SYMBOLIC AND QUANTITATIVE REASONING: DESIGN OF REACTION PATHWAYS THROUGH RECURSIVE SATISFACTION OF CONSTRAINTS

Michael L. Mavrovouniotis

Department of Chemical Engineering
Northwestern University
Evanston, IL 60208.

Given a fixed, predetermined set of elementary reactions, compose reaction pathways (mechanisms) that satisfy given specifications in the transformation of available raw materials to desired products. This is a problem encountered quite frequently during research and development of chemical and biochemical processes. As in the assembly of a puzzle, the pieces (available reaction steps) must fit with each other (i.e., satisfy a set of constraints imposed by the precursor and successor reactions) and conform with the size and shape of the board (i.e., the specifications on the overall transformation of raw materials to products). This chapter draws from *symbolic and quantitative reasoning* ideas of AI which allow the systematic synthesis of artifacts through a *recursive satisfaction of constraints* imposed on the artifact as a whole and on its components. The artifacts in this chapter are mechanisms of catalytic reactions and pathways of biochemical transformations. The former require the construction of *direct* mechanisms, without cycles or redundancies, to determine the basic legitimate chemical transformations in a reacting system. The latter are the chemical engines of living cells, and they represent legitimate routes for the biochemical conversion of substrates to products either desired from a bioprocess or essential for cell survival. The algorithms discussed in this chapter could be used in one of the following two settings: (1) synthesize alternative pathways of chemical or biochemical reactions as a means to interpret overall transformations which are experimentally observed and (2) synthesize reaction pathways in the course of exploring new, alternative production routes. In this chapter, we will discuss examples in both directions. Although we will be concerned only with constraints on the directionality and stoichiometry of elementary reactions, the ideas can be extended to include other types of constraints arising, for example, from kinetics or thermodynamics.

I. Reaction Systems and Pathways

Complex systems of chemical reactions, and the problems associated with their design and analysis, are omnipresent in process engineering applications. In biochemical processes, complex networks of biochemical reactions, catalyzed by enzymes, accomplish both the growth of cells in the bioreactor and the conversion of the raw materials to the target products. Catalytic processes with solid catalysts and fluid-phase raw materials and products are used for many inorganic chemicals as well as refining and petrochemics. Noncatalytic fluid-phase reaction processes involving complex feeds and products are also common.

1. Hierarchical Structure

In all of these systems, chemical reactions have a hierarchical structure. What we normally think of as one reaction is actually composed from several elementary steps, which are often called, collectively, the *mechanism* of the reaction. For processes with a solid catalyst, these steps describe the interaction of the catalyst's active sites with the fluid-phase reactants. For biochemical reactions, there are association and dissociation steps that involve complexes of the enzyme and the substrates. In all cases, unstable, short-lived intermediates may be formed and destroyed in a rapid succession of steps; the overall reaction we observe might not involve these unstable intermediates because it is a suitable composition of such elementary steps.

The hierarchical character of reaction systems does not end with decomposition of reactions into steps. At the next level, many reactions taken together make up pathways. In bioprocesses, such pathways form long chains, cycles, and branching structures that accomplish biologically identifiable functions. In other processes, a pathway is used to describe the sequence of transformations that are needed to obtain the desired product from the available raw materials. One might even envision, at the next level, pathways combined to describe entire chemical plants or families of processing technologies.

2. *Synthesis of Pathways*

This chapter concerns itself with the hierarchical view of reaction systems. We are not interested in the decomposition of a reaction into mechanism steps or a pathway into reactions, but rather the synthesis or design of these composite entities from appropriate building blocks: Construction of reactions and mechanisms from elementary steps, or construction of pathways from reactions.

The common feature is the imposition of constraints on the overall chemical transformation, and the assembly of building blocks so that these constraints are met by the construct (reaction mechanism or pathway) as a whole. The constraints refer to the stoichiometry of the transformations, i.e., the reactants and products that are involved, along with their proportions. As we will see, there are different ways to formulate these constraints, and each way is suitable for different applications.

In all applications, the core problem can be stated as follows. We have a set of constraints that specify what may or may not be used as a starting material and final product of the process, and we are given a set of transformations (steps or reactions) that can be used as building blocks. For each transformation, we know the participating compounds and their stoichiometric coefficients. We need to assemble the transformations into a composite transformation, so that the stoichiometry of this composite transformation satisfies the constraints. This composite transformation may utilize any compound it needs along the way, as long as in the end its net stoichiometry satisfies the constraints. We can think of the construction procedure as assembly of a device from components, or we can abstract it mathematically into the formation of suitable linear combinations of the building-block transformations. In the second view, we should keep in mind that these linear combinations must respect the initial direction of reactions, since some reactions may be permitted only in their forward direction.

3. Application Domains

Two types of applications that involve this core problem will be discussed in this chapter. In the synthesis of biochemical pathways, using enzymatic reactions as building blocks, it is useful to examine each chemical species and decide whether it is required or allowed to be a starting material for the pathway, and likewise whether it is required or allowed to be a final product of the pathway.

The second application is the construction of catalytic reaction mechanisms out of elementary steps, involving only on type of catalytic site. A more useful way to formulate the stoichiometric constraints for these systems is to classify every chemical species as either a terminal species or an intermediate. Terminal species represent stable compounds that can be produced or consumed in significant quantities, while intermediates are short-lived unstable species that participate in the mechanism but are neither raw materials nor final products of the process.

4. Interpretation of Pathway Design

There are various ways in which we can interpret the idea of meeting stoichiometric constraints by combining transformations or assembling steps. We alluded to these interpretations in earlier sections, in our informal statement of the problem. As with the formulation of the constraints, each application domain accepts more naturally a different interpretation.

One interpretation, well suited to the synthesis of biochemical pathways, is process design. In the development and design of a new bioprocess, there are specifications on the raw materials that may be used and the desired and permitted products. One must select a pathway to use and a microorganism that possesses this pathway. With the advent of genetic engineering techniques, pathways and microorganisms are not restricted to existing ones: It is possible to modify a pathway of an organism by introducing biotransformations from other organisms. Therefore, the design of pathways takes on a true synthetic character rather than mere selection among enumerated alternatives.

Another interpretation is related to the dynamic behavior of a closed chemical system and the quasi-steady-state assumption. It is common in the kinetics of reaction systems to assume that the concentrations of intermediates reach a quasi-steady state: The rate of production and the rate of consumption of the intermediate approximately balance each other, and their difference (the net rate that leads to accumulation or depletion of the intermediate) is small compared to each of the two. The quasi-steady-state assumption allows simplification of the equations, by setting the accumulation term in the mass balance for an intermediate to zero. The construction of pathways is tantamount to identification of quasi-steady-state patterns, because the pathway makes sure that the intermediates are consumed at the same rate as they are produced. We are specifically interested in the smallest possible pathways, which cannot be shortened through elimination of a step and cannot be reduced to a superposition of smaller subpathways.

A similar interpretation can be proposed for open systems if only some species are permitted to enter and leave the system: The remaining species are the intermediates, and the construction of pathways arranges the stoichiometries to reflect the fact that, at steady state, the production of intermediates must balance their consumption.

II. Catalytic Reaction Systems

Our discussion of pathway design in the context of catalytic reaction mechanisms will summarize the treatment presented by Mavrovouniotis (1992) and Mavrovouniotis and Stephanopoulos (1992). The interested reader may refer to these for mathematical details, analysis of computational complexity, and comparison to other approaches in the context of model reaction systems.

A. Basic Concepts, Terminology, and Notation

We will follow the terminology and symbols used by Mavrovouniotis (1992) and Mavrovouniotis and Stephanopoulos (1992), who in turn followed the nomenclature of Happel and Sellers (1982, 1983, 1989), Happel (1986), Sellers (1984, 1989), and Happel *et al.* (1990). For the rest of this chapter, H&S will denote the latter set of references and the entire approach of Happel and Sellers. In accordance with H&S, we use the term *mechanism* rather than *pathway* for this class of systems.

1. Intermediate and Terminal Species

Consider a given set of elementary reaction steps that are feasible in a system, and the species involved in these steps. Species can be classified as either *intermediates*, which occur in very small amounts, or *terminal* species that can occur in significant amounts and constitute the raw materials and products of the process. The intermediates are precisely those species for which the quasi-steady-state assumption could be made.

A chemical systems consists of A species, designated as $a_1, a_2, \ldots, a_A$, and S mechanism steps, designated as $s_1, s_2, \ldots, s_S$. Each step s_i accomplishes a specific elementary reaction or transformation, $r_i = R(s_i)$. Let α_{ij} represent the stoichiometric coefficient of species a_j in step s_i, with the usual convention that $\alpha_{ij} > 0$ if and only if a_j is a product of s_i, $\alpha_{ij} < 0$ if and only if a_j is a reactant of s_i, and $\alpha_{ij} = 0$ if and only if a_j does not participate in s_i. The stoichiometry of the transformation $R(s_i)$ can then be written as

$$r_i = R(s_i) = \sum_{j=1}^{A} \alpha_{ij} a_j .$$

The coefficients α_{ij} of each step s_i denote only ratios of species. Thus, the meaning of the reaction r_i is not affected if we multiply all corresponding α_{ij} by a positive constant.

The directionality of a step s_i will be denoted by a label, as either $\leftrightarrows s_i$ or $\rightarrow s_i$. The sign $\leftrightarrows$ denotes a *reversible* step, whose net rate may be either positive (i.e., in the forward direction) or negative (i.e., in the reverse direction). The sign $\rightarrow$ denotes an *irreversible* step that is either thermodynamically irreversible or known to proceed with a net positive rate (i.e., in the forward direction).

The first I species (a_j with $j = 1, \ldots, I$) are assumed to be the intermediates, whereas the remaining $T = A - I$ species (a_j, $j = I + 1, \ldots, I + T$) are terminal species. If intermediates are present in very low concentrations, and their high rate of production is balanced by a high rate of consumption, the quasi-steady-state assumption allows dropping the accumulation term from the mass balance of an intermediate.

2. *Overall Mechanisms*

An *overall* mechanism (Horiuti and Nakamura, 1967; Horiuti, 1973; Temkin 1973, 1979) is one that obeys the quasi-steady-state assumption: It must consist of steps combined in specific proportions, such that the net transformation involves only terminal species.

Although there is significant net production of some terminal species, and significant net consumption of others, within the quasi-steady-state assumption the concentrations of intermediates are low and the rate of production of each intermediate by some steps is approximately balanced by its rate of consumption by other steps. *Overall* reactions can thus be defined as the set of net transformations permissible under the quasi-steady-state assumption; *overall* mechanisms are the combinations of steps that accomplish this.

A *mechanism*, m_k, is a linear combination of steps:

$$m_k = \sum_{i=1}^{S} \sigma_{ki} s_i ,$$

and belongs to an S-dimensional vector space. The coefficients σ_{ki} represent the proportional participation of the steps, and their signs must reflect the proper direction of steps: for irreversible steps, σ_{ki} must be nonnegative.

For clarity, the index i will hereafter be used for steps, the index j for species, and the index k for mechanisms. A reaction vector r_k, associated with each mechanism, describes the net transformation that is accomplished by the mechanism:

$$r_k = R(m_k) = \sum_{i=1}^{S} \sigma_{ki} R(s_i) .$$

Through substitution of the expression of $R(s_i)$ the reaction $R(m_k)$ can be written as

$$R(m_k) = \sum_{j=1}^{A} \beta_{kj} a_j \quad \text{with} \quad \beta_{kj} = \sum_{i=1}^{S} \sigma_{ki} \alpha_{ij} ,$$

where β_{kj} is the stoichiometric coefficient of species a_j in the net reaction accomplished by mechanism m_k.

A reaction $r_k = R(m_k)$ that involves only terminal species (i.e., $\beta_{kj} = 0$ for $j = 1, \ldots, I$) is called an *overall reaction*, and the mechanism m_k is an *overall mechanism*, corresponding to a quasi-steady state.

3. Direct Mechanisms

Clearly, a linear combination of overall mechanisms is also an overall mechanism. We are interested only in the smallest possible overall mechanisms, which describe the most extreme modes of operation of the chemical system. To define these, we examine the set of steps participating in a mechanism—with nonzero coefficients σ_{ki}. We are interested in overall mechanisms whose set of participating steps is not a proper superset of some other overall prime mechanism m'_k. These mechanisms, called *direct* mechanism by H&S, are minimal with respect to inclusion (for the set of participating steps). One might equivalently state that a mechanism m_k is direct if and only if every other overall mechanism includes at least one step not participating in m_k. The definition used here, in contrast to the one given by H&S, makes no particular reference to the *reaction* accomplished by the mechanism, other than the requirement that only overall mechanisms be considered.

Reaction mechanisms in the literature are often direct, because physical intuition encourages avoidance of excess steps. In considering the mechanism of a reaction, one often postulates ahead of time an implicit quasi-steady state, i.e., a combination of steps that explains the observed net reaction and conserves all intermediate species. In our approach, however, we are interested in identifying all possible direct mechanisms that are consistent with the postulated set of steps.

Direct mechanisms are not necessarily linearly independent; it may be possible to express a direct mechanism as a linear combination of other direct mechanisms. However, direct mechanisms are chemically distinct, because each involves a unique combination of steps.

