The Homology Relation between Molecules:
a Revival of an Old Way for Classification of Molecules

Jan Cz. Dobrowolski

Industrial Chemistry Research Institute, 01-793 Warsaw, 8 Rydygiera Street, Poland
National Medicines Institute, 00-725 Warsaw, 30/34 Chełmska Street, Poland
(*E-mail: janek@il.waw.pl, jan.dobrowolski@ichp.pl*)

Abstract. The homology (homolo) relation between molecules was introduced. This relation is a generalization of an old idea of series of homologous compounds. The homolo relation operates on a molecule by removing all the structural fragments that are identical with a certain selected fragment. As a result a multiset of fragments is produced. It was shown that the homolo relation is an equivalence relation in a set of molecules. Thus, by choosing various reference fragments, the molecules can be classified in many different ways. Using the language of homolo operation it is possible to redefine such ideas as constitutional and stereo isomers as well as a generator of a molecule and, for instance, factorization of a molecule onto fragments.

Keywords: equivalence relation, isomerism, multiset, mathematical chemistry, ordering relation, series of homologues

INTRODUCTION

Mathematical chemistry has become an expanding discipline, which, however, has been not developed uniformly.[1] Probably the most important branch of mathematical chemistry is identical with broad part of mathematical physics as chemical structure and chemical reactions have their basis in the quantum mechanics and statistical thermodynamics. Also, as experimental physics, experimental chemistry uses mostly language of linear algebra, mathematical analysis, and differential equations, as well as probability.[2,3] Much more specific for chemistry is chemical graph theory enabling for enumeration of molecules,[4–11] coding molecules,[12–17] assigning chemical graph invariants,[18–26] and correlating chemical graph invariants with diverse important molecular properties such as thermodynamic data and pharmaceutical activity.[21–34] Since early 1960s, topological chemistry has also begun to play a significant role.[35–37] Although, in mathematical chemistry molecular topology is often treated as a synonym of graph theoretical representation of a molecule, with discovery of DNA topological structures,[38–46] synthesis of molecular knots,[47–49] catenanes,[50,51] rotaxanes,[50,52–54] and Moebius strips *etc.*,[55-62] topology in chemistry approaches more central topological problems.[63,64] In this frame of reference, chemistry has been discussed in terms of sole set theory or pure algebra in much less extent.[10,65–73] In this study we just try to introduce some properties of molecules referring only to simple set theory language.

In this study we introduce a homology (homolo) relation between molecules which is a generalization of an old concept of homologous molecules. An elementary question: What is a molecule?[74–77] is transformed into more practical problem: How the molecule can be represented?,[78,79] rather than to be answered unequivocally. However, a unique representation is not applied throughout the study. Instead, it is pointed out that using one representation a molecular property can be defined whereas using the other can be not.

The homology operation acts by cutting out from molecule all the fragments identical with a defined one and gather all the remaining fragments into one multiset. By cutting out all the CH_2 groups from organic molecules, for example aliphatic alcohols, each series of homologous molecules form a well-defined subset of molecules. In other words, the homolo operation introduces the equivalence relation between the molecules and enables one to classify the molecules. Moreover, thanks to inclusion of multisets a partial order can be introduced into the set of molecules. After showing the

homolo operation to be a linear operation one can introduce a composition of the homolo operations and next factorization of the molecules. On the other hand, the homolo relation can be truncated to fragments of a given molecule and such terms as conjugation of the molecular fragments and generator of a molecule can be defined. Using homolo relation with respect to two different representations of molecules some kinds of isomers can be defined.

This is the first paper on the homolo operation and, as such, it is a mere introductory and the presented ideas are not exhaustively explored. The idea of homolo operation can be further explored in some new directions; however, classification of the chemical reactions, seems to be the most promising. The homolo operation can be also useful in molecular databases search.

Introductory Note on Representation of Molecule

Let us assume that it exists *M which is a set of all molecules* $m \in M$. By *m* we shall understand any molecule that has ever been or will be studied; atoms; crystals; molecular fragments as well as ions and radicals; and theoretical molecules included. Assume also, that *an empty molecule* $\in$ *is an element of M*. In practice, chemists deal with a not-bijective representation R of the set *M* assigning particular properties to each molecule rather than the set *M* itself:

$$R : M \rightarrow M_R \;\; ; \;\; R(m) \rightarrow m_R$$

The empirical formula; molecular formula; any structural formula; stick-and-ball model; points in Euclidean 3D space labelled with atomic symbols and with (or without) specified connectivity between the points; as well as solutions of the Schrödinger equation or their approximations at different computational levels; are examples of such representations. Moreover, molecules are represented by chemical graphs; matrices and polynomials as well. Also, the ^{13}C NMR; mid-IR; and X-ray spectra are certain representations of molecules. In true, definition of a particular representation of the set *M* can be a complex task. Further, a specific property of the representation, rather than a precise definition of the entire representation, will be referred to.

Let us recall that the *empirical formula EF* of a chemical compound is a simple expression of the relative number of each type of atom in it.

Let us agree that by *letter formula LF* representation we shall understand the common chemical letter notation of the compounds revealing some basic structural properties, *i.e.*, it showing kinds of atoms in a molecule and the closest surroundings of the most important atoms as in letter formula of alanine, glycolic acid, or nitric acid: $CH_3CH(NH_2)COOH$, $HOCH_2C(=O)OH$, $HONO_2$, respectively. Hereafter, the LF representation serves only as a convenient abbreviation for the structural formula representation.

Let us also agree that by *structural formula SF* representation we shall understand the common chemical structural notation without specifying details of the notation.

Finally, as in mathematical chemistry the chemical graph theory has been intensively developed,[4–34,80–83] the *graph theoretical GT* representation of molecules should be mentioned to be used in a standard way.

Below we shall understand that the representation R is chosen as GT or SF and, if this will not lead to misunderstandings, we shall skip the index R in symbols M_R and m_R.

THE HOMOLOGY RELATION BETWEEN MOLECULES

All the set theory terms used hereafter can be found in most monographs[84] and for the multiset properties used here see Refs. 85–88.

A multiset, or set with repetition, is a set in which multiple occurrence of its (indistinguishable) elements is allowed. For example the following sets are different multisets:

$$[\![a, b, c]\!]; [\![a, b, a, c, a, b]\!] =$$
$$[\![a, a, a, b, b, c]\!]; [\![b, a, a, b, c, c, c, c]\!]$$

We shall use the following definition of multiset.

Definition 1. A multiset **A** is a Cartesian product of a set A and a set of repetition numbers N_R:

$$\mathbf{A} = A \times R = [\![(a, r_a): a \in A \text{ and } r_a \in N_R]\!]$$

and A is then said to be base A^b of the multiset **A**. To distinguish the multisets from the ordinary sets we use the double line brackets $[\![\;]\!]$ instead of the braces $\{\}$. Notice, that the Definition 1. allows N_R to be set of any numbers, however, in most cases, N_R is here the set of non-negative integers *N*.

Definition 2. By homology (*homolo*) operation with respect to the connected molecular fragment *f* we understand the operation (hom*f*) assigning to the molecule $m \in M$ a multiset F_f^m of molecular fragments remaining after removal of all fragments *f* from the molecule *m*:
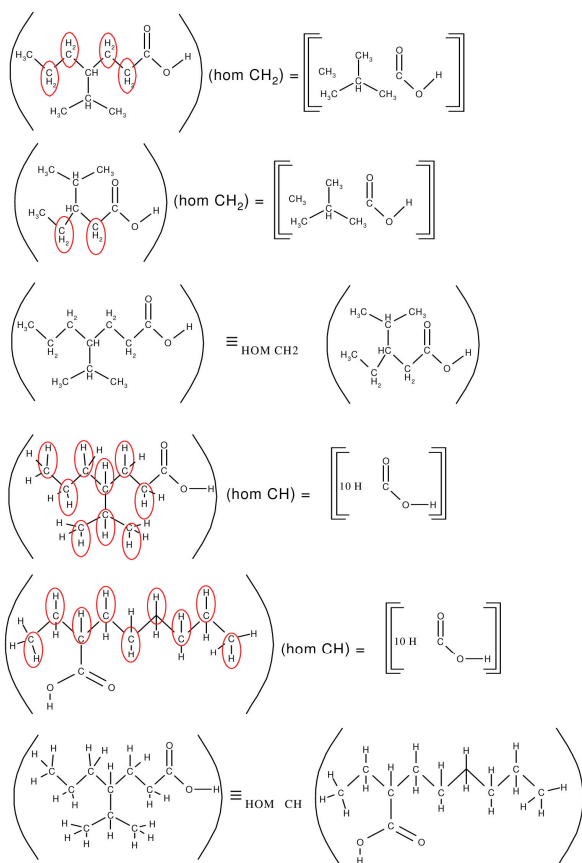
$$m (\text{hom} f) = [\![f_1, f_2, ..., f_n]\!] = F_f^m \text{ , where}$$
$$f_i \text{ may be equal to } f_j$$

$$(\text{hom} f) : M \rightarrow \mathscr{F}_f = \{ \; F_f^m \; \} = \{ [\![f_1^m, f_2^m, ..., f_n^m]\!]_f \}$$
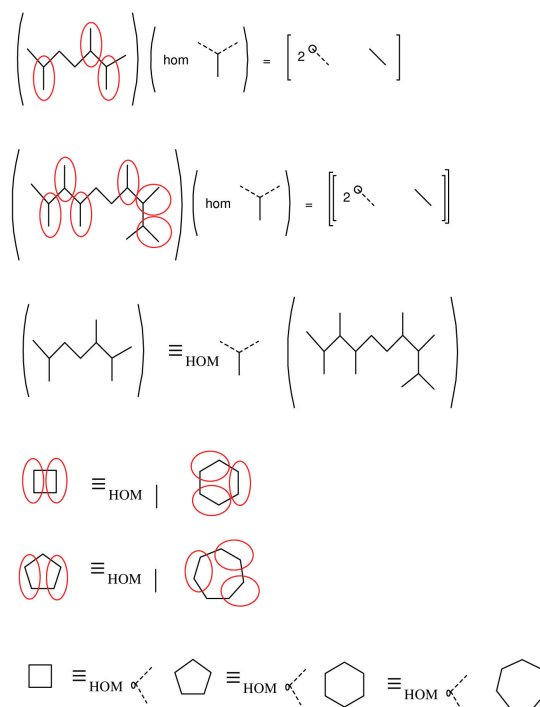
and if $m (\text{hom} f) = F_f^m \neq [\![m]\!]$ we say that *f divides the molecule m* and *f* as well as $f_i^m \in F_f^m$ constitute the molecule *m*.

