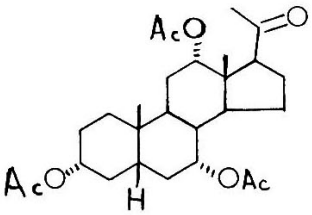
match no. 17 pp. 175-190 1985

The Use of Directed Graphs in Mass Spectrometry.

Jerry Ray Dias
Department of Chemistry
University of Missouri
Kansas City, Missouri 64110

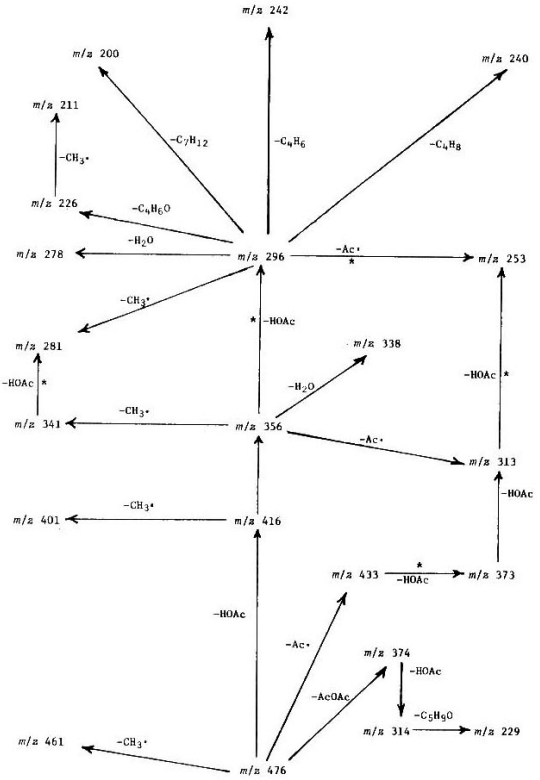
(received: June 1984)

Genesis schemes provide a quicker and more complete summary of the information in a mass spectrum. They are little used by mass spectroscopists, and are completely absent in introductory texts on mass spectrometry.[1] The author of this paper has required all his students to construct probable genesis schemes for every mass spectrum and has found such schemes to be useful aids in student training in interpretive mass spectrometry. Thus this summary of the graph theoretical properties of genesis schemes as applied to a complicated steroid molecule is presented with the intent to encourage their use in student teaching.[2]



1, $C_{27}H_{40}O_7$, $[M]^{+\cdot} = 476$

Scheme I



Genesis Schemes

A genesis scheme should be constructed for a complex mass spectrum to show pictorially the probable or possible origin of the various daughter ions. These pictorial schemes are known in mathematics as acyclic digraphs (directed graphs without rings that result from unidirectional arrows to form a closed clockwise or counterclockwise cycle). Since there are some properties of digraphs which are worthy of discussion in regard to their application in computer interpretation in mass spectrometry, some relevant terminology is now introduced.[3,4] Consider Genesis Scheme I (for the 12eV mass spectrum *1*). If this scheme is complete, then the point (vertex) corresponding to the molecular ion (*m/z* 476) is called the transmitter, points corresponding to *m/z* 433, 374, 373, 341, 314 and 226 are carrier points, points corresponding to *m/z* 416, 356, 313 and 296 are ordinary points, and points corresponding to *m/z* 461, 401, 338, 281, 278, 253, 242, 240, 229, 211, and 200 are receiver points. The number of arrows (edges) into a point i is called the *indegree* id_i of that point and the number of arrows out of a point i is called the *outdegree* od_i; the number of arrows (edges) q is given by: $\sum_{i=1}^{p} id_i = \sum_{i=1}^{p} od_i = q$ where p = total No. of graph points. Thus transmitter points have $id_i = 0$, carrier points have $id_i = od_i = 1$, and receiver points have $od_i = 0$. Every digraph has a defined vertex set [$P = \{v_1, v_2, \ldots\}$] and a directed edge set [$Q = \{(v_a, v_b), \ldots\}$] having ordered vertex pairs (v_a, v_b) determined by some relation (e.g., a genesis scheme).

