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.

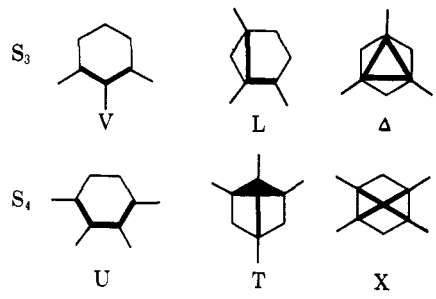


Figure 1. Definitions of aromatic substituent orientations.

Besides clear substituent definition we also require definition of all possible polysubstituent orientations, and a simple symbolism which is easily recognizable visually is offered in Figure 1. There are three possible

Table I. Aromatic Substituent Groups

Activity[a]	Direction[b]	Symbol	Examples
Activating	Ortho	O	O^-, OH, OR, OCOR
	Ortho	N	NH_2, NR_2, NHCOR
	Para	S	SH, SR, SCOR
[c]			[c]
		C	Saturated carbon
[a]		H	H
		X	Cl, Br, I
[b]			
Deactivating	Meta	C_0	CO, CN
		S_0	SO_2
		N_0	NO_2 (NR_3^+)

Divisions by effect on electrophilic substitution: [a] Effect on rate (compared to H); substituents ranged in rough order of activity. [b] Directive influence. [c] Substituents below can be introduced directly by electrophilic substitution.

orientations of disubstituted benzenes (S_2), labeled O (ortho), M (meta), P (para), three each of tri- and tetrasubstituted benzenes (S_3 and S_4), shown in Figure 1, and only one orientation each for S_1, S_5, and S_6.[1] The scope of the synthesis problem may now be seen in the number of combinations and permutations of these variously substituted and oriented benzenes, which are presented in Table II for nine substituents or eight substituents and hydrogen.

The reactions available are conveniently divided into two groups: H ⇄ S, reactions for interchanging substituents and hydrogen; ΔS, reactions for interconversions among substituents.[2] The first group includes substitutions of hydrogen by other substituents (H → S), reactions for which directive influences are important. It will be assumed that electrophilic substitutions give predominantly para substitution to the highest substituent on the activity list of Table I,[3] above the X–C_0 (ortho, para–meta direction) division, and meta substitution if only meta-directing substituents are present. Other reactions than electrophilic substitution, which do not obey these directive influences, are excluded from consideration.[4] Ortho-substitution reactions, however,

(1) The sets of all benzene derivatives with n substituents are designated S_n.

(2) The term substituent will henceforth exclude hydrogen, *i.e.*, eight substituents in Table I.

(3) Ambiguity between groups of comparable activity is treated later.

(4) The few excluded substitution reactions include phenyl carbanion, free-radical, pericyclic, and carbene and insertion reactions, in many cases of less discriminating directive influence.

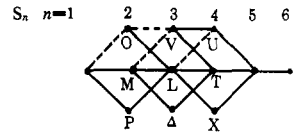


Figure 2. Synthetic pathways graph for aromatic substitution (H⇄S).

include any kind which are ortho specific; these are cyclizations or reactions involving the ortho substituent in cyclic transitions and may include electrophilic substitution or any other substitution which gives predominantly ortho introduction.[5] Reactions replacing substituents by hydrogen (S → H) or other substituents (ΔS) are site specific and involve no problems of orientation.

All possible synthetic pathways for substitution (H ⇄ S) are graphed in Figure 2, with a horizontal coordinate of n = number of substituents and lines indicating reaction in either direction. The H → S reactions ($\Delta n > 0$) are represented as graph lines with direction (vector) to the right and S → H reactions ($\Delta n < 0$) to the left. All S → H reactions are viable, but dotted lines indicate H → S (substitution) reactions which must be ortho specific. There are 19 lines and 12 points on the pathways graph,[6] all lines being characterized by being only $\Delta n = \pm 1$; there are no $\Delta n = 0$ (vertical) lines and four ortho-specific lines (dotted). However, there are also six $\Delta n = \pm 1$ lines missing (OΔ, PV, PΔ, VX, ΔU, ΔX), which are physically impossible transformations. The graph incorporates Körner's classical proof of the orientation of an equivalently disubstituted benzene as well as a parallel proof for trisubstituted benzenes.[7]

The main feature of the graph is its value in showing all possible paths to a given product, and these paths include both $\pm\Delta n$ (forward and back reactions). In general we may presume to start any synthesis from one of the 108 disubstituted benzenes (S_2; Table II), either commercially available or in the literature. The graph requires that the number of steps (lines on the graph) in the synthetic pathway, *i.e.*, the synthetic path length, will be $p = \Delta n + 2i$, where Δn is the difference in number of substituents between product and starting material (*i.e.*, $\Delta n = n_{prod} - 2$) and i = any positive integer. For reasons of economy the only synthetic sequences of interest are direct paths ($p = \Delta n$) and indirect paths ($p = \Delta n + 2$). The kinds of indirect paths may be characterized by the numbers of substituents (n) on each derivative en route.[8] Any path length on the graph (Figure 2) is only the sequence of H ⇄ S steps since substituent interconversions, ΔS, are re-

(5) The present analysis makes clear the need for ortho-specific aromatic substitution reactions. Ordinary electrophilc substitutions are usually not viable for ortho introduction (low yields, separation of para). Cyclizations can be used with subsequent ring cleavage but involve extra steps. Reactions like Claisen rearrangement and ortho-proton removal with alkyllithium are examples of viable ortho-specific substitutions.

