Journal of Mathematical Chemistry 9(1992)1–10

THE ALGEBRAIC AND GRAPH THEORETICAL COMPLETION OF TRUNCATED REACTION EQUATIONS

Ivar UGI and Alf DENGLER
Lehrstuhl I für Organische Chemie der Technischen Universität München,
Lichtenberg-str. 4, 8046 Garching, Germany

Received 29 October 1991

Abstract

In chemical literature, reactions are often represented in truncated form, i.e. some of the educts and products are left out for brevity. A matrix formalism can be used to obtain the stoichiometrically complete representation of chemical reactions in terms of empirical formulas. The latter results lead from truncated constitutional descriptions of chemical reactions to complete constitutional reaction equations. STOECH, a corresponding computer program, has been implemented.

1. Introduction

In a stoichiometrically balanced reaction equation, the educts on the left-hand side of the equation and the products on its right-hand side correspond to isomeric ensembles of molecules (EM). Such a stoichiometrically complete description of chemical reactions is needed for its representation by the fundamental equation

$$\mathbf{B} + \mathbf{R} = \mathbf{E} \tag{1}$$

of the theory of the BE- and R-matrices [1]. Here, **B** and **E** represent BE-matrices (bond-electron-matrices), which contain information about the molecular connectivity (bonds = off-diagonal entries) and the free electrons (diagonal entries) of the ensemble of molecules (EM) at the beginning (**B**) and the end (**E**) of the reaction. The reaction matrix **R** is the difference of the BE-matrices **B** and **E**: $\mathbf{R} = \mathbf{E} - \mathbf{B}$. **R** represents the redistribution of the valence electrons during the reaction of eq. (1). According to this model, the educt EM and the product EM of a stoichiometrically balanced reaction or sequence of reactions can be subjected to an atom-by-atom mapping according to the principle of minimal chemical distance (PMCD) [2–4]. Thus, the basic requirements, prerequisites for the hierarchical classification and mechanistic characterization of reactions can be established, namely the identification of the reactive centers and of the bonds that are broken or made.

An efficient documentation system for chemical reactions is based on the hierarchical classification [5–7] that can be deduced from the above theory. Despite its many advantages, this documentation system is not used as much as it should be because in chemical literature, most chemical reactions are reported in truncated form, i.e. some of the participating reactants are left out, assuming that any knowledgeable reader is capable of restoring the stoichiometrically complete reaction equations from the corresponding truncated representation, as is illustrated by schemes 1 and 2.

When large data bases of truncated reactions must be converted into stoichiometrically balanced reactions, computer assistance is needed.

Scheme 1. Example of a truncated reaction:

$$CH_3{-}CO_2H \rightarrow CH_3{-}CO_2C_2H_5.$$

Scheme 2. Example of a stoichiometrically complete reaction:

$$CH_3{-}CO_2H + HO{-}C_2H_5 \rightarrow CH_3{-}CO_2C_2H_5 + H_2O.$$

In most cases, the complete constitutional reaction equation can be re-established from the truncated representation by first determining the corresponding stoichiometrically balanced reaction equation and then finding the constitutional formulas of the missing molecules. The latter is usually fairly easy to accomplish because, with few exceptions, the omitted molecules are small molecules that are unambiguously characterized by their empirical formulas and the context in which they are mentioned. These molecules are listed as standard *stoichiometric complement molecules.*

In the present article, we review the hitherto published literature on balancing truncated reactions, and report on our own endeavors in this field, including the development of a corresponding computer program.

2. The algebra of chemical stoichiometry

2.1. STATE OF THE ART

The stoichiometric balancing of reaction equations by mathematical methods was an active research topic in the late sixties [8–11]. In recent years, due to computer-assisted classification and documentation of chemical reactions, interest in this subject has resumed [12–16].

When Ugi and Kaufhold [8] investigated the reaction mechanism of the stereoselective four-component condensation [17], they introduced so-called stoichiometric matrices in order to establish a basis set of linearly independent stoichiometric relations.

Fabishak [9] reported the use of the respective (sub)determinants for balancing stoichiometric equations.

