match no. 8 pp. 159 - 192 1980

CHEMICAL GRAPHS. XXXVIII.[1] SYNTHON GRAPHS

Alexandru T. Balaban

Organic Chemistry Department, Polytechnic, Bucharest, Roumania

(received: January 1980)

Abstract

The paper defines a new type of chemical graphs, namely synthon graphs, which are a class of reaction graphs (vertices represent synthons, intermediates, or the target molecule, while edges represent elementary reaction steps). These graphs have, however, in common with constitutional graphs, the mode of linking of various synthons in the target molecule, which thus appears as a pseudoconstitutional graph. A detailed general analysis of synthon graphs involving up to four different synthons is provided, followed by a cursory general analysis of synthon graphs involving more than four different synthons. For linear pseudoconstitutional graphs, induction allows treatment of the general case involving n synthons and similarly for an n-membered cyclic pseudoconstitutional graph, the general case is discussed. A few practical cases are used as examples of applicative value.

1. Graph-Theoretical Preliminaries

Historically, the first recorded type of chemical graphs is represented by constitutional (molecular) graphs, where vertices symbolize atoms, edges symbolize covalent bonds i. e. pairs of shared electrons, and the whole graph is a molecule. As mentioned by Rouvray [2] (cf. also Rouvray and Balaban [3]), the introduction of today's constitutional or structural formulas was pioneered by Boskovič, Higgins, Couper, Butlerov, Crum Brown, Kekulé and others. It was also pointed out [3] that such constitutional graphs were subsequently refined in that edges may nowadays symbolize more (as in the Hückel C_6 graph for benzene) or less than a pair of shared electrons (as in deltahedral

formulas for boranes, carboranes, metal clusters,[4] or other electron-deficient species such as polyhedral organic cations [5]).

The second large group of chemical graphs is represented by reaction graphs, introduced in 1966 by ref.[6], where vertices symbolize molecules or reactive intermediates such as ions or free radicals, edges symbolize elementary reaction steps, and the whole graph represents a reacting chemical system. Soon after the initial paper dealing with 1,2-shifts in carbenium ions was published,[6] Ramirez discovered independently the same reaction graph applying it to intramolecular rearrangements of pentacoordinated systems.[7] Mislow and coworkers unified previous approaches [8] and developed further the subject of reaction graphs for which a rich literature now exists.[9,10]

In the present paper we discuss a third class of chemical graphs, for which we propose the name of synthon graphs, where vertices symbolize synthons (reagent molecules or molecular fragments) as well as intermediate building blocks or the target molecule, edges symbolize assembly or degradation reactions, and the whole graph represents all possible syntheses of the target molecule starting from available synthons via allowed reactions. The term "synthon graph" was first used by Hendrickson.[11]

The concept of synthons was developed from the pioneering work of Corey and Wipke who, first together,[12] then separately, [13,14] devised computer approaches towards converting the art of organic synthetical strategy into a scientific discipline. Today there exists a rich literature on this topic, from which only a few selected references are quoted.[15-30] The forerunner is reported to be Sarret.[31] Spectacular results in the computer-aided design of organic syntheses were obtained by Corey, who called this method "antithetic analysis", and who succeeded in synthesizing prostaglandins [32,33] and gibberellic acid.[34]

Evidently, synthon graphs resemble reaction graphs and may be considered as a particular case of reaction graphs ; however, they also include structural information about how the synthons must be connected to afford the target molecule, which is a pseudoconstitutional graph : just as the constitutional graph expresses the bonding topology of atoms, the target molecule represents the bonding topology of the various possible synthons.

An important difference between constitutional graphs and pseudoconstitutional graphs is that the set of isomers for the former graphs is uniquely defined ; however, for a given target molecule, its decomposition into possible synthons can occur in several distinct ways, each of which corresponds to a pseudoconstitutional graph, and each of which can in turn give rise to "isomers" since synthons may be connected in several topologically distinct ways. The modes of decomposing target molecules into synthons form part of the previously quoted programs [12-34] and will not be discussed further here. In the present paper we shall assume the target molecule to have been already sequenced into synthons which will be denoted by capital letters A, B, C, etc. if different from one another, or by A_2 , A_3 , A_2BC, etc. if several identical synthons must be incorporated into the target molecule. However, it must be borne in mind that a given target molecule may correspond to more than one pseudoconstitutional graph, involving different synthons, just as a molecular formula may correspond to more than one constitutional formula, involving the same atoms.

As defined here, constitutional graphs are non-directed graphs ; reaction graphs are usually non-directed, but occasionally they may be directed graphs (digraphs) ; synthon graphs are usually digraphs, but occasionally they may be non-directed, e. g. when we want to represent not only assembly reactions (syntheses), but also fragmentations such as processes occurring in mass spectrometry, or decomposition of the target molecule into a particular set of synthons.

Vertices in a synthon graph symbolize one of the three following items : (i) the starting materials of the synthesis (synthons), denoted in the following by A, B, C, etc. These vertices are endpoints of synthon trees, which form subgraphs of the synthon graph ; (ii) the fragments formed from two or more synthons, denoted like constitutional formulas by A_2 , AB, ABC, ACB, ABAB, ABBA, A-B$\begin{smallmatrix} A \\ \diagdown B \end{smallmatrix}$, A$\begin{smallmatrix} \diagup B \\ \diagdown C \end{smallmatrix}$, etc., depending on how the synthons are linked. Whenever possible, we shall avoid lines in the symbols of the fragments, but in some cases like the two last examples, lines in the fragment symbols are unavoidable ; (iii) the target molecule (pseudoconstitutional

graph) denoted either by F (final product) or by the pseudo-constitutional assembly of the synthons, like a fragment vertex.

Edges in a synthon graph symbolize elementary assembly reactions, e. g. A ——— AB ——— B indicates that two synthons, A and B, form together by any kind of reaction, the fragment or target vertex AB.

When one and the same fragment or target vertex, say ABC, may be formed by more than one route, the edges of each synthon tree may be differentiated from all other ones by colouring ; in the present paper, coloured edges will be symbolized by drawing them with various kinds of lines (full, dotted, dashed, etc.) as indicated in Figure 1. Thus, vertex ABC can originate by two routes : either from A and BC (full line) or from AB and C (broken line). The full line in Figure 1 thus forms a synthon tree rooted at ABC with endpoints A, B, and C, while the broken line forms another synthon tree with the same endpoints and root. In the case presented in Figure 1, the two synthon trees are spanning trees of the synthon graph, but this is not a general rule.*

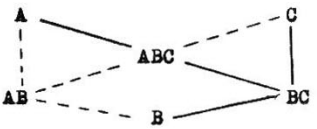


Figure 1. Synthon graph for a target molecule ABC

---

* Graph-theoretical definitions for a few terms encountered in the paragraph : a rooted tree is an acyclic graph (tree) with a vertex (root) distinguished from the other vertices ; in the following, the target molecule F will always be the root. A planted tree is a rooted tree where the root is an endpoint, i. e. a vertex of degree one. A spanning tree (partial tree) of a graph is an acyclic subgraph with all the vertices of the given graph. If in a given graph one deletes edges or vertices, one obtains a subgraph. Edge colouring here does not imply any restriction on colours of adjacent edges.

2. Chemical Preliminaries

We define below the five main types of bond forming or bond breaking chemical reactions :

(i) Colligation is the formation of exactly one bond A-B between two molecular fragments A and B :

$$A + B \longrightarrow A\text{-}B$$

When the two fragments A and B are not linked and different, the reaction is called addition, substitution or condensation, depending on whether smaller molecules are, or are not, split off during the reaction. When the two fragments are not linked and identical, the process is called dimerization or condensation, respectively, according to the previous criterion. When A and B are already linked, the process is no longer intermolecular but becomes intramolecular and is called cyclization.

The colligation is the most important process in synthetical strategy planning, and is the basis of all graphs to be discussed in the present paper. Though we are not concerned with reaction mechanisms here, it may be mentioned that such reactions have mostly ionic mechanisms (with consequences for regio- and stereoselectivity), but sometimes they occur homolytically. As shown by Sinanoglu,[35-37] topological treatment of all possible mechanisms is possible for synthetic pathways.

The reverse reaction (breaking of one bond) is called elimination (fragmentation, scission) with many subdivisions which may be general like decondensation, or particular like decarboxylation or hydrolysis.

(ii) Cycloaddition is the practically simultaneous formation of exactly two bonds between the ends of two molecular chains, leading to a ring system. As a particular case we include here the α-additions to carbenoid-type systems K, and we consider a double bond as a two-membered ring :

$$\begin{matrix} A \\ B \end{matrix} + \begin{matrix} C \\ D \end{matrix} \longrightarrow \begin{matrix} A\text{-}C \\ B\text{-}D \end{matrix}$$

$$K + \begin{matrix} A \\ B \end{matrix} \longrightarrow K\begin{matrix} \diagup A \\ \diagdown B \end{matrix}$$

$$K + K' \longrightarrow K = K'$$

When the reaction occurs in a concerted process, it is regiospecific and stereospecific in agreement with the Woodward-

-Heffmann rules [38-40] (Dewar-Evans postulates [41]). Such reactions are considered as "no-mechanism reactions". Recently, Dewar [42] advocated that no two bonds may be formed simultaneously. For practical purposes, however, there exist criteria differentiating cycloadditions from all other reactions, namely reaction entropy, insensitivity to conditions favouring ionic or radicalic reactions, etc.

