JOURNAL
OF THE AMERICAN CHEMICAL SOCIETY

VOLUME 97, NUMBER 9 APRIL 30, 1975

Theory of Chemical Reaction Networks. All Possible Mechanisms or Synthetic Pathways with Given Number of Reaction Steps or Species

Oktay Sinanoğlu[1]

Contribution from the Sterling Chemistry Laboratory, Yale University, New Haven, Connecticut 06520. *Received May 9, 1974*

Abstract: A method is presented which allows all possible mechanisms with a specified number of elementary reaction steps to be drawn a priori as simple networks for a given overall reaction. For computer assisted organic synthesis design, the method gives the pathways between reagents and synthetic target. The networks and their more compact skeletons, drawn simply from the rules given, are classified in regards to the catalytic, feedback loops, etc., they contain. Steady states, their stabilities, rate laws, or yields depend on these few possible network skeleton types. This method systematizes the various mechanisms, proposed in chemical and in enzyme kinetics, yielding them and many new ones a priori for broad classes of overall reaction types. Definite numbers of intermediates, catalysts, and precursors that are possible for any type of mechanism or synthetic pathway are predicted once the number of steps is given and vice versa. Methods are illustrated on chain reactions, enzyme mechanisms with inhibition, some organic syntheses, biochemical electron transport, and oxidative phosphorylation. Biochemical pathways and laboratory syntheses lead to very different network skeleton types reflecting the regulation requirements of the former.

I. Objectives and the Problem

A chemical reaction network is a set of coupled reaction steps. The steps are the "elementary" reactions in a mechanism, the synthetic steps in an organic synthesis, and the enzymatic or other steps in a metabolic cycle. The steps are coupled by the occurrence, in their reactant and product pools, of chemical species common to two or more steps.

In all such networks, the steps add up to an overall reaction (OVR). The steps are stoichiometric, as is, therefore the OVR.

The analysis below, and in subsequent papers, of networks will be useful: (a) for design, particularly computer assisted, of organic synthesis; (b) in finding all the possible mechanisms given an OVR, and the corresponding gross rate laws; and (c) for studying the steady states, yields, and control and regulation characteristics of kinetic, organic, or biochemical mechanisms or pathways.

Organic Synthetic Pathways. Given a target organic molecule to be synthesized and common organic or industrial reagents available, what are the possible pathways {P} between the start and finish? The first thing we find is that if we specify only the number, ρ, of individual synthetic steps to make the {P}, but not the actual steps, there are still a limited number of pathways. There are also only a limited number of intermediate species, (σ) precursors, that can occur.

We shall broadly classify all the {P} and find general relations between ρ and σ. This will depend only on the types of OVR, not even on the actual species involved. In fact just the specification of ρ alone is sufficient to generate all the possible types of networks and the possible σ numbers.

The analysis would aid computer design of organic synthesis, both in the empirical "reactions bank" approach[2a] and in the a priori "tree generation" approach. In the latter approach of Ugi and coworkers,[2b] one works backwards from the target (plus a reasonable by-product pool), with bond scissions and makings, to generate a growing downward tree of precursors.

The network analysis and approach of the present paper narrows down the problem by specifying both ends of the synthesis, reagents and target. Then from the broad a priori essentially topological pathway classification, further guidelines are obtained before the computer stage, by concentrating on only some types of desirable pathway classes.

It is interesting that biochemical synthesis pathways are generally of different topological type than the organic synthesis pathways for the same end product (see section X).

Kinetics. All Possible Mechanisms for a Given Overall Reaction Type. Given an overall reaction (OVR) and its observed macroscopic rate law, it has always been a problem to come up with the right mechanism, to know which of several mechanisms that may have been proposed by various workers to use, and, particularly, to know a priori if there are other possible mechanisms, which ones, how many? The

net result is that presently, finding mechanisms involves guessing and trial and error. The same difficulty exists also in the enzymatic mechanisms in biochemistry.

In the present paper we find that specifying just the number, ρ, of mechanistic elementary reactions a finite number of and certain types of mechanisms (networks) are obtained. These types allow only certain numbers of intermediate species, catalysts, etc., possible in a mechanism. For each type one also gets only certain general types (like A → B, A + B → C + D, etc.) of overall reactions OVR, given just the number of steps ρ.

The treatment of this paper classifies and systematizes, in addition to all mechanisms of a given OVR, all mechanisms coming also from different OVR. The inverse problem of giving a specific, actual OVR, to find all its mechanisms, though a more restrictive one is treated in greater detail in other papers.[3]

Once the networks or mechanisms are written down, each leads to an exact rate law expression. These rate laws are in general nonlinear and difficult to integrate; however, some general features of them and their approximate solutions may also be explored with the present network theory.

Finally, given the networks, one may also explore their steady states (whether they exist, how many, etc.) and the stability characteristics of such states. The stability characteristics are important in industry, in chemical reactor control, and in biochemistry in metabolic control and regulation, biological clocks, etc. This last is of course a current, active area of research.[4,5] In one of our subsequent papers we shall use the network theory of this paper in that direction too, obtaining a simple procedure to test the stability of mechanistic steady states.

Given some possible networks, synthetic pathways, or mechanisms, one still has the problem of knowing which paths are energetically more favorable than others. Here the problems are quantum mechanical, specific to the actual species involved. Each pathway and each elementary step corresponds to a quantum mechanical potential energy surface[6] $U(\{R\})$, with $\{R\}$ the relative coordinates of all the atoms in the reactant, intermediate, product, etc., set. It would be costly, tedious, and not particularly useful to calculate point by point on the computer each possible pathway's $U(\{R\})$. Instead one needs qualitative but rigorously founded quantum methods for selecting pathways. Important progress has been made in that direction, though the methods used so far have been confined mostly to symmetry situations,[7] or to special systems (like pericyclic reactions)[8] (see, however, the approach of ref 9 which does not require symmetry).

In the present work we deal only with the topological pathway and connectivity aspects of chemical reactions networks. This delineates properties dependent on types of networks, rather than on specific energetics of each passage. The quantum aspects, the energetic pathway selection aspects, are dealt with in a separate work of this writer where a new algebraic approach for organic molecules has been under development for the past several years.[10]

II. A Chemical Reaction Network. Species Mole Lines, Line Blocks, Skeletons

We first draw a "network" representing the given mechanism or pathway.

How to Draw the Network. Consider, for example, the Henry-Michaelis-Menten mechanism for enzyme kinetics.

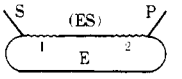
$$\begin{array}{l} \text{(I)}\ \ E + S \longrightarrow (ES) \\ \underline{\text{(II)}\ \ (ES) \longrightarrow E + P} \\ \text{OVR}\ \ S \longrightarrow P \end{array} \qquad \text{(M1)}$$

The overall reaction (OVR) is shown with a boldface arrow (→). The elementary steps are labeled with roman numerals.

Definitions: In drawing a network each mole of each species in each reaction step (i) is shown by a solid line (——). Each reaction (each arrow →) is shown by a wiggly line (∼∼) but with no direction to it (see below).

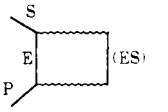
Mechanism M consists of ρ steps $\{1, 2, \ldots, \rho\}$. The network has therefore ρ wiggly lines.

Example M1: The network of M1 is



(N1)

Or, drawn more compactly



(N1′)

Note that in (M1) the *same one* mole of (ES) is involved in steps I and II, so there is only one segment of (ES) solid line in (N1). The same is true of E.

Direction of Reaction Arrows. A network shows mainly how the species mole lines are connected together. Whether an elementary step in a mechanism means only forward reaction, forward and back reactions, or equilibrium does not affect the topology. For a single reaction step

$$A + B \longrightarrow C$$

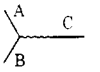
or

$$A + B \longleftarrow C \qquad \text{(M2)}$$

or

$$A + B = C$$

all have the same network



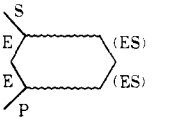
(N2)

However, depending on how the steps are added together the overall reaction (OVR) changes, as do the number of mole lines. The connectivity among species displayed by the network is still unaffected.

