A MATHEMATICAL MODEL OF REALISTIC CONSTITUTIONAL CHEMISTRY. A SYNTHON APPROACH. II: THE MODEL AND ORGANIC SYNTHESIS

Jaroslav KOČA

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

Abstract

The notion of a Synthetic Precursor/Successor (SPS) is introduced, and an algorithm for the generation of all SPSs is given. The recursive algorithm is based on the so-called "stabilization" of a reaction center with respect to the whole synthon. The notion of the immersion of a subsynthon into a synthon and the notion of the order of SPSs are introduced. A mathematical theory of SPSs is given.

0.   Introduction

Organic synthesis is a very important task in organic chemistry. Many problems connected with organic synthesis may be solved with the assistance of computers. Computer-assisted organic synthesis design is one of them. There are three basic tasks in this field: (i) synthesis in the forward direction, (ii) retrosynthesis, and (iii) the reaction network [1]. The third item has already been discussed [2] in connection with reaction mechanisms. An approach to the above first and second items may be divided into two relatively independent steps. In the first step, it is necessary to look for all possible precursors leading to a given substance. In the second step, it is necessary to evaluate the reactivity of reaction paths produced in the first step (i.e. to select appropriate reagents and reaction conditions from the kinetic as well as the thermodynamic point of view). Both of these steps can be solved subsequently or in a parallel way. Two different methods of synthetic precursors/successors (SPSs) generation can be found in the literature. In particular, there are the information approaches and those based on the logical structure [3] of chemistry (non-empirical approaches).

The information approaches use broad chemical experience and exploit it in a satisfactory manner. However, they are not able to find a new type of SPS which has not been included in a more-or-less extensive data base. Typical representatives of these programs are LHASA [4], SECS [5] and many others [6−11].

The approaches based on the logical structure of chemistry, or better, on a mathematical model, can produce, in general, new types of SPSs not yet described. Considerable over-production in the generation of a series of chemically "crazy" solutions is the main shortcoming of this category of programs at present. These are usually based on the Dugundij−Ugi model [3] (or analogues) This category is represented by, for example, the programs of the Munich group [1,12,13], by the program TOSCA [14] and by the programs for chemical reactions and reaction types generation [15,16].

Most programs used at present do not strictly belong to either of the two mentioned categories. There are programs starting from chemical experience, which generalize in such a way that their underlying model is used for the deductive prediction of organic synthesis [17−22].

The prediction of organic synthesis used in our model is based on the mathematical approach. Two principal notions are involved: (i) valence states of atoms, and (ii) a reaction distance RD [2]. The valence states and their interconversions are an empirical basis of the suggested model, whereas the reaction distance serves as a very important and effective heuristic for "oriented" generation of SPSs. A reaction center of a substrate is attacked in the course of chemical reaction by a reaction center of a reagent. At the same time, bonds localized on the reaction center are reorganized. These elementary changes are reflected in the neighbourhood of reaction centers and give rise to a reorganization of valence electrons. This process will be stopped at a moment when all synthon atoms achieve "stable" valence states, corresponding to the formation of notable products or intermediates, formally represented by the so-called stable synthon [2]. We have divided the process into two sub-processes − reorganization of valence electrons at the reaction centers and stabilization of the structures formed.

1.   Definition of the SPS

Before formulating the definition of the SPS, we first define the notions of inside and outside characteristics of the atom x during the change of a synthon; henceforth, we define a stable X-neighbourhood.

DEFINITION 1

Let $S(X)$, $S'(X)$ be two isomeric synthons and let $x \subset X$. We define the positive and the negative inside characteristic $C_x^{I^+}(S(X), S'(X))$ and $C_x^{I^-}(S(X), S'(X))$,

respectively, of the atom x involved in the change $S(X) \to S'(X)$ as the number of created and annihilated bonds of the atom x to other atoms of the set X. We define the inside characteristic $C_x^I(S(X), S'(X))$ of the atom x during the change $S(X) \to S'(X)$ as $C_x^I(S(X), S'(X)) = C_x^{I+}(S(X), S'(X)) + C_x^{I-}(S(X), S'(X))$. If card $X = 1$, then we define $C_x^{I+}(S(X), S'(X)) = C_x^{I-}(S(X), S'(X)) = 0$. Let us denote the number of the created and annihilated outside bonds (i.e. bonds from the outside component of the synthon $S(X), S'(X)$, respectively, cf. definition 7 in [2]), by the symbol k. If each outside bond of the atom x in $S(X)$ is contained in $S'(X)$ too, including multiplicity, then we define the outside characteristic $C_x^E(S(X), S'(X))$ of the atom x assigned to the change $S(X) \to S'(X)$ as $C_x^E(S(X), S'(X)) = k$; in all other cases, we put $C_x^E(S(X), S'(X)) = -k$. At the same time, we include each (σ as well as π) bond separately.

EXAMPLE 1

Let $S(X)$ be $-\text{C} \overset{\diagup \overline{\text{O}}|}{\underset{\diagdown \overline{\text{N}}<}{}}$ and $S'(X)$ be $>\text{C} \overset{\diagup |\overline{\text{O}} -}{\underset{\diagdown \overline{\text{N}}=}{}}$. Then we have:

$$C_C^{I-}(S(X), S'(X)) = C_O^{I-}(S(X), S'(X)) = 1, \; C_N^E(S(X), S'(X)) = -2,$$

$$C_C^{I+}(S(X), S'(X)) = C_O^{I+}(S(X), S'(X)) = 0, \; C_N^I(S(X), S'(X)) = 0,$$

$$C_C^E(S(X), S'(X)) = C_O^E(S(X), S'(X)) = 1.$$

THEOREM 1

Let $X = \{x_1, \ldots, x_n\}$ be a set of atoms. Let $S(X), S'(X)$ be two isomeric synthons. Let $P = (p_{ij})$ and $Q = (q_{ij})$ be the S-matrices of the inside and outside components of the synthon $S(X)$, respectively. Let $R = (r_{ij})$ and $S = (s_{ij})$ be the S-matrices of the inside and outside components of the synthon $S'(X)$, respectively (cf. definition 7 in [2]). Let $V = (v_{ij}) = R - P$, $W = (w_{ij}) = S - Q$, $v_{ii} = (v_1^i, \ldots, v_4^i)$, $w_{ii} = (w_1^i, \ldots, w_4^i)$. Then for each $x = x_i \in X$, $1 \leqslant i \leqslant n$:

(1) $\quad C_x^{I+}(S(X), S'(X)) = \displaystyle\sum_{\substack{1 \leqslant j \leqslant n \\ j \neq i}} F(v_{ij}), \; C_x^{I-}(S(X), S'(X)) = \displaystyle\sum_{\substack{1 \leqslant j \leqslant n \\ j \neq i}} F(-v_{ij}),$

$\qquad C_x^I(S(X), S'(X)) = \displaystyle\sum_{\substack{1 \leqslant j \leqslant n \\ j \neq i}} |v_{ij}|, \;$ where $F(t) = \begin{cases} t & \text{for } t > 0 \\ 0 & \text{for } t \leqslant 0 \end{cases}$

(2) $\quad C_x^E(S(X), S'(X)) = \begin{cases} w_2^i + 2w_3^i + 3w_4^i, & \text{if } w_j^i \geqslant 0 \text{ for } j = 2, 3, 4 \\ -|w_4^i| - |w_3^i| + w_4^i| - |w_2^i| + w_3^i| + w_4^i| & \text{in the other} \end{cases}$

<div align="right">cases.</div>

Proof

The assertion (1) is true because the matrix V is the SR-matrix of the change of the inside component of the synthon $S(X)$ into the inside component of the synthon $S'(X)$. Let us prove assertion (2). If $w_j^i \geqslant 0$, then the outside bonds have been created only on the atom x and their total sum is the sum of single bonds plus twice double bonds plus three times triple bonds created. Let us suppose that at least one bond has been annihilated. We can prove the assertion following the idea used in [23]. Three cases of annihilation of a bond can occur. We can describe these three cases by the vectors t_1, t_2, t_3. For annihilation of a single bond, we have $t_1 = (0, -1, 0, 0)$, for annihilation of a double bond $t_2 = (0, 1, -1, 0)$, and for annihilation of a triple bond $t_3 = (0, 0, 1, -1)$. The vectors t_1, t_2, t_3 form the basis of the subspace in $E^{(4)}$. All vectors w_{ii} are resident in this subspace because the first component of each w_{ii} is zero. Consequently, we can express each vector w_{ii} in the basis t_1, t_2, t_3 as follows: $(w_1^i, w_2^i, w_3^i, w_4^i) = a(0, -1, 0, 0) + b(0, 1, -1, 0) + c(0, 0, 1, -1)$. Solving this equation, we obtain: $a = -w_2^i - w_3^i - w_4^i$, $b = -w_3^i - w_4^i$, $c = -w_4^i$. Analogically, for creating bonds we can take the vectors $-t_1, -t_2, -t_3$. Since each vector t_1, t_2, t_3 describes annihilation of one bond (σ or π), we have

$$C_x^E(S(X), S'(X)) = -(|a| + |b| + |c|)$$

$$= -(|-w_2^i - w_3^i - w_4^i| + |-w_3^i - w_4^i| + |-w_4^i|)$$

$$= -|w_2^i + w_3^i + w_4^i| - |w_3^i + w_4^i| - |w_4^i|$$

and the assertion is proved.

