match no. 27 pp. 191-207 1992

Generation of stereoisomers and their spatial models corresponding to the given molecular structure

L.A. ZLATINA, M.E. ELYASHBERG[1]

All-Russian Research Institute of Organic Synthesis,

Radio str., 12, 107005, Moscow

(received: March 1991)

Abstract

Programs for exhaustive and irredundant generation of stereoisomers, corresponding to the adjacency matrix of a molecular structure have been elaborated. The display of stereoisomers is performed by usage of generally accepted symbols for stereobonds. An algorithm, designed for calculating the Cartesian coordinates of atoms within the molecule on the basis of known internal coordinates and given symmetry, is proposed. The spatial models of a stereoisomer family are automatically generated basing upon the Cartesian coordinates of the initial structure. The programs are written in FORTRAN for IBM PC AT. They are designed for the solution of the molecular modelling problem, for calculation of the molecular spectra in an expert systems, aiming at the elucidation of the molecular structure by their spectra, etc.

1 Introduction

Both physical and chemical properties of chemical substances are known finally to depend upon the spatial disposition of the atoms composing them and linked by the chemical bonds. Each configuration determines certain energy and charge

[1]Financial support from the BIMF (Bayreuther Inst. f. Makromolekülforschung) is gratefully acknowledged

densities on atoms and bonds, which in turn determine force and electrooptical parameters, dipole moment, reactivity of a molecule, etc. Therefore, the elaboration of structural isomer generation algorithms becomes one of the principal tasks of mathematical chemistry, as well as exhaustive and irredundant enumeration of all the theoretically possible stereoisomers, answering some structural formula, and the creation of methods of the stereoisomer spatial model generation.

These investigational trends are closely connected with the development of quantum chemistry, molecular mechanics, molecular spectroscopy, the design of artificial intelligence systems aiming at the search for new ways of synthesis of chemical compounds with prescribed properties, molecular structure elucidation based on spectra, prediction of the biological activity based upon the "structure-property" correlations, etc.

During the recent years our laboratory has been engaged in the development of the RASTR expert system for molecular structure elucidation, based on spectra [1-4]. The elucidation of the most probable structures within the limits of this system is carried out on the basis of the isomer vibrational spectra prediction, which in turn requires the complete list of all possible stereoisomers and data on the Cartesian coordinates of the atoms. In the case when a molecule possesses a symmetry, calculation of its vibrational spectrum is significantly simplified [5]. Therefore, the coordinates of the stereoisomer atoms in the process of constructing its spatial model for the consequent calculation of its vibrational spectrum should be calculated taking into account the conservation of its symmetry.

All of the previously available algorithms of calculating Cartesian coordinates in space using geometrical parameters (bond lengths, valence and dihedral angles) [6-15] either did not take into account the symmetry at all, or they demanded for the active interference of the user for its consideration. The Go-Sheraga algorithm [16] takes into consideration only one of the possible elements of symmetry by solution of nearly a hundred of transcendental equations. We have recently elaborated the original algorithm (algorithm of symmetrization), which uses the coordinates of atoms calculated by the Crippen method [15] as the initial approximation, and then builds a model with the required symmetry. Crippen's method [15] does not evidently consider the molecular symmetry, nevertheless our experience has shown that in many cases the spatial models of the complex polycyclic structures, calculated according to this method possessed the required symmetry. However, there is no guarantee for success. For example, a calculation of coordinates of the trans-decalin atoms by Crippen's method resulted in a spatial model with distorted symmetry (C_2 instead of C_{2h}) and geometric parameters being distinctly unlike the experimental ones. The calculation of the atomic coordinates of the structure, shown on Fig. 1, has produced a model, characterised by the required geometrical parameters, but lacking symmetry. An application of our algorithm of symmetrization made it possible to restore the required symmetry (C_{2h} and C_3 correspondingly) in both cases, retaining the correct geometric parameters.

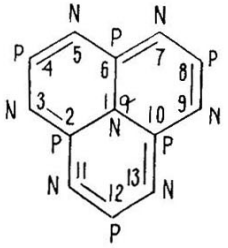


Fig1. "Skeleton" of hexaphosphorheptamin

However, the atomic coordinates, calculated by means of this algorithm, generally do not cover the total variety of spatial models corresponding to the given molecular structure. A molecule possessing n stereocentres may have up to $N = 2^n$ stereoisomers, differing in the spatial orientation of the atoms. Various stereoisomers of a molecule may have different physico-chemical features. Therefore, the detailed identification of a molecule by its spectral and other physical and chemical characteristics demands for the list of all theoretically possible non-coincident stereoisomers, and of their spatial models. The prediction of the spectra within the stereoisomer family is possible if the coordinates of the atoms are calculated. We elaborated both an algorithm and a program for the exhaustive and irredundant generation of stereoisomers. The approach used is based upon the ideas of Nourse [17]. This paper presents the algorithm of symmetrization for the construction of spatial models as well as of methods for the generation and the display of stereoisomers. All the algorithms are realized as Fortran programs for the personal computer, compatible with IBM PC AT.