B. Previous Work on the Construction of Mechanisms

A related concept is that of *laminar* mechanisms, introduced by Sinanoğlu (1975), Sinanoğlu and Lee (1978), and Lee and Sinanoğlu (1981). These mechanisms can be constructed a priori (i.e., without regard to a specific set of mechanism steps) given the number of catalysts involved. Sellers (1971, 1972) also presented procedures for construction of a priori mechanisms for certain chemical systems. In both these instances of mechanism construction, the focus was the a priori enumeration of potential pathway shapes, without regard to the specific reactions permitted by the system. These techniques are thus not applicable to the synthesis of pathways or mechanisms from a given set of reaction steps.

Milner (1964) used the term *direct* mechanisms to describe the smallest possible overall mechanisms. Milner (1964) showed that, under certain conditions, a direct mechanism can involve at most $I + 1$ steps, where I is the number of intermediates; this allows the construction of the mechanisms by examining combinations of $I + 1$ steps at a time. Another algorithm for the construction of direct mechanisms was presented by H&S, under the fundamental assumption that all steps are reversible and the net reaction taking place is actually known. This chapter presents an algorithm characterized by much greater generality and efficiency. The advantages of the algorithms presented here over those of H&S will be pointed out in the description of the algorithms and the examples throughout this section.

C. STRUCTURE OF THE ALGORITHM

We will describe here the general operation of the algorithm for the construction of direct mechanisms for chemical systems. We will examine the algorithm only in its simplest form. Mavrovouniotis (1992) presents various enhancements and discusses computational implementation considerations. We will begin with two examples (ammonia and methanol synthesis) that illustrate the step-by-step operation of the algorithm; these should be particularly useful for readers that are not accustomed to the abstract description of algorithms, who may wish to study the examples first, or study the abstract algorithm and the examples simultaneously.

Initially, the algorithm considers each individual reaction step as a partial mechanism. Then, one intermediate after another are examined, and the set of partial mechanisms is modified so that the intermediate does not appear in the net stoichiometry; the modification of mechanisms is carried out in a way that preserves the correct direction of irreversible reaction steps. By processing all intermediates in this way, a set of overall mechanisms is constructed. This final set of mechanisms might include duplicate mechanisms and even indirect ones; these can be easily discarded. Similar action must be taken in the procedure of H&S. Mavrovouniotis (1992) discusses procedures for eliminating such redundant mechanisms in the end, or even preventing their construction.

The algorithm operates on a set of intermediate species N, the set of terminal species N_T, and a set of partial mechanisms M, which is iteratively modified. Other information maintained by the algorithm during its operation is arranged in a convenient format in Fig. 1, which also explains how the entries are initialized. The algorithm proceeds from this setup by successively eliminating each intermediate species from the system. In order to eliminate an intermediate, we consider all the mechanisms whose reactions involve the intermediate species at hand. We can create combinations of two mechanisms at a time, with combination coefficients such that the intermediate vanishes from the reaction. The new row for a combination of mechanisms has elements σ_{ki} (left portion of Fig. 1) and β_{kj} (right portion of Fig. 1) that are simply the linear combinations of the respective elements of the old mechanisms. To permit the best choice of intermediate to be made, the setup of the algorithm (Fig. 1) lists, below

s_1	s_2	.....	s_S		number of combinations ⇒	n_1	n_2	.....	n_I	–	.....	–
ε_1	ε_2	.....	ε_S	mechanism ⇓	origin ⇓	a_1	a_2	.....	a_I	a_{I+1}	.....	a_A
σ_{11}	σ_{12}	.....	σ_{1S}	ε_1 m_1	—	β_{11}	β_{12}	.....	β_{1I}	$\beta_{1(I+1)}$	.....	β_{1A}
σ_{21}	σ_{22}	.....	σ_{2S}	ε_2 m_2	—	β_{21}	β_{22}	.....	β_{2I}	$\beta_{2(I+1)}$	.....	β_{2A}
⋮	⋮	⋮	⋮	⋮ ⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
σ_{S1}	σ_{S2}	.....	σ_{SS}	ε_S m_S	—	β_{S1}	β_{S2}	.....	β_{SI}	$\beta_{S(I+1)}$	.....	β_{SA}

FIG. 1. Initialized setup for the application of the algorithm for the construction of mechanisms. Rows correspond to mechanisms, and the operation of the algorithm adds and deletes rows. The left portion of the table contains the coefficients σ_{ki}; initially, each mechanism m_k contains only the corresponding step s_k (i.e., σ is the identity matrix). With each step symbol s_i we list the directionality label ϵ_i ($\rightarrow$ for an irreversible step, or $\leftrightarrows$ for a reversible step). Directionality labels are also listed for each mechanism m_k. The right portion of the table shows the reactions accomplished by the mechanisms, and groups intermediate species as a_1 to a_I, and terminal species as a_{I+1} to a_A. The algorithm deletes columns corresponding to intermediates as these are processed. Initially $\beta_{kj} = \alpha_{ki}$ when $i = j$, but this will not be the case for new mechanisms that are constructed. Above each intermediate species symbol the number of combination mechanisms that must be created to eliminate the species is listed; this will be explained further in the description of the algorithm. The column marked "origin" in the middle of the table is not essential; it is used for keeping track of how each new mechanism is constructed; one can list the intermediate that was eliminated and the mechanisms that were combined in the construction. [Reprinted with permission from Mavrovouniotis, M. L., and Stephanopoulos, G. "Synthesis of reaction mechanisms consisting of reversible and irreversible steps: I. A synthesis approach in the context of simple examples". *Ind. Eng. Chem. Res.* 31, 1625–1637. (1992). Copyright 1992 American Chemical Society.]

each species, the number of combination mechanisms that must be created to eliminate the species. The description given below explains the individual operations involved in more detail.

1. Initialization

We initialize N to the set of all intermediates (assuming $I \neq 0$), i.e., $N := \{a_1, a_2, \ldots, a_I\}$. The set M is initialized with one-step mechanisms. Therefore, the initial σ is just the identity matrix, whereas the initial β matrix is equal to the α matrix (Fig. 1):

$$M := \{m_1, m_2, \ldots, m_S\} \quad \text{with} \quad \beta_{kj} := \alpha_{kj} \text{ and } \sigma_{ki} := \begin{cases} 1 \text{ if } k = i, \\ 0 \text{ if } k \neq i. \end{cases}$$

The initial directionalities are the same as those of the corresponding individual steps (Fig.1).

2. *Number of Combinations for Each Intermediate*

We will now define the sets of partial mechanisms in which each intermediate participates. We divide them into irreversible and reversible ones to take directionality restrictions properly into account. For each species $a_j \in N$, let Y_j be a subset of M, defined as follows:

> A mechanism $m_k \in M$ belongs to Y_j if and only if it is irreversible and its net reaction contains a_j as a reactant, i.e., $m_k \in Y_j$ if and only if $m_k \in M$, $\rightarrow m_k$, and $\beta_{kj} < 0$.

Let y_j be the cardinality of Y_j, i.e., $y_j = |Y_j|$. Define similarly a subset Z_j of irreversible mechanisms as follows:

> A mechanism $m_k \in M$ belongs to Z_j if and only if it is irreversible and its net reaction contains a_j as a product (i.e., $\beta_{kj} > 0$).

Let z_j be the cardinality of this set. For reversible partial mechanisms, it does not matter whether a_j is a net reactant or a net product, as long as it appears in the net transformation. Thus, we define another subset, X_j, as follows:

> A mechanism $m_k \in M$ belongs to X_j if and only if it is reversible and its net reaction contains species a_j (i.e., $\beta_{kj} \neq 0$).

Let x_j be the cardinality of X_j, i.e., $x_j = |X_j|$. One can obtain the numbers y_j, z_j, and x_j by scanning the column of a_j in Fig. 1 for nonzero entries. A positive entry corresponding to an irreversible mechanism is counted into z_j. A negative entry corresponding to an irreversible mechanism is counted into y_j. A positive or negative (but not zero) entry corresponding to a reversible mechanism is counted into x_j.

We want to combine partial mechanisms, two at a time, so that a_j is not present in the net stoichiometry of the combination, always respecting directionality restrictions on mechanisms. We can combine a mechanism from X_j with one from Y_j, because the mechanisms in X_j can be reversed if necessary to make a_j a product (to balance with a mechanism from Y_j that uses a_j as a reactant); this gives rise to $x_j y_j$ combinations. Similarly, we can combine a mechanism from X_j with one from Z_j, reversing mechanisms in X_j if necessary to make a_j a reactant; this gives rise to $x_j z_j$ combinations. We can also combine a mechanism from X_j with any other mechanism from X_j in the same way; the number of combinations in this case is

$$\binom{x_j}{2} = \frac{x_j(x_j - 1)}{2}.$$

Finally, a mechanism from Z_j can be combined with any mechanism from Y_j, without any reversal, giving rise to $z_j y_j$ combinations.

The total number of possible combinations, n_j (placed above a_j in the layout of Fig. 1), necessary to eliminate a_j, is thus given by

$$n_j := \frac{x_j(x_j - 1)}{2} + x_j z_j + x_j y_j + z_j y_j.$$

Two limiting cases of this formula will be used later in the ammonia and methanol examples. If all steps are reversible, $n_j := [x_j(x_j - 1)]/2$, while if all the steps are irreversible $n_j := z_j y_j$.

3. *Selection and Elimination of an Intermediate*

The algorithm ultimately eliminates all intermediates, and one could process them in any order. However, to minimize computational effort in the current iteration, we select to eliminate the a_J that corresponds to the smallest n_J. For the intermediate a_J, the elimination constructs precisely those combinations that were counted into n_J in the previous phase of the algorithm. We put the new mechanisms in a set M_J. Specifically, the construction of M_J is governed by the following rules.

Rule 1. For each distinct combination $m_k \in X_J$, $m_b \in X_J$, the combination mechanism is $m_c := \beta_{bJ} m_k - \beta_{kJ} m_b$ and it is reversible ($\leftrightarrows m_c$).

Rule 2. For each $m_k \in Z_J$ and each $m_b \in X_J$, the mechanism $\rightarrow m_c$ must be constructed according to the sign of β_{bJ}, because the sign determines whether we need to reverse the mechanism m_b, and we must ensure that the coefficient that multiples m_k is positive. Thus

(a) If $\beta_{bJ} > 0$, we form $m_c := \beta_{bJ} m_k - \beta_{kJ} m_b$.

(b) If $\beta_{bj} < 0$, we form $m_c := \beta_{kJ} m_b - \beta_{bJ} m_k$.

Rule 3. Similarly, for each $m_k \in Y_J$ and each $m_b \in X_J$, for $m \rightarrow m_c$ as follows:

(a) If $\beta_{bJ} > 0$, we form $m_c := \beta_{bJ} m_k - \beta_{kJ} m_b$.

(b) If $\beta_{bJ} < 0$, we form $m_c := \beta_{kJ} m_b - \beta_{bJ} m_k$.

Rule 4. Finally, for each $m_k \in Y_J$ and each $m_b \in Z_J$, we form the mechanism $\rightarrow m_c$ as $m_c := \beta_{bJ} m_k - \beta_{kJ} m_b$. From the definition of the set Z_J it is easy to see that $\beta_{bJ} > 0$ and $-\beta_{kJ} > 0$, which guarantees that the irreversibility of mechanisms m_k and m_b is respected in the construction of m_c.

The combination mechanisms in all of these categories constitute the set M_J, whose cardinality is $n_J = x_J(x_J - 1)/2 + x_J z_J + x_J y_J + z_J y_J$. The directionality of each new mechanism is $\leftrightarrows$ (reversible) if and only if both mechanisms used in the combination were reversible. The σ and β coefficients of each combination mechanism are obtained as linear combinations of the respective coefficients of the constituent mechanisms. For $m_c := \beta_{bJ} m_k - \beta_{kJ} m_b$, one obtains $\sigma_{ci} := \beta_{bJ}\sigma_{ki} - \beta_{kJ}\sigma_{bi}$ (for all i) and $\beta_{cj} := \beta_{bJ}\beta_{kj} - \beta_{kJ}\beta_{bj}$ (for all j). From the latter expression the coefficient for a_J can be computed by setting $j := J$, which gives $\beta_{cJ} = \beta_{bJ}\beta_{kJ} - \beta_{kJ}\beta_{bJ} = 0$, verifying that the choice of combination coefficients achieves the elimination of a_J, as intended.

A new mechanism is logged as a new row in Fig. 1, and its σ and β portions are computed as linear combinations of the rows of the mechanisms being considered. The "origin" entry in Fig. 1 is used for bookkeeping. In the examples presented later, we record the name of the intermediate eliminated and a symbolic linear expression of the combination of mechanisms (see Figs. 3–7 and 11–16).

4. *Update of Active Sets*

In order to eliminate a_J we can now modify the current set of partial mechanisms M by adding the new combination mechanisms M_J, and removing their ancestors X_J, Y_J, and Z_J, i.e., $M := (M \cup M_J) - (X_J \cup Y_J \cup Z_J)$. We can then remove the intermediate a_J, by setting $N := N - \{a_J\}$. Within Fig. 1, in this phase we remove all rows that have nonzero entries in the a_J-column and then remove the column a_J itself.

