Novel Chirality Descriptors Derived from Molecular Topology

Alexander Golbraikh,† Danail Bonchev,‡ and Alexander Tropsha*,†

Laboratory for Molecular Modeling, Division of Medicinal Chemistry and Natural Products, School of Pharmacy, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina 27599-7360, and Program for Theory of Complex Systems Texas A&M University, Galveston, Texas 77553-1675

Received June 30, 2000

Several series of novel chirality descriptors of chemical organic molecules have been introduced. The descriptors have been developed on the basis of conventional topological descriptors of molecular graphs. They include modified molecular connectivity indices, Zagreb group indices, extended connectivity, overall connectivity, and topological charge indices. These modified descriptors make use of an additional term called chirality correction, which is added to the vertex degrees of asymmetric atoms in a molecular graph. Chirality descriptors can be real or complex numbers. Advantages and drawbacks of different series of chirality descriptors are discussed. These descriptors circumvent the inability of conventional topological indices to distinguish chiral or enantiomeric isomers, which so far has been the major drawback of 2D descriptors as compared to true 3D descriptors (e.g., shape, molecular fields) of molecular structure. These novel chirality descriptors have been implemented in a quantitative structure−activity releationship (QSAR) study of a set of ecdysteroids with a high content of chiral and enantiomeric compounds using the *k* nearest neighbor QSAR method (kNN) recently developed in this laboratory. We show that the results of this study compare favorably with those obtained with the comparative molecular field analysis (CoMFA) applied to the same dataset. The novel chirality descriptors of molecular structure should find their applications in QSAR studies and related investigations of molecular sdatasets.

1. INTRODUCTION: MOTIVATION TO DEVELOP CHIRALITY DESCRIPTORS DERIVED FROM MOLECULAR TOPOLOGY

Asymmetry of atomic configurations is a very important feature in determining the physical, chemical, and biological properties of chemical substances. Two molecules with identical chemical formulas but different states of symmetry of only one atom are referred to as enantiomers. Enantiomers are optical antipodes: they rotate the polarization plane in opposite directions. The molecules with identical 2D structural formulas containing more than one asymmetric atom are referred to as σ-diastereomers.[2] In the literature, the asymmetric atoms are often referred to as chiral atoms, and molecules containing chiral atoms are referred to as chiral molecules. In general, the term "chirality" has a broader sense: for instance, chirality can be caused by a spatial isomerism resulting from the lack of free rotation around single or double bonds such as in derivatives of biphenyl or in allenes,[2,3] rather than due to chiral atoms. However, in this paper, the term chirality is used in the more narrow sense outlined above. We also consider only those cases when it is practically possible to partition asymmetric atoms into homochirality (R and S) classes (see ref 3 for further discussion).

Most of the physical as well as chemical properties of chiral molecules are similar. At the same time, it is well-known that many biological molecules (amino acids, carbohydrates, lipids, many receptor ligands) are chiral and that the chirality plays an essential role in defining biological activity (e.g., ref 1). In quantitative structure−activity relationship (QSAR) studies, taking chirality into account has become possible only after the development of the so-called three-dimensional QSAR methods such as the comparative molecular field analysis (CoMFA).[4,5] The method in its classical form[6] involves (i) a spatial alignment of molecules according to a pharmacophore model or their docking to a receptor, (ii) construction of a grid around the area occupied by the aligned molecules, (iii) calculation of van der Waals and Coulomb fields for each molecule in each grid point, (iv) partial least squares (PLS) (or another multiple-regression-like) procedure[7] with cross-validation[8] to construct a model for the prediction of a biological activity or property (the field values at grid points serve as independent variables), and (v) model validation using the external test set of compounds. Since the development of CoMFA, many varieties of CoMFA-like methods have appeared (e.g., refs 9−12). Evidently, the chirality in CoMFA is taken into account by default, since 3D molecular field values of chiral isomers are different.

Despite its wide popularity, CoMFA (and CoMFA-like) methods are not always applicable, especially, if compounds under investigation are highly flexible and no experimental structural information about the receptor−ligand complexes is available. Several shortcomings of CoMFA have been discussed in previous publications from this and other groups. In many cases, it is difficult to precisely define a pharmacophore model, and even if the alignment is based on the receptor structure, existing docking procedures are very slow or inexact, or both (cf. refs 13 and 14). Several problems related to the use of a grid in CoMFA studies have been

* To whom correspondence should be addressed.
† University of North Carolina at Chapel Hill.
‡ Texas A&M University.

10.1021/ci000082a CCC: $20.00 © 2001 American Chemical Society
Published on Web 01/04/2001

discussed previously, including the sensitivity of CoMFA results to the orientation,[10] the translation,[15] and the size of the grid.[16] These shortcomings have motivated several researchers to consider improvements to standard CoMFA procedures using variable selection approaches (GOLPE,[17] q^2-GRS[10]). In addition, alignment-free variable selection QSAR methods utilizing topological descriptors of chemical structures have been actively sought.[18−21] In fact, it has been shown that 2D descriptors of chemical structures are more efficient than 3D descriptors for chemical diversity and database mining[22] and at least as efficient in QSAR studies.[20,21]

Our recent experience with several datasets[20,21] indicated that the statistical characteristics (cross-validated and non-cross-validated correlation coefficients, standard deviation, *F*-test) and prediction power (statistics calculated for an external test set) of both 2D and 3D QSAR models are similar. In principle, 2D QSAR methods offer a clear advantage over 3D methods since they require no conformational analysis, no alignment, and no 3D pharmacophore hypothesis and, as a result, can be easily fully automated. However, one of the main drawbacks of QSAR approaches utilizing 2D molecular descriptors such as molecular connectivity indices (calculated with the MolconnZ program[23]) and atom pairs[24] has been their inability to take into account chirality of atoms since the latter is a true 3D property. Naturally, this deficiency has severely limited the range of applications of 2D QSAR methods, especially as compared to 3D QSAR approaches.

The idea of modifying conventional 2D descriptors to make them chirality sensitive was put forward several years ago.[25] In ref 25 a so-called chiral factor was introduced equal to +1 or −1 for an atom in R- or S-configuration, respectively. This factor was used to derive several chirality descriptors which appeared to have different values for linear hexose isomers.[25] The first attempt to consider chirality descriptors in QSAR studies was made in 1998,[26] when several chirality-sensitive molecular and charge topological indices were introduced. Both chirality descriptors and conventional MolconnZ[23] descriptors have been implemented in QSAR studies of series of D_2 dopamine and σ receptor ligands.[26] It was shown that the resulting QSAR models had better statistics and predictive power for IC_{50} values of the ligands than those obtained with conventional descriptors only. In addition, using these chirality descriptors, a set of chiral barbiturates were correctly classified as sedatives or stimulants.[26]

In this paper, we introduce several new series of chirality descriptors. They include the modified Zagreb group[27] and molecular connectivity indices[23,28−30] as well as recently introduced extended[31] and overall connectivity indices,[32,33] and topological charge indices.[26,34] All these descriptors make use of an additional term called chirality correction added to vertex degrees of asymmetric atoms in a molecular graph. Chirality correction can be a real or an imaginary number. In the latter case, the chirality descriptors are complex numbers. Advantages and drawbacks of different series of chirality descriptors are discussed. These novel descriptors can find immediate applications in QSAR and database mining studies of datasets that contain chiral compounds.