Examples:

(0)   For illustrations referring to SF and GT representations see Schemes 1 and 2.

(1) $CH_3CH_2CH_3(\hom H) = [\![C–C–C]\!]$

(2) $CH_3CH_2CH_3(\hom C) =$
        $[\![H, H, H, H, H, H, H, H]\!] = [\![8H]\!]$

(3) $CH_3CH_2CH_3(\hom CH) = [\![5H]\!]$

(4) $CH_3CH_2CH_3(\hom CH_2) = [\![2H]\!]$

(5) $CH_3CH_2CH_3(\hom CH_3) = [\![CH_2]\!]$

(6) $CH_3CH_2CH_3(\hom CH_3CH_2CH_3) = \varnothing$

(7) $CH_3CH_2CH_3(\hom CH_4) = [\![CH_3CH_2CH_3]\!]$

(8) $CH_3CH_2CH_3(\hom e) = [\![CH_3CH_2CH_3]\!]$

(9) $C_6H_{12}(\hom CH_2) = \varnothing$

(10) $CH_3CH_2CHF_2(\hom CH_2) = [\![H, CHF_2]\!]$

Scheme 1. Illustrative examples how the homolo operation acts on molecules represented by structural formulas and what is understand by the structural formula homolo equivalence.

Scheme 2. Illustrative examples how the homolo operation acts on molecules represented by graphs and what is understand by the graph homolo equivalence.
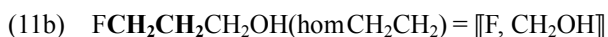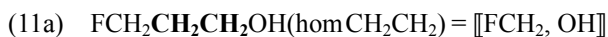
Also:

(i) $m(\hom m) = \varnothing$

(ii) $e(\hom m) = \varnothing$

(iii) $m(\hom e) = [\![m]\!]$

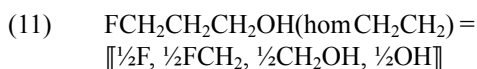(iv)   If $f \nsubseteq m$ then $m(\hom f) = [\![m]\!]$

where e denotes the empty molecule, $\varnothing$ stands for the empty set, and all the structural formula are used in their most common sense.

It is important, however, how we technically define removal of the fragment. Consider two algorithms of realization of the homolo operation. In the first one, a fragment f is recognized in a molecule and immediately cut. Then, the algorithm searches for a next fragment f in each of the fragments resulted from the first cut. In the second, initially all the fragments f are recognized in the molecule and covered by a net of overlapping fragments f, and next, all the covered fragments are removed together.

Let us have a closer look at the first algorithm. The example below and hom 1 operation in Scheme 3 show that by performing the homolo operation stepwise it happens that the removal can be done in more then one way and leads to numerous different sets of fragments:

(11a) $FCH_2\mathbf{CH_2CH_2}OH(\hom CH_2CH_2) = [\![FCH_2, OH]\!]$

(11b) $F\mathbf{CH_2CH_2}CH_2OH(\hom CH_2CH_2) = [\![F, CH_2OH]\!]$

The solution to equivocality of the homolo operation can be addition of all the sets of fragments. Assume that the sum ⊎ of multisets puts all the elements of the added multisets to the resulted multiset union. However, if the multisets are simply added, the repetition numbers in the multiset union do not reflect equivocality the homolo operation. Thus, one can use the rational repetition coefficients $1/n$, where n is the number of different multisets to which the molecule can be fragmentised in each step of operation. Thus, for the above example the multiset of fragments is the following:

(11) $FCH_2CH_2CH_2OH(\hom CH_2CH_2) =$
$[\![\tfrac{1}{2}F, \tfrac{1}{2}FCH_2, \tfrac{1}{2}CH_2OH, \tfrac{1}{2}OH]\!]$
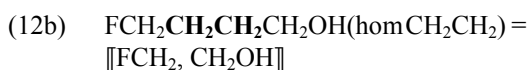
The rational repetition numbers are allowed by Definition 1. This does not disturb the multiset addition. The chemical meaning of the rational repetitions can be seen in understanding the repetition numbers as representing moles rather than sole molecules.

Even though introduction of rational repetition numbers some complications may occur. Consider the following example:

(12a) $FCH_2CH_2CH_2CH_2OH(\hom CH_2CH_2) = [\![F, OH]\!]$

but when one would like to start from the middle of the molecule, then

(12b) $FCH_2\mathbf{CH_2CH_2}CH_2OH(\hom CH_2CH_2) =$
$[\![FCH_2, CH_2OH]\!]$

Thus, the first algorithm should specify both: how to choose the first fragment to be cut, selection of fragment that should be cut next, and what to do when there is more than one such a next fragment. In other words it should define a sequence in which the cutting should be executed. This can be quite an undertaking and is closely related to basis of nomenclature and atom labeling of chemical compounds, which is, or pretend to be, unequivocal.

Therefore, here, instead of giving a precise and general rule for fragment cutting, which probably can be as complicated as chemical nomenclature, for the purposes of this paper (in which we consider only simple illustrative examples and do not intend to go into more details of any of the complex chemical structure) we propose kind of intuitive routine:

1.   to perform a first cut starting from each of the ends (including beginning and end of each of substituents)

2.   for each of the cutting sequences begun, for the next cut choose a fragment with labels the closest

to the labels of the fragment just cut (assuming labeling to be in agreement with the chemical nomenclature),
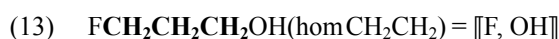
3.   and if there is more then one such a closest fragment, for example because of specific branching, the cut fragment must have the smallest possible labels including agreement concerning the order of the side chain labeling.

More precise description of the above routine, if needed, may be an objective of a further study.
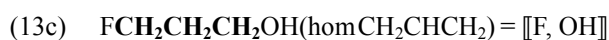
Homolo operations with respect to fragments dividing a molecule in not a univocal way can be called *multivocal* for that molecule.

Remark, however, that in case of multivocal homolo operation acting on more branched molecules determination of the repetition factors can be a very complex combinatorial task. However here, we shall not explore this problem.

Now, have a look at the second algorithm. The example below and operation hom2 in Scheme 3 show that such a definition does not lead to equivocality and multisets with rational repetition numbers do not have to be introduced.
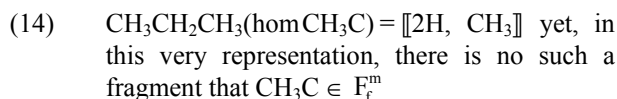
(13) $F\mathbf{CH_2CH_2CH_2}OH(\hom CH_2CH_2) = [\![F, OH]\!]$

Seemingly, the second algorithm looks more convenient. However, as a result selectivity of the homolo operation is lost out:
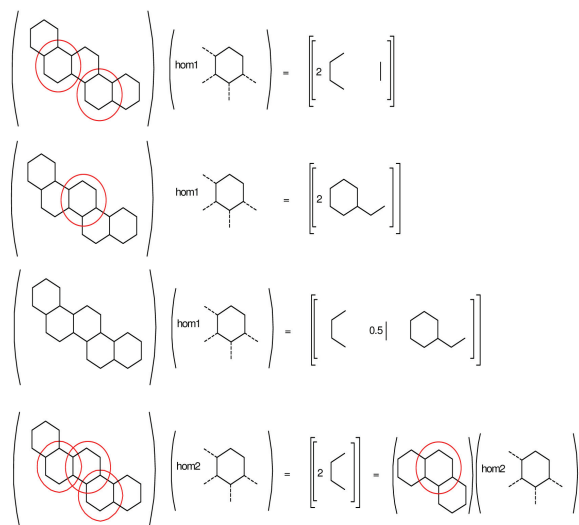
(13a) $F\mathbf{CH_2CH_2CH_2}OH(\hom CH) = [\![F, OH]\!]$

(13b) $F\mathbf{CH_2CH_2CH_2}OH(\hom CHCH_2) = [\![F, OH]\!]$

(13c) $F\mathbf{CH_2CH_2CH_2}OH(\hom CH_2CHCH_2) = [\![F, OH]\!]$
    etc.

Therefore, hereafter, we shall use the former algorithm rather than the later one.

Remark, finally that

(14) $CH_3CH_2CH_3(\hom CH_3C) = [\![2H, CH_3]\!]$ yet, in this very representation, there is no such a fragment that $CH_3C \in F_f^m$
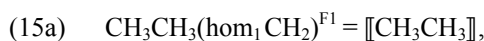
Therefore, it is important that unconnected molecular fragments are not-considered in the homolo operation on molecules. Thus, a definition of fragment connectivity within a given representation must be defined. For SF and GT representations fragment connectivity is well defined, whereas for LF and EF it is not a straightforward problem. However, for the purposes of this paper LF serves for abbreviation of the SF representation and thus the connectivity in the frame of LF needs no separate definition.

Scheme 3. Illustrations of two considered definitions of homolo operation: hom1 is based on a stepwise algorithm and hom2 is based on covering all the fragments declared in the operation and removing them all together.

