ESCHER−A Computer Program for the Determination of External Rotational Symmetry Numbers from Molecular Topology

W. P. Walters[†] and S. H. Yalkowsky*[,‡]

Department of Chemistry and College of Pharmacy, The University of Arizona, Tucson, Arizona 85721

Received November 7, 1995[⊗]

An algorithm for determining the external rotational symmetry number of a molecule from a SMILES[1] string has been developed. ESCHER operates by first locating the center or centers of graphical symmetry for the molecule and the equivalence classes of atoms connected to the center or centers. The center(s) of graphical symmetry is the atom(s) which is(are) most symmetrical with respect to the connections to other atoms. These are then used to calculate the symmetry number, σ.

INTRODUCTION

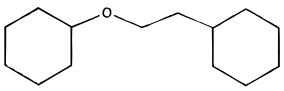
A rapid automated method for determining molecular symmetry can contribute significantly to studies of the behavior of crystalline materials. The molecular external rotational symmetry number, σ, is the number of indistinguishable positions that can be obtained by rigidly rotating a molecule about its center of mass.[2] The greater the rotational symmetry number, the higher the probability of the molecule being in its correct orientation to fit into its crystal lattice or a macromolecular receptor. Symmetry numbers have been used in the prediction of physical properties such as melting point and the entropy of melting.[3,4]

Muller and co-workers[5] presented a method for determining the external symmetry number based on a factored representation of molecules and free radicals. This method begins by locating a molecule's graphical center or "focus" by recursively removing monovalent atoms until no monovalent atoms remain. The remaining atom or atoms are treated as the focus. The layers which are removed during the determination of the focus are then replaced and identical branches are identified. While the above method works well in many cases, it breaks down for structures which contain ring systems connected by chains. In such cases, one can arrive at a focus which may or may not contain centers of symmetry. Figure 1 shows an example of one such case.

This report describes a rapid and more general method for determining σ which has been implemented in the computer program ESCHER. ESCHER uses Golender's[6] graph potentials method to locate the graphical center(s) of symmetry and identify topologically equivalent atoms. The graph potentials method requires only a few simple matrix operations which can be implemented using any one of a number of readily available subroutine libraries. Once the graphical center of symmetry has been located and topological equivalencies have been assigned, ESCHER uses a few simple rules to determine σ.

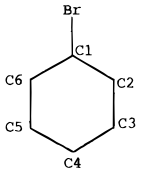
GRAPH POTENTIALS

A brief overview of the graph potentials method is provided here. For a more thorough discussion, the reader should refer to the original work of Golender.[6]



C2(CCCCC2)OCCC1CCCCC1

Figure 1. A structure whose graphical center cannot be located by Muller's method.



	C1	C2	C3	C4	C5	C6	Br
C1	4	−1	0	0	0	−1	−1
C2	−1	3	−1	0	0	0	0
C3	0	−1	3	−1	0	0	0
C4	0	0	−1	3	−1	0	0
C5	0	0	0	−1	3	−1	0
C6	−1	0	0	0	−1	3	0
Br	−1	0	0	0	0	0	2

The Matrix **G**

	C1	C2	C3	C4	C5	C6	Br
C1	0.37	0.14	0.06	0.04	0.06	0.14	0.18
C2	0.14	0.44	0.17	0.07	0.04	0.06	0.07
C3	0.06	0.17	0.45	0.17	0.07	0.04	0.03
C4	0.04	0.07	0.17	0.45	0.17	0.07	0.02
C5	0.06	0.04	0.07	0.17	0.45	0.17	0.03
C6	0.14	0.06	0.04	0.07	0.17	0.44	0.07
Br	0.18	0.07	0.03	0.02	0.03	0.07	0.59

The Matrix **H**

C1	18.00
C2	12.00
C3	12.00
C4	12.00
C5	12.00
C6	12.00
Br	35.00

The Vector **C**

C1	18.43
C2	14.50
C3	13.07
C4	12.71
C5	13.07
C6	14.50
Br	26.71

The Graph Potentials **U**

Figure 2. The graph potentials for bromobenzene.

The first step in determining graph potentials is the creation of a hydrogen suppressed adjacency matrix, **G**. In this matrix each diagonal element (i,i) is set equal to 1 + the valence of atom i. Each off-diagonal element (i,j) is set equal to −1 if atoms i and j are connected, or set equal to 0 if atoms i and j are not connected. The matrix, **G**, is then inverted to create the matrix, **H**. The graph potentials, **U**, are calculated by multiplying **H** by a column vector, **C**, consisting of atomic valences.

The generation of the **G** and **H** matrices, the column vector, and the graph potentials for bromobenzene are illustrated in Figure 2. Note that C1 is connected to C2, C6. The Br and has a valence of 3 and a diagonal matrix entry of 4. The remaining carbons have valences of 2 and matrix entries of 3, and the bromine has a valence of 1 and a matrix entry of 2. Finally the off diagonal entries for C1 are −1 for C2, C6, and Br and 0 for all other atoms.

