

**A MATHEMATICAL MODEL OF REALISTIC CONSTITUTIONAL
CHEMISTRY. A SYNTHON APPROACH.
I: AN ALGEBRAIC MODEL OF A SYNTHON**

Jaroslav KOČA

*Department of Organic Chemistry, Faculty of Science, Purkyně University,
Kotlářská 2, 611 37 Brno, Czechoslovakia*

Received 2 May 1988

(in final form 5 September 1988)

Abstract

A mathematical model of a synthon is suggested. The synthon is modelled by a special so-called *S*-matrix. The notion of isomeric synthons on the set of atoms *A* and that of a Family of Isomeric Synthons *FIS*(*A*) is introduced. The chemical reaction is represented by a matrix equation and it is modelled by the so-called *SR*-matrix. The notion of the reaction distance (RD) between two isomeric synthons is defined. A mathematical theory of the *S* and *SR*-matrices is developed.

0. Introduction

The question of the need for a unified mathematical model of constitutional chemistry has already been discussed in the literature [1]. Chemical synthesis is one of the fundamental and most principal part of experimental chemistry; an endeavour to transfer this chemical "art" to computers gives rise to a new field of theoretical chemistry – computer-assisted organic synthesis design. Two main trends have appeared in this field: (i) information-oriented programs, and (ii) logical structure-oriented programs [1].

The logical structure-oriented programs are based on a mathematical model of chemistry. Under the notion "mathematical model" we mean something other than a model built-up on the basis of quantum chemistry. A formulation of such a model is useful for several reasons, the most consequential one being that, at present, although the methods of quantum chemistry offer a detailed view of elementary chemical processes, its tools, due to the enormous computational difficulties and a lack of simple and straightforward interpretation of numerical results, are still very limited in giving a chemically valuable answer whether a studied reaction is feasible or not. Therefore, we believe that it is more important to form such a model which

can put "reasonable" questions to quantum chemistry concerning some elementary mechanistic steps of chemical synthesis rather than the general problem of finding the overall potential energy hypersurface. Hence, the problem of finding paths on the hypersurface can be reduced to the simpler problem of verification of a concrete path. Obviously, the most tried models in this field are Dugundji–Ugi's matrix model [1,2] and Kvasnička's graph model [3,4], respectively.

The purpose of this communication is to form a deductive mathematical model in which the actions of a synthetic chemistry would be modelled. One of the most fundamental concepts in this topic is the synthon, initially introduced by Corey [5], as that smallest part of the substrate at which some changes occur during the chemical reaction of interest. In our present theoretical consideration, the concept of the synthon will be understood in a much broader sense as a connected or a disconnected molecular substructure [6] without a relationship to a particular chemical reaction. The most rudimentary forms of the synthon are the so-called one-atomic [7] and two-atomic [8] synthons. It is to be emphasized that the suggested model is entirely topological, i.e. stereochemical aspects of synthons and their appropriate manifestation during the chemical reaction are fully ignored. The above mentioned one-atomic synthons simply correspond to the valence states of the atom [2,7,9–12]. If one uses only those valence states of atoms that are loosely classified as "chemically permissible", then the obtained model represents a primary part of realistic constitutional chemistry.

1. An algebraic model of the synthon

The model is based on an extended notion of isomerism [1,13] by the study of the chemistry of a fixed set of atoms A . The chemical constitution of the synthon $S(A)$ is determined by the localization of the chemical bonds and free valence electrons on atoms from the set A . This concept has been derived from that of an ensemble of molecules $EM(A)$ used [1,2] for the set of empirical formulae of one or several molecules constructed from atoms of the set A . Against the definition of $EM(A)$, $S(A)$ is extended so that it may involve free valences, i.e. bonds that do not connect two atoms but which only start from an atom. Such bonds are constructed as bonds between a particular real atom of the synthon and a free, so-called virtual atom, which is not specified. The concept of virtual atoms adds appreciably to the deductive power of the model, various concrete atoms or groups being conceived in place of a virtual atom in particular cases.

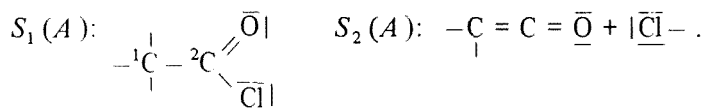
DEFINITION 1

Mathematically, a synthon is denoted by the so-called synthon BE -matrix (SBE , referred below as the S -matrix), $M = (m_{ij})$. Its off-diagonal entries are defined

as the entries of the *BE*-matrix [1,2], and the m_{ii} entry is the four-dimensional vector whose components identify the valence state of the atom A_i [2,7] as specified by the numbers of: unbonded valence electrons, single bonds, double bonds, and triple bonds, cf. example 1. It is clear that M is a symmetric square matrix of dimension n (if $A = \{A_1, \dots, A_n\}$). The headings of the rows and columns are labeled by the individual atoms from the set A .

