

# Atomic Ring Invariant and Modified CANON Extended Connectivity Algorithm for Symmetry Perception in Molecular Graphs and Rigorous Canonicalization of SMILES

Dmytro G. Krotko (✉ [d.krotko@gmail.com](mailto:d.krotko@gmail.com))

---

Research article

**Keywords:** molecular graphs, symmetry perception, SMILES, canonicalization

**Posted Date:** June 10th, 2020

**DOI:** <https://doi.org/10.21203/rs.3.rs-33257/v1>

**License:**  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

---

**Version of Record:** A version of this preprint was published on August 20th, 2020. See the published version at <https://doi.org/10.1186/s13321-020-00453-4>.

# Abstract

We propose new invariant (product of corresponding primes for ring size of each bond of an atom) as a simple unambiguous ring invariant of the atom which allows to distinguish symmetry classes in highly symmetrical molecular graphs (with using of traditional local atom invariants). Also, we propose modifications of Weininger's CANON algorithm which let to avoid its ambiguities (swapping and leveling of ranks, incorrect determination of symmetry classes in non-aromatic annulenes, arbitrary selection of atom for breaking ties). The atomic ring invariant and the Modified CANON algorithm allow us to create a rigorous procedure for the generation of canonical SMILES which can be used for accurate and fast structural search in large chemical databases.

## Introduction

Perception of the symmetry of atoms in the molecular graphs is an important problem for chemoinformatics. Correct determination of the symmetry classes for atoms in a molecular graph is a basis for finding a canonical ordering of the atoms in a molecule, which in turn is necessary for generating a unique representation of the molecule [1]. This approach is used in canonicalization algorithms for linear representations of molecular graphs like SMILES [2, 3, 4, 5] and InChI [4, 6]. Canonical linear notations of molecular graphs are widely used in contemporary chemical databases since they allow to compare chemical structures as plain strings and these strings could be lexicographically ordered, that allow using very fast binary search in very large databases of canonicalized linear representations of chemical graphs [1]. Accuracy and speed of searching in millions of known compounds and billions of virtual compounds for drug discovery depend on the correctness and effectiveness of these algorithms.

The first algorithm for symmetry perception of the atoms in the molecular graphs was developed in 1965 by Morgan [7] for the Chemical Abstracts Service. In the first phase of this algorithm, an initial value is assigned to each atom based on its local properties like degree, atomic number, and bond type. In the second so-called relaxation step, these values are iteratively refined by summing of the current values of the immediate neighbors of each atom and assigning this new value to the atom. The algorithm is terminated when a count number of different values stops to increase from the previous step of the algorithm. This algorithm is a classical realization of extended connectivity algorithm and works pretty well on most of the typical chemical structures. But in 1975 Randić has pointed out [8] that issues exist with this algorithm: for some molecular graphs nonequivalent atoms can be assigned the same extended connectivity values. Some of these issues are because the sum of integers is not unambiguous function and addition various integers can give the same sum. This problem can be solved by a method described by Weininger et al [3] in 1989 by replacing summing in refining step of the algorithm by the products of corresponding primes as an unambiguous function that can be easily shown from the prime factorization theorem. But the method of corresponding primes is not universal: for some 'pathological' molecular graphs (one of them shown in Fig. 1) none of the extended connectivity algorithms, which initial values assigned to each atom based on local properties of atom only, can resolve all symmetry classes of these

highly symmetrical molecular graphs [1]. This fact led Ivanciuc to statement [1] that only algorithms with an explicit determination of graph automorphisms (an isomorphism of a graph with itself is called an automorphism) can accurately determine all symmetry classes in complex molecular graphs. Such algorithms, like Shelley-Munk algorithm [9], HOC (Hierarchically Ordered Extended Connectivities) [10], McKay algorithm [11] (which is implemented in open-source library nauty) and Faulon et al [12, 13] signatures algorithm are quite complex and computationally intensive because they use some sort of label permutations. Simplified and adapted for chemical graphs version of McKay algorithm (nauty) is implemented in the InChI canonicalization algorithm [6]. Another approach is shown by Schneider, Sayle, and Landrum [5] in 2015 and is implemented in an open-source RDKit library. They use some nonlocal invariants of atoms (chiral invariant and special high-symmetry invariant) for resolving symmetry classes of chiral structures and highly symmetrical molecular graphs. Using these invariants allows to obtain an accurate number of symmetry classes in chiral structures and 'pathological' molecular graphs, respectively, but their approach seems to me as overcomplicated and I would propose much more obvious nonlocal invariants of atoms for same purposes.

