

A Bond Flexibility Index Derived from the Constitution of Molecules[†]

Claus-Wilhelm von der Lieth,* Klaus Stumpf-Nothof, and Ursula Prior

German Cancer Research Center, Central Spectroscopy Department,
Im Neuenheimer Feld 280, D-69120 Heidelberg, Germany

Received September 14, 1995[⊗]

The concept of flexibility is a widely used qualitative term in chemistry. However, only a few attempts have been made to quantify this structural property. The basic assumptions used in a description of molecular flexibility are outlined, and their consequences are discussed. A local description is needed to handle molecules which exhibit both flexible and rigid parts. The assumption is made that local molecular flexibility can be described by defining structural features which exhibit an intrinsic flexibility which is largely independent of directly connected functional groups. A flexibility index which can be defined for each bond in a molecule is presented. It allows one to describe local differences between flexible and rigid regions within one molecule. Some potential applications of the bond flexibility index will be discussed.

1. INTRODUCTION

The concept of flexibility and/or rigidity of a complete molecule or of parts of it is widely used in a qualitative sense in chemistry^{1,2} and is often helpful in the discussion of structural features which influence chemical and biological effects of a compound or its interaction with other molecules. Flexible molecules show a higher isothermal compressibility than more rigid molecules, and they tend to be more difficult to crystallize. The physical properties of organic polymers are strongly influenced by the flexibility of the main chains. In receptor-ligand interactions the flexibility or rigidity of either molecule can greatly influence their adaptability for complex formation and the resulting binding constants.

Although the concept of flexibility and/or rigidity is used in many discussions, only a few attempts to quantify this structural property can be found in the literature. Luisi³ proposed a scale of conformational rigidity for acyclic hydrocarbons without heteroatoms on the basis of the three-state isomerism of C-sp³ atoms. Kier⁴⁻⁷ developed a more general flexibility index Φ for the quantification of molecular flexibility. It is based on molecular size, branching, rings, and heteroatom content of a molecule. Φ is stored as a global molecular property in the CAS property data base⁸ and can be used for similarity searching. In addition, local simple (LS) and global simple (GS) flexibility indices have been included in the molecular property data of the CAS registry files⁹ to facilitate searches for molecular properties which are closely related to 3D structures.

Clark et al.² used the number of rotatable bonds in a structure as a simple flexibility measure. They found this simple measure performed as well as the Kier or the GS flexibility index to rank the outputs of several pharmacophoric pattern searches of 3D structures.

All four approaches mentioned so far are based solely on the constitution of the molecule. More recently attempts have been made to use molecular dynamics simulations to derive global or local descriptions for protein surfaces or macromolecular chains.^{10,11}

Because many molecules of interest such as natural products, peptides, and oligosaccharides often show both structurally rigid and flexible regions, an appropriate bond flexibility index may be useful for estimating the number of different conformations of a molecule which need to be evaluated. One global flexibility index for such molecules may be misleading and is certainly not able to describe local differences in flexibility within a molecule. A bond flexibility index may also be useful in overcoming the problems which can occur when ligand molecules taken from a data base are fitted automatically into the binding site of a target receptor and only one, possibly "wrong", conformation of a flexible molecule has been stored.

The aim of our study was to develop a semiquantitative flexibility index for each bond in a molecule, based on the constitution of the molecule alone. The algorithm for index calculation should be fast so that it can be applied to a large number of molecules in a data base application. Ideally, the bond flexibility index should be transferable between molecules of different size and shape.

2. MOLECULAR FLEXIBILITY

Kier's molecular flexibility index⁴⁻⁷ and the LS and the GS flexibility description of CAS⁹ are the best known approaches for the quantification of molecular flexibility. Both are based entirely on the constitution of molecules with all hydrogens omitted. The basic philosophy of both indices can be summarized in the following way: a completely flexible molecule is assumed to be an endless chain of C-sp³ atoms. Its flexibility index is infinite. The structural features mitigating against complete flexibility are as follows: fewer atoms, the presence of rings, the presence of branching, and, in Kier's approach only, the presence of atoms with covalent radii different from that of C-sp³. This list of factors which reduce flexibility also gives the order of their importance: the number of atoms in a molecule is the parameter which influences the flexibility most, followed by the number of branches and the number of bonds to heteroatoms. Ring systems are taken into account indirectly through their number of branches or in terms of nonrotatable bonds.