EXAMPLE 1

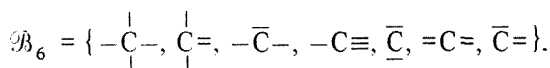
The examples of synthons on the set $A = \{C, C, O, Cl\}$ are, for example, structures:



The *S*-matrix M of the synthon $S_1(A)$ is the following:

	${}^1\text{C}$	${}^2\text{C}$	O	Cl
${}^1\text{C}$	(0, 4, 0, 0)	1	0	0
${}^2\text{C}$	1	(0, 2, 1, 0)	2	1
O	0	2	(4, 0, 1, 0)	0
Cl	0	1	0	(6, 1, 0, 0)

Let the symbol \mathcal{A}_i denote an ordered set of valence states of an atom with atomic number i ; by the symbol \mathcal{B}_i we shall denote a set of so-called "stable" (i.e. chemically permissible) valence states of an atom with atomic number i . Hence, $\mathcal{B}_i \subset \mathcal{A}_i$ and \mathcal{B}_6 is, for example,



We say that the synthon $S(A)$ is stable if each atom $A_j \in A$ with atomic number j has the valence state from the set \mathcal{B}_j . The synthons $S_1(A)$ and $S_2(A)$ from example 1 are stable.

It is often convenient to consider only a part of the given synthon $S(A)$ (e.g. one-atomic reaction center); this will be called the subsynthon [6] if the following conditions are satisfied:

DEFINITION 2

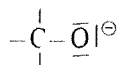
Let $S(A)$ and $S(X)$ be synthons on the set A and X , respectively, and let M and M' be their S -matrices. We say that $S(X)$ is a subsynthon of $S(A)$ (denoted as $S(X) \subset S(A)$) if

- (1) X is a subset of the set A ,
- (2) M' is a submatrix of the matrix M .

It is easy to see from the above definition that the property of "being a subsynthon" preserves the valence states of all atoms with respect to the initial synthon. The synthon



is a subsynthon of the synthon $S_1(A)$ from example 1, whereas the synthon



is not a subsynthon of the synthon $S_1(A)$.

2. Isomerism of synthons and electronic processes on synthons

The S -matrix is the static representation of the synthon. If we study a change of the synthon $S(A)$ into $S'(A)$, then the isomerism of synthons is of basic note. We have taken over the general definition of isomerism [1,2,13], which we have modified as follows.

DEFINITION 3

Let A and A' be two sets of atoms. Let $S(A)$ and $S'(A)$ be a synthon on the set A and A' , respectively. We say that $S(A)$ and $S'(A)$ are isomeric if $A = A'$.

The definition of the isomerism is much looser in our model than in ref. [1] since it does not require conservation of the total number of valence electrons, thereby enabling also nonstoichiometric processes to be modelled.

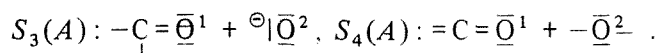
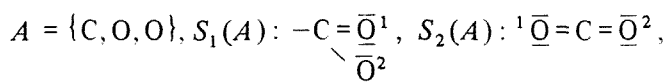
2.1. THE NOTION $FIS(A)$

The set of all isomeric synthons built on a set A will be referred to as a Family of Isomeric Synthons [6] and denoted $FIS(A)$, in analogy with $FIEM(A)$ built up as a set of all isomeric [1,2] $EM(A)$. A formal study of the topology of all synthons

built on the set A is reduced to a study of $FIS(A)$. A chemical reaction can be formally treated as an isomerization [2,13].

EXAMPLE 2

Let



The following reactions are modelled by the following isomerizations:

- (a) $S_1(A) \rightarrow S_2(A)$ decarboxylation,
- (b) $S_1(A) \rightarrow S_3(A)$, for example esterification,
- (c) $S_4(A) \rightarrow S_1(A)$ addition to $C=C$ bond of ketene.

It is easy to see from example 2 that the number of valence electrons need not be conserved. For example, it is seen that during reaction (a), electrons are "consumed", and during reactions (b) and (c), electrons are "disengaged".

2.2. A CHEMICAL REACTION AS A SERIES OF ELEMENTARY ELECTRONIC CHANGES ON THE SYNTHON

The process of a change of one synthon into another is described in our model by a special, so-called SR -matrix.

DEFINITION 4

Let $S(A)$ and $S'(A)$ be two isomeric synthons with S -matrices M and M' , respectively. We say that the SR -matrix (Synthon Reaction Matrix) of the change $S(A) \rightarrow S'(A)$ is the matrix P defined as $P = M' - M$. This operation of subtraction consists of the subtraction of matrices for the off-diagonal entries and the subtraction of vectors for the diagonal entries.

The SR -matrix describes a total electronic process during the change $S(A) \rightarrow S'(A)$. It is an analogy of the R -matrix [1,2,14] or the R -graph [4]. It follows from the definition that the SR -matrix is symmetric (because M and M' are symmetric), its diagonal entries model the global electronic changes, and the off-diagonal entry m_{ij} models the change between the atoms i and j .