Within the framework of set theory, a relation $\preceq$ in a set S is called a *partial order* (or *order*) on S iff, for every a, b, $c \in S$: (i) $a \preceq a$ (reflective criterion); (ii) $a \preceq b$ and $b \preceq a$ implies $a=b$ (symmetric criterion); and (iii) $a \preceq b$ and $b \preceq c$ implies $a \preceq c$ (transitive criterion). The set S together with the partial order, i.e., the pair $(S, \preceq)$ is called a *partially ordered set*. If $a \preceq b$ in an ordered set, then it is said that a precedes b and that b follows a. The ion peaks in a mass spectrum of a compound and the associated genesis scheme represents such a partially ordered set (e.g., *1* and Scheme I). An element $a \in S$, is the *first* element iff $a \preceq s$ for all $s \in S$, and an element $b \in S$ is the *last* element of S iff $s \preceq b$ for all $s \in S$. Scheme I has only the first element *m/z* 476 and no last elements. An element $a \in S$ is *maximal* if no other element follows a, and an element $b \in S$ is *minimal* if no other element precedes b. In Scheme I, *m/z* 476 is a minimal element and *m/z* 461, 401, 338, 281, 278, 253, 242, 240, 229, 211, and 200 are maximal elements. Let A be a subset of a partially ordered set S ($A \subset S$). An element $m \in S$ is a *lower bound* of A iff $m \preceq X$ for all $x \in A$, i.e. if m precedes every element in A. If some lower bound of A follows every other lower bound at A, then it is the *greatest lower bound* (glb) or *infimum* of A [inf(A)]. Similarly, an element $M \in A$ is an *upper bound* of A iff $x \preceq M$ for all $x \in A$, i.e. if M follows every element in A. If some upper bound of A precedes every other upper bound of A, then it forms the *least upper bound* (lub) or *supremun* of A [sup(A)] lon peaks *m/z* 416 and 401 is a subset in Scheme I which has *m/z* 476 as a glb and *m/z* 356 as a lub.

Matrix I. Adjacency matrix, S(R), of genesis Scheme I for the mass spectrum of 3α,7α,12α-triacetoxy-5β-pregnan-20-one

	200	211	226	229	240	242	253	278	281	296	313	314	338	341	356	373	374	401	416	433	461	476
200	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
211	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
226	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
229	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
240	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
242	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
253	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
278	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
281	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
296	1	0	1	0	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0
313	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
314	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
338	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
341	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
356	0	0	0	0	0	0	0	0	0	1	1	0	1	1	0	0	0	0	0	0	0	0
373	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
374	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
401	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
416	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1	0	0	0	0
433	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
461	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
476	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	1	1	0

Given a relation R on a set S of $|P| = p$ points which have been labeled $v_1, v_2, \ldots, v_p$, the adjacency matrix S(R) is defined to be the p x p matrix whose (i,j) entry is one if the ordered pair (v_i, v_j) is in R and zero otherwise.[5] Thus the adjacency matrix for the graph of Scheme I is shown in Matrix I. For example in Scheme I, the 476 ion goes to the 461, 433, 416 and 374 ions, and therefore this row has a 1 in these corresponding columns in Matrix I and 0 in the remaining columns. A *walk* in a directed graph is an alternating sequence of points and arrows (directed edges) such that each arrow is directed from the point preceding it to the point following it. A walk may be denoted by its point sequence, e.g., $v_1 v_2 \ldots v_n$, where the intervening arrows are implied; the number of occurrences of arrows in a walk is its length. The point v is said to be reachable from the point u if there is a walk from u to v. The length of the shortest walk from u to v is the distance between u and v . A *semiwalk* joining u and v is an alternating sequence of points and arrows, $v_1, x_1, v_2, x_2, \ldots, v_p$, where x_i is either the ordered pair (v_i, v_{i-1}) or (v_{i+1}, v_i). A *spanning* walk or semiwalk contains all the points of S. A spanning tree from a point v consists of an acyclic spanning semiwalk or walk from v, and spanning tree to a point v is an acyclic spanning semiwalk or walk to v.