2. METHODOLOGY

2.1. Chirality Correction. A series of molecular descriptors can be calculated using the adjacency matrix for a hydrogen suppressed molecular graph. Namely, a molecule can be described as a nondirected graph, each vertex of which corresponds to a non-hydrogen atom, and each edge represents a chemical bond between two non-hydrogen atoms. An adjacency matrix **A** of the graph is defined as a square $N \times N$ symmetric matrix (N is the number of vertexes). If there is an edge connecting vertexes i and j ($i, j = 1, ..., N$), then $a_{ij} = 1$; otherwise $a_{ij} = 0$, where a_{ij} are the elements of matrix **A**.

Vertex degree a_i is defined as the sum of all elements of matrix A in the ith row:

$$a_i = \sum_{j=1}^{N} a_{ij} \tag{1}$$

The vertex degree for atom i is equal to the number of non-hydrogen atoms connected to it. A series of descriptors for molecules containing chiral atoms can be calculated by using chirality correction c for each chiral atom, when calculating the corresponding vertex degree. For each asymmetric atom in R-configuration, a_i is substituted with $(a_i + c)$, and for each atom in S-configuration a_i is substituted with $(a_i - c)$. This transformation is equivalent to making main diagonal elements a_{ii} of matrix A equal to $+c$ or $-c$, for all chiral atoms in R- or S-configuration, respectively. Chiral vertex degrees for these atoms can be obtained from formula 1.

2.2. Two Classes of Chirality Descriptors. Chirality correction c can be a real or an imaginary number. In the latter case, we will denote the chirality correction as ic, where $i = \sqrt{-1}$, thus always implying that c is a real number. [Thus, if the chirality correction is an imaginary number, the corresponding vertex degree will be a complex number equal to $(a_i + ic)$ or $(a_i - ic)$ for an atom in R- or S-configuration, respectively.] Subsequently, we introduce two classes of chirality descriptors: class I that is based on the real number chirality correction and class II that is based on the imaginary number chirality correction.

Since class I descriptors are real numbers, no additional transformation is necessary to use them in chemical database related applications. It is not the case for class II descriptors, for which several subclasses were defined in this work as follows (Table 1).

(i) If some vertex degrees are complex numbers, the descriptors will also generally be complex numbers. Thus, one of the options is to introduce complex descriptors and use them directly in applications to QSAR and other molecular data processing problems. These complex descriptors will be referred to as *subclass IIa* of class II of chirality descriptors. It is the most natural definition of the class II descriptors. Unfortunately, these descriptors cannot be currently applied in the majority of the QSAR software, which is not adapted yet to complex descriptors.

(ii) The real and imaginary parts of a complex descriptor are used as two real descriptors (*subclass IIb* of chirality descriptors).

Table 1. Comparison of the Chirality Indices

series	features[a] 1	2	3
class I	+	+	−
subclass IIa	−	+	+
subclass IIb	+	−	+
subclass IIc	+	+	+
subclass IId	+	−	+

[a] The following features are compared: (1) usability of descriptors in existing software; (2) values of chirality descriptors for nonchiral compounds must be equal to the corresponding nonchiral descriptors; (3) symmetry of descriptor values for a pair of enantiomers relative to (i) the values of corresponding nonchiral descriptors, or (ii) values that are independent of chirality of the atom defining this pair of enantiomers, or (iii) the real parts of the descriptor values (for class II of the descriptors).

(iii) *Subclass IIc* of the chirality descriptors is defined according to the following formula:

$$D_c = \mathrm{Re}(d) + \mathrm{Im}(d) \tag{2}$$

where d is a complex descriptor, D_c is the corresponding real descriptor. Re(d) and Im(d) are the real and imaginary parts of descriptor d.

(iv) *Subclass IId* of chirality descriptors is defined as follows:

$$D_d = \arctan(\mathrm{Re}(d),\mathrm{Im}(d)) = \begin{cases} \pi/2, \text{ if } \mathrm{Re}(d) = 0, \mathrm{Im}(d) > 0 \\ \arctan(\mathrm{Im}(d)/\mathrm{Re}(d)), \text{ if } \mathrm{Re}(d) > 0 \\ \arctan(\mathrm{Im}(d)/\mathrm{Re}(d)) - \pi, \text{ if } \mathrm{Re}(d) < 0, \mathrm{Im}(d) \le 0 \\ \arctan(\mathrm{Im}(d)/\mathrm{Re}(d)) + \pi, \text{ if } \mathrm{Re}(d) < 0, \mathrm{Im}(d) > 0 \\ -\pi/2, \text{ if } \mathrm{Re}(d) = 0, \mathrm{Im}(d) < 0 \end{cases} \tag{3}$$

Certainly, these subclasses do not exhaust all possibilities for defining class II chirality descriptors. However, the five subclasses introduced above seem to be the most obvious initial choices.

2.3. Symmetry and Asymmetry of Chirality Descriptors for a Pair of Enantiomers. In this paper, we describe symmetry and asymmetry of chiral descriptors for a pair of enantiomers in relation to the corresponding conventional descriptors, or terms independent of chirality of the atom defining this pair of enantiomers. Here we give the necessary definitions. Let d_1 and d_2 ($d_1 > d_2$) be the values of a chirality descriptor for a pair of enantiomers, and D is some number satisfying the condition $d_1 > D > d_2$. If $d_1 - D = D - d_2$, descriptors d_1 and d_2 are symmetrical relative to number D. If $d_1 - D < D - d_2$, the degree of symmetry of d_1 and d_2 relative to the number D is defined as $S = (d_1 - D)/(D - d_2)$; if $d_1 - D > D - d_2$, it is defined as $S = (D - d_2)/(d_1 - D)$. The degree of asymmetry Y of d_1 and d_2 relative to the number D is defined by formula

$$Y = 1 - S = \begin{cases} 1 - (d_1 - D)/(D - d_2), & \text{if } d_1 - D < D - d_2 \\ 1 - (D - d_2)/(d_1 - D), & \text{if } d_1 - D > D - d_2 \end{cases} \tag{4}$$

This definition guarantees that $Y \in [0,1]$.