## Methods

We will discuss use of local invariants, then proposed nonlocal invariants and improvements of refinement procedure which can resolve some ambiguities in Weininger's method of corresponding primes [3]. After that procedure of obtaining canonical absolute SMILES (with chirality) will be described.

**Local invariants.** Degree, atomic number, and bond type are used as local invariants of the atom in the original Morgan algorithm [7]. Weininger proposed to use the number of connections, number of non-hydrogen bonds (degree), atomic number, the sign of charge, absolute charge and number of attached hydrogens as local invariants [3]. For the full representation of local invariants of an atom, we can classify them as properties of atom proper (atomic number – count of protons in atomic nuclei, atomic mass – the sum of counts of protons and neutrons in atomic nuclei, charge – count of protons in atomic nuclei minus count of electrons of atom) and properties of the nearest neighborhood of the atom (degree – count of explicit connections, count of attached hydrogens, connectivity – total count of connections, valence – the sum of bond orders of all connections). For using as an initial rank of the atom, we must combine all local invariants in the atomic vector in some order. Weininger has pointed out [3] that a terminal atom is preferred to start of traversal of the molecular graph (if it is possible). From this reason, degree of the atom must be at the start of a combined atomic vector (1 character). Next positions consecutively: atomic number (3 characters), count of attached hydrogens (1 character), the sign of charge (1 character: '0' – if no charge, '1' – if the charge is positive, '2' – if the charge is negative), absolute charge (1 character), connectivity (1 character), valence (1 character), atomic mass (3 characters, '000' if unspecified). The atomic vector of local invariants has 12 characters.

**Chirality invariant.** In SMILES, a designation of chirality depends on the enumeration order of atoms, and from this reason, simple parity symbols '@' and '@@' cannot be used as invariant because it's meaning change with transposition of atoms. But we can determine the order of atoms around the chiral center

relative to symmetry classes of its neighbors [14, 15]. To determine this order we should use extended connectivity algorithm to atomic invariants (local and ring nonlocal described further) without chirality designation and obtain symmetry classes for non-chiral structure. For all atoms with no chiral parity chirality invariant equal to 0 (by definition). If two or more neighbors of the atom with chiral parity have the same symmetry class then chirality invariant of this atom equals to 0 also. If the order of symmetry classes can be obtained by an even number of swaps from direct ascending order of these numbers then chirality invariant equal to count of the sign '@' in parity, and otherwise chirality invariant equal to 3 - count of the sign '@' in parity. Such a definition of chiral invariant gives us the truly invariant property of chiral atom which is not depending on the order of atoms in SMILES because symmetry classes of atoms are independent of its order. This invariant can be concatenated to the atomic vector of each atom (1 character). After concatenation of chiral invariant, we must recalculate all atom ranks in structure.

Thus complete atomic vector of each atom must contain 13 characters: 12 for local invariants and 1 character for chirality invariant. Since all characters of the atomic vector are digits then we can transform the atomic vector to the number.

**Ring invariant of atom.** If we use only traditional local invariants for each atom of the molecular graph from Fig. 1 we will obtain only two symmetry classes after refinement procedure of extended connectivity algorithm: one for atoms with degree = 3 and another one for atoms with degree = 2. But obviously, there are three symmetry classes in this structure. How do we know that? Because eight atoms with degree = 2 belong to 6-membered rings, but another four atoms with degree = 2 do not. This very simple observation gives us obvious nonlocal invariant for each ring atom – ring size of the smallest ring to which this atom belongs (for non-ring atom this invariant = 0 by definition). This value is invariant, i.e. it does not depend on the order of atoms in the molecular graph. Now, for molecular graph from Fig. 1 for eight atoms with degree = 2 this invariant = 6, and for another four atoms with degree = 2 this invariant = 12. Taking into account local invariants of each atom we obtain three symmetry classes for atoms of this molecular graph at once and refinement procedure do not change this.