The number of walks of length one is the sum of all the ones in the adjacency matrix S(R); e.g., the number of walks of length one is 24 for the graph of Scheme I which is equal to the total number of arrows in it. In general, the number of walks of length n is the sum of the numbers contained in the matrix obtained by raising the adjacency matrix to the $n^{\underline{th}}$ power (S^n), and the number of walks of length n from point v_i to point v_j in S equals the numerical value of the (i,j) entry in S^n. Thus the number of

Matrix II. The 2nd power of the adjacency matrix, $[S(R)]^2$, for genesis Scheme I.

	200	211	226	229	240	242	253	278	281	296	313	314	338	341	356	373	374	401	416	433	461	476
200	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
211	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
226	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
229	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
240	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
242	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
253	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
278	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
281	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
296	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
313	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
314	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
338	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
341	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
356	1	0	1	0	1	1	2	1	2	0	0	0	0	0	0	0	0	0	0	0	0	0
373	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
374	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
401	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
416	0	0	0	0	0	0	0	0	0	1	1	0	1	1	0	0	0	0	0	0	0	0
433	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
461	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
476	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1	1	0	1	0	0	0	0

Matrix III. The reachability matrix, N(S), for the graph of Scheme I.

	200	211	226	229	240	242	253	278	281	296	313	314	338	341	356	373	374	401	416	433	461	476
200	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
211	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
226	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
229	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
240	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
242	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
253	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
278	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
281	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
296	1	1	1	0	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0
313	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
314	0	0	0	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
338	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
341	0	0	0	0	0	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0
356	1	1	1	0	1	1	1	1	1	1	1	0	1	1	1	0	0	0	0	0	0	0
373	0	0	0	0	0	0	1	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0
374	0	0	0	1	0	0	0	0	0	0	0	1	0	0	0	0	1	0	0	0	0	0
401	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0
416	1	1	1	0	1	1	1	1	1	1	1	0	1	1	1	0	0	1	1	0	0	0
433	0	0	0	0	0	0	1	0	0	0	1	0	0	0	0	1	0	0	0	1	0	0
461	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
476	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1

Matrix IV. The distance matrix, D(S), for Scheme I.

