

THE DEDUCTIVE SOLUTION OF CHEMICAL PROBLEMS BY COMPUTER  
PROGRAMS ON THE BASIS OF A MATHEMATICAL MODEL OF CHEMISTRY

Ivar Ugi, Josef Brandt, Josef Friedrich, Johann Gasteiger,  
Clemens Jochum, Peter Lemmen, and Wolfgang Schubert

Organisch-chemisches Institut der Technischen Universität  
München, Lichtenbergstr. 4, 8046 Garching, W.-Germany

**Abstract** - A mathematical model of constitutional chemistry is described which is well suited as a theoretical basis for the deductive solution of a variety of chemical problems by computer programs. Within this framework the chemical constitution of molecules and ensembles of molecules (EM) is represented by BE-matrices, whose rows and columns are assigned to the considered atomic cores, and whose entries represent covalent bonds and free valence electrons. Chemical reactions are represented by transforming the BE-matrix B of the beginning EM into the BE-matrix E of the end EM by addition of an R-matrix R according to the master equation  $B + R = E$  of the present theory.

With a given initial matrix B, those R-matrices R whose addition to B represent chemical reactions can be generated mathematically without any information on individual chemical reactions. The applications of this approach are synthesis design and the prediction of the products which may conceivably be formed from combinations of listed chemical compounds. When the basis elements of the R-matrices are used in a successive mode in this context, results may be obtained which take into account mechanistic aspects of chemical reactions.

A pair (B,E) representing the beginning and end EM of any chemical reaction, or a sequence of chemical reactions yields a difference matrix  $E - B = R = \sum V R_V$ , whose components  $R_V$  may be used to generate networks of reaction pathways that lead from EM(B) to EM(E). This may also be used for synthesis design, or for the elucidation of reaction mechanisms.

From given R-matrices R one may generate those pairs (B,E) which satisfy  $B + R = E$ , and thus systematically "invent" chemical reactions.

The deductive approach yields unprecedented conceivable solutions to chemical problems. These can be screened through selection rules which are based on general chemical experience.

### 1. Computers and Chemistry

The capabilities of computers with respect to numerical calculations as well as data storage and retrieval play an essential role in modern chemistry. This is illustrated by the spectacular successes of quantum chemistry and also by current computer-assisted chemical documentation.

In the foreseeable future problem-solving computer programs with artificial intelligence features will acquire similar importance for chemistry. The synthesis design programs LHASA<sup>1</sup>, SECS<sup>2</sup>, SYNCHEM<sup>3</sup>, and the Leverkusen Peptide Synthesis Program<sup>4</sup> (LPSP) were the first chemical computer programs with artificial intelligence features, e. g. decision-making processes in combination with complex information processing and data manipulation.

At the beginning of this decade a universal mathematical model of constitutional chemistry was introduced which involves the theory of the BE- and R-matrices<sup>5</sup>. This mathematical model may serve as the basis for computer programs with artificial intelligence capabilities which generate solutions to particular chemical problems, starting from a general theory, and simulating deductive reasoning. These computer programs have inventive capacity and are capable of generating solutions to chemical problems which involve unprecedented chemistry.

### 2. Isomeric Ensembles of Molecules and the Chemistry of a Given Set of Atoms

Our mathematical model of constitutional chemistry is based on extending the concept of isomerism from molecules to ensembles of molecules (EM).

With a given set of atoms  $A = A_1, \dots, A_n$  any molecule, or collection of two or more molecules, which contains each atom of A once and only once in an EM belonging to A, is denoted by EM(A). The term FIEM(A) is used for the family of all EM(A).

Within the chemistry of a given set of atoms A, any chemical reaction, or sequence of chemical reactions may be regarded as the conversion of an EM(A) into an isomeric EM'(A). Accordingly the whole chemistry of A is given by the EM(A) and their interconversions within the FIEM(A).

A theory of the FIEM(A) is at the same time a universal theory of chemistry. Any part of chemistry is contained in the FIEM(A) of a suitably chosen set of atoms.

A detailed computation and subsequent analysis of the whole energy hypersurface of any set of atoms A by quantum mechanical methods would make all essential features of the chemistry of A predictable.

This will, however, not be possible in the near future for the chemically interesting cases, due to the vast amount of the required computations.

### 3. Matrix Representation of EM

The concept of a covalent bond as a pair of electrons in an orbital comprising two atomic cores and the conservation of the total number of valence electrons as well as the atomic cores consisting of the atomic nuclei and the inner shell electrons suffice as the starting points of a mathematical theory of the FIEM which is well suited as a foundation of computer programs for the deductive solution of chemical problems.

Within this theory the chemical constitution of a molecule or EM with n atoms is described by an  $n \times n$  BE-matrix B with non-negative entries  $b_{ij}$ . In a BE-matrix the i-th row and column is associated with the core of the

atom  $A_i$ . The formal covalent bond order between the atoms  $A_i$  and  $A_j$  is represented by the off-diagonal entry  $b_{ij}$  ( $i \neq j$ ) in the  $i$ -th row and  $j$ -th column. Since a bond from  $A_i$  to  $A_j$  is also a bond from  $A_j$  to  $A_i$ , we have  $b_{ji}=b_{ij}$ . Thus, a BE-matrix is symmetric. The  $i$ -th diagonal entry  $b_{ii}$  is the number of free valence electrons belonging to  $A_i$ .