3. ZAGREB GROUP INDICES $^nM_1$ AND $^nM_2$

Zagreb group indices $^nM_1$ (n is the order of the index: it is defined as the total number of edges in the corresponding subgraphs) are defined by the following formulas [in ref 27 only $^0M_1$ and $^1M_2$ are defined; for completeness, here Zagreb-group-like indices of higher orders are introduced]:

$$^0M_1 = \sum_{i=1}^{N} a_i^2, \quad ^1M_1 = \sum_{\text{all edges}} a_{i_1}^2 a_{i_2}^2, \ldots, \quad ^{n-1}M_1 = \sum_{\text{all } (n\text{-}1)\text{-edge subgraphs}} a_{i_1}^2 a_{i_2}^2 \ldots a_{i_v}^2 \tag{5}$$

Zagreb group indices $^1M_2$ are defined as follows:

$$^1M_2 = \sum_{\text{all edges}} a_{i_1} a_{i_2}, \; ^2M_2 = \sum_{\text{all 2-edge subgraphs}} a_{i_1} a_{i_2} a_{i_3}, \ldots \; ^{n-1}M_2 = \sum_{\text{all } (n\text{-}1)\text{-edge subgraphs}} a_{i_1} a_{i_2} \ldots a_{i_v} \tag{6}$$

where v is the number of vertexes in the $(n - 1)$-edge subgraph. (In general, for a subgraph containing $(n - 1)$ edges, v vertexes, and k cycles, the following condition is fulfilled for v: $n \ge v \ge (n - k)$.)

3.1. Class I Chirality Zagreb Group Indices. To obtain class I chirality descriptors, vertex degrees a_i must be replaced by $(a_i + c)$ for chiral atoms in R-configuration and by $(a_i - c)$ for chiral atoms in S-configuration. Thus, a_i^2 must be replaced by $(a_i \pm c)^2 = a_i^2 \pm 2a_ic + c^2$. (The plus sign in $\pm$ and minus sign in $\mp$, in this and in all subsequent formulas, refer to an atom in the R-configuration, and the minus sign in $\pm$ and the plus sign in $\mp$ refer to an atom in the S-configuration.) It follows immediately that *for a pair of enantiomers with only one asymmetric atom, the values of all* $^nM_1^{chir}$ *indices are symmetrical with respect to the value of* $^0M_1 + c^2$, *but not to the value of the corresponding nonchiral indices.* If the total number of atoms in R- and S-configurations are n_R and n_S, then

$$^0M_1^{\text{chir}} = {}^0M_1 + 2c\left[\sum_{j=1}^{n_R} a_j - \sum_{j=1}^{n_S} a_j\right] + \sum_{j=1}^{n_R+n_S} c^2 \tag{7}$$

Now, consider the following term: $a_{i_1}^2 a_{i_2}^2 \ldots a_{i_k}^2 \ldots a_{i_v}^2$. A sum of such terms defines $^{n-1}M_1$ for $n > 1$. If vertex i_k corresponds to an asymmetric atom defining the enantiomeric pair, a_{i_k} in this term must be substituted with $(a_{i_k} \pm c)$. Let $a_{i_1}^2 a_{i_2}^2 \ldots a_{i_{k-1}}^2 a_{i_{k+1}}^2 \ldots a_{i_v}^2 = A_s$, where $s = 1, \ldots, S$, and S is the total number of $(n - 1)$-edge subgraphs containing vertex i_k. Then for $^{n-1}M_1^{\text{chir}}$ we obtain

$$^{n-1}M_1^{\text{chir}} = {}^{n-1}M_1^{\text{chir0}} + \sum_{s=1}^{S} A_s(a_{i_k}^2 \pm 2ca_{i_k} + c^2) = {}^{n-1}M_1^{\text{chir*}} \pm 2c\sum_{s=1}^{S} A_s a_{i_k} + c^2 \sum_{s=1}^{S} A_s \tag{8}$$

In (8), $^{n-1}M_1^{\text{chir0}}$ denotes the term independent of i_k, and $^{n-1}M_1^{\text{chir*}} = {}^{n-1}M_1^{\text{chir0}} + \sum_{s=1}^{S} A_s a_{i_k}^2$ is the index value, if the chirality of i_k- atom is not taken into account.

Contrary to $^{n-1}M_1^{\text{chir}}$ indices, for a pair of enantiomers with only one asymmetric atom, the values of all $^{n-1}M_2^{\text{chir}}$ indices are symmetrical relative to the corresponding nonchiral

indices. Indeed, consider product $a_{i_1}a_{i_2}...a_{i_k}...a_{i_v}$, which is a term of ${}^{n-1}M_2$. Let i_k be the asymmetric atom defining a pair of enantiomers. Then a_{i_k} must be substituted with $(a_{i_k} \pm c)$. If $a_{i_1}...a_{i_{k-1}}a_{i_{k+1}}...a_{i_v} = B_s$, where $s = 1, ..., S$, and S is the total number of $(n - 1)$-edge subgraphs containing vertex i_k, then $a_{i_1}a_{i_2}...a_{i_k}...a_{i_v} \rightarrow B_s(a_{i_k} \pm c) = B_s a_{i_k} \pm B_s c$, where the arrow denotes the substitution. $B_s a_{i_k}$ is independent of the chirality of the i_k atom. The final formula for ${}^{n-1}M_2{}^{\text{chir}}$ is

$$ {}^{n-1}M_2{}^{\text{chir}} = {}^{n-1}M_2{}^{\text{chir0}} + \sum_{s=1}^{S} B_s a_{i_k} \pm c\sum_{s=1}^{S} B_s = {}^{n-1}M_2{}^{\text{chir*}} \pm c\sum_{s=1}^{S} B_s \quad (9) $$

In (9), ${}^{n-1}M_2{}^{\text{chir0}}$ denotes the term independent of atom i_k, and ${}^{n-1}M_2{}^{\text{chir*}} = {}^{n-1}M_2{}^{\text{chir0}} + \sum_{s=1}^{S} B_s a_{i_k}$ is the index value, if the chirality of the i_k- atom is not taken into account. If the enantiomers contain additional chiral atoms, they will be accounted for in B_s. Thus, from (9) it follows that *for a pair of enantiomers, the values of all ${}^{n-1}M_2{}^{chir}$ indices are symmetrical relative to the corresponding ${}^{n-1}M_2{}^{chir*}$ values, i.e., part of the index independent of the chirality of the atom determining these enantiomers.*

3.2. Class II Chiral Zagreb Group Indices. Since real and imaginary parts of the chirality descriptors have different properties, these parts will be discussed separately. To obtain class II chirality descriptors, vertex degrees a_i must be substituted with $(a_i + ic)$ for chiral atoms in R-configuration and with $(a_i - ic)$ for chiral atoms in S-configuration. Real parts of ${}^0M_1{}^{\text{chir}}$ for two enantiomers defined by atom j can be calculated from the following formula:

$$ \text{Re}({}^0M_1{}^{\text{chir}}) = \text{Re}({}^0M_1{}^{\text{chir*}}) - c^2 \quad (10a) $$

Imaginary parts of ${}^0M_1{}^{\text{chir}}$ for two enantiomers can be obtained from

$$ \text{Im}({}^0M_1{}^{\text{chir}}) = \text{Im}({}^0M_1{}^{\text{chir*}}) \pm 2a_j c \quad (10b) $$