Consider again, for example, (M1), but with (II) reversed.

$$\begin{array}{l} \text{(I)}\ \ E + S \longrightarrow (ES) \\ \underline{\text{(II′)}\ \ E + P \longrightarrow (ES)} \\ \text{OVR′}\ \ 2E + S + P \longrightarrow 2(ES) \end{array} \qquad \text{(M1t)}$$

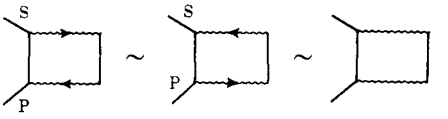
The network now is



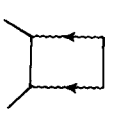
(N1t)

It still shows the same pattern, but E and (ES) have now developed "kinks". A kink shows more than 1 mol of a species is involved in the overall reaction (OVR). In (N1t) we have E⋱E showing the 2E in the OVR.

The display of arrows on wiggly lines in (N1) is unnecessary as

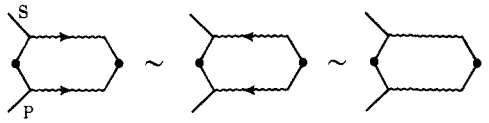


but



(N1′)

would be incorrect as it would imply a network with kinks, i.e.,



(N1t')

The crucial difference between (N1) and (N1t) is, in (N1) there is a circulation around the closed loop, while in (N1t) arrows meet head on and lead to kinks.

Definition: If in a network arrows can be placed on all wiggly lines in such a way that they all lead to full circulations around all loops and free flows along open paths, then the network will be called "laminar". If there is no such arrow assignment possible, i.e., if some arrows meet head on, the network is "turbulent" and depending on the flows some species lines develop kinks.

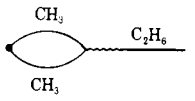
Corollary: In the laminar network's OVR stoichiometric coefficients $|\nu_i|$ are unity.

$$\Sigma \nu_i B_i = 0 \qquad \nu_i \in \{1, -1\} \tag{1}$$

A network is a "turbulent" network if there is more than one mole line of a species in it, as in eq N1t'. Multiple mole lines of the same species in a network can also arise from other causes: (a) if an elementary reaction has some stoichiometric coefficient $|\nu_i| > 1$ and/or (b) if the same species occurs on the same sides ("reactant" side is left of arrow) of more than one elementary step.

Case (a) occurs for example in the chain mechanism decomposition of acetaldehyde, as the step

$$2CH_3 \longrightarrow C_2H_6 \tag{M'3}$$

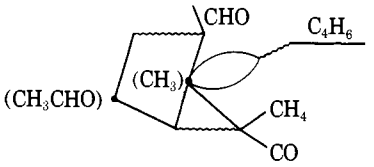


(N'3)

The full mechanism has also a case (b) in it

$$\begin{aligned} &\text{(I)} \quad CH_3CHO \longrightarrow CH_3 + CHO \\ &\text{(II)} \quad CH_3CHO + CH_3 \longrightarrow CH_4 + CO + CH_3 \\ &\text{(III)} \quad 2CH_3 \longrightarrow C_2H_6 \end{aligned} \tag{M3}$$

with the network



(N3)

Note that wherever several mole lines of the same species occur (as in CH_3 in (N3)) they are tied together at *a species vertex.*

The mole lines of a species are tied together because in the entire mechanism and in the rate of the overall reaction there is just one concentration of that species which occurs. All mole lines affect that concentration, hence the rate. Thus the network ought to display this connectivity though mole lines of that species may have been generated in different steps and independently.

Only in the case of a batch, stepwise organic synthesis, where different synthetic steps are carried out in different vessels, the same by-product mole lines coming off in different batch steps would not need to be tied to a species vertex. Each mole line there would act as if it were a distinct species (like different isotopic ones). Thus, for the organic synthesis batch pathways case, we have an option but we may nevertheless wish to stick to species vertices to make the network applicable also to an all in one vessel, e.g., continuous operation synthesis as in an engineering or cellular use of the same steps.

Definition: "Strictly Laminar Network". If the stoichiometric coefficients both in the OVR *and* in each elementary step are all unity, and if only one mole line of a species occurs in the entire mechanism, we shall refer to the network as "strictly laminar"

(M1) is an example of a *strictly laminar network.* (Note again that a *common intermediate* between steps (ES) in (M1) is counted as a single mole line.)

In some of our other papers,[3] a distinction is made between laminar and turbulent networks in developing certain aspects of the theory. The present paper, however, is general, covering both laminar and turbulent networks.

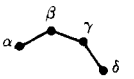
Line Blocks and Skeletons. A network, a mechanism, is a set of ρ relations between pools of species moles. The set of species moles is made up of a number of mutually exclusive subsets. Whenever one or more species are in common between two pools, the two pools belong in the same subset.

If we designate the left (L) and right (R) sides of each reaction step as pools, we have a mechanism which looks like that in Figure 1. The species lines in the network belong in the same pool, if they are joined either at some species vertex or come together at a terminal of a reaction wiggly line.

Definition: A set of species mole lines which are thus joined together is a *line block.*

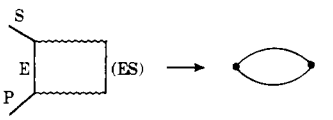
The crucial aspects of a network, whether turbulent or laminar, involve how line blocks (e.g., α, β, γ, δ in Figure 1) are interconnected by the ρ reaction wiggly lines. These features of a network can be shown by drawing a more condensed picture: the *skeleton* (S) of a network (N).

Definition: Each line block of an N is shown as a "dot point" (·) and each wiggly line of N as a solid line (——). Then N gets compressed into a "skeleton". For example, Figure 1 becomes



Other N's given above become:

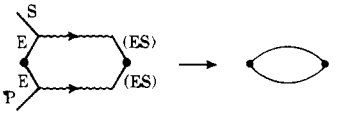
(a)



(S1)

(N1)

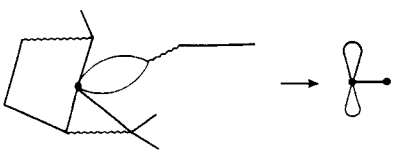
(b)



(S1t)

(N1t)

(c)



(S3)

(N3)

Note that (1) while the network may be rather intricate as in (N3), it leads to a very simple skeletal picture (S3), where reaction loops, etc., are immediately clear. (2) Regardless of the direction of reaction arrows and whether the

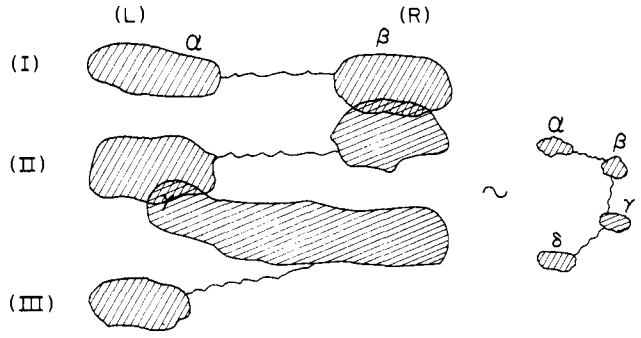


Figure 1. Overlapping of reactant (L) and product ((R) species pools of reaction steps to form a network.

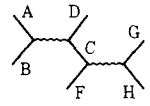
N is laminar or turbulent we get the same skeleton (S1 and S1t). (3) Several quite different networks may lead to the same skeleton, making their inherent similarity in regard to their feedback aspects, presence of catalysts, and the nature of the coupling between the reactions apparent while those may have been obscure in the intricate looking networks.

III. Skeletons of Some Basic Mechanism Types. Nature of Coupling Due to Common Intermediates, Catalysts, Etc.

So as to develop a feeling for what types of mechanisms a given skeleton corresponds to, we examine first some basic mechanism and pathway types.

Consecutive reactions

$$\begin{aligned} &(I)\ A + B \longrightarrow C + D \\ &(II)\ C + F \longrightarrow G + H \end{aligned} \qquad (M4)$$



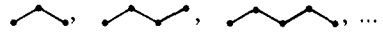
(N4)

(S4)

The same type results from the simpler

$$A \longrightarrow B \longrightarrow C \qquad (M5)$$

Skeletons in the form of linear chains

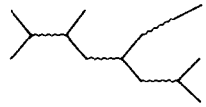


(S5)

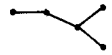
result from simple consecutive reactions.