[†] Department of Chemistry.
[‡] College of Pharmacy.
[⊗] Abstract published in *Advance ACS Abstracts,* May 15, 1996.

S0095-2338(95)00278-2 CCC: $12.00

Table 1. Rules for Determining σ

no. of attached atoms	no. of attached types	hybridization	σ	no. of attached atoms	no. of attached types	hybridization	σ
2	1	sp, sp^2, sp^3	2	3	3	sp^3	1
2	2	sp,sp^2,sp^3	1	4	1	sp^3	12
3	1	sp^2	6	4	2 (3 + 1)	sp^3	3
3	2	sp^2	2	4	2 (2 + 2)	sp^3	2
3	3	sp^2	1	4	3	sp^3	1
3	1	sp^3	3	4	4	sp^3	1
3	2	sp^3	1				

Atoms that are located closer to the graphical center of the molecule will have a higher graph potential than atoms at the outer edges. In the ESCHER program, the atom or atoms with the highest graph potential are the graphical center or centers of symmetry. In the case of bromobenzene the center of symmetry is C1.

The graph potentials method is also used to determine which atoms in a molecule are topologically equivalent. Atoms with the same graph potential are considered to be topologically equivalent. When determining the topological equivalence of atoms, the graph potentials method is augmented by multiplying the atomic number for atom i by element i in the vector, **C**.

ALGORITHM FOR DETERMINING σ

The method by which ESCHER calculates σ consists of the following six steps, all of which require only the connection table provided by a SMILES string.

1. Determine if the Molecule Is Flexible. In the ESCHER program, flexible molecules are considered to have a σ of 1. A molecule is considered flexible if it possesses more than two adjacent rotatable heavy atom torsion angles or a ring having greater than six atoms.

2. Locate the Graphical Center(s) of Symmetry. The graphical center of symmetry for a molecule is located using the graph potentials method. Graph potentials are assigned to all non hydrogen atoms in the molecule and the atom or atoms with the highest graph potential are assigned as the graphical center or centers of symmetry.

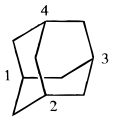
3. Identify Topologically Equivalent Atoms. Topologically equivalent atoms are determined using the graph potential calculated in the previous step. Atoms with the same graph potentials are considered to be topologically equivalent.

4. Determine the Coordination Number and the Hybridization State of the Graphical Centers(s) of Symmetry. The coordination number and hybridization state of each atom in the molecule can be easily determined by examining the atom types and bond orders in the SMILES string.

5. Determine the Symmetry Number from Table 1. Table 1 lists rules for predicting σ_i based on three characteristics of each graphical center of symmetry: the number of atoms attached, the number of topologically distinct atoms attached, and the geometry. The number of attached atoms and the geometry are determined from the smiles string. The number of topologically distinct atoms attached to the graphical center of symmetry is determined in step 3 using the graph potentials method. The few cases which are exceptions to these rules are presented below.

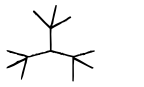
6. Account for Multiple Graphical Centers of Symmetry. In many cases, ESCHER will determine that a molecule has more than one graphical center of symmetry. If a molecule possesses multiple equivalent graphical centers of symmetry, the final value of σ is the sum of the σ_i values for all the graphical centers of symmetry. Figure 3 shows adamantane, which contains four graphical centers of symmetry (labeled 1−4). Each of these graphical centers has three attached non-hydrogen atoms. Since the number and type of the attached atoms are equivalent, each graphical center has a σ of 3. The σ of adamantane is then calculated as 3 × 4, or 12. Similarly *p*-dichlorobenzene has two graphical centers of symmetry and a σ of 4.

If a molecule possesses multiple nonequivalent graphical centers of symmetry, σ_i values are calculated for each of the centers and the lowest σ_i value is used.



C12CC(CC(C3)C2)CC3C1

Figure 3. The graphical centers of symmetry for adamantane.



CC(C)(C)C(C(C)(C)C)C(C)(C)C

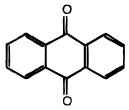
σ = 3

CC(C)(C)CC(C)(C)C

σ = 2

C12=CC=CC(CC3)=C1C3=CC=C2

σ = 2

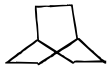


O=C(C1=CC=CC=C12)C3=C(C=CC=C3)C2=O

σ = 4

C12=CC=CC=C1NC=C2

σ = 1



C12CCC(CC2)CC1

σ = 6

Figure 4. The σ values calculated by ESCHER for six examples.

Table 2. Special Rules for Spiro Compounds

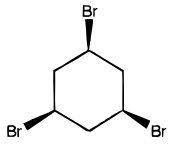
no. of attached types	σ	no. of attached types	σ
1	4	3	1
2	2	4	1

SCOPE AND LIMITATIONS

ESCHER has been tested on a wide variety of cyclic and acylic organic compounds. The σ values calculated for selected compounds are shown in Figure 4.