5. *Termination or New Iteration*

If Fig. 1 still contains active columns corresponding to intermediates (i.e., $N \neq \emptyset$), we go back to phase 2 of the algorithm for the next iteration. If no intermediate columns remain ($N = \emptyset$) the elimination is complete, and all direct mechanisms are in the set M.

D. FEATURES OF THE ALGORITHM

Our algorithm for the synthesis of direct mechanisms initially considers each reaction step as a partial mechanism. Then, one intermediate after another are examined, and the set of partial mechanisms is modified so that the intermediate does not appear in the net stoichiometry; the modification of mechanisms is carried out in a way that preserves the correct direction of irreversible reaction steps. By processing all intermediates in this way, we can construct a set of overall mechanisms.

The set of mechanisms M resulting from the application of the basic algorithm may contain duplicate mechanisms (as in the ammonia example, see next subsection) or even indirect ones (as in the methanol example, see next subsection). A simple procedure can be applied in the end to remove redundant (duplicate or indirect) mechanisms from the final set of mechanisms. Alternatively, incremental checking can be carried out for each new mechanism or each iteration of the algorithm. Redundancy of mechanisms is one of a number of issues that affect the operation of the algorithm, and variations of the basic algorithm can be defined to enhance either its computational efficiency or its conceptual simplicity. For example, when two or more intermediates entail the same number of combinations (n_j), the basic algorithm makes a completely arbitrary choice; a variation of the algorithm might instead make the selection in a manner that removes as many existing mechanisms as possible from the set M, increasing computational efficiency. In the direction of conceptual simplicity, instead of assembling and using the sets X_j, Y_j, and Z_j separately, one may modify the algorithm by dropping directionality labels from mechanisms, and lumping X_j, Y_j, and Z_j into a single set W_j; the construction of combinations that eliminate a selected intermediate a_j using every possible pair of mechanisms $m_k \in W_j$ and $m_b \in W_j$, and rejecting the ones that contain steps in the wrong direction. These and related issues are discussed by Mavrovouniotis (1992).

The algorithm presented here for the construction of direct mechanisms differs significantly in its operation from the method of H&S. First, the irreversibility of mechanism steps is taken into account by our algorithm as mechanisms are constructed; i.e., it is not necessary to reserve directionality considerations until the end. Second, the mechanisms do not have to be restricted a priori to a particular reaction. Third, direct mechanisms are constructed recursively through combination of steps or partial mechanisms, rather than as linear combinations from a basis set of linearly independent mechanisms.

E. EXAMPLES

The mechanism–construction approach is illustrated here through two examples, which have also been treated by H&S: an example involving reversible steps for the synthesis of ammonia, and an example with irreversible steps for the synthesis of methanol. We note, in particular, that in the methanol example H&S assumed that the steps that lead to ethane and dimethylether (which are byproducts) are irrelevant for methanol mechanisms; this assumption was motivated by the need to keep

$N_2+I \leftrightarrows N_2I$	$N_2I+H_2 \leftrightarrows N_2H_2I$	$N_2H_2I+I \leftrightarrows 2NHI$	$N_2+2I \leftrightarrows 2NI$	$NI+HI \leftrightarrows NHI+I$	$NHI+HI \leftrightarrows NH_2I+I$	$NHI+H_2 \leftrightarrows NH_3+I$	$H_2+2I \leftrightarrows 2HI$	$NH_2I+HI \leftrightarrows NH_3+I$	mechanism	species ⇒	N_2I	N_2H_2I	NHI	NI	HI	NH_2I	I	N_2	H_2	NH_3
s_1	s_2	s_3	s_4	s_5	s_6	s_7	s_8	s_9	⇓	number of combinations ⇒	1	1	6	1	6	1	28	–	–	–
⇆	⇆	⇆	⇆	⇆	⇆	⇆	⇆	⇆		origin ⇓	a_1	a_2	a_3	a_4	a_5	a_6	a_7	a_8	a_9	a_{10}
1									⇆ m_1	—	1						−1	−1		
	1								⇆ m_2	—	−1	1							−1	
		1							⇆ m_3	—		−1	2				−1			
			1						⇆ m_4	—				2			−2	−1		
				1					⇆ m_5	—			1	−1	−1		1			
					1				⇆ m_6	—			−1		−1	1	1			
						1			⇆ m_7	—			−1				1		−1	1
							1		⇆ m_8	—					2		−2		−1	
								1	⇆ m_9	—					−1	−1	2			1

FIG. 2. Application of the algorithm on the ammonia mechanism. The mechanism steps for the synthesis of ammonia are shown on the top left portion, above each step symbol s_i. The steps are those used by Happel and Sellers (1982, 1983) and Happel *et al.* (1990), and were originally derived from the work of Horiuti (1973) and Temkin (1973). The active site of the catalyst is denoted as "I," and all species that contain this active site are intermediates; N_2, H_2, and NH_3 are terminal species. The setup follows the description given in Fig. 1, with the entries that are equal to zero left blank. Directionalities are shown, but for this example all steps are considered reversible and the directionality does not matter. We start by taking partial mechanisms $m_1, m_2, \ldots, m_9$, each of which has length 1 and includes only the step with the same index (i.e., $\sigma_{k_i} = 1$ if $k = i$; $\sigma_{k_i} = 0$ if $k \neq i$). Naturally, these are only partial mechanisms, which will be gradually transformed to overall mechanisms. Since all steps are reversible, the number of combination mechanisms that must be created to eliminate each intermediate (listed below the species) is simply $x(x-1)/2$, where x is the number of mechanisms whose reactions involve the species. The algorithm proceeds by successively eliminating each intermediate from the system. If, for example, we pursue the elimination of NHI, there are four existing partial mechanisms, m_3, m_5, m_6, and m_7, to be considered; there are $4 \times 3/2 = 6$ pairwise combinations of these that must be constructed to eliminate NHI. The choice to eliminate the intermediate a_3 (NHI) first, however, is not computationally efficient; if we choose NH_2I (a_6) instead, we only need to construct one mechanism. In fact, a_6 has the minimum number of combinations and its column has been highlighted in Fig. 2, to show that this is the species which will be eliminated first. Intuitively, one can think of this elimination as follows. There are only two steps that involve NH_2I; if any mechanism uses one of the two steps, it must use the other, with a coefficient that would eliminate the intermediate NH_2I; hence, the two steps can be combined. [Reprinted with permission from Mavrovouniotis, M. L., and Stephanopoulos, G. "Synthesis of reaction mechanisms consisting of reversible and irreversible steps: I. A synthesis approach in the context of simple examples". *Ind. Eng. Chem. Res.* **31**, 1625–1637. (1992). Copyright 1992 American Chemical Society.]

the size of the problem small, since the H&S algorithm is not very efficient computationally. As we will see, our algorithm constructs one mechanism that involves simultaneous production of ethane and methanol (i.e., m_{35} in Figs. 16 and 17), showing that the omission of byproducts is misleading.

1. Synthesis of Ammonia

We first consider an example on ammonia synthesis that was used by Happel and Sellers (1982, 1983) and Happel *et al.* (1990) and contains only

$N_2+I \leftrightarrows N_2I$	$N_2I+H_2 \leftrightarrows N_2H_2I$	$N_2H_2I+I \leftrightarrows 2NHI$	$N_2+2I \leftrightarrows 2NI$	$NI+HI \leftrightarrows NHI+I$	$NHI+HI \leftrightarrows NH_2I+I$	$NHI+H_2 \leftrightarrows NH_3+I$	$H_2+2I \leftrightarrows 2HI$	$NH_2I+HI \leftrightarrows NH_3+I$	mechanism	species ⇒	N_2I	N_2H_2I	NHI	NI	HI	NH_2I	I	N_2	H_2	NH_3
s_1	s_2	s_3	s_4	s_5	s_6	s_7	s_8	s_9	⇓	number of combinations ⇒	1	1	6	1	3	##	21	–	–	–
⇆	⇆	⇆	⇆	⇆	⇆	⇆	⇆	⇆		origin ⇓	a_1	a_2	a_3	a_4	a_5	a_6	a_7	a_8	a_9	a_{10}
1									⇆ m_1	—	1					##	−1	−1		
	1								⇆ m_2	—	−1	1				##			−1	
		1							⇆ m_3	—		−1	2			##	−1			
			1						⇆ m_4	—				2		##	−2	−1		
				1					⇆ m_5	—			1	−1	−1	##	1			
##	##	##	##	##	##	##	##	##	⇆ m_6	—	##	##	##	##	##	##	##	##	##	##
						1			⇆ m_7	—			−1			##	1		−1	1
							1		⇆ m_8	—					2	##	−2		−1	
##	##	##	##	##	##	##	##	##	⇆ m_9	—	##	##	##	##	##	##	##	##	##	##
					1			1	⇆ m_{10}	a_6: $m_6 + m_9$			−1		−2	##	3			1

FIG. 3. Setup for the application of the algorithm on the ammonia mechanism, after the elimination of a_6 (NH_2I). In the "origin" column, the construction of the mechanism m_{10} is documented as resulting from the elimination of a_6 through a specific linear combination of the mechanisms m_6 and m_9. The row of m_{10} has been obtained by carrying out the linear combination of the rows of m_6 and m_9. The rows for m_6 and m_9, as well as the column for a_6 have been crossed out from the table. The numbers of combinations have changed for a_5 and a_7 (but remain unchanged for other intermediates). Any one of the species a_1, a_2, and a_4, each entailing only one new combination, could be eliminated next. Both a_2 (N_2H_2I) and a_4 (NI) have been highlighted; their elimination can be carried out in parallel, because the two sets of mechanisms involved are disjoint; i.e., there is no mechanism whose reaction stoichiometry involves both a_4 and a_2. [Reprinted with permission from Mavrovouniotis, M. L., and Stephanopoulos, G. "Synthesis of reaction mechanisms consisting of reversible and irreversible steps: I. A synthesis approach in the context of simple examples". *Ind. Eng. Chem. Res.* **31**, 1625–1637, (1992). Copyright 1992 American Chemical Society.]

$N_2+I \leftrightarrows N_2I$	$N_2I+H_2 \leftrightarrows N_2H_2I$	$N_2H_2I+I \leftrightarrows 2NHI$	$N_2+2I \leftrightarrows 2NI$	$NI+HI \leftrightarrows NHI+I$	$NHI+HI \leftrightarrows NH_2I+I$	$NHI+H_2 \leftrightarrows NH_3+I$	$H_2+2I \leftrightarrows 2HI$	$NH_2I+HI \leftrightarrows NH_3+I$	mechanism	species ⇒	N_2I	N_2H_2I	NHI	NI	HI	NH_2I	I	N_2	H_2	NH_3
s_1	s_2	s_3	s_4	s_5	s_6	s_7	s_8	s_9	⇓	number of combinations ⇒	1	##	6	##	3	##	10	–	–	–
⇆	⇆	⇆	⇆	⇆	⇆	⇆	⇆	⇆		origin ⇓	a_1	a_2	a_3	a_4	a_5	a_6	a_7	a_8	a_9	a_{10}
1									⇆ m_1	—	1	##		##		##	−1	−1		
##	##	##	##	##	##	##	##	##	⇆ m_2	—	##	##	##	##	##	##	##	##	##	##
##	##	##	##	##	##	##	##	##	⇆ m_3	—	##	##	##	##	##	##	##	##	##	##
##	##	##	##	##	##	##	##	##	⇆ m_4	—	##	##	##	##	##	##	##	##	##	##
##	##	##	##	##	##	##	##	##	⇆ m_5	—	##	##	##	##	##	##	##	##	##	##
##	##	##	##	##	##	##	##	##	⇆ m_6	—	##	##	##	##	##	##	##	##	##	##
						1			⇆ m_7	—		##	−1	##		##	1		−1	1
							1		⇆ m_8	—		##		##	2	##	−2		−1	
##	##	##	##	##	##	##	##	##	⇆ m_9	—	##	##	##	##	##	##	##	##	##	##
					1			1	⇆ m_{10}	a_6: m_6+m_9		##	−1	##	−2	##	3			1
	1	1							⇆ m_{11}	a_2: m_2+m_3	−1	##	2	##		##	−1		−1	
			1	2					⇆ m_{12}	a_4: m_4+2m_5		##	2	##	−2	##		−1		

FIG. 4. Application of the algorithm on the ammonia mechanism, after the elimination of a_2 and a_4, which yielded two new partial mechanisms, m_{11} and m_{12}, and eliminated m_2, m_3, m_4, and m_5. The number of combinations has been recomputed for each remaining intermediate species. The intermediate N_2I (a_1) will be eliminated next. [Reprinted with permission from Mavrovouniotis, M. L., and Stephanopoulos, G. "Synthesis of reaction mechanisms consisting of reversible and irreversible steps: I. A synthesis approach in the context of simple examples". *Ind. Eng. Chem. Res.* **31**, 1625–1637, (1992). Copyright 1992 American Chemical Society.]

reversible steps. The initial setup for the application of the algorithm is shown in Fig. 2, which also presents the actual elementary steps above the corresponding symbols s_i. The progression of the algorithm is explained in Figs. 2–7. Each figure shows one or more iterations through phases 2–5 of the algorithm; thus each figure eliminates one or more intermediates and flags those intermediates that will be eliminated next. Figure 7 provides the final results and compares them to the results obtained by Happel *et al.* (1990).