We get *trees* for combinations and branchings of consecutive reactions, e.g.,

$$\begin{aligned} &(I)\ A + B \longrightarrow C + D \\ &(II)\ C \longrightarrow F + G \\ &(III)\ F \longrightarrow K + L \\ &(IV)\ G \longrightarrow M \end{aligned} \qquad (M6)$$

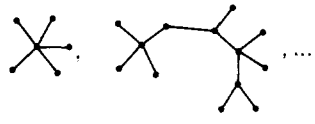


(N6)



(S6)

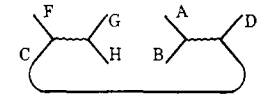
All such trees are made up of linear chains and k stars. A k star is k lines coming out of a dot point.



In laminar networks, e.g., (N6), reactions are coupled through *intermediates*, like C, F, and G.

Note that in (M4) had the order of steps been reversed, we would have

$$\begin{aligned} &(I')\ C + F \longrightarrow G + H \\ &\quad\ A + B \longrightarrow C + D \end{aligned} \qquad (M'4)$$



(N'4)

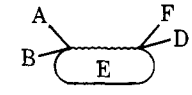
with the skeleton

(S'4)

But this is of course exactly the same as (N4). The C looked like an *intermediate* in (M4). It looks like a "catalyst" in (M'4) in the sense that it is first used up, then regenerated, but it is not. If they occur in the same batch vessel, the order of (I) and (II) is irrelevant. To have a true *catalyst*, the reaction steps must be coupled in such a way as to develop a loop in the skeleton. This means either C is regenerated in the same step, or steps are coupled through with two or more species including C (see below).

Loops, Rings, Catalysts, Chain Reactions. An enzyme or a one-step catalyst leads to a loop in the S; e.g.,

$$(I)\ A + B + E \longrightarrow D + F + E \qquad (M5)$$

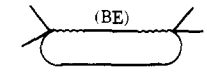


(N5)

(S5)

If we add intermediate species in (M5) before the regeneration of E, we get rings in S, as in the Henry-Michaelis-Menten mechanism, (S1), or in

$$A + B + E \longrightarrow (BE) \longrightarrow D + F + E \qquad (M6)$$

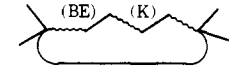


(N6)

(S6)

We can keep on by postulating other intermediates in (M6) and get larger rings; e.g.,

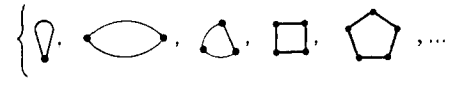
$$A + B + E \longrightarrow (BE) \longrightarrow (K) \longrightarrow D + E + F \qquad (M7)$$



(N7)

(S7)

Thus, essentially the same mechanism can be made to lead to a sequence of rings



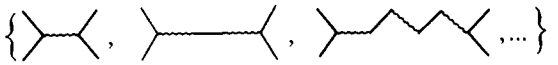
(S7')

by assuming, e.g., shorter and shorter lived "transient states" in sequence. The crucial thing in (S7') is the occurrence of a single ring (by a "ring", it is meant a closed skeletal region with no lines inside; for example,

is a ring, but

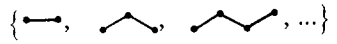
consists of two rings), corresponding to the occurrence of one enzyme.

This expansion of a skeletal region into a sequence can occur in tree portions as well; e.g.,



(N8)

with



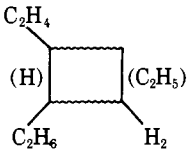
(S8)

To generate a sequence corresponding to basically the same type of mechanism, one inserts more and more dots into a given skeletal line.

Chain reactions also lead to rings. In fact, the action of a chain propagator in a two-step chain is similar to that of a catalyst.

In the original chain reaction

$$\begin{array}{lrcl} \text{(I)} & H + C_2H_4 & \longrightarrow & C_2H_5 \\ \text{(II)} & C_2H_5 + H_2 & \longrightarrow & C_2H_6 + H \\ \hline \text{OVR} & C_2H_4 + H_2 & \rightarrow & C_2H_6 \end{array} \qquad \text{(M9)}$$

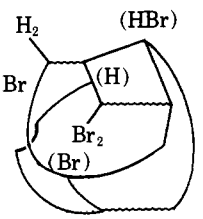


(N9)

(S9)

Multiple rings can arise in multiple step chains, e.g.,

$$\begin{array}{lrcl} \text{(I)} & Br + H_2 & \longrightarrow & HBr + H \\ \text{(II)} & H + Br_2 & \longrightarrow & HBr + Br \\ \text{(III)} & H + Br & \longrightarrow & HBr \end{array} \qquad \text{(M10)}$$



(N10)

(S10)

This example illustrates how it may be difficult to trace through a network, while the topological relationships of reaction steps are very clear in the skeleton.

If chain initiation and/or termination steps are added to the chain mechanism, we may get additional petals (loops) or stems on the flower patterns like (S10).

For example, if we add

$$\text{(IV)}\quad 2Br \longrightarrow Br_2 \qquad \text{(M10′)}$$

(S10) becomes

(S10′)

because we have one more reaction line within the same line block.

If chain termination leads to a product species outside of the chain line block, we get a stem in S, e.g., the C_2H_6 termination of CH_3 radicals in (M3), (N3), (S3).

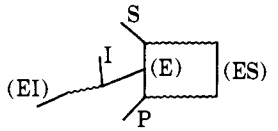
Branched (explosive) chain reactions have similar S's to chain reactions: a flower with one or more stems.

Enzymatic Reactions with Inhibition. We already remarked that simple enzyme mechanism skeletons are similar to chain reactions [see (N9), (S9) vs. (N1), (S1)].

Inhibition steps are similar to chain termination ones and introduce stems.

Competitive inhibition

$$\begin{array}{ll} \text{(I)} & E + S \longrightarrow (ES) \\ \text{(II)} & E + I \longrightarrow (EI) \\ \text{(III)} & (ES) \longrightarrow E + P \end{array} \qquad \text{(M11)}$$

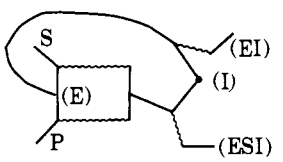


(N11)

(S11)

Noncompetitive inhibition

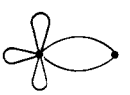
$$\begin{array}{ll} \text{(I)} & E + S \longrightarrow (ES) \\ \text{(II)} & (ES) \longrightarrow E + P \\ \text{(III)} & (ES) + I \longrightarrow (ESI) \\ \text{(IV)} & E + I \longrightarrow (EI) \end{array} \qquad \text{(M12)}$$



(N12)

(S12)

If we add to (M12) a step (ESI) → (EI) + S, the stems in (S12) get tied together. Also, a new petal forms



(S12′)

More examples from organic and biochemical pathways as well as mechanisms will be found in later sections.

We noted that the overall rate law will depend on the skeletal type as will the steady state solutions and their stabilities. While, e.g., (S1) leads to a "saturating" rate law, a tree branch leads to a rate uniformly increasing with reactant concentration.

We turn now to the problem of finding all the possible mechanisms or pathways with the given number ρ of reaction steps and the allowed numbers of species. Clearly the first step in this, and the complementary problem treated by Lee and Sinanoğlu[3] of finding all {N} of a given OVR, is to find all the skeletons of given ρ.

IV. Skeletons Arising from a Given Number ρ of Reaction Steps

For a given $\rho \in \{1, 2, 3, 4, \ldots\}$, networks possible may be many, but the skeletons are few.

We display all the {S} with $\rho = 1, 2, 3$ in Table I, and with $\rho = 4$ in Table II. Recall that each S is for many networks $\{N\}_S$, some laminar, some turbulent.

There are $\text{No.}\{S_\rho\} = 2, 3, 11, 30, \ldots$ skeletons for $\rho = 1, 2, 3, 4, \ldots$.

