What Information about the Conserved–Moiety Structure of Chemical Reaction Systems Can Be Derived from Their Stoichiometry?

2 AUTHORS:

Stefan Schuster
Friedrich Schiller University Jena
221 PUBLICATIONS **8,630** CITATIONS

SEE PROFILE

Claus C Hilgetag
University Medical Center Hamburg - Eppend…
102 PUBLICATIONS **5,747** CITATIONS

SEE PROFILE

What Information about the Conserved-Moiety Structure of Chemical Reaction Systems Can Be Derived from Their Stoichiometry?

Stefan Schuster* and Claus Hilgetag

Faculty of Mathematics and Sciences I, Chair of Theoretical Biophysics, Humboldt University Berlin, Invalidenstr. 42, D-10115 Berlin, Germany

Received: March 2, 1994; In Final Form: September 29, 1994[⊗]

The present paper is devoted to the question what conclusions about invariant building blocks (maximal conserved moieties) in chemical reaction systems can be derived from the stoichiometry of these systems. We define a concept of maximal conserved-moiety vectors in terms of those vectors that possess four essential properties: conservation and integer-element properties, nonnegativity, and nondecomposability, and examine the correspondence of these vectors to physically meaningful molecular units. In particular, we show that the set of all maximal conserved-moiety vectors (which turns out to be the complete set of fundamental nonnegative solutions to a system of linear diophantine equations) corresponds to the set of all potential maximal conserved moieties admissible for the given stoichiometry. Our interpretation of the term moiety to express predominantly conservation of the empirical formula (rather than invariability of the bond structure) is explained and discussed. We propose and substantiate an algorithm for computing all maximal conserved-moiety vectors and compare it with existing procedures for analyzing the moiety structure. The number of these vectors may be equal to, less than, less than, or greater than the number of linearly independent conservation relations. A number of chemical and biochemical examples illustrate our analysis.

1. Introduction

In every reaction mechanism, certain chemical units (atoms, ions, assemblies of atoms or ions) participate without loss of integrity.[1−4] These units are called conserved moieties, irreducible moieties, or building blocks. It is often of interest to find maximal conserved moieties, i.e., those building blocks not being part of larger conserved moieties.[5]

The moiety structure of closed reaction systems can be obtained by factorizing the atomic matrix,[5] which indicates how many atoms of every involved chemical element enter one molecule of every reacting substance.[6,7] This factorizing may yield multiple solutions. In many situations, however, the atomic matrix is not available. For example, in numerous chemical and biochemical reactions involving macromolecules, the atomic structure of some or all reactants is unknown. On the other hand, this information is unnecessary for detecting how many units of how many different, unspecified moieties enter the particular reacting substances.

It is a well-studied problem how the stoichiometric proportions, which enter the stoichiometry matrix, can be constructed from an atomic matrix.[7−9] In the present paper, however, we deal with the situation that a specific stoichiometry matrix, **N**, is given at the outset, suggesting by a kinetic mechanism, while the atomic composition of the participating chemical species may or may not be known. From knowledge of the stoichiometric proportions, considerable information can be derived. First, one can decide whether the system involves conservation relations, i.e., whether linear combinations of concentrations remain constant,[4,6,10] whether it is conservative (i.e., whether it is consistent with conservation of mass),[11−13] and whether a non-trivial steady-state is feasible.[4,14,15] Moreover, attempts have been made to derive information about the conserved-moiety structure.[2,13,16] In an earlier paper, we gave a method for determining the set of all conservation relations with nonnegative coefficients.[13] It was shown that all of these

relations can be written as a nonnegative linear combination of a finite set of extreme semipositive conservation relations, so-called because they can be represented mathematically as extreme rays or edges of a convex polyhedral cone. We argued that these extreme relations often, but not always, correspond to maximal conserved moieties. It is the purpose of the present paper to investigate this correspondence in a systematic way and in much more detail. We give a clear-cut definition of the concept of a maximal conserved-moiety vector. We then develop an algorithm for finding all of these vectors for any given reaction mechanism.
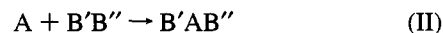
2. Maximal Conserved Moieties

Consider, for example, the hydration of maleic acid (or of fumaric acid) to malic acid,
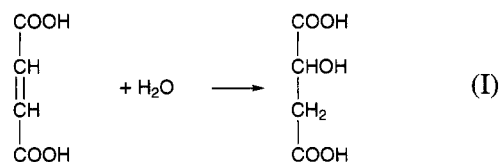
Obviously, the two carboxyl groups are conserved moieties. They are not, however, maximal, because also the maleic group, which encompasses the carboxyl moieties, is conserved. Moreover, the hydrogen and hydroxyl group in the water molecule are conserved moieties.

We can write reaction I in a formal notation as

$$A + B'B'' \rightarrow B'AB'' \tag{II}$$

with A, B', and B'' being conserved moieties and the sequence of symbols in B'AB'' referring to the fact that the hydroxyl group and hydrogen atom arrive at different locations in the molecule.