(6) If the symmetrical but trivial point, $n = 0$ (benzene itself), is included at the left the graph has 13 points and 20 lines.

(7) W. Körner, *Gazz. Chim. Ital.*, **4**, 305 (1874); *Jahresber.*, 299 (1875). The distinction among S_2 and S_3 isomers (with identical substituents) is found in the index of the point on the graph (*i.e.*, the number of lines incident to it), specifically to the index for forward ($\Delta n = +1$) reactions; indexes: O = 2, M = 3, P = 1; V = 2, L = 3, Δ = 1. Körner's proof identifies the orientation of a di- (or tri-) substituted benzene by the number of isomers it forms in a substitution reaction.

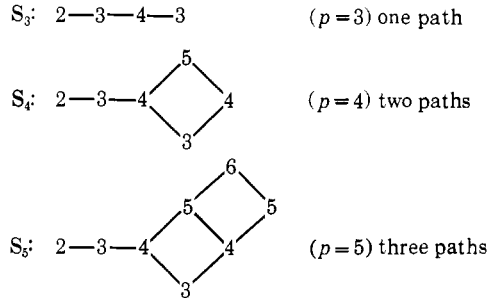
(8) No route returning to S_2 is included since all S_2 derivatives are taken as given starting materials.

Table II. Numbers of Combinations of Polysubstituted Benzenes

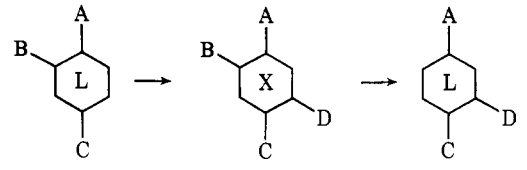
Substituents	A = H n^a	Form	Number of arrangements	Combinations/arrangement 9 subst	Combinations/arrangement 8 subst (A = H)	Total structures 9 subst	Total structures 8 subst (A = H)
A_6	0		1	9	1	9	1
A_5B	1		1	72	8	72	8
A_4B_2	2	O	1			72	8
		M	1	72	8	72	8
		P	1			72	8
		Total	3			216	24
A_4BC	2	O	1			252	28
		M	1	252	28	252	28
		P	1			252	28
		Total	3			756	84
A_3B_3	3	V	1			36	8
		L	1	36	8	36	8
		Δ	1			36	8
		Total	3			108	24
A_3B_2C	3	V	2			1,008	112
		L	3	504	56	1,512	168
		Δ	1			104	56
		Total	6			3,024	336
A_3BCD	3	V	3			1,512	168
		L	6	504	56	3,024	336
		Δ	1			504	56
		Total	10			5,040	560
$A_2B_2C_2$	4	U	4			336	112
		T	4	84	24	336	112
		X	3			252	84
		Total	11			924	308
A_2B_2CD	4	U	6			4,536	1,008
		T	7	756	168	5,292	1,176
		X	3			2,268	504
		Total	16			12,096	2,688
A_2BCDE	4	U	12			7,560	840
		T	12	630	70	7,560	840
		X	6			3,780	420
		Total	30			18,900	2,100
ABCDEF	5		60	84	56	5,040	3,360

[a] Number of substituents (other than H).

actions which remain at one point on the pathways graph.



The numbers of direct and indirect paths available to the several S_n forms from the pathways graph (Figure 2) are listed in Table III, sorted to show the number of ortho-specific steps required in any such path (since such paths are usually less desirable). The numbers of indirect paths[8] include the several path types diagrammed above and also include paths which backtrack when such backtracking is not redundant, as in the example shown below (L → X → L′). The only redundant backtracking is found in (Δ → T → Δ), (V → T → V), and (X → 5 → X),[9] for which return to the original form must give the same compound.



The paths described are set only in terms of the forms (V, L, T, etc.) of the intermediate molecules. When we consider a particular product goal (S_n) with all substituents different and attempt to enumerate the total numbers of possible routes to S_n from all particular S_2 starting materials, we must evaluate the multiplicity of starting materials and of their route choices. Thus *routes* will imply pathways times multiplicity; a pathway in Figure 2 may incorporate several examples since a given S_n product will usually contain more than one O-, M-, or P-S_2 starting material. A pentasubstituted benzene (S_5), labeled ABCDE, will contain the ortho starting materials AB, BC, CD, DE, the meta AC, BD, CE, EA, and the para AD and BE. A complete graph of all the direct routes to S_5 is shown in Figure 3 and includes 60 routes from 10 S_2 starting materials; the 60 routes incorporate the 14 kinds of pathways (on Figure 2) which are listed in Table III. Figure 3 thus represents a complete synthesis tree[10] of direct routes to S_5. The considerable overlapping of intermediates on synthesis trees like this is a common and noteworthy feature.

Direct routes are characterized (as in Figure 3) by a set of S_2 starting materials passing successively to

(9) Redundant backtracking is also implicit in the inverse of two: (T → V → T) and (5 → X → 5).

(10) E. J. Corey and W. T. Wipke, *Science*, **166**, 178 (1969).