In (10a) and (10b) the terms with the asterisks are independent of the chirality of the jth atom. Equations 10a,b can be obtained from (5) by using identity $(a \pm ic)^2 = (a^2 - c^2) \pm 2iac$. Thus, *the real parts for a pair of enantiomers of ${}^0M_1{}^{chir}$ are equal, and the imaginary parts are symmetrical relative to Im(${}^0M_1{}^{chir*}$), i.e. to the part of the index independent of the chirality correction of the atom determining enantiomers.* It can be easily shown that

$$ \text{Re}({}^0M_1{}^{\text{chir}}) = \sum_{j=1}^{N} a_j{}^2 - nc^2 \quad (11a) $$

$$ \text{Im}({}^0M_1{}^{\text{chir}}) = 2c\left(\sum_{j=1}^{n_R} a_j - \sum_{j=1}^{n_S} a_j\right) \quad (11b) $$

where N is the total number of non-hydrogen atoms in a molecule, n is the total number of asymmetric atoms in a molecule, n_R and n_S are the number of R and S chiral atoms, respectively, $n_R + n_S = n$.

Now, consider the following term $a_{i_1}^2 a_{i_2}^2 ... a_{i_k}^2 ... a_{i_v}^2$. A sum of such terms defines ${}^{n-1}M_1$ for $n > 1$. Again, if vertex i_k corresponds to an asymmetric atom defining the enantiomeric pair, a_{i_k} in this term must be substituted with $(a_{i_k} \pm ic)$. In general, some of the vertexes $i_1, i_2, ..., i_{k-1}, i_{k+1}, ..., i_v$ may correspond to other asymmetric atoms in the molecule, so that $a_{i_1}^2 a_{i_2}^2 ... a_{i_{k-1}}^2 a_{i_{k+1}}^2 ... a_{i_v}^2 \rightarrow A_s + iC_s$, where the arrow denotes the substitution $s = 1, ..., S$ and S is the total number of
$(n - 1)$-edge subgraphs containing vertex i_k. Then

$$ a_{i_1}^2 a_{i_2}^2 ... a_{i_k}^2 ... a_{i_v}^2 \rightarrow (A_s + iC_s)[(a_{i_k}^2 - c^2) \pm 2ica_{i_k})] = [A_s(a_{i_k}^2 - c^2) \mp 2C_s c a_{i_k}] + i[C_s(a_{i_k}^2 - c^2) \pm 2A_s c a_{i_k}] \quad (12) $$

Eventually, the following formulas can be obtained

$$ \text{Re}({}^{n-1}M_1{}^{\text{chir}}) = \text{Re}({}^{n-1}M_1{}^{\text{chir0}}) + \sum_{s=1}^{S} A_s a_{i_k}^2 - c^2\sum_{s=1}^{S} A_s \mp 2c\sum_{s=1}^{S} C_s a_{i_k} = \text{Re}({}^{n-1}M_1{}^{\text{chir*}}) - c^2\sum_{s=1}^{S} A_s \mp 2c\sum_{s=1}^{S} C_s a_{i_k} \quad (13a) $$

$$ \text{Im}({}^{n-1}M_1{}^{\text{chir}}) = \text{Im}({}^{n-1}M_1{}^{\text{chir0}}) + \sum_{s=1}^{S} C_s a_{i_k}^2 - c^2\sum_{s=1}^{S} C_s \pm 2c\sum_{s=1}^{S} A_s a_{i_k} = \text{Im}({}^{n-1}M_1{}^{\text{chir*}}) - c^2\sum_{s=1}^{S} C_s \pm 2c\sum_{s=1}^{S} A_s a_{i_k} \quad (13b) $$

where $\text{Re}({}^{n-1}M_1{}^{\text{chir0}})$ and $\text{Im}({}^{n-1}M_1{}^{\text{chir0}})$ denote the real and imaginary parts not containing terms pertaining to atom i_k and $\text{Re}({}^{n-1}M_1{}^{\text{chir*}})$ and $\text{Im}({}^{n-1}M_1{}^{\text{chir*}})$ are independent of the chirality of atom i_k. If i_k is the only chiral atom in a pair of enantiomers, then all $C_s = 0$, and *the real parts of ${}^{n-1}M_1{}^{chir}$ are equal, but the imaginary parts are symmetrical relative to Im(${}^{n-1}M_1{}^{chir*}$)*, i.e., to *the part of the index independent of the chirality correction of this atom.* The stronger condition, when the last assumption is true, is that *no one subgraph of order (n − 1) contains more than one chiral atom.*

Consider product $a_{i_1}a_{i_2}...a_{i_k}...a_{i_v}$, which is a term of ${}^{n-1}M_2$. Let a_{i_k} be the vertex degree of the asymmetric atom defining a pair of enantiomers. Again, it must be substituted with $(a_{i_k} \pm ic)$. Generally, a compound can contain several asymmetric atoms, so the product defined by the other terms can also be a complex number. Let it be equal to $B_s + iD_s$, where $s = 1, ..., S$, and S is the total number of $(n - 1)$-edge graphs containing vertex i_k. Then

$$ a_{i_1}a_{i_2}\cdots a_{i_k}\cdots a_{i_v} \rightarrow (B_s + iD_s)(a_{i_k} \pm ic) = (B_s a_{i_k} \mp D_s c) + i(a_{i_k} D_s \pm B_s c) \quad (14) $$

where the arrow denotes the substitution. B_s and D_s depend on vertex degrees of $i_1, i_2, ..., i_{k-1}, i_{k+1}, ..., i_v$. The following formulas can be obtained

$$ \text{Re}({}^{n-1}M_2{}^{\text{chir}}) = \text{Re}({}^{n-1}M_2{}^{chir0}) + \sum_{s=1}^{S}(B_s a_{i_k} \mp D_s c) = \text{Re}({}^{n-1}M_2{}^{\text{chir*}}) \mp c\sum_{s=1}^{S} D_s \quad (15a) $$

$$\mathrm{Im}({}^{n-1}M_2^{\mathrm{chir}}) = \mathrm{Im}({}^{n-1}M_2^{\mathrm{chir0}}) + \sum_{s=1}^{S}(a_{i_k}D_s \pm B_s c) = \mathrm{Im}({}^{n-1}M_2^{\mathrm{chir*}}) \pm c\sum_{s=1}^{S}B_s \quad (15b)$$

where $\mathrm{Re}({}^{n-1}M_2^{\mathrm{chir0}})$ and $\mathrm{Im}({}^{n-1}M_2^{\mathrm{chir0}})$ denote the parts not containing terms pertaining to atom i_k and $\mathrm{Re}({}^{n-1}M_2^{\mathrm{chir*}})$ and $\mathrm{Im}({}^{n-1}M_2^{\mathrm{chir*}})$ are the parts independent of the chirality of this atom. *If i_k is the only chiral atom in the pair of enantiomers, then all $D_s = 0$, and the real parts of ${}^{n-1}M_2^{chir}$ are equal for both enantiomers.* The same is true, *if no one subgraph of order $(n - 1)$ contains more than one chiral atom.* At the same time, *imaginary parts of ${}^{n-1}M_2^{chir}$ are symmetrical relative to $\mathrm{Im}({}^{n-1}M_2^{chir*})$,* i.e., *to the part of the index, independent of the chirality correction of the atom defining the pair of enantiomers.*