The preceding theorem gives a fast method for the calculation of the inside and outside characteristics.

DEFINITION 2

Let $A = \{A_1, \ldots, A_n\}$ be a set of atoms, $I = 1, \ldots, n$. Let $S(A), S'(A)$ be two isomeric synthons and let $P = (p_{ij})$ be the SR-matrix of the change $S(A) \to S'(A)$. We say that the change $S(A) \to S'(A)$ satisfies the connectivity condition (furthers the condition ($\star$) only) if there does not exist a decomposition of the set $I = I_1 \cup I_2$, $I_1 \cap I_2 = \emptyset$, $I_1, I_2 \neq \emptyset$ such that $p_{ij} = 0$ for each $i \in I_1$ and $j \in I_2$, respectively, $p_{ij} \neq 0$ for some $i, j \in I_1$ and $p_{ij} \neq 0$ for some $i, j \in I_2$, and, finally, $p_{ii} = 0$ if and only if $|p_1^i| + \ldots + |p_4^i| = 0$.

DEFINITION 3

Let $A = \{A_1, \ldots, A_n\}$ be a set of atoms, $S(A)$ a synthon and $S(X) \subset S(A)$. A stable X-neighbourhood of the synthon $S(A)$ generated by the synthon $S'(X)$ is

the set of all stable synthons $S'(A)$ [2] for which $S'(X) \subset S'(A)$ and the isomeric synthons $S(X)$, $S'(X)$ satisfy the following five conditions:

(1) $RD(S(X), S'(X)) > 0$.

(2)   The change $S(X) \rightarrow S'(X)$ satisfies the condition ($\star$).

(3)   If $X'' \subset X$ is a set such that card $X'' > 1$ and the change $S(X'') \rightarrow S'(X'')$ satisfies the condition ($\star$), then $RD(S(X''), S'(X'')) \leqslant$ card $X'' + 1$.

(4) $C_x^I(S(X), S'(X)) + |C_x^E(S(X), S'(X))| \leqslant 2$, for each $x \in X$.

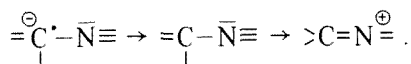(5) $C_x^E(S(X), S'(X)) < 2$, for each $x \in X$.

The stable X-neighbourhood of the synthon $S(A)$ is the union of all stable X-neighbourhoods of the synthon $S(A)$ generated by all synthons $S'(X)$ satisfying conditions $1-5$.

The chemical interpretation of conditions $1-5$ from definition 3 is as follows:

(1) There must be at least one electronic change on the atoms of the reaction center.

(2) The global distribution of electrons assigned to the change $S(X) \rightarrow S'(X)$ must be continuous, i.e. a reaction course consisting of two or more mutally independent reactions is forbidden.

(3) This condition minimizes the number of ESRE [2] which must be realized on the atoms from the set X during the formation of the stable synthon $S'(X)$. For example, the reaction of the hypothetical radical

$$=\overset{\ominus}{\underset{|}{C}}{}^{\bullet}-\overset{}{\underset{|}{\bar{N}}}= \quad \rightarrow \quad >C=N\overset{\oplus}{=}$$
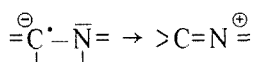
will be split by condition (3) into two reactions

$$=\overset{\ominus}{\underset{|}{C}}{}^{\bullet}-\bar{N}\equiv \quad \rightarrow \quad =\underset{|}{C}-\bar{N}\equiv \quad \rightarrow \quad >C=N\overset{\oplus}{=} \ .$$

Accordingly, the synthon

$$>C=N\overset{\oplus}{=}$$
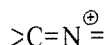
is not in the stable X-neighbourhood of the synthon

$$=\overset{\ominus}{\underset{|}{C}}{}^{\bullet}-\bar{N}\equiv$$
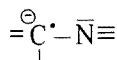
but it is in the stable X-neighbourhood of the synthon

$$=\underset{|}{C}-\bar{N}\equiv$$
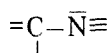
(if $X = A = \{C, N\}$).

(4) Each of the atoms of the reaction center X can take part maximally in two elementary processes on the bonds.

(5) During the change $S(X) \rightarrow S'(X)$ at most one new outside bond at any $x \in X$ can be formed.

EXAMPLE 2

Let us consider the synthons:

$$S(A): \quad -{}^4\overset{|}{\underset{|}{C}}-{}^3\overset{|}{\underset{|}{C}}-{}^2\overset{|}{\underset{|}{C}}-{}^1 C \overset{\nearrow O}{\underset{\searrow \overline{O}-H}{}} , \qquad S_1'(A): \quad -\overset{|}{\underset{|}{C}}-\overset{|}{\underset{|}{C}}-\overset{|}{\underset{|}{C}}-C \overset{\nearrow |\overline{O}-}{\underset{\searrow \underline{O}-H}{}} ,$$

$$S_2'(A): \quad -\overset{|}{\underset{|}{C}}-\overset{|}{\underset{|}{C}}-C=C=\overline{\underline{O}}{}^5 + -\overline{\underline{O}}-H, \qquad S_3'(A): \quad -\overset{|}{\underset{|}{C}}-\overset{|}{\underset{|}{C}}-\overset{\vee}{\underset{|}{C}} + C \overset{\nearrow \overline{O}|}{\underset{\searrow \underline{O}|}{}} + H-,$$

$$S_4'(A): \quad {}^3_{\nwarrow}C=C{}^3_{\nearrow} + {}^3_{\nwarrow}C \overset{\lrcorner}{=} C=\overline{\underline{O}}{}^5_{\nwarrow} + -\overline{\underline{O}}-H, \quad S_5'(A): \quad \equiv C-\overset{|}{\underset{|}{C}}-C=C=\overline{\underline{O}}{}^5 + -\overline{\underline{O}}-H,$$

$$S_6'(A): \quad -{}^2\overset{|}{\underset{\underset{|}{{}^3\overset{}{C}}}{C}}-{}^1 C \overset{\nearrow |\overline{O}-H}{\underset{\searrow |O|^5}{}} \quad \begin{matrix}\\ -{}^3\ove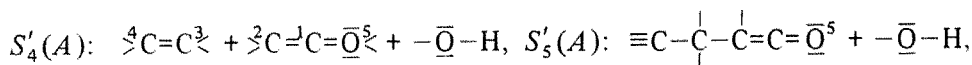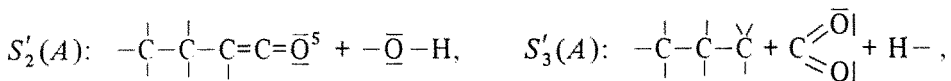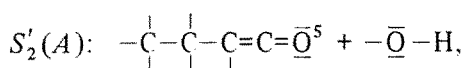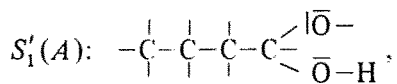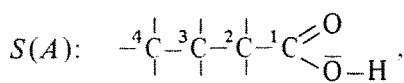rset{|}{\underset{|}{C}}-{}^4\overset{}{C}{\underset{\diagup}{\diagdown}}\end{matrix} , \qquad S_7'(A): \quad -\overset{|}{\underset{|}{C}}-\overset{|}{\underset{|}{C}}=C-C \overset{\nearrow \overline{O}|}{\underset{\searrow |\underline{O}-H}{}} ,$$

$$S_8'(A): \quad -\overset{|}{\underset{|}{C}}-\overset{|}{\underset{|}{C}}-\overset{|}{\underset{|}{C}}=C=\overline{\underline{O}}{}^5- + -\overline{\underline{O}}-H, \quad S_9'(A): \quad -\overset{|}{\underset{|}{C}}-\overset{|}{\underset{|}{C}}-\overset{|}{\underset{|}{C}}=C-\overline{\underline{O}}{}^5- + -\overline{\underline{O}}-H.$$
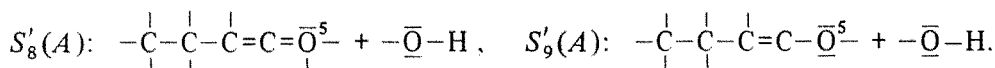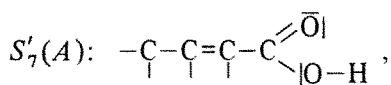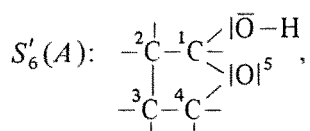
Let $X = \{C^1, O^5\}$. Then $S(X)$ is: $>C=\overline{\underline{O}}$. The elements of the stable X-neighbourhood of the synthon $S(A)$ are the synthons $S_1'(A), \ldots, S_6'(A)$. At the same time $S_1'(A)$ and $S_6'(A)$ are generated by the synthon $S'(X)$ which is $>C-\overline{\underline{O}}-$. The synthons $S_2'(A), \ldots, S_5'(A)$ are generated by the synthon $S'(X)$: $=C=\overline{\underline{O}}$. The synthons $S_7'(A), \ldots, S_9'(A)$ are not elements of the stable X-neighbourhood because $S_7'(A)$ does not satisfy condition (1) of definition 3, $S_8'(A)$ does not satisfy conditions (2) and (5) of this definition, and the synthon $S_9'(A)$ does not satisfy condition (4) (the sum for the atom C^1 is equal to three). An example of a synthon which does not satisfy condition (3) has been mentioned within the framework of the chemical interpretation of the stable X-neighbourhood.