2.1. Kier's Flexibility Index Φ . Kier⁴⁻⁷ has suggested a simple and fast way for calculating a flexibility parameter

[†] Keywords: molecular flexibility, bond flexibility index, computational chemistry.

[⊗] Abstract published in *Advance ACS Abstracts*, May 1, 1996.

Φ for an entire molecule. It does not require experimental measurements and is based on the previously derived shape indices $^1\kappa$ and $^2\kappa$, which he applied in QSAR studies.

$$\Phi = {}^1\kappa^2\kappa/N_A \quad (1)$$

$${}^1\kappa = (N_A + a - 1)(N_A + a - 2)/(p_1 + a)^2 \quad (2)$$

$${}^2\kappa = (N_A + a - 1)(N_A + a - 2)/(p_2 + a)^2 \quad (3)$$

$$a = \sum_j [(r_j/r(\text{sp}^3)) - 1] \quad (4)$$

N_A is the total number of skeleton atoms (omitting hydrogens), r_j is the covalent radius of atom j , $r(\text{sp}^3)$ is the covalent radius of a C-sp³ atom, p_1 is the number of paths of length one in a hydrogen-suppressed skeleton structure, and p_2 is the count of contiguous two-bond pathways passing through each skeleton atom. The first-order κ shape index $^1\kappa$ encodes the number of atoms in a molecule and the relative cyclicity. The second-order $^2\kappa$ index describes the branching and spatial density of a molecule.

2.2. The CAS LS and GS Flexibility Indices.⁹ In order to be able to search for molecular properties in the CAS registry files which are closely related to the 3D structure both local and global molecular flexibility indices have been developed. The local simple (LS) flexibility index between atom i and j in a molecule is given as

$$\text{LS}_{ij} = \text{SPN} - (\text{NRB} + 0.75\text{BX} + 0.50\text{BY})/2 \quad (5)$$

where SPN is number of centers in the shortest path between the atoms i and j , NRB is the number of nonrotatable bonds, BX is the number of branching atoms with four attached non-H atoms, and BY is the number of branching atoms with three attached atoms.

An acyclic, nonbranched chain of C-sp³ atoms is regarded as completely flexible and LS equals simply the number of atoms involved in the path, e.g., LS₁₇ for *n*-heptane is 7. The number of nonrotatable bonds and the number of atoms with four and three attached skeletal atoms reduce the flexibility of the path. These values are appropriately scaled and subtracted from the value for a completely flexible chain of that length.

The local flexibility index is calculated between each pair of atoms in the molecule for the shortest possible path between two atoms. For any given structure with N nonhydrogen atoms, there will be $N(N-1)/2$ such local indices and the mean of these LS_{*ij*} gives the global flexibility index GS for the structure.

$$\text{GS} = 2/N(N-1) \sum_{i=1}^{N-1} \sum_{j=i+1}^N \text{LS}_{ij} \quad (6)$$

2.3. Some Consequences of the Definition of the Molecular Flexibility Index. The approaches described above follow the same philosophy, and in this section we discuss some of the consequences of defining molecular flexibility in such a way. Table 1 lists the flexibility indices Φ for cyclohexane and some substituted cyclohexanes. According to Kier's concept of molecular flexibility substituted rings are generally regarded as more flexible than the corresponding unsubstituted ring. This is due to the basic

Table 1. The Kier Flexibility Index Φ for Substituted Cyclohexanes

molecule	Φ
cyclohexane	1.54
methylcyclohexane	1.74
1,4-dimethylcyclohexane	1.90
ethylcyclohexane	2.38
1,4-diethylcyclohexane	3.20