4. MOLECULAR CONNECTIVITY INDICES ${}^n\chi$

Molecular connectivity indices ${}^n\chi$ are defined as follows:[28–30]

$$ {}^0\chi = \sum_{i=1}^{N}(a_i)^{-0.5}, \quad {}^1\chi = \sum_{\text{all edges}}(a_{i_1}a_{i_2})^{-0.5}, \quad \ldots, {}^{n-1}\chi = \sum_{\text{all }(n\text{-1})\text{-edge subgraphs}}(a_{i_1}a_{i_2}\ldots a_{i_\upsilon})^{-0.5} \quad (16)$$

where υ is the number of vertexes in the $(n - 1)$-edge subgraph.

Division of indices of order higher than 2 into subtypes (path, cluster, path/cluster, cycle indices) is not considered here. The formulas obtained here can be easily extended to these subtypes. Thus, we consider only one connectivity index of order 3, one index of order 4, one index of order 5, etc., and all subgraphs of the corresponding orders contributing to them.

Valence connectivity indices ${}^n\chi^v$ are defined by the same formulas (16) with the vertex degrees $a_i = (Z_i{}^\upsilon - h_i)/(Z_i - Z_i^\upsilon - 1)$, where Z_i is the atomic number in the periodic table of elements, Z_i^υ is the number of valence electrons, and h_i is the number of hydrogen atoms connected to atom i.[23] ${}^n\chi^v$ will not be considered separately, since all of the conclusions obtained below for ${}^n\chi$ are valid for ${}^n\chi^v$ with the above-mentioned substitution.

4.1. Class I Chiral Molecular Connectivity Indices. The minimum nonchiral degree (a_i) of a chiral vertex is 3. Thus, the chirality correction for class I indices cannot be equal to ±3 or the other small whole number with the absolute value higher than 3, since if it is equal to it, some denominators in (16) can be equal to zero. In fact, to avoid having a real number index for one enantiomer and the complex number for the other enantiomer in an enantiomeric pair, *the absolute value of the chirality correction cannot exceed three (or more generally the lowest nonchiral degree of a chiral atom in S-configuration in a dataset, if $c > 0$, or in R-configuration, if $c < 0$).* It is a natural limit imposed on the real number chirality correction. Furthermore, in this subsection, we will assume that $|c| < a_{i_k}$.

Consider index ${}^{n-1}\chi$. It is a sum of S terms of the kind $(a_{i_1}a_{i_2}\ldots a_{i_k}\ldots a_{i_\upsilon})^{-0.5}$ containing a chiral atom i_k defining a pair of enantiomers and other terms, independent of atom i_k, the sum of which is F. Let $(a_{i_1}\ldots a_{i_{k-1}}a_{i_{k+1}}\ldots a_{i_\upsilon})^{-0.5} = H_j$, $j = 1, \ldots, S$. Then for a pair of enantiomers $(a_{i_1}a_{i_2}\ldots a_{i_k}\ldots a_{i_\upsilon})^{-0.5} \rightarrow H_j(a_{i_k} \pm c)^{-0.5} = H_j a_{ik}{}^{-0.5}(1 \pm c/a_{i_k})^{-0.5} = H_j a_{i_k}{}^{-0.5}(1 \pm x)^{-0.5}$, where $x = c/a_{i_k}$. At the same time, if the chirality of atom i_k is not taken into account, $(a_{i_1}a_{i_2}\ldots a_{i_k}\ldots a_{i_\upsilon})^{-0.5} = H_j a_{i_k}{}^{-0.5}$. If $a_{i_k} > c > 0$, the positive deviations of ${}^{n-1}\chi$ from the value $H_j a_{i_k}{}^{-0.5}$, i.e., when the chirality of atom i_k is not taken into account, for enantiomers with atom i_k in R- and S-configuration are

$$[F + \sum_{j=1}^{S}H_j a_{i_k}^{-0.5}] - [F + \sum_{j=1}^{S}H_j a_{i_k}^{-0.5}(1 + x)^{-0.5}] = a_{i_k}^{-0.5}[1 - (1 + x)^{-0.5}]\sum_{j=1}^{S}H_j$$

$$[F + \sum_{j=1}^{S}[H_j a_{i_k}^{-0.5}(1 - x)^{-0.5}] - [F + \sum_{j=1}^{S}H_j a_{i_k}^{-0.5}] = a_{i_k}^{-0.5}[(1 - x)^{-0.5} - 1]\sum_{j=1}^{S}H_j$$

respectively. Their ratio is the degree of symmetry of the corresponding indices for a pair of enantiomers relative to the value $F + \sum_{j=1}^{S}H_j a_{i_k}{}^{-0.5}$, i.e., when the chirality of atom i_k is not taken into account (see subsection 2.3). If this ratio is one, the chirality descriptors are symmetrical relative to this value. As we will see, it is never the case for the descriptors in question, except when $c = 0$. The degree of asymmetry Y is defined as one minus this ratio (eq 4). After cancellation on $a_{i_k}{}^{-0.5}\sum_{j=1}^{S}H_j$, the following formula is obtained:

$$Y = 1 - \frac{1 - (1 + x)^{-0.5}}{(1 - x)^{-0.5} - 1} \quad (17a)$$

In (17a), x belongs to interval (0,1). If the degree of asymmetry Y would be equal to zero, the chirality descriptors would deviate symmetrically from the values $F + \sum_{j=1}^{S}H_j a_{i_k}{}^{-0.5}$ when the chirality of the corresponding atom is not taken into account. We are going to prove that it is impossible, if $c \neq 0$. By expansion in Taylor series, it can be shown that $\lim_{x\to 0} Y = 0$. Indeed, for small $x(1 \pm x)^{-0.5} = 1 \mp 0.5x + o(x)$; thus $\lim_{x\to 0} Y = 1 - \{[1 - (1 - 0.5x + o(x))]/[(1 + 0.5x + o(x)) - 1] = 1 - 1 = 0$. It is obvious also that $\lim_{x\to 1} Y = 1$. It can be shown that Y as a function of x monotonically increases in interval (0,1). Actually, it is a continuous and differentiable function in this interval with the derivative $Y'(x) > 0$ Indeed,

$$Y'(x) = -\frac{(1 + x)^{-1.5}[(1 - x)^{-0.5} - 1] - (1 - x)^{-1.5}[1 - (1 - x)^{-0.5}]}{2[(1 - x)^{-0.5} - 1]^2}$$