Krishnamurthy [12] balanced chemical equations by a generalized inverse matrix approach. He classified the reaction equations according to the nullity of the coefficient matrix into "not feasible" (nullity = zero), "unique" (nullity = one), and "non-unique" (nullity = two or more). In the last case, two or more equations form a basis set, and all of their linear combinations are stoichiometrically allowable representations of the underlying chemical reaction.

Blakley [13] described a procedure for computing whether or not a chemical reaction is stoichiometrically possible between the members of a set of reactants. If a reaction is possible, the procedure yields a basis set of reaction equations by solving a system of linear equations. All of the stoichiometrically conceivable equations are linear combinations of the elements of the basis set.

The BASIC-86 program CAPURS of Ayme et al. [14] generates all combinations of the columns of a given matrix. The quadratic subdeterminants of the resulting matrices are used to determine whether or not they correspond to soluble systems of linear equations (SLE). If this is the case, CAPURS computes the stoichiometric equations.

The contributions mentioned above are based on empirical formulas.

In their search for reactive atoms and bonds, Bargon et al. [16] determine the largest substructure that is, in terms of constitutional formulas, common to the educts and products of a reaction (MCSS = maximum common substructure). The procedure is based on connectivities and identity numbers. (The identity numbers are assigned to the atoms according to their chemical element numbers.) The matching substructures are deleted, and the procedure is repeated until no more atoms of the educts can thus be mapped onto atoms of the products. The second and all of the following substructures that are deleted must belong to the same molecules as the substructures deleted first. In order to balance the reaction equation, the remaining parts of the educt molecules are subsequently completed in the products and vice versa.

2.2. THE PROGRAM STOECH

In order to ensure that chemical reactions are really represented by stoichiometrically complete reaction equations in terms of empirical formulas, the corresponding systems of linear equations are established and then solved by the Gauss–Jordan method [8,13,18]. First, it is determined whether or not a given set of molecules yields a *stoichiometrically meaningful* reaction equation. If this is the case, the number of linearly independent equations is determined.

2.2.1. *The representation of molecules and their chemical elements by matrices (MCE-matrices)*

Each column of an MCE-matrix represents the empirical formula of a molecule; single atoms are also considered as molecules. The symbol *Lad* is used to denote formal electrical charges, which are treated as if they were atoms. Reaction-invariant residues like the phenyl group can also be represented analogously.

This procedure is illustrated by the reaction scheme in fig. 1 that describes the thermodynamic equilibrium system of isobutyraldehyde-(S)-α-phenylethylimine **A** and benzoic acid **B** and their various adducts in methanol at 0 °C. In the original

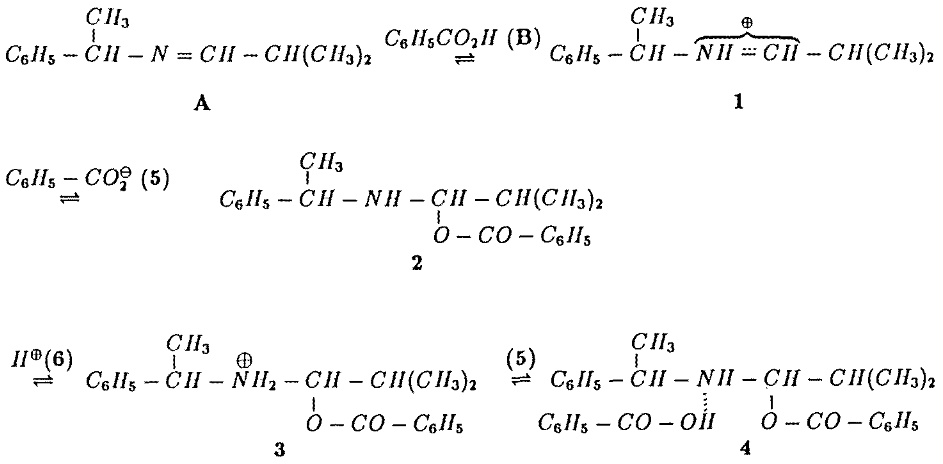

Fig. 1. The equilibrium system of the adducts that are formed from isobutyraldehyde-(S)-α-phenylethylimine (A) and benzoic acid (B).

publication [8], this example was described in abbreviated form. In order to illustrate our procedure, we will now work it out in detail.

The species that participate in this equilibrium system are listed in table 1. This corresponds to the matrix **AM**.