2 The algorithm of symmetrization

The initial point of the symmetrization algorithm work is the calculation of spatial coordinates of the atoms in a molecule on the basis of the internal coordinates, using the method [15]. According to [15], the matrix of the lower and upper bounds of distances between all the atom pairs in the space is produced as an initial step:

$$LU = [L\backslash U],\ l_{ij} \le d_{ij} \le u_{ij},\ \forall\, i,j \qquad (1)$$

This step is followed by the $G(LU)$ metric matrix formation according to Crippen's principle and then its eigenvalues and eigenvectors are calculated. The latter are used in the calculation of the initial Cartesian atomic coordinates' approximation $x_{ik}, i = 1,\ldots,n$, where n is the atom number and $k = 1,2,3$. As this approximation does not commonly fit (1), the further minimization of the function F is required:

$$F = \sum_{j>i=1}^{n-1} \begin{cases} (d_{ij}^2 - u_{ij}^2)^2, & d_{ij} > u_{ij} \\ 0, & l_{ij} \le d_{ij} \le u_{ij} \\ (l_{ij}^2 - d_{ij}^2)^2, & d_{ij} < l_{ij} \end{cases} \qquad (2)$$

where $d_{ij}^2 = \sum_{k=1}^{3} (x_{ik} - x_{jk})^2$. This is the fast step of Crippen's method. The algorithm of symmetrization requires the specification of the point group of the molecular symmetry to be given by researcher in the symbolic form (D_{3h}, C_4, etc.). Constructed according [18], the graph automorphism group (GAG) partitions the set of atoms into the orbits, the number of atoms within an orbit presenting its index. Then, the equation of the presumable principal symmetry axis is derived out of the corresponding GAG permutation by means of regression problem solving, and the Crippen model is oriented in the space by geometric transformations so that to meet the following conditions:

a) the presumable principal axis coincides with the OZ axis;

b) one of the σ_v planes coincides with the YOZ plane;

c) one of the C_2 axes coincides with the OX axis;

d) the fixed point, in case of being the single one, coincides with the coordinates' origin.

The next step of our algorithm results in the formation of the set A of atoms, taking one (as "representative") from each orbit according to a certain rule. Let N be the order of point group, and L the index of the orbit R containing the selected atom - "representative" $a_i \in R$. Then in case of $N = L$ the coordinates of $a_i \in R$ taken from the Crippen model, would be preserved. If $L < N$, the atom $a_i \in R$ is projected onto a symmetry element of the point group. The choice for such an element depends upon several factors: relation between L and N ($L = N/2$, $L = N/4$, $L = 2 \neq N/2$ etc.), the type of the point group, the location of the atoms, previously placed in the space and in some cases on the value of the third coordinate of a_i in the Crippen model. Coordinates of the other atoms of the orbit are found by the application of generator combinations to the "representative" of this orbit. The correlation between the orbits and symmetry elements are illustrated by the following example:

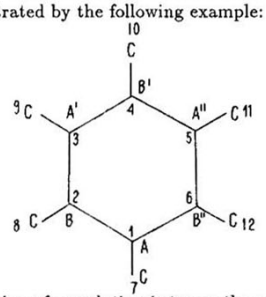


Fig. 2. Illustration of correlation between the orbits and symmetry elements

Here the symmetry group is C_{3v}, $N = 6$, and $L(A) = L(B) = 3 = N/2$

$$\begin{array}{ll} \text{orbit 1}: & (A, A', A''), L(A) = 3, A \in \sigma_{v_1}, \sigma_{v_1} = YOZ \\ \text{orbit 2}: & (B, B', B''), L(B) = 3, B \in \sigma_{v_2}, \sigma_{v_2} \perp [AA'] \end{array}$$