The denominator of $Y'(x)$ is positive, and the numerator can be rewritten as follows:

$$\frac{2}{(1 - x^2)^{3/2}} - \left[\frac{1}{(1 + x)^{3/2}} + \frac{1}{(1 - x)^{3/2}}\right] = \frac{2 - [(1 - x)^{3/2} + (1 + x)^{3/2}]}{(1 - x^2)^{3/2}}$$

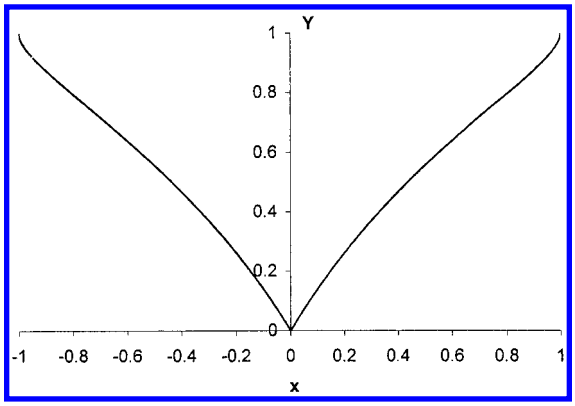

Figure 1. Dependence of the asymmetry of class I molecular connectivity indices Y on the ratio of the chirality correction to the vertex degree of a chiral atom.

The denominator of the right side of the last formula is positive; if $x = 0$, the numerator of the right side of the last equation is 0, and its derivative is $^3/_2[(1-x)^{1/2} - (1+x)^{1/2}] < 0$. Taking into account the minus sign before the fraction in the equation for $Y'(x)$, we obtain that in interval (0,1) *$Y'(x) > 0$ and $Y'(x)$ monotonically increases in the interval (0,1) from 0 to 1* (Figure 1). A case when the chirality correction is negative ($-a_{i_k} < +c < 0$) can be considered in a similar way. In this case, asymmetry Y can be defined as

$$Y = 1 - \frac{1-(1-x)^{-0.5}}{(1+x)^{-0.5} - 1} \tag{17b}$$

where x belongs to the interval (−1,0) (Figure 1). Thus, we proved that *the higher the absolute value of chirality correction is, the higher is the degree of asymmetry.* This is another drawback of the class I chiral molecular connectivity indices.

4.2. Class II Chiral Molecular Connectivity Indices. The terms in the sums in (16) contain square roots. The square root of a complex number has two values. Let z be a complex number. It can be represented as a sum of the real and imaginary parts; i.e., $z = x + iy$. For calculating the powers of complex numbers, a polar representation is used. Namely, each complex number, except for zero, can be represented in a form $z = \rho e^{i\varphi} = \rho(\cos\varphi + i\sin\varphi)$, where $\rho = (x^2 + y^2)^{1/2}$ is the modulus or the magnitude of z, $\cos\varphi = x/\rho$, $\sin\varphi = y/\rho$, and φ is the argument or the phase of z. We will use the φ values from the interval $[-\pi,\pi)$. $z^{1/2}$ has two values, one equal to $\rho^{1/2}e^{i\varphi/2}$ and the other $\rho^{1/2}e^{i(\varphi/2+\pi)}$, if $\varphi < 0$, or $\rho^{1/2}e^{i(\varphi/2-\pi)}$, if $\varphi > 0$, where $\pi = 3.14159....$ For our molecular connectivity indices we will use the first value so that the argument of $z^{1/2}$ would belong to the interval $[-\pi/2, \pi/2)$.

Let i be a chiral atom. Then a_i in eqs 16 must be substituted with $z_i = a_i \pm ic$. The modulus of $z_i\rho_i = (a_i^2 + c^2)^{1/2}$, and the argument $\varphi = \pm\arctan(c/a_i)$ since $a_i > 0$ by definition. After simple transformations, the following formula is obtained:

$$z_I^{-0.5} = \frac{1}{\sqrt[4]{a_i^2 + c^2}} \exp\left(\mp 0.5i \arctan\frac{c}{a_i}\right) = \frac{1}{\sqrt{a_i^2 + c^2}}\left[\sqrt{\frac{\sqrt{a_i^2 + c^2} + a_i}{2}} \mp i\sqrt{\frac{\sqrt{a_i^2 + c^2} - a_i}{2}}\right] = \frac{1}{\rho_i}\left[\sqrt{\frac{\rho_i + a_i}{2}} \mp i\sqrt{\frac{\rho_i - a_i}{2}}\right] \tag{18}$$

For $^0\chi^{\text{chir}}$ the following formulas are obtained:

$$\text{Re}(^0\chi^{\text{chir}}) = \sum_{i=1}^{n} \frac{1}{\rho_i}\sqrt{\frac{\rho_i + a_i}{2}} \tag{19a}$$

$$\text{Im}(^0\chi^{\text{chir}}) = -\sum_{i=1}^{n_r} \frac{1}{\rho_i}\sqrt{\frac{\rho_i - a_i}{2}} + \sum_{i=1}^{n_S} \frac{1}{\rho_i}\sqrt{\frac{\rho_i - a_i}{2}} \tag{19b}$$

where n is the total number of vertexes and n_R and n_S are the number of atoms in R- and S-configuration, correspondingly. In the case of a symmetrical atom, the corresponding term calculated with (18) or (19a) will be equal to $(a_i)^{-0.5}$, as it must be, since for symmetrical atom $c = 0$, and $\rho_i = a_i$, and the contribution to $\text{Im}(^0\chi^{\text{chir}})$ (formula 19b) will be zero. Formula 19a can be rewritten as follows:

$$\text{Re}(^0\chi^{\text{chir}}) = \sum_{i=1}^{n_1} (a_i)^{-0.5} + \sum_{i=1}^{n_2} \frac{1}{\rho_i}\sqrt{\frac{\rho_i + a_i}{2}} \tag{20a}$$

where n_1 and n_2 are the numbers of symmetrical and asymmetrical atoms in a molecule, $n_1 + n_2 = N$. Let j be a chiral atom defining a pair of enantiomers. Formula 19b can be rewritten as follows:

$$\text{Im}(^0\chi^{\text{chir}}) = \text{Im}(^0\chi^{\text{chir0}}) \mp \frac{1}{\rho_j}\sqrt{\frac{\rho_j - a_j}{2}} \tag{20b}$$

where $\text{Im}(^0\chi^{\text{chir0}})$ is a part that includes all terms in sums 19b except for that for atom j. From (19b) and (20b) it follows that *for a pair of enantiomers real parts of $^0\chi^{\text{chir}}$ are equal and imaginary parts are symmetrical relative to the $\text{Im}(^0\chi^{\text{chir0}})$ value, i.e., part independent of the chirality of the atom defining a pair of enantiomers.*