But if we look at the structure in Fig. 2, we see that this invariant is insufficient because all atoms belong to 3-membered rings and therefore have the same ring size of the smallest ring to which this atom belongs.

To solve this problem, we can first calculate invariant for each bond – the size of the smallest ring to which it belongs. We can use a breadth-first search (BFS) from the first atom which is incident to this bond to second atom which is incident to this bond for obtaining the shortest non-trivial path (if such path does not exist then this is non-ring bond). After obtaining this ring invariant for each bond we can choose for each atom minimal and maximal values of ring sizes of the bonds which are incident to this atom. The minimal ring size of the bonds which are incident to the atom will be ring size of the smallest ring to which this atom belongs, but maximal will be maximum of minimal ring sizes of the bonds to which this atom is incident. Both of these values will be invariants of each atom. For the molecular graph in Fig. 2, four atoms have both these invariants equal to 3, but four another have minimal ring size equals

to 3, but maximal ring size equals to 6. The minimal and maximal ring sizes of bonds which are incident to the atom in combination with local invariants are sufficient to distinguish symmetry classes of atoms in all known to us 'pathological' molecular graphs but to avoid any possible ambiguity we propose to use the product of corresponding primes for ring size of each bond of the atom as unambiguous ring invariant of the atom (contribution for a non-ring bond to the product will be equal to 1). As an example, for the molecular graph in Fig. 2 this ring invariant will be  $5*5*5 = 125$  for the atoms with the maximum ring size of bonds equals to 3 and  $5*5*13 = 325$  for the atoms with the maximum ring size of bonds equals to 6 (since 5 is third prime and 13 is the sixth prime number). Ring invariant will be unambiguous because reverse factorization of it gives us such invariants as the count of ring bonds of atom and the minimal sizes of rings to which all bonds of atom belong.

Now, we have a 13-digit atomic vector and numeric ring invariant for each atom. The product of the numeric value of the atomic vector and ring invariant will be the initial rank of the atom in further refinement step of extended connectivity algorithm.

**Refinement procedure.** After obtaining initial ranks for all atoms we can rank them for obtaining consecutive small integer ranks, and this manipulation does not change their relative order. Further manipulation, in general, corresponds to the CANON algorithm described by Weininger et al [3], but some changes in this algorithm are necessary. If you look on simple structure in Fig. 3 then you see that after obtaining initial atomic vectors, transforming them to numbers and ranking we obtain four symmetry classes. If we follow the original description of the CANON algorithm then after the first iteration we obtain products of corresponding primes for atoms: for  $\text{CH}_3 - 7$ ,  $\text{CH} - 2*3*5 = 30$ ,  $\text{CH}_2 - 5*7 = 35$ ,  $\text{O} - 3*7 = 21$ . After ranking of these integers we obtain new ranks:

As you can see there is rotation of ranks in the ring. This is an infinite loop of ranks rotations that cannot give us stable numbers for symmetry classes for atoms in such type of structures. The first idea for solving this problem is to multiply the corresponding prime of the current rank of each atom with the product of corresponding primes for ranks of its neighbors. Try this approach, products for first iteration:  $\text{CH}_3 - 2*7 = 14$ ,  $\text{CH} - 7*2*3*5 = 210$ ,  $\text{CH}_2 - 3*5*7 = 105$ ,  $\text{O} - 5*3*7 = 105$ . After ranking we obtain that now O and  $\text{CH}_2$  have the same rank, although previously they had different ranks! This issue arises because the products of primes are an unambiguous function only if we do not consider transposition of number in the product, but in 3-membered rings this fact led to the leveling of previously different ranks. A solution of this problem for 3-membered rings is to multiply the square of the corresponding prime of the current rank of each atom with the product of corresponding primes for ranks of its neighbors. Try this approach, products for first iteration:  $\text{CH}_3 - (2*2)*7 = 28$ ,  $\text{CH} - (7*7)*2*3*5 = 1470$ ,  $\text{CH}_2 - (3*3)*5*7 = 315$ ,  $\text{O} - (5*5)*3*7 = 525$ . After ranking:  $\text{CH}_3 - 1$ ,  $\text{CH} - 4$ ,  $\text{CH}_2 - 2$ ,  $\text{O} - 3$ . As you can see ranks are stable after such calculation for this structure. It can be easily proved that such a method of obtaining new ranks from previous has never led to leveling or swapping of previously different ranks in 3-membered rings. But for structure in Fig. 4 squaring of the corresponding prime of the current rank do not help us: after the first round of the refinement, we have previously different ranks leveled.