The homolo operation does depend on representation R of the set of all molecules M and the way of identifying fragment f_R in a molecule m_R. For instance, the ordinary GT representation of molecules does not differentiate atoms. Therefore, the graphs of the benzene and pyridine molecules are identical and so are the results of any homolo operation on these graphs. On the other hand, within the SF representation benzene(hom CH) = $\varnothing$, whereas pyridine(hom CH) = $[\![N]\!]$.

Also, consider F1 representation of the CH_2 group in which it is recognized in a molecule when the other two bonds of the carbon atom bind it with the non-hydrogen atoms, and the other representation F2, recognizing the CH_2 group regardless the kind of the other two atoms. As a result of homolo operation with respect to the former representation:

(15a) $CH_3CH_3(hom_1 CH_2)^{F1} = [\![CH_3CH_3]\!]$,

whereas concerning the later representation we obtain a different result:

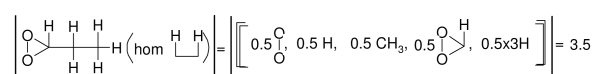(15b) $CH_3CH_3(hom_2 CH_2)^{F2} = [\![H, H]\!]$.

Instead of multiset, the ordinary set can be used for the definition of the homolo operation. In such a case, however; in comparison to the proposed multiset approach, the ability for differentiation between molecules is inferior. For example: if $\mathscr{F}_f$ is an ordinary set, then $CH_3CH_2CH_3(homCH)=\{H\}=CH_3CH=CH_2$ (homCH) and they are CH-homologous, whereas if $\mathscr{F}_f$ is the multiset $CH_3CH_2CH_3(homCH) = [\![5H]\!]$ and $CH_3CH=CH_2(homCH) = [\![4H]\!]$ and, in agreement with chemical intuition, they are not CH-homologous.

Definition 3.   By power of the homolo operation on the molecule, $|m^1(hom\, m^2)|$, we understand cardinality of the multiset of fragments, card ($\mathscr{F}_f^{\,1}$), i.e., sum of the repetition coefficients of all multiset elements.

$$|m^1(hom\, m^2)| = \text{card}\,(\mathscr{F}_f^{\,1}) = \sum k_{f,\,i}^1,$$

where $i = 1, 2, \ldots, n$ and $n = \text{card}(\mathscr{F}_f^{\,1\ b})$ *i.e.*, n is the number of elements in the base of the multiset $\mathscr{F}_f^{\,1}$.

Remark, that in the light of Definition 1, the so defined cardinality can be, *e.g.*, rational instead of a natural number. As was shown by the reviewer of this paper, this is indeed the case for some molecules and some multivocal homolo operations (Scheme 4).

Scheme 4. Exemplary molecule for which the defined multiset cardinality is not a natural number.

At the end of this section we show three facts:

Fact 1.   If f divides m^1 and g divides f then g divides m^1.

$$m^1(hom\, f) = \mathscr{F}_f^{\,m1} \neq [\![m^1]\!] \quad \text{and}$$
$$f(hom\, g) = \mathscr{F}_g^{\,f} \neq [\![f]\!] \quad \text{then}$$
$$m^1(hom\, g) = \mathscr{F}_g^{\,m1} \neq [\![m^1]\!].$$

Proof.   Suppose that $m^1(hom\, f) = \mathscr{F}_f^{\,m1} \neq [\![m^1]\!]$ and $f(hom\, g) = \mathscr{F}_g^{\,f} \neq [\![f]\!]$ but $m^1(hom\, g) = [\![m^1]\!]$. Assume, that for m^1 the (hom f) operation is unequivocal. Let us form a multiset $\mathcal{M}$ including $\mathscr{F}_f^{\,m1}$ and k-times the f fragment where k is the f fragment repetition number which is counting how many times f was removed from m^1. Thus, $\mathcal{M}$ is the multiset equivalent of the m^1 molecule i.e. m^1 is decomposed by the (hom f) operation from which the f fragments were not yet removed. Now, let us refer to a property that is formally introduced in the Definition 9 and which allows one to execute the homolo operation on multisets. Then, $\mathcal{M}(hom\, f) = \mathscr{F}_f^{\,m1}$ and $\mathcal{M}(hom\, g) \neq [\![m^1]\!]$ because, $\mathcal{M}(hom\, g) = [\![\mathscr{F}_f^{\,m1}, k \cdot f]\!]$ $(hom\, g) = [\![\mathscr{F}_f^{\,m1}(hom\, g), k \cdot f(hom\, g)]\!] \neq [\![m^1]\!]$ because $f(hom\, g) = \mathscr{F}_g^{\,f} \neq [\![f]\!]$. Thus indeed also g divides the m^1 molecule. Finally, if for m^1 the (hom f) operation is multivocal the argumentation is similar because of $k \cdot f(hom\, g)$ element of the $\mathcal{M}$ multiset.

Examples:

Let $m^1 = CH_3CH_2CH_2CH_3$ and $f = CH_2CH_2$ then $CH_3CH_2CH_2CH_3(homCH_2CH_2) = [\![2H]\!] = \mathscr{F}_f^{\,C_4H_{10}}$ and $\mathcal{M} = [\![2H, 2CH_2CH_2]\!]$. Then, $\mathcal{M}(homCH_2CH_2) = [\![2H, 2CH_2CH_2]\!](homCH_2CH_2) = \mathscr{F}_f^{\,C_4H_{10}}$. Now, let $g = CH$, then $CH_2CH_2(homCH) = [\![2H]\!]$ and $CH_3CH_2CH_2CH_3(homCH) = [\![6H]\!] \neq [\![CH_3CH_2CH_2CH_3]\!]$

Let $m^1 = CH_3CH_3$ and $f = CH_3CH_3$ then $CH_3CH_3(\text{hom } CH_3CH_3) = \varnothing = \mathscr{F}_{C_2H_6}^{C_2H_6}$ and $\mathscr{M} = [\![\varnothing, CH_3CH_3]\!]$. Then $\mathscr{M}(\text{hom } CH_3CH_3) = [\![\varnothing, CH_3CH_3]\!](\text{hom } CH_3CH_3) = \varnothing$. Now, let $g = CH_2$, then $CH_3CH_3(\text{hom } CH_2) = [\![2H]\!] \neq [\![CH_3CH_3]\!]$

$CH_3CH_2CH_3(\text{hom } CH_2) = [\![2H]\!]$ and $CH_2(\text{hom } CH) = [\![H]\!]$ then $CH_3CH_2CH_3(\text{hom } CH) = [\![5H]\!] \neq [\![CH_3CH_2\text{-}CH_3]\!]$ $CH_3CH_2CH_3(\text{hom } C_2) = [\![0.5CH_3, 2.5H]\!] \uplus [\![0.5 CH_3, 2.5H]\!] = [\![CH_3, H]\!]$ and $C_2(\text{hom } C) = \varnothing$ then $CH_3\text{-}CH_2CH_3(\text{hom } C) = [\![8H]\!] \neq [\![CH_3CH_2CH_3]\!]$

Fact 2. The homolo operation on a set of molecules M constitutes an equivalence relation $\equiv_{\text{hom } f}$ between the molecules $m^1, m^2 \in M$.

Proof. Indeed, the $\equiv_{\text{hom } f}$ relation is (i) reflexive, (ii) symmetric, and (iii) transitive:

(i) $m \equiv_{\text{hom } f} m$

           $m(\text{hom} f) = m(\text{hom} f)$

(ii) $m^1 \equiv_{\text{hom } f} m^2$ then $m^2 \equiv_{\text{hom } f} m^1$

           $m^1(\text{hom} f) = m^2(\text{hom} f)$ then $m^2(\text{hom} f) = m^1(\text{hom} f)$

(iii) $m^1 \equiv_{\text{hom } f} m^2$ and $m^2 \equiv_{\text{hom } f} m^3$ then $m^1 \equiv_{\text{hom } f} m^3$

           $m^1(\text{hom} f) = m^2(\text{hom} f)$ and $m^2(\text{hom} f) = m^3(\text{hom} f)$ then $m^1(\text{hom} f) = m^3(\text{hom} f)$
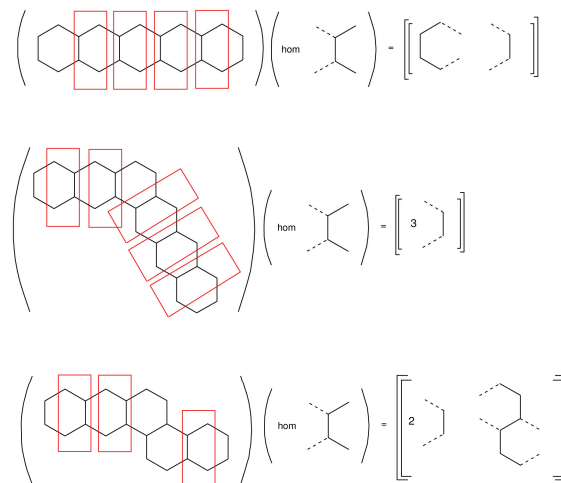
As any equivalence relation, the homolo relation splits the set of molecules M into cosets of equivalent molecules - equivalence classes, *i.e.*, sets of homologous molecules. The molecules m^1 and $m^2 \in M$ are said to be homologous with respect to the fragment f, m^1, $m^2 \in H^f \subset M$ iff $m^1 \equiv_{\text{hom } f} m^2$. In such a case we say that m^1 and m^2 are *f-homologous*.

If two molecules m^1 and m^2 are not *f*-homologous they are called *f-dishomologues*. If there is not such a fragment f that two molecules m^1 and m^2 are *f*-homologous they are called *strictly dishomologues* or, by analogy to coprime numbers, *coatomic*. Two molecules (fragments) are said to be coprime if $m^1(\text{hom} m^2) = [\![m^1]\!]$ and $m^2(\text{hom} m^1) = [\![m^2]\!]$.

