

On the Misinterpretation of Our Algorithm for the Perception of Constitutional Symmetry

CLEMENS JOCHUM and JOHANN GASTEIGER*

Institut für Organische Chemie, Technische Universität München, D-8046 Garching, West Germany

Received September 6, 1978

It is shown that claims that our algorithm for canonical numbering and the perception of constitutional symmetry fails in certain cases are ill-founded. We demonstrate that the "counterexamples" given are correctly treated.

A recent paper¹ in this journal contained critical comments on two methods^{2,3} for the perception of topological symmetry. Carhart's claim¹ that our algorithm³ does not correctly perceive topological symmetry and fails with certain structures is the result of a misinterpretation of our algorithm.

Unfortunately, the author did not contact us directly to help him clarify his misunderstanding. This failure is unusual and difficult to understand. Thus, it was not until we received the recent issue of this journal that we learned of this misinterpretation.

In our paper we were particularly aiming at catching the interest of the organic chemist for the problems of uniquely numbering the atoms of a molecule. Therefore, we put particular emphasis on the criteria for determining priorities among atoms to enable the chemist to manually number the atoms of molecules according to our procedure. We restrained from giving all small details of the algorithm to keep the paper concise, working under the assumption that persons interested in the details would contact us directly. It is astonishing that Carhart at the point where we did not fully elaborate on the details works with the premise that we misconceived the problem. Initially one should rather assume that other people, too, understand a problem. Only if explicit errors are found should one digress from this conviction.

To clarify any misunderstanding we want here to elaborate on our algorithm.

CONSTITUTIONAL SYMMETRY

Carhart¹ uses the expression "symmetry equivalence" in a rather vague manner as it does not specify the kind of representation used for a molecule. We adhere to the term "constitutional symmetry" as it implies that only the constitution of a molecule is considered. Specifically we are talking about constitutionally equivalent atoms. The detection of the constitutional equivalence of atoms is a prerequisite for the perception of constitutionally equivalent bonds,⁴ for the correct treatment of stereochemistry, and for canonically numbering the atoms of a molecule.

In our paper³ we have given a definition for the constitutional equivalence of atoms. To further develop an understanding for the constitutional equivalence of atoms in a molecule we present here two additional definitions, a somehow more intuitive one and a graph-theoretical definition. Both definitions are equivalent to our previous one.

Definition: Two atoms, A_i and A_j , of a molecule M are constitutionally equivalent if a duplicate structure M' of M can be laid upon M in such a manner that (i) atom A'_i lies on atom A_j , and (ii) each other atom of M' lies on an atom of M with the same atomic number and the same number of free electrons, and each bond of M' lies on a bond of M with the same bond order.

Definition: Let M be a molecule consisting of a set of atoms $M = \{A_1, \dots, A_n\}$, and let $B = (b_{kl})$ be its bond matrix. Two atoms, A_i and A_j , are constitutionally equivalent if there exists an automorphism⁵

$$\pi: \{1, \dots, n\} \rightarrow \{1, \dots, n\}$$

such that

$$(i) \pi(i) = j$$

$$(ii) AN(A_i) = AN(A_{\pi(i)}) \quad (AN = \text{atomic number})$$

$$FE(A_i) = FE(A_{\pi(i)}) \quad FE = (\text{no. of free electrons})$$

$$(iii) \sum_{k,l} |b_{kl} - b_{\pi(k)\pi(l)}| = 0$$

Even Carhart believes that our canonicalization algorithm could be proven correct if our perception of constitutional symmetry is accurate. Thus we want to limit our discussion here to demonstrate that we do correctly perceive constitutional symmetry. If two atoms are recognized as being constitutionally nonequivalent, one of these atoms can be given priority.

Carhart is wrong when he states¹ that we "let ad hoc rules substitute for true symmetry perception". These rules were chosen only to give priorities to nonequivalent atoms. They allow one to label a graph without destroying the symmetry of the graph. These rules are heuristic, but they were chosen to make the algorithm for canonically numbering as fast as possible.

To prevent any further misinterpretations we present here the algorithm for detecting constitutionally equivalent atoms in more detail.

ALGORITHM

Constitutionally equivalent atoms are recognized in step (7) of the function flow of our algorithm.³

Let A_i and A_j be two atoms of a molecule M which are checked for constitutional equivalence:

1. A necessary condition for constitutional equivalence of A_i and A_j is that they have the same NOON³ (number of outermost occupied neighbor sphere = shell of highest order).
2. All atoms of M are ordered into neighbor spheres (shells) of the atoms A_i and A_j .
3. The neighbor spheres of A_i and A_j are compared in their natural order.