For example, with  $A_1 = \text{H}$ ,  $A_2 = \text{C}$ , and  $A_3 = \text{N}$  the chemical constitution of 1 is represented by the BE-matrix  $B_{\text{HCN}}$

$$\begin{array}{c} \text{H} - \text{C} \equiv \text{N}: \\ \underline{1} \end{array}$$

$$B_{\text{HCN}} = \begin{array}{ccc} & \begin{matrix} 1 & 2 & 3 \end{matrix} \\ \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 3 \\ 0 & 3 & 2 \end{pmatrix} & \begin{matrix} 1 \\ 2 \\ 3 \end{matrix} \end{array}$$

In  $B_{\text{HCN}}$  the entries  $b_{12} = b_{21} = 1$  correspond to the single bond between H and C, and  $b_{33} = 2$  to the pair of free electrons at the N atom.

In a BE-matrix the row/column sum

$$s_i = \sum_j b_{ij} = \sum_j b_{ji} \quad (1)$$

indicates the number of valence electrons which belong formally to  $A_i$ , and the cross sum

$$\hat{s}_i = 2s_i - b_{ii} \quad (2)$$

is the total number of valence electrons in the orbitals associated with  $A_i$ . For atoms obeying the octet rule<sup>6</sup> we have  $\hat{s}_i = 8$ . The sum over all entries

$$S = \sum_{ij} b_{ij} \quad (3)$$

is the total number of valence electrons in the EM.

In stable molecules each atom can only have certain combinations of values for the numbers  $s_i$  and  $\hat{s}_i$  characteristic for the valence state for a given chemical element, and  $b_{ij}$  for a pair of elements<sup>7</sup>. In the case of reactive intermediates some further combinations are acceptable. Table 1 contains some characteristic combinations of these numbers for the first 18 chemical elements which may be called their valence chemical stability rules. Table 2 is a survey of the latter.

The row/column indices  $i = 1, \dots, n$  of an  $n \times n$  BE-matrix can be assigned to the  $n$  atoms of an EM in  $n!$  different ways. Accordingly an EM can be represented by up to  $n!$  equivalent BE-matrices which differ by row/column permutations. Thus, any two  $n \times n$  BE-matrices  $B$  and  $B'$  belong to the same EM, if there exists an  $n \times n$  permutation matrix  $P^t$  which transforms  $B'$  into  $B$  according to (4).

$$B = P^t \cdot B' \cdot P \quad (4)$$

With a unique assignment of BE-matrices to molecules, or EM, the BE-matrices would be ideally suited for chemical documentation, not only for internal documentation within computer programs on the basis of the present theory of constitutional chemistry.

In order to achieve a unique representation of an EM by a BE-matrix, one assigns the row/column indices  $1, \dots, n$  of the BE-matrix in a canonical manner. For this purpose the programs NOON<sup>8</sup> and CANON<sup>9</sup> have been developed. NOON results in an indexing which in most cases parallels the one obtained by the Morgan algorithm<sup>10</sup> but avoids its pitfalls. Both CANON and NOON are well suited for chemical documentation, since they may also be used in the description of stereochemical features.

Generally, the stereochemical nomenclature systems involve the conceptual dissection of molecules into skeletons and ligands, and sequential ordering of the latter. This cannot only be achieved through sequence rules, like the CIP rules<sup>12</sup>, but also through the atomic indices of the  $\alpha$ -atoms of the ligands, i. e., the atoms which are directly connected with the molecular skeleton<sup>11,13</sup>. CANON assigns indices to the atoms in a molecule which simultaneously give the order of the ligands through the indices of their  $\alpha$ -atoms whose order corresponds to their CIP sequential order, with a few well defined exceptions. Furthermore CANON uniquely determines the sequence of ligands in those cases where there are apparent difficulties in applying the CIP rules.

Another type of equivalency of BE-matrices arises from resonance. Molecules, or EM, with delocalized  $\pi$ -electron systems can be represented by two or more resonance formulas. These correspond to distinct BE-matrices which are equivalent descriptions of the considered EM. This resonance equivalence of BE-matrices must be taken into account in chemical computer programs involving BE-matrices. Resonance equivalent BE-matrices differ only by the values of their non-zero off-diagonal entries, but not by their placement. Accordingly, it is easy to obtain a whole set of resonance equivalent BE-matrices from any member of the set.

#### 4. Matrix Representation of Chemical Reactions

A chemical reaction, or sequence of chemical reactions, begins with an EM(B) and ends with an isomeric EM(E). It proceeds with conservation of the atomic cores and leads to a redistribution of the valence electrons, corresponding to a change in the location of the covalent bonds and free electrons.

The reaction  $EM(B) \longrightarrow EM(E)$  corresponds to a transformation  $B \longrightarrow E$  of the BE-matrix B of EM(B) into the BE-matrix E of EM(E). Yoneda<sup>14</sup> was the first to describe a chemical reaction by a matrix transformation, however, the type of matrices which he used is different from the BE- and R-matrices.