EXAMPLE 3

The SR -matrix of reaction (a) from example 2 is

$$\begin{array}{rcccccc}
 & \text{C} & (0,0,2,0) & 2 & 2 & (0,2,1,0) & 2 & 1 \\
 P = & \text{O} & 2 & (4,0,1,0) & 0 & - & 2 & (4,0,1,0) & 0 \\
 & \text{O} & 2 & 0 & (4,0,1,0) & 1 & 0 & (4,2,0,0) \\
 \\
 & & (0, -2, 1, 0) & 0 & 1 & & & & \\
 = & & 0 & (0,0,0,0) & 0 & & & & \\
 & & 1 & 0 & (0, -2, 1, 0) & & & &
 \end{array}$$

It is easy to see from the definition of the *SR*-matrix that a chemical reaction is expressed in the model by matrix equation (1)

$$M' = M + P, \quad (1)$$

where *M* and *M'* are the *S*-matrices of the educt and the product, respectively, and *P* is the *SR*-matrix.

The total electronic process represented by the change $S(A) \rightarrow S'(A)$ is a set composed of the elementary steps of valence electrons reorganization (ESRE) [9,10]. The idea is based on the treatment of the elementary processes of reorganization of electrons [7,9,10] and the definition of the elementary electronic processes in the ASSOR program [15]. Four types of elementary matrix operators for modelling the elementary electronic processes [6] are introduced.

- (1) Operator α_k^{ij} for the electrofugal (from the viewpoint of the *i*th atom) dissociation of the bond between the *i*th and *j*th atoms, and operator $-\alpha_k^{ij}$ for association of the same pair of atoms; *k* denotes the multiplicity of the bond.
- (2) Operator β_k^{ij} for the nucleofugal (from the viewpoint of the *i*th atom) dissociation of the bond between the *i*th and *j*th atoms, and operator $-\beta_k^{ij}$ for association of the same pair of atoms. It is clear that $\alpha_k^{ij} = \beta_k^{ji}$ ($i \neq j$).
- (3) Operator γ_k^{ij} for the homolysis of the bond between the *i*th and *j*th atoms, and $-\gamma_k^{ij}$ for association of the radicals.
- (4) Operator δ_l^{ij} and $-\delta_l^{ij}$ for redox processes associated with the oxidation and reduction, respectively, of the *i*th atom ($l = 1$ or 2 according to the number of involved electrons).

Operators α_k^{ii} , $-\alpha_k^{ii}$, β_k^{ii} , $-\beta_k^{ii}$, γ_k^{ii} , $-\gamma_k^{ii}$, δ_l^{ii} , $-\delta_l^{ii}$ refer to situations where a virtual atom is involved instead of the *j*th atom.

The application of the elementary operators is realized by making use of eq. (1). Formally, the elementary operators are the *SR*-matrices. The concrete forms of the operators are given in table 1.

Table 1

Matrix form of elementary operators. All elements that are not given explicitly are zeroes; $\alpha_k^{ij} = \beta_k^{ji}$

Operator	$k = 1$	$k = 2$
α_k^{ij}	$ \begin{array}{ccc} & i & j \\ & \vdots & \vdots \\ i \dots & (2, -1, 0, 0) \dots & -1 \\ & \vdots & \vdots \\ j \dots & -1 \dots & (0, -1, 0, 0) \end{array} $	$ \begin{array}{ccc} & i & j \\ & \vdots & \vdots \\ i \dots & (2, 1, -1, 0) \dots & -1 \\ & \vdots & \vdots \\ j \dots & -1 \dots & (0, 1, -1, 0) \end{array} $
	$k = 3$	Process to be modelled
α_k^{ij}	$ \begin{array}{ccc} & i & j \\ & \vdots & \vdots \\ i \dots & (2, 0, 1, -1) \dots & -1 \\ & \vdots & \vdots \\ j \dots & -1 \dots & (0, 0, 1, -1) \end{array} $	$ \begin{array}{l} k = 1: I - J \rightarrow \bar{I} + J \\ k = 2: I = J \rightarrow \bar{I} - J \\ k = 3: I \equiv J \rightarrow \bar{I} = J \end{array} $
	$k = 1$	$k = 2$
α_k^{ii}	$ \begin{array}{ccc} & i \\ & \vdots \\ i \dots & (2, -1, 0, 0) \dots \end{array} $	$ \begin{array}{ccc} & i \\ & \vdots \\ i \dots & (2, 1, -1, 0) \dots \end{array} $
	$k = 3$	Process to be modelled
α_k^{ii}	$ \begin{array}{ccc} & i \\ & \vdots \\ i \dots & (2, 0, 1, -1) \dots \end{array} $	$ \begin{array}{l} k = 1: I - \rightarrow \bar{I} \\ k = 2: I = \rightarrow \bar{I} - \\ k = 3: I \equiv \rightarrow \bar{I} = \end{array} $
	$k = 1$	$k = 2$
β_k^{ii}	$ \begin{array}{ccc} & i \\ & \vdots \\ i \dots & (0, -1, 0, 0) \dots \end{array} $	$ \begin{array}{ccc} & i \\ & \vdots \\ i \dots & (0, 1, -1, 0) \end{array} $
	$k = 3$	Process to be modelled
β_k^{ii}	$ \begin{array}{ccc} & i \\ & \vdots \\ i \dots & (0, 0, 1, -1) \end{array} $	$ \begin{array}{l} k = 1: I - \rightarrow I \\ k = 2: I = \rightarrow I - \\ k = 3: I \equiv \rightarrow I = \end{array} $