Fact 3. If the powers of the homolo operations on two molecules m^1 and $m^2 \in M$ with respect to the molecular fragment f are different then the molecules are *f*-dishomologues.

Remark finally, that if the homolo operation is taken with respect to the CH_2 moiety, it defines the ordinary homology relation between the a *séries homologues* of the aliphatic compounds introduced by Gerhardt in 1844.[89] The aliphatic molecules are said to be homologous when they differ by a natural number of the CH_2 groups.

In Scheme 5, the series of *n*-cene, *n*-fene, and chrysene homologues are shown to be defined in terms of abstraction classes of the homolo relation with respect to the CC–CC fragment.

Scheme 5. *n*-cene, *n*-fene, and chrysene homologues can be defined in terms of abstraction classes of the homolo relation with respect to the CC-CC fragment.

THE MOLECULES ARE PARTIALLY HOMOLO ORDERED

A partial order can be introduced to the set of molecules M taking the advantage from inclusions of multisets.

Definition 4. If a pair (a, n_a) is an element of the multiset **A** then a pair (a, k_a) is also an element of **A** iff $k_a \leq n_a$.

$\forall (a, n_a) \in \mathbf{A}$ iff $(a, k_a): a \in \mathbf{A}^b$ and $k_a \leq n_a$ then $(a, k_a) \in \mathbf{A}$

Remark:

$(a, k_a) \notin \mathbf{A}$ iff $a \notin \mathbf{A}^b$ or $k_a > n_a$

Definition 5. Two elements $(a, n_a) \in \mathbf{A}$ and $(b, n_b) \in \mathbf{B}$ of two multisets **A**, **B** are equal iff $(a, n_a) = (b, n_b)$ iff $a = b$ and $n_a = n_b$

Definition 6. The multiset **B** is included in a multiset **A**, $\mathbf{B} \subset \mathbf{A}$, iff for any element $(a, k_a) \in \mathbf{B}$ $a \in \mathbf{B}^b$ exists an element $(a, n_a) \in \mathbf{A}$ such that $a \in \mathbf{A}^b$ and $k_a \leq n_a$

$\mathbf{B} \subset \mathbf{A}$ iff $\forall (a, k_a) \in \mathbf{B}$, $a \in \mathbf{B}^b \exists (a, n_a) \in \mathbf{A}$, $a \in \mathbf{A}^b$: $a \in \mathbf{A}^b$ and $k_a \leq n_a$

Definition 7. The molecule $m^1 \in M$ is said to be homolo greater than or equal to the molecule $m^2 \in M$ with respect to the fragment f:

$m^1(\text{hom} f) \geq_{\text{hom } f} m^2(\text{hom} f)$

if the multiset of fragments F_f^2 is included in multiset of fragments F_f^1, *i.e.* $F_f^2 \subset F_f^1$.

Fact 4. The $\geq_{\hom f}$ relation is (i) reflexive, (ii) transitive, and (iii) antisymmetric which is a simple consequence of multiset inclusion properties.

Proof.

(i) $m \geq_{\hom f} m$
         $m(\hom f) \geq_{\hom f} m(\hom f)$

(ii) $m^1 \geq_{\hom f} m^2$ and $m^2 \geq_{\hom f} m^3$ then $m^1 \geq_{\hom f} m^3$
         $m^1(\hom f) \geq_{\hom f} m^2(\hom f)$ and
         $m^2(\hom f) \geq_{\hom f} m^3(\hom f)$ then
         $m^1(\hom f) \geq_{\hom f} m^3(\hom f)$

(iii) $m^1(\hom f) \geq_{\hom f} m^2(\hom f)$ and
         $m^2(\hom f) \geq_{\hom f} m^1(\hom f)$ then $m^1 =_{\hom f} m^2$
         $m^1(\hom f) \geq_{\hom f} m^2(\hom f)$ and
         $m^2(\hom f) \geq_{\hom f} m^3(\hom f)$ then
         $m^1(\hom f) =_{\hom f} m^2(\hom f)$

The following examples illustrate properties of the $\geq_{\hom f}$ relation:

1. $CH_3CH(CH_3)CH_3 \geq_{\hom CH_2} CH_3(CH_2)_2CH_3$
     because $[\![H, CH, H, H]\!] \supset [\![H, H]\!]$

2. $CH_3CH(CH_3)CH_3 =_{\hom CH} CH_3(CH_2)_2CH_3$
     because $[\![6H]\!] = [\![6H]\!]$

3. $CH_3CH(CH_3)CH_3 \not\geq_{\hom CH_3} CH_3(CH_2)_2CH_3$ and
     $CH_3CH(CH_3)CH_3 \not\leq_{\hom CH_3} CH_3(CH_2)_2CH_3$
     because $[\![CH]\!] \not\subset [\![(CH_2)_2]\!]$ and
     $[\![CH]\!] \not\supset [\![(CH_2)_2]\!]$

The molecules which cannot be compared by the *f* non-strict homolo inequality, as in example 3., are said to be incomparable homolo *f*.

In conclusion of these examples, one can state that the same molecule m^1 can simultaneously be greater than the other m^2 with respect to one fragment, equal to m^2 with respect to the second fragment, and incomparable with m^2 with respect to the third one. Thus, each fragment *f* orders the set of molecules *M* differently.

The set of molecules *M* has no the greatest molecule homolo *f* because for each molecule yielding a given multiset of fragments it exists another molecule yielding a greater multiset of fragments. However, the empty molecule *e* is the smallest element.

In a set of molecules ordered (homolo *f*) there are subsets which are *chains* (every two elements of the subset are comparable) as well as *antichains* (every two elements of the subset are incomparable). The set of alkanes (homolo C) is a chain whereas (homolo CH_3) is an antichain.

The ordered set is called *directed* if for every pair of its elements it exist another element that is greater then or equal to these two elements. The question whether the set of molecules *M* ordered homolo *f* is directed or not is ques-

tion of the basic assumptions made and choice of a representation. In the GT representation, in which the elements loose their identities, the M_{GT} is ordered because it always exists a graph which is greater (homolo *f*) than the other two graphs. On the other hand, it seems that in the SF representation, every nonsense molecule can be written and so it would always exists another nonsense molecule that is greater (homolo *f*) then the former one. However, we believe that most of chemists will agree that only selected subsets of M_{SF} are directed whereas M_{SF} itself is not.

LINEARITY OF THE HOMOLO OPERATION

Let $\uplus$ be the addition of multisets which can be illustrated as follows:

$$[\![a,a,b,b,c,c,c]\!] \uplus [\![b,b,b,a]\!] = [\![a,a,a,b,b,b,b,c,c,c]\!].$$

Definition 8.   Let $m^1(\hom f) = [\![f_1^1, f_2^1, ..., f_{n_1}^1]\!]$ and $m^2(\hom f) = [\![f_1^2, f_2^2, ..., f_{n_2}^2]\!]$, then by taking the advantage of the $\uplus$ multiset addition the homolo operation can be defined on *sum of two molecules* and can be defined as the homolo-addition of two molecules:

$$(m^1 + m^2)(\hom f) = m^1(\hom f) + m^2(\hom f) =$$
$$[\![f_1^1, f_2^1, ..., f_{n_1}^1]\!] \uplus [\![f_1^2, f_2^2, ..., f_{n_2}^2]\!] =$$
$$[\![f_1^1, f_2^1, ..., f_{n_1}^1, f_1^2, f_2^2, ..., f_{n_2}^2]\!] = m^1 +_{\hom} m^2.$$

Fact 5. As a simple consequence of the $\uplus$ multiset addition operation

(i) $(m^1 + m^2)(\hom f) = (m^2 + m^1)(\hom f)$

        $m^1 +_{\hom} m^2 = m^2 +_{\hom} m^1$

(ii) $((m^1 + m^2) + m^3)(\hom f) =$
         $(m^1 + (m^2 + m^3))(\hom f)$
         $(m^1 +_{\hom} m^2) +_{\hom} m^3 = m^1 +_{\hom} (m^2 +_{\hom} m^3)$

(iii) $(m^1 + e)(\hom f) = (e + m^1)(\hom f) = m^1(\hom f)$,
         where e is the empty molecule,
         $m^1 +_{\hom} e = e +_{\hom} m^1 = m^1(\hom f)$
         as well as

(iv) $(k \cdot m^1)(\hom f) = k \cdot (m^1)(\hom f)$,
         where *k* is a natural number.

The properties (i) and (ii) mean that homolo addition of molecules is commutative and associative and the property (iii) that there exists a unit element - the empty molecule. The properties (i) and (iv) say that the homolo operation is linear.

Remark. Definition of the addition of two molecules facilitates discernment of the homolo operation on certain classes of topological molecules as catenanes and rotaxanes.

The homolo addition can be also introduced in a different way.

Definition 9. Let us now allow the homolo operation to act not only on molecules but also on multisets containing the molecules and fragments in the same way as it does on molecules:

$$[\![m]\!](\mathrm{hom}f) \equiv [\![m(\mathrm{hom}f)]\!] \equiv$$
$$[\![[\![f_1^m, f_2^m, ..., f_n^m]\!]_f]\!] \equiv [\![f_1^m, f_2^m, ..., f_n^m]\!]_f \equiv$$
$$m(\mathrm{hom}f)$$

i.e., that homolo operation acting on multiset with molecule does not only generate fragments but also it reduces the external brackets making from multiset of multiset(s) the simple multiset containing all the fragments. The homolo operation on sum of molecules introduced by this way is linear as well.

Fact 6. When the homolo operation on multiset is defined it is possible to compose the homolo operations.