In drawing the skeletons of $\rho + 1$ it is convenient to gen-

Table I. All Skeletons of $\rho \in \{1, 2, \text{and } 3\}$

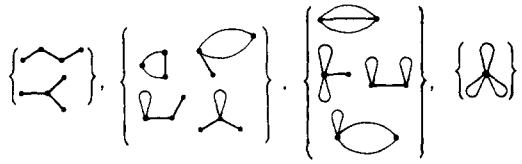
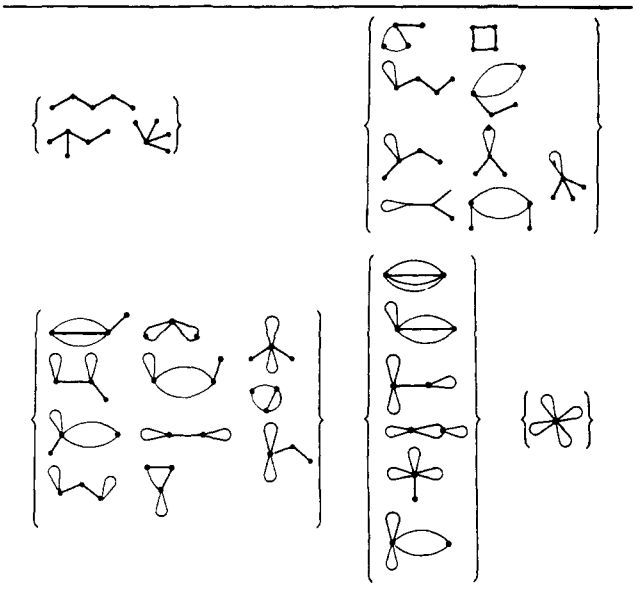
ρ	Skeletons $\{S\}_\rho$
1	
2	
3	

Table II. All Skeletons with $\rho = 4$



erate first all those that can be obtained from those of ρ by adding one more stem, one more loop, etc.

Having all skeletons for a given ρ, we can then expand each S into its networks, first laminar, then turbulent. That procedure will be demonstrated in other papers.[3] Here the writer will concern himself more with the problem of finding the number of species, networks of a given S can have. We shall find strongly restricting relations between the number of species possible in a mechanism or pathway and the number of steps postulated. Clearly this would provide a useful guide in coming up with plausible mechanisms or pathways for an overall synthesis or observed rate law.

V. The Species Count. Internal Species, External Species

Given only the number of steps ρ desired in a mechanism or pathway, there are a finite and quite small number of topological skeletons as shown in Tables I and II. In fact most mechanisms and quite a few organic syntheses involve ρ = 2-6 steps. More steps are fairly rare.

To go from an S to an N, each dot point of S must be expanded into all possible line blocks. The first and very general limitation on the possible number of species in all {N} that would have the same S_ρ comes therefore from the number of dot points, γ, in S_ρ.

To find $\rho(\sigma)$, a step number vs. species number, we have the sequence

$$\rho \longrightarrow \gamma \longrightarrow \sigma \tag{2}$$

i.e., we need to find γ first, given the ρ, and then find the σ. Also we note that σ total is made up of internal species, σ_{int}, and external ones, σ_{ext}.

$$\sigma = \sigma_{\text{int}} + \sigma_{\text{ext}} \tag{3}$$

Definition: Internal species (mole lines) have both ends on wiggly lines in N (reaction line in S). External species have only one end on a wiggly line in N. The other end is either free or on a species vertex.

Examples: (1) In (M9), internal species (int) = {H, C_2H_5}; external species (ext) = {C_2H_4, C_2H_6, H_2}, σ = 5 = 2 + 3.

(2) In (M10), int = {none}, ext = {H, Br, HBr, H_2, Br_2}, $\sigma = \sigma_{\text{ext}}$ = 5. Note that if we exclude step III, we get int = {H, Br}; ext = {H_2, Br_2, HBr}.

(3) In (N1) (laminar case), int = {(ES), E}, ext = {S, P}, σ_{total} = 4, but in (N1t) (turbulent version of (N1)), we get int = {none}, ext = {S, P, E, (ES)}, though σ_{total} = 4.

Remark 1: External species appear in the OVR, while internal ones do not. Enzymes, intermediates, and fully used up precursors are internal species.

Remark 2: Whether it be in a mechanism or in a synthesis, one would like to keep the number of steps, ρ, and/or internal species (and by-products), hence σ, to a minimum, for economy. In designing syntheses or in postulating mechanisms it is important to know therefore the minimum and maximum possible σ's (and σ_{int} and σ_{exl}'s) given the ρ.

VI. Number of Species Line Blocks, γ, Given the Number of Reaction Steps, ρ

Line blocks in N are dot points in S. Reaction lines in S have dot-point ends. Given the ρ lines, there is only a certain number, γ, of dot points possible in any skeleton S_ρ. The γ is given by the following general result obtained by this writer.

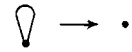
Theorem 1: In a skeleton S made of ρ lines, the number of dot points is

$$\gamma = \rho - r + 1 \tag{4}$$

where r = the number of rings in S.

This theorem is similar to Euler's theorem relating the numbers of faces, corners, and edges of polyhedra (see, e.g., the book on elementary topology[11a] and also ref 11b). Although more rigorous and elegant proofs may exist in mathematics, we give here a simple proof in the form of several basic steps which give some insight into the skeletons.

(a) Consider any S stripped first of its loops, i.e., put



and multiple lines, i.e., put

(b) Then any *stripped S* is made of only trees and larger empty rings (i.e., rings of three or more dots). (c) Any tree is made up of linear segments

and k stars

(d) A linear segment has $\gamma = \rho + 1$. Any k star has $\gamma = k + 1$. (e) Everytime we join a line segment and a star at a dot, we loose one dot, so we still get $\gamma = \rho + 1$. (f) When a segment of l lines ($\gamma = l + 1$) is turned into a single ring, one dot point is lost, hence $\gamma \rightarrow l$. Thus for r separate rings in S, r dots are lost. (g) Next one considers how trees and rings and rings and other rings are joined together (e.g.,

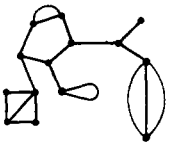
making a ring by adding a ρ_i segment's two ends to a ring, we loose two dots and gain one ring and ρ_i lines). With all such junctions, the result up through (f) remains intact, namely $\gamma' = \rho' - r' + 1$. (h) Adding now the loops and multiple lines back on, ρ is increased by one for each such addition, but so is r. Thus γ remains $\gamma = \rho - r + 1$. (Q.E.D.)

Examples:

(1) (S3): $\rho = 3$, $r = 2$, $\gamma = 2$

(2) Flower: $\rho = 5$, $r = 5$, $\gamma = 1$

(3) $\rho = 20$, $r = 7$, $\gamma = 14$



Remark: It is interesting to note right away that the more rings and loops for given ρ, the less the γ, hence the smaller the σ. We will note below that in turn, the more enzymes, the more rings.

Classification of Skeletons. The $\gamma(\rho)$ eq 4 allows us to classify the skeletons.

With fixed ρ, each r defines the r-ringed class of skeletons $S_{(\rho,r)}$.

All skeletons in the same r-ring class have the same γ, but not all skeletons with the same γ are in the same r-ring class.

Corollary to Theorem 1: For given ρ, there are $\rho + 1$ classes (ring number classes)

$$r \in \{0, 1, 2, 3, \ldots, \rho\}$$

with

$$\gamma \in \{\rho + 1, \rho, \rho - 1, \rho - 2, \ldots, 1\} \tag{5}$$

Examples: The $\rho = 4$ skeletons of Table II result in Table III.

We note that the number of skeletons vs. r goes through a maximum, the most possibilities being around the midpoint, i.e., $r \approx \rho/2$.

On the other hand, the least number No.$\{S_\rho\}$ is for $r = p$, i.e., the flower pattern. We will see a special significance of this case for biochemical pathways (see section X).

It is also interesting to classify the *skeletal sequences*, e.g., (S8) and (S′7), which had special mechanistic significance.

An *r-ring skeletal sequence* is generated, one ring at a time, by enlarging just that ring, giving a subsequence within the fixed r increasing ρ sequence. For example, (S8) is a no star, $r = 0$ subsequence

$$\{r = 0; \rho = 1, 2, 3, \ldots\}$$

Other subsequences within ($r = 0$, $\{\rho\}$) would be obtained specifying numbers of k stars and k's. Similarly for (S′7), the $\{r = 1; \rho = 1, 2, 3, \ldots\}$ subsequence.

VII. Possible Numbers of Species in a Network Given the Number of Reaction Steps

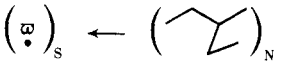
For any network we found the possible numbers of line blocks given the ρ. (The $\rho \rightarrow \gamma$ problem.)