Whereas in reaction I the atomic composition and bond structure of the reacting species are known, this is not so in many association reactions involving protein molecules, e.g., the association of the α and $\beta\gamma$ subunits of the G-protein.[17]

This reaction can be written as in (II), with B′AB″ modified to AB′B″. This notation is an example of a so-called atom-free stoichiometry,[18] i.e., a stoichiometric equation using general symbols instead of names or structural formulas of concrete substances. The representation (II) can be simplified to
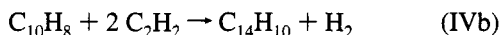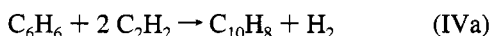
$$A + B \rightarrow AB \qquad\qquad (III)$$

Note that systems (I), (II), and (III) have the same stoichiometry matrix,

$$N = \begin{pmatrix} -1 \\ -1 \\ 1 \end{pmatrix} \qquad\qquad (2.1)$$

It is not clear from notation (III) whether the moieties B′ and B″ are always linked to each other. As this decision cannot be made from knowledge of the stoichiometry matrix alone, moieties B′ and B″ can formally be combined into one unit, B. We prefer therefore to interpret the condition of invariance of conserved moieties in a wide sense, in that the empirical formula of the moiety remains the same. The conserved moieties do not remain strictly unchanged in example (I) either, because the double bond in the maleic group transforms into a single bond. From now on, we will only consider maximal blocks of conserved moieties, even if parts of these moieties are located at different sites in the molecule but are transferred together in the reactions.

To demonstrate what information may be derived from the stoichiometry matrix, we consider the first two reactions of a mechanism for the formation of polycyclic aromatic hydrocarbons in a benzene flame:[9]

$$C_6H_6 + 2\,C_2H_2 \rightarrow C_{10}H_8 + H_2 \qquad\qquad (IVa)$$

$$C_{10}H_8 + 2\,C_2H_2 \rightarrow C_{14}H_{10} + H_2 \qquad\qquad (IVb)$$

The stoichiometry matrix reads

$$N = \begin{pmatrix} -1 & 0 \\ -2 & -2 \\ 1 & -1 \\ 1 & 1 \\ 0 & 1 \end{pmatrix} \qquad\qquad (2.2)$$

with benzene, acetylene, naphthaline, molecular hydrogen, and anthracine in this order represented by the rows from top to the bottom. One can calculate the following extreme semipositive conservation vectors, using the algorithm presented in our earlier paper[13] (see also section 3),

$$e_1 = (2\ 0\ 1\ 1\ 0) \qquad\qquad (2.3a)$$

$$e_2 = (1\ 0\ 1\ 0\ 1) \qquad\qquad (2.3b)$$

$$e_3 = (0\ 1\ 2\ 0\ 4) \qquad\qquad (2.3c)$$

$$e_4 = (0\ 1\ 0\ 2\ 0) \qquad\qquad (2.3d)$$

The vector e_3, for example, can be interpreted in that the linear combination of concentrations $[C_2H_2] + 2[C_{10}H_8] + 4[C_{14}H_{10}]$ remains constant in time. All non-negative conservation vectors for the considered system are nonnegative linear combinations of e_1 to e_4. While e_1 is not immediately amenable to physical interpretation, e_2 to e_4 correspond to conservation of the phenyl group, the C_2H group stemming from the acetylene and the hydrogen atom transferred from acetylene to H_2, respectively. Note that the vectors e_i have been calculated without using any knowledge about the atomic structure of the participating

substances. Interestingly, the mechanism (IV) involves more maximal conserved moieties than chemical elements. Other examples with this property are the hydrodealkylation reaction of toluene[7,19] and a system in which only mass transfer of some substances between two phases occurs.[7] This situation also occurs in biochemistry, where enzymes may couple reactions.

For complex mechanisms (e.g., for the complete mechanism of formation of polycyclic aromatic hydrocarbons), it is very difficult to detect the conserved moieties by inspection. A mathematical treatment suitable for use on computer would therefore be very helpful.

The question arises whether, for arbitrary reaction mechanisms, the set of extreme semipositive conservation vectors contains vectors corresponding to maximal conserved moieties. Generalizing considerations by Park Jr.,[16] we first give a definition to formalize the concept of conserved moiety from a purely stoichiometric viewpoint. We regard the conserved moieties (which we denote by M_k), as some physicochemical entities unspecified in their concrete structure. The constitution of a reaction system by these atom groups can be written in terms of vectors, whose elements, z_i, indicate how many entities of a moiety are contained in one molecule of the reacting species i.

Let the symbol **0** stand for the null matrix of suitable dimension.

Definition 1. For a given reaction mechanism, a row vector **z** with n components, z_i (n denotes the number of reacting species), is called a maximal conserved-moiety vector if, and only if, it has the following four properties:

(P1) conservation property, $zN = 0$ (2.4)

(P2) integer-element property, z_i are integers for all i (2.5)

(P3) non-negativity, $z_i \geq 0$ for any i, $z \neq 0$ (2.6a,b)

(P4) non-decomposability (maximal size of moiety);

$$\qquad\qquad z \text{ cannot be written as}$$

$$z = \eta' z' + \eta'' z'', \quad \eta', \eta'' \geq 1, \text{ integer} \qquad (2.7)$$

with z' and z'' fulfilling conditions (P1)−(P3).

Note that if the conserved moieties coincide with single atoms, eq 2.4 means that the product of the atomic matrix and stoichiometry matrix equals zero.[6−8] Condition (P2) follows from the fact that the invariant moieties are indivisible by definition, so that they constitute the reacting substances in integer proportions. The mathematical description of the moiety structure therefore requires integer solutions, and we will have to use the theory of diophantine equations.[20,21] In contrast, relations expressing conservation of mass involve the molecular weights, which are positive real numbers.[11,12,16]

To explain the meaning of condition (P4), we have to consider two cases according to whether or not z' and z'' are identical (cf. Figure 1). If condition (P4) were not fulfilled in the former case, $\eta' + \eta''$ entities of some moiety could be combined into one moiety, which then would be larger than the one corresponding to z.