Table III. Numbers of Pathways: $S_2 \rightarrow S_n$

		Direct paths: Path length ($p = \Delta n$)	Ortho-specific steps included: 0	1	Total	Indirect paths: Path length ($p = \Delta n + 2$)	Ortho-specific steps included: 0	1	2	Total	
S_3	V	1	0	2	2	3	4	5	0	9	
	L	1	3	0	3	3	7	7	0	14	
	Δ	1	1	0	1	3	3	2	0	5	
S_4	U	2	0	5	5	4[a]	4	12	7	23	37
						4[b]	3	11	0	14	
	T	2	4	2	6	4[a]	7	12	0	19	33
						4[b]	3	11	0	14	
	X	2	3	0	3	4[a]	7	7	0	14	25
						4[b]	0	11	0	11	
S_5		3	3	11	14	5[c]	11	26	19	56	
						5[d]	6	25	11	42	112
						5[e]	3	11	0	14	
S_6		4	3	11	14						

Indirect paths enumerated: [a] 2—3—4—3—4. [b] 2—3—4—5—4. [c] 2—3—4—3—4—5. [d] 2—3—4—5—4—5. [e] 2—3—4—5—6—5.

Table IV. Total Possible Routes to Substituted Benzenes (S_n)

Product S_n	$S_2(S_2')$ form[a]	Direct routes ($p = n - 2$): Ortho-specific steps included: 0	1	Total[b]	$\|S_2\|$[c]	Indirect routes ($p = n$): $\|S_2'\|$[d]	Total routes[e]
S_3 V	O	0	2	2	2	2	10
	M	0	1	1	1	4	11
	P	0	0	0	0	3	6
	Total	0	3	3	3	9	27
L	O	1	0	1	1	4	11
	M	1	0	1	1	4	11
	P	1	0	1	1	1	5
	Total	3	0	3	3	9	27
Δ	O	0	0	0	0	6	12
	M	3	0	3	3	0	9
	P	0	0	0	0	3	6
	Total	3	0	3	3	9	27
S_4 U	O	0	6	6	3	2	60
	M	0	4	4	2	4	72
	P	0	2	2	1	2	36
	Total	0	12	12	6	8	168
T	O	2	2	4	2	4	72
	M	5	1	6	3	2	60
	P	2	0	2	1	2	36
	Total	9	3	12	6	8	168
X	O	4	0	4	2	4	72
	M	4	0	4	2	4	72
	P	4	0	4	2	0	24
	Total	12	0	12	6	8	168
S_5	O	4	20	24	4	2	288
	M	4	20	24	4	2	288
	P	4	8	12	2	1	144
	Total	12	48	60	10	5	720
S_6	O	24	120	144	6		
	M	24	120	144	6		*f*
	P	24	48	72	3		
	Total	72	288	360	15		

[a] Form of the S_2 or S_2' (indirect only) starting material. [b] Total of direct routes only. [c] Number of S_2 starting materials (used direct or indirect). [d] Number of S_2' starting materials. [e] Total indirect routes from both S_2 and S_2'. [f] No indirect routes to S_6.

sets of S_3, S_4, etc., to the final product, S_n; all members of all sets en route contain only those substituted positions which are substituted in the final product, S_n. These substituted positions may be termed *inclusive* positions; in the pentasubstituted (S_5) benzene ABCDE above, A, B, C, D, and E are inclusive positions when they appear substituted in precursors (like AB, CE, ADE, etc.), but F is an *exclusive* position and is never substituted in direct routes.[11] With indirect routes, however, there must be at least one (actually two in practice) set with its members containing one (and only one) exclusive substituent. Indirect routes imply substitution of one position which is not substituted in the final

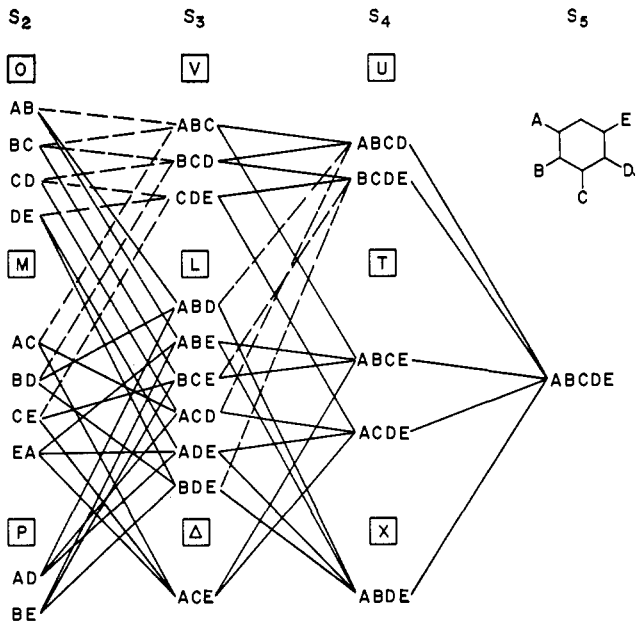
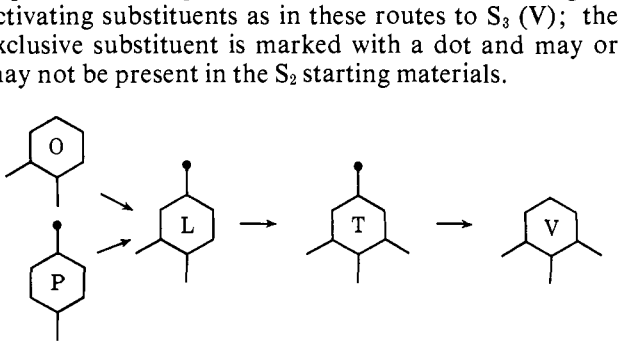


Figure 3. Graph of direct routes to S_5.

product and is removed (S → H) at some stage in the sequence. Examples include the use of blocking or activating substituents as in these routes to S_3 (V); the exclusive substituent is marked with a dot and may or may not be present in the S_2 starting materials.