2. *Synthesis of Methanol*

Consider, as a second example, the mechanism for the synthesis of methanol (Fig. 8), analyzed by Happel *et al.* (1990). All the steps are

$N_2+I \leftrightarrows N_2I$	$N_2I+H_2 \leftrightarrows N_2H_2I$	$N_2H_2I+I \leftrightarrows 2NHI$	$N_2+2I \leftrightarrows 2NI$	$NI+HI \leftrightarrows NHI+I$	$NHI+HI \leftrightarrows NH_2I+I$	$NHI+H_2 \leftrightarrows NH_3+I$	$H_2+2I \leftrightarrows 2HI$	$NH_2I+HI \leftrightarrows NH_3+I$	mechanism	species ⇒	N_2I	N_2H_2I	NHI	NI	HI	NH_2I	I	N_2	H_2	NH_3
s_1	s_2	s_3	s_4	s_5	s_6	s_7	s_8	s_9	⇓	number of combinations ⇒	##	##	6	##	3	##	6	–	–	–
⇆	⇆	⇆	⇆	⇆	⇆	⇆	⇆	⇆		origin ⇓	a_1	a_2	a_3	a_4	a_5	a_6	a_7	a_8	a_9	a_{10}
##	##	##	##	##	##	##	##	##	⇆ m_1	—	##	##	##	##	##	##	##	##	##	##
##	##	##	##	##	##	##	##	##	⇆ m_2	—	##	##	##	##	##	##	##	##	##	##
##	##	##	##	##	##	##	##	##	⇆ m_3	—	##	##	##	##	##	##	##	##	##	##
##	##	##	##	##	##	##	##	##	⇆ m_4	—	##	##	##	##	##	##	##	##	##	##
##	##	##	##	##	##	##	##	##	⇆ m_5	—	##	##	##	##	##	##	##	##	##	##
##	##	##	##	##	##	##	##	##	⇆ m_6	—	##	##	##	##	##	##	##	##	##	##
						1			⇆ m_7	—	##	##	−1	##		##	1		−1	1
							1		⇆ m_8	—	##	##		##	2	##	−2		−1	
##	##	##	##	##	##	##	##	##	⇆ m_9	—	##	##	##	##	##	##	##	##	##	##
					1			1	⇆ m_{10}	a_6: $m_6 + m_9$	##	##	−1	##	−2	##	3			1
##	##	##	##	##	##	##	##	##	⇆ m_{11}	a_2: $m_2 + m_3$	##	##	##	##	##	##	##	##	##	##
			1	2					⇆ m_{12}	a_4: $m_4 + 2m_5$	##	##	2	##	−2	##		−1		
1	1	1							⇆ m_{13}	a_1: $m_1 + m_{11}$	##	##	2	##		##	−2	−1	−1	

FIG. 5. Application of the algorithm on the ammonia mechanism, after the elimination of a_1, with only five mechanisms (m_7, m_8, m_{10}, m_{12}, and m_{13}) remaining active. This reduction in the number of mechanisms as the first few intermediates are eliminated is quite common; however, the number of mechanisms tends to increase again at the end. The intermediate HI (a_5) will be eliminated next. [Reprinted with permission from Mavrovouniotis, M. L., and Stephanopoulos, G. "Synthesis of reaction mechanisms consisting of reversible and irreversible steps: I. A synthesis approach in the context of simple examples". *Ind. Eng. Chem. Res.* **31**, 1625–1637, (1992). Copyright 1992 American Chemical Society.]

irreversible, providing an opportunity to show how directionality of steps is taken into account as the mechanisms are constructed. In the beginning of their analysis, Happel *et al.* (1990) remove steps s_9 and s_{11}, because they lead to byproducts (ethane and dimethylether) rather than methanol. Here, we will keep all steps; it turns out that the products are not formed independently of each other, and isolation of the main product from the byproducts of the mechanism can be misleading.

The application of the algorithm is shown in Figs. 9–16, with explanation of the differences arising from the directionality of the steps; each figure eliminates one or more intermediates and flags those intermediates

$N_2+I \leftrightarrows N_2I$	$N_2I+H_2 \leftrightarrows N_2H_2I$	$N_2H_2I+I \leftrightarrows 2NHI$	$N_2+2I \leftrightarrows 2NI$	$NI+HI \leftrightarrows NHI+I$	$NHI+HI \leftrightarrows NH_2I+I$	$NHI+H_2 \leftrightarrows NH_3+I$	$H_2+2I \leftrightarrows 2HI$	$NH_2I+HI \leftrightarrows NH_3+I$	mechanism	species ⇒	N_2I	N_2H_2I	NHI	NI	HI	NH_2I	I	N_2	H_2	NH_3
s_1	s_2	s_3	s_4	s_5	s_6	s_7	s_8	s_9	⇓	number of combinations ⇒	##	##	10	##	##	##	10	–	–	–
⇆	⇆	⇆	⇆	⇆	⇆	⇆	⇆	⇆		origin ⇓	a_1	a_2	a_3	a_4	a_5	a_6	a_7	a_8	a_9	a_{10}
						1			⇆ m_7	—	##	##	−1	##	##	##	1		−1	1
##	##	##	##	##	##	##	##	##	⇆ m_8	—	##	##	##	##	##	##	##	##	##	##
##	##	##	##	##	##	##	##	##	⇆ m_{10}	a_6: m_8+m_9	##	##	##	##	##	##	##	##	##	##
##	##	##	##	##	##	##	##	##	⇆ m_{12}	a_4: m_4+2m_5	##	##	##	##	##	##	##	##	##	##
1	1	1							⇆ m_{13}	a_1: m_1+m_{11}	##	##	2	##	##	##	−2	−1	−1	
					1		1	1	⇆ m_{14}	a_5: m_8+m_{10}	##	##	−1	##	##	##	1		−1	1
			1	2			1		⇆ m_{15}	a_5: m_8+m_{12}	##	##	2	##	##	##	−2	−1	−1	
			1	2	−1			−1	⇆ m_{16}	a_5: $m_{12}-m_{10}$	##	##	3	##	##	##	−3	−1		−1

FIG. 6. Application of the algorithm on the ammonia mechanism, after the elimination of a_5. We can generally, at any time, drop inactive (crossed out) mechanisms or eliminated intermediates, but we have so far maintained them so that one can easily keep track of the progress of the algorithm. At this point, in order to reduce the size of the setup, we drop from the figure all those mechanisms that had been crossed out in Fig. 5 or earlier. The two intermediates (a_3 and a_7) remaining in Fig. 6 involve the same number of combinations; in fact elimination of either one results automatically in the elimination of the other (Fig. 7), because the columns for a_3 and a_7 contained precisely opposite numbers; a_3 is arbitrarily chosen. [Reprinted with permission from Mavrovouniotis, M. L., and Stephanopoulos, G. "Synthesis of reaction mechanisms consisting of reversible and irreversible steps: I. A synthesis approach in the context of simple examples". *Ind. Eng. Chem. Res.* **31**, 1625–1637, (1992). Copyright 1992 American Chemical Society.]

that will be eliminated next. Figure 17 analyzes the results and compares them to those of Happel *et al.* (1990). With respect to the mechanisms m_{29} and m_{35}, which lead to byproducts, we observe (Figs. 16 and 17) that m_{29} can be thought of as unrelated to the synthesis of methanol, but omission of m_{35} may be misleading: This mechanism leads to simultaneous production of methanol and ethane, in stoichiometric proportions. Thus, the mechanism m_{35} should properly be viewed as one of the mechanisms that lead to methanol—with the drawback, of course, that it also leads to an equal number of moles of an undesired byproduct. The omission of byproducts is therefore risky simplification. We also note that the inclusion of directionality considerations appears particularly cumbersome in the treatment of Happel *et al.* (1990, p. 1061): The direct submechanisms for one of the basis reactions violate directionality restrictions; thus, an

$N_2+I \leftrightarrows N_2I$	$N_2I+H_2 \leftrightarrows N_2H_2I$	$N_2H_2I+I \leftrightarrows 2NHI$	$N_2+2I \leftrightarrows 2NI$	$NI+HI \leftrightarrows NHI+I$	$NHI+HI \leftrightarrows NH_2I+I$	$NHI+H_2 \leftrightarrows NH_3+I$	$H_2+2I \leftrightarrows 2HI$	$NH_2I+HI \leftrightarrows NH_3+I$	mechanism	corresponding mechanism from Happel *et al.* (1990)	species ⇒	N_2I	N_2H_2I	NHI	NI	HI	NH_2I	I	N_2	H_2	NH_3
s_1 ⇆	s_2 ⇆	s_3 ⇆	s_4 ⇆	s_5 ⇆	s_6 ⇆	s_7 ⇆	s_8 ⇆	s_9 ⇆	⇓	⇓	number of combinations ⇒	##	##	##	##	##	##	0	–	–	–
											origin ⇓	a_1	a_2	a_3	a_4	a_5	a_6	a_7	a_8	a_9	a_{10}
1	1	1				2			⇆ m_{17}	m_4,Table V	a_3: $m_{13}+2m_7$	##	##	##	##	##	##		–1	–3	2
					1	–1	1	1	⇆ m_{18}	null,Table IV	a_3: $m_{14}-m_7$	##	##	##	##	##	##		0	0	0
			1	2		2	1		⇆ m_{19}	m_1,Table V	a_3: $m_{15}+2m_7$	##	##	##	##	##	##		–1	–3	2
			1	2	–1			–1	⇆ m_{20}	m_3,Table V	a_3: $m_{16}+3m_7$	##	##	##	##	##	##		–1	–3	2
1	1	1			2		2	2	⇆ m_{21}	m_5,Table V	a_3: $m_{13}+2m_{14}$	##	##	##	##	##	##		–1	–3	2
–1	–1	–1	1	2			1		⇆ m_{22}	null,Table IV	a_3: $m_{15}-m_{13}$	##	##	##	##	##	##		0	0	0
3	3	3	–2	–4	2			2	⇆ m_{23}	m_6,Table V	a_3: $3m_{13}-2m_{16}$	##	##	##	##	##	##		–1	–3	2
			1	2	2		3	2	⇆ m_{24}	m_2,Table V	a_3: $m_{15}+2m_{14}$	##	##	##	##	##	##		–1	–3	2
			1	2	2		3	2	⇆ m_{25}	m_2,Table V	a_3: $m_{16}+3m_{14}$	##	##	##	##	##	##		–1	–3	2
			1	2	2		3	2	⇆ m_{26}	m_2,Table V	a_3: $3m_{15}-2m_{16}$	##	##	##	##	##	##		–1	–3	2

FIG. 7. The final set of mechanisms produced here and their correspondence to the mechanisms constructed by Happel *et al.* (1990). As was explained earlier, the intermediate a_7 does not need to be eliminated because its column contains only zeros. All the direct mechanisms identified by Happel *et al.* (1990) have been produced. However, the last three mechanisms in Figure 7 are actually identical. Hence, the simple procedure discussed here does not preclude multiple occurrences of the same mechanism, or, as the next example will show, the occurrence of mechanisms that are not direct. For small studies this is not a significant drawback, because one can easily eliminate the redundancies in the end. The potential duplication of mechanisms and construction of indirect mechanisms are addressed by Mavrovouniotis (1992). [Reprinted with permission from Mavrovouniotis, M. L., and Stephanopoulos, G. "Synthesis of reaction mechanisms consisting of reversible and irreversible steps: I. A synthesis approach in the context of simple examples". *Ind. Eng. Chem. Res.* **31**, 1625–1637, (1992). Copyright 1992 American Chemical Society.]

alternative reaction is manually constructed which gives acceptable direct submechanisms.