Table 1

Molecules of the reaction scheme of fig. 1.

Label	Molecule	Label	Molecule
A	$C_{12}H_{17}N$	B	$C_7H_6O_2$
1	$C_{12}H_{18}N^{\oplus}$	2	$C_{19}H_{23}NO_2$
3	$C_{19}H_{24}NO_2^{\oplus}$	4	$C_{26}H_{29}NO_4$
5	$C_7H_5O_2^{\ominus}$	6	$H^{\oplus}$

Table 2

Element-species table of the reaction scheme of fig. 1.

Atomic symbol	Molecule label							
	A	B	1	2	3	4	5	6
C	12	7	12	19	19	26	7	0
H	17	6	18	23	24	29	5	1
N	1	0	1	1	1	1	0	0
O	0	2	0	2	2	4	2	0
Lad	0	0	1	0	1	0	−1	1

$$\mathbf{AM} = \begin{pmatrix} 12 & 7 & 12 & 19 & 19 & 26 & 7 & 0 \\ 17 & 6 & 18 & 23 & 24 & 29 & 5 & 1 \\ 1 & 0 & 1 & 1 & 1 & 1 & 0 & 0 \\ 0 & 2 & 0 & 2 & 2 & 4 & 2 & 0 \\ 0 & 0 & 1 & 0 & 1 & 0 & -1 & 1 \end{pmatrix}.$$

2.2.2. *Solution of the systems of linear equations*

The SLE are solved by determining the Hermitian canonical form according to the Gauss–Jordan approach:

(1) Exchange of two rows of **AM**.

(2) Multiplication of a row of **AM** with a non-zero integer.

(3) Addition of a multiple of a row to *another* row.

The result is the matrix **SR**

$$\mathbf{SR} = \begin{pmatrix} 1 & 0 & 0 & 1 & 0 & 1 & 1 & -1 \\ 0 & 1 & 0 & 1 & 1 & 2 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 & 0 & -1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

The zero rows (fourth and fifth row) may be deleted without affecting the set of solutions [13, 18, 19]. These rows were linear combinations of other rows and were transformed into zero rows by the Gauss–Jordan procedure.

The number of linearly independent stoichiometrically *allowable* equations is:

$$N = C - R, \tag{2}$$

with

N = number of linearly independent equations,

C = number of columns (molecules),

R = number of non-zero rows after applying the Gauss–Jordan method.

Since the matrix **SR** contains eight columns and three non-zero rows, we obtain five stoichiometrically allowable equations. Any stoichiometrically allowable relation corresponds to linear combinations of the equations mentioned above.

The solutions of eq. (3)

$$\begin{bmatrix} 1 & 0 & 0 & 1 & 0 & 1 & 1 & -1 \\ 0 & 1 & 0 & 1 & 1 & 2 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 & 0 & -1 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \\ a \\ b \\ c \\ d \\ e \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \tag{3}$$

are obtained by assigning the value 1 to one of the independent variables *a, b, c, d, e* and a value of 0 to all other independent values; the values of the dependent variables x, y and z are then determined.

In the case of $a = 1$, $b = 0$, $c = 0$, $d = 0$, $e = 0$, we obtain

$$1 * x + 0 * y + 0 * z + 1 * 1 + 0 * 0 + 1 * 0 + 1 * 0 + (-1) * 0 = 0$$

$$0 * x + 1 * y + 0 * z + 1 * 1 + 1 * 0 + 2 * 0 + 1 * 0 + 0 * 0 = 0$$

$$0 * x + 0 * y + 1 * z + 0 * 1 + 1 * 0 + 0 * 0 + (-1) * 0 + 1 * 0 = 0$$

$$\Rightarrow x + 1 = 0 \rightarrow x = -1, \quad y + 1 = 0 \rightarrow y = -1, \quad z = 0.$$

The result is $x = -1$, $y = -1$, $z = 0$, $a = 1$, $b = 0$, $c = 0$, $d = 0$, $e = 0$, and we have eq. (4). Equations (5)–(8) are obtained analogously.