Table 2. The Kier Flexibility Index Φ for Simple Ring Systems with Different Degrees of Unsaturation

ring size	saturated ring	one double bond	two double bonds	three double bonds	four double bonds
3	0.09	0.06			
4	0.42	0.26	0.23		
5	0.92	0.74	0.63		
6	1.54	1.23	1.18	0.98	
7	2.24	1.98	1.83	1.59	
8	3.01	2.72	2.55	2.28	2.02
9	3.82	3.50	3.20	2.91	2.64
10	4.66	4.32	4.00	3.69	3.39
11	5.53	5.17	4.83	4.51	4.19

Table 3. The Kier Flexibility Index Φ for Simple Acyclic Molecules with Different Degrees of Unsaturation

molecule	Φ
<i>n</i> -butane	3.00
2-butene	2.74
butadiene	2.31
<i>n</i> -hexane	5.00
1,4-hexadiene	4.09
2,4-hexadiene	4.09
2,5-hexadiene	4.09
1,3,5-hexatriene	3.67
<i>n</i> -octane	7.00
octatetraene	5.10

assumption that flexibility always increases with the number of atoms. Experimentally it is well-known that the conformation of a ring is generally stabilized by substitutions. Experience would normally let us to assume that methylcyclohexane and 1,4-dimethylcyclohexane are less flexible than cyclohexane itself.

Since the Φ index considers the neighborhood of a given atom only to a depth of one and two bonds, the rigidity introduced by conjugated double bonds or aromatic systems cannot be described adequately by this concept. Table 2 displays Φ values for simple ring systems with different degrees of unsaturation. It is generally accepted that benzene ($\Phi = 0.98$) is a very rigid molecule and should be even more rigid than cyclobutane ($\Phi = 0.42$). 1,3-Cycloheptadiene and 1,4-cycloheptadiene certainly exhibit different conformational behaviors, but both molecules have the same Φ index of 1.83.

The molecular flexibility of conjugated systems is generally overestimated. The effects due to the loss of conformational freedom through conjugation effects can best be demonstrated for acyclic conjugated systems. Table 3 presents the Φ index for simple acyclic molecules with different degrees of unsaturation. The flexibility of molecules with conjugated double bonds compared to the corresponding saturated molecules seems to be overestimated. The reduction of conformational freedom through conjugation cannot be treated by Kier's method.

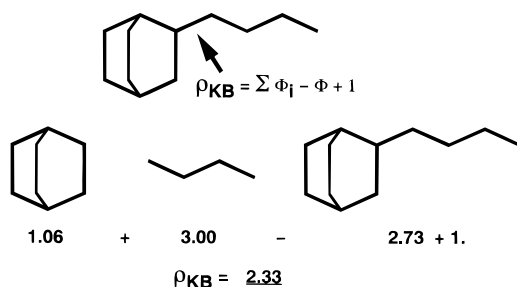


Figure 1. Calculation of the Kier bond rigidity index ρ_{KB} according to eq 7. The bond rigidity is defined as the difference of the Φ_{Kier} of the whole molecule minus the sum of Φ_{Kier} for the two parts which are formed when breaking a bond.

3. BOND FLEXIBILITY

Some of the factors involved in determining the flexibility of a bond within a molecule are its bond order, the nature of the atoms forming the bond, its participation in one or more cyclic structures, the branching of the adjacent atoms, and the location of a "rigid" bond or cyclic structures in the neighborhood. A single acyclic bond formed by two C-sp³ atoms is normally regarded as "freely" rotatable (three allowed rotamers) at room temperature. Cyclic structures are considered to be much less flexible. (Poly-) aromatic ring systems as well as double- and triple-bonds are normally regarded as rigid.

3.1. Kier Bond Rigidity Index ρ_{KB} Based on Kier's Molecular Flexibility Φ . Despite the shortcomings of the molecular flexibility index Φ discussed above, we decided, as a first approach, to base the bond flexibility and/or rigidity on the original parametrization of Kier since the Kier index is well-known, widely accepted, easy to implement, and very fast.