If we try to use the third power of the corresponding prime of current rank for this structure then there is no leveling. Generally speaking, this issue arises when we have two or more neighbors of the atom with the same rank. Since in molecular graphs we never have atom degree higher than 6 (structures like SF<sub>6</sub>), we propose to use the eighth power of the corresponding prime of current rank to guarantee that leveling or swapping of ranks in structure never occurs on refinement step (since  $8 = 2^3$ , the eighth power can be calculated quickly).

Another issue with the original CANON algorithm is that it does not take into account bond orders between atoms. For example, it determines only four symmetry classes for structure in Fig. 5 while there are five.

This fact has led to arbitrary choosing between atoms and generation of different SMILES by traversal of molecular graph. We propose to solve this problem by multiplying the eighth power of the corresponding prime of current rank with the corresponding primes for ranks of its neighbors in power equal to the order of bond which connects current atom with the neighbor.

Pseudocode of Modified CANON algorithm:

Set atomic ranks to initial invariants;

Rank them to obtaining consecutive small integer ranks;

Do

Set for each atom the new value of rank as a product of the eighth power of the corresponding prime of current rank with the corresponding primes for ranks of its neighbors in the power equal to the order of bond which connects current atom with the neighbor;

Rank them to obtaining consecutive small integer ranks;

Loop until no invariant partitioning observed;

The Modified CANON algorithm terminates when the count number of different values of ranks stop to increase from the previous step of the algorithm. After its termination we must have the same ranks only on symmetrically equivalent atoms, i.e. ranks of atoms are their symmetry classes [3]. These symmetry classes can be used for calculation of chirality invariant. For acyclic and simple cyclic structures they can be used also directly for further generation of canonicalized SMILES. If after use of the Modified CANON algorithm the count of various ranks in the molecular graph is less than the count of atoms, then to obtain unambiguous order of molecular graph traversal we need to make breaking ties procedure, as Weininger et al pointed out [3]. Our method for this is the same as described: doubling all ranks and reducing the value of the atom, which is tied, by one. The set is then treated as a new invariant set, and the previous algorithm for generating an invariant partitioning is repeated. But Weininger et al assumption that "the double-and-tie-break step does not introduce ambiguity into the ordering since only

otherwise equivalent atoms will be tied at any point” can be wrong in the general case. Ivanciuc pointed out [1]: “Two vertices from different atom invariant class cannot be automorphic, while two vertices from the same atom invariant class are not necessarily automorphic. Despite numerous efforts, no vertex graph invariant is known which is sufficient to establish the automorphism partitioning, because for certain graphs non-automorphic vertices are partitioned in the same class.” For this reason we suggest that only a rigorous way to produce canonical ordering for any chemical graph is to generate permutations for all atoms with the same atom invariant class, obtain SMILES for each permutation by traversal and select lexicographically minimal (or maximal) [1]. Our approach to doing this is: after obtaining symmetry classes choose ring atoms with maximal tied ranks, make tiebreaking for each of them, refine each partition by the Modified CANON algorithm, and save it. If these new partitions have atoms with tied ranks then repeat tiebreaking for each of them as long as ring atoms with tied ranks exist in partitions. For all partitions with fully partitioned ranks for ring atoms generate SMILES by traversal and select lexicographically minimal as canonical SMILES for this molecular graph.