In the case of molecular connectivity indices of higher orders, terms $a_{i_1}a_{i_2}...a_{i_k}...a_{i_v}$ must be considered. Let i_k be a chiral atom. Generally, among atoms $i_1, i_2, ..., i_{k-1}, i_{k+1}, ..., i_v$ other asymmetric atoms can be found. So the general formula 14 can be applied with all designations pertaining to (14) (see above). In this case,

$$\rho_{s,ik} = [(B_s a_{i_k} \mp D_s c)^2 + (a_{i_k} D_s \pm B_s c)^2]^{0.5} = [(B_s a_{i_k})^2 + (D_s c)^2 + (a_{i_k} D_s)^2 + (B_s c)^2]^{0.5} \tag{21a}$$

$$\varphi_{s,i_k} = \begin{cases} \operatorname{acrtan}\dfrac{a_{i_k}D_s \pm B_s c}{B_s a_{i_k} \mp D_s c}, \text{ if } B_s a_{i_k} \mp D_s c > 0 \\ \arctan\dfrac{a_{i_k}D_s \pm B_s c}{B_s a_{i_k} \mp D_s c} - \pi, \text{ if } B_s a_{i_k} \mp D_s c < 0,\ a_{i_k}D_s \pm B_s c < 0 \\ \arctan\dfrac{a_{i_k}D_s \pm B_s c}{B_s a_{i_k} \mp D_s c} + \pi, \text{ if } B_s a_{i_k} \mp D_s c < 0,\ a_{i_k}D_s \pm B_s c > 0 \end{cases} \tag{21b}$$

After simple transformations, the following formula can be obtained:

$$(a_{i_1}a_{i_2}...a_{i_k}...a_{i_v})^{-0.5} \rightarrow \begin{cases} \dfrac{1}{\rho_{s,i_k}}\left[\sqrt{\dfrac{\rho_{s,i_k} + |B_s a_{i_k} \mp D_s c|}{2}} \mp i\sqrt{\dfrac{\rho_{s,i_k} - |B_s a_{i_k} \mp D_s c|}{2}}\right], \\ \qquad\qquad \text{if } a_{i_k}D_s \pm B_s c > 0 \\ \dfrac{1}{\rho_{s,i_k}}\left[\sqrt{\dfrac{\rho_{s,i_k} + |B_s a_{i_k} \mp D_s c|}{2}} \pm i\sqrt{\dfrac{\rho_{s,i_k} - |B_s a_{i_k} \mp D_s c|}{2}}\right], \\ \qquad\qquad \text{if } a_{i_k}D_s \pm B_s c < 0 \end{cases} \tag{22}$$

where the arrow denotes the substitution. At last, the following formulas are obtained

$$\mathrm{Re}({}^{n-1}\chi^{\mathrm{chir}}) = \mathrm{Re}({}^{n-1}\chi^{\mathrm{chir0}}) + \sum_{s=0}^{S}\frac{1}{\rho_{s,i_k}}\sqrt{\frac{\rho_{s,i_k} + |B_s a_{i_k} \mp D_s c|}{2}} \tag{23a}$$

$$\mathrm{Im}({}^{n-1}\chi^{\mathrm{chir}}) = \mathrm{Im}({}^{n-1}\chi^{\mathrm{chir0}}) + \sum_{s=0}^{S_1}(\mp)\frac{1}{\rho_{s,i_k}}\sqrt{\frac{\rho_{s,i_k} - |B_s a_{i_k} \mp D_s c|}{2}} + \sum_{s=0}^{S_2}(\pm)\frac{1}{\rho_{s,i_k}}\sqrt{\frac{\rho_{s,i_k} - |B_s a_{i_k} \mp D_s c|}{2}} \tag{23b}$$

In (23a) and (23b) $\mathrm{Re}({}^{n-1}\chi^{\mathrm{chir0}})$ and $\mathrm{Im}({}^{n-1}\chi^{\mathrm{chir0}})$ are real and imaginary parts of the indices independent of chiral atom i_k. S_1 and S_2 are the numbers of terms for which $a_{i_k}D_s \pm B_s c > 0$ and $a_{i_k}D_s \pm B_s c < 0$, respectively. *If i_k is the only chiral atom in the pair of enantiomers, all $D_s = 0$, and the real parts of the molecular connectivity indices for both enantiomers are equal.* Since B_s will be positive, for enantiomers with the atom in R- or S-configuration, only the first or the second sum in (23b) will remain, and they will be equal to each other; consequently, *the imaginary parts of ${}^{n-1}\chi^{\mathrm{chir}}$ will be symmetrical relative to zero. If in the pair of enantiomers there are other chiral atoms, but all (n − 1)-edge subgraphs contain no more than one of them, the real parts of the indices will also be equal, and the imaginary parts will be symmetrical relative to $\mathrm{Im}({}^{n-1}\chi^{\mathrm{chir0}})$,* i.e. *the component of the index independent of atom i_k.*

5. EXTENDED AND OVERALL CONNECTIVITY INDICES.

First, the definitions of nonchiral extended and overall connectivity indices are given. Then the properties of the corresponding chirality descriptors are considered.

5.1. Extended Connectivity Indices ^{n}EC.[31] The sums of vertex degrees of all vertexes connected to each vertex are calculated. If vertex i is connected to j_i vertexes, then the following sums are obtained:

$${}^n b_i = \sum_{k=1}^{j_i} {}^{n-1}b_k \tag{24a}$$

where n is the order of extended connectivity index ^{n}EC and ${}^0b_k = a_k$, $k = 1, ..., N$ (the vertex degrees of order 0 are equal to that obtained from the adjacency matrix). ^{n}EC is defined as a sum of ${}^n b_i$ values for all vertexes:

$${}^n\mathrm{EC} = \sum_{i=1}^{N} {}^n b_i = \sum_{i=1}^{N}\sum_{k=1}^{j_i} {}^{n-1}b_k \tag{24b}$$

5.2. Overall Connectivity Indices ^{n}TC$_0$ and ^{n}TC$_1$.[32,33] For all subgraphs of a molecular graph, vertex degrees are obtained using only adjacency matrixes for these subgraphs (for ^{n}TC$_0$) or the adjacency matrix for the molecular graph (for ^{n}TC$_1$). If the total number of subgraphs of order n is equal to j_n, and the total number of vertexes of a subgraph i is k_i,

$${}^n\mathrm{TC}_f = \sum_{i=1}^{j_n}\sum_{j=1}^{k_l} {}^n b^i_j \tag{25}$$

where ${}^n b^i_j$ is the vertex degree of atom j in subgraph i of order n and f can be equal to zero or one. In the case of $f = 1$, ${}^n b^i_j$ is equal to the corresponding vertex degree for the whole molecular graph. Evidently, if chiralities of atoms are not taken into account, ${}^0\mathrm{TC}_0 = 0$, and ${}^1\mathrm{TC}_0 = {}^0\mathrm{TC}_1$ and is equal to the total number of edges of the whole molecular graph, multiplied by 2. An additional two indices were used, which are the sums of the ${}^n\mathrm{TC}_f$ ($f = 1, 2$) over all orders 0 to m:

$$\mathrm{TC}_f = \sum_{n=0}^{m} {}^n\mathrm{TC}_f \tag{26}$$

m is the maximum order of a subgraph, i.e., the graph itself, or, since it is practically impossible to obtain all subgraphs of a large graph, m can be an external parameter.