The flexibility and/or rigidity of a particular bond is given by the difference between the flexibility of the whole molecule (Φ) and the sum of the flexibilities (Φ_i) of the molecular fragments obtained by breaking the bond of interest (see Figure 1).

$$\rho_{KB} = \sum \Phi_i - \Phi + 1 \quad (7)$$

where ρ_{KB} is the Kier bond rigidity, Φ_i is the Kier flexibility parameter for molecular fragment i , and Φ is the Kier flexibility for the intact molecule. Thus, ρ_{KB} is a rigidity and not a flexibility index since it represents the increase in flexibility of the fragments relative to the parent molecule. This difference increases as the rigidity of the broken bond increases. The term +1 in eq 7 results in a minimum of $\rho_{KB} = 0$ for a single bond in an alkyl chain.

When the rigidity of a multiple bond is calculated, it is first transformed to single bonds by adding the additional partner atoms and single bonds at each site according to the CIP rules.¹² This approach does not work for aromatic and conjugated bonds. Therefore, such bonds must be considered as alternating single and double bonds. This is the reason why topological equivalent bonds are sometimes assigned different rigidity indices (Figure 2). Keeping this in mind the above procedure works satisfactorily within a molecule.

The ρ_{KB} values are generally in the agreement with chemical experience: fused ring atoms and multiple bonds in a ring are the most rigid parts, large rings are more flexible than small ones, and the flexibility of chains is greater than that of rings and mainly determined by the number of

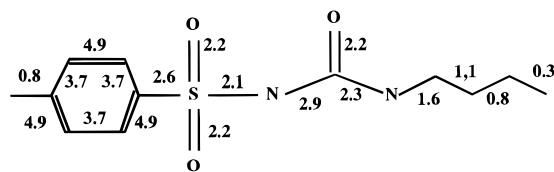


Figure 2. Kier bond rigidity index ρ_{KB} of *N*-(butylamino)carbonyl-4-methylbenzene sulfonamide.

branches and attachments to more rigid parts of the molecule (see Figure 2).

However, since the value of Kier's index Φ depends on the number of atoms in a molecule or fragment, the ρ_{KB} indices for homologous bonds in different molecules may differ. Relative differences in flexibility within one molecule, or in a series of structurally closely related molecules, can be described in a consistent way by this approach. However, the ρ_{KB} values are not transferable between different molecules, especially when the number of atoms differ considerably.

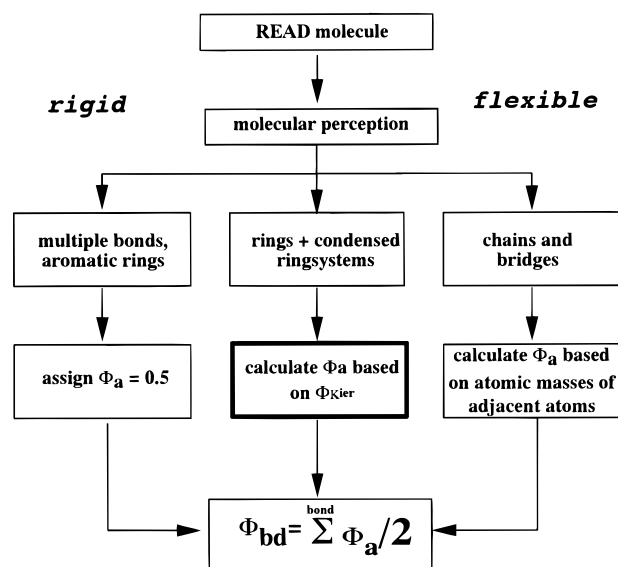
3.2. Bond Flexibility Index Φ_{BD} . Our approach assumes that a molecule can be described by structural features which exhibit intrinsic flexibilities, which are largely independent of directly connected functional groups or atoms. We define six basic structural features: aromatic rings, multiple bonds, conjugated systems, simple rings, chains, and bridges.