	200	211	226	229	240	242	253	278	281	296	313	314	338	341	356	373	374	401	416	433	461	476
200	0	∞	∞	∞	∞	∞	∞	∞	∞	∞	∞	∞	∞	∞	∞	∞	∞	∞	∞	∞	∞	∞
211	∞	0	∞	∞	∞	∞	∞	∞	∞	∞	∞	∞	∞	∞	∞	∞	∞	∞	∞	∞	∞	∞
226	∞	1	0	∞	∞	∞	∞	∞	∞	∞	∞	∞	∞	∞	∞	∞	∞	∞	∞	∞	∞	∞
229	∞	∞	∞	0	∞	∞	∞	∞	∞	∞	∞	∞	∞	∞	∞	∞	∞	∞	∞	∞	∞	∞
240	∞	∞	∞	∞	0	∞	∞	∞	∞	∞	∞	∞	∞	∞	∞	∞	∞	∞	∞	∞	∞	∞
242	∞	∞	∞	∞	∞	0	∞	∞	∞	∞	∞	∞	∞	∞	∞	∞	∞	∞	∞	∞	∞	∞
253	∞	∞	∞	∞	∞	∞	0	∞	∞	∞	∞	∞	∞	∞	∞	∞	∞	∞	∞	∞	∞	∞
278	∞	∞	∞	∞	∞	∞	∞	0	∞	∞	∞	∞	∞	∞	∞	∞	∞	∞	∞	∞	∞	∞
281	∞	∞	∞	∞	∞	∞	∞	∞	0	∞	∞	∞	∞	∞	∞	∞	∞	∞	∞	∞	∞	∞
296	1	2	1	∞	1	1	1	1	1	0	∞	∞	∞	∞	∞	∞	∞	∞	∞	∞	∞	∞
313	∞	∞	∞	∞	∞	∞	1	∞	∞	∞	0	∞	∞	∞	∞	∞	∞	∞	∞	∞	∞	∞
314	∞	∞	∞	1	∞	∞	∞	∞	∞	∞	∞	0	∞	∞	∞	∞	∞	∞	∞	∞	∞	∞
338	∞	∞	∞	∞	∞	∞	∞	∞	∞	∞	∞	∞	0	∞	∞	∞	∞	∞	∞	∞	∞	∞
341	∞	∞	∞	∞	∞	∞	∞	∞	1	∞	∞	∞	∞	0	∞	∞	∞	∞	∞	∞	∞	∞
356	2	3	2	∞	2	2	2	2	2	1	1	∞	1	1	0	∞	∞	∞	∞	∞	∞	∞
373	∞	∞	∞	∞	∞	∞	2	∞	∞	∞	1	∞	∞	∞	∞	0	∞	∞	∞	∞	∞	∞
374	∞	∞	∞	2	∞	∞	∞	∞	∞	∞	∞	1	∞	∞	∞	∞	0	∞	∞	∞	∞	∞
401	∞	∞	∞	∞	∞	∞	∞	∞	∞	∞	∞	∞	∞	∞	∞	∞	∞	0	∞	∞	∞	∞
416	3	4	3	∞	3	3	3	3	3	2	2	∞	2	2	1	∞	∞	1	0	∞	∞	∞
433	∞	∞	∞	∞	∞	∞	3	∞	∞	∞	2	∞	∞	∞	∞	1	∞	∞	∞	0	∞	∞
461	∞	∞	∞	∞	∞	∞	∞	∞	∞	∞	∞	∞	∞	∞	∞	∞	∞	∞	∞	∞	0	∞
476	4	5	4	3	4	4	4	4	4	3	3	2	3	3	2	2	1	2	1	1	1	0

Matrix V. Matrix (M_i) for determining the number of different spanning trees from any point in the graph of Scheme I.

	200	211	226	229	240	242	253	278	281	296	313	314	338	341	356	373	374	401	416	433	461	476
200	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
211	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
226	0	-1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
229	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
240	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
242	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
253	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
278	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
281	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0
296	-1	0	-1	0	-1	-1	-1	-1	-1	1	0	0	0	0	0	0	0	0	0	0	0	0
313	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0
314	0	0	0	-1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
338	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
341	0	0	0	0	0	0	0	0	-1	0	0	0	0	1	0	0	0	0	0	0	0	0
356	0	0	0	0	0	0	0	0	0	-1	-1	0	-1	1	0	0	0	0	0	0	0	0
373	0	0	0	0	0	0	0	0	0	0	-1	0	0	0	0	1	0	0	0	0	0	0
374	0	0	0	0	0	0	0	0	0	0	0	-1	0	0	0	0	1	0	0	0	0	0
401	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0
416	0	0	0	0	0	0	0	0	0	0	0	0	0	0	-1	0	0	-1	1	0	0	0
433	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	-1	0	0	0	1	0	0
461	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
476	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	-1	0	-1	-1	-1	0

walks in Scheme I of length two (i.e., the numerical sum in Matrix II) is 21, length three is 17, length four is 11, length five is 1 and length six or greater is 0.