The reverse reaction (breaking of two bonds simultaneously) is called retrocycloaddition, or sometimes fragmentation.

(iii) Insertion is the simultaneous replacement of a bond B-C by two new bonds A-B and A-C to a group A, which thus becomes "inserted" into the previous bond B-C :

$$A + B\text{-}C \longrightarrow B\text{-}A\text{-}C$$

(iv) Polymerization and polycondensation is the formation of n-1 bonds between n monomer units A :

$$n\,A \longrightarrow A_n$$

(in polycondensations, n-1 smaller molecules are also formed). The reactions are ionic or radicalic, often with chain mechanism. One has to exclude the cases n = 2 (this belongs to (i), since only one bond is formed) and n = 3, because trimerizations are formally equivalent either to cycloadditions (ii), or to insertions (iii). The reverse reaction is called depolymerization.

(v) Rearrangement or isomerization [43,44] is the intermolecular or intramolecular reorganization of bonds and atoms :

$$A \longrightarrow B$$
$$A + BCD \longrightarrow ACBD \text{ , etc.}$$

For the purpose of the present discussion, we are interested in building up more complicated molecules from simpler aggregates, i. e. synthons. Though all five types of reactions are used by chemists, in the present paper we shall limit ourselves to colligations, including exceptionally cycloadditions in a few cases.

In actual syntheses, in addition to the two types of reactions which will be considered here, eliminations which are the reverse of type (i) reactions are also important : though they do not increase the molecule, in organic syntheses regiospecifity and activation are often achieved by means of substi-

tuents or groups which are eliminated after they performed their part. Examples are amino or sulphonic groups for directing substitutions in aromatic systems, carbalkoxy groups for activating adjacent hydrogens in condensations, etc.

3. Discussion of various cases of synthon graphs

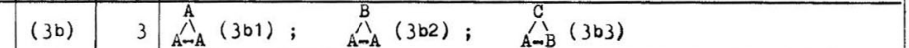
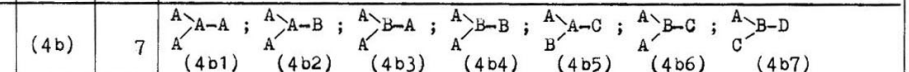
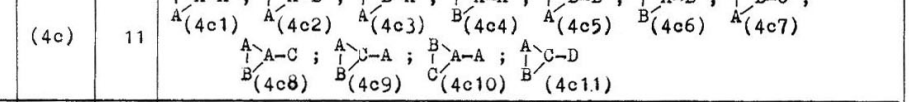
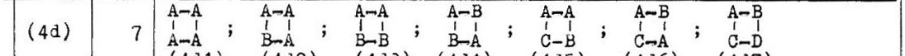
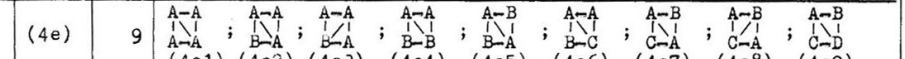
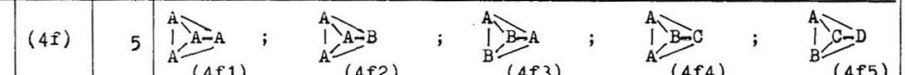
The purpose of the present paper is to discuss complete mappings of assembling molecules from synthons via fragments. Target molecules which differ only in permutations of synthon symbols such as ABBA = BAAB, or AAAB = BBBA will be considered as equivalent as indicated by the equal sign, and their synthon graphs will be described only once ; however, this only applies to targets, but not to fragments.

Only cases with at most four different synthons will be discussed in detail. Structures of target molecules are surveyed in Table 1. In a target which has to be assembled from several synthons, not all of them need be different, however ; this gives rise to several pseudoconstitutional graphs corresponding to the same type of structure (in terms of synthons) of the target molecule. The synthon graphs which will be discussed all bear an indication consisting of three symbols : the number of synthons in the target molecule, a small letter distinguishing among targets with the same number of synthons, and a serial number distinguishing among targets with the same structure and number of synthons, but differing in the number and arrangement of identical synthons.

In drawing the synthon graph, whenever possible, we shall indicate synthon trees rooted at the target F, by different kinds of lines ; this will be always be done for 2 or 3 synthons in the target, and occasionally for 4 synthons. In many cases, it will be observed that the two branches of the rooted synthon tree meet at the root with angles of 180° ; this observation makes it easier to find the rooted synthon trees for 4 synthons in the target, when synthon trees are no longer distinguished by different types of lines.

To simplify the following representations of synthon graphs, they will be drawn as non-directed graphs. If desired, arrows may be added on edges of synthon trees starting from endpoints (synthons) towards the target molecule (vertex F) to convert these graphs into directed graphs (digraphs).

Table 1. Survey of target molecules whose synthon graphs will be selectively discussed

Pseudoconstitutional graph			Target molecule	
No. of vertices	Representation	Code	Total no.	Representation
2		(2a)	2	AA (2a1) ; AB (2a2)
3		(3a)	4	AAA (3a1) ; ABA (3a2) ; AAB (3a3) ; ABC (3a4)
		(3b)	3	 (3b1) ; (3b2) ; (3b3)
4		(4a)	11	AAAA (4a1) ; AAAB (4a2) ; AABA (4a3) ; AABB (4a4) ; ABAB (4a5) ; ABBA (4a6) ; ABBC (4a7) ; ABCA (4a8) ; ABAC (4a9) ; AABC (4a10) ; ABCD (4a11)
		(4b)	7	 (4b1) (4b2) (4b3) (4b4) (4b5) (4b6) (4b7)
		(4c)	11	 (4c1) (4c2) (4c3) (4c4) (4c5) (4c6) (4c7) (4c8) (4c9) (4c10) (4c11)
		(4d)	7	 (4d1) (4d2) (4d3) (4d4) (4d5) (4d6) (4d7)
		(4e)	9	 (4e1) (4e2) (4e3) (4e4) (4e5) (4e6) (4e7) (4e8) (4e9)
		(4f)	5	 (4f1) (4f2) (4f3) (4f4) (4f5)

There is no limitation to the degree of a vertex in a synthon graph. As it will be seen, owing to the various modes of decomposing a given target molecule into synthons, there exist synthon graphs with vertices of quite high degree. Owing to the fact that synthon vertices usually have lower degrees than target vertices, synthon graphs have low, if any, symmetries.

From the cases presented in Table 1, owing to space limitations, we selected only a few synthon graphs for detailed presentation. It should be noted that in the following treatment synthon graphs corresponding to cyclic target molecules may, or may not, be allowed to involve simultaneous formation of more than one bond at a time, i. e. cycloadditions.

3.1. Target molecules with two synthons

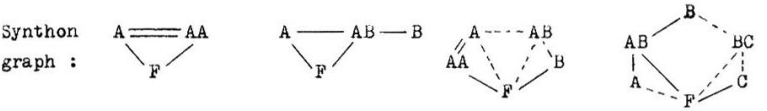
Figure 2 presents the two synthon graphs of the simplest target molecules A-A and A-B involving two synthons which may, or may not, be identical (cases 2a1 and 2a2 from Table 1).

(2a1) (2a2)

Figure 2. Synthon graphs for target molecules F = A-A (2a1), and A-B (2a2).

3.2. Target molecules with three synthons

3.2.1. Target molecules coded (3a). Figure 3 presents the four synthon graphs corresponding to the target molecules indicated in Table 1. We shall discuss in more detail only the last synthon graph corresponding to a target molecule ABC, case (3a4) in Table 1, which was also included in Figure 1.



Target F : A-A-A A-B-A A-A-B A-B-C

(3a1) (3a2) (3a3) (3a4)

Figure 3. Synthon graphs of target molecules involving three synthons in an acyclic arrangement.

The synthon graph corresponding to case (3a4) involves two possible intermediates, AB and BC, depending upon which lar-

ger fragment is formed first. The target molecule F = ABC is the root of two trees with endpoints representing the synthons A, B, and C : one rooted tree is drawn with full lines, the other with broken lines, as in Figure 1. For instance, if the target molecule is para-bromochlorobenzene to be obtained from A = bromine, B = benzene and C = chlorine, the intermediates are AB = bromobenzene and BC = chlorobenzene.

3.2.2. Target molecules coded (3b). Since the pseudoconstitutional graph for decomposing the target molecule into synthons is cyclic, we shall have to discuss separately reactions involving, or not involving, cycloadditions. Figure 4 presents the nine synthon graphs corresponding to the three (3b)-type target molecules and to the three possibilities relative to cycloadditions : involving only, involving also, or not involving, cycloadditions.