(11) The inclusive substituted positions constitute a set of *n* positions; the exclusive positions are the remaining set of (6 − *n*).

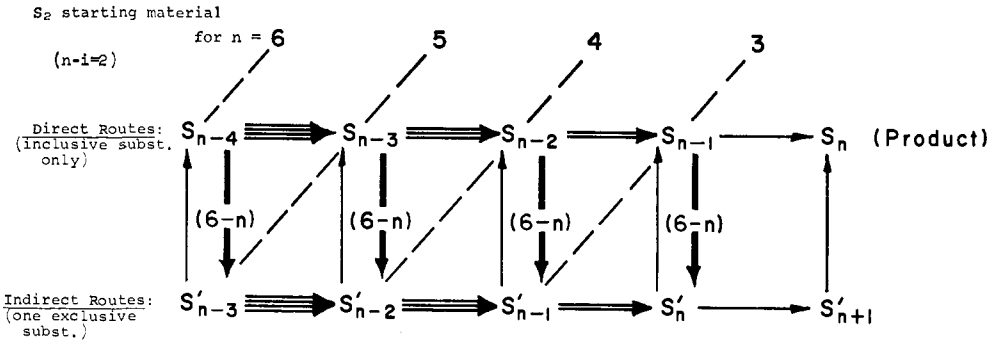


Figure 4. Route enumeration graph for aromatic substitution (H⇄S).

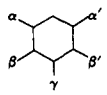
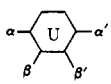
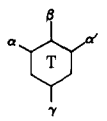
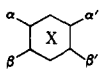
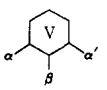
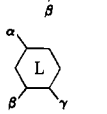
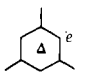
Number of members in set:

$$|S_{n-i}| = \binom{n}{n-i} \text{ combinations} \qquad |S'_{n-i}| = \binom{n}{n-i-1}(6-n) \text{ combinations}$$

	Number of routes per path	Number of paths
$S_2 \rightarrow S_n$ direct ($p = n - 2$)	$n!/2$	1
$S_2 \rightarrow S_n$ indirect ($p = n$)	$(6 - n)n!/2$	$(n - 1)(n - 2)/2$
$S'_2 \rightarrow S_n$ indirect ($p = n$)	$n(6 - n)(n - 1)!$	$n - 2$

Enumeration of the possible routes for all forms is now conveniently derived from the graph in Figure 4 showing the sets, S_{n-i}, of intermediates which constitute the direct route to the product S_n on the upper row and the sets, S_{n-i}', comprising indirect routes on the lower row. All members of sets S_{n-i} (upper row) contain only inclusive substituted positions; all members of sets S_{n-i}' (lower row) contain one exclusive substituted position. The subscript $(n - i)$ indicates the total number of substituted positions on each member of set S_{n-i}.[12] The route links between sets are shown

Table V. Orientation Table for Substitution (H → S)

	Product (S_n)	Product positions D[a]	Product positions S[b]	Starting material[c] Form	Starting material[c] D[a]	Priority[d] Ordinary	Priority[d] Ortho specific
S_5		α	β	T	γ	3	3
		α	β′	T	α	2	
		β	α	U	α	1	1
		β	α′	U	β	1	
		β	γ	X		1	1
		γ	β	T	α	4	1
S_4		α	α′	V	α	2	
		α	β	L	α	4	4
		β	α	V	α	4	1
		β	β′	L	β	4	1
		α	β	Δ		4	3
		β	γ	V	β	1	
		β	α	L	γ	1	1
		γ	β	Δ		2	
		α	β	L	α	2	2
		α	β′	L	β	2	
S_3		α	β	M		4	4
		β	α	O		4	1
		α	γ	M		2	
		β	γ	M		4	2
		γ	α	O		2	
		γ	β	P		1	1
				M		1	

[a] Position of major directive substituent in product (S_n) and starting material (S_{n-1}). [b] Position of substituent introduced. [c] The form of the starting material (S_{n-1}) and position of the major directive substituent (D) on it. [d] Priority scale: 1 = unequivocal product; 2 = probable major product; 3 = uncertain (or ~50:50); 4 = probable minor product. [e] All three substituents are electron withdrawing.

(12) Vertical columns—one direct (S) and one indirect (S′) set—have the same number of inclusive substituted positions in each, *i.e.*, equal to the direct set (S_{n-i}) subscript.

