Topology and Group Theory-Tools for Determinating the Stereochemistry of Molecules

Stefan Reichelt, Antje Reichelt, Nicole Müller, and Ivar Ugi

*Technical University of Munich – Organic Chemical Institute Chair I,
Lichtenbergstraße 4, D-85748 Garching, Germany*

Modern Chemistry requires to handle complex stereochemical problems with a computer. Therefore, it is necessary to find a suitable way of modelling ensembles of molecules and their reactions. This kind of modelling must also consider dynamic structural properties of molecules. This cannot be described sufficiently by rigid models.

Developing concepts and computer programs for describing stereochemical and constitutional regards simultanously requires an exact chemical nomenclature. Thus, a nomenclature must be found which is easily applicable and produces only a small amount of data.

Especially concerning stereochemistry this nomenclature must be applicable to atoms of any validity. No rules for special cases should be defined. For a definite description of ensembles of molecules our algorithms profit by topological and graph theoretical methods. Based on uncoloured molecular graphs, first the chemical nature of the vertices and finally by group theoretical aspects stereochemistry is introduced. This hierachical ordering allows to compare structural properties of chemical different molecules. Thus, for example reaction patterns of molecules can be transferred between chemical different but structural identical ensembles.

GRAPHS OF MOLECULES

Many properties of molecules are represented by their (chemical) graphs.[1–4]
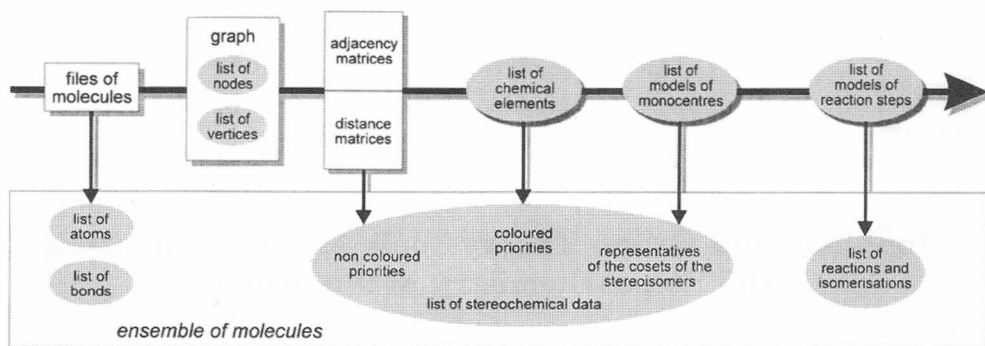
Scheme 1. Stereochemical specification of ensembles of molecules.

Let M be an ensemble of molecules consisting of a set of atoms $A = \{a_1, a_2, ..., a_n\}$ and a set of bonds $B = \{b_1, b_2, ..., b_m\}$. All bonds between two connected atoms are described by only *one* edge e. Bonds with delocalized electrons and multicentre bonds are separated by an according number of edges. The style of the bond is characterized by labels. The set V of vertices which corresponds to A and the set E build the Graph G.
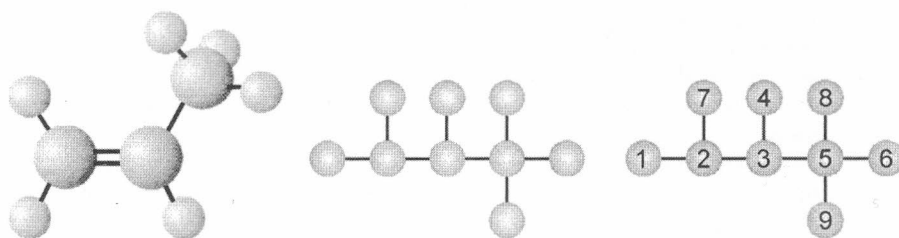

Figure 1. Molecule **1**   The graph of **1**   The arbitrarily labeled graph of **1**

The adjacency matrix and the distance matrix of a molecule

The connectivity of the atoms of a given molecule represents the topological properties of the graph.

The adjacency matrix A(G) of the graph G(M) reflects the connectivities of the molecule M. For the calculation of this matrix the nodes of G must be enumerated.