THEOREM 2

Let $S(A)$ and $S'(A)$ be two isomeric synthons. Then,

$$RD(S(A), S'(A)) \geqslant \left[\sum_{x \in A} C_x^I(S(A), S'(A)) \right] \Big/ 2 + \sum_{x \in A} |C_x^E(S(A), S'(A))| .$$

Proof of theorem 2 follows directly from the definition of the inside and outside characteristics, respectively, because at least one step ESRE is needed for formation/annihilation of one bond, and each formed/annihilated inside bond increases the inside characteristic of two atoms.

COROLLARY

Let $S(A), S'(A)$ be two isomeric synthons, and M and M' their S-matrices, respectively. Let $P = (p_{ij}) = M' - M$. Then,

$$RD(S(A), S'(A)) \geqslant \sum_{\substack{i = 1, \ldots, n-1 \\ j = i+1, \ldots, n}} |p_{ij}| .$$

Proof

From theorems 1 and 2, we have:

$$RD(S(A), S'(A)) \geqslant \left[\sum_{x \in A} C_x^I(S(A), S'(A)) \right] \Big/ 2 + \sum_{x \in A} |C_x^E(S(A), S'(A))|$$

$$\geqslant \left[\sum_{x \in A} C_x^I(S(A), S'(A)) \right] \Big/ 2 = \sum_{\substack{i = 1, \ldots, n-1 \\ j = i+1, \ldots, n}} |p_{ij}| .$$

THEOREM 3

Let $S(A), S'(A)$ be two isomeric synthons, let $X \subset A, S(X) \subset S(A)$ and $S'(X) \subset S'(A)$. Then,

$$RD(S(A), S'(A)) \geqslant RD(S(X), S'(X)).$$

Proof

If $k = RD(S(X), S'(X))$, then we need at least a series of k elementary operators for modelling the change $S(X) \to S'(X)$. Each of them is localized between two atoms. Let k_1 be a number of the operators which are localized only between the atoms from the set X. Let k_2 be a number of operators which are localized between an

atom x from X and a virtual atom. Evidently, $k = k_1 + k_2$. Let $l = RD(S(A), S'(A))$ and let $l = l_1 + l_2$ where l_1, l_2 are defined in the same way as k_1, k_2 but for the set A. Then obviously, $l_1 \geqslant k_1$, $l_2 + (l_1 - k_1) \geqslant k_2$ and consequently, $l = l_1 + l_2 \geqslant k_1 + k_2 = k$.

THEOREM 4

Let $S(A), S'(A)$ be two isomeric one-atomic synthons. Let $A = \{x\}$ and x be an atom of a main group element. Let us suppose that the change of the formal charge[*] on the atom x is during the change $S(A) \rightarrow S'(A)$ two, maximally. Let $RD(S(A), S'(A)) > 2$. Then $|C_x^E(S(A), S'(A))| > 2$.

Proof

Let us suppose that $RD(S(A), S'(A)) = 3$, i.e. three steps of ESRE have been realized on the atom x. Then the following four cases can set in:

(a) All steps have been redox $\Rightarrow$ the change of the charge is five, minimally. This is in contradiction with the assumption. Accordingly, (a) cannot set in.

(b) Two steps have been redox $\Rightarrow$ the change of the charge is three, minimally. This is in contradiction, too, and (b) cannot set in.

(c) One step has been redox and two next steps have been connected with the formation/annihilation of bonds. Then there must be $|C_x^E(S(A), S'(A))| \geqslant 2$. Let $|C_x^E(S(A), S'(A))| = 2$. Then the following three cases can set in:

(c1) $C_x^E(S(A), S'(A)) = 2 \Rightarrow$ two new bonds have been formed and one step has been redox. If this is an oxidation, then at least five electrons had to be in the valence shell before the process, and there have to be, in this case, two electrons there after the process. If the process has been reduction, then we can derive in the same way that there must be five electrons in the valence shell, minimally, after the process, and there were two electrons in the core, in this case, before the process. In both cases the change of the charge has been three, in contradiction with the assumption.

(c2) $C_x^E(S(A), S'(A)) = -2$ and one bond has been formed/annihilated, respectively, and one step has been redox. Similarly, as above we can show that in this case the change of the charge must be at least three, too.

(c3) For the case where $C_x^E(S(A), S'(A)) = -2$ and two bonds have been annihilated and one step has been redox, we can prove the assertion analogically as in (c1).

---

[*]If the vector v of the valence state of the atom x is $v = (v_1, \ldots, v_4)$ and x is a main group element from the nth group of the periodic system, then we express the formal charge c as $c = n - (v_1 + v_2 + 2v_3 + 3v_4)$.

(d) No step has been redox. Because the increment of RD by each formation/annihilation of a bond is one, then there must be $|C_x^E(S(A), S'(A))| = 3 > 2$.

We can use the above-mentioned considerations also for $RD(S(A), S'(A)) > 3$.

COROLLARY

Let $S(A)$, $S'(A)$ be two isomeric synthons, $S(X) \subset S(A)$. Let $S'(A)$ be from the stable X-neighbourhood of the synthon $S(A)$. Then, for each $x \in X$ it holds that $RD(S(\{x\}), S'(x)) \leqslant 2$.

Proof

Let X' be a set of all atoms from X on which a change has been realized during the process $S(A) \to S'(A)$. For $x \in X - X'$, the assertion is clear because we have for such x: $RD(S(\{x\}), S'(x)) = 0$. Suppose, on the contrary, that $RD(S(x), S'(\{x\})) > 2$ for some $x \in X'$. Then theorem 4 implies

$$2 < |C_x^E(S(x), S'(\{x\}))| \leqslant C_x^I(S(\{x\}), S'(\{x\})) + |C_x^E(S(\{x\}), S'(\{x\}))|,$$

which is in contradiction to condition 4 from definition 3. Therefore, the corollary is proved.

The corollary of theorem 4 is important for the restriction of a combinatorial explosion during the generation of the stable X-neighbourhood of the synthon $S(A)$ since we can take only valence states for which the condition is satisfied for the generation.

THEOREM 5

Let $S(A)$, $S'(A)$ be two isomeric synthons, $S(X) \subset S(A)$. Let $S'(A)$ be from the stable X-neighbourhood of the synthon $S(A)$. Let $S'(X) \subset S'(A)$ be isomeric to $S(X)$. Then,

$$\sum_{x \in X} |C_x^E(S(X), S'(X))| \leqslant 2.$$

Proof

Let X' be the set of all atoms from X on which changes occurred, in any way, during the process $S(A) \to S'(A)$. Let $k = \text{card } X'$. From definition 3, theorem 2 and theorem 4, we have:

$$k + 1 \geqslant RD(S(X'), S'(X'))$$

$$\geqslant \left[\sum_{x \in X'} C_x^I(S(X'), S'(X')) \right] \Big/ 2 + \sum_{x \in X'} |C_x^E(S(X'), S'(X'))|.$$

Further, from condition 2 of definition 3 and from theorem 1, we have.

$$\left[\sum_{x \in X'} C_x^I(S(X'), S'(X')) \right] \Big/ 2 \geqslant k - 1.$$

Accordingly,

$$k + 1 \geqslant k - 1 + \sum_{x \in X'} |C_x^E(S(X'), S'(X'))|,$$

and following

$$2 \geqslant \sum_{x \in X'} |C_x^E(S(X'), S'(X'))|.$$

Since the increment of each atom from the set $X - X'$ to the sum of the absolute values of the outside characteristics is zero, the assertion is proved.

COROLLARY

Let $S(A)$, $S'(A)$ be two isomeric synthons, $S(X) \subset S(A)$. Let $S'(A)$ be from the stable X-neighbourhood of the synthon $S(A)$. Then only the following possibilities can set it:

(1) There does not exist $x \in X$ such that $C_x^E(S(X), S'(X)) \neq 0$.

(2) There exists only one $x \in X$ such that $|C_x^E(S(X), S'(X))| = 1$ or $|C_x^E(S(X), S'(X))| = 2$, and $C_y^E(S(X), S'(X)) = 0$, for any $y \neq x$, $y \in X$.

(3) There exist $x_1, x_2 \in X$ ($x_1 \neq x_2$) such that $|C_{x_1}^E(S(X), S'(X))| = 1$ and $|C_{x_2}^E(S(X), S'(X))| = 1$.