This step in our algorithm was misunderstood by Carhart. When comparing the k th neighbor spheres $S_k(A_i)$ and $S_k(A_j)$ ($1 \leq k \leq \text{NOON}$), we do not cut the shells completely apart and treat them independently as he assumes. We rather conserve the relationships of atoms between two shells by taking into account bonds between spheres and previous assignments of atoms.

In our original paper³ in the paragraph between rule b and rule c we briefly described the pairwise comparison of all atoms in each neighbor sphere. This part of the algorithm which is in effect the assignment of atoms to each other and thus the detection of the automorphisms is now presented in more detail:

1. If $\text{NOON}(A_i) = \text{NOON}(A_j)$ then go to step 2; else atoms A_i and A_j are constitutionally nonequivalent.
2. $\pi(i) = j$; $\pi(l) = 0$ ($l = 1, \dots, i-1, i+1, \dots, n$); $k \leftarrow 0$.
3. $k \leftarrow k + 1$. If $k > \text{NOON}$ then stop: A_i and A_j are

constitutionally equivalent; else go to step 4.

4. Assign each atom $A_p \in S_k(A_i)$ to an atom $A_q \in S_k(A_j)$ if the following conditions are satisfied:

- (i) There was no assignment $\pi(p) = q$ for all $A_p \in S_k(A_i)$ in a previous run through step 4 for the same sphere k .
- (ii) $AN(A_p) = AN(A_{\pi(p)})$ (AN = atomic number); $FE(A_p) = FE(A_{\pi(p)})$ (FE = no. of free electrons)
- (iii) For $l = k - 1$ and $l = k$ do: For each atom $A_r \in S_l(A_i)$ which is bonded to A_p with bond order b there exists an atom $A_s \in S_l(A_j)$ with $\pi(r) = s$ which is connected to A_q with a bond of order b . In addition, A_r and A_s must have the same atomic number and the same number of free electrons.

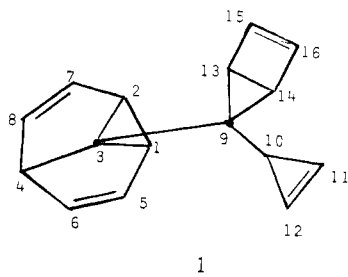
If all atoms of $S_k(A_i)$ can be assigned to atoms of $S_k(A_j)$ then go to step 3; else go to step 5.

5. $k \leftarrow k - 1$. If $k < 1$ then stop: A_i and A_j are constitutionally nonequivalent; else go to step 4.

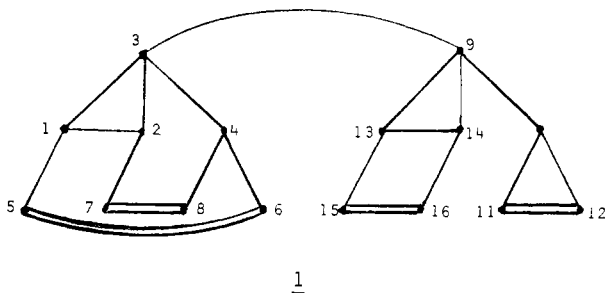
Note that the algorithm is always finite since there is only a finite number of different assignments between the atoms of $S_k(A_i)$ and $S_k(A_{\pi(i)}) \equiv S_k(A_j)$.

EXAMPLE

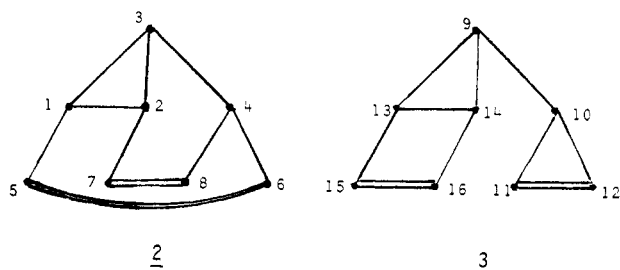
To further the understanding of our algorithm for perceiving constitutional symmetry, we illustrate its execution with an example. We select one of the "counterexamples" given by Carhart (his structure IV).