For reasons which will become obvious in the subsequent discussion the transformation  $B \longrightarrow E$  is best accomplished through a so called R-transformation by the addition of an R-matrix R to B (see (5)), according to the master equation (5) of the present theory.

$$B + R = E \quad (5)$$

Thus, the entries  $e_{ij}$  of E are determined by the entries  $b_{ij}$  of B, and the entries  $r_{ij}$  of R (see (6)).

$$b_{ij} + r_{ij} = e_{ij} \quad (6)$$

The total number of valence electrons does not change during a chemical reaction. Therefore we have

$$S = \sum_{ij} e_{ij} = \sum_{ij} b_{ij} + \sum_{ij} r_{ij} = \sum_{ij} b_{ij} \quad (7)$$

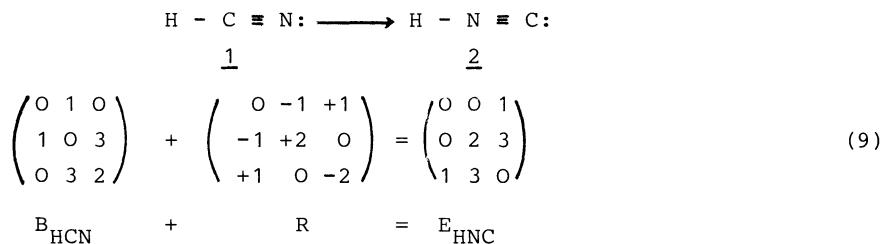
From this follows

$$\sum_{ij} r_{ij} = 0 \quad (8)$$

The R-matrices are symmetric; with  $b_{ij} = b_{ji}$  and  $e_{ji}$ , eq. (6) requires  $r_{ij} = r_{ji}$ .

The off-diagonal entries  $r_{ij} = r_{ji}$  ( $i \neq j$ ) of an R-matrix indicate the changes in formal covalent bond order, and the associated change in free electron distribution is given by the diagonal entries  $r_{ii}$ .

For example, the reaction  $\underline{1} \rightarrow \underline{2}$  is represented by  $B_{\text{HCN}} \rightarrow E_{\text{HNC}}$



Recently a series of papers<sup>15</sup> was published in which the representation of a large number of chemical reactions by R-matrices is listed. Since equations (5) and (10) represent the same reaction, they are equivalent.

$$P^t B P + P^t R P = P^t E P \quad (10)$$

In this context any two  $n \times n$  R-matrices, R and R', are equivalent, if

$$R' = P^t R P \quad (11)$$

An R-matrix corresponds to a pattern of valence electron flow. Any two R-matrices which are interconvertible by insertion or removal of rows and columns which contain only zeros represent the same type of electron flow. An R-matrix from which all empty rows and columns have been removed is called an irreducible R-matrix. A canonical form of an irreducible R-matrix may serve as the representation of all chemical reactions with the same electron flow. Accordingly they are said to belong to the same R-category. Since the description of chemical reactions may range from its R-category down to a detailed one by eq. (5), combined with a statement about its stereochemical features, a hierarchic documentation system of chemical reactions may be based on the present concept.

An R-matrix fits eq. (5) only, if its negative entries  $r_{ij} = -r_{ij}$  are placed to match some positive entries  $b_{ij}$ , i. e.

$$b_{ij} + r_{ij} = e_{ij} \geq 0 \quad (12)$$

With a given beginning matrix B, the mathematical fitting condition<sup>5</sup> in combination with the mathematical structure of the R-matrices, as given by eq. (8), and the symmetry requirement permits to manufacture all R-matrices which are potential candidates for the representation of conceivable chemical reactions which EM(B) may undergo, or by which EM(B) may be formed.

In order to confine the R-transformations of B to those which lead to a chemically meaningful result E or EM(E), one selects only those R-matrices, for which the  $E = B + R$  fulfil those valence chemical stability rules which may also be called a boundary condition. (Table I). We refer to this as the valence chemical fitting condition, which may also be called a boundary condition.

With beginning matrix B and fitting conditions given mathematical structure of the R-matrices suffices to determine the R-matrices which can be added to B to yield representations of conceivable chemical reactions. Thus, representation of conceivable chemical reactions are deducible from general principles without any specific prior knowledge about these reactions.

Accordingly, a computer program could generate from  $b_{12} = b_{21} = 1$  and  $b_{33} = 2$  of  $B_{\text{HCN}}$  the entries  $r_{12} = r_{21} = -1$  and  $r_{33} = -2$  of an R-matrix, which fits  $B_{\text{HCN}}$  mathematically, and then "fill up" the R-matrix by the mathematically and valence chemically acceptable entries  $r_{13} = r_{31} = +1$  and  $r_{22} = +2$ . Thereby the reaction  $\underline{1} \longrightarrow \underline{2}$  would be "reinvented" by a computer without any specific input concerning this reaction.

This unique "creative" capability of the present theory is essential to the design of computer programs for the deductive solution of chemical problems.

## 5. Chemical Metric

An  $n \times n$  BE-matrix B with entries  $b_{ij}$  corresponds to a vector with  $n^2$  components  $(b_{11}, \dots, b_{1n}; b_{21}, \dots; \dots; b_{n1}, \dots, b_{nn})$ . This leads to an embedding of the BE-matrices of an FIEM in  $R^{n^2}$ , an  $n^2$ -dimensional metric space on the real numbers. The entries  $b_{ij}$  of B may be seen as the cartesian coordinates of a so-called BE-point P(B) in  $R^{n^2}$ .