To find the number of species mole lines and of species, σ, given the ρ, we must find the number of species $\bar{\omega}$ in each line block.

A dot point of S may represent different size line blocks.

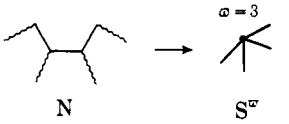
Definition: The *weight* $\bar{\omega}$ of a dot point of S is the number of species in the corresponding line block of N.

We have when N → S

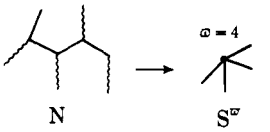


(6)

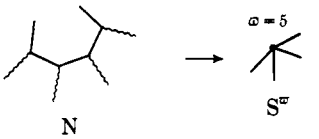
for example



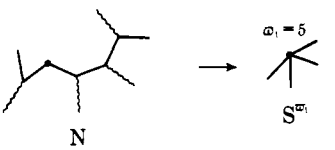
(7a)



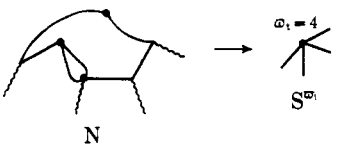
(7b)



(7c)



(7d)



(7e)

All the {N} in eq 7 have the same S =

The strictly laminar cases (7a-c) have different weights $\bar{\omega}$. Cases in (7c and 7d) have the same weight, though their N's are different, one laminar, the other turbulent.

We have the many-to-one mappings

$$\{\{N\}\} \longrightarrow \{S^{\bar{\omega}}\} \longrightarrow S \tag{8}$$

with $S^{\bar{\omega}}$ = weighted skeleton.

The weights in an $S^{\bar{\omega}}$ add up to the total species count σ

$$\sum_{i=1}^{\gamma} \bar{\omega}_i = \sigma \tag{9}$$

The least number of species mole lines occurs for some laminar N, but not the least number of species.

There are many more possible types of line blocks to consider in the turbulent case. However, though the species

Table III. The Four-Step Skeletons by Ring Number Class

$\rho = 4$		
r	γ	No. of skeletons No. $\{S_\rho\}$
0	5	3
1	4	9
2	3	11
3	2	6
4	1	1

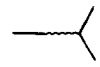
mole line number can then increase quite a bit, the species count remains in the same range as laminar.

Given the number of steps ρ, we now find the maximum, σ_{max}, and the minimum, σ_{min}, number of species possible. These will be obtained from the laminar case. We shall then show that if the network is made turbulent (same skeleton) in any way possible, σ_{max} is still the maximum, but σ_{min} is lowered. Also more species can become external.

More important than turbulence for the species count problem is the types of elementary reaction steps allowed.

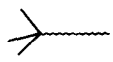
Types of Elementary Steps. In a mechanism, elementary steps correspond to molecular processes. The steps therefore are uni- or bimolecular.

$$\left.\begin{array}{l} A \longrightarrow B + C \\ B + C \longrightarrow A \end{array}\right\} \qquad \text{(M)} \qquad \text{(N)} \qquad (10a)$$



$$A + B \longrightarrow C + D \qquad \text{(M)} \qquad \text{(N)} \qquad (10b)$$

Three or more body collisions



are unlikely.

In organic synthetic pathways, known reactions are used. There are 20-25 such reactions most commonly occurring. A list of syntheses reveals steps by and large of the eq 10a and 10b type. In fewer cases, steps of type

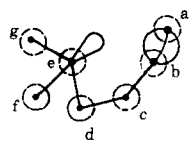
$$A + B \longrightarrow D + E + F \qquad (11)$$

are seen.

We will develop the theory for all steps being of the types in eq 10a and 10b. Each wiggly line end has one or two species mole lines, no more (the "step complexity index" $\equiv n = 2$). This covers almost all that is of practical interest. (The writer extended the work to $n = 3$, but omits it. Added tedium is far more than added benefit.) For organic pathways it is simpler to break up such $n = 3$ or more steps into two or more $n = 2$ steps using an actual (mechanistic) or hypothetical intermediate as in

$$\begin{array}{l} n = 2 \\ A + B \longrightarrow D + X \\ X \longrightarrow E + F \\ \text{combined step } (n = 3) \\ A + B \longrightarrow D + E + F \\ \qquad n = 3 \end{array} \qquad (11')$$

Weight of the k Stars of a Skeleton. Consider a skeleton S, e.g.,



(12)

with the dot points labeled for convenience.

If we draw a tiny circle around each dot point, we find the k value of each k star; for example, in (12)

$$k_a = 3,\ k_b = 4,\ k_c = 2,\ k_d = 2,\ k_e = 5,\ k_f = 1,\ k_g = 1 \qquad (13)$$

For convenience of presentation we state this as a lemma: the sum of the star k values of a skeleton S is twice the number of steps ρ, i.e.

$$\sum_{i=1}^{\gamma} k_i = 2\rho \qquad (14)$$

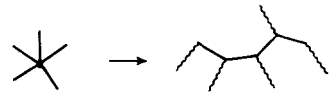
Proof: Stars of S are made by cutting each line in two. Counting star lines we double count the lines.

Theorem 2: The possible weights $\{\omega_k\}$ of a k star dot point in S are

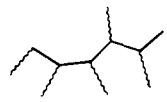
$$\bar{\omega}_k = k - 1, k, k, k + 1 \text{ for } k \geq 2 \quad (\text{for } k = 1,\ \bar{\omega}_1 = 1,\ 2 \text{ only}) \qquad (15)$$

if the dot point represents laminar line blocks ("turbulence index" $\equiv 0$) or turbulent line blocks with turbulence caused by arrow directions only (turbulence index $\equiv 2$).

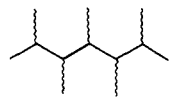
Proof: The only types of laminar line blocks possible for a k star are



$(k\text{-star})_S$ $\qquad \omega_k = k - 1$ $\qquad (16a)$

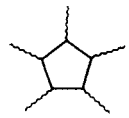


$\omega_k = k$ $\qquad (16b)$



$\omega_k = k + 1$ $\qquad (16c)$

and



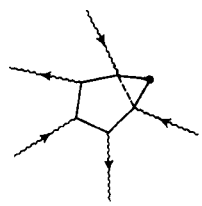
$\omega_k = k\,(k = \text{even only})$ $\qquad (16d)$

(see "note" to theorem 2).

The linear line blocks segments with zero (mechanism 16a), with one (mechanism 16b), and two (mechanism 16c) external species lines, and a species ringed line block (k star giving k-membered species ring) (structure in 16d) are the only possibilities. (Q.E.D.)

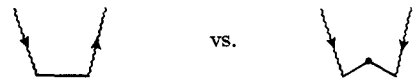
For one stars $k - 1 = 0$ is not possible, since all wiggly lines should end on a species line. (—〰—) Thus for one stars $\bar{\omega}_1 \in \{1, 2\}$.

Note 1. Putting possible reaction arrows on the wiggly lines we see that, for (16d), k = odd, an odd-membered species ring cannot be laminar, but will at best have one turbulent species kink, as at least two arrows will meet. But this does not change the number of species $\bar{\omega}_k = k$.



(16d′)

Note 2. Turbulence Index $\equiv$ 2 Cases. This means species vertices with two mole lines exist. Such are caused by two arrows meeting.

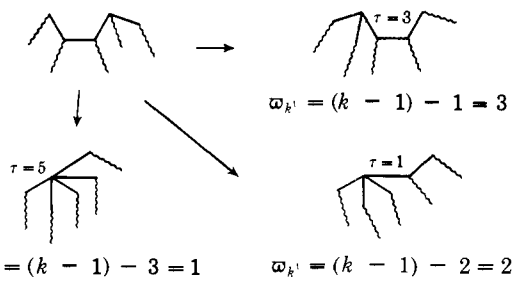


vs. (17)

(turbulence index $\equiv \tau = 0$) (turbulence index $\equiv \tau = 2$)

These arrow turbulences do not change the weights in eq 16; so eq 15 is for 2 turbulence as well as laminar. (Q.E.D.)