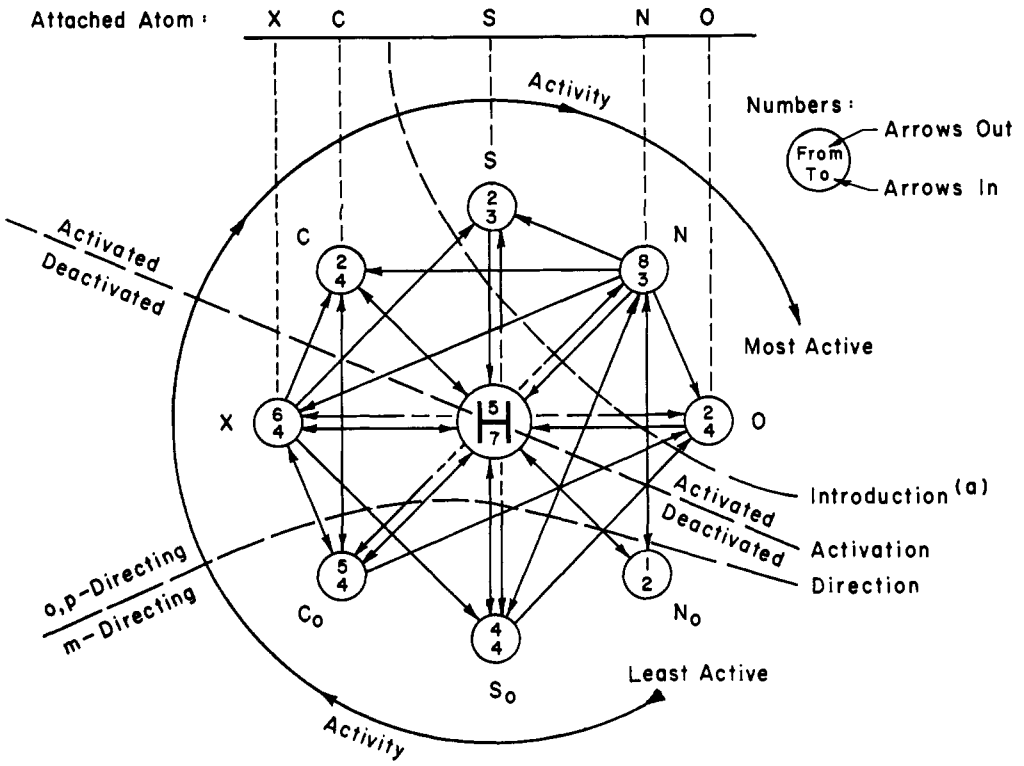


Figure 5. Interconversion graph for substituents. (a) Substituents below line can be introduced directly (H → S) by electrophilic substitution.

as multiple arrows equivalent to the number of route choices any single member of a set has for the conversion, *e.g.*, conversion of any S_{n-3} to S_{n-2} can occur in three ways since any one of three bonds may be formed. Vertical arrows down all have multiplicity of $6 - n$, implying that any of the $6 - n$ exclusive positions are open to substitution in converting a direct-route intermediate (S_{n-i}; all inclusive substituted positions) to an indirect-route one (S_{n-i+1}') with one exclusive substituted position.

The numbers of members in each set are given by the relations at the bottom of Figure 4. The total routes to S_n from S_2 (or S_2') may be computed from the numbers in the S_2 (or S_2') sets of starting materials, the multiplicity of the links en route, and the numbers of paths that can be followed from S_2 (or S_2') to S_n. These relations are also formulated at the bottom of Figure 4 and the numerical results for particular products are assembled in Table IV. There are no indirect route totals for S_6 since there cannot be any exclusive substituted positions in the intermediates; the number of S_2 starting materials is 15, the total number of substituent pairs possible for a fully substituted benzene: $\binom{6}{2} = 15$.

The actual number of *viable* routes for a particularly substituted product is not nearly so great as that in Table IV, however, since the directive effects of existing substituents put limitations on the positions in which substitution occurs. In order to chart these we shall make the simple assumption that a substituent with major directive influence (labeled "D"), *i.e.*, those at the top of Table I, will dominate the choice of position for subsequent substitution, and in the order para > ortho ≫ meta to itself.[13] An orientation table (Table V) may easily be developed from this assumption, showing the position of the major directive substituent (D) in the product (S_n) and then in the starting material (S_{n-1}) as well as the position ("S") of the substituent introduced. For each viable pair (D and S) a selection priority is offered, first for ordinary electrophilic substitution, then for ortho-specific reactions. Questions of steric hindrance may be evaluated in a crude way, by assuming a more favorable priority for substitution in a position adjacent to hydrogen than for one between two existing substituents. If this is done it will be found to affect the priority in less than half of the cases and then always only to enhance a high (2) electronic priority and lower a low one (4); *i.e.*, steric hindrance at that level only enforces the priorities shown. Use of the orientation table to generate viable routes to a given S_n is outlined below.

The second kind of available aromatic reaction is the interconversion of substituents (ΔS) at a single ring position. Such reactions are required in the following situations: (a) best directive position (Table V) does not contain D so that D is placed there for activation and later changed (ΔS) to product substituent; (b) no directive substituent is major or two are of competitive activity; one must be enhanced or depressed *via* ΔS to allow a clear directive influence; (c) desired substituent cannot be introduced by (H → S) so that another (N_0, S_0, C_0, X, C; Table I) is introduced and then transformed by ΔS.

(13) The assumption is simplistic but probably fairly general. It is adopted to show the development of a first-order protocol for route selection. Later modification of this assumption from a systematic study of the literature of aromatic substitution will be desirable but will not change the selection logic. It may be possible, for example, to utilize Hammett $\sigma\rho$ values to calculate overall relative rates for the several open positions to be substituted, and so select the preferred site. The much simpler assumption used in the text is probably usually valid, however, although cases in which the combined directive effects of *two* lesser substituents could overwhelm that of the single dominant substituent must be cause for caution.