An R-matrix R can similarly be represented as a vector in  $R^{n^2}$ . A chemical reaction  $EM(B) \longrightarrow EM(E)$  may be visualized as a vector R from P(B) to P(E) in  $R^{n^2}$ .

It is expedient to define the distance between P(B) and P(E) as

$$D(B,E) = \sum_{ij} b_{ij} - e_{ij} = \sum_{ij} r_{ij} \quad (13)$$

This distance has a straightforward chemical interpretation: the value of  $D(B,E)$ , the chemical distance between P(B) and P(E), is twice the number of valence electrons which participate directly in the reaction  $EM(B) \longrightarrow EM(E)$ .

The origin of the coordinates in  $R^{n^2}$  is given by O, the  $n \times n$  zero matrix. The chemical distance of a BE-point P(B) from the origin is therefore equal to

$$D(B,O) = \sum_{ij} b_{ij} = S \quad (14)$$

i. e. the number of valence electrons in EM(B). Since all EM(A) contain the same number of valence electrons, the BE-points of an FIEM(A) must lie on the surface of a "hypersphere" whose "chemical radius" is S. Further, these points belong to a lattice of points in  $R^{n^2}$  with positive integer coordi-

nates. The chemical nature of the atoms in A determines, through the valence chemical stability rules which points of the above lattice are representing stable EM, which ones are representing reactive intermediates, and which ones do not belong to existing molecular systems.

Since each coordinate in the representation space of an FIEM(A) is determined by a pair (i,j) of atomic indices, the coordinates of a BE-point of an EM(B) are determined by the assignment of the atomic indices. A permutation of the atomic indices induces a permutation of the coordinates of the BE-points.

Let EM(B) and EM(E) be the beginning and end EM of a chemical reaction, or sequence of reactions. If the fate of the individual atoms in the process  $EM(B) \rightarrow EM(E)$  is not known, this reaction is not representable by the master equation  $B + R = E$ , unless a method is found to correlate the assignment of atomic indices in EM(B) with that of EM(E), such that each atom in these EM has the same index in both EM. This can be achieved by the postulate of minimal electron flow, i. e. the postulate that those reaction mechanisms are preferred for which the number of the participating valence electrons is as small as possible. From this it follows that the atoms in interconvertible EM must be indexed so that the chemical distance  $D(B,E)$  has a minimal value  $D(B_0,E_0)$ . Let the atoms in EM(E) be endowed with fixed indices and EM(E) be represented by  $E_0$ , while the atomic indices of EM(B) are permuted, then any BE-matrix

$$B_k = P_k^t B_0 P_k = P_k^t B_a B_k,$$

may be used for EM(B).

Beginning with an arbitrary indexing of the atoms in EM(B) and a corresponding BE-matrix  $B_a$ , the desirable indices of the atoms in EM(B) are found through the permutation matrices  $P_k'$  for which  $D(P_k' B_a P_k', E_0)$  has a minimum value. Since this minimum is difficult to determine by numerical methods<sup>16</sup>,  $P_k'$  is found through the maximum value of

$$Q(E_0, P_k', B_a P_k',) = \sum_{ij} (e_{ij})_0 \cdot (b_{ij})_k \quad (15)$$

This correlation of atomic index assignments in two isomeric EM is useful for the elucidation of reaction mechanisms and the design of syntheses where not only the target and its coproducts but also the starting materials are known<sup>17</sup>.

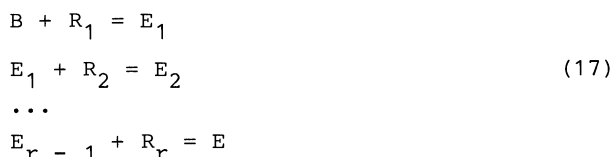
## 6. The Components of R-Matrices

An R-matrix R can be written as a sum of components according to eq. (16)

$$R = R_1 + R_2 + \dots + R_r \quad (16)$$

This decomposition of R can be carried out to yield components  $R_j$  which belong to a basis of the additive free Abelian group  $R(n)$  to which the  $n \times n$  R-matrices belong<sup>5</sup>.

When an R-transformation  $B + R = E$  can be carried out stepwise under the fitting conditions

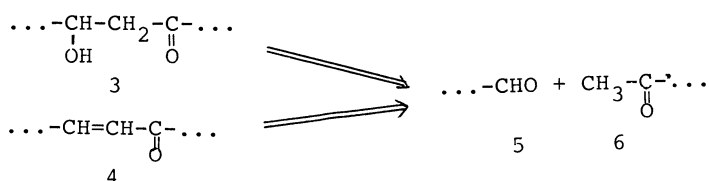


the intermediate BE-matrices  $E_1, \dots, E_r$  may be interpreted as intermediates of the reaction represented by  $B + R = E$ .