DEFINITION 4

Let $A = \{A_1, \ldots, A_n\}$ be a set of atoms, and let $S(A)$, $S'(A)$ be two isomeric synthons on the set A. Let $S(X) \subset S(A)$. Let $M = (m_{ij})$ and $M' = (m_{ij}')$ be the S-matrices of the synthons $S(A)$ and $S'(A)$, respectively. Let $P = (p_{ij})$ be the SR-matrix of the change $S(A) \to S'(A)$. Let $S'(A)$ be an element of the stable X-neighbourhood of the synthon $S(A)$. Let I, J be two sets defined as follows: $J = \{1, 2, \ldots, n\}$,

$I = \{i \mid A_i \in X\}$. We say that $S'(A)$ is the SPS of $S(A)$ with respect to the reaction center X if:

(1) the change $S(A) \to S'(A)$ satisfies the condition ($\star$),

(2) $C_x^I(S(A), S'(A)) + |C_x^E(S(A), S'(A))| \leqslant 2$, for each $x \in A$,

(3) if $m_{ij} = 0$, then $m'_{ij} = 0$, for any $i \in J$ and $j \in J - I$ such that $i \neq j$,

(4) let us define the sets:

$$Y_0 = X,$$

$$Y_1 = \{A_j \in A - X \mid \text{there exists } A_i \in X \text{ such that } p_{ij} \neq 0\},$$

$$\vdots$$

$$Y_l = Y_{l-1} \cup \{A_j \in A - (Y_{l-1} \cup X) \mid \text{there exists } A_i \in Y_{l-1} \text{ such that } p_{ij} \neq 0\} \quad \text{for} \quad l = 2, 3, \ldots$$

Since the set A is finite, there must exist a number k such that $Y_{k+1} = Y_k$. Then we request that the following implication be satisfied:

$$k \geqslant 1 \Rightarrow \text{for each } l = 1, 2, \ldots, k \text{ and for each } y \in Y_l:$$

$$RD(S(X \cup Y_{l-1} \cup \{y\}), S'(X \cup Y_{l-1} \cup \{y\}))$$

$$- RD(S(X \cup Y_{l-1}), S'(X \cup Y_{l-1})) \leqslant 1 .$$

We denote the set of all SPSs of the synthon $S(A)$ with respect to the reaction center X by $\mathscr{S}(S(A/X))$.

REMARK 1

It follows from the corollary of theorem 4 that if $S'(A) \in \mathscr{S}(S(A/X))$, then $RD(S(\{x\}), S'(\{x\})) \leqslant 2$, for each $x \in A$.

The chemical interpretation of the conditions from definition 4 is as follows.

(1) and (2) The chemical meaning of these conditions is identical to the meaning of conditions (2) and (4), respectively, from definition 3 for the whole synthon $S(A)$.

(3) The formation of a new bond is forbidden between an atom outside X and an arbitrary atom from A. The formation of this one is realized by the connection of a new virtual atom. Condition (3) stops a combinatorial explosion, and it adds to the deductive power of the model. It follows also from condition (3) that substitutions and rearrangements must be modelled on two atomic reaction centers, minimally. The connection of atoms forbidden by condition (3) can be realized by the physicalization of the model.

(4) This condition is analogous to condition (3) from definition 3.

EXAMPLE 3

Let us consider the synthons $S(A)$ and $S'_1(A), \ldots, S'_6(A)$ from example 2. Then only the synthons $S'_1(A), \ldots, S'_4(A)$ are from the set $\mathcal{G}(S(A/X))$. The synthons $S'_5(A)$ and S''_6 do not satisfy conditions (1) and (3) of definition 4, respectively (the atoms C^4 and O^5 are bonded in $S'_6(A)$ and they are not in $S(A)$).

It is clear from the definition SPS that the set $\mathcal{G}(S(A/X))$ is dependent on the choice of the reaction center X. By the design of retrosynthesis, it is suitable to select as the reaction center the part of the skeleton or functional group which may be synthesized in the last step.

2. Stabilization and construction of the set $\mathcal{G}(S(A/X))$

We can obtain the set $\mathcal{G}(S(A/X))$ in two ways. The first one is to find all synthons $S'(A)$ forming the stable X-neighbourhood and to select all synthons satisfying the conditions of definition 4. Such a way is very time and memory consuming because the stable X-neighbourhood is very large. Therefore, we select the second one, i.e. the generative way.

The choice of the generative way induces also the separation of the definition of the SPS into two parts. The first one is the definition of the stable X-neighbourhood, and the second one is the appropriate definition of the SPS. The problem of the generation of the synthons $S'(X)$ from definition 3 is closely related to a well known and solved problem in chemistry, the problem of the generation of structures (cf. [24–27]). However, we cannot use these methods for the appropriate generation of the whole SPS because, in this case, they are very time consuming. Therefore, we introduce the notions of a stabilization and an immersion of synthons.

DEFINITION 5

Let $A = \{A_1, \ldots, A_n\}$ be a set of atoms. Let $S(A)$ and $S'(A)$ be two isomeric synthons. Let $M = (m_{ij})$ and $M' = (m'_{ij})$ be the S-matrices of the synthons $S(A)$ and $S'(A)$, respectively. Let $S(X) \subset S(A)$, $S'(X) \subset S'(A)$. Let I, J be sets defined as follows: $I = \{i \mid A_i \in X\}$, $J = \{1, 2, \ldots, n\}$. We say that the synthon $S'(A)$ was formed by an immersion of the synthon $S'(X)$ into the synthon $S(A)$ if the equality $m_{ij} = m'_{ij}$ holds for each $i \in J - I$ and $j \in J$.

EXAMPLE 4

$$S(A) : -\overset{|}{\underset{|}{C}}-\overset{|}{\underset{|}{C}}-C\overset{\overline{O}|}{\underset{|\underline{O}-H}{\diagup}} , \qquad S(X) : -C\overset{\diagup}{\underset{\underline{O}-}{\diagdown}} ,$$

$$S'(A) : -\overset{|}{\underset{|}{C}}-\overset{|}{\underset{|}{C}}-C\overset{\overline{O}|}{\underset{\diagdown}{\diagup}} + -\underline{O}-H, \quad S'(X) : -C\overset{\diagup}{\diagdown} + -\underline{O}- .$$

The synthon $S'(A)$ has been formed by immersion of $S'(X)$ into $S(A)$.

A sufficient condition for the existence of the immersion is formulated by the following theorem.

THEOREM 6

Let $S(A)$ be a synthon, $S(X) \subset S(A)$. Let $S'(X)$ be isomeric to $S(X)$ and $C_x^E(S(X), S'(X)) \geqslant 0$ for each $x \in X$. Then we can immerse $S'(X)$ into $S(A)$.

Proof

Let the assumptions of the theorem be satisfied. Hence, we have that all outside bonds on all atoms from the set X are conserved, and we can immerse $S'(X)$ into $S(A)$.

The necessary and sufficient condition for the existence of the immersion is stated in the the following theorem.

THEOREM 7

Let $A = \{A_1, \ldots, A_n\}$ be a set of atoms, $S(A)$ be a synthon, $S(X) \subset S(A)$ and $S'(X)$ be isomeric to $S(X)$. Let $M = (m_{ij})$ be the S-matrix of the inside component of the synthon $S(A)$ with the diagonal entry $m_{ii} = (m_1^i, \ldots, m_4^i)$. Let $P = (p_{ij})$ and $Q = (q_{ij})$ be the S-matrices of the inside and outside components of the synthon $S(X)$ and $S'(X)$, respectively, with diagonal entries $p_{ii} = (p_1^i, \ldots, p_4^i)$ and $q_{ii} = (q_1^i, \ldots, q_4^i)$. Suppose that the numberings of the atoms in $S(A), S(X)$ and $S'(X)$ are identical. Let us define the set I as $I = \{i \mid A_i \in X\}$. Then it is possible to immerse the synthon $S'(X)$ into $S(A)$ iff the inequalities $m_j^i - p_j^i \leqslant q_j^i$ are satisfied for each $i \in I$ and $j = 2, 3, 4$.

Proof

Theorem 7 follows evidently from the definition of the inside and outside components [2] and from definition 5.

DEFINITION 6

Let $A = \{A_1, \ldots, A_n\}$ be a set of atoms. Let $S(A), S(X)$ be two synthons, $S(X) \subset S(A)$. Let $S'(X)$ be isomeric to $S(X)$ and satisfying conditions (1)–(5) from definition 4. Let $Q = (q_{ij})$ be the S-matrix of the synthon $S(A)$. Let us define the following sets: $I = \{i \mid A_i \in X\}$, $C = \{x \in X \mid C_x^E(S(X), S'(X)) < 0\}$, $Y = \{A_j \in A \mid A_j \notin X$ and there exists $i \in I$ such that $A_i \in C$ and $q_{ij} \neq 0\}$, $E = \{j \mid A_j \in Y\}$, $K = \{k \mid A_k \in C\}$, $L = I \cup E$. Let $M = (m_{ij})$ and $M' = (m'_{ij})$ be the S-matrices of the synthons $S(X \cup Y)$ and $S'(X \cup Y)$, respectively. Put $P = (p_{ij}) = M' - M$.

We say that the synthon $S'(X \cup Y)$ has been created by the stabilization of the synthon $S'(X)$ with respect to the synthon $S(A)$ into first neighbourhood if:

(1) the synthon $S'(X \cup Y)$ is stable;

(2) the change $S(X \cup Y) \rightarrow S'(X \cup Y)$ satisfies the condition ($\star$);

(3) the inequality

$$C_y^I(S(X \cup Y, S'(X \cup Y)) + |C_y^E(S(X \cup Y), S'(X \cup Y))| \leqslant 2$$

holds for each $y \in X \cup Y$;

(4) there exists $k \in K$ such that $C_{A_k}^I(S(A_k \cup Y), S'(A_k \cup Y)) > 0$;

(5) $p_{ij} = 0$, for each $i \in I - K$ and for each $j \in E$;

(6) $m_{ij} = 0 \Rightarrow m'_{ij} = 0$, for each $i \in E \cup K, j \in E, i \neq j$;

(7) the inequality

$$RD(S(X \cup \{y\}), S'(X \cup \{y\})) - RD(S(X), S'(X)) \leqslant 1$$

holds for each $y \in Y$.