The adjacency matrix is simply calculated from the list of bonds of M. For any atoms connected by an edge the corresponding entries a_{ij} are set to **1**.

$$a_{ij} = \begin{cases} 1, \text{ if } i \text{ and } j \text{ are connected} \\ 0, \text{ else} \end{cases} \qquad d_{ij} = \begin{cases} 0, \text{ if } i = j \\ \text{else, the shortest distance between } i \text{ and } j \end{cases}$$

Figure 2. Calculation of the adjacency matrix and the distance matrix.

The sum of all entries in the row j is called the degree of the vertex. Note that it may differ from the *chemical* valence of the corresponding atom.

Another important matrix for the topological analysis of chemical graphs is the distance matrix $\boldsymbol{D}$.[5,6,7] $\boldsymbol{D}$ contains the shortest distances d_{ij} between all vertices of G.

TOPOLOGICAL BASIS FOR STRUCTURING CHEMICAL GRAPHS

Topology of the non-coloured chemical graph

Let the topology of a single atom be defined as the shortest distances to all other atoms of an ensemble of molecules. This is expressed by the number of edges between the corresponding vertices. In an ensemble of molecules with n atoms all vertices which are not connected will get the entry n.

The entries of the distance matrix reflect the distances between the corresponding atoms. The atoms with the same topological surrounding get the same entries, however, in a different order. To find such equivalent atoms, the number of neighbours in a given distance is calculated. Thus, a new matrix — the layer matrix[8-10] – is compiled. The numbers of the first, second, .. neigh-

TABLE I

Priorities in **1**

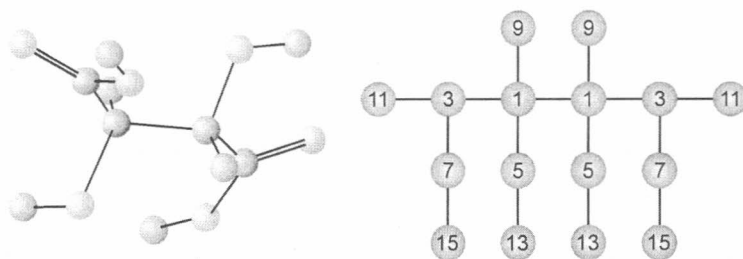| Atom	Neighbours				Priority
	1st	2nd	3rd	4th	
5	4	2	2	0	1
3	3	5	0	0	2
2	3	2	3	0	3
6	1	3	2	2	4
8	1	3	2	2	4
9	1	3	2	2	4
4	1	2	5	0	7
1	1	2	2	3	8
7	1	2	2	3	8

Figure 3. Tartaric acid and the corresponding graph with priorities.

bours creat a new order by which temporary priorities are given to the atoms. If there are atoms with the same priority those of their neighbours are analysed. These are sorted and then used to calculate new priorities. These steps are recursively repeated, until there are no more changes.

The application of this algorithm results in the topological priorities of the non-coloured chemical graph. Topologically equivalent atoms get the same priority.

The topological priorities provide an ordering of the atoms. Atoms with equivalent indices may be interchanged. This may be interpreted as free rotations (*e.g.* CH$_3$-groups) or symmetries (ethen, benzene, tartaric acid, *etc.*).

Important stereochemical aspects of these non-coloured graphs are shown for the example of tartaric acid (Figure 3). The atom with the priority 1 has four topologically different ligands (1, 3, 5, 9) and *may* be chiral. The existence of an equivalent atom in the same molecule results here in an innermolecular symmetry. Thus the number of possible stereoisomers is reduced.

Topology of the coloured chemical graph

So far, the chemical nature of the atoms has been suppressed for the following advantages:

- The machinery of graph theory is completely applicable.
- Molecules with similar skeletons but different chemical composition are comparable.
- Patterns of reactions are applicable regardless of the chemical nature of the atoms and bonds.

As following, the chemical nature of the vertices will be introduced as an additional differentiation. Vertices with a lower atomic number are given a lower priority.