The network of all conceivable reaction paths which may lead from a beginning EM(B) to an end EM(E) is obtained from B through successive action of the components  $R_v$  of  $R = E - B$  under the fitting conditions. Here the assignment of the row/column indices to the atoms must be maintained for all considered BE-matrices. If necessary, this can be ascertained by the aforementioned procedure.

### 7. The Design of Syntheses

In the retrieval oriented synthesis design programs<sup>1-4</sup> synthetic precursors are generated from the target molecule through the action of transforms. The transforms are contained in the reaction libraries of such programs and represent synthetically useful chemical reactions, as their retro-reactions. The selection of transforms from the reaction library is guided by the perceived structural features of the target molecule. The structural feature 3, or 4, in a target induces the application of transforms referring to the aldol condensation and leads to 5 and 6 as precursors for 3, or 4, respectively.



By successive applications of transforms one obtains, starting from the target, a tree of synthetic pathways which correspond to sequences of reactions stored in the reaction library. For the reactions in the reaction library the scope is known and can be taken into account. Therefore the synthetic pathways which are generated by a retrieval oriented synthetic design program are quite likely to succeed experimentally.

The drawback is, that no synthetic pathways are obtained which contain reactions without precedent. A computer program for synthesis design would need the capability of generating reactions leading to a given target which are not contained in, or implied by, the reaction library in order to be able to design syntheses which contain hitherto unknown reactions. This necessitates deriving from general principles transforms which correspond to the conceivable reactions leading to the formation of a given target molecule. These general transforms require the redistribution of the valence electrons in a manner which does not lead for any one of the involved atoms to a "forbidden" electronic configuration. Furthermore the allowable electron redistribution must not be applied to the target molecule alone, but to the target together with its coproducts; otherwise one would be limited to syntheses by rearrangement and addition reactions and exclude e. g. condensation reactions.

The coproducts of synthetic reactions are generally small low-energy molecules such as water, carbon dioxide etc. With a coproduct list of less than 30 such compounds, it is hard to find any known organic synthesis which yields coproducts which are not contained in the list!

Thus, one could consider the various combinations of the target molecule



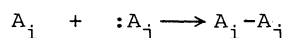
with members of the coproduct list and carry out all of the conceivable redistributions of valence electrons which are valence chemically permissible.

Such a procedure of generalizing the transforms of a reaction library, however, would at best lead to the concepts of the mathematical model outlined above<sup>+</sup>. Existent reaction libraries could then be used - like any other computer file on chemical reactions - to check whether or not any one of the generalized transforms corresponds to known reactions.

Computer programs which can generate all conceivable reactions exist already at a rather advanced stage of development in the form of some deductive synthesis design programs on the basis of the mathematical model of constitutional chemistry. In the pilot program CICLOPS<sup>7</sup> and the synthesis design program EROS<sup>18,19</sup> which evolved from the latter, the synthetic pathways are generated by a formal procedure. In these programs the BE-matrix E(Z) of the target molecule Z is augmented by the conceivable coproducts C which may be formed during the synthesis of Z. Thus, the BE-matrix E(Z,C) of an EM(Z,C) is obtained which consists of Z and some of its conceivable coproducts. The R-transformations of E(Z,C) by certain mathematically and valence chemically fitting R-matrices, whose negative off-diagonal entries correspond to one, two or three of the "breakable bonds" in the target EM(Z,C), lead to BE-matrices of precursors which are analysed according to selection rules which are based on general chemical experience. In EROS the enthalpy changes which accompany the proposed synthetic reactions are emphasized.

In order to evaluate the capabilities of EROS some synthetic targets were subjected to the program, e. g.  $\beta$ -propiolactam, perfluoro-n-propyl ether, and pyridoxal<sup>18</sup>. Without prior knowledge of the synthetic pathways which had been found for these targets, the output of EROS contained among others also some recently published synthetic pathways by which the above target compounds are formed.

The R-matrices with even diagonal entries have basis elements  $R_{ij}$  with the non-zero entries  $r_{ij} = r_{ji} = +1$ ,  $r_{jj} = -2$  ( $i \neq j$ ). The basis elements  $R_{ij}$  represent elementary chemical reactions of the type



The successive action of the basis elements  $R_{ij}$  to a BE-matrix E under the mathematical and valence chemical fitting conditions generates a sequence of BE-matrices  $B_1, B_2, \dots$  which correspond either to the characteristic intermediates of chemical reactions or to the stable EM(B), from which EM(E) may conceivably be obtained.

On this basis a new type of deductive synthesis design program is being developed<sup>20</sup>. The first results which have been obtained with this program are most encouraging. For example, from ethyl acetate as the synthetic target and water as the coproduct all of the known syntheses for ethyl acetate are generated.

---

<sup>+</sup> W. T. Wipke pointed out during a discussion in Sept. 1976 that the capabilities of a retrieval oriented synthesis design program may be extended by the use of generalized, or what he termed *ab initio* transforms, which are in a one-to-one correspondence with certain valence chemically allowable R-transformations. Thus, our model of constitutional chemistry might also serve as a theoretical foundation of an extension of the retrieval type chemical computer programs.

Synthesis design by these programs may be visualized in terms of pathways in  $R^{n^2}$  which begin with the BE-point  $P(Z,C)$  and proceed via intermediate BE-points to the BE-point  $P(A)$  of those EM which consist of available starting materials only. These pathways are generated stepwise beginning with a scan of the vicinity of  $P(Z,C)$  within a suitably chosen chemical distance, and proceeding successively to further vicinities of the BE-points of intermediates.