Since we intend to describe the flexibility measure as a membership function in the sense of fuzzy logic theory,^{13,14} we define a scale for Φ_{BD} between 0 (no flexibility) and 10 (completely flexible). The bond rigidity index is defined as $\rho_{BD} = 10 - \Phi_{BD}$.

The Algorithm. As a first step the molecule must be "perceived" in order to assign the structural features as defined above (see Figure 3 (parts a and b)). In a subsequent step a flexibility index Φ_a for each atom is calculated. The bond flexibility index Φ_{BD} is then given as the mean of the atom flexibility indices Φ_a of the two atoms forming the bond.

Calculation of Φ_a for Each Structural Feature. The molecular perception is accomplished with a routine implemented in our SPEKTREN II software.^{15,16} The procedure to assign flexibility value Φ_a for each individual atom is the following: for all atoms forming a double or triple bond or which are part of an aromatic ring a flexibility index of $\Phi_a = 0.5$ is assigned. For rings and condensed ring systems Kier's molecular flexibility index Φ with a correction for conjugated systems is used to calculate the flexibility index Φ_a for each atom in a ring. If condensed ring systems are present, Φ_a is also calculated for each condensed ring system. Each atom in a condensed ring system is assigned the smaller of the two calculated Φ_a values (individual or condensed rings). For example in the case of fused 6- and 9-membered ring systems, the atoms of the 9-membered ring are assigned the Φ_a value of the condensed system ($\Phi_a = 2.7$ or $\rho_a = 6.3$), whereas the atoms of the 6-membered ring are assigned the Φ_a value of the isolated 6-membered ring ($\Phi_a = 1.5$ or $\rho_a = 8.5$) (see Figure 5d). This procedure reflects that the flexibility of the rings in a condensed ring system is reduced compared to an isolated ring of that size.

Since the conformational freedom of the complete ring is reduced by the number of substitutions, we subtract from Φ_a for each atom in a ring or condensed ring a value of 0.2 units per substitution.



Φ_{bd} = Bond flexibility index
 Φ_{Kier} = Kier's molecular flexibility
 Φ_a = flexibility index for each atom

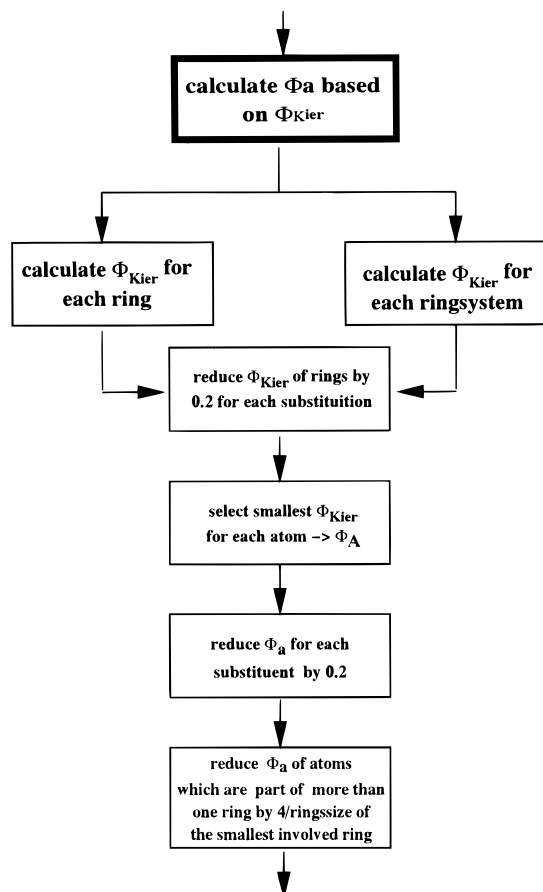


Figure 3. (a, top) Algorithm for calculating the bond flexibility index Φ_{BD} and (b, bottom) special treatment for the calculation of Φ_a of atoms in rings and ring systems.