Indeed, let f and g be two fragments and $m(\mathrm{hom}f) = [\![f_1, f_2, ..., f_n]\!]$, and $f_i(\mathrm{hom}g) = [\![g_1^i, g_2^i, ..., g_{k_1}^i]\!]$ then

$$[m(\mathrm{hom}f)](\mathrm{hom}g) = [\![[\![g_1^1, g_2^1, ..., g_{k_1}^1]\!],$$
$$[\![g_1^i, g_2^i, ..., g_{k_2}^i]\!], ..., [\![g_1^n, g_2^n, ..., g_{k_n}^n]\!]]\!] =$$

$$[\![g_1^1, g_2^1, ..., g_{k_1}^1]\!] \uplus [\![g_1^i, g_2^i, ..., g_{k_2}^i]\!] \uplus, ..., \uplus$$

$$[\![g_1^n, g_2^n, ..., g_{k_n}^n]\!] = [\![g_1^1, g_2^1, ..., g_{k_1}^1,$$
$$g_1^i, g_2^i, ..., g_{k_2}^i, ..., g_1^n, g_2^n, ..., g_{k_n}^n]\!]$$

Fact 7. Composition of the homolo operation on molecule m, with respect to the same fragment *f, is idempotent.*

If $m(\mathrm{hom}f) = [\![f_1, f_2, ..., f_n]\!]$, then

$$[m(\mathrm{hom}f)](\mathrm{hom}f) = [\![f_1, f_2, ..., f_n]\!] \text{ and}$$

$$m(\mathrm{hom}f)^n = m(\mathrm{hom}f).$$

HOMOLO CONJUGATION AND HOMOLO PARTIAL ORDER BETWEEN FRAGMENTS OF A MOLECULE

Definition 10. Two fragments $f^1, f^2 \subset m$ are said to be *homolo conjugated* in molecule m:

$$f^1 \sim_{\mathrm{H,m}} f^2 \text{ iff } m(\mathrm{hom}f^1) = m(\mathrm{hom}f^2).$$

Fact 8. As the homolo operation on molecule, the homolo conjugation of fragments in molecule m is an equivalence relation between the fragments because it is (i) reflexive, (ii) symmetric, and (iii) transitive:

(i) $\qquad f^1 \sim_{\mathrm{H,m}} f^1$

(ii) $\qquad$ if $f^1 -_{\mathrm{H,m}} f^2$ then $f^2 -_{\mathrm{H,m}} f^1$

(iii) $\qquad$ if $f^1 -_{\mathrm{H,m}} f^2$ and $f^2 -_{\mathrm{H,m}} f^3$ then $f^1 -_{\mathrm{H,m}} f^3$

The homolo conjugation relation $\sim_{\mathrm{H}}$ is in fact the homolo relation truncated to the fragments of molecule m:

$$\sim_{\mathrm{H,m}} = \equiv_{\mathrm{hom f}} *_{m\mathrm{R}}$$

which splits the set of fragments of the molecule m into cosets of equivalent fragments.

Example:

p-$C_6H_4Cl_2(\mathrm{hom}CH) = [\![2CCl]\!] = $ p-$C_6H_4Cl_2(\mathrm{hom}C_2H_2)$

Remark that the same fragments can be conjugated in one molecule whereas not conjugated in the other:

$C_6H_{12}(\mathrm{hom}CH_2) = [\![\varnothing]\!] = C_6H_{12}(\mathrm{hom}CH_2CH_2)$
thus $CH_2 \sim_{\mathrm{H,C_6H_{12}}} C_2H_4$

$CH_3CH_2CH_3(\mathrm{hom}CH_2) = [\![2H]\!] \neq$
$CH_3CH_2CH_3(\mathrm{hom}CH_2CH_2) = [\![H,CH_3]\!]$ thus
$CH_2 \nsim_{\mathrm{H,C_3H_8}} C_2H_4$.

Fact 9. The fragments f of molecule m are partially homolo ordered by the $\geq_{\mathrm{hom f}}$ non-strict inequality with respect to the chosen fragment f.

Indeed, the homolo relation $\geq_{\mathrm{hom f}}$ on molecules is (i) reflexive, (ii) transitive, and (iii) antisymmetric:

(i) $\qquad f \geq_{\mathrm{hom f}} f$

(ii) $\qquad f^1 \geq_{\mathrm{hom f}} f^2$ and $f^2 \geq_{\mathrm{hom f}} f^3$ then $f^1 \geq_{\mathrm{hom f}} f^3$

(iii) $\qquad f^1 \geq_{\mathrm{hom f}} f^2$ and $f^2 \geq_{\mathrm{hom f}} f^2$ then $f^1 \sim_{\mathrm{H,m}} f^2$

The molecule m is the greatest fragment of the $\geq_{\mathrm{hom f}}$ relation on the molecule m and the empty molecule e is the smallest element.

GENERATOR OF A MOLECULE

Definition 11. Fragment $g \subset m$ is said to be a *generator of the molecule m* iff $m(\mathrm{hom}g) = \varnothing$.

For each molecule, the molecule itself is a trivial generator. A generator is said to be non-trivial iff $g \subsetneq m$.

Once again, depend on representation; the number of molecular generators varies. For instance, if the atoms are represented as balls, and the C-atom in the CH_2 fragment is represented as half of the ball then $CH_4(\mathrm{hom}CH_2) = \varnothing$ whereas if the C-atom is represented by entire ball then $CH_4(\mathrm{hom}CH_2) = [\![H, H]\!]$.

The molecule m for which a generator g exists is called to be a power of the generator g, and if $(g)_n$ is also the generator of the molecule m (where $(g)_n$ stands for the n-times connected g fragment) then the molecule is said to be the nth power of the generator g.

if $m (\mathrm{hom}(g)_n) = \varnothing$ then $m = (g)^n$

The greatest n is called the power of the generator. The non-trivial generator which is the smallest, according to the $\geq_{\mathrm{hom g}}$ inequality between the generators, is called the *prime generator* of the molecule m and its powers are called the *composite generators*.

For example, CH_2, CH_2CH_2, and $CH_2CH_2CH_2$ are generators of the cyclohexane molecule, which are the 1st, 2nd, and 3rd power of the CH_2 prime generator. The CH_3NO_2 has only trivial generator, whereas in Scheme 6 two different prime generators of hexafluorobenzene are illustrated. The coronene molecule has at least two prime generators (Scheme 7).
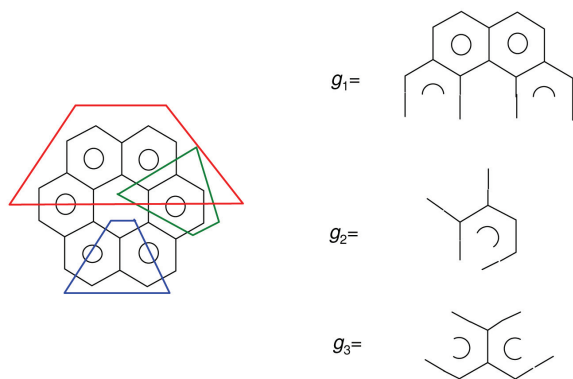
Molecule that has k non-trivial prime generators is called *k-generable* otherwise the molecule is *ingenerable or 0-generable*.

Scheme 6. Two prime generators of the hexafluorobenzene molecule.

If the molecule is generated by a single elemental fragment then it is called the *elemental molecule*. Whether the elemental molecule is generated by an element or not depends on the molecule structure and representation applied.
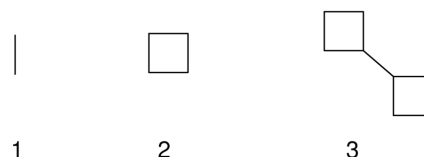

Scheme 7. Three generators of the coronene molecule. Notice that $g_1 = (g_2)^2$.

Schemes 6 and 7 show also that although different generators are conjugated may not have a common generator.

Also, one might erroneously suppose that if generator g^1 of molecule m^1 is also generator of generator g^2 of another molecule m^2 than g^1 is the generator of molecule m^2. However, as was shown by reviewer of this paper, in GT representation of acetylene, cyclobutadiene, and

diradical of bicyclobutyl-diene, 1, 2, and 3, respectively (Scheme 8), 1 is generator of 2 and 2 is generator of 3, but 1 is not a generator of 3. However in the LT representation of these molecules, $(CH)_2$, $(CH)_4$, and $(CH)_8$, respectively, it is not the case.

Scheme 8. Illustration of the fact that 1 is generator of 2 and 2 is generator of 3 but 1 is not a generator of 3.

ISOMERIC MOLECULES

Definition 12. Two molecules m_R^1 and m_R^2 are said to be *indistinguishable* according to the representation R (m_R^1 and m_R^2 are *R-identical*)

$$m_R^1 \equiv_R m_R^2 \text{ iff } m_R^1 (\text{hom } m_R^2) = m_R^2 (\text{hom } m_R^1) = \varnothing,$$

otherwise they are *R-different*.

Fact 10. Two molecules m_R^1 and m_R^2 are R-identical iff for any fragment f_R the molecules m_R^1 and m_R^2 are homolo f_R equivalent:

$$m_R^1 \equiv_R m_R^2 \text{ iff } \forall f_R \ m_R^1 \equiv_{\text{hom f}} m_R^2$$

Proof. Suppose that $m_R^1 \equiv_R m_R^2$ that means m_R^1 (hom m_R^2) = m_R^2 (hom m_R^1) = $\varnothing$. Then from Fact 1 we know that for any f_R if m_R^2 (hom f_R) $\neq [\![m_R^2]\!]$ then f_R divides also what is divided by m_R^2: m_R^1 (hom f_R) $\neq [\![m_R^1]\!]$. The same holds true for the second molecule. Thus, m_R^1 and m_R^2 are divided by the same fragments and there is no fragment that divides one and does not divide the other. So, the implication from left to right is proven.