The following examples show the calculation of the priorities:

$$H^1 - N^2 = C^3 = O^4 \qquad H^4 - N^1 = C^2 = O^3$$

Figure 4. Molecule **2** arbitrarily enumerated   Priorities in **2**

TABLE II

Calculation of the priorities in **2**

Atom	Distance matrix				Neighbours			Priority (non coloured)	Atom	Priority (coloured)
	1	2	3	4	1st	2nd	3rd			
1	0	1	2	3	1	1	1	3	H	4
2	1	0	1	2	2	1	0	1	N	1
3	2	1	0	1	2	1	0	1	C	2
4	3	2	1	0	1	1	1	3	O	3

$$H^1 - O^2 = C^3 = N^4 \qquad H^4 - O^1 = C^2 = N^3$$

Figure 5. Molecule **3** arbitrarily enumerated   Priorities in **3**

TABLE III

Calculation of the priorities in **3**

Atom	Distance matrix				Neighbours			Priority (non coloured)	Atom	Priority (coloured)
	1	2	3	4	1st	2nd	3rd			
1	0	1	2	3	1	1	1	3	H	4
2	1	0	1	2	2	1	0	1	O	1
3	2	1	0	1	2	1	0	1	C	2
4	3	2	1	0	1	1	1	3	N	3

Important topics are:

- The group with the highest priority must be recalculated first.
- Within one step only the priorities of a single group of equivalent indices are changed.

- Atoms within a single group of equivalent indices and with the *same* sort of atoms are *not* distinguishable.
- The steps are repeated until the priorities are constant.

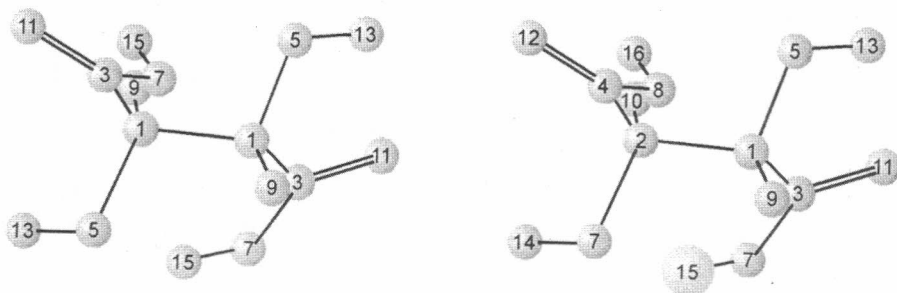The chemical distinction of the vertices cancels certain symmetries.

Figure 6. Priorities in tartaric acid (*A*)   in a derivate (*B*)

The symmetry in tartaric acid A is canceled by replacing a single atom (15).

CHARACTERIZING THE STEREOCHEMISTRY
OF MOLECULES BY THEIR TOPOLOGY

The theory of the chemical identity groups provides algebraic solutions to stereochemical questions. An introduction is found in Refs. 11 and 12. The permutational isomerism is the central part of this theory. This enables the application of group theoretical methods to stereochemistry.

The new aspects of our approximation are the intensive analysis of the chemical graphs of molecules and the modelling of stereochemical properties by the stereochemical data of its monocentres. Particularly, an advantage is that the cumbersome determination of the identity preserving permutations of the molecule as a whole can be avoided. In the following context we only consider monocentric permutation isomers, since the monocentres may be combined to polycentric skeletons.

As models for the monocentres serve idealized structures like octahedra for hexavalent atoms. They are not based on any metric data like bond length or bond angles. Once the priorities are determined, the ligands of an atomic centre can be characterized by the corresponding indices. Each ligand occupies a topological position of this model. Permutations of ligands
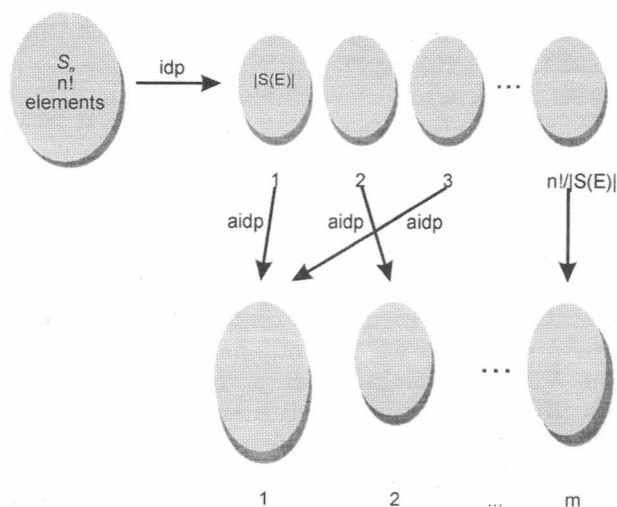
Figure 7. Specification of the permutational isomers.

mean exchanges of the ligands on these positions. The relative arrangement of the positions (*e.g.* axial) is fixed. Variations from the idealized model resulting from different ligands or intramolecular or external forces can be taken into consideration by an appropriate choice of the identity preserving permutations. Molecular symmetries are caused exclusively by the relative arrangement of identical ligands. The model of each topology must be provided only once time.