We now address the case that z' and z'' are different. Consider again the stoichiometry matrix (2.1). The vector (1 1 2), for example, can be written as the sum of (1 0 1) and (0 1 1). The latter two vectors fulfil properties (P1)−(P4). Then the moieties corresponding to the vectors (1 0 1) and (0 1 1) can each be combined with one unit of the moiety corresponding to the vector (1 1 2), so that larger moieties obtain (cf. Figure 1). Strictly speaking, this is true as long as at least one of the
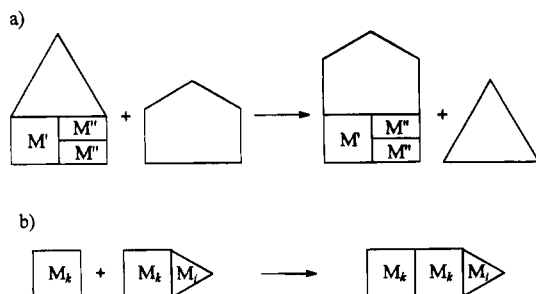
Conserved-Moiety Structure of Reaction Systems

J. Phys. Chem., Vol. 99, No. 20, 1995 **8019**

Figure 1. Schematic illustration of condition (P4). (a) The moieties M' and M'' correspond to the vectors (1 0 1 0) and (2 0 2 0). Since M' and the two entities of M'' always occur together and in the same proportion, they can be combined into one maximal conserved moiety, which is represented by the moiety vector (1 0 1 0). (b) The moieties M_k and M_i correspond to the vectors (1 1 2) and (0 1 1). Those units of M_k and M_i that are transferred together can be combined into one maximal conserved moiety, which is represented by the moiety vector (0 1 1). Therefore, the vector (1 1 2) splits into the sum (0 1 1) + (1 0 1), the latter term of which corresponds to that unit of M_k that is not combined with M_i.

two former moiety vectors has a physical representation in the system. If this is not so, isomerization reactions have been neglected in the model. Consider the following reaction mechanism
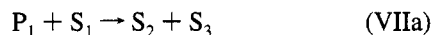
$$S_1 \rightarrow S_2, \quad S_1 + S_2 \rightarrow S_3 \qquad \text{(Va,b)}$$
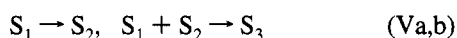
By methods derived below, one obtains the unique maximal conserved-moiety vector (1 1 2). It correctly reflects the fact that S_1 and S_2 are isomers. (For the case that external reactants participate in reaction Va, see the treatment of open systems below.) If the first reaction is so slow that it can be neglected in the scheme, only reaction Vb is left. This reduced system is characterized by two conserved-moiety vectors, (1 0 1) and (0 1 1). These vectors would imply that the moieties in S_1 and S_2 are different. They have been assigned also to mechanism (III), which is equivalent to (Vb). Only when including reaction Va, we obtain a moiety structure that reflects the chemical similarity of S_1 and S_2. This example shows that, occasionally, it cannot be decided from the stoichiometry matrix alone whether some moieties are identical or different. To distinguish between the two cases, one needs information about the molecular structure of the substances involved or about neglected isomerization reactions.

For the polymerization reaction

$$A + (A)_m \rightarrow (A)_{m+1} \qquad \text{(VI)}$$
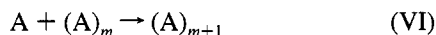
the vector (1 m m+1) satisfies conditions (P1)−(P3). It can, however, be decomposed into the vector (1 0 1) and m times the vector (0 1 1), which correspond to the moieties A and $(A)_m$ and fulfil property (P4). They are more consistent with our approach of detecting maximal preserved units than the vector (1 m m+1).

The following example shows that the present analysis also applies to open systems.

$$P_1 + S_1 \rightarrow S_2 + S_3 \qquad \text{(VIIa)}$$

$$S_3 \rightarrow P_2 \qquad \text{(VIIb)}$$

$$P_3 + S_2 \rightarrow S_1 + P_4 \qquad \text{(VIIc)}$$

The P_i denote species with concentrations kept or considered constant. It is sometimes convenient to ignore such "external" species in limited model equations that summarize the behavior of a small number of species in the overall reaction system.[11] Closed systems, by definition, do not contain any external species. In the above example, S_1 and S_2 may stand for ATP
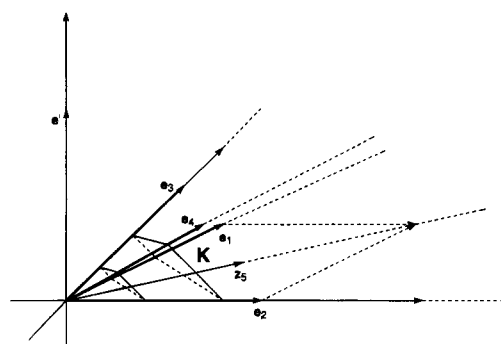

Figure 2. Three-dimensional cone of semi-positive conservation relations for a system with the stoichiometry matrix $N = (-2 \ -1 \ 1 \ 1)^T$. $e_1 = (1 \ 0 \ 2 \ 0)$, $e_2 = (1 \ 0 \ 0 \ 2)$, $e_3 = (0 \ 1 \ 1 \ 0)$, and $e_4 = (0 \ 1 \ 0 \ 1)$ are maximal conserved-moiety vectors and generating vectors of the cone. In addition, the maximal conserved-moiety vector $z_5 = (1 \ 0 \ 1 \ 1)$ arises, which lies on a face of the cone (cf. section 4). In this picture, the four-dimensional space of conservation vectors is projected into a three-dimensional space spanned by the vectors e_2, e_3, and $e' = (-4 \ 5 \ -5 \ 2)$, which is orthogonal to e_2 and e_3. Note that the semipositivity condition need not be fulfilled in the transformed space.

and ADP. The (here unique) conserved-moiety vector is (1 1 0). It appropriately reflects conservation of the adenosine group contained in ATP and ADP.