With a list  $L$  of available starting materials, and an algorithm to choose  $EM(A) \leq L$  from this list as starting materials, it is also possible to design syntheses from both ends  $E(Z,C)$  and  $B(A)$ . Suitable  $EM(A)$  can be found by comparison of the substructures of  $Z$  and then members of  $L$ , and when the  $EM(A)$  and  $EM(Z,C)$  are known, then the atoms can be appropriately indexed by minimizing the chemical distance  $D(B,E)$ . The components or basis elements of  $R = E - B$  are then used under fitting conditions to generate networks of pathways between  $P(B)$  and  $P(E)$ .

Synthesis design from "both ends" is most promising, because of the complete "look ahead" and the relatively small number of intermediates and pathways to be taken into consideration.

#### 8. The Scope of Deductive Chemical Computer Programs

Due to its universal nature, the mathematical model of constitutional chemistry has many other potential uses in computer chemistry besides the design of syntheses.

It is possible to predict the products which may conceivably be obtained from a given chemical compound when it is reacted with some listed reaction partners. The considered chemical may be an agricultural chemical which is used in an environment containing certain reactive partners, or it may be the coproduct of a synthesis which is not directly usable, and for which one tries to find chemical uses. One could predict the products which may be obtained from this coproduct by reacting it with listed inexpensive reactants. For a given set of reactants not only the desired products but also the potential formation of undesirable by-products could be foreseen<sup>21</sup>.

From the initial and final EM of a known reaction the network of all conceivable reaction mechanisms could be generated by the use of R-matrix basis elements.

Even the systematic discovery of new reactions is within the scope of the present theory. A given R-matrix  $R$  may be used to find all pairs  $(B,E)$  of BE-matrices which satisfy the master equation (5) under mathematical and valence chemical fitting conditions. These matrices  $B$  and  $E$  are found from  $R$  via the matrices  $B_0$  and  $E_0$  by addition of matrices  $X$ .

$$B = B_0 + X \quad (18)$$

$$E = E_0 + X \quad (19)$$

The matrix  $B_0$  is obtained from  $R$  through replacing its positive entries by zeros, and the negative entries by positive ones with the same number values. The matrix  $E_0$  results from  $R$  by replacement of the negative entries by zeros. The matrix  $X$  is found from  $B_0$  and  $E_0$  by associating the rows/columns of  $B$  and  $E$  with certain chemical elements and selecting the entries of  $X$  in accordance with the mathematical and chemical fitting rules for  $B + R = E$ . Thus, one would find all conceivable reactions which are represented by  $R$  and by all R-matrices which are equivalent to  $R$  by belonging to the same irreducible R-matrix.

Since there exists only a relatively small number (<100) of non-equivalent irreducible R-matrices, it does not seem to be unrealistic to expect that it

will be possible to generate representatives for all conceivable basic types of chemical reactions, as long as these involve EM's that are representable by BE-matrices.

As soon as adjacency matrices, topological matrices, or connectivity matrices<sup>22-26</sup> were used in the representation of molecules, it became obvious that the known chemical reactions could be described by suitably defined transformations of such matrices. Without our mathematical model, constitutional chemistry and its master equation

$$B + R = E$$

would be merely another device for storing, retrieving, and manipulating information on chemical reactions. Although it would provide definite advantages within retrieval oriented synthesis design programs, the representation of known chemical reactions by the master equation would not add totally new qualities to computer chemistry.

If, however, the master equation is seen within the framework of the present theory, its various types of solutions under the mathematical fitting conditions of the R-matrices and the valence chemical boundary conditions open up the entire field of computer programs for the deductive solution of chemical problems. Then this equation which expresses some general principles can be used for the deductive solution of a great variety of chemical problems. The need for detailed information on chemical reactions is eliminated by the fact that for the chemistry of any set of  $n$  atoms the R-matrices belong to the set  $R(n)$  of  $n \times n$  symmetric matrices whose entries add up to zero. The members of the set  $R(n)$  can be generated according to their mathematical structures without any chemical knowledge. Those members of  $R(n)$  which represent the chemical reaction of a given EM can be picked from  $R(n)$  according to the mathematical and valence chemical fitting conditions, with due consideration of the chemical distance aspect. It is also possible to find those BE-matrices which can be combined with a given R-matrix to give a chemically meaningful solution of the master equation. Further, the additive group theoretical properties of the R-matrices afford the decomposition of overall chemical changes into a sequence of reaction steps. Thus, with a given BE-matrix  $B$  the master equation may be solved under the mathematical fitting conditions and the valence chemical boundary conditions to yield  $R$  and  $E$ . When two BE-matrices,  $B$  and  $E$ , are known the R-matrix  $R = E - B$  may be obtained, as well as its components in permissible order. Also one can begin with an R-matrix and find those pairs  $(B, E)$  of BE-matrices which satisfy the master equation in a chemically meaningful way.