Generally, the identity preserving permutations (idp) are defined as permutations that transform an ensemble of molecules into a chemically identical ensemble. The most important identity preserving permutations are outer rotations, but they can also be caused by the identity of ligands, isomerisations like tautomerism and fluctuations or delocalized electrons.

Considering the monocentre E with n different ligands L, there exist $n!$ permutation isomers, that form the family of permutation isomers P(E). Some of these permutation isomers can be interconverted by identity preserving permutations. The following group theoretical representation for permutational isomerism is valid, if exclusively the outer rotations are taken into account:

The family of permutation isomers forms the symmetric group S_n with the cardinality $n!$. The identity preserving permutations build a subgroup $S(E)$ of S_n. S_n is divided into cosets of $S(E)$, each of it represents a single permutational isomer. The cardinality of each of these cosets corresponds to $|S(E)|$, since this is the number of outer rotations. Therefore the number of different cosets is $n!/|S(E)|$.

Permutations of ligands are denoted as permutations of the numbers of the topological positions of the corresponding model. The resulting permutations of each coset are ordered lexicographically. In the following context, the first element shall represent its coset. This criterion is fixed arbitrarily. A definite and consistent assignment is required in order to avoid cumbersome comparisons of the complete cosets. Nevertheless, if it is necessary, the whole cosets can be generated through the identity preserving permutations.

If some ligands of a monocentre are equivalent, there are additional identity preserving permutations (aidp). These describe the interchange of such ligands. Therefore, some cosets are combined, and the number of different stereoisomers m decreases. The number and the structure of these stereoisomers and additional identity preserving permutations can be recognized automatically by analysing substitutional patterns and the relative arrangements of the ligands. Generally, the number of the united cosets and the additional identity preserving permutations can be different for each new coset. Each stereoisomer is denoted as a single permutation. The substitutional pattern is made out of the priorities of the ligands.

The algorithm is:

1. Copy the substitutional pattern to all representatives of the cosets, and put it into a list L.
2. Get the first element of L and put this in a list L'. Construct the complete coset by applying the identity preserving permutations to this element.
3. Compare the coset with all elements of L.
4. Unite equivalent cosets, remove them from L, and generate the additional identity preserving permutations.
5. If there still exist more than two representatives go to step 2.
6. Place the last representative in L'.

In L' the number of the elements is the number of the different stereoisomers, and the stereoisomers can be visualized by the corresponding model.
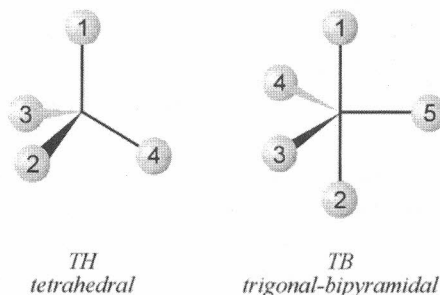

TH
tetrahedral

TB
trigonal-bipyramidal

Figure 8. Models of monocentres.

Let M be a monocentre with a trigonal bipyramidal skeleton. The ligands are: L_1, L_2 = A, L_3 = B, L_4 = C, L_5 = D. Thus it has the substitutional pattern AABCD. For the determination of the identity preserving permutations, first all ligands are regarded as distinguishable. The labelling of the positions of the trigonal bipyramidal model and the distribution of the ligands are as shown above. There exist 5! = 120 permutational isomers. The number of the identity preserving permutations is 6. That means 6 outer rotations. They divide the 120 permutational isomers into 20 cosets (stereoisomers). The identity of two ligands in this example unites two cosets. Thus, there exist 10 cosets/stereoisomers and for each coset one additional identity preserving permutation. Further cosets can be united if for example isomerisations between two stereoisomers are allowed. The permutations of the example are shown in the following table:

TABLE IV

The representatives and the idp of a trigonal bipiramidal model with the substitutional pattern AABCD