Table 1 (continued)

Operator	$k = 1$	$k = 2$
γ_k^{ij}	$i \dots (1, -1, 0, 0) \dots -1$	$i \dots (1, 1, -1, 0) \dots -1$
	$j \dots -1 \dots (1, -1, 0, 0)$	$j \dots -1 \dots (1, 1, -1, 0)$
	$k = 3$	Process to be modelled
γ_k^{ij}	$i \dots (1, 0, 1, -1) \dots -1$	$k = 1: I - J \rightarrow I^* + J^*$
	$j \dots -1 \dots (1, 0, 1, -1)$	$k = 2: I = J \rightarrow I^* - J^*$
		$k = 3: I \equiv J \rightarrow I^* = J^*$
	$k = 1$	$k = 2$
γ_k^{ii}	$i \dots (1, -1, 0, 0) \dots$	$i \dots (1, 1, -1, 0) \dots$
	$k = 3$	Process to be modelled
γ_k^{ii}	$i \dots (1, 0, 1, -1) \dots$	$k = 1: I - \rightarrow I^*$
		$k = 2: I = \rightarrow I^*$
		$k = 3: I \equiv \rightarrow I^* =$
	$l = 1$	$l = 2$
δ_l^{ij}	$i \dots (-1, 0, 0, 0) \dots 0$	$i \dots (-2, 0, 0, 0) \dots 0$
	$j \dots 0 \dots (1, 0, 0, 0)$	$j \dots 0 \dots (2, 0, 0, 0)$
Process to be modelled	$I^* + J \rightarrow I + J^*$	$\bar{I} + J \rightarrow I + \bar{J}$
	$l = 1$	$l = 2$
δ_l^{ii}	$i \dots (-1, 0, 0, 0) \dots$	$i \dots (-2, 0, 0, 0) \dots$
Process to be modelled	$I^* \rightarrow I$	$\bar{I} \rightarrow I$

2.3. PROPERTIES OF *S*-MATRICES AND *SR*-MATRICES

We show some properties of *S*-matrices similar to those of *BE*-matrices previously given in [1].

DEFINITION 5

Let $B(n)$ be the set of all $n \times n$ symmetric matrices with nonnegative integer off-diagonal entries. Let each diagonal entry be a four-dimensional vector with nonnegative integer entries. Let $S[B(n)]$ be the additive Abelian group of all $n \times n$ symmetric matrices with integer off-diagonal entries. Let each diagonal entry be a four-dimensional vector with integer entries.

DEFINITION 6

Let Q_{rs} ($1 \leq r < s \leq n$) denote an $n \times n$ symmetric matrix $Q_{rs} = (q_{lp})$ whose diagonal entries are four-dimensional vectors $(0,0,0,0)$. Suppose also that each off-diagonal entry of Q_{rs} is zero except $q_{rs} = q_{sr} = 1$. Let S_{ij} ($1 \leq i \leq n$, $1 \leq j \leq 3$) denote an $n \times n$ symmetric matrix $S_{ij} = (s_{lp})$ with zero off-diagonal entries, whose diagonal entry s_{ii} is a four-dimensional vector given as follows: $s_{ll} = (0,0,0,0)$ for $i \neq l$ and $s_{ii} = (y_1, \dots, y_4)$, where $y_{j+1} = 1$ and $y_k = 0$ for $k \neq j + 1$. Let $L_{ii} = (l_{lp})$ be a matrix of the same type as matrices Q and S , with all zero entries except $l_{ii} = (1,0,0,0)$.

The following theorem may be easily proved.

THEOREM 1

- (a) The matrices Q_{rs}, S_{ij}, L_{ii} form the basis of the group $S[B(n)]$.
- (b) The group $S[B(n)]$ is a free Abelian group* of rank $n(n + 7)/2$.

DEFINITION 7

Let $A = \{A_1, \dots, A_n\}$ be a set of atoms. Let $S(A)$ be the synthon and $M = (m_{ij})$ its *S*-matrix. The following symmetric $n \times n$ matrices are defined:

* A finitely generated additive free Abelian group B is an Abelian group with a basis b_1, \dots, b_n , i.e. each element of B can be written in one, and only one, way as the linear combination $x_1 b_1 + \dots + x_n b_n$, where each x_i is integer. The number n is called the rank of B . Two free Abelian groups are isomorphic iff they have the same rank [1].

- (a) $M^E = (m_{ij}^E) : m_{ij}^E = 0$ for each $1 \leq i, j \leq n, i \neq j$;
if we denote $m_{ii}^E = (y_1, \dots, y_4)$, $m_{ii} = (x_1, \dots, x_4)$, then we define

$$y_1 = 0, y_l = x_l - \sum_{\substack{k=1 \\ k \neq i}}^n G_l(m_{ik}) \text{ for } 2 \leq l \leq 4$$

and

$$G_l(x) = \begin{cases} 1 & \text{for } x = l - 1, \\ 0 & \text{pro } x \neq l - 1. \end{cases}$$

- (b) $M^N = (m_{ij}^N) : m_{ij}^N = 0$ for each $1 \leq i, j \leq n, i \neq j$;
if we denote $m_{ii}^N = (y_1, \dots, y_4)$, $m_{ii} = (x_1, \dots, x_4)$, then we define
 $y_l = 0, y_1 = x_1$ for $2 \leq l \leq 4$.