3. *Importance of Reaction Directionality*

We provide here an abstract example (Fig. 18) to discuss the complexity characteristics of the method of H&S, and in particular that method's assumption at the outset that all reactions are reversible. This assumption might be insignificant for very simple reaction mechanisms or those that in fact do not contain irreversible steps. However, the assumption constitutes

s_1: $CH_4 + O_2 \rightarrow CH_3 + HO_2$

s_2: $CH_3 + O_2 \rightarrow CH_3O_2$

s_3: $CH_3O_2 \rightarrow CH_2O + OH$

s_4: $CH_3O_2 + CH_4 \rightarrow CH_3O_2H + CH_3$

s_5: $CH_3O_2H \rightarrow CH_3O + OH$

s_6: $CH_3O \rightarrow CH_2O + H$

s_7: $CH_3O + CH_4 \rightarrow CH_3OH + CH_3$

s_8: $OH + CH_4 \rightarrow CH_3 + H_2O$

s_9: $CH_3 + CH_3 \rightarrow C_2H_6$

s_{10}: $CH_3 + OH \rightarrow CH_3OH$

s_{11}: $CH_3 + CH_3O \rightarrow CH_3OCH_3$

s_{12}: $CH_2O + CH_3 \rightarrow CH_4 + CHO$

s_{13}: $CHO + O_2 \rightarrow CO + HO_2$

s_{14}: $CH_2O + CH_3O \rightarrow CH_3OH + CHO$

s_{15}: $CHO + CH_3 \rightarrow CO + CH_4$

FIG. 8. Mechanism steps for the synthesis of methanol, as used in an example by Happel *et al.* (1990). The mechanism was proposed by Yarlagadda *et al.* (1988) and assumes that all steps have a net rate in the indicated direction. The species CH_4, O_2, CH_3OH, CO, H_2O, C_2H_6, and CH_3OCH_3 are terminal and all others are intermediates; formaldehyde (CH_2O) is an intermediate because it is present in small amounts (Happel *et al.*, 1990). [Reprinted with permission from Mavrovouniotis, M. L., and Stephanopoulos, G. "Synthesis of reaction mechanisms consisting of reversible and irreversible steps: I. A synthesis approach in the context of simple examples". *Ind. Eng. Chem. Res.* **31**, 1625–1637, (1992). Copyright 1992 American Chemical Society.]

a significant drawback if many steps are irreversible and one starts with a very large reaction system, a large portion of which eventually does not participate in the final solution. In particular, the presence of many parallel irreversible routes leading from intermediates to other intermediates can have devastating consequences on efficiency. Figure 18 shows an idealized schematic of these features; there is actually only one mechanism that converts A to E, but a method that postpones consideration of reaction directionality until the final mechanisms have been constructed will produce an enormous number of possibilities. Thus, the procedure of H&S, as well as any method that ignores directionality, faces great difficulties for large systems of mostly irreversible reactions.

															mechanism	species ⇒	CH_3	HO_2	CH_3O_2	OH	CH_3O_2H	CH_3O	H	CHO	CH_2O
s_1	s_2	s_3	s_4	s_5	s_6	s_7	s_8	s_9	s_{10}	s_{11}	s_{12}	s_{13}	s_{14}	s_{15}	⇓	combinations⇒	24	0	2	4	1	4	0	4	4
→	→	→	→	→	→	→	→	→	→	→	→	→	→	→		origin ⇓	a_1	a_2	a_3	a_4	a_5	a_6	a_7	a_8	a_9
1															→m_1	—	1	1							
	1														→m_2	—	−1		1						
		1													→m_3	—			−1	1					1
			1												→m_4	—	1		−1		1				
				1											→m_5	—				1	−1	1			
					1										→m_6	—						−1	1		1
						1									→m_7	—	1					−1			
							1								→m_8	—	1			−1					
								1							→m_9	—	−2								
									1						→m_{10}	—	−1			−1					
										1					→m_{11}	—	−1					−1			
											1				→m_{12}	—	−1							1	−1
												1			→m_{13}	—		1						−1	
													1		→m_{14}	—						−1		1	−1
														1	→m_{15}	—	−1							−1	

FIG. 9. Initial setup for the application of the algorithm on the methanol mechanism. Terminal species are not shown in this figure; they will be incorporated in Fig. 10. The initial arrangement shows the mechanisms m_1 to m_{15}, each using only the step with the same index. The arrows below each step and to the left of each mechanism serve as indicators of directionality; here, all steps are irreversible (→), and the algorithm that is proposed in this chapter can take directionality into account. In the formation of combinations of partial mechanisms to eliminate an intermediate species, it is no longer possible to take *any* combination of two mechanisms whose net reactions involve the species. If, for example, we attempt to combine m_1 and m_{13} to eliminate a_2 (HO_2), then the combination expression $\beta_{b2}m_k - \beta_{k2}m_b$ would lead to either $m_1 - m_{13}$ (for $k = 1$ and $b = 13$) or $m_{13} - m_1$ (for $k = 13$ and $b = 1$), violating the directionality of one of the mechanisms m_1 and m_{13}. This happens because a_2 participates as a net product in both mechanisms; it is simply not possible to eliminate a_2 if we insist on using both mechanisms in the forward direction. Thus, for irreversible steps and mechanisms, a legitimate combination that eliminates an intermediate must include one mechanism in which the species is a net product and one mechanism in which the species is a net reactant. This makes each intermediate's number of combinations equal to yz, where z is the number of mechanisms for which the species is a net product, and y is the number of mechanisms for which the species in question is a net reactant. The somewhat more complicated procedure for chemical systems that include both irreversible and reversible steps was given in the algorithm. As in the previous example, the intermediates with the smallest numbers of combinations are highlighted. The species a_2 (HO_2) and a_7 (H) are chosen for elimination, because they give rise to zero combinations (they occur only as products). [Reprinted with permission from Mavrovouniotis, M. L., and Stephanopoulos, G. "Synthesis of reaction mechanisms consisting of reversible and irreversible steps: I. A synthesis approach in the context of simple examples". *Ind. Eng. Chem. Res.* **31**, 1625–1637. (1992), Copyright 1992 American Chemical Society.]

															mechanism	species ⇒	CH_3	CH_3O_2	OH	CH_3O_2H	CH_3O	CHO	CH_2O	CH_4	O_2	CH_3OH	CO	H_2O	C_2H_6	CH_3OCH_3
s_1	s_2	s_3	s_4	s_5	s_6	s_7	s_8	s_9	s_{10}	s_{11}	s_{12}	s_{13}	s_{14}	s_{15}	⇓	combinations⇒	18	2	4	1	3	2	2							
→	→	→	→	→	→	→	→	→	→	→	→	→	→	→		origin ⇓	a_1	a_3	a_4	a_5	a_6	a_8	a_9	a_{10}	a_{11}	a_{12}	a_{13}	a_{14}	a_{15}	a_{16}
	1														→m_2	—	−1	1							−1					
		1													→m_3	—		−1	1				1							
			1												→m_4	—	1	−1		1				−1						
				1											→m_5	—			1	−1	1									
						1									→m_7	—	1				−1			−1		1				
							1								→m_8	—	1		−1					−1				1		
								1							→m_9	—	−2												1	
									1						→m_{10}	—	−1		−1							1				
										1					→m_{11}	—	−1				−1									1
											1				→m_{12}	—	−1					1	−1	1						
													1		→m_{14}	—					−1	1	−1			1				
														1	→m_{15}	—	−1					−1		1			1			

FIG. 10. Setup for the methanol mechanism, after the elimination of a_2 (HO_2) and a_7 (H). Mechanisms that have been abolished and intermediates that have been eliminated are immediately removed from the figures; thus, the rows of m_1, m_6, and m_{13}, and the columns of a_2 and a_7 are not present in Fig. 10 and subsequent figures. The terminal species, not shown in Fig. 9, have been included in the setup. The number of combinations has been recalculated for each species. The intermediates a_5 (CH_3O_2H) and a_8 (CHO), each of which gives rise to only two combinations, are next eliminated in parallel, since their sets of mechanisms (m_4 and m_5 for a_5; m_{12}, m_{14}, and m_{15} for a_8) are disjoint. [Reprinted with permission from Mavrovouniotis, M. L., and Stephanopoulos, G. "Synthesis of reaction mechanisms consisting of reversible and irreversible steps: I. A synthesis approach in the context of simple examples". *Ind. Eng. Chem. Res.* **31**, 1625–1637. (1992). Copyright 1992 American Chemical Society.]

III. Biochemical Pathways

The algorithm described for direct reaction mechanisms can also be used for the synthesis of biochemical pathways from basic bioreactions in the pursuit of quasi-steady-state behaviors of bioprocesses. More importantly, the algorithm identifies pathways leading from available raw materials to desired target products, enabling more informed decisions in the early stages of the design of a bioprocess. In translating the algorithm to biochemical systems, mechanism steps would map to individual bioreactions (usually catalyzed by enzymes), whereas overall reaction mechanisms correspond to acceptable biochemical pathways for a bioprocess.

Here, we will discuss a different formulation of the problem, which is better suited to the preliminary-design interpretation of the synthesis of biochemical pathways. This section will be based on the algorithm and

s_1 →	s_2 →	s_3 →	s_4 →	s_5 →	s_6 →	s_7 →	s_8 →	s_9 →	s_{10} →	s_{11} →	s_{12} →	s_{13} →	s_{14} →	s_{15} →	mechanism ⇓	species ⇒	CH_3	CH_3O_2	OH	CH_3O	CH_2O	CH_4	O_2	CH_3OH	CO	H_2O	C_2H_6	CH_3OCH_3
																combinations⇒	18	2	4	3	2							
																origin ⇓	a_1	a_3	a_4	a_6	a_9	a_{10}	a_{11}	a_{12}	a_{13}	a_{14}	a_{15}	a_{16}
	1														→m_2	—	-1	1					-1					
		1													→m_3	—		-1	1		1							
						1									→m_7	—	1			-1		-1		1				
							1								→m_8	—	1		-1			-1				1		
								1							→m_9	—	-2										1	
									1						→m_{10}	—	-1		-1					1				
										1					→m_{11}	—	-1			-1								1
			1	1											→m_{16}	a_5: m_4+m_5	1	-1	1	1		-1						
											1			1	→m_{17}	a_8: $m_{12}+m_{15}$	-2				-1	2			1			
													1	1	→m_{18}	a_8: $m_{14}+m_{15}$	-1			-1	-1	1		1	1			

FIG. 11. Setup for the application of the algorithm on the methanol mechanism, after the elimination of a_5 (CH_3O_2H) and a_8 (CHO). Elimination of the intermediate a_9 (CH_2O), which involves two combination mechanisms, is carried out next. [Reprinted with permission from Mavrovouniotis, M. L., and Stephanopoulos, G. "Synthesis of reaction mechanisms consisting of reversible and irreversible steps: I. A synthesis approach in the context of simple examples". *Ind. Eng. Chem. Res.* **31**, 1625–1637. (1992). Copyright 1992 American Chemical Society.]

s_1 →	s_2 →	s_3 →	s_4 →	s_5 →	s_6 →	s_7 →	s_8 →	s_9 →	s_{10} →	s_{11} →	s_{12} →	s_{13} →	s_{14} →	s_{15} →	mechanism ⇓	species ⇒	CH_3	CH_3O_2	OH	CH_3O	CH_4	O_2	CH_3OH	CO	H_2O	C_2H_6	CH_3OCH_3
																combinations⇒	18	3	6	3							
																origin ⇓	a_1	a_3	a_4	a_6	a_{10}	a_{11}	a_{12}	a_{13}	a_{14}	a_{15}	a_{16}
	1														→m_2	—	-1	1				-1					
						1									→m_7	—	1			-1	-1		1				
							1								→m_8	—	1		-1		-1				1		
								1							→m_9	—	-2									1	
									1						→m_{10}	—	-1		-1				1				
										1					→m_{11}	—	-1			-1							1
			1	1											→m_{16}	a_5: m_4+m_5	1	-1	1	1	-1						
		1									1			1	→m_{19}	a_9: m_3+m_{17}	-2	-1	1		2			1			
		1											1	1	→m_{20}	a_9: m_3+m_{18}	-1	-1	1	-1	1		1	1			

FIG. 12. State of the algorithm, for the mechanism, after the elimination of a_9 (CH_2O). Similarly, elimination of the intermediate a_3 (CH_3O_2) leads to Fig. 13. [Reprinted with permission from Mavrovouniotis, M. L., and Stephanopoulos, G. "Synthesis of reaction mechanisms consisting of reversible and irreversible steps: I. A synthesis approach in the context of simple examples". *Ind. Eng. Chem. Res.* **31**, 1625–1637. (1992). Copyright 1992 American Chemical Society.]

s_1 →	s_2 →	s_3 →	s_4 →	s_5 →	s_6 →	s_7 →	s_8 →	s_9 →	s_{10} →	s_{11} →	s_{12} →	s_{13} →	s_{14} →	s_{15} →	mechanism ⇓	species ⇒	CH_3	OH	CH_3O	CH_4	O_2	CH_3OH	CO	H_2O	C_2H_6	CH_3OCH_3
																combinations⇒	10	6	3							
																origin ⇓	a_1	a_4	a_6	a_{10}	a_{11}	a_{12}	a_{13}	a_{14}	a_{15}	a_{16}
						1									→ m_7	—	1		−1	−1		1				
							1								→ m_8	—	1	−1		−1				1		
								1							→ m_9	—	−2								1	
									1						→ m_{10}	—	−1	−1				1				
										1					→ m_{11}	—	−1		−1							1
	1		1	1											→ m_{21}	a_3: m_2+m_{16}		1	1	−1	−1					
	1	1									1			1	→ m_{22}	a_3: m_2+m_{19}	−3	1		2	−1		1			
	1	1											1	1	→ m_{23}	a_3: m_2+m_{20}	−2	1	−1	1	−1	1	1			

FIG. 13. State of the algorithm, for the methanol mechanism, after the elimination of a_3 (CH_3O_2). Next, elimination of the intermediate a_6 (CH_3O) yields Fig. 14. [Reprinted with permission from Mavrovouniotis, M. L., and Stephanopoulos, G. "Synthesis of reaction mechanisms consisting of reversible and irreversible steps: 1. A synthesis approach in the context of simple examples". *Ind. Eng. Chem. Res.* **31**, 1625–1637. (1992). Copyright 1992 American Chemical Society.]

