending on the h, f, σ characteristics of the site: if $h_0 > 0$: $\Delta c = 1, 10$; $f_0 > 0$: $\Delta c = -1, 9$; $\sigma_0 > 0$: $\Delta c = -9, -10$. The allowed values of Δc may then be separately added to c_0 to give two, four, or six single-path termini, c_1. For each of these the procedure is repeated, adding allowed Δc values for each c_1 to obtain the set of c_2 (binary) termini; those reached by diamond and double triangle paths will appear twice, of course. Thus, from the quaternary site, $c_0 = 40$, only $\Delta c = -9$ and -10 are allowed, yielding $c_1 = 30$ and 31. For $c_1 = 30$, $\Delta c = 1, 10, -9, -10$, yielding $c_2 = 31, 40, 21, 20$, and for $c_1 = 31$, $\Delta c = -1, 9, -9, -10$, yielding $c_2 = 30, 40, 22, 21$. The binary termini collected are therefore $c_2 = 20, 21, 22, 30, 31, 40$ (trivial), with 21 and 40 having been yielded twice; thus there are five nontrivial termini for $c_0 = 40$ (group I), as noted in Table V.

This protocol can be used to determine the number of pathways to any given site as well as the termini at which these pathways originate. Thus, a carboxylic acid is $c_0 = 13$. Single paths to it must arise from $c_1 = 03, 04, 12, 22$. The first two are σ_{01} class reactions, *i.e.*, attachment of single carbons: 03(HCN, CN^-) or 04 (CO_2, $COCl_2$, and derivatives varying F). The other two represent aldehyde (12) oxidation (FH) (or oxidation of other $c = 12$ derivatives such as C≡CH, >C=CHOR, etc.) and cleavage (FR) of $c = 22$, *i.e.*, ketones and their derivatives. The number of single paths to an amine at a secondary site ($c = 21$, group IV) is six and there are therefore 6 × 4 = 24 ways to make an α-amino acid utilizing one step to reach each function. Similar analyses can afford a systematic description of all possible ways to arrive synthetically at any given site of particular functionality (f) and substitution (σ) level.

The total number of possible last steps in any synthesis of camphor (above), for example, is then the sum of possible single paths (one reaction or step) to each of its ten sites: C-1 ($c = 30$), 4 paths; C-5 ($c = 22$), 4 paths; C-7 ($c = 40$), 2 paths, etc., for a total of 40 last steps to synthetic camphor. This affords a clear measure of the scope of possibility in any synthesis.

A synthetic sequence, like a structure, can be represented in a condensed and linear format, as shown for the classical synthesis of camphor (Figure 2)[17] in Figure 3. The connectivity list for camphor[18] is shown at the top of Figure 3 and below it a reaction notation indicating that sites 6 and 11 were cleaved (HR, FR, respectively) in the final reaction producing camphor. The connectivity list for the penultimate molecule (P-1) in the sequence can be generated from this, as shown, by subtracting Δc for HR (-10) from the character ($c_0 = 20$) of site 6; the character for site 6 in the penultimate molecule (P-1) is thus $c_1 = 20 - (-10) = 30$, and similarly for site 11, $c_1 = c_0 - \Delta c = 04 - (-9) = 13$ since for FR, $\Delta c = -9$. Beneath the connectivity list for the penultimate molecules is shown

---

(17) G. Komppa, *Ber.*, **36**, 4332 (1903); *Justus Liebigs Ann. Chem.*, **370**, 209 (1909); the multiste preduction of the cyclopentanedione is shown as one step and interconversions with $\Delta c = 0$, as RCN → RCO-OR in (P-2), are not listed separately. The last step, pyrolysis of the diacid salt, is shown with an intermediate (P-1) equivalent to that in the parallel Dieckmann condensation in order to clarify the concomitant construction and cleavage reactions involved. Starting material carbons are numbered according to their numbers in the product camphor.

(18) The camphor molecule is arbitrarily numbered in Figure 2 as before and a double slash is introduced to separate site 11 as a disconnected molecule, *i.e.*, the separate product, CO_2.

the remaining series of reactions, reading backwards in sequence with each operation separated by a semicolon.[19] For each intermediate step the connectivity list could be generated as it was for compound A, from the reaction description.