- (c) $M^I = M - M^E - M^N$.

The respective structures $S^E(A)$, $S^N(A)$ and $S^I(A)$ associated with M^E , M^N and M^I are called the outside, N and inside components of the synthon $S(A)$.

THEOREM 2

(a) Let $B^E(n)$ denote the set of all $n \times n$ symmetric matrices with zero off-diagonal entries and diagonal entries as four-dimensional integer vectors $x^i = (x_1^i, \dots, x_4^i)$, where $x_1^i = 0$ and $x_2^i, x_3^i, x_4^i \geq 0$, for $1 \leq i \leq n$. Let $C^E(n)$ be an additive Abelian group of all $n \times n$ symmetric matrices with zero off-diagonal entries and with four-dimensional integer vectors in the main diagonal such that the first entry of each vector is 0.

(b) Let $B^N(n)$ denote the set of all $n \times n$ symmetric matrices with zero off-diagonal entries and diagonal $x^i = (x_1^i, \dots, x_4^i)$, where $x_1^i \geq 0$ and $x_2^i, x_3^i, x_4^i = 0$, for $1 \leq i \leq n$. Let $C^N(n)$ be an additive Abelian group of all $n \times n$ symmetric matrices with zero off-diagonal entries and with four-dimensional integer vectors in the main diagonal such that the first entry of each vector is an integer and the others are zero.

(c) Let $B^I(n)$ denote the set of all $n \times n$ symmetric matrices with nonnegative integer off-diagonal entries and diagonal entries as four-dimensional integer vectors $x^i = (x_1^i, \dots, x_4^i)$, where $x_1^i = 0$ and $x_2^i, x_3^i, x_4^i \geq 0$, for $1 \leq i \leq n$. Let $C^I(n)$ be an additive Abelian group of all $n \times n$ symmetric matrices with integer off-diagonal entries and with four-dimensional integer vectors in the main diagonal where the first entry of each vector is zero. Then:

- (i) $C^E(n)$ is a free Abelian group of rank $3n$ with basis S_{ij} ,
- (ii) $C^N(n)$ is a free Abelian group of rank n with basis L_{ii} ,
- (iii) $C^I(n)$ is a free Abelian group of rank $n(n+5)/2$ with basis Q_{rs}, S_{ij} .

The proof of theorem 2 is analogous to that of theorem 1 and we omit it here.

EXAMPLE 4

Let $A = \{^1C, ^2C, O, N\}$, the synthon $S(A) : \begin{matrix} \bar{O} \\ \downarrow \\ ^1C \text{---} ^2C \\ \downarrow \\ \bar{N} \end{matrix}$.

We can express the S -matrix M of the synthon $S(A)$, and matrices M^I, M^E, M^N as follows:

$$M = \begin{matrix} ^1C & (0,4,0,0) & 1 & 0 & 0 \\ ^2C & 1 & (0,2,1,0) & 2 & 1 \\ O & 0 & 2 & (4,0,1,0) & 0 \\ N & 0 & 1 & 0 & (2,1,1,0) \end{matrix} = \begin{matrix} Q_{12} + 2Q_{23} + Q_{24} + 4L_{33} \\ + 2L_{44} + 4S_{11} + 2S_{21} + S_{22} \\ + S_{32} + S_{41} + S_{42} \end{matrix}$$

$$M^I = \begin{matrix} ^1C & (0,1,0,0) & 1 & 0 & 0 \\ ^2C & 1 & (0,2,1,0) & 2 & 1 \\ O & 0 & 2 & (0,0,1,0) & 0 \\ N & 0 & 1 & 0 & (0,1,0,0) \end{matrix} = \begin{matrix} Q_{12} + 2Q_{23} + Q_{24} + S_{11} \\ + 2S_{21} + S_{22} + S_{32} + S_{41} \end{matrix}$$

$$M^E = \begin{matrix} ^1C & (0,3,0,0) & 0 & 0 & 0 \\ ^2C & 0 & (0,0,0,0) & 0 & 0 \\ O & 0 & 0 & (0,0,0,0) & 0 \\ N & 0 & 0 & 0 & (0,0,1,0) \end{matrix} = 3S_{11} + S_{42}$$

$$M^N = \begin{matrix} ^1C & (0,0,0,0) & 0 & 0 & 0 \\ ^2C & 0 & (0,0,0,0) & 0 & 0 \\ O & 0 & 0 & (4,0,0,0) & 0 \\ N & 0 & 0 & 0 & (2,0,0,0) \end{matrix} = 4L_{33} + 2L_{44}$$

DEFINITION 8

Let $V_{rs} = Q_{rs}$. Let W_{ik} ($1 \leq i \leq n, 1 \leq k \leq 3$) denote an $n \times n$ symmetric matrix with zero off-diagonal entries. Suppose also that each diagonal entry x^{jk} of W_{ik} is a four-dimensional integer vector $x^{jk} = (x_1^{jk}, \dots, x_4^{jk})$, where $x^{jk} = (0, 0, 0, 0)$ for $j \neq i, 1 \leq j \leq n, x^{i1} = (0, -1, 0, 0), x^{i2} = (0, 1, -1, 0)$ and $x^{i3} = (0, 0, 1, -1)$.