The flexibility of an atom which is part of more than one ring will be considerably reduced. Therefore, the Φ_a index for such an atom is reduced by an amount defined as $4/RS$, where RS is the size of the smallest of the participating rings.

In a third step the flexibility for atoms in chains and in bridges is established. We found that it is not appropriate to apply Kier's molecular flexibility index Φ of Kier to

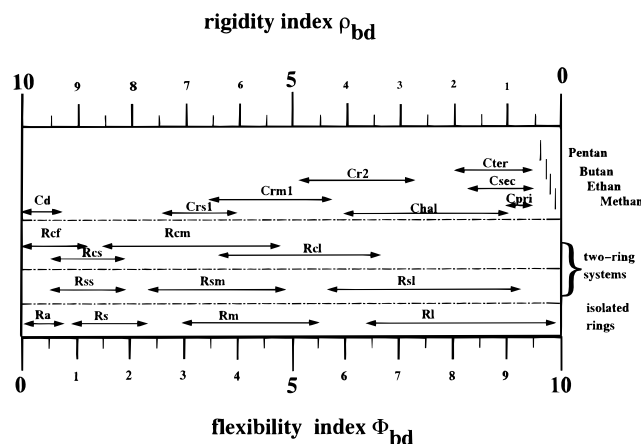


Figure 4. Overview of ρ_{BD} and Φ_{BD} value for characteristic structural features. **R**: rings; **C**: chains (carbon atoms); **a**: aromatic bonds, double bonds, ring size < 5 ; **c**: condensed; **cf** at point of condensation; **s**: small ($n = 4-8$); **m**: medium ($n = 8-11$); **l**: large ($n > 11$); **d**: double or triple bonds; **hal**: carbon halogen bond; and **r(s,m,l)**: single bond connected to a ring.

chains and bridges. In our approach only the hybridizations of the atom of interest and its directly connected atoms are taken into account. As mentioned already above, C-sp and C-sp² atoms are always assigned a flexibility index Φ_a of 0.5. C-sp³ atoms in chains are regarded as completely flexible and with a $\Phi_a = 10$. This value is corrected by subtracting the sum of the atom masses of all atoms in the next four adjacent shells, divided by 100 in order to maintain limits of 0–10 for the defined flexibility scale.

The bond flexibility index Φ_{BD} is calculated as the mean of the Φ_a for the two atoms forming a given bond. Figure 4 gives an overview of the characteristic regions of Φ_{BD} and ρ_{BD} ($\rho_{BD} = 10 - \rho_{BD}$) for some characteristic structural features. Figure 5 displays the bond rigidity index ρ_{BD} for some representative types of familiar structures: acyclic linear and branched molecules as well as cyclic and bicyclic structures of increasing complexity.

The Φ_{BD} index depends on the nature of local structural features. For rings it shows the same trend as the molecular flexibility index: the flexibility measure increases with the ring size but is reduced by conjugation effects, substitutions, fusion of rings, and heteroatoms. For chains short unbranched structures of C-sp³ bonds are regarded as the most flexible part of a molecule, while the flexibility of an aliphatic chain is reduced by an increasing number of connected atoms, branch points, adjacent rigid structural feature, and heteroatoms.

4. SUMMARY AND DISCUSSION

We present a flexibility index which can be calculated for each bond in a molecule and which describes the local differences that are expected between the flexible and rigid parts of a molecule. Since the algorithm is based on the molecular constitution alone, it is very fast. Several thousand molecules can be treated in a few seconds of CPU time using state-of-the-art workstations. Thus the algorithm is suitable for data base applications. The bond flexibility index is transferable between molecules of different size and molecular shape.