The most general case (3b3) is also the one to lead to the most symmetrical synthon graph. In the case where the last step can only be a cycloaddition, the target molecule is the root of three trees drawn with different kinds of lines. It is easy to see that the case where cycloadditions may also be involved leads to a synthon graph which is a supergraph of the synthon graph for the case involving only cycloadditions, and which is a spanning supergraph of the synthon graph for the case not involving cycloadditions. According to graph-theoretical definitions, a supergraph of a given graph has all points and lines of the given graph (in addition to other points and lines), while a spanning supergraph of a given graph has only the set of points of the given graph (and more lines than the given graph). Conversely, the given graph is a subgraph of the larger graph (or a spanning subgraph, respectively).

It can also be observed that the synthon graph for the case (3b3), involving only cycloadditions as the last step, contains as a subgraph the synthon graph of the acyclic case (3a4) presented in Fig. 3.

In order not to complicate unduly the drawing for Fig. 4, the synthon graph for (3b3) without cycloadditions, where six intermediates may exist, has six possible trees rooted and planted at F, only one of which is shown in thick lines. No rooted trees are delineated for the more complicated case (3b3) with cycloadditions possible, but not mandatory.

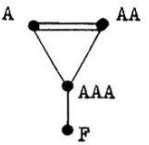
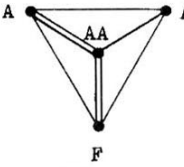
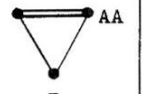
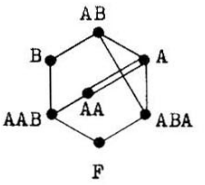
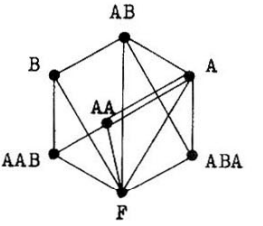
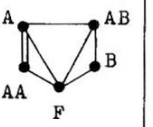
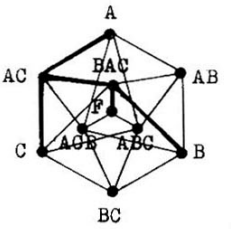
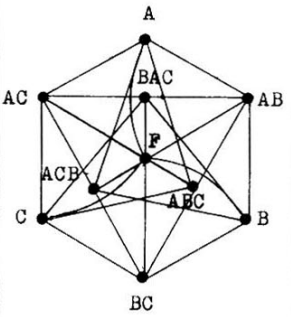
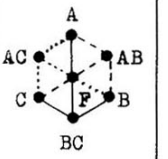
Target & code	Is cycloaddition involved in the last step ?		
	No	Yes, it may	Yes, it must
A, A—A (3b1)			
B, A—A (3b2)			
B, A—C (3b3)			

Figure 4. Synthon graphs of cyclic target molecules consisting of three synthons (3b) arranged circularly.

3.3. Target molecules with four synthons

Owing to the increasing complexity of synthon graphs when the number, the number of types, and the dissymmetry in the arrangement, of synthons in the target molecule increases, it will no longer be possible to present exhaustively a complete analysis of synthon graphs involving four synthons, as it was done for target molecules composed of two or three synthons.

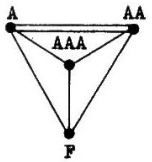
3.3.1. Target molecules coded (4a). When a target molecule can be decomposed into, and reconstructed from, four synthons arranged linearly, these four synthons may be of one, two, three, or four different types. According to their arrangement, the eleven possible cases presented in Fig. 5 may occur, with their codes from Table 1. It may be seen that the synthon graphs have between four and ten vertices, representing the synthons, the intermediates (with two or three synthons), and the final vertex F (target molecule). For the most general case, of the molecule involving four different synthons ABCD (coded 4a11), there are five possible intermediates, and five possible trees rooted at F. One such tree has branches AB + CD, and two pairs have branches A + BCD and D + ABC, respectively ; each pair decomposes differently the three-synthon intermediate, e.g. B + CD and D + BC.

3.3.2. Target molecules coded (4b). Fig. 6 presents the synthon graphs for all possible cases (4b1) - (4b7). The most general case, (4b7) with four different synthons, has three possible trees rooted at F (different lines in Fig. 6).

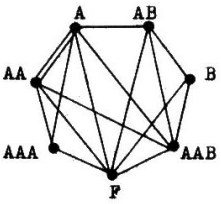
3.3.3. Target molecules coded (4c). Only the five simplest cases from the eleven possibilities are shown in Figure 7, and only when cycloadditions are not allowed.

3.3.4. Target molecules coded (4d). Target molecules consisting of four synthons arranged circularly (e. g. corrins) can give rise to seven cases, presented in Fig. 8 if cycloadditions are not allowed; if cycloadditions are allowed, but are not mandatory, Figures 9 and 10 present the synthon graphs for selected cases. Comparing Fig. 10 with the last part (continuation) of Fig. 8, it becomes evident that cycloadditions increase the complexity of the synthon graph for the same target molecule coded (4d7).

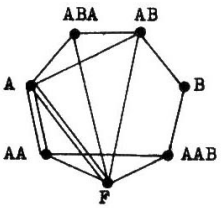
3.3.5. Target molecules coded (4e) and (4f). The synthon graphs are too complicated, and the target molecules too improbable to warrant discussion in the present paper.



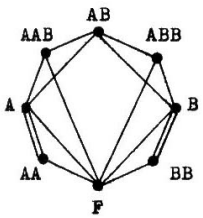
AAAA (4a1)



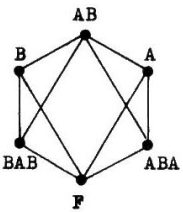
AAAB (4a2)



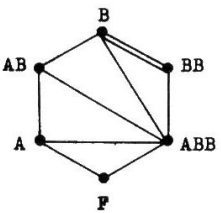
AABA (4a3)



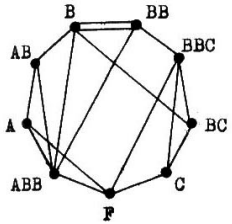
AABB (4a4)



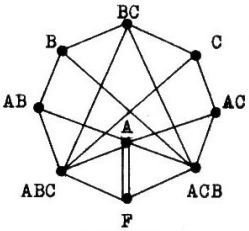
ABAB (4a5)



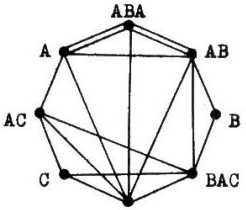
ABBA (4a6)



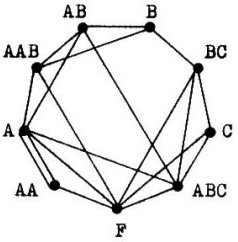
ABBC (4a7)



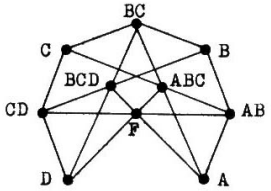
ABCA (4a8)



ABAC (4a9)



AABC (4a10)



ABCD (4a11)

Figure 5. Synthon graphs of target molecules F composed of four synthons arranged linearly.

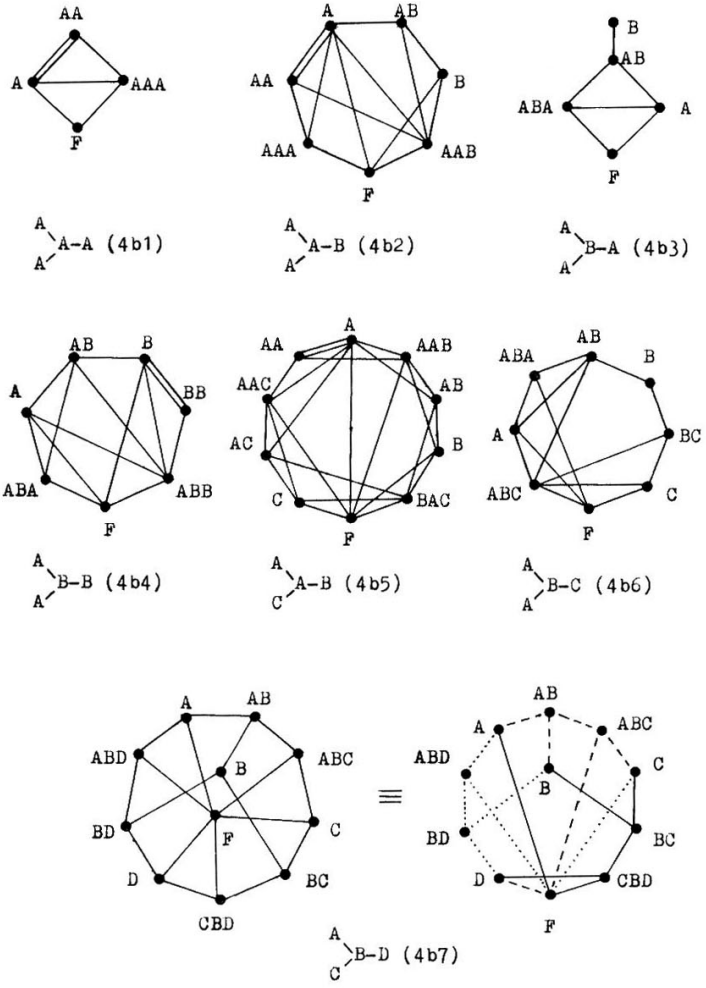
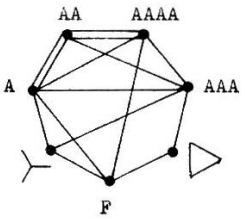
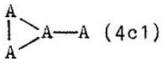


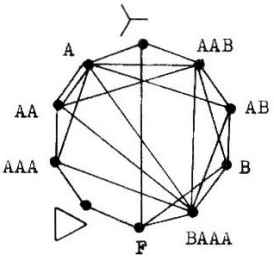
Figure 6. Synthon graphs of target molecules F = 

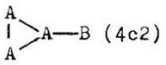
Under each synthon graph, the corresponding target molecule (pseudoconstitutional graph) is indicated, together with the code from Table 1.