After establishing of canonical order of atoms by the Modified CANON algorithm with, possibly, the tiebreaking further procedure to obtain canonical SMILES corresponds described by Weininger et al [3] (procedure GENES): the structure is treated as a tree and a SMILES string is generated that corresponds to a depth-first search (DFS) of that tree. The lowest canonically numbered atom is chosen as the starting point. This atom becomes the root of a tree for a subsequent DFS. Branching decisions are making by directs branching toward the lowest labeled atom at the fork in the branch. For cyclic structures, Weininger et al use a two-pass method: on the first pass, the DFS algorithm detects which bonds will become ring bonds (chords) and on the second pass, the ring closure indicators (digits) could be appended to the chords node symbols.

After the second pass of the DFS algorithm, there is full information about the canonical order of atoms in the neighborhood of chiral centers. We use the third pass of the DFS algorithm to obtain canonical SMILES with correct parity (absolute SMILES): if canonical order of atoms on the second pass of the DFS algorithm can be obtained by even number of swaps from the original order of atoms in the molecular graph then parity symbol ('@' or '@@') is preserved, if not – then parity symbol is inverted. We cannot use order relatively canonical ranks for correct designation of chirality because of the existence of structures that have only relative configurations and have no chiral centers (structure in Fig. 6 for example).

For such structure, all chirality invariants equal to 0 and that led to two possible traversals of this molecular graph. Previously described tiebreaking procedure allows us to avoid this ambiguity by the selection of lexicographically minimal SMILES as canonical.

## Results And Discussion

Now, we consider the accuracy and computational complexity of the described algorithms.

**Symmetry perception.** Use of ring invariant and the Modified CANON algorithm allow predicting an accurate count of vertex equivalence classes for all graphs from the article of Razinger et al [16] without

any explicit determination of graph automorphisms: for all test graphs in Fig. 6 in this article we have the same count of symmetry classes as in Table 1 [16] for HOC and Shelley-Munk algorithms. Correct counts of vertex equivalence classes are obtained also for all “complex” graphs in Fig. 4 and all regular graphs in Fig. 8 in this article [16], as well for counterexample graphs from Faulon’s article [17]. Due to explicit use of bond orders in the main loop of the modified CANON algorithm, we can correctly determine the count of symmetry classes in non-aromatic annulenes and fullerenes (Ivanciuc pointed out [1] to issues for these classes for classical Morgan algorithm). We are not aware of any molecular graph for which our algorithm would incorrectly determine the count of symmetry classes, but we cannot rule out such possibility for some molecular and particularly general graphs (with an arbitrary large degree of vertices and, correspondently, arbitrary large rings count for each edge and each vertex). We have constant computational complexity for determination of local invariants and chiral invariant of atoms and quadratic in worst-case computational complexity for determination of ring invariant: count of bonds in molecular graphs is approximately equal to the count of atoms and we must calculate for each bond the shortest non-trivial path between atoms which are incident to it by breadth-first search with linear complexity in worst-case. The Modified CANON algorithm has the same computational complexity as Weininger’s original CANON algorithm [3] and thus we have  $O(N^2)$  general computational complexity for the procedure of symmetry perception in molecular graphs.