The ultimate starting material molecules (P-7) may be generated in the same way by reading all the successive Δc values for involved sites in each reaction. These molecules (P-7) are listed in such a way as to retain the camphor site numbers and it will be seen that this final connectivity list allows the actual starting materials of Figure 2 to be read (dimethylglutaric ester, sites 1, 4, 5, 6, 7, 8, 9; oxalate, sites 2, 3; methyl iodide, site 10; cyanide, site 11); the double slash is again used to signify disconnection of the molecules containing the sites shown at either side of the double slash.

Beneath the synthesis in Figure 3 are shown overall sequences for several individual sites through the synthesis, with the reaction and Δc value shown for each transformation. These sequences can show up redundancies in the scheme; site 6 could pass from 6^{11} to 6^{20} in one step (RF, $\Delta c = 9$) instead of three if a reductive coupling RF·RF between sites 5 and 6 could be engineered, or 6^{13} could pass to 6^{22} *via* RF, $\Delta c = 9$; these observations focus on the possibility of shortening the sequence by using an acyloin reaction as the fourth reaction in the sequence, *i.e.*, RF·RHF^2: $(5^{13}\text{–}6^{13}) \rightarrow (5^{22}\text{–}6^{21})$.

In principle synthetic sequences may be developed by selecting all possible (one-step) routes to all sites (Table V) of the desired product, in all combinations. Such a procedure would generate the complete synthesis tree[8] but the number of sequences obtained would obviously be prohibitive without the introduction of selection criteria.

In conclusion, then, a general system has been developed unambiguously from a simple fundamental view of structure. The system is specifically developed for the particular needs of synthesis design. Both carbon skeleton and functionality are separated and equally treated (*via* $c = 10\sigma + f$ notation). The organization of carbon site characteristics and their interconversions forms a basis for a systematic catalog of all possible as well as available synthetic reactions, the data required for synthesis design. Furthermore, it affords a way to comprehend quickly all possible reaction types which can give rise to any given structural feature. Similarly, available starting materials could also be catalogued in a structurally and functionally meaningful way for synthesis use. The simplicity arising from condensing all heteroatom attachments to carbon as a single feature (F) can be subsequently expanded as required to distinguish the subsets of F as olefinic or attachment to the several particular heteroatoms of interest. Finally, the system proposed is easily described in a simple linear notation which makes it adaptable to computer manipulation without an elaborate and tedious apparatus for coping with actual structural formulas by machine. The characterization developed here forms a basis for an exploration of systematic synthesis design which is currently under way.

Acknowledgment. The author wishes to express his gratitude to Professor E. R. H. Jones for the opportunity of developing these ideas in an uninterrupted atmosphere during a sabbatical stay at the Dyson-Perrins Laboratory, Oxford.

---

(19) In some instances several reactions occurred in one operation and these are merely listed with commas between the semicolons.

Synthesis Design for Substituted Aromatics

James B. Hendrickson

Contribution from Brandeis University, Department of Chemistry, Waltham, Massachusetts 02154. *Received December 28, 1970*

Abstract. A systematic protocol is presented for deriving the various viable synthetic pathways to polysubstituted benzenes from available mono- and disubstituted ones.

The simplest problem in systematic synthesis design should be that of creating polysubstituted benzenes from benzene, or from available mono- and disubstituted benzene derivatives. The problem is relatively simple because the methods are dominated by the relatively clear orienting influences of existing substituents in electrophilic substitution and because no stereochemical ambiguities are involved. The solution to this, as to any, problem of synthesis design is to devise a protocol for systematically developing all possible, or viable, synthetic pathways to a product structure, and, ideally, for ordering them in terms of effectiveness.

For the synthesis of substituted aromatics we shall focus on the making and breaking of bonds between benzene ring carbons and their attached substituent atoms. This removes from consideration syntheses that must include elaboration of links in substituent side chains, except for transformations at the substituent atom itself which change its directive influence on the ring. We may now divide the substituents into nine groups (eight substituent types plus hydrogen) based on the atom attached to the ring and its directive influence. These groups are listed in Table I in approximate order of decreasing activity in facilitating electrophilic substitution and are divided with respect to activation–deactivation of the ring, directive influence, and capability of being introduced in electrophilic substitution.