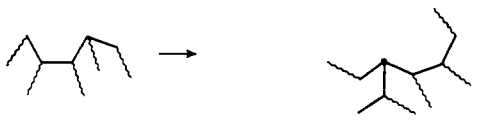
Weights for Turbulence Index > 2. Different type line blocks arise if more than two species lines meet in a species vertex. A skeletal k star may represent then, e.g.,



$\varpi_k{}^t = (k - 1) - 1 = 3$ (16t)

$\varpi_k{}^t = (k - 1) - 3 = 1$ $\varpi_k{}^t = (k - 1) - 2 = 2$

Each of (13t) can have more external species added to bring the total back up to $(k - 1)$; e.g., by



(16t′)

$$\varpi_k{}^t = (k - 1) - 1 + 1 = (k - 1)$$

Formation of a species vertex has decreased the number of distinct species by one, but addition of a new external species line has increased it back up by one. Thus we get an *addendum to theorem 2.* For a k star dot point with $\tau > 2$ turbulence allowed, the weights can be

$$\bar{\omega}_k{}^t \in \{1, 2, 3, \ldots, k - 1, k, k + 1\} \quad (k \geq 2) \quad (18)$$

Thus the maximum $\bar{\omega}_k$ has remained at $k + 1$ in the most general network, but the minimum $\bar{\omega}_k$ can be as low as 1 (this for $\tau = k$).

For one stars again $\bar{\omega}_1 \in \{1, 2\}$ in the general turbulent case.

Maximum Number of Species Possible. From eq 9 and 15

$$\sigma_{max} = \sum_S \omega_k{}^{max} = \sum_S (k + 1) \quad (19)$$

The sum is over all dot points of S. (Σ_S means the sum of the value that follows Σ for each dot point.) Thus one part of the sum $\Sigma_S(k + 1)$ becomes

$$\sum_S (1) = \gamma \quad (20)$$

since this term amounts to counting the dot points of the skeleton S. Using also eq 14

$$\sigma_{max} = 2\rho + \gamma \quad (21)$$

or using $\gamma = \rho - r + 1$

$$\sigma_{max} = 3\rho - r + 1 \quad (21')$$

Equation 21 applies to all networks (laminar; turbulent; $\tau = 0, 2, 3, \ldots$).

Minimum Number of Species. For laminar and their isomorphic

$\tau = 2$ turbulent networks (most common case), from theorem 2

$$\sigma_{min} = \sum_{\substack{S \\ \text{(all dots including } k = 1\text{)}}} (k - 1) + \sum_{(k=1)} {}_*(1) = 2\rho - \gamma + f \quad (22)$$

or

$$\{\sigma_{min} = \rho + r - 1 + f \geq (\rho + r - 1)\} \text{ for } \tau \in \{0, 2\} \quad (22')$$

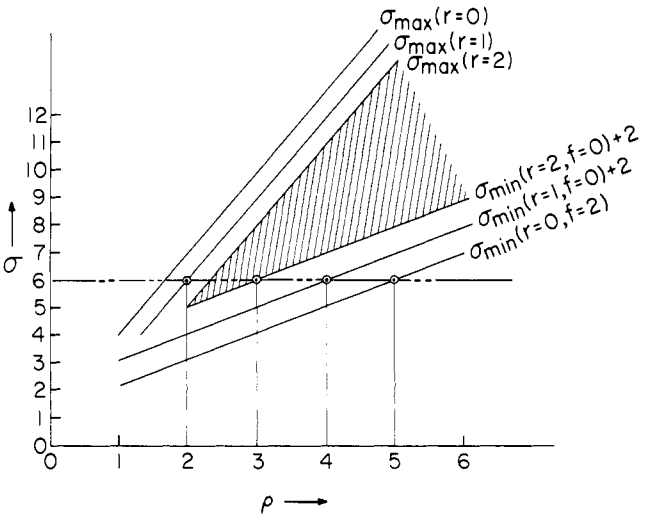


Figure 2. The σ-ρ, $\rho(\sigma)$, plot. The plot gives the range of the number of species possible in a mechanism of ρ reaction steps. Conversely, given the number of chemical species σ, one can read off the number of steps any possible mechanism would have. These ranges show the dependence also on the number of catalysts or feedback loops (number of rings r) in the mechanism. For a given ρ or σ only the intersections with σ_{max}, σ_{min} lines which yield integers are the solutions (for the $\sigma = 6$ case shown circled).

where $f \equiv \Sigma_{(k=1)*}(1)$ is the number of one stars, i.e., free ends in the skeleton. For any S with ρ we have $f \in \{0, 1, 2, \ldots, \rho\}$.

For the maximal turbulence case, from the addendum to theorem 2, eq 17

$$\sigma_{min}{}^t(\tau = k_{max} > 2) = \sum_S (1) = \gamma$$

or

$$\sigma_{min}{}^t(\tau > 2) = \rho - r + 1 \quad (22t)$$

If we have a tree S, i.e., $r = 0$ (as in most organic syntheses), or $r = 1$, eq 22′ and 22t give the same or very close σ_{min}. As $\tau > 2$ cases are rare, we shall take eq 22′ as an indication of σ_{min} to be expected for an S in general.

The Range of σ and Average σ. From eq 22 and 21, the average $\bar{\sigma}(\rho)$ is (for $f \approx 0$)

$$\bar{\sigma} = 2\rho \quad (23)$$

With the range $(\sigma_{max} - \sigma_{min}) = \Delta\sigma$ (for f small)

$$\Delta\sigma = 2\gamma$$
$$\Delta\sigma = 2(\rho - r + 1) \quad (24)$$

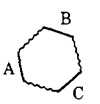
In the biochemically important case of $\rho \approx r$ (cf. last section, around theorem 4) we have

$$\sigma(\rho \approx r) \in \{2\rho - 1, 2\rho, 2\rho + 1\} \approx 2\rho \quad (25)$$

So, in that case, we predict from the number of reaction steps the number of species involved almost with certainty!

The Species-Reaction Step Plot. Figure 2 shows a $\sigma(\rho)$ plot. For each skeletal r-ring class [given (ρ, r)] there is a $(\sigma_{min}, \sigma_{max})$ range. For each r-ring skeletal sequence (given r), there is a triangular region (e.g., shaded area for $r = 2$) of possible (σ, ρ) values.

In the plot, for $r = 1, 2, \ldots$, we added a 2 to σ_{min}, so as to eliminate "unproductive networks" like

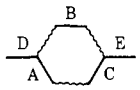


(26a)

(A ⟶ B ⟶ C ⟶ A)

OVR 0

With 2 added to such σ_{min}, we start with instead, e.g.,



$$\begin{array}{l} A + D \longrightarrow B \\ B \longrightarrow F + C \\ C \longrightarrow A \\ \hline \text{OVR } D \rightarrow A \end{array} \qquad (26b)$$

The σ_{max} lines in the plot apply to all networks, the σ_{min} lines to laminar and $\tau = 2$ turbulent. For $\tau > 2$ turbulent, all σ below the σ_{max} are possible.

VIII. Reaction Steps Possible Given the Number of Species

The inverse question is of much practical interest. One may have an idea of the number of species, σ, to be considered (number of species in the OVR plus intermediates, catalysts, precursors). Then, how many reaction steps are possible in any mechanism or synthesis; i.e., $\rho(\sigma)$ = what?

The possible ρ are read off from Figure 2 along a σ line. For example, in the figure, for $\sigma = 6$, we get $\rho = 2, 3, 4, 5$ if $r = 0$; $\rho = 2, 3, 4$ if $r = 1$, and $\rho = 3$ if $r = 2$, and that is all (for $\tau \leq 2$). We can narrow down the possible ρ more if we know how many species out of the $\sigma = 6$ are catalysts, or if we know how many catalytic or chain reaction loops would be reasonable. These numbers, essentially measured by r, can only be, we see also from Figure 1 for $\rho = 6$, $r \in \{0, 1, 2\}$. If we require $r = 2$, then $\rho = 3$ is the only possibility.

Knowing the number of steps (and r and σ) it will surely be much easier to guess the mechanism or synthetic pathway for the stoichiometry (and OVR) at hand.

In fact, the search for possible mechanisms is narrowed even further when we look at the possible numbers of internal and external species within a fixed $\sigma = \sigma_{int} + \sigma_{ext}$.