THEOREM 3

The matrices V_{rs}, W_{ik}, L_{ii} form the basis of the free Abelian group $C(n)$.

Proof

It is easy to see that matrices V_{rs}, W_{ik}, L_{ii} are linearly independent. Their number is $n(n + 7)/2$. Then from this and theorem 1, the conclusion of theorem 3 follows.

THEOREM 4

The matrices V_{rs} and W_{ik}, W_{ik} , form the basis of the free Abelian group $C^I(n)$ and $C^E(n)$, respectively.

Proof

The number of matrices V_{rs} and W_{ik}, W_{ik} , is $n(n + 5)/2, 3n$, respectively. Then from this and theorem 2, theorem 4 follows.

COROLLARY

Since the SR -matrices of the elementary operators $\alpha, \beta, \gamma, \delta$ are from $C(n)$, they must be linear combinations of the matrices V_{rs}, W_{ik}, L_{ii} . The expression of the linear combinations is shown in table 2.

Table 2
The expression SR -matrices of elementary operators by basis matrices V_{rs}, W_{ik}, L_{ii}

Operator	Linear combination
$\alpha_k^{ij} = \beta_k^{ji}$	$V_{ij} + W_{ik} + W_{jk} + 2 \cdot L_{ii}$
α_k^{ii}	$W_{ik} + 2 \cdot L_{ii}$
β_k^{ii}	W_{ik}
γ_k^{ij}	$V_{ij} + W_{ik} + W_{jk} + L_{ii} + L_{jj}$
γ_k^{ii}	$W_{ik} + L_{ii}$
δ_l^{ij}	$-l \cdot L_{ii} + l \cdot L_{jj}$
δ_l^{ii}	$-l \cdot L_{ii}$

THEOREM 5

Let $A = \{A_1, \dots, A_n\}$ be a set of atoms. Let $S(A)$ and $S'(A)$ be two isomeric synthons with S -matrices M and M' , respectively. Let $P = M' - M$. Then P is a sum of the elementary operators α^{ij} , α^{ii} , β^{ij} , γ^{ij} , γ^{ii} , δ^{ij} , δ^{ii} .

Proof

Let us consider the equality between the first and the second column of table 2. We obtain a system of seven equations with five unknowns. This system has a solution on the condition that $\gamma_k^{ij} = \alpha_k^{ij} + \delta_l^{ij}$, which is always satisfied (cf. table 2). Solving this system, we obtain, for example,

$$L_{ii} = \alpha_1^{ii} - \gamma_1^{ii}.$$

$$V_{ij} = \alpha_1^{ij} + \alpha_1^{jj} - \alpha_1^{ii} - 2\gamma_1^{jj},$$

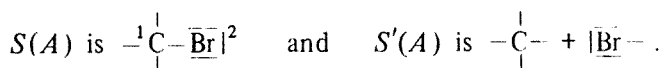
$$W_{ik} = \alpha_k^{ii} + 2\gamma_k^{ii}.$$

We can express P as a linear combination of the matrices V_{ij} , W_{ik} , L_{ii} . This proves theorem 5.

It is clear that the set of all S -matrices of all synthons $S(A) \in FIS(A)$ on the set of atoms $A = \{A_1, \dots, A_n\}$ is a subset of the set $C(n)$. Analogically, the set of all SR -matrices of changes $S(A) \rightarrow S'(A)$ is a subset of the set $C(n)$ too. So, we can express each SR -matrix as a unique linear combination of the matrices V_{ij} , W_{ik} , L_{ii} and, therefore, as a linear combination of the SR -matrices of the elementary operators α , β , γ , δ .

EXAMPLE 5

Let us consider the reaction $S(A) \rightarrow S'(A)$, where



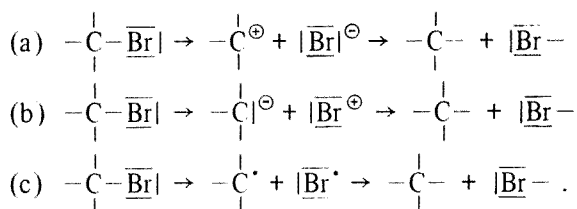
Let M, M' be the S -matrices of the synthons $S(A)$ and $S'(A)$, respectively. Then $P = M - M'$ is the SR -matrix of the change $S(A) \rightarrow S'(A)$: P can then be expressed, for example, as follows:

$$(a) \quad P = \beta_1^{12} - \beta_1^{11} - \alpha_1^{22};$$

$$(b) \quad P = \alpha_1^{12} - \alpha_1^{11} - \beta_1^{22};$$

$$(c) \quad P = \gamma_1^{12} - \gamma_1^{11} - \gamma_1^{22}.$$

Or, in chemical symbols:



It is clear that each expression of an *SR*-matrix by *SR*-matrices of the elementary operators $\alpha, \beta, \gamma, \delta$ has a different chemical meaning. In case (a), it is nucleophilic substitution, in case (b) an electrophilic substitution, and in case (c) a radical substitution.