**Generation of Canonical SMILES.** The procedure of graph traversal by depth-first search has linear computational complexity. But in case of symmetrical structures, we must make the procedure of breaking ties and such count of graph traversals as count of partitions with fully partitioned ranks for ring atoms we obtain after breaking ties. This count depends on the structure of the molecular graph, but for the great majority of “usual” chemical structures with  $N$  non-overlapping bilaterally symmetrical rings, it will be  $2^N$ . We prefer to begin breaking ties from ring atoms with the highest ranks because they have as rule higher degree and thus breaking ties on them after refinement procedure give us faster symmetry breaking in the original molecular graph. For complex molecular graphs with overlapping symmetrical rings count of traversals is hard to predict, but in general, it is not very large. As examples: for structure in Fig. 1 (SMILES: C1CC2CCC1CCC3CCC(CC3)CC2) we have 16 traversals, for structure in Fig. 2 (SMILES: C12C3C1C4C5C4C5C23) – 16 traversals too, for adamantane (SMILES: C1C2CC3CC1CC(C2)C3) – 24 traversals, for cubane (SMILES: C12C3C4C1C5C2C3C45) – 48 traversals, for buckminsterfullerene ( $C_{60}$ ) – 120 traversals. Since traversals are independent of each other, they can be easily parallelized in the computation. We must say that all generated SMILES are the same for each structure from molecular graphs without chirality investigated by us. For this reason, it can seem redundant to generate all traversals, but we cannot guarantee that this will be true for any molecular graph and thus a rigorous way to obtain canonical SMILES is to generate SMILES by all traversals and select lexicographically minimal SMILES as canonical [1]. This is particularly necessary for structures with only relative (dependent) chirality because all chirality invariants on their atoms equal to 0 and they can have several equal traversals which led nevertheless to different SMILES. For example, for structure in Fig. 6 we have two traversals and one of them led to SMILES C[C@H]1CC[C@H](N)CC1, but another one led to SMILES C[C@@H]1CC[C@@H](N)CC1. We have no other rule of choice between these two SMILES except to select

lexicographically minimal SMILES as canonical, i.e. C[C@@H]1CC[C@@H](N)CC1 will be canonical SMILES for this structure.

It should be emphasized that our procedure of generation of canonical SMILES is complete in the sense as Ivanciuc pointed out [1]: it contains the graph invariant atom partitioning step (calculation of the atomic local and nonlocal invariants and refining them by the Modified CANON algorithm for obtaining of the symmetry classes of the atoms) and the canonical code generation by automorphism permutation step (the tiebreaking procedure and generation of SMILES by deep-first traversal for each partition). This means that regardless of the possible existence of the molecular graphs, for which refinement by the Modified CANON algorithm can give an inaccurate count of vertex equivalence classes, our procedure always gives the same (lexicographically minimal) canonical SMILES for all label permutations of one molecular graph. Its only drawback is that in the rare worst cases (highly symmetrical molecular graphs) the count of traversals can have an exponential dependence on the number of rings in the molecular graph. If it will be proved that our local and nonlocal invariants with the refinement by the Modified CANON algorithm are sufficient for the accurate determination of all symmetry classes in all molecular graphs then the automorphism permutation step will be facultative (except structures with only relative chirality) and the overall computational complexity of generation of canonical SMILES will be  $O(N^2)$  in worst case as the computational complexity of the Modified CANON algorithm.

## Conclusion

We propose new invariant (product of corresponding primes for ring size of each bond of an atom) as simple unambiguous ring invariant of the atom which allows to distinguish symmetry classes in highly symmetrical molecular graphs (with help of traditional local atom invariants). Also, we propose modifications of Weininger's CANON algorithm which let to avoid its ambiguities (swapping and leveling of ranks, incorrect determination of symmetry classes in non-aromatic annulenes, arbitrary selection of atom for breaking ties). Atomic ring invariant and Modified CANON algorithm allow us to create a rigorous procedure for generation of canonical SMILES which can be used for accurate and fast structural search in large chemical databases.

## Associated Content

Supporting Information. Additional file Graphs.xlsx contains examples of test and complex molecular graphs with counts of symmetry classes for them using Modified CANON algorithm and algorithms with an explicit determination of graph automorphisms.

## Declarations

## Availability of data and materials

Not applicable

# Competing interests

The author declare that he has no competing interests.

# Funding

Not applicable

# Authors' contributions

DK put forward the idea, wrote the program code and tested it. DK also wrote the manuscript. DK read and approved the final manuscript.