s_1 →	s_2 →	s_3 →	s_4 →	s_5 →	s_6 →	s_7 →	s_8 →	s_9 →	s_{10} →	s_{11} →	s_{12} →	s_{13} →	s_{14} →	s_{15} →	mechanism ⇓	species ⇒	CH_3	OH	CH_4	O_2	CH_3OH	CO	H_2O	C_2H_6	CH_3OCH_3
																combinations⇒	10	8							
																origin ⇓	a_1	a_4	a_{10}	a_{11}	a_{12}	a_{13}	a_{14}	a_{15}	a_{16}
							1								→ m_8	—	1	−1	−1				1		
								1							→ m_9	—	−2							1	
									1						→ m_{10}	—	−1	−1			1				
	1	1									1			1	→ m_{22}	a_3: m_2+m_{19}	−3	1	2	−1		1			
	1		1	1		1									→ m_{24}	a_6: $m_{21}+m_7$	1	1	−2	−1	1				
	1		1	1						1					→ m_{25}	a_6: $m_{21}+m_{11}$	−1	1	−1	−1					1
	2	1	1	1									1	1	→ m_{26}	a_6: $m_{21}+m_{23}$	−2	2		−2	1	1			

FIG. 14. Setup of the methanol mechanism, after the elimination of a_6 (CH_3O). The intermediate a_4 (OH) will be eliminated next. [Reprinted with permission from Mavrovouniotis, M. L., and Stephanopoulos, G. "Synthesis of reaction mechanisms consisting of reversible and irreversible steps: 1. A synthesis approach in the context of simple examples". *Ind. Eng. Chem. Res.* **31**, 1625–1637. (1992). Copyright 1992 American Chemical Society.]

s_1 →	s_2 →	s_3 →	s_4 →	s_5 →	s_6 →	s_7 →	s_8 →	s_9 →	s_{10} →	s_{11} →	s_{12} →	s_{13} →	s_{14} →	s_{15} →	mechanism ⇓	species ⇒	CH_3	CH_4	O_2	CH_3OH	CO	H_2O	C_2H_6	CH_3OCH_3
																combinations⇒	5							
																origin ⇓	a_1	a_{10}	a_{11}	a_{12}	a_{13}	a_{14}	a_{15}	a_{16}
								1							→m_9	—	−2						1	
	1	1					1				1			1	→m_{27}	a_4: m_8+m_{22}	−2	1	−1		1	1		
	1		1	1		1	1								→m_{28}	a_4; m_8+m_{24}	2	−3	−1	1		1		
	1		1	1			1			1					→m_{29}	a_4: m_8+m_{25}		−2	−1			1		1
	2	1	1	1			2						1	1	→m_{30}	a_4: $2m_8+m_{26}$		−2	−2	1	1	2		
	1	1							1		1			1	→m_{31}	a_4: $m_{10}+m_{22}$	−4	2	−1	1	1			
	1		1	1		1			1						→m_{32}	a_4: $m_{10}+m_{24}$		−2	−1	2				
	1		1	1					1	1					→m_{33}	a_4: $m_{10}+m_{25}$	−2	−1	−1	1				1
	2	1	1	1					2				1	1	→m_{34}	a_4: $2m_{10}+m_{26}$	−4		−2	3	1			

FIG. 15. Setup of the methanol mechanism, after the elimination of a_4 (OH). Elimination of the only remaining intermediate, a_1 (CH_3), is carried out next. [Reprinted with permission from Mavrovouniotis, M. L., and Stephanopoulos, G. "Synthesis of reaction mechanisms consisting of reversible and irreversible steps: I. A synthesis approach in the context of simple examples". *Ind. Eng. Chem. Res.* **31**, 1625–1637. (1992). Copyright 1992 American Chemical Society.]

s_1 →	s_2 →	s_3 →	s_4 →	s_5 →	s_6 →	s_7 →	s_8 →	s_9 →	s_{10} →	s_{11} →	s_{12} →	s_{13} →	s_{14} →	s_{15} →	mechanism ⇓	species ⇒	CH_4	O_2	CH_3OH	CO	H_2O	C_2H_6	CH_3OCH_3
																origin ⇓	a_{10}	a_{11}	a_{12}	a_{13}	a_{14}	a_{15}	a_{16}
	1		1	1			1			1					→m_{29}	a_4: m_8+m_{25}	−2	−1			1		1
	2	1	1	1			2						1	1	→m_{30}	a_4: $2m_8+m_{26}$	−2	−2	1	1	2		
	1		1	1		1			1						→m_{32}	a_4: $m_{10}+m_{24}$	−2	−1	2				
	1		1	1		1	1	1							→m_{35}	a_1: $m_{28}+m_9$	−3	−1	1		1	1	
	2	1	1	1		1	2				1			1	→m_{36}	a_1: $m_{28}+m_{27}$	−2	−2	1	1	2		
	3	1	2	2		2	2		1		1			1	→m_{37}	a_1: $2m_{28}+m_{31}$	−4	−3	3	1	2		
	2		2	2		1	1		1	1					→m_{38}	a_1: $m_{28}+m_{33}$	−4	−2	2		1		1
	4	1	3	3		2	2		2				1	1	→m_{39}	a_1: $2m_{28}+m_{34}$	−6	−4	5	1	2		

FIG. 16. The final results for the methanol mechanism. [Reprinted with permission from Mavrovouniotis, M. L., and Stephanopoulos, G. "Synthesis of reaction mechanisms consisting of reversible and irreversible steps: I. A synthesis approach in the context of simple examples". *Ind. Eng. Chem. Res.* **31**, 1625–1637. (1992). Copyright 1992 American Chemical Society.]

s_1	s_2	s_3	s_4	s_5	s_6	s_7	s_8	s_9	s_{10}	s_{11}	s_{12}	s_{13}	s_{14}	s_{15}	mechanism	direct?	mechanism in Table X of Happel *et al.* (1990)
	1		1	1			1			1					m_{29}	√	not found, because s_{11} was omitted
	2	1	1	1			2						1	1	m_{30}	√	m_2/r_0; m_3/r_0 (lines 5 and 7)
	1		1	1		1			1						m_{32}	√	m_1/d_4; m_3/d_4; m_4/d_4; m_6/d_4 (lines 2,6,8,10)
	1		1	1		1	1	1							m_{35}	√	not found, because s_9 was omitted
	2	1	1	1		1	2				1			1	m_{36}	√	m_4/r_0 (line 7)
	3	1	2	2		2	2		1		1			1	m_{37}	$=m_{36}+m_{32}$	—
	2		2	2		1	1		1	1					m_{38}	$=m_{29}+m_{32}$	—
	4	1	3	3		2	2		2				1	1	m_{39}	$=m_{30}+2m_{32}$	—

FIG. 17. Analysis of the results for the methanol mechanism. The last three mechanisms that were produced (m_{37}, m_{38}, and m_{39}) are not direct; they can be formed from the direct ones. The last column shows the correspondence of the direct mechanisms to those of Happel *et al.* (1990). It is important to note that the same mechanism can be constructed many times in the procedure of Happel and Sellers (1983, p. 290)—as well as the simple algorithm presented here. It should also be noted that, in Table X of Happel *et al.* (1990), the mechanisms m_1/r_0 (line 1), m_2/d_4 (line 4), m_5/r_0 (line 9), m_5/d_4 (line 10), and m_6/r_0 (line 11) are all infeasible because they use either s_1 or s_5 in the wrong direction. The algorithm presented in this chapter never constructs mechanisms that violate the directionality of irreversible steps. [Reprinted with permission from Mavrovouniotis, M. L., and Stephanopoulos, G. "Synthesis of reaction mechanisms consisting of reversible and irreversible steps: I. A synthesis approach in the context of simple examples". *Ind. Eng. Chem. Res.* **31**, 1625–1637. (1992). Copyright 1992 American Chemical Society.]

examples presented by Mavrovouniotis (1989) and Mavrovouniotis *et al.* (1990, 1992). There are two differences between this view and the one adopted in the previous section. First, the classification of compounds as either intermediates or terminal species, introduced by H&S and used in Section II, does not distinguish between appearance of a compound as a reactant (raw material) and appearance as a product; in this section, we will permit separate specifications on net reactants and net products. Second, in Section II reversibility considerations were taken into account in the construction of mechanisms. Since most bioreactions are irreversible, we will treat reversible bioreactions as an exception, by splitting them into forward and reverse portions.

A. FEATURES OF THE PATHWAY SYNTHESIS PROBLEM

Biochemical pathway synthesis is the construction of consistent sets of enzyme-catalyzed bioreactions meeting certain specifications. One seeks to construct pathways which produce certain target bioproducts, under partial constraints on the available substrates (reactants), allowed byproducts, desired yield, productivity, etc. The pathway must include all reactions needed to convert initial substrates supplied to the bioprocess into final

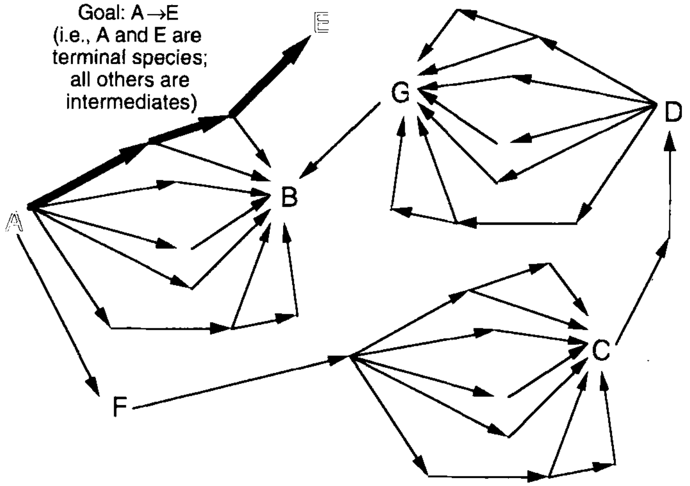

FIG. 18. The consequences of disregarding reaction irreversibility, in an abstract example. The species A and E are terminal and all others are intermediates. There are many parallel irreversible mechanisms that transform various intermediates to others (e.g., from A to B, from F to C, and from D to G). The direction of these internal mechanisms, however, is such that they do not participate in the final solution—which requires mechanisms only between A and E. There is actually only one mechanism (shown in thick arrows) that converts A to E in three steps. A method like that of H&S, which disregards directionality of steps and treats all steps as reversible initially, will come up with a very large number of mechanisms, going through the intermediates F, C, D, G, B (in that order). Our algorithm would *immediately* prune out the irrelevant portions of the network (starting with B) and identify the one feasible mechanism.

products that will be recovered from the process. Even if a computationally synthesized pathway is not present in any single known microorganism, it is of interest because of the possibilities offered by genetic engineering, which achieves the transfer of reactions and pathways from one microorganism to another. Through genetic engineering, a missing step can be inserted in a microorganism so that the synthesized new biochemical pathway (and hence a new bioprocess) is realizable.

Just like a direct reaction mechanism discussed in Section II, a pathway is not merely a set of reactions, because many distinct pathways can be constructed to include the same reactions but achieve different transformations. A pathway must include a *coefficient* (denoted by σ_{ki} in Section II) for each reaction, indicating the proportions in which the stoichiometries are combined. In this section, we will call these coefficients the *reaction stoichiometry* of the pathway. The overall transformation of net reactants to net products will be called the *molecular stoichiometry* of the pathway, and was denoted by β_{kj} in Section II.

Whereas in Section II we assumed that the participating steps will be given for a particular application, here it is useful to have a *database* with the most common enzymatic reactions and compounds, so that one only needs to add the specialized reactions for a given investigation.

B. FORMULATION OF CONSTRAINTS

A whole class of specifications in the synthesis of biochemical pathways can be formulated by classifying each compound and each reaction from the database according to the role it is required or allowed to play in the synthesized pathways' stoichiometries.

A given compound may occur in a pathway in any of three capacities: (1) as a net reactant or substrate of the pathway, (2) as a net product of the pathway, or (3) as an intermediate in the pathway. Constraints state whether a compound is allowed or possibly required to participate in the synthesized pathways in each of these three capacities.

For compounds as net reactants in the pathways, these conditions are inequality constraints on the coefficients β_{kj} (see Section II) that describe the participation of each species a_j in a pathway m_k. The specifications take the following forms:

(a) The species a_j is called a required reactant if it *must be consumed* by the pathway; this corresponds to a requirement that the coefficient β_{kj} be negative, i.e., $\beta_{kj} < 0$.
(b) A species is designated as an excluded reactant if it *cannot be consumed* by the pathway, and the corresponding restriction on the coefficient is $\beta_{kj} \geq 0$.
(c) Finally, the classification of a_j as an *allowed reactant* introduces no restriction on the coefficient β_{kj}.

In a realistic synthesis problem, the default constraint for compounds is that of excluded reactants, since most compounds in the database would not be available as economical raw materials.

The constraints for potential products of the pathways involve similar inequalities. A species a_j can be specified as (1) a required product, $\beta_{kj} > 0$, (2) an excluded product, $\beta_{kj} \leq 0$, which is the default, or (3) an allowed product, which involves no restriction on the sign of β_{kj}. Each compound a_j thus acquires one constraint from the reactant specifications and one from the product specifications. One of these combinations, however, the designation of the same compound as a required reactant ($\beta_{kj} < 0$) and a required product ($\beta_{kj} > 0$), is inconsistent. Of the remaining combinations, the designation of a compound as an allowed reactant and allowed product corresponds to the absence of any constraint on β_{kj}. The reader may compare these sets of constraints to the specification of intermediate and terminal species in Section II. There, we classified each species a_j using only two categories, as follows:

(a) If a species a_j is an intermediate, its coefficient in m_k is restricted to $\beta_{kj} = 0$. Clearly, this corresponds to the combination (conjuction) of an excluded reactant constraint ($\beta_{kj} \geq 0$) and an excluded product constraint ($\beta_{kj} \leq 0$).