Definition 2. A matrix **Z** with n columns is called complete conserved-moiety matrix, if, and only if, it contains, as rows, all vectors fulfilling definition 1 and has the property

(P5) dissimilarity, all row vectors of **Z** are different:

$$z_{i\cdot} \neq z_{k\cdot} \text{ for all } i, k \qquad (2.8)$$

with $z_{i\cdot}$ and $z_{k\cdot}$ denoting the ith and kth columns of **Z**.

Reaction II illustrates the relevance of condition (P5). If the moieties B' and B'' were considered separately, two identical vectors **z** would occur. Condition (P5) thus reflects the approach to combine into one unit those moieties that are "inherited" together.

3. Extreme Semipositive Conservation Relations

For many reaction systems, conservation relations hold true. These relations state that linear combinations of concentrations remain constant, and they arise if there is a matrix, **B**, that fulfills the equation[4,6,10,22]

$$\mathbf{BN} = \mathbf{0} \qquad (3.1)$$

The existence of this conservation matrix **B** is guaranteed, if

$$\varrho = \text{rank}(\mathbf{N}) \qquad (3.2)$$

is smaller than n. Note that matrix **B** is primarily defined in an algebraic way by eq 3.1 and need not equal the atomic matrix **A** nor the complete conserved-moiety matrix **Z**. To represent conserved moieties, **B** may only contain nonnegative elements. We can speak of semipositive conservation relations, because we exclude null vectors. If the description is to include the conservation of electric charge, which may be negative, the entries can nevertheless be transformed to be nonnegative.[13] The set of all nonnegative vectors $b_{k\cdot}$ is therefore determined by a linear equation/inequality system, so that it forms a convex polyhedral cone, **K**.[23] As shown in convex analysis,[23] such a cone can be represented as a nonnegative linear combination of generating vectors, e_k,

$$\mathbf{K} = \{\mathbf{y} \in \mathbf{R}^n | \mathbf{y} = \sum_{k=1}^{g} \eta_k \mathbf{e}_k, \quad \eta_k \geq 0, \quad k = 1, ..., g\} \qquad (3.3)$$

(cf. Figure 2). Clarke[4] called this linear combination an external

representation in contrast to the internal representation as given by eq 3.1 together with the nonnegativity condition. The generating vectors (extreme vectors, edges) are those vectors that cannot be expressed as nonnegative linear combinations of other vectors of the cone.[23] A complete set of such vectors can be computed by using an algorithm described earlier,[13] which is based on a method given by Nožička et al.[24] The number of generating vectors can be greater than, or equal to, the dimension of the cone, which in turn can be smaller than, or equal to, $n - \varrho$.

Extending the algorithm developed earlier,[13] we shall impose that for each generating vector, the greatest common divisor of its components be unity, which can easily be fulfilled by appropriate reduction. The vectors thus obtained shall be called reduced generating vectors.

We now show that all reduced generating vectors fulfil properties (P1)−(P4). (P1) is fulfilled owing to eq 3.1. (P2) is granted because the algorithm for calculating the generating vectors[13] starts with integers and contains, as operations, only integer multiplication and summation. Non-negativity, i.e., (P3), is guaranteed because the cone $\mathbf{K}$ contains only nonnegative vectors. (P4) is fulfilled, in the case $\mathbf{z}' \neq \mathbf{z}''$, on account of the above definition of generating vectors. In the case $\mathbf{z}' = \mathbf{z}''$, it is satisfied because the vectors have been reduced to have the smallest integer components possible.

All reduced generating vectors are admissible row vectors of the conserved-moiety matrix $\mathbf{Z}$, since they meet properties (P1)−(P4). For many reaction mechanisms, finding all reduced generating vectors is sufficient to obtain all maximal conserved-moiety vectors (e.g., mechanisms I and III−VII).

The vectors given in eq 2.3 are the generating vectors of cone $\mathbf{K}$ for the mechanism (IV). That example shows that there may be more generating vectors than actual conserved moieties. Strictly speaking, the vectors $\mathbf{z}$ introduced in definition 1 are in a sense *potential* conserved-moiety vectors. This is in line with assertions of other authors[2,16] that there may exist more than one moiety structure that conforms with a particular stoichiometry.

For closed reaction systems, a possible way to detect which conserved-moiety vectors have a physical meaning in a concrete mechanism is by using knowledge of the molecular weights, w_i, of reacting species.[16] Let the w_i be assembled in a vector $\mathbf{w}$, and the molecular weights of the maximal conserved moieties, m_j, in the vector $\mathbf{m}$. We than have

$$\mathbf{w} = \mathbf{Z}^{\mathrm{T}}\mathbf{m} \qquad (3.4)$$

Only the nonzero components of the solution, $\mathbf{m}$, of this equation correspond to a conserved moiety. Note, however, that a unique solution only exists if the number of conserved-moiety vectors is less than, or equal to, $n - \varrho$. Furthermore, this method is not applicable to open systems and it is not robust to errors in measurement of $\mathbf{w}$, since for solutions m_j close to zero one cannot decide whether it corresponds to a small moiety (such as H^+) or to no moiety at all. At any event, this point is beyond the scope of our paper, since we wish to use knowledge of the stoichiometry matrix only.