IX. Possible Numbers of Internal and External Species

Going through the derivations of the $\sigma(\rho)$ results above, we see that, in

$$\sigma = \sigma_{int} + \sigma_{ext} \qquad (27)$$

for laminar networks, in fact,

$$\sigma_{min} = \sigma_{min,\,int} = 2\rho - \gamma \qquad (28)$$

$$\sigma_{max,\,ext} = 2\gamma \qquad (29)$$

Further, still with no turbulence $\tau = 0$)

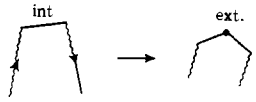
$$2\rho - \gamma \leq \sigma_{int} \leq 2\rho - \gamma_{odd} \qquad (30a)$$

$$f + 0 \leq \sigma_{ext} \leq 2\gamma \qquad (30b)$$

where $\gamma = \gamma_{odd} + \gamma_{even}$, γ_{odd} = number of dot points with odd k stars in the skeleton. Also a σ is made up of only certain combinations (subject to $\sigma_{int} + \sigma_{ext} \leq 2\rho + \gamma$) of σ_{int} and σ_{ext} within their ranges ($\sigma_{ext} = 2\gamma - 2$ goes with $\sigma_{int} = 2\rho - \gamma + 1$ and correspondingly for the others).

Equations 28 and 29 follow easily from eq 16. The γ_{odd} occurs in eq 30a because only for even stars can we get species rings ($\omega_k = k$ case, eq 6) still retaining laminarity. The odd k stars cannot lead to laminar species rings with any arrow arrangement. Thus ($2\rho + \gamma_{even}$) is the most for internal. (Q.E.D.)

With turbulence, internals decrease, externals increase. For example, for each



(31)

we get $\sigma_{int} \rightarrow \sigma_{int} - 1$ while $\sigma_{ext} \rightarrow \sigma_{ext} + 1$. So

$$\sigma_{int}{}^{t} \in \{0,\ 1,\ 2,\ \ldots,\ 2\rho - \gamma\} \qquad (32a)$$

$$\sigma_{ext}{}^{t} \in \{f,\ f + 1,\ \ldots,\ 2\rho + \gamma\} \qquad (32b)$$

such that in each (σ_{int}, σ_{ext}) combination

$$\gamma \leq \sigma_{int} + \sigma_{ext} = \sigma \leq 2\rho + \gamma \qquad (32c)$$

With these σ_{int}, σ_{ext} ranges and Figure 2, possible types of ρ step or σ species networks are much narrowed down. If in addition to know the OVR type (like A + B + C → E + F, etc.) we know the σ_{ext} (all externals appear in the OVR). Then $\sigma \geq \sigma_{ext}$. Choosing such a σ, we get the ρ range. Or if we select a ρ, we have to see from Figure 2 and eq 32, if it is compatible with σ_{ext}, i.e., does σ_{ext} fall in the possible $\Delta\sigma$ and $\Delta\sigma_{ext}$ ranges for this ρ.

Possible Overall Reaction Types Given ρ or σ. Deducing the possible σ_{ext} from (ρ, r, σ) we have also the possible types of OVR. If σ_{ext} is in the laminar range, the stoichiometric coefficients in the OVR are unity. Thus, e.g., for σ_{ext} ($\tau = 0$) = 5, the only possibilities are

$$A \longrightarrow B + C + D + E$$

or

$$A + B \longrightarrow C + D + E \qquad (33)$$

(reverses are the same).

Writing $\sigma_{ext} = \sigma_{ext}{}^{L} + \sigma_{ext}{}^{R}$, OVR "reactants" on the left, "products" on the right, all laminar OVR's are obtained from the partitions of the integer σ_{ext} (e.g., 5 = 4 + 1, 2 + 3). The number No. of such OVR is

$$\text{No. \{OVR\}} = \frac{\sigma_{ext}}{2} \text{ (for } \sigma_{ext} = \text{even)}$$

$$\text{No. \{OVR\}} = \frac{\sigma_{ext} - 1}{2} \text{(for } \sigma_{ext} = \text{odd)} \qquad (34)$$

The same applies for the turbulent, but the σ_{ext} given must be in the turbulent $\Delta\sigma_{ext}$ range and for the desired turbulence index τ. For example, $\tau = 3$ means, the *highest* stoichiometric coefficients occurring will be 3. Then in addition to eq 33 we will have OVR's like

$$A + 2B \longrightarrow 3C + D + E \qquad (33t)$$

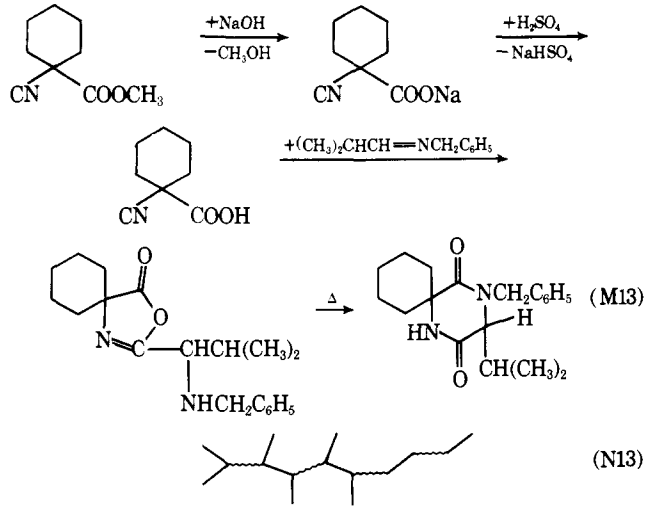
etc.

Once actual chemical species are assigned, the possible stoichiometric coefficients are of course very much narrowed down or determined.

Clearly now one can also work backward from the OVR to σ_{ext}, to $\{\sigma\}$, to $\{(\rho,r)\}$, to all skeletons, to all possible networks, finding thus all possible mechanisms or pathways.

X. Some Examples and Remarks on Organic Vs. Biochemical Synthetic Pathways

Example 1: (example supplied by Dieter Marquading) (organic pathway)



(M13)

(N13)

(S13)

Here all $\nu_i = 1$, a strictly laminar N, with

$$\rho = 4$$
$$r = 0;\ f = 2$$
$$\gamma = \rho - r + 1 = 5;\ \gamma_{odd} = 2$$

From eq 30 we expect

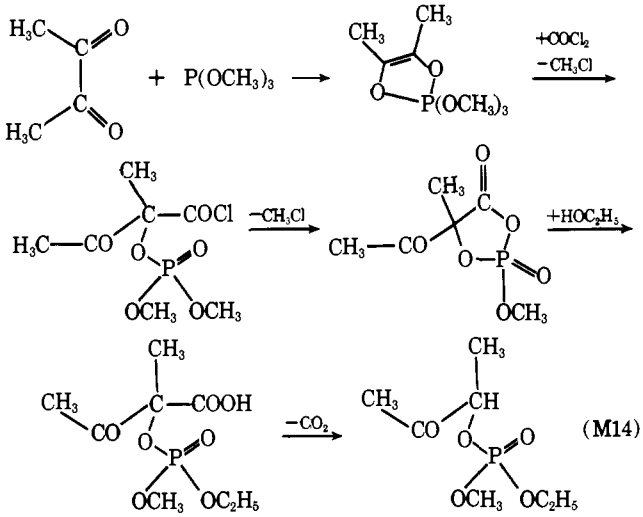
$$3 \leq \sigma_{int} \leq 6$$

subject to $5 \leq \sigma \leq 13$

$$2 \leq \sigma_{ext} \leq 10$$

In the case at hand, we have $\sigma_{int} = 3 = \sigma_{min,int}$ and $\sigma_{ext} = 7$ which check.

Example 2: (example supplied by Ivar Ugi) (organic pathway)



(M14)

(N14)

(S14)

Here

$$\rho = 5$$
$$r = 1;\ f = 2$$
$$\gamma = \rho - r + 1 = 5;\ \gamma_{odd} = 2$$

From eq 32 we expect

$$5 \leq \sigma \leq 15$$
$$\sigma_{int} \in \{0 \text{ to } 5 \text{ to } 8\}$$
$$\sigma_{ext} \in \{2 \text{ to } 10 \text{ to } 15\}$$

Here $\sigma_{int} = 4$; $\sigma_{ext} = 7$; $\sigma = 11$ which check.