Table VI. Adjacency Matrix and Squared Matrix for Interconversion Graph

Adjacency matrix (A): one-step conversions $H \rightleftarrows S$ and ΔS[a]

Reactants	Products: H	O	N	S	C	X	C_0	S_0	N_0	Routes from
H	0	0	0	0	1	1	1	1	1	5
O	1	0	0	0	0	1	0	0	0	2
N	1	1	0	1	1	1	1	1	1	8
S	1	0	0	0	0	0	0	1	0	2
C	1	0	0	0	0	0	1	0	0	2
X	1	1	0	1	1	0	1	1	0	6
C_0	1	1	1	0	1	1	0	0	0	5
S_0	1	1	1	1	0	0	0	0	0	4
N_0	0	0	1	0	0	0	0	0	0	1
Routes to	7	4	3	3	4	4	4	4	2	$\Sigma = 35$

$(A-H)^2$: two-step conversions ΔS only

Reactants	Products: O	N	S	C	X	C_0	S_0	N_0	Routes from
O	-	0	1	1	0	1	1	0	4
N	3	-	2	2	2	2	2	0	13
S	1	1	-	0	0	0	0	0	2
C	1	1	0	-	1	0	0	0	3
X	2	2	1	1	-	1	1	0	8
C_0	2	0	2	2	2	-	2	1	11
S_0	1	0	1	1	2	1	-	1	7
N_0	1	0	1	1	1	1	1	-	6
Routes to	11	4	8	8	8	6	7	2	$\Sigma = 54$

[a] First row $H \rightarrow S$, first column $S \rightarrow H$, others ΔS.

The eight substituent types may be grouped in a circle of roughly increasing activation in a graphical presentation which summarizes the possible interconversions between them. Such a graph for ΔS is presented in Figure 5 with vertical rows (labeled at top) denoting the same atom attached to the benzene ring. The graph is a reorientation of Table I and shows the ΔS reactions as directed links among the eight substituents on the periphery. Hydrogen is placed in the center and links into the center are ($S \rightarrow H$) reactions while those out from it are substitutions ($H \rightarrow S$). (Some ΔS reaction links (arrows) which must cross the center are dotted to indicate their bypassing the center hydrogen.)

Each substituent is represented by a circle showing the numbers of its interconversions with other substituents (including hydrogen), *i.e.*, the number of arrow links in and out from it. The actual number of such links shown is 35, a rather conservative estimate of the usually traditional, well-tested transformations. Since the total possible number of links is $8 \times 9 = 72$, more than half of the potential interconversions remain to be invented.

The interconversions are of course not all viable in all cases since the presence of other groups on the ring may preclude use of some ΔS procedure. This is obvious in most of the six conversions noted from halogen (X), which generally proceed through Grignard and other organometallic intermediates. The five reactions from H are $H \rightarrow S$ and are those below the substitution line (note a, Figure 5), substituents sufficiently deactivating not to cause polysubstitution;[14] Friedel-Crafts alkylation ($H \rightarrow C$) is included despite its lack of viability in many instances.[15] The reverse ($C \rightarrow H$) has only a few examples, such as acidic loss of *tert*-butyl groups. The most versatile source of ΔS reactions (all eight) is N, largely because of diazonium salt transformations, and this is reflected in the common tradition of use of ($NO_2 \rightarrow NH_2 \rightarrow N_2^+ \rightarrow$ etc.) sequences in aromatic synthesis. The group N_0 (*i.e.*, NO_2), however, has virtually no other potential. Oxygen groups also occupy a notably nonversatile position; it is a common corollary that, as oxygen substituents can virtually be neither created nor changed, they must be present in the starting material.

The numbers of transformations on the interconversion graph may also be presented as an adjacency matrix (9×9) of that graph,[16] illustrated in Table VI. The adjacency matrix is created by filling the rows for each substituent with 1 for each route *from* the substituent and 0 for no link, and the columns with 1 for each route *to* that substituent. The sum of each row is the number of routes from the corresponding substituent, while the sum of any column equals the number of transformations which create the corresponding substituent. The square of the adjacency matrix is also shown in Table VI since this shows the numbers of two-step paths between any pair of substituents.[16] The squared matrix is only (8×8), with the hydrogen removed, since any two-step route ($S \rightarrow H \rightarrow S'$) passes through an intermediate with hydrogen at the site and resubstitution must compete with the other hydrogen sites on the molecule, *i.e.*, site specificity is lost. Hence, while the one-step (9×9) adjacency matrix includes $H \rightleftarrows S$ reactions (first row, $H \rightarrow S$; first column, $S \rightarrow H$; dotted lines) as well as ΔS, the two-step (8×8) squared matrix includes only ΔS reactions.

The tools developed thus far allow protocols to be written for finding all the routes (Table IV) from various S_2 to a given S_n. In effect these protocols will constitute ways of creating a viable selection of routes, constituting a synthesis tree, by working backwards on Figure 4. For the direct routes the given S_n is the starting point. For indirect routes, the $(6 - n)$ S_{n+1}' compounds are first created as starting points by placing in each of the $(6 - n)$ exclusive positions a substituent which may be removed ($S \rightarrow H$, Figure 5) at the end of the sequence; these may be major directive (D) groups for activation, or blocking groups (non-D).