The spatial model (let us denote it by $M(X^0_{ik})$) obtained by means of this technique, possesses the required symmetry, but highly different geometrical parameters if they are compared to the real ones. It is necessary to optimize the coordinates of all the "representatives" in order to restore the given geometrical parameters, keeping the required symmetry. So we arrive at the optimization problem:

$$min\ F'(x),\ x \in R^m,\ p^- \leq x \leq p^+,$$

where $F' = F((LU)') + \sum_i \varphi_i$, F is similar to (2) in the Crippen method, but with new parameters (elements of a new matrix $(LU)'$ of the lower and upper bounds of distances) and additional terms φ_i, which serve for preserving the required symmetry; x means a vector of the coordinates of the point set E, where $E = A \cup B \cup C$ (let $\tilde{A} = A \cup B$) where A is a set of "representatives", B a set of certain atoms, added to the set of atoms A in order to fix all given geometrical parameters, C is a set of additional points; p^- and p^+ are the vectors of lower and upper bounds of variables' variation accordingly. We shall describe the rules for construction of C, the formation of the terms φ_i, the calculation of the matrix $(LU)'$, the bounds of the variables' variation and the initial values for the variables. The set C of auxiliary points, situated at the symmetry elements, is added to the set $\tilde{A}$. Let the atom a' be an image of the atom $a \in A$ relatively the symmetry element α : $a \xrightarrow{\alpha} a'$, the distance between the image and the prototype being fixed, i.e. $l_{aa'} = u_{aa'}$. Then the C set includes the point $c \in \alpha$, the distance between c and a and between c and $a_i \in \tilde{A}$ atoms, participating in the formation of the valence angles $\angle a_i a a'$, being calculated. The method of these distances' calculating depends upon the kind of symmetry element α. If α possesses the order $k = 2$ the distance between a and c equals $d_{ac} = 1/2\ d_{aa'}$, and the distance between a_i and c equals $d_{a_i c} = (d^2_{a_i a} + d^2_{ac} - 2d_{a_i a} \cdot d_{ac} \cdot cos\angle a_i a a')^{1/2}$. The Fig. 3 illustrates the case, when $\alpha = C_2, i = 1$:

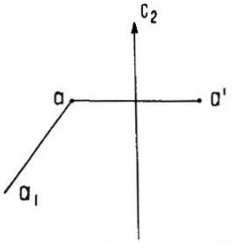


Fig. 3. Distances between atoms and auxiliary point (when $k = 2$)

If α is a proper axis, possessing the order $k > 2$ (let $a \xrightarrow{\alpha^2} a''$), the distance d_{ac} and d_{a_1c} is calculated as follows (Fig. 4):

$$\begin{aligned} d_{ac} &= d_{aa'}/(2sin\ (\pi/k)) \\ cos\ \beta &= (d^2_{a_1a'} + d^2_{a'a''} - d^2_{a_1a''})/(2d_{a_1a'}d_{a'a''}) \\ |a_1f|^2 &= d^2_{a_1a'} + |a'f|^2 - 2d_{a_1a'}|a'f|\ cos\ \beta \end{aligned}$$

where f is the middle of $[a'a'']$

$$\begin{aligned} |af| &= d_{aa'}cos[\pi(k-2)/2k] \\ cos\ \gamma &= (d^2_{a_1a} + |af|^2 - |a_1f|^2)/(2d_{a_1a}|af|) \\ d_{a_1c} &= (d^2_{a_1a} + d^2_{ac} - 2d_{a_1a}d_{ac}cos\ \gamma)^{1/2} \end{aligned}$$

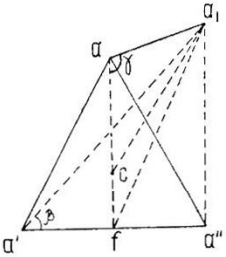


Fig. 4. Distances when $k > 2$ and a symmetry element is proper axis

If α is an improper axis of the order $k > 2$, the calculation of d_{ac} and d_{a_1c} distances demands for more complicated formulas, and $l_{aa''} = u_{aa''}$, where $a \xrightarrow{\alpha^2} a''$, becomes the necessary condition for calculating d_{ac} and d_{a_1c}. If $a \xrightarrow{\alpha} a'$ and $a' \in B$, then $d_{a'c} = d_{ac}$.

In all these cases the lower bound of distance is equal to the upper bound ($l'_{ac} = u'_{ac} = d_{ac}$).

The distance bounds between the $\tilde{A}-$set atoms are taken from the $LU-$matrix, built according to Crippen ($l'_{a_1a_2} = l_{a_1a_2}, u'_{a_1a_2} = u_{a_1a_2}$, when $a_1, a_2 \in \tilde{A}$).