The components of the column vectors are the stoichiometric coefficients; column entries with negative algebraic signs belong to the educts, and positive algebraic signs refer to the products. Here, the educts are represented on the left-hand side of the reaction equation and the products on the right-hand side, both with positive stoichiometric coefficients. Since stoichiometric relations are equivalence relations, and thus symmetric, a sign of equality is used in stoichiometric equations instead of reaction arrows.

$$C_{12}H_{17}N + C_7H_6O_2 = C_{19}H_{23}NO_2, \tag{4}$$

$$C_7H_6O_2 + C_{12}H_{18}N^{\oplus} = C_{19}H_{24}NO_2^{\oplus}, \tag{5}$$

$$C_{12}H_{17}N + 2C_7H_6O_2 = C_{26}H_{29}NO_4, \tag{6}$$

$$C_{12}H_{17}N + C_7H_6O_2 = C_{12}H_{18}N^{\oplus} + C_7H_5O_2^{\ominus}, \tag{7}$$

$$C_{12}H_{18}N^{\oplus} = C_{12}H_{17}N + H^{\oplus}. \tag{8}$$

The equations of the reaction scheme in fig. 1 are represented by the following combinations of eqs. (4)–(8):

$$\begin{aligned}
&\text{eq. (7)} && \rightarrow \mathbf{A} + \mathbf{B} = \mathbf{1} + \mathbf{5},\\
&\text{eq. (4)} - \text{eq. (7)} && \rightarrow \mathbf{1} + \mathbf{5} = \mathbf{2},\\
&\text{eq. (5)} - \text{eq. (4)} - \text{eq. (8)} && \rightarrow \mathbf{2} + \mathbf{6} = \mathbf{3},\\
&\text{eq. (6)} - \text{eq. (5)} - \text{eq. (7)} && \rightarrow \mathbf{3} + \mathbf{5} = \mathbf{4}.
\end{aligned}$$

2.2.3. *Incomplete systems of equations*

When the given molecules do not suffice to yield balanced stoichiometric equations, i.e. when the reaction equations are given in truncated form, the program STOECH tries to balance the stoichiometric equations by adding molecules from a file of molecular complements. This file contains the empirical formulas of small molecules that are often encountered as reactants or co-products of chemical reactions. If no suitable complement is listed, it is supplied by the user, who also inserts it into the file of molecular complements. This stoichiometric complementation is illustrated by the example in fig. 2.

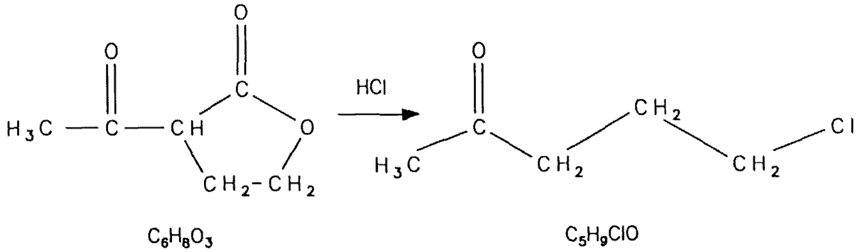

Fig. 2. Preparation of 5-chloropentane-2-one from α-acetyl-γ-butyrolactone: truncated reaction equation.

This reaction corresponds to the stoichiometric matrix **AM** and its Hermitian canonical form **SR**:

$$\mathbf{AM} = \begin{pmatrix} 6 & 5 \\ 8 & 9 \\ 0 & 1 \\ 3 & 1 \end{pmatrix}, \quad \mathbf{SR} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \\ 0 & 0 \end{pmatrix}.$$

According to **SR**, the above system of equations only has a trivial solution where all stoichiometric coefficients are zero. This means that the representation of the respective reaction is truncated. If the reaction of fig. 2 is complemented by HCl and CO_2, members of the molecular complement file, the matrices $\mathbf{AM}_1$ and $\mathbf{SR}_1$ result, together with the reaction equation of fig. 3:

$$\mathbf{AM}_1 = \begin{pmatrix} 6 & 5 & 1 & 0 \\ 8 & 9 & 0 & 1 \\ 0 & 1 & 0 & 1 \\ 3 & 1 & 2 & 0 \end{pmatrix}, \quad \mathbf{SR}_1 = \begin{pmatrix} 1 & 0 & 0 & -1 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$

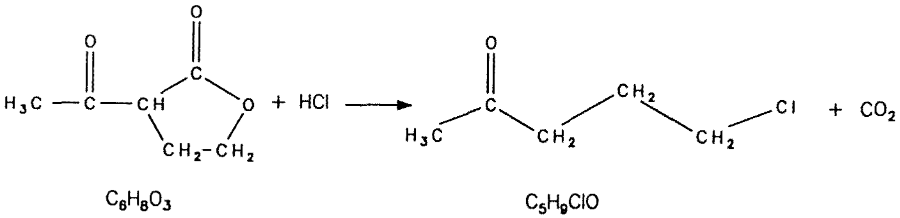

Fig. 3. Preparation of 5-chloropentane-2-one from α-acetyl-γ-butyrolactone: complete reaction equation.