If $Y = \emptyset$ or if there does not exist the synthon $S'(X \cup Y)$ satisfying conditions (1)–(7), then we say that there is no possibility of stabilizing the synthon $S'(X)$ with respect to the synthon $S(A)$ into first neighbourhood. The stabilization into first neighbourhood we shall also call 1-stabilization.

The chemical interpretation of conditions (1)–(7) from the definition of 1-stabilization is as follows. The meaning of condition (1) is clear, conditions (2) and (3) have the same meaning as conditions (1) and (2) from definition 4, respectively. Condition (3) is important for the algorithm SPS-GEN mentioned later. Without this condition, the algorithm would not be finite. Conditions (5) and (6) have a similar meaning to condition (3) from definition 4, and condition (7) is analogous to condition (4) from definition 4. Generally, the stabilization expresses a projection of changes of the reaction center into its nearest neighbourhood. The aim of this approach is to obtain a new stable synthon.
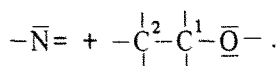
EXAMPLE 5

Let $A = \{C^1, C^2, C, C, C, N, O, H, H\}$, $X = \{C^1, N\}$. Let $S(A)$ be:
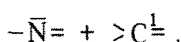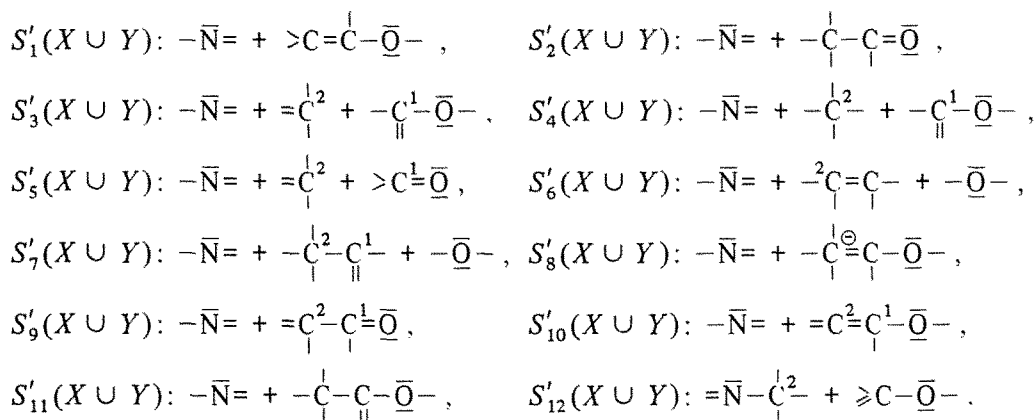
$$H-\bar{N}=\overset{|}{C}-\overset{|}{C}-\overset{|}{C}-\overset{|}{C}{}^2\!-\overset{|}{C}{}^1\!-\underline{\bar{O}}-H.$$

Then $S(X)$ is:

$$-\bar{N}= \; + \; -\overset{|}{\underset{|}{C}}{}^1\!-$$

and $S(X \cup Y)$ is:

$$-\bar{N}= \; + \; -\overset{|2}{\underset{|}{C}}-\overset{|1}{\underset{|}{C}}-\underline{\bar{O}}- \; .$$

Let $S'(X)$ be:

$$-\bar{N}= \; + \; >C\overset{1}{=} \; .$$
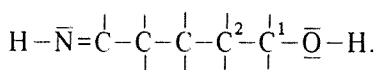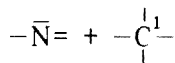
Further, let us consider the synthons:

$$S'_1(X \cup Y): \; -\bar{N}= \; + \; >C=\overset{|}{C}-\underline{\bar{O}}- \; , \qquad S'_2(X \cup Y): \; -\bar{N}= \; + \; -\overset{|}{\underset{|}{C}}-C=\bar{O} \; ,$$

$$S'_3(X \cup Y): \; -\bar{N}= \; + \; =\overset{|2}{\underset{|}{C}} \; + \; -\overset{1}{\underset{\|}{C}}-\underline{\bar{O}}- \; , \qquad S'_4(X \cup Y): \; -\bar{N}= \; + \; -\overset{|2}{\underset{|}{C}}- \; + \; -\overset{1}{\underset{\|}{C}}-\underline{\bar{O}}- \; ,$$

$$S'_5(X \cup Y): \; -\bar{N}= \; + \; =\overset{|2}{\underset{|}{C}} \; + \; >C\overset{1}{=}\underline{\bar{O}} \; , \qquad S'_6(X \cup Y): \; -\bar{N}= \; + \; -\overset{2}{C}=\underset{|}{C}- \; + \; -\underline{\bar{O}}- \; ,$$

$$S'_7(X \cup Y): \; -\bar{N}= \; + \; -\overset{|2}{\underset{|}{C}}-\overset{1}{\underset{\|}{C}}- \; + \; -\underline{\bar{O}}- \; , \qquad S'_8(X \cup Y): \; -\bar{N}= \; + \; -\overset{\ominus}{C}\underset{|}{=}\underset{|}{C}-\underline{\bar{O}}- \; ,$$

$$S'_9(X \cup Y): \; -\bar{N}= \; + \; =\overset{2}{\underset{|}{C}}-\overset{1}{\underset{|}{C}}\overset{}{=}\underline{\bar{O}} \; , \qquad S'_{10}(X \cup Y): \; -\bar{N}= \; + \; =C\overset{2}{=}\overset{1}{\underset{|}{C}}-\underline{\bar{O}}- \; ,$$

$$S'_{11}(X \cup Y): \; -\bar{N}= \; + \; -\overset{|}{\underset{|}{C}}-\overset{}{\underset{\|}{C}}-\underline{\bar{O}}- \; , \qquad S'_{12}(X \cup Y): \; =\bar{N}-\overset{|2}{\underset{|}{C}}- \; + \; \geqslant C-\underline{\bar{O}}- \; .$$

The synthons $S'_1(X \cup Y), \ldots, S'_7(X \cup Y)$ are the elements of the 1-stabilization $S'(X)$ with respect to $S(A)$. The synthon $S'_8(X \cup Y)$ does not satisfy condition (1) from definition 6. the synthon $S'_9(X \cup Y)$ does not satisfy condition (2) from definition 6, the synthon $S'_{10}(X \cup Y)$ does not satisfy condition (3) (the sum of the inside and the outside characteristics of the atom C^2 is four), the synthon $S'_{11}(X \cup Y)$ does not satisfy condition (4), and the synthon $S'_{12}(X \cup Y)$ does not satisfy conditions (5) and (6).

The algorithms SPS-GEN and STAB

Notations: $A = \{A_1, \ldots, A_n\}$, $S(A)$ is a synthon, $S(X) \subset S(A)$, Y is the set defined analogously as in definition 6.

The input structures are $S(A)$ and $S(X)$ for the algorithm SPS-GEN, and $S(A)$ and $S'(X)$ for the algorithm STAB, respectively. The output is the set N and M, respectively. The algorithm STAB is recursive.

The algorithm SPS-GEN

Step 1 (initialization). Let the set M be a set of all synthons $S'(X)$ isomeric to $S(X)$ and satisfying conditions (1)–(5) from definition 3. Let $N = \emptyset$.

Step 2. For each synthon $S'(X)$ from M, do step 3.

Step 3. STAB$(S(A), S'(X), N)$.

Step 4. The end of the algorithm.

The algorithm STAB

Step 1. If it is possible to immerse $S'(X)$ into $S(A)$ by formation $S'(A)$ then do this immersion and put $N = N \cup S'(A)$.

Step 2. Define the set Y. If $Y = \emptyset$, then go to step 6.

Step 3. Let C be a set of all synthons $S'(X \cup Y)$ which have been formed by the 1-stabilization of the synthon $S'(X)$ with respect to the synthon $S(A)$. If $C = \emptyset$, then go to step 6.

Step 4. For each synthon $S'(X \cup Y)$ from the set C do step 5.

Step 5. Let Y' be a set of all atoms $y \in Y$ on which any change has been realized during the isomerization $S(X \cup Y) \to S'(X \cup Y)$. Put $X = X \cup Y'$ and do STAB$(S(A), S'(X), N)$.

Step 6. The end of the algorithm.

Schematically, we can demonstrate the work of the algorithms SPS-GEN and STAB as follows:
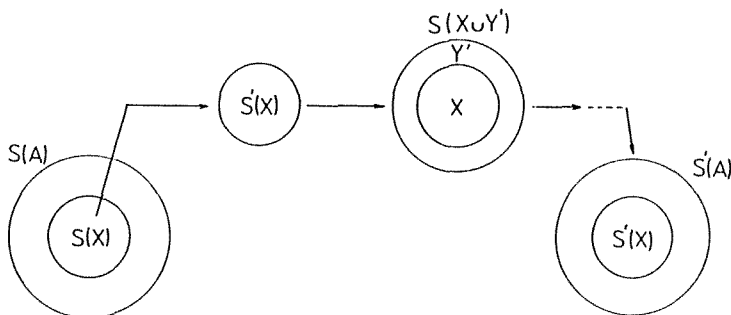
Fig. 1.