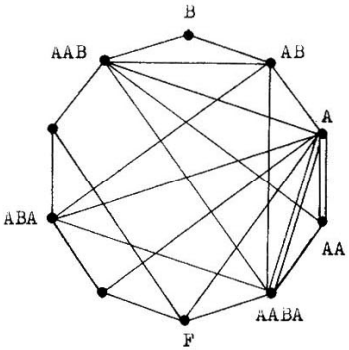


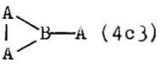
(4c1)



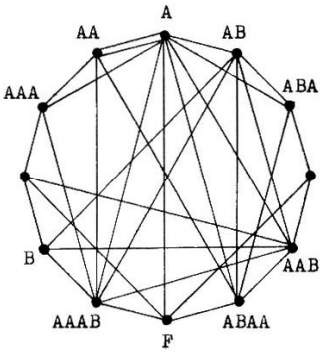


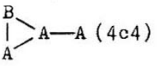
(4c2)





(4c3)

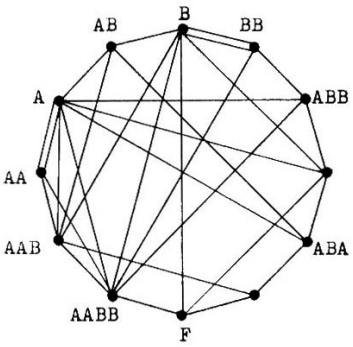




(4c4)

Figure 7. Synthon graphs for some of the target molecules coded (4c), when cycloadditions are not allowed.

(4c5)



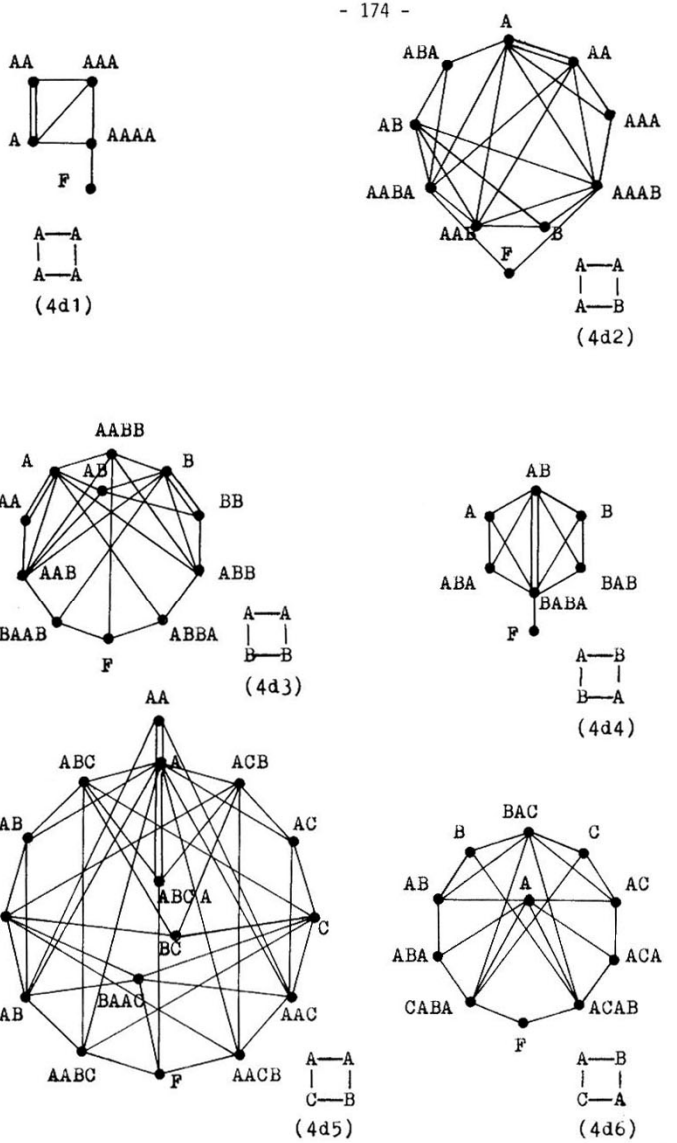
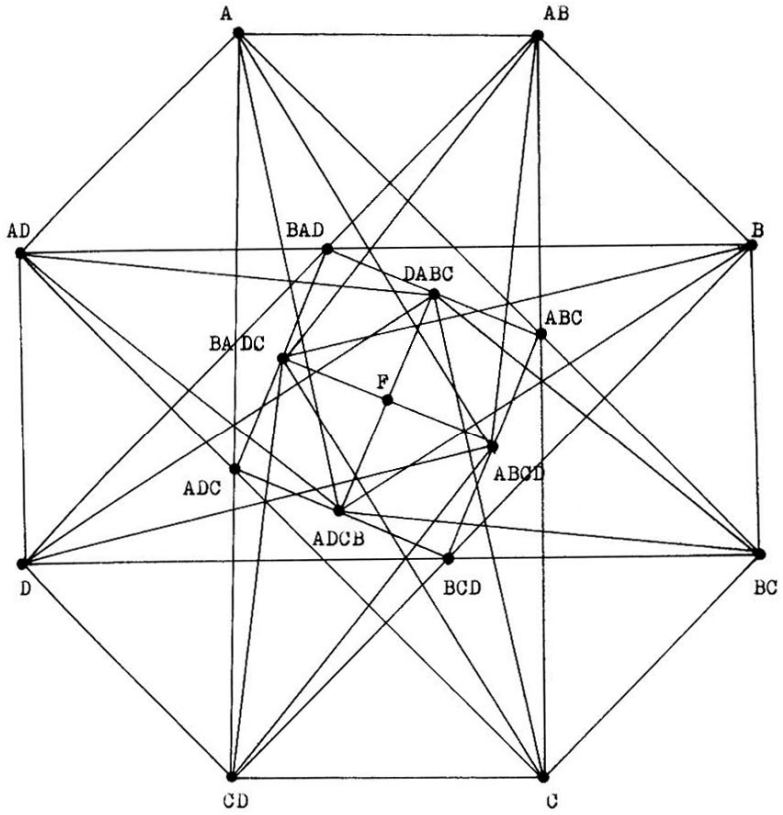


Figure 8. Synthon graphs of target molecules coded (4d) with no cycloadditions allowed (continued on next page).



(continued). Synthon graph for target molecule F =  with no cycloadditions allowed, code (4d7).

Continuation of Figure 8, presenting the last case

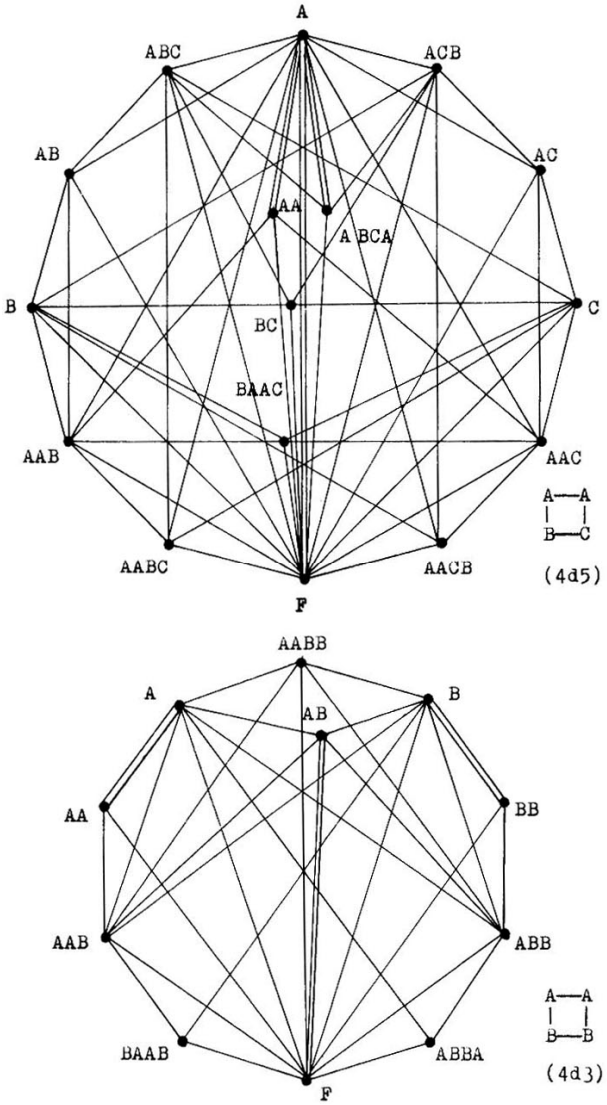


Figure 9. Synthon graphs for target molecules indicated with their codes, when cycloadditions may take place.