(b) If a_j is a terminal species, its coefficient in m_k is not restricted at all. This corresponds to the combination of an allowed reactant and an allowed product.

Thus, in the case of biochemical pathways we have allowed a richer vocabulary for stoichiometric constraints but we can still reproduce the earlier specifications.

Notice the relation between the specification of a required reactant (negative stoichiometric coefficient—a strict inequality) and an excluded product (negative or zero stoichiometric coefficient—a loose inequality). In the operation of the algorithm, the first type of constraint is initially not fully satisfied; instead, the corresponding loose inequality constraint is satisfied in its place. Thus, for most of the phases in the algorithm, required reactants are treated merely as excluded products. As we will see, the distinction is eventually enforced, in the last phase of the algorithm. A similar observation holds for the strict inequality arising from required products, in relation to the loose inequality arising from excluded reactants.

A reaction can participate in pathways in either its forward or its reverse direction, giving rise to additional possible specifications: Reactions may be required, allowed, or prohibited to participate in the synthesized pathways in each of the two directions. Most reactions are likely to be *excluded* in the backward direction, because of prior knowledge about the (thermodynamic or mechanistic) irreversiblity of the reaction.

C. ALGORITHM

The algorithm for the synthesis of biochemical pathways follows closely the logic of the algorithm for the synthesis of catalytic reactions, i.e., it synthesizes biochemical pathways from a set of enzymatic reactions through an iterative satisfaction of constraints. A few minor differences reflect the richer vocabulary of specifications, given in the preceding section, and the biochemical context of compounds and enzymatic reactions.

Given a set of stoichiometric constraints and a database of biochemical reactions, the algorithm carries out iterative satisfaction of constraints, just like the algorithm in Section II. The algorithm proceeds as follows.

1. *Initialization and Reaction Processing*

The set M contains a one-step pathway for each individual enzymatic reaction available in the database, which has been compiled from known biochemistry. The set N contains the metabolites present in the enzymatic reactions of the set M. Naturally, these include the substrates (raw materials) that can be used, the desired products, and a large number of other compounds that occur in the bioreactions but will not serve as raw materials or desired products. This last set of compounds will carry the restrictions of excluded reactants and excluded products.

In order to account for the reversibility of some reactions, the algorithm decomposes reversible reactions into a forward and a backward step; the two directions are thereafter prohibited from participating together in the same pathway. This approach, unlike Section II, treats reversible reactions as the exception.

2. *Number of Combinations for Each Intermediate*

We note that the iteration through phases 2–4 was called compound-processing in the original description of this algorithm (Mavrovouniotis *et al.*, 1990). In these phases, partial pathways are gradually combined to satisfy the constraints (just as in Section II). This takes place as the algorithm tackles one compound at a time.

The objective of phase 2, in particular, is the selection of the most suitable compound for constraint satisfaction. We want to given priority to compounds that participate (as reactants or products) in only a small number of pathways that are active in the current state of the problem.

The computation of combinations (n_j as defined in Section II) facilitates this selection. Since all pathways are irreversible in our biochemical formulation, we only need to assemble (for each species a_j) the set Y_j, which contains all the partial pathways that include a_j as a net reactant, and the set Z_j, which contains all the partial pathways that include a_j as a net product. The cardinalities of these sets are y_j and z_j, respectively. It is clear that, in order to eliminate the metabolite a_j, we need to combine any pathway from Y_j, giving rise to $n_j = x_j y_j$ combinations.

3. *Elimination of an Intermediate Metabolite*

Partial pathways (which initially are just individual enzymatic reactions) are gradually combined in an effort to satisfy the constraints on the role of each metabolite. To minimize the computational effort, we choose in each iteration that metabolite that has the smallest number of combinations n_j. Let a_J be this metabolite. For the satisfaction of the constraints, a set M_J of new pathways is constructed through combinations of exactly one partial pathway from Z_J and exactly one pathway from Y_J. These pathways have the form

$$m_c = \beta_{bJ} m_k - \beta_{kJ} m_b,$$

where m_k are m_b are taken from the sets Y_J and Z_J, respectively (i.e., $m_k \in Y_J$ and $m_b \in Z_J$) and β_{bJ}, β_{kJ} are the net coefficients with which a_J participates in m_b and m_k, respectively. The net coefficient of a_J in the new pathway, m_c, is $\beta_{bJ}\beta_{kJ} - \beta_{kJ}\beta_{bJ} = 0$, verifying that, in the resulting new pathway, a_J has been eliminated.

A subtle precaution must be taken in the formation of the combinations, because of the way in which reversible reactions were decomposed into forward and reverse portions. Specifically, if m_k and m_p involve the same reaction (*any* reaction) *in different directions*, then the combination is rejected (or rather not formed at all).

4. *Update of Active Sets*

In order to modify the set of pathways such that the constraints on the compound a_J are met, we make use of the sets M_J, Y_J, and Z_J. The set of active pathways is modified by selecting the first applicable modification from the following list:

1. If a_J is an excluded product and a required or allowed reactant, $M := M \cup M_J - Z_J$.
2. If a_J is an excluded reactant and a required or allowed product, $M := M \cup M_J - Y_J$.
3. If a_J is an excluded reactant and an excluded product, $M := M \cup M_J - Y_J - Z_J$.
4. If none of the previous conditions describe the constraints on a_J, then M remains unchanged.

Then, we can remove the metabolite a_J and set $N := N - \{a_J\}$.

5. *Termination or New Iteration*

At this point, if no metabolite remains in N (i.e., if $N \neq \varnothing$), we can proceed to the final phase. If metabolites remain, then we return to phase 2, with a new computation for the selection of the next metabolite to process. In effect, the procedure in phases 3–4 is eventually carried out for all compounds.

6. *Pathway Processing*

The necessity for this phase, which did not exist in the algorithm of Section II, is a consequence of the richer vocabulary of stoichiometric constraints used for biochemical systems. The set of active pathways resulting from the processing of all metabolites as above satisfies all the requirements, except the constraints designating required reactants, products, or reactions. For these, only the corresponding loose-inequality constraints are satisfied; for example, instead of a required reactant and prohibited product, a compound will have been treated, up to this point, as an allowed reactant and prohibited product. The pathways satisfying the original stoichiometric constraints are those combinations of pathways from the final set that use at least one constituent pathway satisfying each of the strict constraints. The combination must thus include at least one pathway consuming each required reactant, at least one pathway producing each required product, etc.

D. EXAMPLES

We will discuss briefly some examples, to illustrate the features of the pathway synthesis algorithm and the results it has produced.

1. *Computational Efficiency*

The structured character of biochemical reaction networks is exploited by the synthesis algorithm in early pruning and abstraction, with significant gains in computational efficiency. This happens because the algorithm processes first those compounds and reactions that lead to few or no new combinations. This is shown schematically in Fig. 19, where some irrelevant portions of the network are pruned and some linear chains of reactions are compacted. We should note that this kind of reaction sequences are very common in biochemical reaction systems.

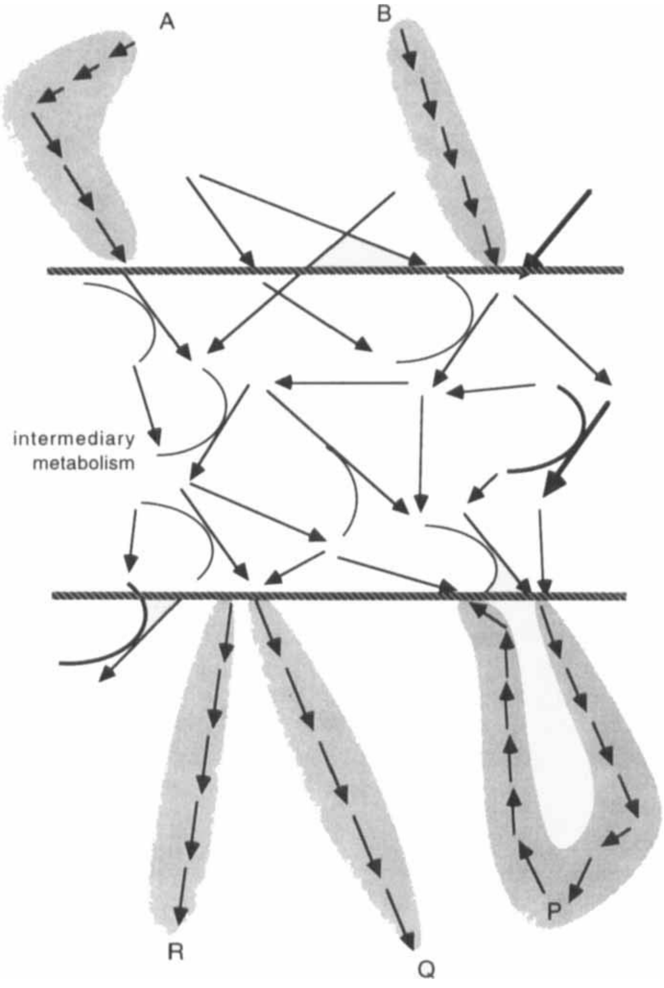

FIG. 19. An instance of efficient processing by the algorithm for the synthesis of biochemical pathways. Suppose that P, A, and Q are excluded reactants and excluded products; B is an allowed reactant; R is an allowed product; and all other compounds are excluded reactants and excluded products. In the early stages of the compound-processing phase, the algorithm will discard, one after the other, all reactions involved in the metabolism of B and R; it will construct the whole pathway that synthesizes and degrades P, discarding the individual reactions, and thus reducing the number of reactions/ pathways that are being considered; it will also construct pathways for the consumption of A and the production of Q, discarding the individual reactions for further reduction of the pathway space.

2. *Synthesis of Alanine*

The synthesis of alanine from glucose will be discussed here. In the database of 250 reactions used by Mavrovouniotis *et al.* (1992), six reactions involve glucose, and four involve alanine (*alanine dehydrogenase*, *methylserine hydroxymethyltransferase*, *alanine aminotransferase*, and *β-alanine aminotransferase*).

The initial formulation of the problem has glucose as a required reactant, alanine as a required product, NH_3 as an allowed reactant, and CO_2 as an allowed product. Additionally, a set of compounds that serve as internal currencies in the cell are designated as allowed reactants and allowed products. These are likely to occur in the problem specification, and they include NAD, NADH, NADP, NADPH, ATP, ADP, AMP, coenzyme-A (CoA), phosphate, and pyrophosphate. This initial formulation is actually too tight to generate a manageable number of pathways. By examining the constraints that are difficult to satisfy, we can modify the formulation of the problem to permit a solution. Here, the problem is resolved by designating the compounds malate and acetyl–CoA as allowed reactants and allowed products, reaching a formulation that constructs 1446 useful pathways.

Figures 20 and 21 show two pathways, suppressing many of the side reactants and side products for simplicity. Figure 20 depicts the normal

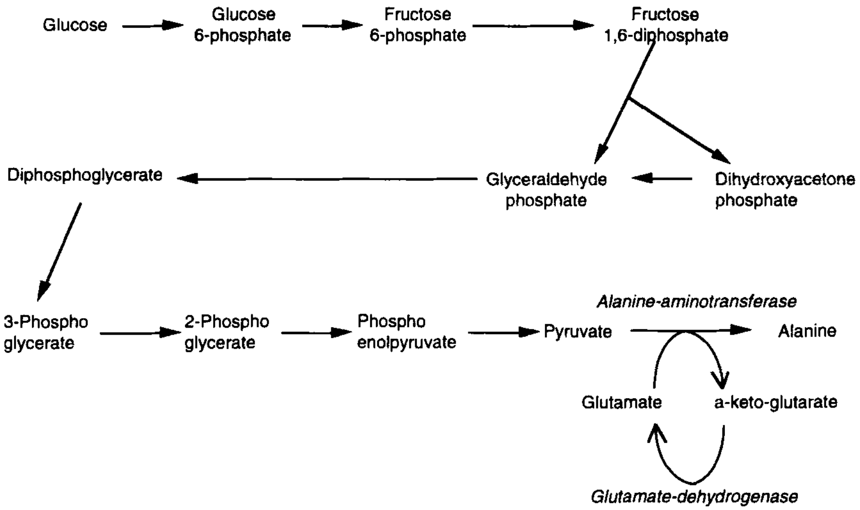

FIG. 20. Synthesis of alanine from glucose that follows the glycolysis pathway to pyruvate, which is in turn converted to alanine by *Alanine aminotransferase*. The glutamate required by this reaction is produced by *glutamate-dehydrogenase*.