Each of these starting points is now evaluated on the orientation table (Table V) to afford a priority for substituting a group S in the position directed by the major directive group, D, which is present on the given S_n (or S_{n+1}') molecule. These choices lead to S_{n-1} (or S_n') precursors which are then evaluated in the same fashion from Table V to create the S_{n-2} (or S_{n-1}') set, etc. Working back on Figure 4 in this way, we may also move from the direct route at any later stage

(14) N and O may sometimes be introduced ($H \rightarrow S$) by benzidine-type rearrangements but these appear to have seen little synthetic use and are not included; creation of the necessary hydroxylamine or hydrazine starting materials is also troublesome.

(15) ($H \rightarrow C$) also includes viable reactions like chloromethylation with CH_2O–HCl.

(16) F. Harary, "Graph Theory," Addison-Wesley, Reading, Mass., 1969.

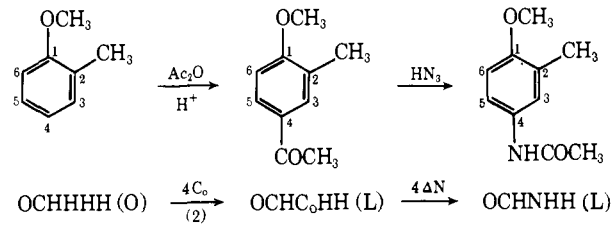
to an indirect route by adding an activating (directive, D) or blocking substituent in any of the $(6 - n)$ exclusive positions to create an S′ set of intermediates, in the same way as was noted for creating the primary indirect route above *via* S_{n+1}'.

At any stage evaluation of the structure on the orientation table (Table V) may lead to choices: (a) two groups may be competitive for D; one must be changed by ΔS to ensure predictable substitution; (b) the group to be introduced may not be suitable for H → S (Figure 5) and another must be introduced and converted to it (ΔS). The interconversion graph will show all the choices to consider for H → S and ΔS as subgraphs incorporating only the arrows to (or from) the substituent in question; these subgraphs are also the separate columns (or rows) of the matrices in Table VI.

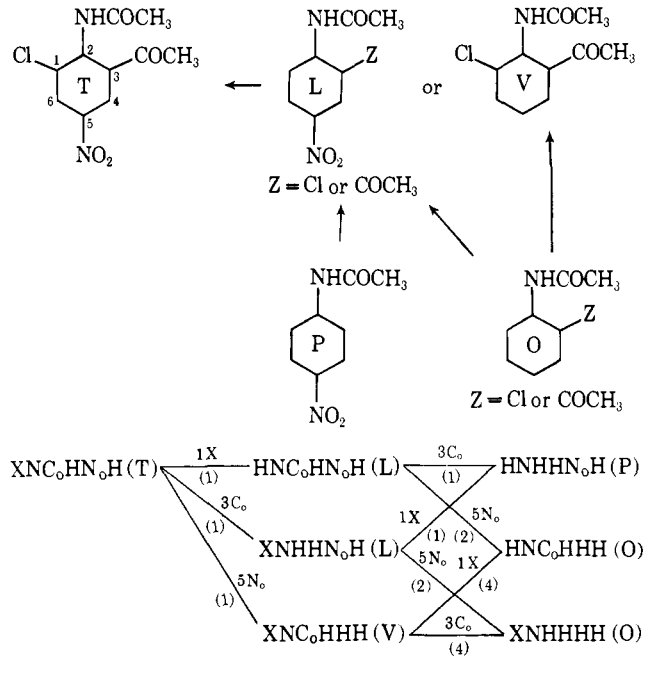
This protocol should result in creating all the viable routes from Table IV in a survey showing the length of the sequence and its overall priority. The sum of the priorities for each (H → S) step divided by the number of these steps ($p = n$ or $n - 2$) will yield an overall priority for judging the value of each sequence, but the number and viability of ΔS steps required will modify this priority. In principle an expanded protocol is possible in which all the given substituents of S_n are considered to be replaced by all their ΔS counterparts and all the molecules so generated evaluated in the same way (there will be $\leq 7n$ such molecules depending on the limits of Figure 5), but such a procedure is wasteful.

It should be noted that any D group selected for directive influence in S_n continues to be present and directive back through the sequence to S_2. This may result only in routes of unacceptable priority, however. In such cases another strategy may be applied which may be longer by virtue of more ΔS reactions but yield preferable overall priorities. The principle of this secondary protocol is to convert the dominant D group into one of lower activity by ΔS and so leave the group next lower in activity as the dominant D group by default. This may be done at any stage and offer opportunity for a better priority in the next lower substitution step by virtue of the different direction of orientation of the new D group. In its fullest form this secondary strategy implies use of ΔS at the start to place a D group in that position in S_n which yields the best priority substitution step no matter where the "natural" D group is located in the largest S_n. As an example, if S_5 with its D group at the γ position demands normal ortho–para substitution, it is rated as priority 4 (Table V). If, however, use of ΔS reactions can reduce it to a lower activity and allow a β substituent to become the dominant influence, one can start the protocol with top priorities. If the same problem occurs farther down the sequence, of course similar changes can be examined there.