The main task is that we have to demonstrate that atoms A_3 and A_9 are constitutionally nonequivalent as Carhart claims that this cannot be perceived by our algorithm. First, let us redraw structure 1:



Although our algorithm considers all atoms in each neighbor sphere, a clearer picture of the main features in the perception of constitutionally equivalent or nonequivalent atoms is obtained when atoms A_3 and A_9 are disconnected and the two entities 2 and 3 thus resulting are compared.



This procedure conserves the essential problem for if our

algorithm works as Carhart claims it does, no differentiation between atoms A_3 and A_9 could be made. Thus it suffices to show that contrary to that claim our algorithm does perceive atoms A_3 and A_9 as being constitutionally nonequivalent.

$$S_1(A_3) = \{A_1, A_2, A_4\} \quad S_1(A_9) = \{A_{13}, A_{14}, A_{10}\}$$

$$S_2(A_3) = \{A_5, A_7, A_8, A_6\} \quad S_2(A_9) = \{A_{15}, A_{16}, A_{11}, A_{12}\}$$

Step 1: $NOON(A_3) = NOON(A_9)$

Step 2: $\pi(3) = 9$

Step 3: $k \leftarrow 1$

Step 4: $\pi(1) = 13$; $\pi(2) = 14$; $\pi(4) = 10$

Step 3: $k \leftarrow 2$

Step 4: $\pi(5) = 15$; $\pi(7) = 16$; $\pi(8) \neq 11$. The last assignment is not allowed, since A_8 is connected to A_7 , but A_{11} is connected to $A_{12} \neq A_{\pi(7)}$. $\pi(8) \neq 12$. Analogous reason as above: A_8 is connected to A_7 , but A_{12} is connected to $A_{11} \neq A_{\pi(7)}$. No other assignment possible.

Step 5: $k \leftarrow 1$

Step 4: $\pi(1) = 14$; $\pi(2) = 13$; $\pi(4) = 10$

Step 3: $k \leftarrow 2$

Step 4: $\pi(5) = 16$; $\pi(7) = 15$; $\pi(8) \neq 11$. See above.

$\pi(8) \neq 12$. See above.

Step 5: $k \leftarrow 1$

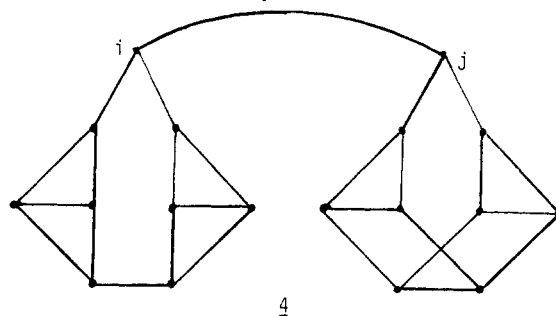
Step 4: $\pi(1) = 10$; $\pi(2) = 14$; $\pi(4) \neq 13$. The last assignment not allowed, since A_2 is not connected to A_4 , but $A_{14} = A_{\pi(2)}$ is connected to $A_{13} = A_{\pi(4)}$. $\pi(2) = 13$

$\pi(4) \neq 14$. Analogous reason as above.

Step 5: $k \leftarrow 0$ Stop: Atoms A_3 and A_9 are constitutionally nonequivalent.

The small number of steps even with such an example is indicative that our algorithm is fast. It suffices to say that also with the other "counterexample" given by Carhart (his structure V) our algorithm correctly recognizes the two atoms in question as being constitutionally nonequivalent.

In developing our algorithm we were aware of the difficulties involved in correctly perceiving constitutional symmetry. One of the structures we had kept in mind was structure 4.



Structure 4 is a regular asymmetric graph. In designing our algorithm we took care that atoms A_i and A_j are perceived as constitutionally nonequivalent.

CONCLUSION

We hope that we have now removed any point which could lead to a misinterpretation of our algorithm. In demonstrating that this algorithm correctly perceives constitutional symmetry we have also supported our proof that our canonicalization algorithm leads to a unique numbering of the atoms of a molecule.

REFERENCES AND NOTES

- (1) R. E. Carhart, *J. Chem. Inf. Comput. Sci.*, **18**, 108 (1978).
- (2) C. A. Shelley and M. E. Munk, *J. Chem. Inf. Comput. Sci.*, **17**, 110 (1977).
- (3) C. Jochum and J. Gasteiger, *J. Chem. Inf. Comput. Sci.*, **17**, 113 (1977).
- (4) J. Gasteiger and C. Jochum, unpublished results.
- (5) F. Harary, "Graph Theory", Addison-Wesley Publishing Co., Reading, Mass., 1972.