THEOREM 8

Let $A = \{A_1, \ldots, A_n\}$ be a set of atoms. Let $S(A)$ be a synthon on the set A, $S(X) \subset S(A)$. Let N be a set of the results of algorithm SPS-GEN by input $S(A)$ and $S(X)$. Then $N = \mathscr{S}(S(A/X))$.

Proof

Let the sets I, J be defined as follows: $J = \{1, 2, \ldots, n\}$, $I = \{i \mid A_i \in X\}$.

(A) Let $S'(A) \in N \Rightarrow S'(X)$ satisfy conditions (1)–(5) of definition 3 and at the same time $S'(X) \subset S'(A)$. $S'(A)$ is stable because it has been formed by the stabilization of the synthon $S'(X)$. Accordingly, $S'(A)$ is an element of the stable X-neighbourhood of the synthon $S(A)$. We show that conditions (1)–(4) from definition 4 are satisfied. Conditions (1) and (2) follow from conditions (2) and (3) of definition 6 and from the mechanism of algorithm STAB. Now we verify condition (3). We distinguish the following cases:

(a) if $Y = \emptyset$, then it is possible to immerse $S'(X)$ into $S(A)$ (since $S'(A) \in N$) and the condition is satisfied from the definition of the immersion;

(b) if $Y \neq \emptyset$, then we have three possibilities:

(b1) it is possible to immerse $S'(X)$ into $S(A)$ and the condition is satisfied from the definition of the immersion;

(b2) if $A = X \cup Y$, then the condition is satisfied from conditions (5) and (6) of definition 6;

(b3) if $A \supset X \cup Y$, then the condition is satisfied (cf. (b1)) in the synthon $S'(X \cup Y')$ (Y' is the set defined in algorithm STAB) and by its further stabilization, we obtain one of the cases (a), (b1) (b2).

In any case, condition (3) is satisfied. Now we verify condition (4). Let us consider the set Y' which is defined in algorithm STAB during the lth immersion. Then $Y' = Y_l - Y_{l-1}$, where $l = 1, 2, \ldots, k - 1$, k is the maximal level of the immersion of algorithm STAB by the generation of $S'(A)$, and the sets Y_l are defined by condition (4) of definition 4. The equality is satisfied with respect to the definition of the set Y in definition 6, and with respect to the validity of condition (3) of definition 4. From condition (7) of definition 6 and from condition (5) of algorithm STAB, we have: $RD(S(X'' \cup \{y\}), S'(X'' \cup \{y\})) - RD(S(X''), S'(X'')) \leqslant 1$ for each $y \in Y$ where the set X'' answers the function of the set X during the lth immersion of algorithm STAB. Since $Y' \subset Y$, this inequality is satisfied also for each $y \in Y'$. There follows from the mechanism of algorithm STAB that

$$X'' = X \overset{1}{\underset{i=2}{\cup}} (Y_{i-1} - Y_{i-2}) = X \cup Y_1 \ .$$

After the introduction of X'' into the inequality mentioned above, we have the assertion. Accordingly, $S'(A) \in \mathcal{G}(S(A/X))$.

(B) Let $S'(A) \in \mathcal{G}(S(A/X)) \Rightarrow S'(A)$ be an element of the stable X-neighbourhood of the synthon $S(A)$ generated by the synthon $S'(X)$, and let it satisfy

conditions (1)–(4) of definition 4. Then there exist two possibilities:

(i) $S'(A)$ is formed by the immersion of $S'(X)$ into $S(A)$ and, accordingly, it is generated in step 1 of algorithm STAB. Then $S'(A) \in N$.

(ii) $S'(A)$ is not formed by the immersion of $S'(X)$ into $S(A) \Rightarrow$ there exists a set Z, $A \supset Z \supset X$ such that $S'(A)$ is formed by the immersion of $S'(Z)$ into $S(A)$. Let us consider minimal Z with this property. Then there exist two possibilities, $Z = X \cup Y'$ or $Z \supset X \cup Y'$ (from the definition of Y' in step 5 of algorithm STAB there cannot be $Z \subset X \cup Y'$).

(ii1) Let $Z = X \cup Y'$. We show that $S'(Z)$ satisfies the conditions of the stabilization (definition 6). Conditions (1), (2), (3), (7) follow directly from definition 4. We show conditions (4), (5), (6). First we prove condition (4) by contradiction. Suppose that there does not exist $k \in K$ such that $C_{A_k}^I(S(Y' \cup \{A_k\})$, $S'(Y' \cup \{A_k\})) > 0$, then for each $k \in K$: $C_{A_k}^I(S(Y' \cup \{A_k\})$, $S'(Y' \cup \{A_k\})) = 0$ and $S'(A)$ is formed by the immersion of $S'(X)$ into $S(A)$, which is case (i), a contradiction. Consequently, condition (4) is satisfied. Now we show (5). Suppose that $i \in I - K$ and $j \in E$ such that $p_{ij} \neq 0$. It follows from the definition of the sets I, K that there cannot exist $j \in J$ (and $j \in E$) such that $p_{ij} \neq 0$. If $C_{A_i}^E(S(X), S'(X)) = 1$, then it means that one new bond has been formed, which is in contradiction with the requirement of condition (3) of definition 4. The $C_{A_i}^E(S(X), S'(X)) > 1$ during 1-stabilization is impossible because of condition (5) of definition 3. During the following stabilizations, condition (1) of definition 4 must be satisfied, so $C_{A_i}^E(S(X), S'(X)) \leqslant 1$ must hold. The same follows from condition (3). Accordingly, condition (5) is satisfied. Condition (6) follows directly from condition (3) of definition 3. Accordingly, $S'(A)$ is formed by the immersion of $S'(Z)$ into $S(A)$ and $S'(Z)$ is formed by the stabilization of the $S'(X)$ with respect to $S(A)$, and $S'(A) \in N$.

(ii2) Let $Z \supset Z' = X \cup Y_1'$, where $Y_1' = Y'$ and Y' is defined in the first way through algorithm STAB. We show that $S'(X \cup Y')$ is formed by the 1-stabilization of the synthon $S'(X)$ with respect to the synthon $S(A)$, i.e. we show that conditions (1)–(7) from definition 6 are satisfied. The first condition is satisfied because each subsynthon of the stable synthon is stable. Condition (2) follows automatically from definition 4 (condition (2)). Condition (3) follows from condition (2) of definition 3, and further from condition (1) of definition 4 and from the construction of the set Y' in algorithm STAB. Condition (4) we can show analogically, as in case (ii1). In the same way, we can show condition (5). Condition (6) follows from condition (3) of definition 4, and condition (7) follows from condition (4) in definition 4 (because there is $Y' = Y$). Accordingly, the synthon $S'(X \cup Y')$ is formed by the 1-stabilization of the synthon $S'(X)$ with respect to $S(A)$, and $S'(X \cup Y')$ is the product of the first way through algorithm STAB. Further, let Y_l' ($l = 1, 2, \ldots, k - 1$) denote the set Y' defined in the lth way through algorithm STAB. It follows from condition (1) of definition 4 that there exists l such that

$$Z = X \cup \bigcup_{i=1}^{l} Y_i' \ .$$

With the help of the same consideration as that used above, we show that the synthon

$$S'((X \cup \bigcup_{i=1}^{l-1} Y_i') \cup Y_l')$$

is formed by the 1-stabilization of the synthon

$$S'(X \cup \bigcup_{i=1}^{l-1} Y_i')$$

with respect to the synthon $S(A)$. From this, we have that $S'(Z)$ has been formed by the stabilization of $S'(X)$ with respect to $S(A)$, and because $S'(A)$ is formed by the immersion of $S'(Z)$ into $S(A)$, it holds that $S'(A) \in N$ and the theorem is proved.

3. Construction of the reduced set of SPS

In practical applications, it is very important to determine not only the atoms of the reaction center but also the atoms which we want to exclude from the chemical process (e.g. the skeleton creating atoms). The conditions of the conservation of the skeleton or its part or of a functional group is very important in, for example, the synthesis design of biologically active compounds. Therefore, we introduce the notion of the set of all SPSs of the synthon $S(A)$ with respect to the synthon $S(X)$ reduced by the synthon $S(\bar{X})$.

DEFINITION 7

Let $A = \{A_1, \ldots, A_n\}$. Let $S(X)$ and $S(\bar{X})$ be its subsynthons, respectively. Let $X \cap \bar{X} = \emptyset$. Let $S'(A) \in \mathcal{G}(S(A/X)$. Let $M = (m_{ij})$ and $M' = (m_{ij}')$ be the S-matrices of the synthons $S(A)$ and $S'(A)$. Let $P = (p_{ij}) = M' - M$. Let $I = \{i \mid A_i \in \bar{X}$ or there exists j such that $A_j \in \bar{X}$ and $m_{ij} \neq 0\}$, $I' = \{i \mid A_i \in \bar{X}\}$. We call the set

$$\mathcal{G}(S(A/X/\bar{X}) = \{S'(A) \in \mathcal{G}(S(A/X)) \mid p_{ij} = 0 \text{ for each } i \in I', \ j \in I\}$$

the set of all SPSs of the synthon $S(A)$ with respect to the reaction center $S(X)$ reduced by the subsynthon $S(\bar{X})$ (denoted as $\mathcal{G}(S(A/X/\bar{X}))$).