Generally any one of these approaches yields a large number of solutions to the master equation. In order to reduce the number of solutions as much as possible without relinquishing valuable information, one needs selection rules which take into account chemical experience. Accordingly, chemical knowledge is not neglected in the sense of CNPE (complete neglect of prior experience)<sup>27</sup>, although it is not used in the primary processes of information production.

Empirical data may not only be used to establish general selection rules for deductive chemical computer programs, but suitably organized detailed files of chemical compounds and reactions may serve well for determining whether or not some computer generated solutions of a chemical problem have already been reported in the literature, as direct precedent, or at least in the form of analogs. Thereby part of the solutions of a problem can be related to the literature, while others are pointed out as being without precedence. This type of approach is to be recommended very much for future developments, since it provides the best of both worlds, namely on the one hand the advantages of the mathematical model with full coverage of all conceivable solutions to a chemical problem and the expediency of matrix addition, and on the other hand access to existent documented chemical information, without the need of converting it into a device for retrieval oriented problem solving.

Under the assumption that in the future computer programs for the deductive solution of chemical problems will be routine tools for the chemist which are used in combination with suitably organized files of empirical data, chemists will be relieved of some routine in chemical problem solving and be reminded of some solutions of problems of which they would not have thought due to the natural limitations of human experience and thought. The chemist would then have more time to concentrate on issues in which computers may at least provide some help, such as the selection of worthwhile research problems, or the search for optimum conditions for experiments.

#### Acknowledgements

We acknowledge gratefully the financial support of our work in this field by the Deutsche Forschungsgemeinschaft (deductive computer programs), Stiftung Volkswagenwerk e.V. (mathematical models and theoretical stereochemistry), and the European Communities (chemical information retrieval and manipulation systems). We also wish to express our appreciation of the advice and guidance which Professor J. Dugundji provided.

#### References

- 1.) E.J. Corey, W.T. Wipke, R.D. Cramer, W.J. Howe, *J. Amer. Chem. Soc.*, 94, 421 (1972)
- 2.) W.T. Wipke, in "Proceedings of the NATO/CNA ASI on Computer Representation and Manipulation of Chemical Information", Wiley, New York, 1974
- 3.) H. Gelernter, N.S. Sridharan, A.I. Hart, S.C. Yen, F.W. Fowler, H.I. Shue, *Topics Curr. Chem.*, 41, 113 (1973)  
H.L. Gelernter, A.F. Sanders, D.L. Larsen, K.K. Agarwal, R.H. Boivie, G.A. Spritzer, J.E. Searleman, *Science*, 4308, p. 1041 (1977)
- 4.) G. Kaufhold and I. Ugi (unpublished results 1966/67);  
see also: I. Ugi, *Rec. Chem. Progress*, 30, 289 (1969)
- 5.) J. Dugundji, I. Ugi, *Topics Curr. Chem.*, 39, 19 (1973)
- 6.) G.N. Lewis, *J. Am. Soc.*, 38, 762 (1916); I. Langmuir, *J. Amer. Chem. Soc.*, 41, 868 (1919)
- 7.) J. Blair, J. Gasteiger, C. Gillespie, P.D. Gillespie, I. Ugi, in W.T. Wipke, S. Heller, R. Feldmann, E. Hyde, "Proceedings of the NATO/CNA ASI on Computer Representation and Manipulation of Chemical Information" Wiley, New York, 1974
- 8.) C. Jochum, J. Gasteiger, *J. Chem. Inf. Comput. Sci.*, 17, 113 (1977)
- 9.) W. Schubert, I. Ugi, *J. Amer. Chem. Soc.*, in press
- 10.) H.L. Morgan, *J. Chem. Doc.*, 5, 107 (1965)
- 11.) J. Blair, J. Gasteiger, C. Gillespie, P.D. Gillespie, I. Ugi, *Tetrahedron*, 30, 1845 (1974)
- 12.) R.S. Cahn, C.K. Ingold, *J. Chem. Soc.*, 612 (1951);  
R.S. Cahn, C.K. Ingold, V. Prelog, *Experientia*, 12, 81 (1956)

- 12 -

- R.S. Cahn, C.K. Ingold, V. Prelog, *Angew. Chem.*, 8, 413 (1966)  
*Angew. Chem., Int. Ed.*, 5, 385 (1966)
- 13.) H.W. Davis, "Computer Representation of the Stereochemistry of Organic Molecules", Birkhäuser, Basel and Stuttgart, 1976
- 14.) Y. Yoneda, *Saikin no Kagakukogaku*, Maruzen, Tokyo (1970) p. 435
- 15.) J.C.J. Bart, E. Garagnani, *Z. Naturforsch.*, 31 b, 1646 (1976); 32 b, 455, 456 (1976); 32 b, 678 (1977)
- 16.) R.E. Burkhard, "Methoden der ganzzahligen Optimierung", Springer, Wien - New York, 1972
- 17.) J. Dugundji, J. Gasteiger, C. Jochum, I. Ugi, unpublished results
- 18.) J. Gasteiger, C. Jochum, *Topics Curr. Chem.*, in press
- 19.) J. Brandt, J. Friedrich, J. Gasteiger, C. Jochum, W. Schubert, I. Ugi, in E.V. Ludena, N.H. Sabelli, A.C. Wahl, "Computers in Chemical Education and Research", Plenum Publishing Corporation, New York, 1977
- 20.) W.Schubert, I. Ugi, *Topics Curr. Chem.*, in preparation
- 21.) I.Ugi, *IBM-Nachrichten* 24 180 (1974); I. Ugi, J. Gasteiger, J. Brandt, J. Brunnert, W. Schubert, *ibid*, 24, 185 (1974)
- 22.) H.J. Hiz, *J. Chem. Doc.*, 4, 173 (1964)
- 23.) L. Spialter, *J. Am. Chem. Soc.*, 85, 2012 (1963)
- 24.) L. Spialter, *J. Chem. Doc.*, 4, 261 (1964)
- 25.) L. Spialter, *J. Chem. Doc.*, 4, 269 (1964)
- 26.) E. Meyer, *Angew. Chem.*, 82, 605 (1970)
- 27.) A.J. Thakkar, *Topics Curr. Chem.*, 39, 3 (1973)