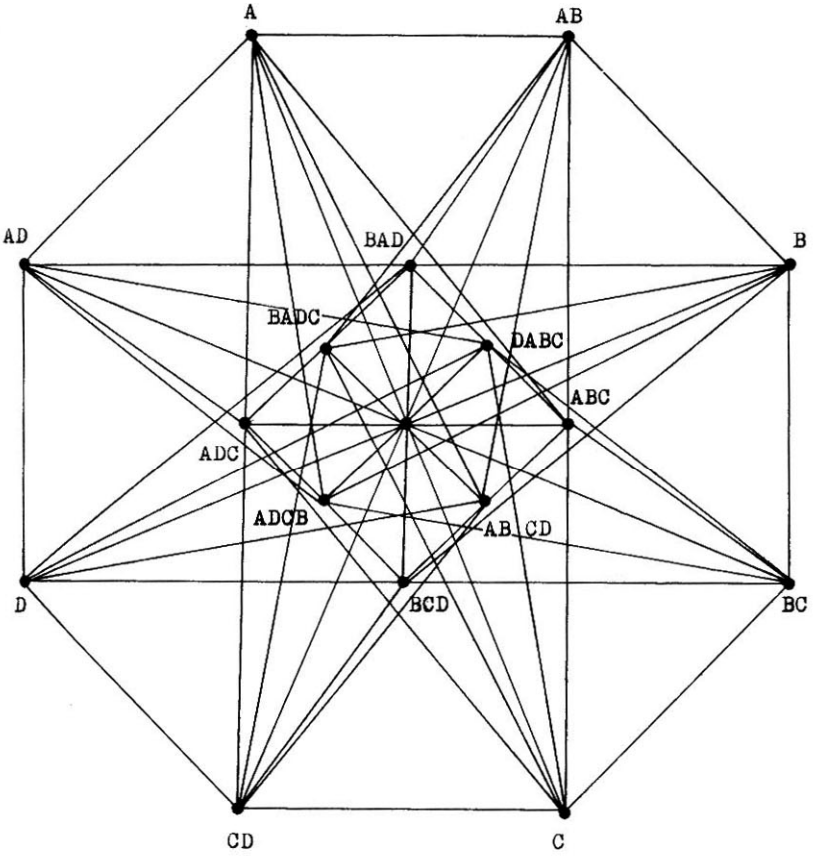


Figure 10. Synthon graph for target molecule $\begin{matrix} A—B \\ | \quad | \\ D—C \end{matrix}$ (represented by the central unlabelled vertex) obtained in reactions which may, or may not, involve cycloadditions.

The code of the target molecule is (4d7) in Table 1.

4. Two general synthon graphs of targets involving n synthons

The preceding Section has presented a discussion of synthon graphs for molecules involving at most four synthons. From the examples displayed so far, it is seen that relations between pseudoconstitutional graphs (i. e. between the structure of target molecules in terms of synthons) and the topology of the corresponding synthon graphs are quite complicated.

In the following we evidence, however, such relations in two simple general cases, namely acyclic linear, and monocyclic, pseudoconstitutional graphs consisting of n synthons A_i, where $i = 1,2,\ldots,n$.

4.1. Acyclic linear pseudoconstitutional graph. The target molecule is $A_1A_2A_3\ldots A_n$. Figure 11 presents the first four terms in this series. Under each synthon graph and target, we list the numbers N of vertices in the synthon graph, according to their genus : synthon vertices (involving one synthon per definitionem), intermediate vertices involving intermediates with 2, 3, or 4 synthons (if any), and the target molecule involving n synthons where n is 2, 3, 4, or 5. The number of synthons in the denomination of the vertex is denoted by S, and the degree of the vertex by D.

From the data in Fig. 11 we generalize easily, in order to provide a demonstration by induction, that the synthon graph for the target molecule with n synthons will contain :

n	synthon vertices for synthons	A_i	(S = 1), degree n-1
n-1	intermediate " " pairs	A_iA_{i+1}	(S = 2), " n
n-2	" " " triples	$A_iA_{i+1}A_{i+2}$	(S = 3), " n+1
.....			
2	" " "	$A_1A_2\ldots A_{n-1}$	(S=n-1), " 2n-3
1	final vertex for the target molecule		(S = n), " 2n-2

This results in a total of n(n+1)/2 vertices for the synthon graph. The algorithm for the construction of this graph by recurrence results from Fig. 11. When we enlarge the synthon graph corresponding to a pseudoconstitutional target molecule of n-1 synthons $A_1A_2\ldots A_{n-1}$ in order to obtain the synthon graph for a target involving n synthons $A_1A_2\ldots A_{n-1}A_n$, we have to add n new vertices (denoted in Fig. 11 by black points, while the

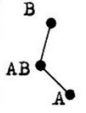
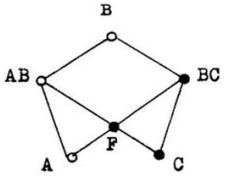
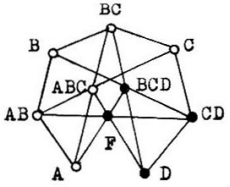
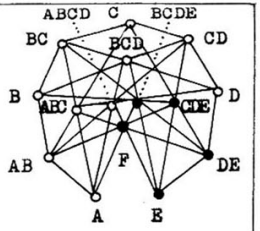
Synthon graph				
Target F	AB	ABC	ABCD	ABCDE
Vertices in synthon graph	N S D	N S D	N S D	N S D
Synthon vertices	2 1 1	3 1 2	4 1 3	5 1 4
Interm. " of 2	- - -	2 2 3	3 2 4	4 2 5
" " " 3	- - -	- - -	2 3 5	3 3 6
" " " 4	- - -	- - -	- - -	2 4 7
Target "	1 2 2	1 3 4	1 4 6	1 5 8
Total "	3	6	10	15

Figure 11. Stepwise formation of synthon graphs for linear arrangements of n synthons (n is the underlined value). The white points indicate vertices of the preceding synthon graph in the series, involving n - 1 synthons ; the newly added n black points form a complete graph K_n . Abbreviations : N = number of vertices ; S = number of synthons in intermediate or target ; D = degree of vertex.

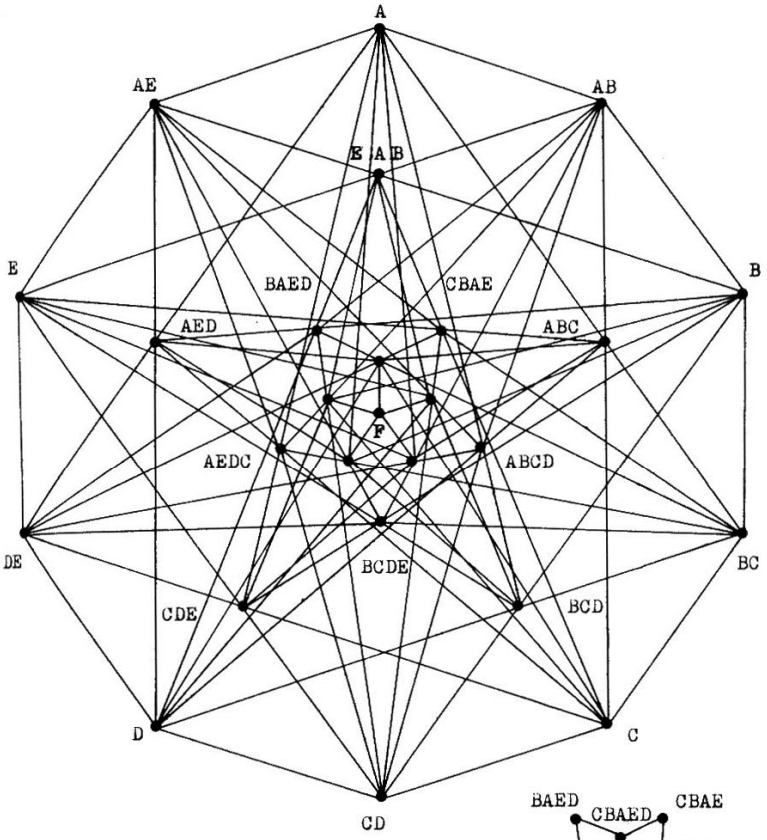
"old" points inherited from the previous synthon graph were white). The new, black, vertices correspond each to groups of 1, 2, ..., n synthons with notations A_n ; $A_{n-1}A_n$; $A_{n-2}A_{n-1}A_n$; ... ; $A_2A_3...A_n$; and the new target $A_1A_2...A_n$, respectively. These new vertices with the edges connecting them are seen to form a complete graph K_n where each black vertex is connected to every other black vertex by an edge.