2.4. REACTION DISTANCE

In the Dugundji–Ugi model, the notion of chemical distance (*CD*) has been introduced as the sum of absolute values of entries of the *R*-matrix [1,16]. Hence, the chemical distance reflects the total number of valence electrons that “migrate” during the reaction, and to some extent it rather expresses the thermodynamic point of view of the process under study. Our model is based mainly on kinetic aspects of the process, and so we define the new notion called the reaction distance (*RD*) of two isomeric synthons. The starting formal structure for the definition of *RD* is the graph $G_{FIS(A)}$ defined as follows.

DEFINITION 9

Let $A = \{A_1, \dots, A_n\}$ be a set of atoms. We define a graph $G_{FIS(A)}$ as the ordered couple

$$G_{FIS(A)} = (V, E),$$

where

$$V = \{S_1(A), \dots, S_m(A) \mid S_i(A) \in FIS(A)\},$$

$$E = \{ \{ x, y \}, x, y \in V, \text{ and if } M \text{ and } M' \text{ are the } S\text{-matrices of the synthon } x \text{ and } y, \text{ respectively, then } o = M' - M \text{ is from the set } O \},$$

$$O = \{ \alpha_k^{ij}, \alpha_k^{ii}, \beta_k^{ij}, \beta_k^{ii}, \gamma_k^{ij}, \gamma_k^{ii}, \delta_l^{ij}, \delta_l^{ii} \} \quad \begin{array}{l} 1 \leq i \leq n, \quad i < j \leq n, \\ 1 \leq k \leq 3, \quad 1 \leq l \leq 2 \end{array}$$

is the set of elementary operators.

Assuming that the above defined graph is connected, then we may proceed to the definition of the so-called reaction distance (*RD*).

DEFINITION 10

Let $A = \{A_1, \dots, A_n\}$ be a set of atoms. Let $S(A)$ and $S'(A)$ be two isomeric synthons. We define a reaction distance $RD(S(A), S'(A))$ between $S(A)$ and $S'(A)$ as follows:

$$RD(S(A), S'(A)) = D(x, y),$$

where x and y are the vertices of the graph $G_{FIS(A)}$, corresponding to the synthon $S(A)$ and $S'(A)$, respectively, and $D(x, y)$ is the length of the shortest path between vertices x and y (the so-called graph distance between vertices x and y).

Because $D(x, y)$ is a metric [17] on $G_{FIS(A)}$, RD is a metric on $FIS(A)$.

A part of the graph $G_{FIS(A)}$ for $A = \{C, O, O\}$ is shown in fig. 1.

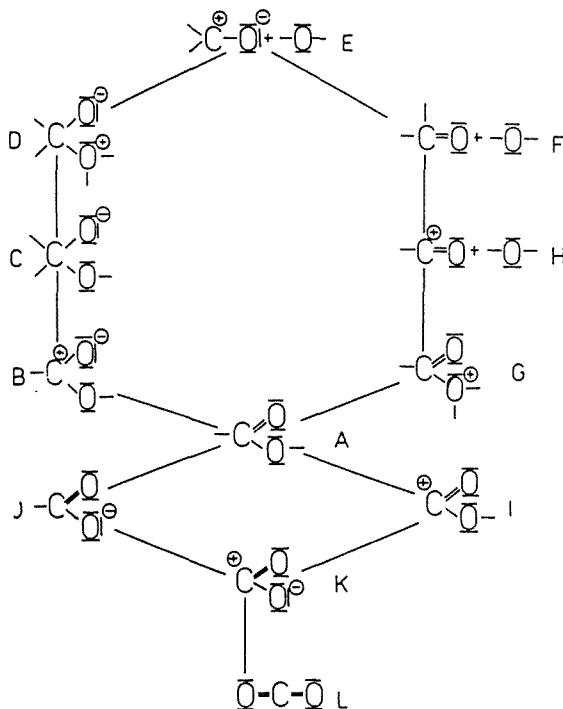


Fig. 1. A section of the graph $G_{FIS(A)}$ for $A = \{C, O, O\}$.

Much important chemical information can be obtained from the graph $G_{FIS(A)}$, for example, precursors/successors, reaction mechanisms, and reaction networks. For illustration, we obtain from the graph in fig. 1 that, for the carboxy group, the

successors/precursors can be carbon dioxide and functional derivatives of carboxylic acids.

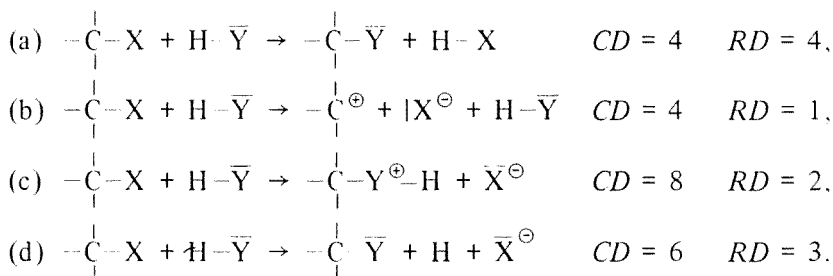
The reaction distance of two isomeric synthons is the smallest number of ESRE. For example, the RD of the synthons A and F is three and the possible path is $A \rightarrow G \rightarrow H \rightarrow F$.