4. Determining the Maximal Set of Conserved-Moiety Vectors

While it is easy to check whether a given set of conserved-moiety vectors (e.g., $\mathbf{e}_1$−$\mathbf{e}_4$ for the mechanism IV) fulfils properties (P1)−(P3), this is not so with respect to condition (P4). It is therefore desirable to have a procedure, based on

definition 1, for computing all conserved-moiety vectors in a constructive, algorithmic way. Such a procedure will be developed in this section.

There exist maximal conserved moieties that do not directly correspond to reduced generating vectors, as an example with the following stoichiometry matrix shows:

$$\mathbf{N} = \begin{pmatrix} -2 \\ 1 \\ 1 \end{pmatrix} \qquad (4.1)$$

The reduced generating vectors are $(1\ 2\ 0)^{\mathrm{T}}$ and $(1\ 0\ 2)^{\mathrm{T}}$. These correspond to the maximal conserved moieties in the mechanism

$$2AB \rightarrow A_2 + B_2 \qquad (\text{VIII})$$

but not in the system

$$2AB \rightarrow AB_2 + A \qquad (\text{IX})$$

which has the same stoichiometry matrix and may stand for the reaction converting carbon monoxide to carbon dioxide and carbon. For the latter mechanism, the conserved-moiety vectors read

$$\mathbf{z}_1 = (1\ 1\ 1), \quad \mathbf{z}_2 = (1\ 2\ 0) \qquad (4.2\text{a,b})$$

Note that $\mathbf{z}_1$ is half the sum of the two generating vectors given above. Generalizing this result, one can suppose that any nonnegative linear combination of generating vectors that can be divided by an integer greater than unity (which is prime relative to the coefficients of the linear combination) is a potential moiety vector.

From definitions 1 and 2, we derive the following theorem. The set of all maximal conserved-moiety vectors is the minimal set of vectors in terms of which any nonnegative solution, $\mathbf{b}$, of eq 3.1 with integer components can be expressed by nonnegative, integer linear combination,

$$\mathbf{b} = \sum_k \eta_k \mathbf{z}_{k'}, \quad \eta_k \geq 0, \text{ integer} \qquad (4.3)$$

Proof: Any nonnegative solution, $\mathbf{b}$, of eq 3.1 with integer components fulfills properties (P1)−(P3). If it also satisfies (P4), eq 4.3 is trivially fulfilled with $\mathbf{b} = \mathbf{z}_{i'}$. If some vector $\mathbf{b}$ does not meet condition (P4), it would be decomposable into two vectors $\mathbf{z}'$ and $\mathbf{z}''$, fulfilling properties (P1)−(P3). If $\mathbf{z}'$ and $\mathbf{z}''$ satisfy (P4), they are conserved-moiety vectors, so that $\mathbf{b}$ can be written as in eq 4.3. If $\mathbf{z}'$ (or $\mathbf{z}''$) does not satisfy (P4), it can in turn be decomposed, and the same reasoning applies again. This recursion will end after a finite number of steps, yielding maximal conserved-moiety vectors, since in each decomposition step, η' and η'' are greater than, or equal to, unity, and the vectors have to remain nonnegative.

Now suppose that the set of moiety vectors is not minimal, i.e., that we can cancel at least one vector, $\mathbf{z}_{i'}$, so that the remaining vectors are sufficient to generate all solutions $\mathbf{b}$. Setting $\mathbf{b} = \mathbf{z}_{i'}$, we see that this assumption leads to a contradiction, since $\mathbf{z}_{i'}$ cannot be represented by integer linear combination of other moiety vectors because of (P4). This completes the proof.

In the theory of diophantine equations, a complete set of solutions, in terms of which all other solutions can be expressed by integer linear combination is called a fundamental set.[10,20,21] From the above theorem, it follows that the rows of the complete conserved-moiety matrix $\mathbf{Z}$ form a fundamental set of semi-positive solutions to eq 3.1. For systems of homogeneous diophantine equations without the semi-positivity restriction, an algorithm for detecting the fundamental set of solutions has been

Conserved-Moiety Structure of Reaction Systems

J. Phys. Chem., Vol. 99, No. 20, 1995 **8021**

known for a long time.[25] This set contains $n - \varrho$ different vectors. Under the semipositivity restriction, the number of fundamental solutions can be greater than $n - \varrho$, because only nonnegative integer combinations of the fundamental solutions are admissible (cf. eq 4.3). This can be seen with the help of the stoichiometry matrix (4.1), which has the fundamental set (1 0 2), (1 2 0), (1 1 1).

For every closed system, at least one maximal conserved-moiety vector exists owing to mass conservation, as can be seen from eq 3.4.[11-13] For open systems, there may be no such vector, for example, if it does not involve any conservation relation.

Bombieri and Vaaler[26] showed that the number of fundamental semi-positive solutions to homogeneous diophantine equation systems is always finite or zero. Results have also been obtained on sharp bounds on these solutions.[27,28] To our knowledge, however, an algorithm for finding the complete set of fundamental solutions has not yet been developed. Below, we propose such an algorithm and give its justification.

Algorithm for Determining the Complete Conserved-Moiety Matrix.

1. Apply the algorithm for finding a complete set of generating vectors of cone $\mathbf{K}$.[13,24] The matrix $\mathbf{T}^{(r)}$ (called tableau) obtained by this algorithm contains a null submatrix with r columns (r denotes the number of reactions). Cancel this submatrix.