The *reachability matrix* N(S) has $n_{ij} = 1$ if v_j is reachable from v_i and $n_{ij} = 0$ otherwise. For the digraph of Scheme I, this is given by Matrix III. Letting s^n_{ij} denote the (i,j) entry (ith row, jth column) of S^n, then $n_{ij} = 1$ for $i \neq j$ if and only if $s^n_{ij} > 0$ for some n; note that $n_{ii} = 1$ for all i. Since the molecular ion, *m/z* 476, can reach all other ions a 1 appears in every column of the 476 row in Matrix III; whereas, the *m/z* 461, 401, 338, 281, 278, 253, 242, 240, 229, 211, and 200 ions can only reach themselves and thus their corresponding rows have only one 1 in matrix diagonal position and the rest 0's. In the *distance matrix* D(S) the d_{ij} entry is the reachable minimum distance from v_i to v_j. For $i \neq j$ the d_{ij} is the value of $s^n_{ij} > 0$ for the smallest n of S^n, if any exists, and ∞ otherwise; $d_{ii} = 0$ for all i. Matrix IV is the distance matrix for the labeled digraph of Scheme I. In D(S) (Matrix IV) a 0 will appear in the diagonal positions and there will be an ∞ at each corresponding matrix position in which a 0 appears in the N(S) (*cf.* with Matrix III). In Matrix IV there are 24 ones corresponding to the number of walks of length one in the digraph of Scheme I. The number of spanning trees of a labeled digraph S from one of its points can be calculated from the matrix $M_i(S) = id(S) - S(R)$ where id is the indegree matrix which contains the indegrees of the points of S on its diagonal and 0 elsewhere. The value of the cofactor of any element in the jth column of M_i is the number of different spanning trees from v_j. Matrix V gives M_i for the digraph of Scheme I. Since the indegree for the *m/z* 476 point is zero, the cofactors for all the columns except the *m/z* 476 column are zero; for any element in the 476 column,

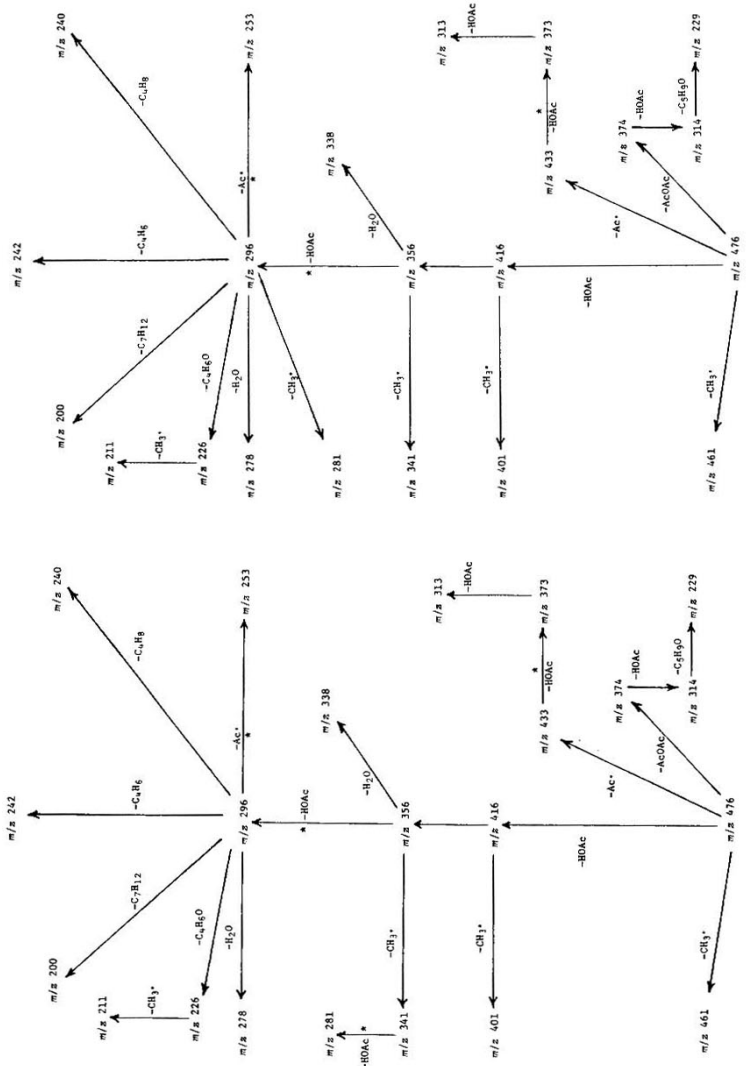


Figure 1. The 8 different spanning trees of Scheme I from the *m/z* 476 point.