Permutational isomers	Idp	Cosets-stereoisomers	Additional idp	Cosets-stereoisomers
[1 2 3 4 5]		[1 2 3 4 5]	[1 2 3 5 4]	
[1 2 3 5 4]		[1 2 3 5 4]	[1 2 3 5 4]	
[1 3 2 4 5]		[1 3 2 4 5]	[3 2 1 4 5]	
[1 2 5 4 3]		[1 3 2 5 4]	[3 2 1 4 5]	
		[1 4 2 3 5]	[3 2 1 4 5]	
		[1 4 2 5 3]	[3 2 1 4 5]	[1 2 3 4 5]
		[1 5 2 3 4]	[3 2 1 4 5]	[1 3 2 4 5]
	[1 2 3 4 5]	[1 5 2 4 3]	[3 2 1 4 5]	[1 3 2 5 4]
	[1 2 4 5 3]	[2 3 1 4 5]	[3 2 1 4 5]	[1 4 2 3 5]
	[1 2 5 4 3]	[2 3 1 5 4]	[3 2 1 4 5]	[1 4 2 5 3]
	[2 1 3 5 4]	[2 4 1 3 5]	[3 2 1 4 5]	[1 5 2 3 4]
	[2 1 4 3 5]	[2 4 1 5 3]	[3 2 1 4 5]	[1 5 2 4 3]
	[2 1 5 4 3]	[2 5 1 3 4]	[3 2 1 4 5]	[3 4 1 2 5]
		[2 5 1 4 3]	[3 2 1 4 5]	[3 5 1 2 4]
		[3 4 1 2 5]	[1 2 3 5 4]	[4 5 1 2 3]
[5 3 2 4 1]		[3 4 1 5 2]	[1 2 3 5 4]	
[5 2 3 4 1]		[3 5 1 2 4]	[1 2 3 5 4]	
[5 1 2 3 4]		[3 5 1 4 2]	[1 2 3 5 4]	
[5 4 1 2 3]		[4 5 1 2 3]	[1 2 3 5 4]	

Different stereoisomers are denoted as permutations. A further specification like chirality can follow. After analysing the monocentres they will be recombined to the molecule. Also conformational isomers can be described.

The developed algorithms allow the computation of all cosets. They belong to the stereoisomers with a given monocentre and further identity preserving permutations. This algorithm is not only suitable for comparing identical ligands but also for analysing other relations of equivalence. Another advantage of these algorithms is that not only single molecules but also ensembles of molecules and even libraries can be described and analyzed.

Accumulation of the representatives of the cosets

If a monocentre has ligands with identical priorities these ligands are distinguishable if the resulting permutation is not found in the idps.

Let M be a monocentre with a trigonal bipyramidal skeleton and its ligands have the priorities 5, 4, 5, 12, 8 (Figure 9).
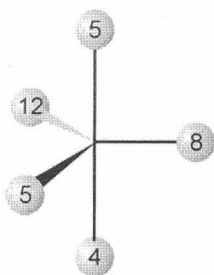

Figure 9. Model M.

These are taken to create the corresponding permutation. There exist two permutations [2 1 3 5 4] and [3 1 2 5 4], because of the identical priorities. The additional permutation (1 3) is not found in the idps and therefore we get two representatives of the cosets:

$$[2\ 1\ 3\ 5\ 4] => [1\ 2\ 3\ 4\ 5]$$
$$[3\ 1\ 2\ 5\ 4] => [1\ 3\ 2\ 4\ 5]\ .$$

With [1 2 3 4 5] < [1 3 2 4 5] we get the new priority 5 for the axial and 6 for the equatorial ligand.

Within groups of equivalent priorities this procedure is used to find inter-dependent stereochemical properties like in 1,2,3-cyclopropantriol 4 (Figure 10).

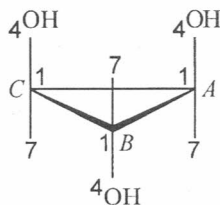The equivalent monocentres A, B and C are calculated and the resulting priorities are accumulated:

Figure 10. Molecule 4.