2. Reduce each generating vector by the greatest common divisor of its components.

3. Calculate additional tableaux $\mathbf{T}^{(m)}$ ($m = r+1, r+2, ...$) by the following steps:

3a. Find all pairs of vectors $\mathbf{t}_{i\cdot}^{(m)}$, $\mathbf{t}_{k\cdot}^{(m)}$ for which one of these has a zero entry at a position where the other has not, i.e.,

$$t_{ij}^{(m)} = 0, \, t_{kj}^{(m)} \neq 0 \text{ for at least one } j \qquad (4.4a)$$

or

$$t_{ih}^{(m)} \neq 0, \, t_{kh}^{(m)} = 0 \text{ for at least one } h \qquad (4.4b)$$

3b. Determine $t_{min} = \min(t_{kj}^{(m)}, t_{ih}^{(m)})$ for each of these pairs and for all j, h with (4.4a) and (4.4b), respectively, being fulfilled.

3c. Test for all pairs of vectors found in 3a, and for all pairs of positive integers η' and η'' that are relatively prime and not greater than t_{min}, whether all components of the combined vector $\eta' \mathbf{t}_{i\cdot}^{(m)} + \eta'' \mathbf{t}_{k\cdot}^{(m)}$ can be divided by the same, largest possible integer, μ, to give integer components again:

$$\hat{z}_j = \frac{\eta' t_{ij}^{(m)} + \eta'' t_{kj}^{(m)}}{\mu}, \, \mu > 1, \text{ integer, } \hat{z}_j \text{ integer for all } j$$

$$(4.5)$$

Gather all vectors $\hat{\mathbf{z}}$ in the tableau $\mathbf{T}^{(m+1)}$. If the same vector $\hat{\mathbf{z}}$ obtains twice or more, select it only once.

3d. The algorithm is completed when the rows of some $\mathbf{T}^{(m)}$ cannot be combined to yield further vectors $\hat{\mathbf{z}}$. Finally, we assemble all tableaux $\mathbf{T}^{(m)}$ ($m = r, r+1, ...$) to obtain the complete conserved-moiety matrix $\mathbf{Z}$.

The justification of this algorithm is given in the Appendix.

Extending the program CORE developed earlier,[13] we wrote the Turbo-Pascal program COMO (COnserved MOieties), which executes the algorithm set out above. We tested this program for a large variety of mechanisms.

For example, for mechanisms having the stoichiometry matrix

$$\mathbf{N} = \begin{pmatrix} -3 \\ 1 \\ 1 \\ 1 \end{pmatrix} \qquad (4.6)$$

we can write the generating (row) vectors as (1 3 0 0), (1 0 3 0), and (1 0 0 3). They can be combined into the fundamental solutions (1 1 0 2), (1 2 0 1), (1 0 1 2), (1 0 2 1), (1 1 2 0), and (1 2 1 0), which form the tableau $\mathbf{T}^{(2)}$ in the algorithm. The first and fifth of these additional vectors give the fundamental solution (1 1 1 1), which is the only row of the tableau $\mathbf{T}^{(3)}$. Combination of other vectors does not yield any further solution.

For the mechanism (IV), the extended algorithm gives a set of fundamental solutions which contains the vectors $\mathbf{e}_i$ given in eq 2.3 and the vector

$$\mathbf{z}_5 = (0 \, 1 \, 1 \, 1 \, 2) \qquad (4.7)$$

which is half the sum of $\mathbf{e}_3$ and $\mathbf{e}_4$ and is, hence, not a generating vector of cone $\mathbf{K}$. $\mathbf{z}_5$ does not, however, correspond to a conserved moiety for the considered mechanism.

There are reactions for which only nonextreme vectors correspond to conserved chemical units, e.g.,[29]

$$3HCl \rightarrow H_2Cl^+ + HCl_2^- \qquad (X)$$

This reaction has the maximal conserved moieties H^+ and Cl^-. They are represented by the vectors (1 2 1) and (1 1 2), neither of which is a generating vector.

5. Conclusions

Whereas other authors used an intuitive notion of conserved moiety, we have suggested a mathematical definition for this concept, which is applicable both to closed and open reaction systems. We have coined the term maximal conserved-moiety vector and have invoked that it fulfills the conservation equation, contains integer, nonnegative elements only, and satisfies a nondecomposability condition.

The complete conserved-moiety matrix, $\mathbf{Z}$, containing all maximal conserved-moiety vectors for a given mechanism, is related to the modified formula matrix, $\mathbf{A}'$, proposed by Smith and Missen.[7] They had, however, devised that matrix with a different purpose. In fact, the two matrices differ in the following aspects: (a) the matrix $\mathbf{A}'$ includes, as a submatrix, the atomic matrix, $\mathbf{A}$, while matrix $\mathbf{Z}$ involves only those rows of $\mathbf{A}$ that correspond to atoms being maximal conserved moieties; (b) matrix $\mathbf{Z}$ is always nonnegative, while the additional rows in $\mathbf{A}'$ (i.e., those not contained in $\mathbf{A}$) may have negative elements, which are difficult to interpret; (c) the matrix $\mathbf{A}'$ contains $n - \varrho$ (linearly independent) rows, while matrix $\mathbf{Z}$ may contain more rows (because some row can correspond to an irreducible moiety although it is a linear combination of other rows) or less rows (because rows with negative entries are not allowed).

We have proven that all extreme semipositive conservation relations (expressed by generating vectors of a convex cone) fulfil the four properties (P1)−(P4) used to define the maximal conserved-moiety vectors, provided that each of them has been divided by an integer so that its components are relatively prime. In addition, there may be more vectors fulfilling the four properties, which can be obtained from the generating vectors by linear combination.