The last new final (target) vertex is connected to n-1 former "old" vertices, namely to all those whose notations include A_1 ; the vertex $A_2A_3...A_n$ is connected to n-2 former vertices, namely all those with notations including A_2 and excluding A_1 ; and so on ; finally, vertex $A_{n-2}A_{n-1}A_n$ is connected to former vertices $A_{n-2}A_{n-1}$ and A_{n-2} , while the last remaining vertex $A_{n-1}A_n$ is connected only to one former vertex A_{n-1} . In addition to these connections between black and white points, we also have the connections between black points only within the new complete graph K_n.

Since the former graph had $n(n-1)/2$ vertices, the addition of the new n black points is easily seen to afford the total of $n(n+1)/2$ vertices, completing the demonstration by induction.

4.2. Cyclic pseudoconstitutional graph. The situation for the cyclic pseudoconstitutional graph is more complex than the preceding one, because on inserting another synthon into a former pseudoconstitutional graph, not only are there new adjacencies formed in the new target molecule, but also previous adjacencies are partly removed, namely one adjacency. Therefore in the new synthon graph, not only are there new vertices in addition to the former vertices, but some former vertices are missing, unlike the case discussed in the preceding Section. On passing from the synthon graph of a target with n-1 synthons arranged circularly to a target with n synthons, 2n-1 extra vertices have to be added.

Table 2 illustrates the stepwise increase of the synthon graph corresponding to a circular arrangement of n synthons with n = 3, 4, and 5. The last synthon graph is presented in Figure 12, the former two were contained in Figures 4 and 8. Notations in Table 2 are similar to those of Fig. 11. Table 2 also indicates the extrapolated values for the general case involving n synthons.



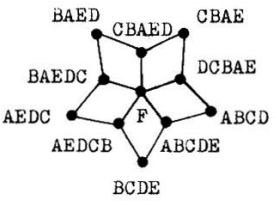


Figure 12. Synthon graph for the pseudoconstitutional graph

F =  without involving cycloadditions. Vertices close to F are labelled as indicated in the right-hand corner.

Table 2. Stepwise formation of synthon graphs for target molecules containing n synthons, all different, in circular arrangement, when cycloadditions are not allowed.

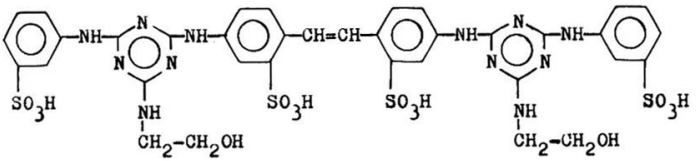
Target												
n	3			4			5			n		
Synthon graph	Fig. 4			Fig. 8			Fig. 12			-		
Vertices in synthon graph	N	S	D	N	S	D	N	S	D	N	S	D
Synthon vertices	3	1	4	4	1	6	5	1	8	n	1	2n-2
Interm. " of 2 synth.	3	2	4	4	2	6	5	2	8	n	2	2n-2
" " " 3 "	3	3	5	4	3	6	5	3	8	n	3	2n-2
" " " 4 "	-	-	-	4	4	7	5	4	8	n	4	2n-2
" " " 5 "	-	-	-	-	-	-	5	5	9	n	5	2n-2
.....	.....			.....			.....			.....		
" " " n-1 "	-	-	-	-	-	-	-	-	-	n	n-1	2n-2
" " " n "	-	-	-	-	-	-	-	-	-	n	n	2n-1
Target "	1	3	3	1	4	4	1	5	5	1	n	n
Total "	10			17			26			n^2+1		

It should be emphasized that the previous discussion of synthon graphs for target molecules consisting of n different synthons arranged circularly only applies to the simplest case, when no cycloadditions are allowed. If cycloadditions must, or may, be involved, the situation becomes even more complicated.

5. Discussion of a practical case involving 7 synthons

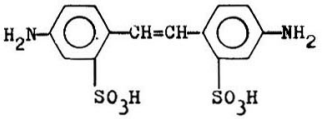
The molecule depicted on top of Figure 13 is a fluorescent whitening agent used in the textile/paper industry.[45]

Despite its complicated appearance, it is easily seen that the above molecule can be synthesized from seven synthons, labelled A - D (the molecule is symmetric so that only four synthon types occur). On constructing the synthon graph in simplified form (effecting only symmetrical operations relative to the symmetry of the pseudoconstitutional graph), the synthon

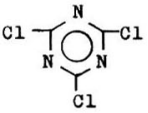


Target molecule : F = C–B(D)–A–B(D)–C or A(B(D)–C)$_2$

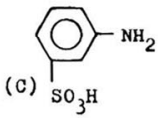
Synthons :



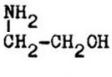
(A)



(B)



(C)



(D)

Synthon graph (symmetry-simplified) :

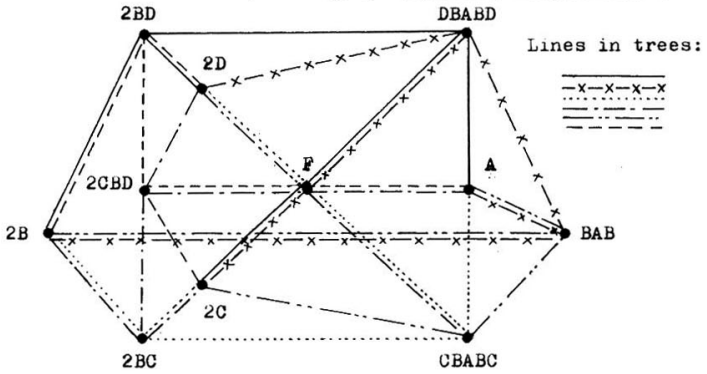


Figure 13. Illustration of a practical example for a target molecule F consisting of seven synthons of four types, and the associated synthon graph involving only symmetrical steps.

graph presented at the bottom of Fig. 13 is obtained. It has 11 vertices, and the target is the root of six different rooted trees, depicted with six different kinds of lines in Fig. 13. Therefore, one has to consider six different temporal sequences in combining the given synthons to afford the target molecule. In practical applications, the merits of each of these six different approaches should be considered separately.

If one considers that a given molecule may be decomposed into, and assembled from, different sets of synthons, and that for each set a synthon graph with many rooted trees will be found, the combinatorial complexity of synthesis design may be fully appreciated, and the need for a computer approach to solving this problem may be explained.

6. Use of synthon graphs and power graphs in synthesis design

Before applying the concept of synthon graphs, the dissection of the target molecule into synthons must be considered. An approach towards making this dissection mathematically complete is to break one bond at a time in all possible ways. This leads us to power graphs.

In an earlier paper,[46] power graphs $G(r^s)$ were defined as a particular case of reaction graphs ; such power graphs have r^s vertices, all of which have degree $s(r-1)$; such graphs describe for instance substitution reactions at s non-equivalent sites, each of which can bind one from r types of reagents. For example the chlorination of $R_2CH-CHR'-CHR''_2$ where $R \neq R' \neq R''$ is described by a power graph $G(2^3)$ which is a 3-cube with eight vertices of degree three ; on the other hand, the alcoholysis of 1,2-dichloropropane with a mixture of ethanol and methanol is described by a power graph $G(3^2)$ which has nine vertices of degree four ; the same last graph also describes the conformational interconversions of staggered propane rotamers.