In general, the factors involved in determining the rigidity of a bond in a molecule are its bond order, its participation in one or more cyclic structures, the branching of the adjacent

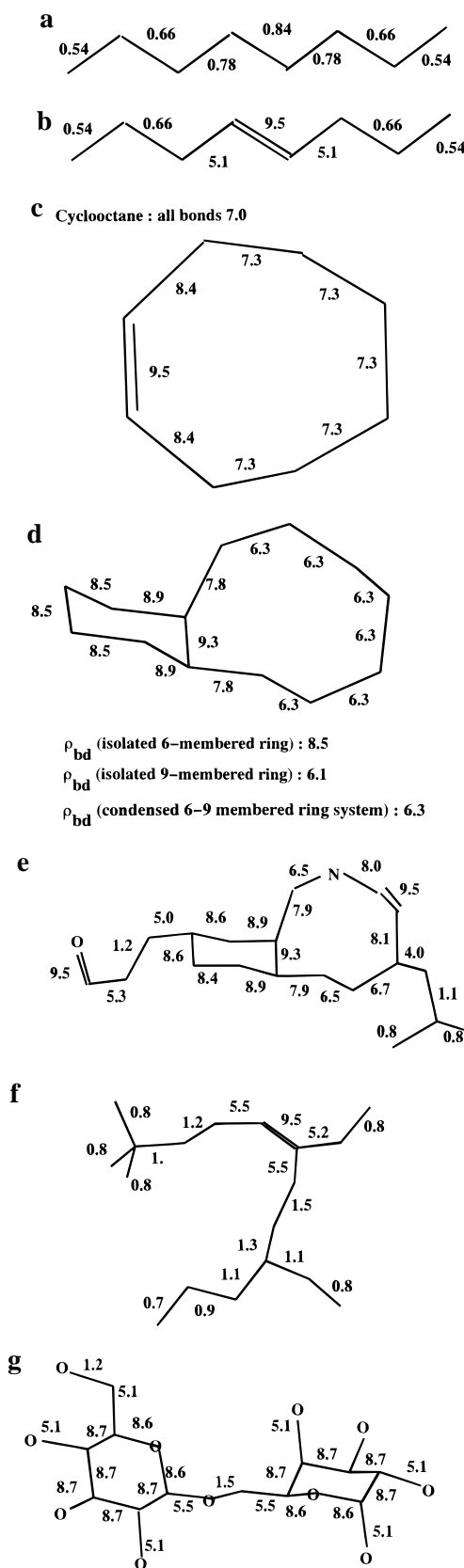


Figure 5. Bond rigidity indices ρ_{BD} for several molecules exemplifying various common structure elements: (a) *n*-octane, (b) 4-octene, (c) cyclooctene, (d) condensed 6-9-membered rings, (e) condensed 6-9-membered rings with substitutions, (f) complicated branch structure, and (g) α DGlC(1-6) α DGal.

atoms, and the location of a "rigid" bond or cyclic structure in the neighborhood. Six basic structural features have been

considered: multiple bonds, aromatic rings, condensed ring systems, single rings, chains, and bridges.

Our bond flexibility index can be used for all applications where an automatic decision must be made as to whether differences in internal coordinates should be regarded as a new conformation of a molecule. For example, small differences in the dihedral angles in rigid parts (e.g., in condensed ring systems) may represent a significant new structure, whereas large differences in the torsion angles in a flexible part (e.g., long aliphatic chains) should not be regarded as representing a significant new conformation.

With the help of the bond flexibility index decisions concerning the number of different conformations (3D structures) of one molecule which should be stored in a data base can be treated in a reproducible and automatic way. Moreover, the bond flexibility index provides an objective way to determine the reliability of 3D molecular structures generated with different 2D-to-3D conversion programs in comparison with experimental data. Normally the RMS value for the distances between corresponding atoms in two overlaid structures are used for this task.¹⁷ But this may produce misleading results when structures containing flexible regions are compared.

The bond flexibility index can also be used to detect automatically significant conformational changes through a large series of 3D structures generated during a molecular dynamics or Monte Carlo simulation, for example. It is a time consuming task, normally done manually by the investigator, to decide which of the many generated conformations belong to a specific conformational cluster and which to another significantly different family. If the bond flexibility index indicates that a given bond is highly flexible, only large differences in its torsion angle justify the definition of a new conformation.