Now, suppose that for any f_R m_R^1 (hom f_R) = m_R^2 (hom f_R). Thus also m_R^1 (hom m_R^1) = m_R^2 (hom m_R^1) and m_R^2 (hom m_R^2) = m_R^1 (hom m_R^2). But, because molecule homolo divided by itself generates the empty set, in the two cases the operation leads to the empty set: m_R^2 (hom m_R^1) = $\varnothing$ and m_R^2 (hom m_R^1) = $\varnothing$. So, indeed this implies the left-hand side of the Fact, *i.e.*, m_R^1 (hom m_R^2) = m_R^2 (hom m_R^1) = $\varnothing$.

Definition 13. Two molecules m^1 and m^2 are said to be *R-isomeric*:

$$m^1 \equiv_{\text{iso}} m^2$$

iff (i) they are identical according to empirical formula representation, EF-identical, and (ii) there exist a representation R in which they are not homolo equal with respect to whole molecules m^1 and m^2:

(i) $m_{EF}^1 (\hom m_{EF}^2) = m_{EF}^2 (\hom m_{EF}^1) = \varnothing$ and

(ii) $\exists R$ such that $m_R^1 (\hom m_R^2) \neq m_R^2 (\hom m_R^1)$

The empirical formulae of the benzene and acetylene molecules explain why the sole $m_{EF}^1 (\hom m_{EF}^2) = \varnothing$ condition is not sufficient, and the existence of a representation R is necessary for m^1 and m^2 to not be the same molecule.

The single EF representation does not enable distinguishing of isomers because another representation must be known to confirm the molecules have a different property.[90] Usually, for the empirical formula it is possible, however, to find more then just one homolo relation. Thus, the homolo equivalence seems to be more fundamental than the isomeric equivalence: it allows one to judge about mathematical structure of chemical species and to define isomerism.

Fact 11. Two molecules m_R^1 and m_R^2 are R-isomeric if they are EF-identical and it exists f_R that they are not equivalent homolo f_R.

Proof. If for any f_R the m_R^1 and m_R^2 molecules would be equivalent homolo f_R, $m_R^1 (\hom f_R) = m_R^2 (\hom f_R)$, then $m_R^1 (\hom m_R^1) = m_R^2 (\hom m_R^1)$ and $m_R^1 (\hom m_R^2) = m_R^2 (\hom m_R^2)$. Because, $m_R^1 (\hom m_R^1) = \varnothing$ and $m_R^2 (\hom m_R^2) = \varnothing$, then $m_R^2 (\hom m_R^1) = m_R^1 (\hom m_R^2)$ which contradicts the second point of the definition of R-isomeric molecules which says that

$$m_R^1 (\hom m_R^2) \neq m_R^2 (\hom m_R^1).$$

Definition 14. Let m^1 and m^2 be EF-identical and R-different there exist a fragment $f_R \neq m_R^1, m_R^2$ for which $m_R^1 (\hom f_R) = [\![m_R^1]\!]$ and $m_R^2 (\hom f_R) \neq [\![m_R^2]\!]$ or oppositely then the molecules m_R^1 and m_R^2 are said to be *pair of R-constitutional (structural) isomers*:
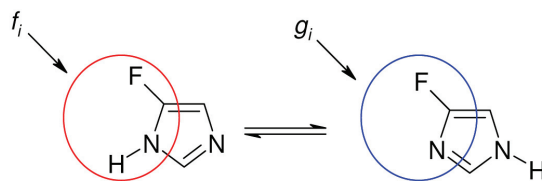
$$m_R^1 \equiv_{c\text{-iso } R} m_R^2$$

and f_R is called *constitutional difference*.

Example. $CH_3CH(CH_3)_2$ and $CH_3(CH_2)_2CH_3$ are EF-identical and $(CH_2)_2$ is the constitutional difference $CH_3CH(CH_3)_2 (\hom (CH_2)_2) = [\![CH_3CH(CH_3)_2]\!]$ and $CH_3(CH_2)_2CH_3 (\hom (CH_2)_2) = [\![1CH_3, 5\ H, 5\ CH_2CH_3]\!]$ (the $(CH_2)_2$ fragment divides both CH_3CH_2 and CH_2CH_2 moiety) whereas CH is not a constitutional difference $CH_3CH(CH_3)_2 (\hom CH) = [\![6H]\!]$ and $CH_3(CH_2)_2CH_3 (\hom CH_3) = [\![6H]\!]$. Remark also that there are fragments for which the molecules are not homologous which may not necessarily reveal the basic structural difference between isomers however in the following example $CH_3CH(CH_3)_2 (\hom CH_3) = [\![CH]\!]$ and $CH_3(CH_2)_2CH_3 (\hom CH_3) = [\![CH_2CH_2]\!]$ the CH_3 fragment reveals the difference quite well. See also an example of constitutional differences in Scheme 9.

Fact 12. Let $m_R^1 \equiv_{c\text{-iso } R} m_R^2$ and f_R be the constitutional

difference between a pair of R-constitutional isomers for which $m_R^1 (\hom f_R) \neq [\![m_R^1]\!]$ then there exist a fragment g_R that satisfies $m_R^2 (\hom g_R) \neq [\![m_R^2]\!]$ and $m_R^1 (\hom g_R) = [\![m_R^1]\!]$.

Scheme 9. An example of constitutional differences between two constitutional isomers: 5-fluoro-1*H*-imidazole and 4-fluoro-1*H*-imidazole.

Proof. Let m_R^1 and m_R^2 be the R-constitutional isomers and the not-empty fragment f_R be the constitutional difference between them that satisfies $m_R^1 (\hom f_R) = [\![m_R^1]\!]$ and $m_R^2 (\hom f_R) \neq [\![m_R^2]\!]$. Assume, that there is not such a (not-empty) fragment g_R that $m_R^2 (\hom g_R) \neq [\![m_R^2]\!]$ and $m_R^1 (\hom g_R) = [\![m_R^1]\!]$. Thus for every g_R it is not true that $m_R^2 (\hom g_R) \neq [\![m_R^2]\!]$ and $m_R^1 (\hom g_R) = [\![m_R^1]\!]$. So, for every g_R if $m_R^2 (\hom g_R) \neq [\![m_R^2]\!]$ then $m_R^1 (\hom g_R) \neq [\![m_R^1]\!]$. This means that also $m_R^1 (\hom m_R^2) \neq [\![m_R^1]\!]$. Because, $m_{EF}^1 \equiv m_{EF}^2$ then $m_R^1 (\hom m_R^2) = \varnothing$. But if $m_R^1 (\hom m_R^2) = \varnothing$ and $m_{EF}^1 \equiv m_{EF}^2$ then also $m_R^2 (\hom m_R^1) = \varnothing$ which means that the $m_R^1 \equiv m_R^2$. However, this contradicts the assumption that they are R-constitutional isomers.

Definition 15. Let m_R^1 and m_R^2 be R-constitutional isomers and $f_i \subsetneq m_R^1$, $i = 1,...,n_1$ and $g_j \subsetneq m_R^2$, $j = 1,...,n_2$ be the constitutional differences between m_R^1 and m_R^2. Some of the constitutional differences f_i may be conjugated constitutional differences f_j, $i \neq j$, and some of them may be not-conjugated. The same holds true for g_j. From each series of conjugated constitutional differences let us choose prime generators f_i^0. Then set of non-conjugated constitutional differences $[\![f_i^0]\!]$ $i = 1, ..., k_1$ is called the m_R^1-molecule basis of the constitutional isomerism between the molecules m_R^1 and m_R^2, and similarly $[\![g_i^0]\!]$ $i = 1, ..., k_2$ is called the m_R^2-molecule basis of the constitutional isomerism between the molecules m_R^1 and m_R^2.

Definition 16. A pair (f^0, g^0) of the non-conjugated constitutional differences $f^0 \subsetneq m_R^1$ and $g^0 \subsetneq m_R^2$ $f^0 \neq g^0$ for which

$$m_R^1 (\hom f^0) = m_R^2 (\hom g^0)$$

is said to precise type of the R -constitutional isomerism of the molecules m_R^1 and m_R^2, *i.e.*, they define *R-(f⁰, g⁰)-constitutional isomerism of the molecules m_R^1 and m_R^2*.

Definition 17. Let m^1 and m^2 be EF-identical and R-isomeric and for every fragment $f_R \subsetneq m_R^1$ and $f_R \subsetneq m_R^2$ they are f_R-homologous, although they are m_R^1- and m_R^2-dishomologous then they are said to be *pair of R-stereoisomers*.

$$m_R^1 \equiv_{\text{s-iso R}} m_R^2$$

Obviously each pair of chiral isomers, as for example right- and left-handed tartaric acids, represented by their structural 3D models, S3D-representation, is a pair of S3D-stereoisomers. Yet, in historical van't Hoff tetrahedral representation of organic molecules [91] (Th) they are Th-stereoisomers as well.

In fact, Definition 17 covers not only enantiomers but a larger class of stereoisomers such as some of the topological isomers. For example molecular knots can have the same constitution and connectivity, yet different number of crossings, thus different topological type.[55,90,92] In this way, not only left- and right handed C_{60} trefoil knots can be distinguished, but also C_{60} trefoil knot and figure-eight knot, in common 3D representation, are pair of R-stereoisomers. The same holds true for certain molecular catenanes of the same constitution and different number of crossings.[93] Last but not least, rotaxanes exhibiting different insertions of the non-cyclic component into the cyclic one may be in some representations pair of R-stereoisomers.

COMPOSITION OF THE HOMOLO OPERATIONS

In the section Linearity of Homolo Operation we show that composition of the homolo operations can be defined and that composition with respect to the same fragment f is idempotent. Now let us make a closer examination of the composition of the homolo operations.
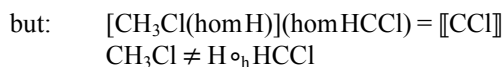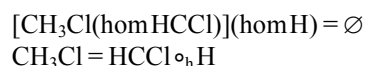
Remark that, in full analogy to facts described in Sections the Homology Relation between molecules and The Molecules are Partially Homolo Ordered, composition of the homolo relation with respect to two (several) distinct fragments f_1 and f_2 determine both equivalence and ordering relations.