The distance bounds between the other points of the $\tilde{A}-$set and points of the $C-$set are defined quite widely. Therefore, a new matrix $(LU)'$ of the distance bounds between the points of the set $E = \tilde{A} \cup C$ is being formed. $(LU)'$ elements serve as parameters of the F function (2).

We should then add to the function F the terms containing the condition of $a \xrightarrow{\alpha} a'$ transits. The direction vector $\vec{b}$ may correspond to the symmetry element α if α is an axis, or it is the perpendicular vector, when α is a plane.

Let's suppose vector $\vec{b}$ coincides with one of the coordinate axes, i.e. $b_{i_1} = 0$ $b_{i_2} = 0$, $b_{i_3} = 1$, and the transition $a \xrightarrow{\alpha} a'$ produces the point $c \in \alpha$, $l_{aa'}$ equals to $u_{aa'}$. Then, if α is a proper axis, the i_3 coordinates of the points a and c should be equal, but in case $a' \in B$, the i_3 coordinates of the points a, a' and c should be

equal. Therefore, the function F is supplemented by terms which become zero if the conditions mentioned above are met:

$$\varphi_t = (x_{ai_3} - x_{ci_3})^2, \qquad \varphi_s = (x_{a'i_3} - x_{ci_3})^2$$

In the case when α is a plane, two analogous terms are used instead of each of the given ones: i_1 is used instead of i_3 within one of them, and i_2 - instead of i_3 within another. When α is an improper axis of order $k > 2$, then $\alpha \| OZ$, $a \xrightarrow{\alpha} a'$, $a' \in B$, the i_3 coordinates of the points a and a' become antipodal, and the corresponding term $\varphi_r = (x_{ai_3} + x_{a'i_3})^2$ should be added to the function F.
This is illustrated by the following example, where $\alpha_1 = C_2$, $\alpha_2 = \sigma_h$, and the coordinates of vector $\vec{b}$ are $(0,0,1)$ for both elements of symmetry (Fig. 5).

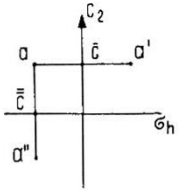


Fig. 5. Location of auxiliary points

Let $a \in A$, $a \xrightarrow{\alpha_1} a'$, $a \xrightarrow{\alpha_2} a''$, $a' \notin B$, $a'' \notin B$ and $\bar{c} \in \alpha_1$, $\bar{\bar{c}} \in \alpha_2$. Then $F' = F(LU') + (x_{a3} - x_{\bar{c}3})^2 + (x_{a1} - x_{\bar{c}1})^2 + (x_{a2} - x_{\bar{\bar{c}}2})^2 + \ldots$
Simultaneously, the bounds of the coordinates' variation are being determined for those points of the set E, which are situated on the symmetry elements. In case $e \in E$, $e \in \alpha$ and α is an axis, $x_{ei_1} = x_{ei_2} = 0$, i.e. the bounds of these variables' variation are zero. When α is a plane, $x_{ei_3} = 0$. The Fig. 5 illustrates, that if $\alpha_1 = C_2$ and $e = \bar{c}$, $x_{\bar{c}i_1}$ and $x_{\bar{c}i_2}$ go to zero, and if $\alpha_2 = \sigma_h$ and $e = \bar{\bar{c}}$; $x_{\bar{\bar{c}}i_3}$ goes to zero. When the transition $a \xrightarrow{\alpha} a'$, where α is an improper axis, produces the point c, the point $c \in C$ is situated in the origin of the coordinates (point 0), and all the bounds of its coordinates' variation are zero. The bounds of the variables which are not located on the symmetry elements are accepted to be rather wide.
Let us consider $a \xrightarrow{\alpha} a'$ ($l_{aa'}$ being equal to $u_{aa'}$) and a point $c \in \alpha$ which is produced by this transition as well assume a vector $\vec{b}$, not coinciding with any of coordinate axes, which corresponds to the element α. Then the function F should be supplemented by terms containing expressions for the cosines of the angles γ_1, γ_2 between two pairs of vectors $\vec{b}, \vec{0c}$ and $\vec{b}, \vec{ac}$, correspondingly. Let α be a proper axis, then $cos\ \gamma_1 = \pm 1$, and $cos\ \gamma_2 = 0$, if α is a plane, then $cos\ \gamma_1 = 0$; $cos\ \gamma_2 = \pm 1$. Therefore, in case when α is a proper axis, the function F is supplemented by the terms $(cos^2\gamma_1 - 1)$ and $cos^2\gamma_2$, which turn into zero, when $cos\ \gamma_1 = \pm 1$ and $cos\ \gamma_2 = 0$. The values $cos\ \gamma_1$ and $cos\ \gamma_2$ are expressed by coordinates of the vector $\vec{b}$ and points a and c. Analogous terms are added to the function F by the points