# Acknowledgments

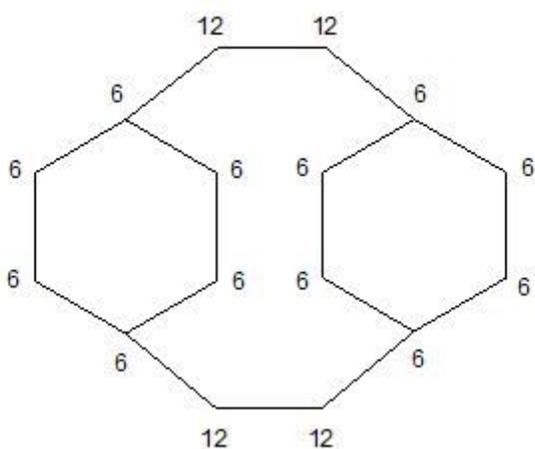
Author thanks Andriy Tolmachev for supporting this research, Yurii Moroz for useful discussions and my wife Svitlana Stetsenko for support during the COVID-19 quarantine.

# References

1. Ivanciuc, O. Canonical Numbering and Constitutional Symmetry. In Gasteiger J. (ed.), Handbook of Chemoinformatics, **2003**, Wiley, pp. 139–160.
2. Weininger, D. SMILES, a Chemical Language and Information System. 1. Introduction to Methodology and Encoding Rules. J. Chem. Inf. Model. **1988**, 28, 31–36.
3. Weininger, D.; Weininger, A.; Weininger, J. L. SMILES. 2. Algorithm for Generation of Unique SMILES Notation. J. Chem. Inf. Model. **1989**, 29, 97–101.
4. O'Boyle, N. M. Towards a Universal SMILES representation – A standard method to generate canonical SMILES based on the InChI. J. Cheminf. **2012**, 4, 22.
5. Schneider, N.; Sayle, R. A.; Landrum G. A. Get Your Atoms in Order – An Open-Source Implementation of a Novel and Robust Molecular Canonicalization Algorithm. Chem. Inf. Model. **2015**, 55 (10), 2111–2120.
6. Heller, S. R.; McNaught, A.; Pletnev, I.; Stein, S.; Tchekhovskoi, D. InChI, the IUPAC International Chemical Identifier. J. Cheminf. **2015**, 7, 23.
7. Morgan, H. L. The Generation of a Unique Machine Description for Chemical Structures - A Technique Developed at Chemical Abstracts Service. J. Chem. Doc. **1965**, 5, 107–113.
8. Randić, M. On Unique Numbering of Atoms and Unique Codes for Molecular Graphs. J. Chem. Inf. Model. **1975**, 15, 105–108.

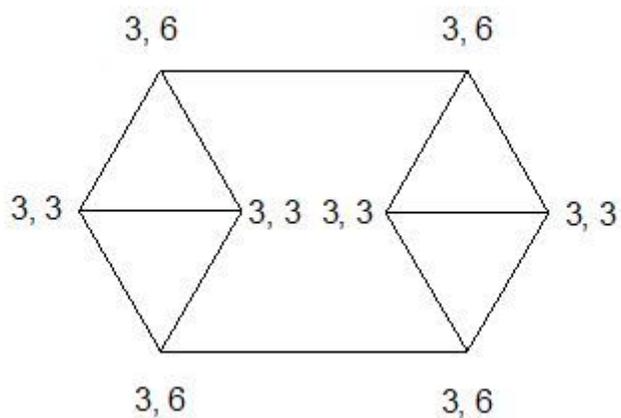
9. Shelley, C. A.; Munk, M. E. An Approach to the Assignment of Canonical Tables and Topological Symmetry Perception. *J. Chem. Inf. Comput. Sci.* **1979**, 19, 247–250.
10. Balaban, A. T.; Mekenyan, O.; Bonchev, D. Unique Description of Chemical Structures Based on Hierarchically Ordered Extended Connectivities (HOC Procedures). I. Algorithms for Finding Graph Orbits and Canonical Numbering of Atoms. *J. Comput. Chem.* **1985**, 6, 538–551.
11. McKay, B. D. Practical Graph Isomorphism. *Congr. Numer.* **1981**, 30, 45–87.
12. Faulon, J. L.; Collins, M. J.; Carr, R. D. The Signature Molecular Descriptor. 4. Canonizing Molecules using Extended Valence Sequences. *J. Chem. Inf. Model.* **2004**, 44, 427–436.
13. Koichi S, Iwata S, Uno T, Koshino H, Satoh H. Algorithm for Advanced Canonical Coding of Planar Chemical Structures That Considers Stereochemical and Symmetric Information. *J. Chem. Inf. Model.* **2007**, 47, 1734–1746.
14. Petrarca, A. E.; Lynch, M. F.; Rush J. E. A Method for Generating Unique Computer Structural Representations of Stereoisomers. *J. Chem. Doc.* **1967**, 7, 3, 154–165.
15. Rohde, B. Representation and Manipulation of Stereochemistry. In Gasteiger J. (ed.), *Handbook of Chemoinformatics*, **2003**, Wiley, pp. 206–230.
16. Razinger, M.; Balasubramanian, K.; Munk, M. E. Graph Automorphism Perception Algorithms in Computer-enhanced Structure Elucidation. *J. Chem. Inf. Model.* **1993**, 33, 197–201.
17. Faulon, J. L. Isomorphism, Automorphism Partitioning, and Canonical Labeling can be Solved in Polynomial-time for Molecular Graphs. *J. Chem. Inf. Model.* **1998**, 38, 432–444.