A simple notation may be used to denote the molecules involved. A sequence of six symbols from Table I expresses the clockwise sequence of substituents on the ring, and the ring form may be added in parentheses if desired, as in these examples. The numbering of substituents is understood to start at 1 with the first symbol in the sequence and proceed clockwise. Reaction arrows can then be labeled with the number of the site reacting and the group being introduced, with perhaps a Δ to distinguish ΔS from H ⇄ S reactions, as in this sequence (priorities are shown under the arrow):



Some examples may serve to illustrate the protocols. The S_4(T) example (XNC_0HN_0H) has D ($NHCOCH_3$) in the β position dictating substitution of either α group or the γ group with priority 1 and yielding two L and one V S_3 precursor as shown; subsequent priorities to the L precursors from S_2 are good but that to V is poor. The derived synthesis tree, limited to direct routes and no ΔS variants, is shown below and has



three S_2 materials and six routes; the other three (meta) S_2 precursors (and six direct routes) enumerated in Table IV cannot be employed without ΔS reactions en route.

If the product is changed from 5-NO_2 to 5-OCH_3 (*i.e.*, XNC_0HOH) there is now ambiguity in the choice of D between $NHCOCH_3$ and OCH_3; both are examined separately. Choice of $NHCOCH_3$ for D as before implies a final ΔS to create 5-OCH_3 from some other group, which is either present in the S_2(P) precursor or introduced in S_2(O) → S_3(L) or S_3(V) → S_4(T). In the latter cases the choices of the group introduced may be quickly made from the adjacency matrices (Table VI) by accepting any of the last five reactants which show non-zero in the O column, *i.e.*, groups which can

themselves be introduced and can also be transformed to OCH_3 (one-step ΔS: from X, C_0, S_0; seven two-step ways: from C, X, C_0, S_0, and N_0). Alternatively, OCH_3 in the product may be chosen as D and another group at site 2 converted to $NHCOCH_3$ (Table VI: choices of ΔS to N = three one-step from C_0, S_0, N_0).[17] The examination of substitutions with the ambiguity retained may assign a reasonable priority to introduction of Cl, *i.e.*,

$$HNC_0HOH(L) \xrightarrow{1X} XNC_0HOH(T)$$

is probably acceptable owing to the meta direction of $COCH_3$, but

$$XNHHOH(L) \xrightarrow{3C_0} XNC_0HOH(T)$$

is probably not. Finally the much greater number of indirect routes may be explored by starting with the two ($6 - n = 2$) S_2 starting materials which will have blocking groups[18] in either exclusive position (4 and 6), but qualitative examination shows that these do not solve the ambiguity problem. Other indirect routes are generated by adding groups to exclusive positions farther back in the sequence, as with the generation of routes like P′-L′-T′-L-T, etc.

A more difficult example is provided by 2-methoxy-3-methylaniline, an S_3(V) derivative (NOCHHH) with no high priority routes in Table V if ortho-specific reactions are not available. Here indirect routes must be explored by placing in each of the three exclusive positions (4, 5, and 6) either a blocking or activating group, thus creating three S_4 forms, one T and two U. The ambiguity about the primary directive influence (NH_2 *vs.* OCH_3) is resolved by examining the preferred sites for D in each S_4 form. The U forms will be found to be redundant or nonviable except for ortho-specific reactions; the T form is best used (priority 1) with OCH_3 as D and a group introduced in the NH_2 position and then converted to NH_2. The following indirect route is one resultant solution

$$HOCHHH(O) \xrightarrow[(2)]{5N_0} HOCHN_0H(L) \xrightarrow[(1)]{1C_0} C_0OCHN_0H(T) \xrightarrow{5\Delta N}$$
$$C_0OCHNH(T) \xrightarrow{5\Delta H} C_0OCHHH(V) \xrightarrow{1\Delta N} NOCHHH(V)$$

i.e., nitration and acetylation of *o*-cresol methyl ether, removal of nitro *via* diazonium reduction, and Schmidt conversion of acetyl to amino.

In conclusion it may be noted that use of these protocols still involves the generation of many routes of varying validity, that it is time consuming to generate them and still necessary to evaluate their relative merits in order to select the best routes. Nonetheless, it is much less time consuming to do this evaluation first than to embark on an unnecessarily long laboratory operation when the system may reveal a simpler one. The protocols, however, do lend themselves to computerized mechanization which can shorten the time considerably and this development is presently under investigation. The use of linear notation here makes the bulky apparatus for computer handling of full structures unnecessary so that an ordinary computer should suffice to generate routes with their respective priorities.

Finally the use of these protocols would be much simplified if distributors of aromatic starting materials would organize their lists of available compounds by S_n and broken down into the substituent types of Table I roughly after the fashion of Chemical Abstracts Formula Indexes.

Acknowledgment. The author wishes gratefully to acknowledge the hospitality shown him by Professor E. R. H. Jones and members of the Dyson-Perrins Laboratory, Oxford, during a sabbatical stay which made these ideas possible in an uninterrupted atmosphere.

(17) In counting sequence steps it will be apparent that there is a kind of reaction which must be added in counting but is not mentioned in the protocol, *i.e.*, those which interconvert specific substituents within the same type class, as $—COCH_3 \rightarrow —COOH$ or $—NH_2 \rightarrow —NHCOCH_3$, etc.

(18) Introduction of activating ("D") groups only confuses the ambiguity already present.