of the set $\tilde{A}$, situated on the symmetry elements, non-coinciding with eigther any coordinate axis or any coordinate plane. Thus, for the structure, illustrated by Fig. 2, we have: $A = \{1,2,8,7\}$, $B = \{\emptyset\}$, $2 \in \sigma_{v_2}$ $(cos(2\pi/3),\ sin(2\pi/3), 0)$.
In this case, the function F should be supplemented by the following term:

$$\varphi_v = \frac{(x_{21} \cdot cos(2\pi/3) + x_{22} \cdot sin(2\pi/3) + x_{23} \cdot 0)^2}{(x_{21}^2 + x_{22}^2 + x_{23}^2) \cdot 1}$$

Now let us assume, that $a \xrightarrow{\alpha} a'$, $a \in A$, $a' \in B$, $l_{aa'} \neq u_{aa'}$, i.e. the distance between the image and the prototype is not settled. Then the function F is supplemented by the term $\varphi_w = \sum_{i=1}^{3} [\psi_i(a) - x_{a'i}]^2$, where ψ is a vector-function of geometrical transformation which corresponds to the symmetry element α.
The initial coordinates of the set $\tilde{A}$ of points are taken from the symmetrized spatial model $M(X_{ik}^0)$ and the initial coordinates of the set of points C defined in the following way. Let $c \in C$ be produced by the transition $a \xrightarrow{\alpha} a'$, then $X_{cj}^0 = \sum_{i=1}^{k} X_{a_i}^0 j/k$ $(j = 1,2,3)$, where a_i are the images of a relatively k degrees of the symmetry element α, the coordinates in the right part being derived from the symmetrized model $M(X_{ik}^0)$. For example, in the structure, displayed on Fig. 1, the point q, produced by $2 \xrightarrow{\sigma_3} 6$ transit, receives initial coordinates $X_{qj}^0 = 1/3\ (X_{2j}^0 + X_{6j}^0 + X_{10j}^0)$, $j = 1,2,3$.
Coordinates of atoms of the set A, gained in the process of optimization, are used for calculation of coordinates of all the other atoms within the molecule by means of "reproducing" the "representatives" relatively generators.
The algorithm of symmetrization can be discribed in short as a sequence of the following steps:

1. Obtain the graph automorphism group according to the adjacency matrix.
2. Construct spatial model of the molecule by means of Crippen's method using the adjacency matrix and geometrical parameters.
3. Input symbols of the point symmetry group. Identify the point group generators and symmetry elements.
4. Separate the atom set into the orbits.
5. Orient the Crippen model in the space.
6. Determine coordinates of atoms belonging to the first orbit.

6.1 Select an atom - "representative" and determine its coordinates.

6.2 Calculate coordinates of the rest orbit atoms "reproducing" the "representative" by means of the generator combinations.

7. Repeat step 6 for the remaining orbits.
8. Optimize the coordinates of all "representatives".

8.1 Set up the optimization problem:

i) Form the set of the independent variables presenting the coordinates of a definite part of atoms and additional points,

ii) Form the target function,

iii) Form the bounds of the variables' variation,

iv) Calculate the initial values for variables.

8.2 Solve the optimization problem.

9. Determine coordinates of all the rest atoms applying generator combinations to the "representatives".

In a majority of cases it doesn't make any sense to optimize the coordinates of all the atoms in molecule according to Crippen's method. It is more profitable to induce the symmetry, basing upon the coordinates, expressed by the eigenvalues and eigenvectors of the matrix G, as upon the initial approximation. This allows to save time. For example, it was found that calculation of adamantan molecule by Crippen's method required 5 times as much of processor time as in the case when the symmetrizing algorithm without optimization of all molecular atom coordinates was used.

Crippen's algorithm and the algorithm of symmetrization, worked out by us, were realized as a Fortran program, effectivity of which has been checked up on a great number of complicated polycyclic organic and inorganic structures, examples of which are given on Fig. 6. While applying the algorithm of the symmetrization to the structures given on Fig. 6, we discovered, that the maximal deviation of the calculated geometrical parameters from the prescribed ones is within the average error, usual for the programs in quantum chemistry. In special cases, when the algorithm of symmetrization also gives not satisfactory results the spatial structure may be built with the aid of programs in quantum chemistry and molecular mechanics. With the initial approximation, obtained, using the algorithm of symmetrization, the computer time of calculation with the help of quantum-chemical programs can be shortened essentially.