## Figures



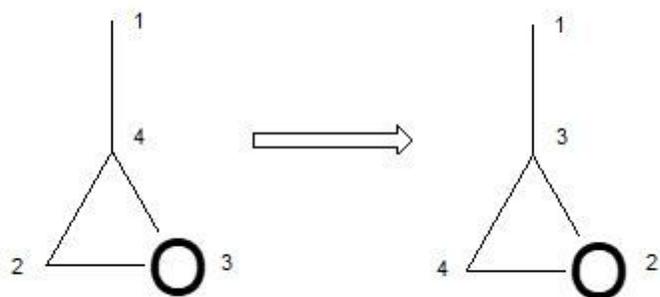
**Figure 1**

'Pathological' graph with sizes of the smallest rings for atoms



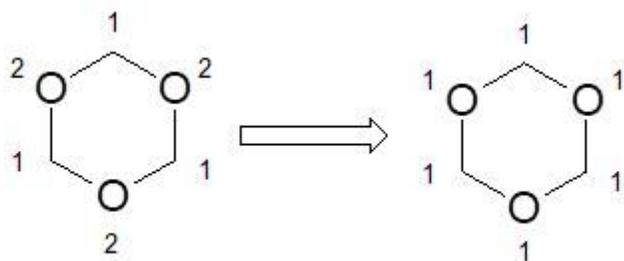
**Figure 2**

Complex graph with minimal and maximal ring sizes of bonds which are incident to all atoms



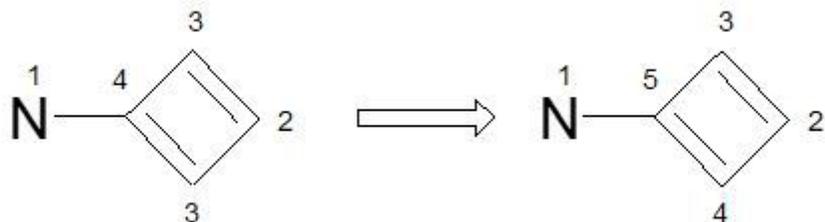
**Figure 3**

Example of ranks rotation by original CANON algorithm



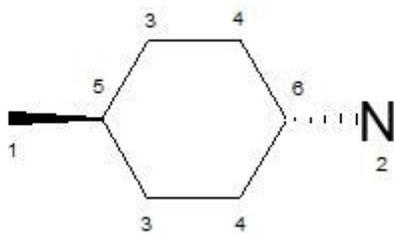
**Figure 4**

Example of leveling of ranks at the refinement step



**Figure 5**

Incorrect (left) and correct (right) symmetry classes for example of annulene



**Figure 6**

Example of the structure with only dependent chirality and no chiral centers

## Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [supplement1.xlsx](#)
- [supplement2.zip](#)