TABLE V

Accumulation of the priorities in 4

Monoce ntre	Neighbours	Permutations	Representative	Priorities			Accumulation		
				A	B	C	A	B	C
A	4 B C 7								
	3 1 1 4	[3 1 2 4]	[1 2 3 4]	−	1	2	0	1	2
		[3 2 1 4]	[1 2 4 3]						
B	4 7 A C								
	3 4 1 1	[3 4 1 2]	[1 2 3 4]	1	−	2	1	1	4
		[3 4 2 1]	[1 2 4 3]						
C	4 A B 7								
	3 1 1 4	[3 1 2 4]	[1 2 3 4]	1	2	−	2	3	4
		[3 2 1 4]	[1 2 4 3]						

The new priorities in 4 are A = 1, B = 2 and C = 3. This leads to new priorities for the whole molecule.
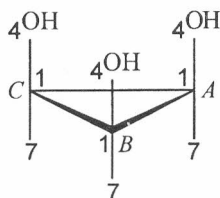
The calculation of the stereoisomer 5 (Figure 11) is:

Figure 11. Molecule 5.

TABLE VI

Accumulation of the priorities in **5**

Mono-centre	Neighbours	Permutations	Representative	Priorities			Accumulation		
				A	B	C	A	B	C
A	4 B C 7								
	3 1 1 4	[3 1 2 4]	[1 2 3 4]	–	1	2	0	1	2
		[3 2 1 4]	[1 2 4 3]						
B	4 7 A C								
	3 1 1 4	[3 1 2 4]	[1 2 3 4]	2	–	1	2	1	3
		[3 2 1 4]	[1 2 4 3]						
C	4 A B 7								
	3 1 1 4	[3 1 2 4]	[1 2 3 4]	1	2	–	3	3	3
		[3 2 1 4]	[1 2 4 3]						

The (new) priorities in **5** are A = 1, B = 1 and C = 1.

A procedure for the accumulation of the priorities is created. Thus, it is possible to determine an absolute stereochemical descriptor even for molecules where the stereochemistry of a single atom depends on the stereochemistry of another atom.

CONCLUSION

A way to analyse and describe stereochemical properties of molecules is found. Grouptheoretical methods are applicable because the stereochemical informations are noted as permutations. If one stereoisomer is found the other stereoisomers can be created and described easily.

The mentioned algorithms are implemented in Pascal. Data of molecules can be taken from programs like ALCHEMY, HYPERCHEM, *etc.*.

REFERENCES

1. R. B. King (Ed.), *Studies in Physical and Theoretical Chemistry*, Elsevier, Amsterdam, Vol. 28 (1983).
2. A. T. Balaban, *J. Chem. Inf. Comput. Sci.* **25** (1985) 334–343.
3. D. H. Hansen and P. C. Jurs, *J. Chem. Educ.* **65** (1988) 547–580.
4. K. Balasubramanian, *Chem. Rev.* **85** (1985) 599–618.
5. Z. Mihalić, D. Veljan, D. Amić, S. Nikolić, D. Plavšić, and N. Trinajstić, *J. Math. Chem.* **11** (1992) 223–258.
6. W. R. Müller, K. Szymanski, J. V. Knop, and N. Trinajstić, *J. Comput. Chem.* **8** (1987) 170–173.

7. M. Bersohn, *J. Comput. Chem.* **4** (1983) 110–113.
8. V. A. Skorobogatov and A. A. Dobrynin, *MATCH* **23** (1988) 385–390.
9. M. V. Diudea, *J. Chem. Inf. Comput. Sci.* **34** (1994) 1064–1071.
10. C. Y. Hu and L. Xu, *Anal. Chim. Acta* **295** (1994) 127–134.
11. I. Ugi, J. Dugundji, R. Kopp, and D. Marquading, *Lecture Notes in Chemistry*, Vol. 36, Springer, Berlin, 1984.
12. I. Ugi, J. Bauer, K. Bley, A. Dengler, A. Dietz, E. Fontain, B. Gruber, R. Herges, M. Knauer, K. Reitsam, and N. Stein, *Angew. Chemie Int. Ed. Engl.* **32** (1993) 201.

SAŽETAK

Topologija i teorija grupa-oruđa za određivanje stereokemije molekula

Stefan Reichelt, Antje Reichelt, Nicole Müller i Ivar Ugi

Pronađena je osnova za definitivnu notaciju molekula, koja uključuje i dinamička stereokemijska te konformacijska svojstva. Topološki i grupno-teorijski pristup omogućio je definitivan opis molekula i laku karakterizaciju njihovih stereokemijskih svojstava.