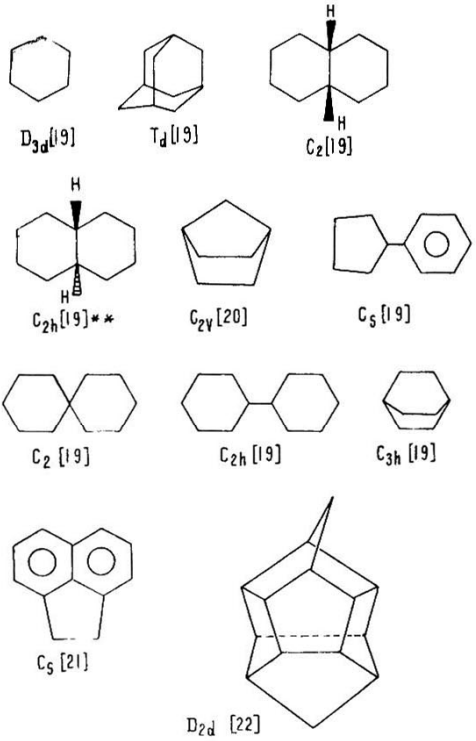


Fig. 6a. Molecules for which spatial models were built.

Digits in the square brackets refer to the sources of the values of inner coordinates of molecules. The structures marked by one star were calculated by method [15] the symmetry being disturbed. The structures marked by two stars were calculated by the same method both the symmetry and geometric parameters being disturbed.

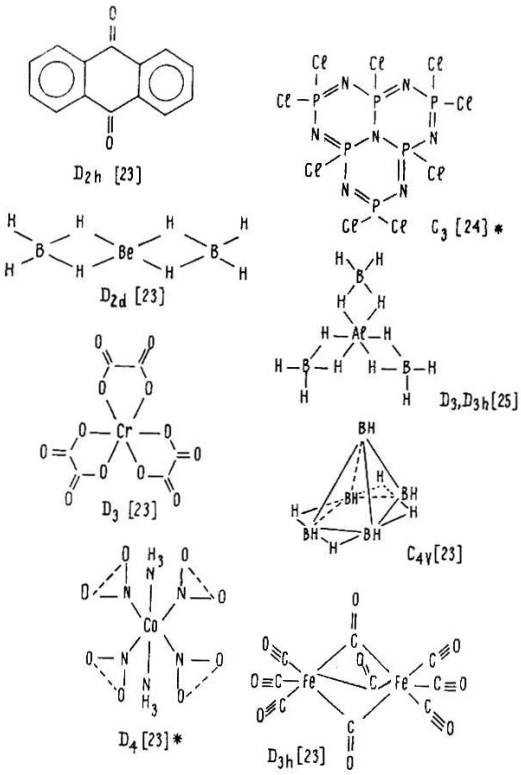
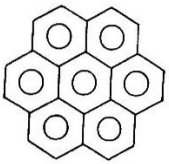
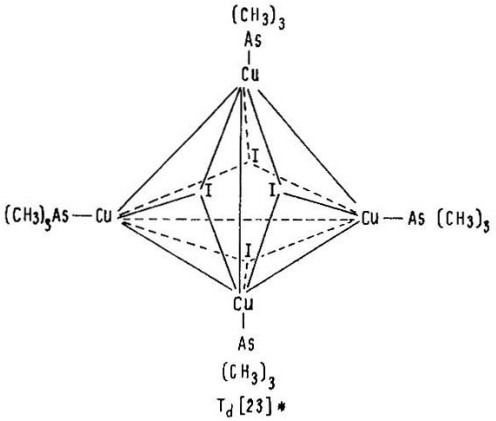


Fig. 6b. Further molecules for which spatial models were built.



D_{6h} [23]

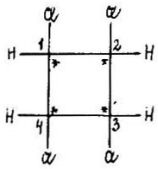


T_d [23] *

Fig. 6c. Further molecules for which spatial models were built.