EXAMPLE 6

Let us consider the same synthon $S(A)$ and the same set X as in example 2. Let $\bar{X} = \{C^2, C^3, C^4\}$. Then only the synthon $S_1'(A)$ is an element of $\mathcal{I}(S(A/X/\bar{X}))$.

THEOREM 9

Let $S(A)$ be a synthon and $S(X)$ its subsynthon. Let $A \supset Z \supset X$. Let $S'(A) \in \mathcal{I}(S(A/X))$, $I = \{i \mid A_i \in Z\}$. Let $P = (p_{ij})$ be the SR-matrix of the isomerization $S(A) \to S'(A)$. Suppose that the isomerization $S(Z) \to S'(Z)$ satisfies the condition $(\star)$. Then $S'(Z) \in \mathcal{I}(S(Z/X))$.

Proof

$S'(Z)$ is an element of the stable X-neighbourhood of the synthon $S(Z)$ because $S'(A)$ is an element of the stable X-neighbourhood of the synthon $S(A)$. Therefore, the synthons $S(X)$ and $S'(X)$ satisfy conditions $(1)-(5)$ from definition 3. We show conditions $(1)-(4)$ of definition 4 for $S'(Z)$. Condition (1) is formulated in the assumptions. Condition (2) follows from the fact that $S'(Z) \subset S'(A)$, $Z \subset A$, and condition (2) is satisfied for each $x \in A$ and, consequently, for each $x \in Z$. We can show condition (3) analogically as for condition (2). Condition (4) follows from the satisfaction of the condition $(\star)$ for the change $S(Z) \to S'(Z)$ and from the satisfaction of condition (4) for the isomerization of the whole synthon, and the theorem is proved.

THEOREM 10

Let $S(X)$ and $S(\bar{X})$ be two subsynthons of the synthon $S(A)$ such that $X \cap \bar{X} = \emptyset$. If $S''(A) \in \mathcal{I}(S(A/X/\bar{X}))$ then there exists $S'(A - \bar{X}) \in \mathcal{I}(S(A - \bar{X}/X))$ such that $S''(A)$ is formed by the immersion of $S'(A - \bar{X})$ into $S(A)$.

Proof

If $S''(A) \in \mathcal{I}(S(A/X/\bar{X}))$ then $S''(A) \in \mathcal{I}(S(A/X))$ and $p_{ij} = 0$ for each $i, j \in I$ (the notation is the same as in definition 5). Let us put $Z = A - \bar{X}$. Since $A = A - \bar{X} \cup \bar{X} = Z \cup \bar{X}$, the atoms of the set $\bar{X}$ are without a change, and the isomerization $S(A) \to S'(A)$ satisfies the condition $(\star)$, the isomerization $S(Z) \to S'(Z)$ must also satisfy the condition $(\star)$. Then from theorem 4 we have $S''(Z) \in \mathcal{I}(S(Z/X)) = \mathcal{I}(S(A - \bar{X}/X))$. Since $p_{ij} = 0$ for each $i, j \in I$, and since $S''(A) \in \mathcal{I}(S(A/X))$ and, therefore, condition (3) from definition 4 is satisfied, it is possible to immerse the synthon $S''(Z)$ into $S(A)$ by the formation of $S''(A)$. Accordingly, we can put $S'(A - \bar{X}) = S''(Z) = S''(A - \bar{X})$ and the theorem is proved.

The assertion of theorem 10 is very important for the convergent generation of $\mathcal{I}(S(A/X/\bar{X}))$ because it follows from this that we need not generate the whole $\mathcal{I}(S(A/X))$ but only $\mathcal{I}(S(A - \bar{X}/X))$.

Now, let us formulate algorithm SPS-GEN-1. It will generate the SPS of the synthon $S(A)$ with respect to the synthon $S(X)$ reduced by the synthon $S(\bar{X})$. The input structures are $S(A)$, $S(X)$ and $S(\bar{X})$, where $A \supset X$, $A \supset \bar{X}$, $X \cap \bar{X} = \emptyset$. The output is the set of results $\Re$.

The algorithm SPS-GEN-1

Step 1. Initialization: $\Re = 0$.

Step 2. Do algorithm SPS-GEN for the synthons $S(A - \bar{X})$, $S(X)$. Let the results of algorithm SPS-GEN (i.e. the synthons $S'(A - \bar{X})$) be saved into the set N.

Step 3. For each synthon $S'(A - \bar{X})$ from the set N, do step 4.

Step 4. If it is possible to immerse the synthon $S'(A - \bar{X})$ into $S(A)$, then do this immersion (the synthon $S'(A)$ is formed) and put $\Re = \Re \cup S'(A)$.

Step 5. The end of the algorithm.

THEOREM 11

Let $A = \{A_1, \ldots, A_n\}$ be a set of atoms, and $S(X)$ and $S(\bar{X})$ be two subsynthons of $S(A)$ such that $X \cap \bar{X} = \emptyset$. Let $\Re$ denote the set of the results of algorithm SPS-GEN-1 by input $S(A)$, $S(X)$, $S(\bar{X})$. Then:

$$\Re = \mathcal{S}(S(A/X/\bar{X})).$$

Proof

(a) Let $S'(A) \in \mathcal{S}(S(A/X/\bar{X}))$. We have, from theorem 10, that there exists the synthon $S'(A - \bar{X}) \in \mathcal{S}(S(A - \bar{X}/X))$ such that $S'(A)$ is formed by the immersion of $S'(A - \bar{X})$ into $S(A)$, i.e. in step 4 of algorithm SPS-GEN-1. From theorem 9, we have that the synthon $S'(A - \bar{X})$ is formed by algorithm SPS-GEN (by the input $S(A - \bar{X})$ and $S(X)$), i.e. in step 2 of algorithm SPS-GEN-1. Accordingly, $S'(A) \in \Re$.

(b) If $S'(A) \in \Re$, then $S'(A)$ is formed by the immersion of the synthon $S'(A - \bar{X})$, generated by algorithm SPS-GEN (by input $S(A - \bar{X})$, $S(X)$), into $S(A)$ and, consequently, $S'(A) \in \mathcal{S}(S(A/X))$ (because $(A - \bar{X}) \subset A$). If $P = (p_{ij})$ is the SR-matrix of the isomerization $S(A) \rightarrow S'(A)$, and I and I' are the sets defined analogically as in definition 5, then it follows from the definition of the immersion that $p_{ij} = 0$ for each $i \in I'$, $j \in I$. Consequently, $S'(A) \in \mathcal{S}(S(A/X/\bar{X}))$ and the assertion is proved.

It follows from theorem 11 that algorithm SPS-GEN-1 may be used for the generation of the whole set $\mathcal{S}(S(A/X/\bar{X}))$.

The following theorem introduces an important property of the set $\mathcal{S}(S(A/X/\bar{X}))$.

THEOREM 12

Let $S(X)$ and $S(\bar{X})$ be two subsynthons of the synthon $S(A)$ such that $X \cap \bar{X} = \emptyset, \bar{X}' \subset \bar{X}$. Then:

(1) $\mathcal{G}(S(A/X)) \supset \mathcal{G}(S(A/X/\bar{X}))$.

(2) $\mathcal{G}(S(A/X/\bar{X}')) \supset \mathcal{G}(S(A/X/\bar{X}))$.

Proof

(1) If $S'(A) \in \mathcal{G}(S(A/X/\bar{X}))$, then from definition 7 we have that $S'(A) \in \mathcal{G}(S(A/X))$.

(2) Let $S'(A) \in \mathcal{G}(S(A/X/\bar{X}))$. Let M and M' be the S-matrices of the synthons $S(A)$ and $S'(A)$, respectively. Let $P = (p_{ij}) = M' - M$. Let I, J, I', J' be sets defined as follows:

$I = \{i \mid A_i \in \bar{X}$ or there exists j such that $A_j \in \bar{X}$ and $m_{ij} \neq 0\}$,
$J = \{i \mid A_i \in \bar{X}'$ or there exists j such that $A_j \in \bar{X}'$ and $m_{ij} \neq 0\}$,
$I' = \{i \mid A_i \in \bar{X}\}, J' = \{i \mid A_i \in \bar{X}'\}$. Evidently, $J \subset I, J' \subset I'$.

From the definition, we have $p_{ij} = 0$ for each $i \in I', j \in I$ and, therefore, $p_{ij} = 0$ for each $i \in J', j \in J$.

4.   An order of SPS

We introduce, to improve the regulation of changes which are accompanied by a splitting of the skeleton, the notion of the order of SPS.

DEFINITION 9

Let $S(A)$ be a synthon, $S(X) \subset S(A), S'(A) \in \mathcal{G}(S(A/X))$. We define the order N of the synthon $S'(A)$ with respect to $S(A)$ as follows:
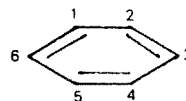
$$N = ((k + 1) \div 4) + 1,$$

where

$$k = \sum_{y \in A - X} RD(S(\{y\}), S'(\{y\}))$$

and the operation $\div$ denotes the so-called integer division.

The order of SPS is a measure of changes of valence states of atoms out of the set X.