TABLE 1. Example of a Set of Valence Chemical Values for the First 18 Chemical Elements

Element symbol	Atomic Number	Core Charge	Row/Column $s_i$ Sum	$s_i$ Cross Sum	Formal Charge	Allowed Covalent Bond Order	Example
H	1	1	0 1 1 2	0 1 2 2	+1 0 0 -1	1	H <sup>+</sup> H· H-R H <sup>-</sup>
He	2	0(2)	0(2)	0(2)	0	0	
Li	3	1	0 1 1	0 1 2	+1 0 0	1	Li <sup>+</sup> Li· Li-R
Be	4	2	0 2 4	0 4 8	+2 0 -2		Be <sup>2+</sup> BeX <sub>2</sub> <sup>2-</sup> BeF <sub>4</sub> <sup>2-</sup>
B	5	3	3 4	6 8	0 -1	1, 2	BX <sub>3</sub> BF <sub>4</sub>
C	6	4	3 4 4 4 5 8	6 6 7 8 8 8	+1 0 0 0 -1 -4	1, 2, 3	C <sup>+</sup> C· C· C· C· C <sup>-</sup> C <sup>4-</sup>
N	7	5	4 4 5 6 7 8	6 8 8 8 8 8	+1 +1 0 -1 -2 -3	1, 2, 3	N <sup>+</sup> N <sup>+</sup> N· N· N <sup>-</sup> N <sup>2-</sup> N <sup>3-</sup>
O	8	6	5 6 7 8	8 8 8 8	+1 0 -1 -2	1, 2	O <sup>+</sup> O· O· O <sup>-</sup> O <sup>2-</sup>
F	9	7	6 7 8	8 8 8	+1 0 -1	1, 2	F <sup>+</sup> F· F <sup>-</sup>

TABLE 1 (continued)

Element symbol	Atomic Number	Core Charge	Row/Column Sum $s_i$	Cross Sum $s_i$	Formal Charge	Allowed Covalent Bond order	Example
Ne	10	O(8)	O(8)	O(8)	0	0	
Na	11	1	0 1 1	0 1 2	+1 0 0	1	Na <sup>+</sup> Na· Na-R
Mg	12	1	0 1 2 4	0 1 4 8	+2 +1 0 -2	1	Mg <sup>2+</sup> ·Mg <sup>+</sup> -Mg- Mg<2-
Al	13	3	0 2 3 4	0 2 6 8	+3 +1 0 -1	1	Al <sup>3+</sup>  Al <sup>+</sup> AlR <sub>3</sub> AlL <sub>4</sub> <sup>-</sup>
Si	14	5	4 5	8 10	0 -1	1 2 (Si=P)	$\begin{array}{c} \diagup \text{Si} \diagdown \\ \text{P SiH}_3 \\ \text{Ph}_3\text{SiDipy} \end{array}$
P	15	5	4 5 8 5 6	8 8 8 10 12	+1 0 -3 0 -1	1,2 (P=O) (P=N) (P=S)	$\begin{array}{c} \diagup \text{P} \diagdown \\ \text{P} \\ \text{PF}_3 \\ \text{PF}_5 \\ \text{PF}_6^- \end{array}$
S	16	7	5 6 7 8 5 6 6 7	8 8 8 10 10 12 12	+1 0 -1 -2 +1 0 0 -1	1 2 (S=O) 3 (S=N)	$\begin{array}{c} \diagup \text{S} \diagdown \\ \text{S} \\ \text{S} \\ \text{S} \\ \text{S} \\ \text{S} \\ \text{S} \\ \text{S} \\ \text{F}_3\text{S}=\text{N} \\ \text{SF}_5^- \end{array}$
Cl	17	7	6 7 7 8 6 6 7 7	8 7 8 8 9 12 12 14	+1 0 0 -1 +1 +1 0 0	1,2 (Cl-O)	$\begin{array}{c} =\text{Cl}^+ \\ \text{Cl} \\ \text{Cl} \\ \text{Cl} \\ \text{Cl} \\ \text{Cl} \\ \text{Cl} \\ \text{Cl} \\ \text{Cl} \\ \text{Cl}_2\text{O}_7 \\ \text{ClO}_3^- \\ \text{ClO}_4^- \end{array}$
Ar	18	O(8)	O(8)	O(8)	0		