3 Generation of the steroisomers

The generation of the stereoisomers of the structure assigned only by its adjacency matrix was performed using method [17] based on the molecular topology analysis. The main idea of a method proposed by Nourse is the identification of the stereocentres using the analysis of their local topological symmetry and construction of the configuration symmetry group (CSG) (on the basis of GAG), containing permutations of the stereocentres, the definite stereocentres of which are supplied with the symbols of inversion. The steroisomers are generated as binary $n-$dimensional vectors, n being the number of stereocentres.
Generation is a process of multiplication of such vectors beginning with the zero one by the elements of the CSG. The binary vector representing a lowest integer which has not been involved in the process of multiplication before is a code of next stereoisomer from the list of the different ones. The process stops when all the 2^n $n-$dimensional vectors are exhausted. Below you can see a structure where the stereocenters are marked by stars.



The results of multiplication of the stereoisomers' codes by the elements of the CSG are shown in Table 1. Every column of Table 1 contains a class of equivalence referring to every distinct stereoisomer. Thus, four different steroisomers have been obtained: 0000; 0001; 0011; 0110. Only the assymetrical stereocenters and double bonds are decoded by the program - R/S configuration is given to the former, cis-trans configuration is given to the latter.

Table 1:

$CSG \backslash STEREOISOMERS$	0000	0001	0011	0110
(1)(2)(3)(4)	0000	0001	0011	0110
(1′)(24)(3′)	1010	1110	1100	1001
(12)(34)	0000	0010	0011	1001
(12′ 34′)	0101	1101	1100	0110
(13)(2′)(4′)	0101	0100	1100	1001
(1′3′)(2′4′)	1111	1011	0011	0110
(1′43′2)	1010	1000	1100	0110
(1′4′)(2′3′)	1111	0111	0011	1001

In the development of the Nourse method we've offered an approach, which combines in itself both the analysis of the molecular topological symmetry and the

consideration of the geometrical qualities. Accounting of the spatial structure lets us identify the symmetrical group (global symmetry) and the stereocenters of the molecule more precisely. The atomic coordinates are used by the program to ascertain the molecular point symmetry group (this group is used instead of *GAG*) and to analyse the potential stereocenters *local* symmetry.. Our approach allows to generate non-coinciding stereoisomers which can't be distinguished by Nourse method.

4 Determination of the coordinates of the stereoisomer atoms

On the basis of the heuristic criteria the algorithm allows us to reject the impossible structures containing the transdouble bonds in the $n-$membered cycle with $n < 8$, the small cycles, condensed in trans-position, inverted bicycles [26]. The atomic coordinates are calculated for all the rest of the stereoisomers generated by the program.

The spatial models of cis- and transisomers of the polycyclic structures with the condensed cycles and the bridged fragments, as well as of the unsaturated structures with double bonds in the cycles, are built independently with the aid of the symmetrization algorithm or Crippen's method. In the rest of the cases the initial stereoisomer atomic coordinates, found by these algorithms are used in the determination of the atomic coordinates of the generated steroisomers. The main idea of the algorithm of the determination of the atomic coordinates of the stereoisomer family using the origin stereoisomer coordinates is the following. The initial stereoisomer is coded by means of the $n-$dimentional binary vector the $i-$th component of which is the configuration code of $i-$th stereocentre. Configuration I corresponds to code 0, configuration II - to 1:

Here 4 is the biggest ligand number. It is characteristic for configuration I that the sequence of atoms 1-2-3 is carried out counterclockwise if one looks at it from the atom number 4 position, and clockwise for the configuration II. The establishment of the configuration type is fulfilled with the aid of the certain geometrical transformations. The code of each stereoisomer is compared with the code of the original one in order to find the atomic coordinates. The stereocentres with the

non-coinciding codes have to be inverted. If the stereocentre's valency is 4, the ligand pair with the least number of bonds with the other stereocentres has to be chosen. Afterwards a plane is drawn through the other ligand pair and invertable stereocenter, reflecting the first pair together with the bonded atoms. This assumes the possibility of the other stereocenters inversion, reasoning the necessity of the constant revision of the code range from the left to the right until the complete coincidence of the codes of the initial and the consequent stereoisomers is achieved.

5 The stereoisomer graphic image construction

Special coordinates are calculated for the graphic image construction independently of the occurrence of the 3-dimensional space coordinates. Here the usage of the 2-dimensional coordinates of the original molecular picture is possible. The fictitious third coordinates are assigned to the stereocentres and their ligands. The obtained conventional space-model is exposed to the procedure of coding and the stereocenter inversion as it was described above. In this way we manage to possess the conventional 3-dimensional coordinates of all the stereocenters and their ligands, necessary for construction of their graphic image: the first two coordinates are used for the plane depiction, the third ones for the introduction of the stereobond symbols (if the third coordinate of stereocentre is greater than the third coordinate of its ligand we use the symbol  between them, otherwise - the symbol ). The final stage includes the coloured stereoisomer display in the graphic regime using the generally accepted symbols of the stereobonds. The stereobond symbols are not to be used for the bonds connecting the two stereocenters in order to avoid ambiguity and therefore some of the stereocenters are placed in the separate picture together with the ligands.