Our results imply interesting epistemological issues. First, we have to decide to what extent features of the molecular

structure of a conserved moiety should be conserved. The empirical formula (i.e., with no reference to structure) of the moiety should no doubt remain constant. As far as the bond structure is concerned, changes necessarily have to be admitted since the bonds to the adjacent moieties alter. Whether multiple bonds inside a moiety are allowed to change to single ones or vice versa is subject to personal preference. We have decided to allow for such changes. Going a step further, we may even allow for complete breaking of bonds inside a moiety. There are two ways of interpreting the conserved-moiety vectors. We may consider such a vector as a representative of a class of conserved moieties, such as the hydrogen and hydroxyl group in reaction I. On the other hand, if one is interested in having a one-to-one correspondence between vectors and moieties, each conserved-moiety vector should be considered to represent one moiety only, thus considering conservation in terms of the empirical formula, regardless of the fact that the moiety might consist of several subunits at different sites in the molecule.

Second, for mechanisms including isomerization reactions, isomers may or may not be identified with different moieties, depending on whether or not the isomerization reaction is neglected in the model. Following our approach, one should therefore include as many reactions as possible, especially all isomerization reactions.

Third, some conserved-moiety vectors calculated from stoichiometry alone may not correspond to molecular units. The mathematical representation may thus include more potential solutions than necessary to reflect physical reality. This situation is similar to the multiple solutions of polynomial equations, where usually only some of them are physically meaningful.

The question posed in the title can be answered as follows. The full set of maximal conserved-moiety vectors coincides with a complete set of fundamental nonnegative solutions to a system of linear diophantine equations derived from the stoichiometry matrix. This set can be calculated by an algorithm based on a method for finding the generating vectors of a convex cone. The algorithm includes tests of whether integer linear combinations of two generating vectors can be reduced by common integer factors to give additional moiety vectors and whether these vectors, in turn, can be combined to give additional moiety vectors, and so on. In each step, only those pairs of vectors need to be considered that have a different pattern of zero components. A computer program COMO performing this algorithm has been written in Turbo-Pascal by the authors and is available on request.

The present paper widens earlier approaches and procedures. The algorithms presented by Park[2] and Park Jr.[16] do not take advantage of the facts that all moiety vectors lie in a convex polyhedral cone and that all reduced extreme vectors of this cone are conserved-moiety vectors. Furthermore, these algorithms are incomplete because they are based on the erroneous assumption that the number of maximal conserved moieties could not be greater than the dimension of the vector space of conservation relations $(n - \varrho)$. Indeed, Clarke[4] used the theory of convex polyhedra but dealt with the different problem of determining the admissible region for concentrations (concentration polyhedron).

Our algorithm is related to an earlier procedure for finding all extreme nonnegative flux vectors.[15] Other algorithms for computing these vectors[14,30] are much less efficient because their running time increases rapidly with the number of reactions.

Appendix. Substantiation of the Proposed Algorithm

Since all reduced generating vectors of cone **K** are fundamental solutions, the algorithm for finding these generating

vectors can serve as a starting point. If the system has fundamental solutions other then the generating vectors, these are situated within cone **K** or on one of its faces. One can obtain them by intersecting the cone with a hyperplane, which is given by a linear equation. Additional moiety vectors can therefore be found by adding an extra column to the stoichiometry matrix. For example, augmenting the stoichiometry matrix of mechanism (VIII) by the column $(0\ -1\ 1)^T$ yields

$$\mathbf{N}' = \begin{pmatrix} -2 & 0 \\ 1 & -1 \\ 1 & 1 \end{pmatrix} \qquad (\text{A}.1)$$

and $\mathbf{z} = (1\ 1\ 1)$, which is the sought additional moiety vector. We thus have to calculate tableaux $\mathbf{T}^{(m)}$ ($m = r+1, r+2, ...$) to an extended stoichiometry matrix. We circumvent the difficulty of finding the appropriate extra columns of **N** by the following reasoning. In step m of the standard algorithm ($m \le r$), some pairs of row vectors of $\mathbf{T}^{(m-1)}$ are linearly combined with each other. We now do the same with the tableaux $\mathbf{T}^{(m)}$ ($m = r, r+1, ...$), but with a larger number of pairs of row vectors ($\mathbf{t}_{i\cdot}^{(m)}$, $\mathbf{t}_{k\cdot}^{(m)}$) and all suitably chosen positive integer weighting factors η' and η'', which are to be relatively prime. In testing all combinations of row vectors that fulfil condition (4.4), we need not construct particular additional columns of **N**.

The vector resulting from eq 4.5 obviously fulfills properties (P1)−(P3). It fulfills property (P4) with $\mathbf{z}' = \mathbf{z}''$, since it has been reduced by the greatest possible integer μ. It is also understandable that the vector $\hat{\mathbf{z}}$ fulfills property (P4) with $\mathbf{z}' \ne \mathbf{z}''$, although this is more difficult to prove in a rigorous way. It is thus an additional conserved-moiety vector.

The conditions in the particular steps of the algorithm were derived from the following arguments. Both η' and η'' have to be smaller than μ, since otherwise we could write

$$\hat{\mathbf{z}} = \mathbf{t}_{i\cdot}^{(m)} + \frac{(\eta' - \mu)\mathbf{t}_{i\cdot}^{(m)} + \eta''\mathbf{t}_{k\cdot}^{(m)}}{\mu} \qquad (\text{A}.2)$$

and similarly for the case $\eta'' > \mu$. Thus, we could decompose $\hat{\mathbf{z}}$, which would be a contradiction to property (P4). The greatest common divisor of η' and μ must be a divisor of all the components of $\mathbf{t}_{k\cdot}^{(m)}$, since η' and η'' are relatively prime. This divisor is unity because otherwise $\mathbf{t}_{k\cdot}^{(m)}$ would not be a reduced vector. A similar reasoning holds for μ and η''.