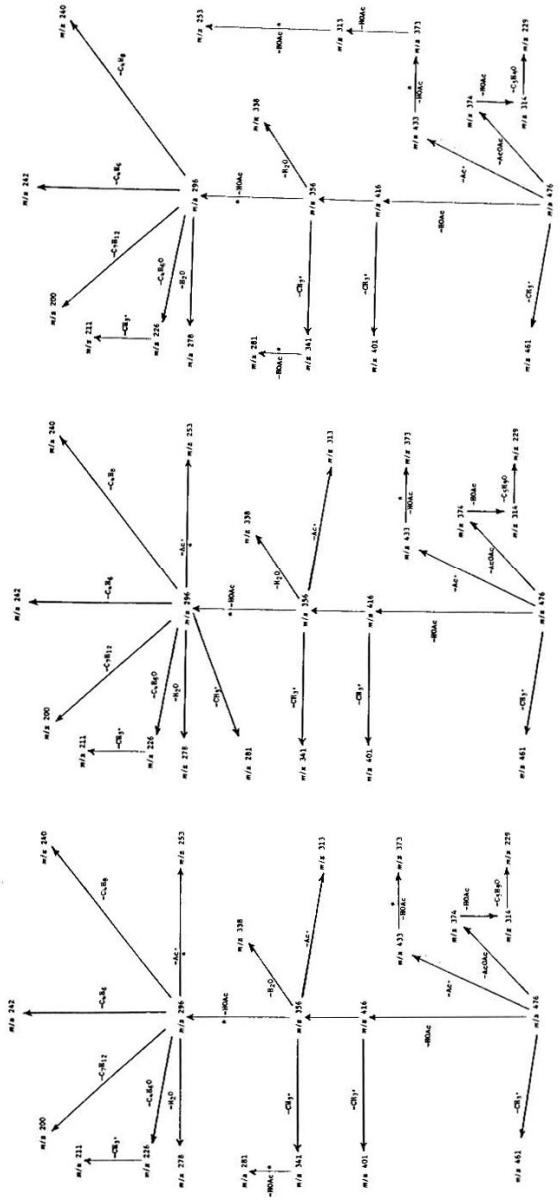


Figure 1 (cont'd)

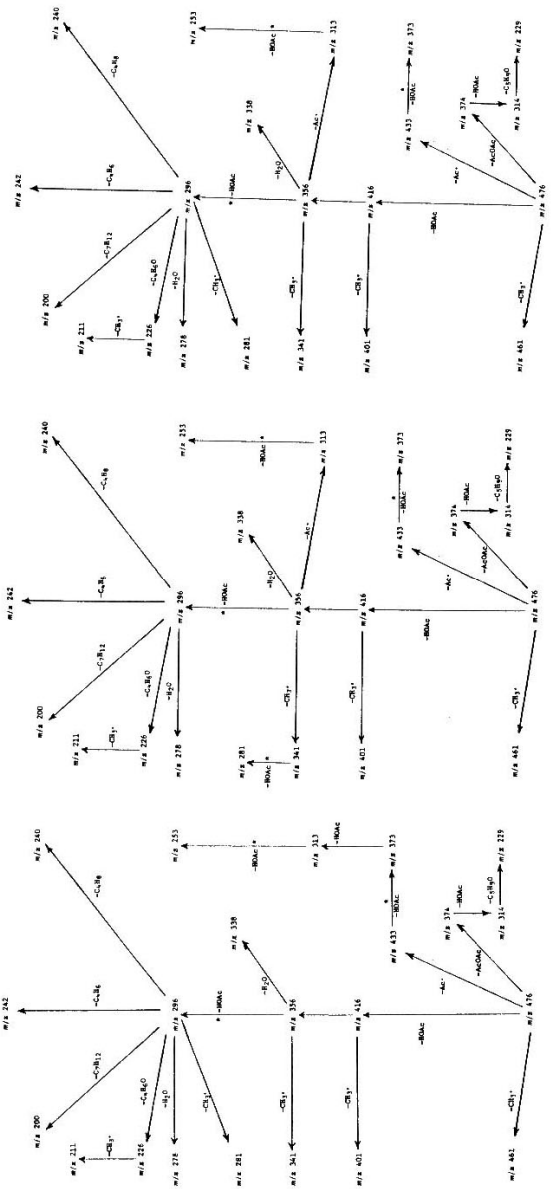


Figure 1 (cont'd)