If we consider a linear acyclic chain of n atoms and the various modes of forming this chain , we have to analyze how the n-1 bonds may be formed one at a time, two at a time, etc. These possibilities represent elementary reaction steps when only one bond is formed at a time, and they correspond to a power graph $G(2^{n-1})$. This is illustrated in Fig. 14 for n = 3 (upper left

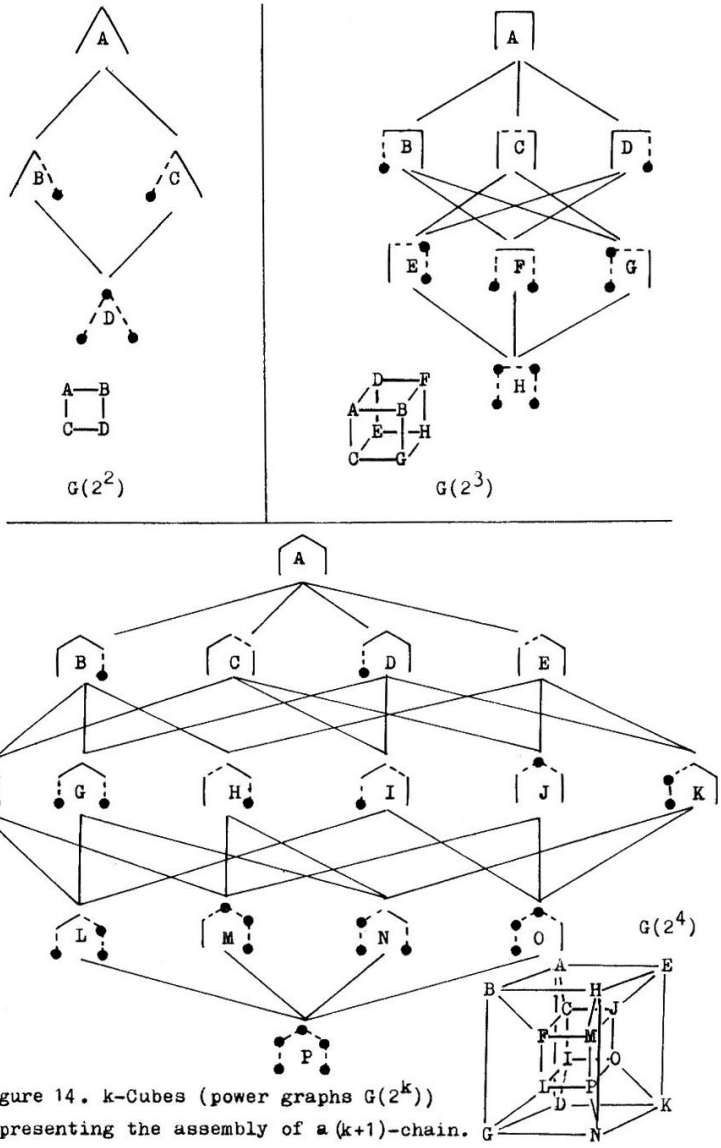


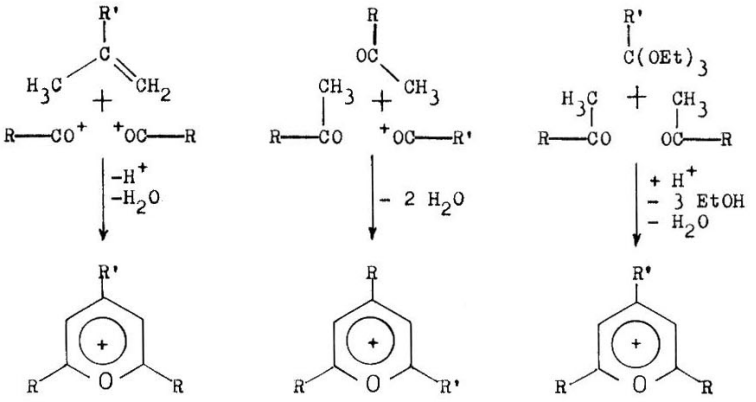
Figure 14. k-Cubes (power graphs $G(2^k)$) representing the assembly of a (k+1)-chain.

corner) : the three-atom chain may be obtained in one piece (vertex denoted by A in the power graph) or from a two-atom synthon and an one-atom synthon by forming one bond (vertices denoted by B and C), or from three one-atom synthons by forming two bonds (vertex denoted by D). The last vertex is adjacent (connected) to the two preceding vertices B and C, but not to A, because an elementary reaction step involves forming or breaking of one bond only. The bond(s) formed or broken is (are) represented in Fig. 14 by broken lines ; solid lines represent covalent bonds linking atoms in synthons, and also edges of power graphs. For each of the three power graphs from Fig. 14, two isomorphic representations are given : one in the conventional manner for a k-cube (n = k+1) only with the letters denoting the vertices, the second with the explicit nature of each vertex, together with its notation, is given. For n = 4 the power graph is a 3-cube, or a usual cube. For n = 5 the power graph $G(2^4)$ is a 4-cube or a hypercube.

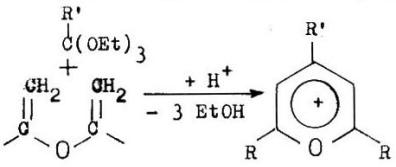
This last power graph corresponds to several chemical applications :[47-49] we shall discuss in more detail one of these corresponding to an analysis of reactions leading to pyrylium salts,[49] whose five-carbon chain can be built from 1-, 2-, 3-, or 4-carbon fragments in various combinations. The ring closure follows automatically from the two carbonyl groups at the ends of the 1,5-enedione chain, in acid medium (the 1,5-enedione is a pyrylium pseudobase). The real two-synthon pathways are in practice 4 + 1 (B or D) or 3 + 2 (C or E) fragment combinations in $G(2^4)$, and the real three-synthon pathways are only 1 + 3 + 1 (vertex denoted G), 2 + 2 + 1 (H or I), and 2 + 1 + 2 (J) where the fragments with the same number of carbon atoms can be identical molecules ; in this case only, a two-component reactant (binary) system results leading to a reaction mixture whose yield in the desired product is significant owing to the restricted number of possible products. In the remaining three-fragment combinations, F and K, or in the four- and five-carbon fragment syntheses L - P, there is no real functional group which could lead to a two-component reactant system, leading to very complicated reaction mixtures, hence to insignificant yields of desired product and to difficult separation problems.[50]

The real three-fragment syntheses of pyrylium salts are presented below :[50]

1 + 3 + 1 | 2 + 2 + 1 | 2 + 1 + 2



The above pyrylium salt syntheses have purposely only considered all possible modes for the formation of C-C bonds in the five-carbon chain, ignoring the C-O bonds which are formed automatically from the carbonyl functions of 1,5-enediones in acid medium. Actually, so far no pyrylium synthesis is known which involves the formation of a C-C bond in a system already possessing the C-O-C bonds, although such possibilities may be definitely considered :



If the C-O bonds would be also taken into consideration, the graph would become more complex.

The pyrylium cation, once formed, is itself a very useful synthon because it reacts readily with various nucleophiles, lengthening their chain with an unsaturated C_5-fragment. In many cases, the product initially formed cyclizes again into an aromatic benzenoid or heterocyclic (six-membered or five-membered) system. A systematic survey of such reactions was published.[49]

It should be stressed that although Fig. 14 denotes the vertices of power graphs with capital letters (like the synthons in all other figures in this paper), power graphs and synthon graphs are quite different kinds of graphs. Chemically, power graphs $G(2^{n-1})$ represent, among other chemical applications like substitutions, all possible modes of assembling a linear acyclic chain of n atoms from fragments of 1 to n atoms through elementary reaction steps of forming one bond at a time. Mathematically, power graphs of the above type are (n-1)-cubes, i. e. regular graphs. On the other hand, synthon graphs are irregular graphs which represent the temporal sequence of bond-forming reactions for a given set of synthons.

One vertex in a power graph represents one possible way of dissecting the n-atom chain, i. e. this vertex corresponds to a whole synthon graph. For a complete mathematical analysis of synthetic strategy, the best approach is to analyze the target molecule in terms of possible decompositions into synthons by means of power graphs or of analogous constructions. Then the real solutions should be selected, e. g. for the C_5-chain of the pyrylium salts the 1-, 2-, or 3-fragment cases. According to the chemical problem at hand, one of these solutions should be then further developed by means of its synthon graph, e. g. the three-fragment approach involving 2 + 2 + 1 or 2 + 1 + 2 atom fragments. In terms of synthon graphs, these correspond to target molecules of type AAB or ABA, analyzed in Figure 3. The former synthon graph is seen in Fig. 3 to possess two different trees rooted at F : one of these involves the intermediate AA (corresponding in real terms to an α,ß-unsaturated ketone formed from two moles of methyl ketone, e. g. dypnone from acetophenone) and the other involves a 1,3-diketone as intermediate AB (e. g. dibenzoylmethane formed from acetophenone and benzoyl chloride) ; indeed, both types of intermediates do lead to pyrylium salts (i. e. benzoylation of dypnone, or condensation of dibenzoylmethane with acetophenone, leads in both cases to 2,4,6-triphenylpyrylium). This example illustrates the heuristic value of the synthon graph approach for unveiling new synthetic approaches, or for elucidating the reaction mechanisms, in addition to the systematic exploration of all possible synthetic routes.