After augmentation by SO_2 and HCl, the reaction equation of the synthesis of acetylchloride according to fig. 4 is represented by the matrices $\mathbf{AM}_2$ and $\mathbf{SR}_2$:

$$\mathbf{AM}_2 = \begin{pmatrix} 2 & 2 & 0 & 0 & 0 \\ 4 & 3 & 0 & 0 & 1 \\ 0 & 1 & 2 & 0 & 1 \\ 2 & 1 & 1 & 2 & 0 \\ 0 & 0 & 1 & 1 & 0 \end{pmatrix}, \quad \mathbf{SR}_2 = \begin{pmatrix} 1 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & -1 \\ 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

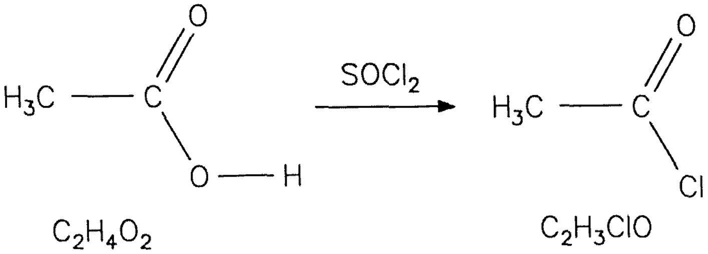

Fig. 4. Preparation of acetylchloride from acetic acid: truncated reaction equation.

Although five chemical elements participate in this reaction, five molecules suffice to solve the system of equations. This is due to the fact that a zero row is formed. The complete reaction eq. (9) results:

$$C_2H_4O_2 + SOCl_2 = C_2H_3ClO + SO_2 + HCl. \tag{9}$$

After balancing the reaction equation, the oxidation of 1-amino-2-naphthol to 1, 2-naphthochinone (fig. 5) is represented by the matrices $\mathbf{AM}_3$ and $\mathbf{SR}_3$:

$$\mathbf{AM}_3 = \begin{pmatrix} 10 & 10 & 0 & 0 & 0 & 0 & 0 \\ 9 & 6 & 4 & 0 & 0 & 2 & 1 \\ 0 & 0 & 1 & 3 & 2 & 0 & 1 \\ 0 & 0 & 0 & 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 2 & 0 & 0 & 0 & 1 & 0 \end{pmatrix}, \quad \mathbf{SR}_3 = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 & 0 & -1 \\ 0 & 0 & 1 & 0 & 0 & 0 & -1 \\ 0 & 0 & 0 & 1 & 0 & 0 & 2 \\ 0 & 0 & 0 & 0 & 1 & 0 & -2 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 \end{pmatrix}.$$

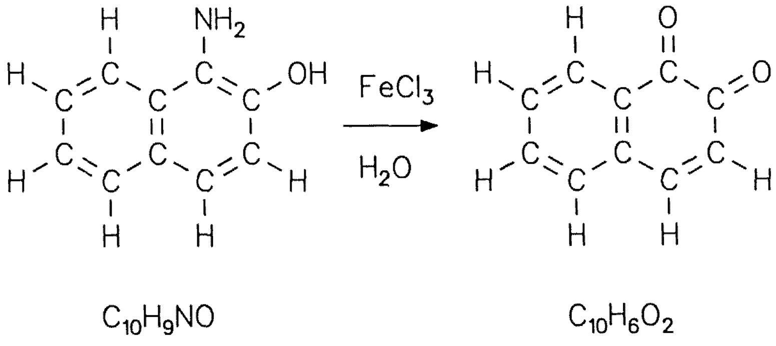

Fig. 5. Preparation of 1, 2-naphthochinone from 1-amino-2-naphthol: truncated reaction equation.