It is easy to see that the graph $G_{FIS(A)}$ is very large (for example, the very rough lower estimate of the number of vertices in the graph $G_{FIS(A)}$ for $A = \{C, O, O\}$ is 23 660), but subgraphs of the graph $G_{FIS(A)}$ are manageable for practical applications. In the same way, the calculation of RD is a complicated problem for more than two-atomic synthons (for one- and two-atomic synthons, the fast algorithm has been found [8]). Therefore, the graph model of the synthon has been given in [18] and the algorithm of the calculation of RD has been implemented [19]. The algorithm is based on the algebraic approach to the generation of the minimal SR -graphs [18], with an employment of the maximal common subgraph [20] of two isomeric SR -graphs.

The concept of the reaction distance and the concept of the chemical distance are not identical. The reaction distance is associated with ESRE and reflects mainly the kinetic aspect of the process. The difference between the two distances is illustrated by example 6.

EXAMPLE 6

Let us consider reaction (a) and partial reactions (b), (c), (d) and corresponding RD and CD :



We can see that RD describes quite satisfactorily the extent of rearrangement from the educts to the products, but CD oscillates, i.e. the products are from the viewpoint of CD nearer to educts than the transition states.

The concept of reaction distance has been employed in a similar concept as the minimal number of unit reactions [21]. In our model, the definition of RD is oriented to the elementary processes of the reorganizations of electrons and, therefore, it is more useful for a reaction mechanism study.

The reaction mechanism can be understood as the time-ordered series of ESRE, i.e. the path in the graph $G_{FIS(A)}$. For example, there exist two paths from

A (e.g. carboxylic acid) to F (e.g. ester) in the graph in fig. 1. The first one, i.e. $A \rightarrow B \rightarrow C \rightarrow D \rightarrow E \rightarrow F$, models a bimolecular mechanism of esterification, the second one, i.e. $A \rightarrow G \rightarrow H \rightarrow F$, models a mechanism of creation of ester through a monomolecular mechanism via the acylium ion $-\overset{\oplus}{C}=\overline{O}$ and its interaction with alcohol. Formally, the paths from A to L represent mechanisms of decarboxylation of an acid. Since the reaction mechanism is determined as a path in the graph $G_{FIS(A)}$, the correspondence with graphs of reaction mechanisms G , which have been defined previously [22], and with the ANCOD strings [22] is possible.

References

- [1] J. Dugundji and I. Ugi, *Top. Curr. Chem.* 39(1973)19.
- [2] I. Ugi, J. Bauer, J. Brandt, J. Friedrich, J. Gasteiger, C. Jochum and W. Schubert, *Angew. Chem. Int. Ed. Engl.* 18(1979)111.
- [3] V. Kvasnička, *Coll. Czech. Chem. Commun.* 48(1983)2097; 48(1983)2118; 49(1984)1090.
- [4] V. Kvasnička, M. Kratochvíl and J. Koča, *Coll. Czech. Chem. Commun.* 48(1983)2284.
- [5] E.J. Corey, *Pure and Appl. Chem.* 14(1967)19.
- [6] J. Koča, *Coll. Czech. Chem. Commun.* 53(1988)1007.
- [7] J. Koča, M. Kratochvíl, M. Kunz and V. Kvasnička, *Coll. Czech. Chem. Commun.* 49(1984) 1247.
- [8] J. Koča, M. Kratochvíl, L. Matyska, V. Kvasnička, *Coll. Czech. Chem. Commun.* 51(1986) 2637.
- [9] M. Kratochvíl, *Chem. Listy* 77(1983)225.
- [10] M. Kratochvíl, J. Koča and V. Kvasnička, *Chem. Listy* 78(1984)1.
- [11] A. Weise, *Z. Chem.* 17(1977)100.
- [12] M. Kratochvíl, J. Koča and V. Kvasnička, *Chem. Listy* 79(1985)807.
- [13] Z. Slanina, *Teoretické aspekty javlenija izomerii v chimii* (Mir, Moscow, 1984).
- [14] M. Kratochvíl, *Chem. Listy* 75(1981)675.
- [15] W. Schubert, *MATCH* 6(1979)213.
- [16] C. Jochum, J. Gasteiger, I. Ugi and J. Dugundji, *Z. Naturforsch.* 37b(1982)1205.
- [17] Cf. J. Bosák, *Grafy a ich aplikácie* (Alfa, Bratislava, 1980).
- [18] J. Koča, *Coll. Czech. Chem. Commun.* 53(1988)3108.
- [19] J. Koča, *Coll. Czech. Chem. Commun.* 53(1988)3119.
- [20] V. Baláž, J. Koča, V. Kvasnička and M. Sekanina, *Čas. pěst. mat.* 111(1986)431.
- [21] J.B. Hendrickson, *Acc. Chem. Res.* 19(1986)274.
- [22] J. Koča, M. Kratochvíl and V. Kvasnička, *Coll. Czech. Chem. Commun.* 50(1985)1433.