Now assume that two vectors $\mathbf{t}_{i\cdot}^{(m)}$ and $\mathbf{t}_{k\cdot}^{(m)}$ have zero components at the same positions. Then, if a combined vector $\hat{\mathbf{z}}$ could be obtained with a $\mu > 0$, it would have the same pattern of zero components and, hence, would have been computed already in that step in which $\mathbf{T}^{(m)}$ has been calculated, if $m > r$. If $m = r$, this case cannot occur since any pair of generating vectors have different patterns of zero elements. This is the justification of condition (4.4).

If condition (4.4a) is fulfilled, μ is a divisor of $\eta''t_{kj}^{(m)}$. μ has to be smaller than, or equal to, $t_{kj}^{(m)}$, since η'' and μ are prime to each other. Also, μ must not be greater than any $t_{ih}^{(m)}$, fulfilling (4.4b). The minimum value (t_{min}) of all those $t_{kj}^{(m)}$ and $t_{ih}^{(m)}$ that have zero counterparts in the other vector is therefore an upper bound on μ, which in turn is an upper bound on η' and η''. Consequently, in each step of the algorithm only a finite number of linear combinations needs to be tested. The algorithm involves a finite number of steps, since the end of the algorithm is reached at the latest when all elements of some $\mathbf{T}^{(m)}$ contain nonzero elements only.

References and Notes

(1) Schott, G. L. *J. Chem. Phys.* **1964**, *40*, 2065.
(2) Park, D. J. M. *J. Theor. Biol.* **1974**, *46*, 31.

Conserved-Moiety Structure of Reaction Systems

J. Phys. Chem., Vol. 99, No. 20, 1995 **8023**

(3) Cavallotti, P.; Celeri, G.; Leonardis, B. *Chem. Eng. Sci.* **1980**, *35*, 2297.

(4) Clarke, B. L. In *Advances in Chemical Physics*; Prigogine, I., Rice, S. A., Eds.; John Wiley: New York, 1980; Vol. 43, pp 1−216.

(5) Park, D. J. M., Jr. *Comp. Meth. Prog. Biomed.* **1986**, *22*, 293.

(6) Aris, R. *Arch. Rational Mech. Anal.* **1965**, *19*, 81.

(7) Smith, W. R.; Missen, R. W. *Chemical Reaction Equilibrium Analysis. Theory and Algorithms*; Wiley-Interscience: New York, 1982; Chapter 2.

(8) Pethö, A. *Chem. Eng. Sci.* **1965**, *20*, 791.

(9) Alberty, R. A. *J. Phys. Chem.* **1991**, *95*, 413.

(10) Corio, P. L.; Johnson, B. G. *J. Phys. Chem.* **1991**, *95*, 4166.

(11) Ridler, G. M.; Ridler, P. F.; Sheppard, J. G. *J. Phys. Chem.* **1977** *81*, 2435.

(12) Oliver, P. *Int. J. Chem. Kinet.* **1980**, *12*, 509.

(13) Schuster, S.; Höfer, T. *J. Chem. Soc., Faraday Trans.* **1991**, *87*, 2561.

(14) Clarke, B. L. *J. Chem. Phys.* **1981**, *75*, 4970.

(15) Schuster, R.; Schuster, S. *Comp. Appl. Biosci.* **1993**, *9*, 79.

(16) Park, D. J. M., Jr. *Comput. Chem.* **1988**, *12*, 175.

(17) Alberts, B.; Bray, D.; Lewis, J.; Raff, M.; Roberts, K.; Watson, J. D. *Molecular Biology of the Cell*; Garland Publishing: New York, London, 1989.

(18) Érdi, P.; Tóth, J. *Mathematical Models of Chemical Reactions*; Manchester University Press: Manchester, 1989.

(19) Björnbom, P. H. *AIChE J.* **1977**, *23*, 285.

(20) Mordell, L. J. *Diophantine Equations*; Academic Press: London, New York, 1969.

(21) Schmidt, W. M. *Analytische Methoden für Diophantische Gleichungen*; Birkhäuser: Basel, 1984.

(22) Alberty, R. *Biophys. Chem.* **1994**, *49*, 251.

(23) Rockafellar, R. *Convex Analysis*; Princeton University Press: Princeton, NJ, 1970.

(24) Nožička, F.; Guddat, J.; Hollatz, H.; Bank, B. *Theorie der linearen parametrischen Optimierung*; Akademie-Verlag: Berlin, 1974; Chapter 11.

(25) Smith, H. J. S. *Philos. Trans. R. Soc.* **1861**, *151*, 372.

(26) Bombieri, E.; Vaaler, J. *Invent. Math.* **1983**, *73*, 11.

(27) Borosh, I.; Flahive, M.; Rubin, D.; Treybig, L. B. *Proc. Am. Math. Soc.* **1989**, *105*, 844.

(28) Borosh, I.; Treybig, L. B. *SIAM J. Matrix Anal. Appl.* **1992**, *13*, 454.

(29) Cotton, F. A.; Wilkinson, G.; Gaus, P. L. *Basic Inorganic Chemistry*; Wiley and Sons: New York, 1987.

(30) Von Hohenbalken, B.; Clarke, B. L.; Lewis, J. E. *J. Comput. Appl. Math.* **1987**, *19*, 231.