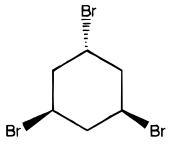
Since ESCHER relies on a purely topological description of the molecule there are a few cases where the program is unable to correctly calculate σ. Instead it provides the maximum symmetry number that is possible for a compound having a particular set of connections among its atoms. The program is designed to warn the user in these cases. In such cases it is necessary to use a program such as Wizard,[7] Corina,[8] or Concord[9] to generate a three-dimensional structure for the molecule and compare atom to atom mappings. There are two general cases where the rules used to calculate σ are not applicable. These cases are described below.

1. Stereochemistry and Conformational Information. The ESCHER program currently has a few limitations due



BrC1CC(Br)CC(Br)C1

σ = 3



BrC1CC(Br)CC(Br)C1

σ = 1

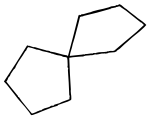
Figure 5. Two isomers of 1,3,5-tribromocyclohexane with the same SMILES string.

to its reliance on the original SMILES notation. Since SMILES does not contain the specification of stereochemistry, ESCHER cannot distinguish among stereoisomers. Figure 5 shows two isomers of 1,3,5-tribromocyclohexane. Although both isomers have the same SMILES string, the cis isomer has a σ of 3, while the trans isomer has a σ of 1. Similarly ESCHER cannot distinguish between the cis and chair forms of cyclohexane and benzene because they have similar connections among their atoms. In cases where the stereochemical or conformational ambiguity exists, ESCHER calculates the maximum symmetry for the SMILES string, and the user is warned that the predicted σ may not be accurate.

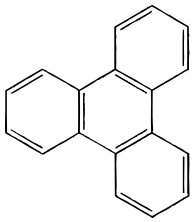
2. Compounds with Graphical Centers of Symmetry in More Than One Ring. ESCHER considers only the type of atoms that comprise noncenters of symmetry. Since it does not consider connections between these noncenters of symmetry, it does not account for the reduction in symmetry produced by the formation of rings containing noncenters of symmetry. This is true for both bond fusion and spiro fusion as shown in Figure 6. While it gives correct symmetry numbers for anphthalene, anthracene, coronene, and most other polycyclic aromatic hydrocarbons, ESCHER overestimates the symmetry numbers of both triphenylene and spirononane.

CONCLUSION

The ESCHER program provides a rapid and accurate means of determining the rotational symmetry number for rigid molecules from a topological description such as a SMILES string. Since the current ESCHER program is limited in that it cannot fully deal with stereochemistry or conformation the σ values calculated are maximum values.



C12(CCCC2)CCCC1



C1(C=CC=C3)=C3C(C=CC=C4)=C4C2=C1C=CC=C2

Figure 6. Two compounds with graphical centers of symmetry that are ring fusion sites.

REFERENCES AND NOTES

(1) Weinenger, D. SMILES, a Chemical Language for Information Systems. 1. Introduction to Methodology and Encoding Rules. *J. Chem. Inf. Comput. Sci*. **1988**, *28*, 31−36.
(2) Dannenfelser, R. M.; Surendran, N.; Yalkowsky, S. H. Molecular Symmetry and Related Properties. *SAR QSAR Environ. Res*. **1993**, *1*, 273−292.
(3) Yalkowsky, S. H.; Dannenfelser, R. M.; Myrdal, P.; Simamora, P. Unified Physical Property Estimation Relationships (UPPER). *Chemosphere* **1994**, *28*, 1657−1673.
(4) Yalkowsky, S. H.; Myrdal, P.; Dannenfelser, R. M.; Simamora, P. UPPER II: Calculation of Physical Properties of the Chlorobenzenes. *Chemosphere* **1994**, *28*, 1675−1688.
(5) Muller, C.; Scacchi, G.; Come, G. M. A Topological Method for Determining the External Symmetry Number of Molecules. *Comput. Chem*. **1991**, *15*, 17−27.
(6) Golender, V. E.; Drboglav, V. V.; Rosenblit, A. B. Graph Potentials Method and Its Application for Chemical Information Processing. *J. Chem. Inf. Comput. Sci*. **1981**, *21*, 196−204.
(7) Dolata, D. P.; Leach, A. R.; Prout, K. WIZARD, AI in Conformational Analysis. *J. Comput. Aided Molecular Design* **1987**, *1*, 73−75.
(8) Gasteiger, J.; Sadowski, C. R. J. Automatic Generation of 3D-Atomic Coordinates for Organic Molecules. *Tetrahedron Comput. Method*. **1992**, *3*, 537−547.
(9) Pearlman, R. S. Rapid Generation of High Quality Approximate 3-D Molecular Structures. *Chem. Design Automation News* **1987**, *2*, 5−6.

CI950278O