ACKNOWLEDGMENT

We gratefully acknowledge funding from Chemical Concepts, Weinheim, Germany in support of the research described here. The authors would like to thank William E. Hull for inspired and valuable discussion.

REFERENCES AND NOTES

- (1) Oshiro, C.; Kuntz, I.; Dixon, J. Flexible ligand docking using a genetic algorithm. *J. Comput.-Aided Mol. Design* **1995**, *6*, 113-130.
- (2) Clark, D.; Willet, P.; Kenny, P. Pharmacophoric pattern matching in files of three-dimensional chemical structures: Implementation of flexible searching. *J. Mol. Graphics* **1993**, *11*, 146-156.
- (3) Luisi, P. Molecular Conformational Rigidity: An Approach to Quantification. *Naturwissenschaften* **1977**, *64*, 569-574.
- (4) Kier, L. B. Shape Index from Molecular Graphs. *Quant. Struct.-Act. Relat.* **1985**, *4*, 109-116.
- (5) Kier, L. B. Shape Indexes of Orders One and Three from Molecular Graphs. *Quant. Struct.-Act. Relat.* **1986**, *5*, 1-7.
- (6) Kier, L. B. Distinguishing Atom Differences in a Molecular Graph Shape Index. *Quant. Struct.-Act. Relat.* **1986**, *4*, 7-12.
- (7) Kier, L. B. An Index of Molecular Flexibility from Kappa Shape Attributes. *Quant. Struct.-Act. Relat.* **1989**, *8*, 218-221.
- (8) Fisanick, W.; Cross, K.; Rusink, III, A. Similarity Searching on CAS Registry Substances. 1. Global Molecular Property and Generic Atom Triangle Geometric Searching. *J. Chem. Inf. Comput. Sci.* **1992**, *32*, 664-674.
- (9) Fisanick, W.; Cross, K.; Rusinko, II, A. Characteristics of Computer-Generated 3D and Related Molecular Property Data for CAS Registry Substances. *Tetrahedron Comput. Methodol.* **1990**, *3*, 635-652.
- (10) Zachmann, C.-D.; Kast, S.; Brickmann, J. Quantification and visualization of molecular surface flexibility. *J. Mol. Graphics* **1995**, *13*, 89-97.

- (11) Arteca, G. Global Measure of Molecular Flexibility and Shape Fluctuations about Conformation Minima. *J. Comput. Chem.* **1993**, *14*, 718–727.
- (12) Prelog, V.; Helmchen, G. Grundlagen des CIP-Systems and Vorschläge für eine Revision. *Angew. Chem.* **1982**, *94*, 614.
- (13) Kosko, B.; Isaka, S. Fuzzy Logic. *Sci. Am.* **1993**, 62–67.
- (14) Bandemer, H.; Otto, M. Fuzzy Theory in Analytical Chemistry. *Mikrochim. Acta* **1986**, 93–124.
- (15) von der Lieth, C. W.; Förster, T. Datenbankanwendungen in der Spektroskopie: Datenbankgestützte Spektrenauswertung. *GIT Fachz. Lab.* **1990**, *35*, 1245–1253.
- (16) Förster, T.; von der Lieth, C. W.; Opferkuch, H. Datenbankanwendungen in der Spektroskopie: Datenbankaufbau und -retrieval, *GIT Fachz. Lab.* **1989**, *33*, 319–328.
- (17) Stumpf-Nothof, K.; Prior, U.; von der Lieth, C. W. 2D/3D-Conversion Programs used in Data Base Application: How to get a Quality Factor for Data Conversion? In *Software-Entwicklung in der Chemie 8/Software-Development in Chemistry 8*, Jochum, C., Ed.; GDCh: Frankfurt/Main, 1994; pp 157–168.

CI9501204