Equation (10) represents the reaction of fig. 5:

$$C_{10}H_9NO + 2FeCl_3 + H_2O = C_{10}H_6O_2 + NH_4Cl + 2FeCl_2 + HCl. \quad (10)$$

The stoichiometrically balanced reaction equation (or system of linearly independent equations) is subsequently translated into a constitutional representation with the help of a structure file. In this file, which resembles the co-product file of CICLOPS [22], small molecules like H_2O, CO_2, etc., are listed. If more than one constitutional formula is in the file of molecular complements for a given empirical formula, then a decision is sometimes possible on the basis of the principle of minimal chemical distance (PMCD) [2–4,20,21]. The PMCD may also serve to obtain an atom-by-atom mapping of the educts onto the products.

The system of computer programs that accomplishes the algebraic and graph-theoretical complementation of truncated chemical reactions will be available upon request by mid-1992.

References

[1] I. Ugi and J. Dugundij, Top. Curr. Chem. 39(1973)19.

[2] C. Jochum, J. Gasteiger and I. Ugi, Angewandte Chemie 92(1980)503.

[3] C. Jochum, J. Gasteiger and I. Ugi, Z. Naturforsch. 37b(1982)1205.

[4] M. Wochner, J. Brandt, A. v. Scholley and I. Ugi, Chimia 42(1988)217.

[5] J. Brandt, J. Bauer, R.M. Frank and A. v. Scholley, Chem. Scr. 18(1981)53.

[6] A. v. Scholley, Hierarchische Klassifizierung und Dokumentation von chemischen Reaktionen, Doctoral Thesis, Technische Universität München (1981).

[7] I. Ugi, J. Brandt, A. v. Scholley, S. Minker, M. Wochner, H. Schönmann and B. Straupe, Hierarchisch strukturierte Speicherung und Ermittlung von chemischen Reaktionen, Research Report BMFT-FB-ID-85-005, FRG-BMFT, Bonn (1985).

[8] I. Ugi and G. Kaufhold, Liebigs Ann. Chem. 709(1967)11.

[9] V.L. Fabishak, CHEMISTRY 40 [11] (1967) 18.

[10] H.W. Ferguson, CHEMISTRY 40 [2] (1967) 18.

[11] G.N. Copley, CHEMISTRY 41 [9] (1968) 22.

[12] E.V. Krishnamurthy, Int. J. Math. Educ. Sci. Technol. 9 [3] (1978) 323.

[13] G.R. Blakley, J. Chem. Educ. 59 [9] (1982) 728.

[14] J.L. Ayme, M. Arbelot and M. Chanon, Comp. Chem. 10(1986)85.

[15] J.L. Larice, in: *Computer Aids to Chemistry*, ed. G. Vernin and M. Chanon (Ellis Horwood, Chichester, UK, 1986), p. 169.

[16] C. Behnke and J. Bargon, J. Chem. Inf. Comput. Sci. 30(1990)228.

[17] I. Ugi, S. Lohberger and R. Karl, in: *Comprehensive Organic Synthesis: Selectivity for Synthetic Efficiency*, Vol. 2, ed. B.M. Trost and C.H. Heathcock (Pergamon, Oxford, UK, 1991), ch. 4.6, in press.

[18] N. Rösch, *Mathematische Methoden der Chemie*, printed lecture notes, Technische Universität München (1990).

[19] H.G. Zachmann, *Mathematik für Chemiker* (Verlag Chemie, Weinheim, 1974).

[20] M. Wochner, Die exakte Lösung der Zuordnungsaufgabe für chemische Reaktionen, Doctoral Thesis, Technische Universität München (1985).

[21] I. Ugi, M. Wochner, E. Fontain, J. Bauer, B. Gruber and R. Karl, in: *Concepts and Application of Molecular Similarity*, ed. M.A. Johnson and G.M. Maggiora (Wiley, New York, 1990).

[22] J. Blair, J. Gasteiger, C. Gillespie, P.D. Gillespie and I. Ugi, in: *Computer Representation and Manipulation of Chemical Information*, ed. W.T. Wipke, S.R. Heller. R.J. Feldmann and E. Hyde (Wiley, New York, 1974), p. 129.