Definition 18. The molecule m is said to be product of two distinct fragments f_1 and f_2, none of which is generator of the molecule m or the empty fragment e, $m = f_1 \circ_h f_2$, if superposition of the homolo operation with respect to the fragments f_1 and f_2 is equal to the empty set.
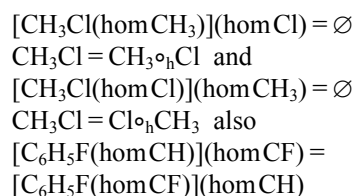
$$m = f_1 \circ_h f_2, \text{ if } \exists f_1, f_2 \subsetneq m \text{ such that}$$
$$f_1 \neq f_2 \text{ and } f_1 \neq e \neq f_2, m(\text{hom} f_1) \neq \varnothing \text{ and}$$
$$m(\text{hom} f_2) \neq \varnothing \text{ and } [m(\text{hom} f_1)](\text{hom} f_2) = \varnothing$$

Then, the fragments f_1 and f_2 are said to be *factors of the molecule*.

Examples:

$$[CH_3Cl(\text{hom} HCCl)](\text{hom} H) = \varnothing$$
$$CH_3Cl = HCCl \circ_h H$$

but:
$$[CH_3Cl(\text{hom} H)](\text{hom} HCCl) = [\![CCl]\!]$$
$$CH_3Cl \neq H \circ_h HCCl$$

Thus generally, factorization $\circ_h$ is not commutative. However, for some factors the homolo $\circ_h$ multiplication can be commutative:

$$[CH_3Cl(\text{hom} CH_3)](\text{hom} Cl) = \varnothing$$
$$CH_3Cl = CH_3 \circ_h Cl \text{ and}$$
$$[CH_3Cl(\text{hom} Cl)](\text{hom} CH_3) = \varnothing$$
$$CH_3Cl = Cl \circ_h CH_3 \text{ also}$$
$$[C_6H_5F(\text{hom} CH)](\text{hom} CF) =$$
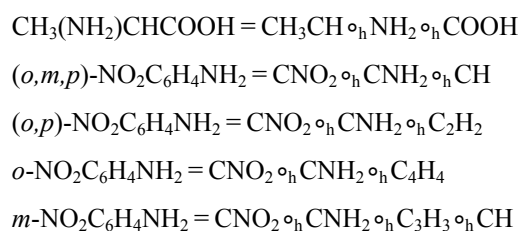$$[C_6H_5F(\text{hom} CF)](\text{hom} CH)$$

Analogously to Definition 18. we can consider factorization of the molecule onto more than two factors.

Definition 19. The molecule m_R is said to be a product of n distinct factors f_1, f_2, ..., $f_n \subsetneq m$, $f_i \neq f_j$, $m = f_1 \circ_h f_2 \circ_h \ldots \circ_h f_n$ iff: (i) superposition of the homolo operation with respect to these fragments is equal to the empty set, and (ii) if for $k < n$ none of the compositions of k the homolo operation with respect to any of these fragments, preserving the order of the subscripts, is equal to the empty set.

(i) $m = f_1 \circ_h f_2 \circ_h \ldots \circ_h f_n$ iff
         $[[m(\text{hom} f_1)](\text{hom} f_2)..]_{n-1}(\text{hom} f_n) = \varnothing$

(ii)   and for $k < n$,
         $[m(\text{hom} f_1)](\text{hom} f_2)..]_{k-1}(\text{hom} f_k) \neq \varnothing$

Examples:

$$CH_3(NH_2)CHCOOH = CH_3CH \circ_h NH_2 \circ_h COOH$$

$$(o,m,p)\text{-}NO_2C_6H_4NH_2 = CNO_2 \circ_h CNH_2 \circ_h CH$$

$$(o,p)\text{-}NO_2C_6H_4NH_2 = CNO_2 \circ_h CNH_2 \circ_h C_2H_2$$

$$o\text{-}NO_2C_6H_4NH_2 = CNO_2 \circ_h CNH_2 \circ_h C_4H_4$$

$$m\text{-}NO_2C_6H_4NH_2 = CNO_2 \circ_h CNH_2 \circ_h C_3H_3 \circ_h CH$$

Remark. The factorization of the non-elemental molecule always exists because each molecule is composed from its elements.

The factorization of a molecule is connected with a property of the dividing fragments mentioned in Example 14, section The Homology Relation Between Molecules, that some fragments can divide the molecule however cannot be element of any of the multisets of remaining fragments. According to the assumption made in this paper, only a connected fragment of a molecule can be the dividing fragment. Thus, the (ordinary) set of all possible divisors of the molecule, D^m, is simply the set of all connected subsets of the molecule.

On the other hand, the (ordinary) set of all remaining fragments, R_f^m, different elements of all multisets F_f^m, *i.e.*, base of union $(\uplus \ F_{fi}^m)^b$, is a subset of D^m. $R^m \subseteq D^m$ and $R^m = D^m$ only for atoms and diatomic heteronuclear molecules.

For example:

$$D^{OCS} = \{O, C, S, OC, CS, OCS, e\} \ \text{ and}$$
$$R^{OCS} = \{CS, O, S, CO, OCS, e\} \ \text{ and } \ C \notin R^{OCS}$$

Finally, let us mention that the introduced homolo concept generates many questions that seems to be worth to be solved and proven, however, this goes far beyond this introductory study.

CONCLUSIONS

The homology (homolo) relation between molecules was introduced. This relation is a generalization of an old idea of series of homologous compounds. The homolo relation operates on a molecule by removing all the structural fragments that are identical with a certain selected fragment. As a result a multiset of fragments is produced. It was shown that the homolo relation is an equivalence relation in a set of molecules. Thus, by choosing various reference fragments, the molecules can be classified in many different ways. On the basis of this relation a partial order can be introduced to the set of molecules as well as in a set of fragments of a molecule. Using the language of homolo operation it is possible to redefine such old ideas as constitutional and stereo isomers as well as to introduce some new terms such as conjugation of fragments, generator of a molecule, and, for instance, factorization of a molecule onto fragments.

REFERENCES

1. Journal of Mathematical Chemistry, http://springerlink.metapress.com/content/101749/
2. R. B. King, *Foundations of Chemistry* **2** (2000) 221–236.
3. P. Tebbutt, *Basic Mathematics for Chemists*, 2nd Ed., Wiley, 1998.
4. A. Cayley, *Philosoph. Mag.* **67** (1874) 444–447.
5. A. T. Balaban and F. Harary, *Tetrahedron* **24** (1968) 2505–2516.
6. F. Harary, E. M. Palmer, R. W. Robinson, and R. C. Read, *Pólya's Contribution to Chemical Enumeration,* in: A. T. Balaban (Ed.), *Chemical Applications of Graph Theory*, Academic Press, London, 1976.
7. K. Balasubraminian, *Chem. Rev.* **85** (1985) 599–618.
8. J. Cioslowski, *J. Comput. Chem.* **8** (1987) 906–915.
9. I. Gutman, S. J. Cyvin, *Introduction to the Theory of Benzenoid Hydrocarbons*, Springer, Berlin, 1989.
10. J. Cz. Dobrowolski, *J. Chem. Inf. Comp. Sci.* **42** (2002) 490–499.
11. N. Trinajstić, S. Nikolić, J. V. Knop, W. R. Müller, and K. Szymanski, *Computational Chemical Graph Theory: Characterization, Enumeration and Generation of Chemical Structures by Computer Methods*, Simon & Schuster/Horwood, Chichester, 1991, pp. 356.
12. M. L. Contreras, R. Valdivia, and R. Rozas, *J. Chem. Inf. Comput. Sci.* **32** (1992) 483–491.
13. M. L. Contreras, R. Valdivia, and R. Rozas, *J. Chem. Inf. Comput. Sci.* **32** (1992) 323–330.
14. E. C. Kirby, *J. Math. Chem.* **11** (1992) 187–197.
15. D. Plavšić D. Vukičević, and M. Randić, *Croat. Chem. Acta*, **78** (2005) 493–502.
16. M. Randić, S. Nikolić, and N. Trinajstić, *J. Chem. Inf. Comput. Sci.* **35** (1995) 357–365.
17. J. Aires-de-Sousa, J. Gasteiger, I. Gutman, and D. Vidović, *J. Chem. Inf. Comput. Sci.* **44** (2004) 831–836.
18. R. B. King and D. H. Rouvray, (Eds.), *Graph Theory And Topology in Chemistry*, Elsevier, 1987.
19. E. Estrada, *Chem. Phys. Lett.* **336** (2001) 248–252.
20. I. V. Stankevich, M. I. Skovortsova, and N. S. Zefirov, *J. Mol. Struct. (Theochem)* **342** (1995) 173–179.
21. M. Randić, P. J. Hansen, and P. C. Jurs, *J. Chem. Inf. Comput. Sci.* **28** (1988) 60–68.
22. M. Randić and J. Cz. Dobrowolski, *Int. J. Quantum. Chem.*, **70** (1998) 1209–1215.
23. L. Pogliani, *Chem. Rev.* **100** (2000) 3827–3858.
24. M. Randić, J. Zupan, M. Novič, B. D. Gute, and S. C. Basak, *SAR QSAR Env. Res.* **13** (2002) 689–703.
25. N. Trinajstić, D. Babić, S. Nikolić, D. Plavšić, D. Amić, and Z. Mihalić, *J. Chem. Inf. Comput. Sci.* **34** (1994) 368–376.
26. D. Janežič, A. Miličević, S. Nikolić, and N. Trinajstić, Ivan Gutman (Ed.), *Mathematical Chemistry Monographs MCM-3* Faculty of Science, University of Kragujevac, Kragujevac, 2007.
27. A. T. Balaban, A. Chiriac, I. Motoc, and Z. Simon, *Lecture Notes in Chemistry*, 15, Springer, NY 1980.
28. E. Estrada and H. Gonzalez, *J. Chem. Inf. Comput. Sci.* **43** (2003) 75–84.
29. O. Ivanciuc, T. Ivanciuc, D. J. Klein, W. A. Seitz, and A. T. Balaban, *J. Chem. Inf. Comput. Sci.* **41** (2001) 536–549.
30. D. Bonchev, *J. Chem. Inf. Comput. Sci.* **41** (2001) 582–592.
31. M. Randic, S. C. Basak, *J. Chem. Inf. Comput. Sci.* **41** (2001) 614–618.
32. N. Trinajstić, *Chemical Graph Theory*, 2nd Ed., CRC Press, 1992.
33. R. Todeschini and V. Consonni, *Handbook of Molecular Descriptors*, Wiley, 2000.
34. A. Sabljić and N. Trinajstić, *Acta Pharm. Jugosl.* **31** (1981) 189–214.
35. E. Wasserman, *J. Am. Chem. Soc.* **82** (1960) 4433–4434.
36. H. L. Frish and E. Wasserman, J. *Am. Chem. Soc.* **83** (1961) 3789–3795.
37. E. Wasserman, *Sci. Amer.* **207** (1962) 94–100.
38. J. C. Wang and H. Schwartz, *Biopolymers* **5** (1967) 953–966.
39. B. Hudson and J. Vinogrd, *Nature* **216** (1967) 647–652.
40. D. A. Clayton and J. Vinogrd, *Nature* **216** (1967) 652–657.
41. M. A. Krasnow, A. Stasiak, S. J. Spengler, F. B. Dean, T. Koller, and N. R. Cozzarelli, *Nature* **304** (1983) 559–560.
42. S. A. Wasserman, J. M. Dungan, and N. R. Cozzarelli, *Science* **229** (1985) 171–174.
43. F. B. Dean, A. Stasiak, T. Koller, and N. R. Cozzarelli, *J. Biol. Chem.* **260** (1985) 4975–4983.
44. J. Cz. Dobrowolski, *Polimery* **48** (2003) 3–15.
45. J. Cz. Dobrowolski and A. P. Mazurek, *J. Chem. Inf. Comp. Sci.* **43** (2003) 861–869.
46. J. Cz. Dobrowolski and A. P. Mazurek, *J. Chem. Inf. Model.* **45** (2005) 1030–1038.
47. C. O. Dietrich-Buchecker and J.-P. Sauvage, *Angew. Chem.*, Int. Ed. Engl. **28** (1989) 189–192.
48. C. O. Dietrich-Buchecker, J. G. Guilhem. C. Pascard, and J.-P. Sauvage, *Angew. Chem.*, Int. Ed. Engl. **29** (1990) 1154–1156.