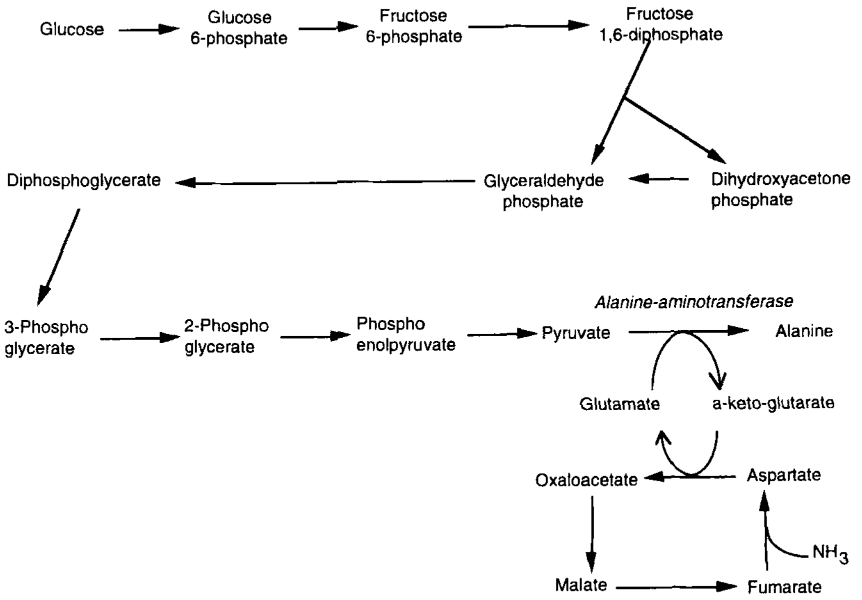

FIG. 21. Synthesis of alanine from glucose, with recovery of glutamate through a loop involving oxaloacetate, aspartate, fumarate, and malate.

pathway for the synthesis of alanine, with glucose as the main reactant. The pathway of Fig. 21 shows an alternative recovery of glutamate through a set of four reactions. The second pathway represents an interesting way for the cell to attach ammonia to an organic backbone; this is a variation that may be relevant for many aminoacid processes.

3. *Synthesis of Lysine*

In another study (Mavrovouniotis *et al.*, 1990), the biosynthesis of the aminoacid lysine was extensively investigated. In that case study, several reactions (shown in gray in Fig. 22) were assumed nonfunctional, and alternatives that avoid these steps were sought.

Figure 22 shows one of the resulting pathways, which involves the use of glyoxylate as an intermediate to bypass one of the nonfunctional steps. The pathway also includes conversion of malate to fumarate (which is the direction opposite to that normally used by microorganisms) and subsequent conversion of fumarate into aspartate and on to lysine. The use of the latter step is the key to the existence of this innovative pathway.

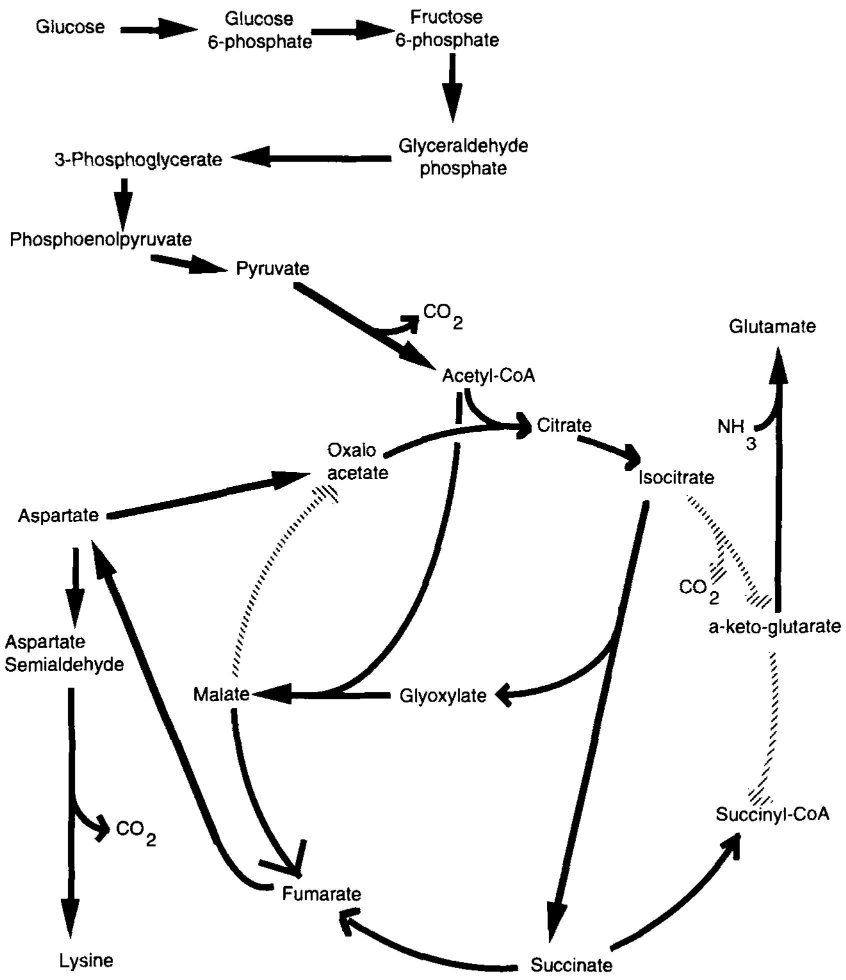

Fig. 22. A pathway for the production of lysine from glucose, drawn in a simplified form (many side reactants and side products are not shown, and many reaction sequences are lumped together).

IV. Properties and Extensions of the Synthesis Algorithm

Both forms of the algorithm (i.e., the basic version of Section II and the alternative of Section III) guarantee that each combination of pathways from the final set (derived by the synthesis algorithm) satisfies all the initial stoichiometric requirements. The correctness of the algorithm is based on the fact that each of the original requirements is satisfied in one of the iterations, and after a constraint is satisfied it is guaranteed to remain satisfied. As was pointed out earlier, constraints on required reactants or required products are satisfied in two phases. Initially, only the corresponding loose-inequality form is satisfied; the satisfaction of the strict-inequality version is ensured in the last phase (6) of the algorithm.

The second important property of the algorithm is its completeness. The synthesis algorithm creates a final set of pathways such that any pathway satisfying the original stoichiometry constraints is present in the final set. The basis of this property is that the elimination procedure is guaranteed to preserve legitimate pathways. A more detailed discussion of these properties is given by Mavrovouniotis *et al.* (1992).

We note that the algorithm may give rise to duplicate or indirect mechanisms (or pathways). Procedures can be applied in the end to remove such redundant mechanisms from the final set of mechanisms, or even to remove them incrementally during each iteration of the algorithm. These and other variations and implementational choices in the algorithm can be defined to enhance either its computational efficiency or its conceptual simplicity (Mavrovouniotis, 1992).

With respect to its computational complexity, the algorithm would require time, in the worst case, exponential in the number of reactions (Mavrovouniotis *et al.*, 1990). This worst-case complexity, however, is not encountered in practice, because the algorithm exploits the structure of biochemical networks, as was shown in Fig. 19. Another example, in Fig. 18, shows how the careful inclusion of directionality (irreversibility) considerations in the algorithm also serves to curb computational complexity.

A last feature of the algorithm that is worth discussing is its extendibility to other types of constraints. Additional specifications for the pathways or mechanisms under construction might arise from thermodynamics, kinetics, yield or productivity restrictions, biological regulation, etc. The question is whether such additional constraints can easily be incorporated into the synthesis procedure.

Some of these constraints are linear in nature. For example, a restriction on the endothermic or exothermic character of a pathway is a linear constraint on the enthalpy of reaction. If we take a linear combination of reactions (or pathways) the enthalpy of the resulting pathway is the corresponding linear combination of the enthalpies of the constituent reactions or pathways. For these linear situations, the quantity being constrained can be considered as merely another species, whose stoichiometry is subject to the type of constraints already tackled by the algorithm. Thus, the synthesis algorithm can deal with these specifications with little or no modification.

Other types of constraints, not corresponding to quantitative linear relations, can also be included, if they fulfill two requirements. The first requirement is that the constraint must possess the following convexity-like property: If the constraint is satisfied for two individual pathways, m_k and m_b, then it should be satisfied for any positive linear combination of these, $m_c = \mu_k m_k + \mu_b m_b$ (where $\mu_k > 0$, $\mu_b > 0$). The second requirement is that, given a set of partial mechanisms that do not satisfy the constraint, there must be a clear way to identify combination mechanisms that do satisfy it. These two requirements allow a constraint to be processed in much the same way that the stoichiometric constraints were, except that the details of construction of combinations would depend on the nature of the constraint. In one of the iterations, the constraint would be processed (through a procedure dictated by the second requirement), and it would thereafter remain satisfied (because of the first requirement).

V. Summary

In a chemical system we often discriminate between intermediate species and terminal species, and the latter are the only ones that appear in appreciable amounts in the stoichiometry of the overall transformation. Alternatively, we designate compounds as allowed or required to appear as reactants or products in the overall stoichiometry. These formulations are significant in the identification of quasi-steady-state behaviors or the synthesis of pathways to accomplish a desired transformation in a process.

We presented here a conceptual framework and algorithms for the synthesis of pathways or mechanisms given a set of steps. The algorithms have been applied to catalytic reaction systems and to biochemical pathways. The basic approach is based on successive processing and elimination of reaction intermediates that should not appear in the net stoichiometry of the overall reactions accomplished.

References

Happel, J., "Isotopic Assessment of Heterogeneous Catalysis." Academic Press, Orlando, FL, 1986.

Happel, J., and Sellers, P. H., Multiple reaction mechanisms in catalysis. *Ind. Eng. Chem. Fundam.*, **21**, 67–76 (1982).

Happel, J., and Sellers, P. H., Analysis of the possible mechanisms for a catalytic reaction system. *Adv. Catal.* **32**, 273–323 (1983).

Happel, J., and Sellers, P. H., The characterization of complex systems of chemical reactions. *Chem. Eng. Commun.*, **83**, 221–240 (1989).

Happel, J., Sellers, P. H., and Otarod, M., Mechanistic study of chemical reaction systems. *Ind. Eng. Chem. Res.* **29**· 1057–1064 (1990).

Horiuti, J., Theory of reaction rates as based on the stoichiometric number concept. *Ann. N.Y. Acad. Sci.* **213**, 5–30 (1973).

Horiuti, J., and Nakamura, T., On the theory of heterogeneous catalysis. *Adv. Catal.* **17**, 1–74 (1967).

Lee, L.-S., and Sinanoğlu, O., Reaction mechanisms and chemical networks—Types of elementary steps and generation of laminar mechanisms. *Z. Phys. Chem.* **124**, 129–160 (1981).

Mavrovouniotis, M. L., Computer-aided design of biochemical pathways. Ph.D. Thesis, Massachusetts Institute of Technology (1989).

Mavrovouniotis, M. L., Synthesis of reaction mechanisms consisting of reversible and irreversible steps: II. Formalization and analysis of the synthesis algorithm. *Ind. Eng. Chem. Res.* **31**, 1637–1653 (1992).

Mavrovouniotis, M. L., and Stephanopoulos, G., Synthesis of reaction mechanisms consisting of reversible and irreversible steps: I. A synthesis approach in the context of simple examples. *Ind. Eng. Chem. Res.* **31**, 1625–1637 (1992).

Mavrovouniotis, M. L., Stephanopoulos, G., and Stephanopoulos, G., Computer-aided synthesis of biochemical pathways. *Biotechnol. Bioeng.* **36**, 1119–1132 (1990).

Mavrovouniotis, M. L., Stephanopoulos, G., and Stephanopoulos, G., Synthesis of biochemical production routes. *Comput. Chem. Eng.* **16**, 605–619 (1992).

Milner, P. C., The possible mechanisms of complex reactions involving consecutive steps. *J. Electrochem. Soc.* **111**, 228–232 (1964).

Sellers, P. H., An introduction to a mathematical theory of chemical reaction networks I. *Archive for Rational Mech. Anal.* **44**, 23–40 (1971).

Sellers, P. H., An introduction to a mathematical theory of chemical reaction networks II. *Archive for Rational Mech. Anal.* **44**, 376–386 (1972).

Sellers, P. H., Combinatorial classification of chemical mechanisms. *SIAM J. Appl. Math.* **44**, 784–792 (1984).

Sellers, P. H., Combinatorial aspects of enzyme kinetics. In "Applications of Combinatorics and Graph Theory in the Biological and Social Sciences" (F. Roberts, ed.). Springer-Verlag, Berlin, 1989.

Sinanoğlu, O., Theory of chemical reaction networks. All possible mechanisms or synthetic pathways with given number of reaction steps or species. *J. Am. Chem. Soc.* **97**, 2309–2320 (1975).

Sinanoğlu, O., and Lee, L.-S., Finding all possible *a priori* mechanisms for a given type of overall reaction. *Theor. Chim. Acta* **48**, 287–299 (1978).

Temkin, M. I., The kinetics of transfer of labeled atoms by reaction. *Ann. N.Y. Acad. Sci.* **213**, 79–89 (1973).

Temkin, M. I., The kinetics of some industrial heterogeneous catalytic reactions. *Adv. Catal.* **26**, 173–291 (1979).

Yarlagadda, P. S., Morton, L. A., Hunter, N. R., and Gesser, H. D., Direct conversion of methane to methanol in a flow reactor. *Ind. Eng. Chem. Res.* **27**, 252–256 (1988).