Example 3: (biochemical pathway). Electron transport chain

(I) metabolite + DPN $\longrightarrow$ oxidized metabolite + $DPNH_2$

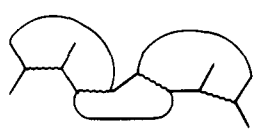
(II) $DPNH_2$ + FP $\longrightarrow$ DPN + FPH_2

(III) FPH_2 + cyt *c* Fe^{3+} $\longrightarrow$ FP + cyt *c* Fe^{2+}

(IV) cyt *c* Fe^{2+} + O $\longrightarrow$ cyt *c* Fe^{3+} + H_2O

OVR metabolite + O $\rightarrow$ oxidized metabolite + H_2O

(M15)



(N15)

(S15)

where we get

$$\rho = 4;\ r = 3;\ f = 0$$
$$\gamma = \rho - r + 1 = 2;\ \gamma_{odd} = 0$$

The case being laminar, we expect

$$6 \leq \sigma_{int} \leq 8$$

such that $6 \leq \sigma \leq 10$

$$0 \leq \sigma_{ext} \leq 4$$

In fact we have $\sigma_{int} = 6 = \sigma_{min,int}$; $\sigma_{ext} = 4 = \sigma_{max,ext}$; $\sigma = 10 = \sigma_{max}$.

Example 4: (biochemical pathway). Oxidative phosphorylation

(I) X-OH + inorg phosphate $\longrightarrow$ X-O-Ⓟ + H_2O

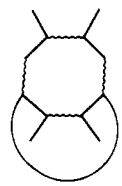
(II) X-O-Ⓟ $\longrightarrow$ X^{2+}-O ~ Ⓟ + $(2e^-)$

(III) X^{2+}-O ~ Ⓟ + ADP $\longrightarrow$ X^{2+}-OH + ATP

(IV) X^{2+}-OH + $(2e^-)$ $\longrightarrow$ X-OH

OVR inorg phosphate + ADP $\rightarrow$ ATP + H_2O

(M16)



(N16)

(S16)

with

$$\rho = 4, r = 2$$
$$\gamma = \rho - r + 1 = 3;\ \gamma_{odd} = 0;\ f = 0$$

The case is again laminar, so from eq 29

$$5 \leq \sigma_{int} \leq 8$$

subject to $5 \leq \sigma \leq 11$

$$0 \leq \sigma_{ext} \leq 6$$

which check with $\sigma_{int} = 5 = \sigma_{min,int}$, $\sigma_{ext} = 4$, and $\sigma = 9$. (Figure 2 shows $\sigma \in \{7, 8, 9, 10, 11\}$ for $\rho = 4$ and $r = 2$).

Many more turbulent or laminar examples can be given. One thing these show is the following.

$$\gamma_{org\ syn} \approx \rho + 1 \qquad (35)$$

Remark 1: Organic syntheses have almost always tree skeletons (no rings or very few, $r \approx 0$)

Moreover, in a laboratory synthesis, we want and have mostly linear tree branches, not stars

as $k > 2$ stars

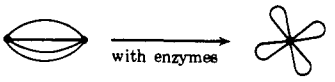
etc., would mean more products coming off, decreasing the yield of the target.

The straight chain tree branch aspect of laboratory synthesis is due also to the batchwise operation. Syntheses are carried out in a succession of batch stages.

In industrial syntheses, in the continuous operation plant, we would have some products recycled, more catalysts used,

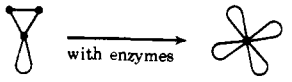
etc., getting more rings and loops in the skeleton. That is, increase of yield, continuous operation, and regulation and control of steady states requires pathway with rings rather than trees.

Remark 2: In fact, nature has designed the biochemical machinery such that the pathway skeletons have often the maximum possible number of rings. Examples above already show this contrast between biochemical and laboratory organic pathways, even though we omitted the enzymes. If to some or all of the elementary steps in the biochemical cases we add the enzymes, we get the maximal ring flower patterns, e.g., (M15) becomes (with n_{max}, reaction step index going up to 3)



(S_E15)

The example (M16) becomes

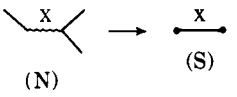


(S_E16)

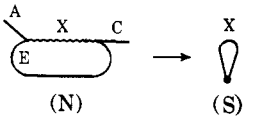
This is general, and we state it as a theorem.

Theorem 4 (Theorem of Enzymes): Whenever an enzyme is added to a reaction step, that reaction line in the skeleton becomes a loop (sufficient, but not necessary) (this is not the only way a loop can arise of course).

Proof: Consider a wiggly reaction line



e.g., A → C. With enzyme E added, A + E → C + E, and



Corollary: As biochemical pathways have enzymes on most of their steps, their skeletons are almost always flower patterns.

Proof: $\gamma = \rho - r + 1$; if each step has an enzyme, by theorem 4, reaction lines become loops, therefore $\rho = r$. Then $\gamma = 1$ and we get a flower with ρ petals

In later papers, we hope to explore the rate law features, steady state, and stability aspects of skeletal patterns.

Acknowledgment. This work was supported by the U.S. National Science Foundation. Its last stages were completed under an Alexander von Humboldt Senior Scientist Award of the German Federal Republic. It is a pleasure to thank the Humboldt Foundation and Dr. Pfeiffer and Dr. H. Hanle for their hospitality in Germany and Professor Ivar Ugi at the Organische Chemie Institut, Technische Universität, München. It is a pleasure also to thank L. S. Lee of our laboratory, Professor I. Ugi, and Drs. D. Marquading, P. Gillespie, and H. Gasteiger of Münich for stimulating discussions. This work was first presented at a lecture at the Chemistry Faculty, University of Istanbul, then at the Fifth Canadian Symposium on Theoretical Chemistry in the C.A. Coulson Memorial Session, June 1974, Ottawa, Canada. The author wishes to dedicate this article to the memory of the late Professor C.A. Coulson.

References and Notes

(1) Alexander von Humboldt Senior Scientist Awardee, 1973-1974, West Germany. Also Danişman, Boğazici, Evrenkentl, Istanbul, Türkiye.
(2) (a) E. J. Corey, *Pure Appl. Chem.*, **14,** 19 (1967); E. J. Corey and W. T. Wipke, *Science*, **166,** 178 (1969), and later papers; (b) I. Ugi, *Intra-Sci. Chem. Rep.*, **5,** 229 (1971); I. Ugi and P. Gillespie, *Angew. Chem., Int. Ed. Engl.*, **10,** 915 (1971), and other papers.
(3) L. S. Lee and O. Sinanoğlu, to be published.
(4) (a) For example, B. Chance, R. W. Estabrook, and A. Ghosch, *Proc. Natl. Acad. Sci. U.S.A.*, **51,** 1244 (1964); (b) J. Higgins, *ibid.*, **51,** 989 (1964).
(5) (a) I. Prigogine, G. Nicolis, and A. Babloyantz, *Phys. Today*, **25,** 38 (1972); (b) P. Glansdorff and I. Prigogine, "Structure, Stability and Fluctuations", Wiley-Interscience, New York, N.Y., 1971.
(6) For a brief survey of the modern theoretical and experimental developments cf.: O. Sinanoğlu, *Comments At. Mol. Phys.*, **3,** 53 (1971); also, O. Sinanoğlu, *J. Mol. Struct.*, **19A,** 81 (1973).
(7) R. B. Woodward and R. Hoffmann, "The Conservation of Orbital Symmetry", Academic Press, New York, N.Y., 1970.
(8) (a) W. J. van der Hart, J. C. Mulder, and L. J. Oosterhoff, *J. Am. Chem. Soc.*, **94,** 5724 (1972); (b) M. J. S. Dewar, *Angew. Chem., Int. Ed. Engl.*, **10,** 761 (1971); (c) H. E. Zimmerman, *Acc. Chem. Res.*, **4,** 272 (1971); (d) R. G. Pearson, *Acc. Chem. Res.*, **4,** 152 (1971); (e) cf. also K. Fukui, *Top. Current Chem.*, **15,** 1 (1970), and others.
(9) C. Trindle and O. Sinanoğlu, *J. Am. Chem. Soc.*, **91,** 4054 (1969).
(10) O. Sinanoğlu, a different series of papers to be published.
(11) (a) B. H. Arnold, "Intuitive Concepts in Elementary Topology", Prentice-Hall, Englewood Cliffs, N.J., 1962, p 77; (b) F. Harary, "Graph Theory", Addison-Wesley, Reading, Mass., 1969.