6 Conclusion

Thus in the present work a unified approach has been developed, which allows to solve the following tasks:

a) to compute the Cartesian coordinates of a molecule if its geometric parameters and point symmetry group in symbolic form are given;

b) exhaustively to generate and graphically to construct all the non-coincident stereoisomers;

c) to generate the spatial models of the stereoisomer family using the Cartesian coordinates of one family member.

The elaborated programs are part of an expert system (RASTR) designed for elucidation of the molecule structure according to its IR; PMR and NMR C-13 spectra.

They can be used independently for solution of the molecular modelling problem, for molecular spectra prediction, as well as in the capacity of the components of the artificial intelligence systems, etc.

7 References

1. M.E. Elyashberg, L.A. Gribov and V.V. Serov, "Molecular spectral analysis and computer", Moscow, Nauka, 1980 (in Russian).
2. L.A. Gribov, M.E. Elyashberg and M.H. Raikhshtat, J. Mol. Struct, **53**, 81 (1979).
3. L.A. Gribov, M.E. Elyashberg, V.N. Koldashov and V.V. Pletnjov, anal. Chim. Acta, **1948**, 159 (1983).
4. V.V. Serov, L.A. Gribov and M.E. Elyashberg, J. Mol. Struct., **129**, 183 (1985).
5. M.A. Elyashevitch, "Atomic Molecular Spectroscopy", Moscow, Nauka, 1962 (in Russian).
6. R.L. Hilderbrandt, J. Chem. Phys., **51**, 1654 (1969).
7. J.E. Norlander, A.F. Bond and M. Bader, Computers Chemistry, **9**, 209 (1986).
8. H. Essen, Computational Chemistry, **4**, 136 (1983).
9. H.B. Thompson, J. Chem. Phys., **53**, 1968 (1970).
10. L.P. Trigub and U.A. Cruglyac, Zhurn. Strukt. Khim (in Russian), **24**, 167 (1983).
11. E. Gavuzzo and S. Paghuca, Acta Cryst., **B28**, 1968 (1972).
12. J. Dillen and H. Gase, Computers Chemistry, **4**, 113 (1980).
13. A.V. Bandura and N.P. Novoselov, Zhurn. Strukt. Khim (in Russian), **19**, 357 (1978).
14. A.A. Zenkin and V.I. Mochalov, Vestnik Mosk. Univers., Series 2, Chemistry, **28**, 88 (1987) (in Russian).
15. G.M. Crippen, "Distance Geometry and Conformational Calculations", N.Y., Research Press, 1981.
16. N. Go and H.A. Sheraga, Macromolecules, **6**, 273 (1973).

17.a) I.G. Nourse, J. Am. Chem. Soc., **101**, 1210 (1973).

17.b) I.G. Nourse, R.E. Carhart, D.H. Smith and C. Djerassi, J. Am. Chem. Soc., **101**, 1216 (1973).

18. M. Uchino, J. Chem. Comput. Sci., **20**, 116 (1980).

19. M.E. Elyashberg, Ya.R. Karasev, V.A. Dementyev and L.A. Gribov, "Interpretated Vibrational Spectra of Hydrocarbons-Derivatives aof Cyclohexane and Cyclopropane", Moscow, Nauka, 1988.

20. L. Doms, L. Van den Enclen, H.I. Geise and C. Van Alsenoy, J. Amer. Chem. Soc., **105**, 158 (1983).

21. V.G. Dashevsky, "Conformations of organic molecules", Moscow, 1974.

22. I. Hargitai, J. Brunvoll, S.J. Cyvin and A.P. Marchand, J. Mol. Struct, **140**, 219 (1986).

23. "Tables of Interatomic Distances and Configuration in Molecules and Ions", completed by H.J.M. Bowen, M.A.D. Phil and J. Donohue, Special Publication N 11, London, The Chemical Society. Burlington Hourse, 1988.

24. W. Harrison and J. Trotter, J.C.S. Dalton, N 5, 623 (1972).

25. L. Alenningen, G. Cundersen and A. Haaland, Acta Chemica Scandinavica, **22**, 328 (1968).

26. J.G. Nourse, D.H. Smith, R.E. Carhart and C. Djerassi, J. Am. Chem. Soc., **102**, 6289 (1980).