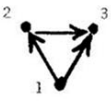
the cofactor is 2·2·2 = 8, and thus there are 8 different spanning trees from the *m/z* 476 point. Similarly, to compute the number of spanning trees to each point in a graph, we define $od(S)$ to be the matrix with the outdegrees of the points on its diagonal and matrix $M_o(S) = od(S) - S(R)$ and calculate the cofactors of elements in rows of M_o. These matrices convey mass spectral information concerning parent, daughter, granddaughter, et cetera ion relationships in lattice form amenable to computer programming.

In mass spectral interpretation, the first step in the analysis of a mass spectrum is to determine the parent ion or ions for each ion in the spectrum (except for the molecular ion). All ions in a spectrum form a set S (S for spectrum) which are points in a digraph of a genesis scheme. This process comprises the finding of all subsets (S_i) of S which are mechanistically related. The system of all statistically possible subsets (A_i) of S is called the power set of S: $P(S) = \{A_i : A_i \subseteq S\}$. In Scheme 1, S = {*m/z* 476, 461, 433, 416, 401, 374, 373, 356, 341, 338,314, 313, 296, 281, 278, 253, 242, 240, 229, 226, 211, 200} has $|S| = 22$ elements; the power set of S, P(S), is all combinations of the elements of S and has $|P(S)| = [22! \left(\sum_{N=0}^{22} \frac{1}{N!\ (22-N)!}\right)] = 2^{22} = 4{,}194{,}304$ subsets of S as elements. Fortunately, not all statistical possibilities need to be considered as general mechanistic principles allow the mass spectroscopist to narrow the number of subsets. For example in Scheme I, the *m/z* 229 ion cannot derive from *m/z* 240 because loss of 11 atomic mass units is impossible for a typical organic compound, but it can derive from *m/z* 314 by loss of the stable neutral $C_3H_6COCH_3$. Thus A_1 = {*m/z* 240,229} is an element of the power set P(S) but is mechanistically unacceptable and $A_2 = S_1$ = {*m/z* 314, 229} is an acceptable element of P(S) and is an interpretable subset of S since its

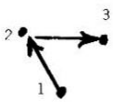
elements are mechanistically related. In Scheme I the total number of interpretable subsets of S is equal to the sum of all the walks of length 1,2, . . . which is 24 + 21 + 17 + 11 + 1 = 75. A subset reduction from 4,194,304 to 75 represents an exemplary intellectual task performed by a mass spectoscopist in his or her interpretation of a mass spectrum for a large molecule such as steroid *1*.

References

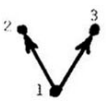
1. For example, see F. W. McLafferty, "Interpretation of Mass Spectra," Univ. Sci Books, Mill Valley, CA, 3rd Edition, 1980.
2. J. R. Dias and B. Nassim, *J. Org. Chem.*, *45*, 337 (1980).
3. S. S. Anderson, "Graph Theory and Finite Combinatorics," Markham Publishing Co., Chicago, IL, 1970.
4. R. E. Prather, "Discrete Mathematic Structures for Computer Science," Houghton-Mifflin Co., Boston, MA, 1976.
5. The reader can easily verify all the following relationships with the simple directed graph (*a*) below which has the point set S = {1,2,3}, the power set P(S) = {ϕ,(1),(2),(3),(1,2),(1,3),(2,3),(1,2,3)}, three walks of length one, one walk of length two, two spanning trees (*b* and *c*) from point 1.



a



b



c