EXAMPLE 11

Let us consider the Kekulé structure of benzene

Here. $A = \{C^1, \ldots, C^6\}$. Let $X = \{C^1, C^2\}$.

(a) An SPS of the order 1 is, for example, $^6C \equiv C^1- + {}^2C \dot{=} C^3 {}^{-4}C \dot{=} C^5-$.

(b) An SPS of the order 2 is, for example, $^6C \equiv C^1- + {}^2C \dot{\equiv} C {}^{4}C \dot{=} C^5$.

(c) An SPS of the order 3 is, for example, $^6C \equiv C^1- + {}^2C \equiv C^3- + {}^4C \equiv C^5-$.

Intuitively, the order of the SPS is connected with the length of the reaction path. However, this connection is not always linearly proportional. If we study, for example, the tele-elimination reaction $S(A) \rightarrow S'(A)$:

$$-\overset{|}{C}-C=C-C\overset{*}{\equiv}\overset{*}{C}-C=C-C= \rightarrow -C=C-C=C\overset{*}{=}\overset{*}{C}-C=C= ,$$
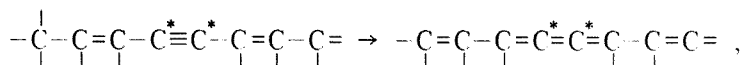
and we denote $X = \{C^*, C^*\}$, then the order of $S'(A)$ with respect to $S(A)$ is two and $RD(S(A), S'(A)) = 9$. However, in example 11(c), the order of the SPS is equal to three and $RD(S(A), S'(A)) = 6$. The order is added fast if the skeleton is split, and it is added slowly if, for example, an electrocyclic reaction takes place. Upper bounds of the order of SPS are formulated by the following theorem.

THEOREM 13

Let $S(A)$ be a synthon, $S(X) \subset S(A)$. Let N_{max} denote the maximal order of $S'(A) \in \mathscr{S}(S(A/X))$ with respect to $S(A)$. Then:

$$N_{max} \leqslant ((2(\text{card } A - \text{card } X) + 1) \div 4) + 1.$$

Proof

This follows immediately from the definition of the order of SPS, from condition (2) of definition 4, and from the corollary of theorem 4.

5. Conclusions

The purpose of this and previous [2] communications is to form a mathematical model of chemical reality which is potentially applicable as a basis for computer-assisted organic synthesis design. The principal notions and concepts defined in the model correspond to the actions or thoughts of a synthesizing chemist. There-

fore, the model is defined for synthons and not only for whole molecules, and the chemical distance [3,28] is substituted by a much more appropriate reaction distance. The idea of the virtual atoms considerably increases the deductive power of the model. The ability of the model to find ESRE enables us to use the theory for an exhaustive construction of reaction mechanisms.

The ability of the model to serve as a basis for the formation of computer programs for organic synthesis design was verified during the implementation of the program for the generation of the set of all SPSs [29].

Acknowledgements

References

[1]   E. Fontain, J. Bauer and I. Ugi, Chem. Lett. (1987)37.
[2]   J. Koča, Part I, J. Math. Chem., this issue.
[3]   J. Dugundji and I. Ugi, Top. Curr. Chem. 39(1973)19.
[4]   E.J. Corey and A.K. Long, J. Organ. Chem. 43(1978)2208;
        E.J. Corey, A.K. Long, J. Mulzer, H.W. Orf, A.P. Johnson and A.P.W. Hewett, J. Chem. Inf. Comp. Sci. 20(1980)221;
        A.K. Long, S.D. Rubenstien and L.J. Joncas, Chem. Eng. News 61(1983)22;
        E.J. Corey, A.K. Long and S.D. Rubenstein, Science 228(1985)408.
[5]   W.T. Wipke and T.M. Dyott, J. Amer. Chem. Soc. 96(1974)4825;
        P. Gund, E.J.J. Grabowski, D.R. Hoff, G.M. Smith, J.D. Andose, J.B. Rhodes and W.T. Wipke, J. Chem. Inf. Comp. Sci. 20(1980)88;
        R.E. Carter and W.T. Wipke, Kem. Tidskr. 93(1981)20.
[6]   E.J. Corey and W.T. Wipke, Science 3902(1969)178.
[7]   H. Gelernter, N.S. Sridharan, A.J. Hart, S.C. Yen, F.W. Fowler and H.J. Shue, Top. Curr. Chem. 41(1973)113;
        H.L. Gelernter, S.S. Bhagwat, D.L. Larsen and G.A. Miller, Anal. Chem. Symp. Ser. 15 (1983)35.
[8]   F. Choplin, C. Laurenco, R. Marc, G. Kaufmann and W.T. Wipke, Nouv. J. Chim. 2(1978) 285;
        C. Laurenco, L. Villien and G. Kaufmann, Tetrahedron 40(1984)2731.
[9]   M. Bersohn, J. Chem. Soc. Japan 45(1972)1897;
        M. Bersohn, A. Esack and J. Luchini, Comput. Chem. 2(1978)105;
        M. Bersohn, ACS Symp. Ser. 112(1979)341.
[10] G. Moreau, Nouv. J. Chim. 2(1978)187.

[11] R. Barone, M. Chanon and J. Metzger, Tetrahedron Lett. 32(1974)2761;
R. Barone, M. Chanon and J. Metzger, Chimia 32(1978)216;
R. Barone, M. Chanon, P. Cadiot and J.M. Cense, Bull. Soc. Chim. Belg. 91(1982)333;
R. Barone, M. Chanon and M.L. Contreras, Nouv. J. Chim. 8(1984)311.

[12] W. Schubert, MATCH 6(1979)213.

[13] J. Gasteiger and C. Jochum, Top. Curr. Chem. 74(1978)93;
J. Gasteiger, M.G. Hutchings, B. Christoph, L. Gann, C. Hiller, P. Low, M. Marsili, H. Saller and K. Yuki, Top. Curr. Chem. 137(1987)19.

[14] R. Doenges, B.T. Groebel, H. Nickelsen and J. Sander, J. Chem. Inf. Comp. Sci. 25(1985) 425.

[15] J. Bauer, R. Herges, E. Fontain and I. Ugi, Chimia 39(1985)43.

[16] N.S. Zefirov and S.S. Tratch, Zh. Org. Khim. 11(1975)225, 1785; ibid. 12(1976)7, 697; ibid. 18(1982)1561; ibid. 20(1984)1121;
N.S. Zefirov, Acc. Chem. Res. 20(1987)237;
S.S. Tratch and N.S. Zefirov, in: *Principles of Symmetry and Systemology in Chemistry*, ed. N.F. Stepanov (Moscow University Publishers, Moscow 1987) p. 54.

[17] A. Weise, Z. Chem. 17(1977)100.

[18] A. Weise, Z. Chem. 13(1973)155; ibid. 15(1975)333;
G. Westphal, A. Klebsch, A. Weise, U. Sternberg and A. Otto, Z. Chem. 17(1977)295;
A. Weise and H.G. Scharnow, Z. Chem. 19(1979)49;
A. Weise, J. Prakt. Chem. 322(1980)761;
A. Weise, G. Westphal and H. Rabe, Z. Chem. 21(1981)218.

[19] J.B. Hendrickson, Acc. Chem. Res. 19(1986)274.

[20] J.B. Hendrickson, Top. Curr. Chem. 62(1976)49;
J.B. Hendrickson, J. Amer. Chem. Soc. 99(1977)5439;
J.B. Hendrickson, J. Chem. Inf. Comp. Sci. 19(1979)129;
J.B. Hendrickson, D.L. Grier and A.G. Toczko, J. Amer. Chem. Soc. 107(1985)5228.

[21] C.E. Peishoff and W.L. Jorgensen, J. Organ. Chem. 50(1985)1056, 3174;
M.G. Bures, B.L. Roos-Kozel and W.L. Jorgensen, J. Organ. Chem. 50(1985)4490;
P. Metivier, A.J. Gushurst and W.L. Jorgensen, J. Organ. Chem. 52(1987)3724.

[22] Cf. for explanation: Z. Hippe, Anal. Chim. Acta 133(1981)677;
Z. Hippe, Stud. Phys. Theor. Chem. 16(1981)249;
Z. Hippe, Przem. Chem. 64(1985)285.

[23] J. Koča, M. Kratochvil, M. Kunz and V. Kvasnička, Coll. Czech. Chem. Commun. 49(1984) 1247.

[24] R. Lindsay, B.G. Buchanan, E.A. Feigenbaum and J. Lederberg, *Applications of Artificial Intelligence for Organic Chemistry* (McGraw-Hill, New York, 1980).

[25] R.E. Carhart, D.H. Smith, N.A.B. Gray, J.G. Nourse and C. Djerassi, J. Organ. Chem. 46 (1981)1708.

[26] M.E. Munk, R.J. Lind and M.E. Clay, Anal. Chim. Acta 184(1986)1.

[27] M.E. Munk, M. Farkas, A.H. Lipkis and B.D. Christie, Microchim. Acta II(1986)199.

[28] I. Ugi, J. Bauer, J. Brandt, J. Friedrich, J. Gasteiger, C. Jochum and W. Schubert, Angew. Chem. Int. Ed. Engl. 18(1979)111.

[29] L. Matyska and J. Koča, J. Chem. Inf. Comp. Sci., in preparation.