References

1. Preceding part, A. T. Balaban, I. Moţoc and R. Vancea, to appear.
2. D. H. Rouvray, Endeavour (New Series), 1, 23 (1977) ; idem, in "Chemical Applications of Graph Theory" (ed. A. T. Balaban), Academic Press, London, 1976, p. 175.
3. D. H. Rouvray and A. T. Balaban, in "Applications of Graph Theory" (eds. R. J. Wilson and L. W. Beineke), Academic Press, London, 1979, p. 177.
4. R. B. King and D. H. Rouvray, J. Am. Chem. Soc. 99, 7834 (1977).
5. W. D. Stohrer and R. Hoffmann, J. Am. Chem. Soc. 94, 1661 (1972) ; M. J. S. Dewar and R. C. Haddon, ibid., 95, 5836 (1973) ; W. J. Hehre and P. von R. Schleyer, ibid. 95, 5837 (1973) ; H. Hogeveen and P. W. Kwant, Acc. Chem. Res. 8, 413 (1975) ; A. T. Balaban and D. H. Rouvray, Tetrahedron (in press) and references quoted therein.
6. A. T. Balaban, D. Fărcaşiu and R. Bănică, Rev. Roumaine Chim. 11, 1025 (1966).
7. P. C. Lauterbur and F. Ramirez, J. Am. Chem. Soc. 90, 6722 (1968).
8. K. Mislow, Acc. Chem. Res. 10, 231 (1970) ; G. Zon and K. Mislow, Fortschr. Chem. Forsch. 19, 61 (1971).
9. M. Gielen, "Stéréochimie dynamique", Freund Publication House, Tel Aviv, 1974.
10. R. Luckenbach, "Dynamic Stereochemistry of Pentacoordinated Phosphorus and Related Elements", G. Thieme Verlag, Stuttgart, 1973, p. 214.
11. J. B. Hendrickson, J. Am. Chem. Soc. 99, 5439 (1977).
12. E. J. Corey and W. T. Wipke, Science, 166, 178 (1969).
13. E. J. Corey, W. Howe and D. A. Pensack, J. Am. Chem. Soc. 98, 210 (1976) ; 96, 7724 (1974) and earlier papers ; D. A. Pensack and E. J. Corey, in "Computer-Assisted Organic Synthesis" (eds. W. T. Wipke and W. J. Howe), A.C.S. Symp. Series No. 61, Am. Chem. Soc., Washington, D. C. 1977.
14. W. T. Wipke and P. Gund, J. Am. Chem. Soc. 98, 8107 (1976)

and earlier papers.

15. S. Turner, "The Design of Organic Syntheses", Elsevier, Amsterdam, 1976.

16. H. Gelernter et al., Topics Curr. Chem. 41, 113 (1973) ; Science, 197, 1041 (1977).

17. J. Gasteiger and C. Jochum, Topics Curr. Chem. 74, 93 (1978) ; I. Ugi, Intra-Science Chem. Repts. 5, 229 (1971) ; I. Ugi, P. Gillespie and C. Gillespie, Ann. N. Y. Acad. Sci. 34, 416 (1972) ; J. Brandt, J. Friedrich, J. Gasteiger, C. Jochum, W. Schubert and I. Ugi, in "Computers in Chemical Education and Research" (eds. E. V. Ludena, N. H. Sabelli and A. C. Wahl), Plenum Press, New York, 1971 ; J. Brandt et al., in "Computer-Assisted Organic Syntheses" (eds. W. T. Wipke and W. J. Howe), A. C. S. Symp. Series No. 61, Am. Chem. Soc., Washington, D. C. 1977 ; I. Ugi et al., Angew. Chem. 91, 99 (1979) ; H. Gelernter et al., Topics Curr. Chem. 41, 113 (1973) ; J. Blair et al., in "Computer Representation and Manipulation of Chemical Information" (eds. W. T. Wipke, S. Heller, R. Feldmann and E. Hyde), Wiley, New York, 1974, p. 129 ; J. Gasteiger, C. Jochum, M. Marsili and J. Thoma, Math. Chem. to appear.

18. A. J. Thakkar, Topics Curr. Chem. 39, 3 (1973).

19. M. Bersohn and A. Esack, Chem. Revs. 76, 269 (1976) ; M. Bersohn, Bull. Chem. Soc. Japan, 45, 1897 (1972) ; A. Esack and M. Bersohn, J. Chem. Soc. Perkin I, 2463 (1974).

20. Y. Yoneda, Bull. Chem. Soc. Japan, 52, 8 (1979) ; idem, in "Information Chemistry. Computer-Assisted Chemical Research Design" (eds. S. Fujiwara and H. B. Mark Jr.), Univ. of Tokyo Press, 1975, p. 239.

21. F. Choplin, P. Bennet, M. H. Zimmer and G. Kaufmann, Nouveau J. Chim. 3, 223 (1979).

22. P. Barone, M. Chanon and J. Metzger, Rev. Inst. Fr. Petrole, 28, 771 (1973) ; R. Barone and M. Chanon, Nouveau J. Chim. 2, 659 (1978) ; R. Barone, Ph. D. Thesis, Marseille, 1976.

23. P. E. Blower Jr. and H. W. Whitlock, J. Am. Chem. Soc. 98, 1499 (1976).

24. A. Weise, Z. Chem. 15, 333 (1975) ; A. Weise and H.-G. Scharnow, ibid. 19, 249 (1979).

25. J. B. Hendrickson, Topics Curr. Chem. 62, 49 (1976) and further references by the same author quoted therein.
26. W. T. Wipke, in "Computer Representation and Manipulation of Chemical Information" (eds. W. T. Wipke, S. Heller, R. Feldmann and E. Hyde), Wiley, New York, 1974.
27. J. Blair, J. Gasteiger, C. Gillespie, P. D. Gillespie and I. Ugi, Tetrahedron, 30, 1845 (1974).
28. G. Stork and K. Isobe, J. Am. Chem. Soc. 97, 6260 (1975).
29. T. F. Browscombe, Diss. Abstr. Int. B 34, 1035 (1973).
30. G. J. Powers, R. L. Jones, G. A. Randall and M. H. Caruthers, J. Am. Chem. Soc. 97, 875 (1975) ; P. Benedek, Magy. Chem. Lapja, 30, 48 (1975).
31. L. H. Sarret, quoted after ref. 32.
32. Cs. Szántay and L. Novak, "The Synthesis of Prostaglandins", Akadémiai Kiado, Budapest, 1978, p. 165.
33. E. J. Corey, W. T. Wipke, R. D. Cramer and W. J. Howe, J. Am. Chem. Soc. 94, 421, 431 (1972).
34. E. J. Corey et al., J. Am. Chem. Soc. 100, 8034 (1978).
35. O. Sinanoglu, J. Am. Chem. Soc. 97, 2309 (1975).
36. O. Sinanoglu, Theoret. Chim. Acta (Berl.), 48, 287 (1978).
37. O. Sinanoglu and L. S. Lee, Theoret. Chim. Acta (Berl.), 51, 1 (1979).
38. R. B. Woodward and R. Hoffmann, Angew. Chem. Int. Ed. Engl. 8, 781 (1969) ; "Die Erhaltung der Orbitalsymmetrie", Akademische Verlagsges. Geest & Portig K.-G., Leipzig, 1970.
39. N. T. Anh, "Les règles de Woodward-Hoffmann", Ediscience, Paris, 1970 ; German translation, Verlag Chemie, 1972.
40. P. Wieland and H. Kaufmann, "Die Woodward-Hoffmann-Regeln. Einführung und Handhabung", Uni-Taschenbücher Nr. 88, Birkhäuser Varlag, Basel, 1972.
41. M. J. S. Dewar, Angew. Chem. Int. Ed. Engl. 10, 761 (1971).
42. M. J. S. Dewar, S. Olivella and H. S. Rzepa, J. Am. Chem. Soc. 100, 5650 (1978).
43. A. T. Balaban and D. Färcaşiu, in "Proceedings of the Second International Conference on Methods of Preparing and Storing Labelled Compounds", Brussels, Nov. 28-Dec. 3, 1966. Euratom, Brussels, 1968, p. 611.

44. A. Barabas and A. T. Balaban, Rev. Roumaine Chim. 19, 1927 (1974).

45. B. H. Carroll and J. E. Jones (to Eastman Kodak), U. S. Pat. 2875058, Feb. 24, 1959 ; Chem. Abs. 53, 9867 (1959) ; J. C. McFall and P. R. Crookshank (to Eastman Kodak), U. S. Pat. 2933390, Apr. 19, 1960 ; Chem. Abs. 54, 20596 (1960) ; F. G. Villaume (to Amer. Cyanamid Co.), U. S. Pat. 31332106, May 5, 1964 ; Chem. Abs. 61, 7150 (1964) ; B. Noll et al., East Germ. Pat. 55668, May 5, 1967 ; Chem. Abs. 67, 101025 (1967).

46. A. T. Balaban and F. Kerek, Rev. Roumaine Chim. 19, 631 (1974).

47. K. Mislow, D. Gust, P. Finocchiaro and J. Boettcher, Topics Curr. Chem. 47, 1 (1974).

48. P. Murray-Rust and F. G. Riddell, Tetrahedron, 32, 427 (1976).

49. A. T. Balaban, in "New Trends in Heterocyclic Chemistry" (eds. R. B. Mitra, N. R. Ayyangar, V. N. Gogte, R. M. Acheson and N. Cromwell), Elsevier, Amsterdam, 1979, p. 79.

50. A. T. Balaban, W. Schroth and G. Fischer, Adv. Heterocyclic Chem. 10, 241 (1969).