49. J.-C. Chambron, C. O. Dietrich-Buchecker, and J.-P. Sauvage, *Top. Curr. Chem.* **165** (1993) 131–162.

50. G. Schill, *Catenanes, Rotaxanes, and Knots*, Academic Press, New York, 1971.

51. C. O. Dietrich-Buchecker and J.-P. Sauvage, *Chem. Rev.* **87** (1987) 795–810.

52. F. Vögtle, T. Dünnwald, and T. Schmidt, *Acc. Chem. Res.* **29** (1995) 451–460.

53. R. Jäger, M. Händel, J. Harren, K. Rissanen, and F. Vögtle, *Liebigs Ann.* (1996) 1201–1207.

54. J. Cz. Dobrowolski, *Croat. Chem. Acta.* **76** (2003) 145–152.

55. D. M. Walba, R. M. Richards, and R. C. Haltiwanger, *J. Am. Soc.*, **104** (1982) 3219–3221.

56. D. M. Walba, in: R. B. King (Ed.) *Chemical Application of Topology and Graph Theory*, Elsevier, Amsterdam, 1983. pp. 17–32.

57. D. M. Walba, Q. Y. Zheng, and K. Schilling, *J. Am. Chem. Soc.* **114** (1992) 6259–6260.

58. H. S. Rzepa, *Chem. Rev.* **105** (2005) 3697–3715.

59. D. Ajami, O. Oeckler, A. Simon, and R. Herges, *Nature* **426** (2003) 819.

60. R. Herges, *Chem. Rev.* **106** (2006) 4820–4842.

61. J. Kang Park, Z. S. Yoon, M.-C. Yoon, K. S. Kim, S. Mori, J-Y. Shin, A. Osuka, and D. Kim, *J. Am. Chem. Soc.* 2008 ASAP Article 10.1021/ja7100483.

62. A. Graovac and N. Trinajstić, *Croat. Chem. Acta* **47** (1975) 95–104.

63. K. Murasugi, *Knot Theory and Its Applications*, Birkhäuser, Boston, Basel, Berlin, 1996.

64. J. H. Przytycki, *Węzły. Podejście Kombinatoryczne do Teorii Węzłów*, SCRIPT, Warszawa 1995.

65. D. J. Klein, *Match* **42** (2000) 7–21.

66. D. Bonchev and E. Gordeeva, *Match* **42** (2000) 85–117.

67. I. Gutman, J. Rada, and O. Araujo, *Match* **42** (2000) 145–154.

68. O. Ivanciuc, T. Ivanciuc, D. Cabrol-Bass, and A. T. Balaban, *Match* **42** (2000) 155–180.

69. M. Randić, M. Vračko, M. Novič, and S. C. Basak, *Match* **42** (2000) 181–231.

70. M. Randić, *J. Math. Chem.* **24** (1998) 345–358.

71. E. Ruch and D. J. Klein, *Theoret. Chem. Acc.: Theory, Comput. Model.*, **63** (1983) 447–472.

72. R. Chauvin, *J. Math. Chem.*, **16** (1994) 245–256.

73. J. Dugundji, P. D. Gillespie, D. Marquarding, I. Ugi, and F. Ramirez, in: A.T. Balaban (Ed.) *Chemical Applications of Graph Theory*, Academic Press, London, 1976. pp. 107–174.

74. R. G. Woolley, *J. Math. Chem.* **23** (1998) 3–12.

75. E. R. Scerri, *J. Chem. Educ.* **77** (2000) 522–526.

76. E. Francoeur, *HYLE – Int. J. Philos. Chem.* **6** (2000) 63–84.

77. P. Zeidler, *HYLE – Int. J. Philos. Chem.* **6** (2000) 17–34.

78. R. Hoffmann and P. Laszlo, *Diogenes* **147** (1989) 23–50.

79. R. Hoffmann and P. Laszlo, *Angew. Chem. Int. Ed.* **30** (1991) 1–16.

80. N. Trinajstić and I. Gutman, *Croat. Chem. Acta*, **75** (2002) 329–356.

81. M. Randić, *Chem. Rev.* **103** (2003) 3449–3605.

82. A.T. Balaban (Ed.), *Chemical Applications of Graph Theory*, Academic Press, London 1976.

83. I. Gutman and O. E. Polansky, *Mathematical Concepts in Organic Chemistry*, Springer, Berlin, 1986.

84. K. Kuratowski and A. Mostowski, *Set Theory, with an Introduction to Descriptive Set Theory*, PWN, Warszawa, North-Holland, Amsterdam 1976.

85. W. D. Blizard, *Notre Dame J. Formal Logic* **30** (1989) 36–66.

86. W. D. Blizard, *Notre Dame J. Formal Logic* **31** (1990) 346–368.

87. A. Syropoulos, *Mathematics of Multisets*, in: C. Calude, G. Păun, G. Rozenberg, A. Saloma (Eds.), *Multiset Processing,* no. 2235 in *Lecture Notes in Computer Science*, Springer, Berlin, 2001. pp. 347–358.

88. N. J. Wildberger, *A new Look at Multisets*, preprint, University of New South Wales, Sydney, Australia, 2003.

89. M. Ch. Gerhardt, *Précis de Chimie Organique*, Tome premiere, Chez Fortin, Maison et Cie, Libraries, Paris, 1844. pp 25.

90. J. J. Berzelius, *Ann. Phys. Chem.*, **19** (1830) 305–335.

91. J. H. van't Hoff, *Archives neerlandaises des sciences exactes et naturelles*, **9** (1874) 445–454.

92. J. Cz. Dobrowolski, A. P. Mazurek, *Int. J. Quantum. Chem.* **70** (1998) 1009–1015.

93. J. Cz. Dobrowolski, A. P. Mazurek, *Int. J. Quant. Chem.* **75** (1999) 839–846.

SAŽETAK

Homološki odnosi među molekulama: oživljavanje starog načina klasifikacije molekula

Jan Cz. Dobrowolski

Industrial Chemistry Research Institute, 01-793 Warsaw, 8 Rydygiera Street, Poland
National Medicines Institute, 00-725 Warsaw, 30/34 Chełmska Street, Poland

Predstavljeni su homološki (homolo) odnosi među molekulama. Ovaj odnos je poopćenje stare ideje o seriji homolognih spojeva. Homolo operacija djeluje na molekulu tako da uklanja sve strukturne fragmente koji su identični određenom odabranom fragmentu. Kao rezultat dobije se komplet skupova fragmenata. Pokazano je da je homolo odnos ekvivalentni odnos u kompletu molekula. Tako se odabirom različitih polaznih fragmenata molekule mogu klasificirati na mnogo različitih načina. Korištenjem jezika homolo operacije moguće je redefinirati ideje poput konstitucijskih i stereo izomera, kao i generator molekule, npr. faktorizacija molekule na fragmente.