5.3. Class I Chirality Extended and Overall Connectivity Indices. Since all vertex degrees in ^{n}EC, ${}^n\mathrm{TC}_f$, and TC_f are in the first power, for two enantiomers the corresponding chirality descriptors are symmetrical relative to their values calculated without taking into account the chirality of the atom defining enantiomers. In other words,

$${}^n\mathrm{EC}^{\mathrm{chir}} = {}^n\mathrm{EC}^{\mathrm{chir*}} \pm {}^n\mathrm{pc} \tag{27a}$$

$${}^n\mathrm{TC}_f^{\mathrm{chir}} = {}^n\mathrm{TC}_f^{\mathrm{chir*}} \pm {}^n q_f c \tag{27b}$$

$$\mathrm{TC}_f^{\mathrm{chir}} = \mathrm{TC}_f^{\mathrm{chir*}} \pm q_f c = \mathrm{TC}^*_f \pm c\sum_{n=0}^{m} {}^n q_f \tag{27c}$$

where the indices denoted with asterisks are independent of

the chirality correction of atom defining enantiomers and ${}^{n}p$, ${}^{n}q_f$, and q_f are the counts the atom makes as a contribution to the corresponding descriptor. These counts depend also on the molecular graph (${}^{0}q_1 = 1$, and ${}^{1}q_0$ is equal to the nonchiral vertex degree of the chiral atom defining enantiomers).

5.4. Class II Chiral Extended and Overall Connectivity Indices. For the same reason, as in the previous subsection, for two enantiomers the real parts of chirality descriptors are equal to the corresponding nonchiral descriptors. The imaginary parts of chirality descriptors are symmetrical relative to their values, calculated without taking into account the chirality of the atom that defines enantiomers. Thus, it can be written

$$\mathrm{Re}({}^{n}\mathrm{EC}^{\mathrm{chir}}) = \mathrm{Re}({}^{n}\mathrm{EC}^{\mathrm{chir}*}) \tag{28a}$$

$$\mathrm{Re}({}^{n}\mathrm{TC}_f^{\mathrm{chir}}) = \mathrm{Re}({}^{n}\mathrm{TC}_f^{\mathrm{chir}*}) \tag{28b}$$

$$\mathrm{Re}(\mathrm{TC}_f^{\mathrm{chir}}) = \mathrm{Re}(\mathrm{TC}_f^{\mathrm{chir}*}) \tag{28c}$$

$$\mathrm{Im}({}^{n}\mathrm{EC}^{\mathrm{chir}}) = \mathrm{Im}({}^{n}\mathrm{EC}^{\mathrm{chir}*}) \pm {}^{n}p_f c \tag{28d}$$

$$\mathrm{Im}({}^{n}\mathrm{TC}_f^{\mathrm{chir}}) = \mathrm{Im}({}^{n}\mathrm{TC}_f^{\mathrm{chir}*}) \pm {}^{n}q_f c \tag{28e}$$

$$\mathrm{Im}(\mathrm{TC}_f^{\mathrm{chir}}) = \mathrm{Im}(\mathrm{TC}_f^{\mathrm{chir}*}) \pm ic\sum_{n=0}^{m} {}^{n}q_f \tag{28f}$$

where the indices denoted with asterisks are independent of the chirality correction of the atom defining enantiomers and ${}^{n}p$, ${}^{n}q_f$, and q_f are the numbers of times each atom makes a contribution to the corresponding descriptor. As it follows from the definitions, class I and subclass IIc of extended and overall connectivity indices of the same order are equivalent. Achiral and subclass IIb (corresponding to real parts of complex descriptors only) of extended and overall connectivity indices of the same order are also equivalent. Subclass IIb ${}^{n}\mathrm{TC}_0$ and ${}^{n}\mathrm{TC}_1$ indices corresponding to imaginary parts of complex descriptors are equal as well.

6. TOPOLOGICAL CHARGE INDICES

In ref 34 chiral topological charge and valence topological charge indices were introduced. The definition of the nonchiral charge indices is given in ref 26. Let **D** be a distance matrix. Its elements d_{ij} are defined as the number of bonds of the minimal path connecting vertexes i and j ($i, j = 1, ..., N$, where N is the total number of vertexes in the molecular graph). Coulombic matrix **Q** is defined by its elements: $q_{ii} = 0$, and for $i \neq j$ $q_{ij} = 1/d_{ij}^2$. Further, define matrix $\mathbf{M} = \mathbf{AQ}$. Let $g_{ij} = m_{ij} - m_{ji}$. Topological charge indices G_k and J_k are defined as[28,34]

$$G_k = \sum_{i=1}^{N-1}\sum_{j=1}^{N} |g_{ij}|\delta_{k,d_{ij}}, \quad J_k = \frac{G_k}{N-1} \tag{29}$$

where δ is the Kronecker's delta.

Valence topological charge indices G_k^v and J_k^v are defined by the same formulas (29), but matrix $\mathbf{A}'$ is used instead of matrix **A** with all nondiagonal elements equal to those of matrix **A**, and the diagonal elements equal to atomic electronegativities.[26] G_k^v and J_k^v will not be considered separately, since all of the conclusions obtained below for G_k and J_k are valid for G_k^v and J_k^v.

6.1. Class I Topological Charge Indices. In ref 26, for each carbon in R- or S-configuration one was added to or subtracted from the corresponding main diagonal element of matrix **A** or $\mathbf{A}'$ (see above). Here this approach is generalized, so if i is a chiral atom in R- or S-configuration, element $a_{ii} = 0$ of matrix **A** (a'_{ii} of $\mathbf{A}'$) is substituted with $\pm c$ ($a'_{ii} \pm c$), the upper sign referring to R- and the lower sign to S-configuration, as before. It can be shown then that g_{ij} in (28) must be substituted with $g_{ij} \pm (1/d_{ij}^2)c$. Thus, *if for all atoms j with* $d_{ij} = k$ *all* $|g_{ij}| \geq |(1/d_{ij}^2)c|$, *the chiral topological charge index for two enantiomers will be symmetrical relative to the index value, calculated without taking the chirality of atom i into account.* Unfortunately, this condition is not satisfied even for the simplest example, triple substituted methane with all different substituents, even for $c = 1$. The chiral adjacency matrix $\mathbf{A}^{\mathrm{chir}}$ and the **Q** matrix for this example are (carbon atom is atom 1):

$$\mathbf{A}^{\mathbf{chir}} = \begin{bmatrix} \pm c & 1 & 1 & 1 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix}, \quad \mathbf{Q} = \begin{bmatrix} 0 & 1 & 1 & 1 \\ 1 & 0 & 1/4 & 1/4 \\ 1 & 1/4 & 0 & 1/4 \\ 1 & 1/4 & 1/4 & 0 \end{bmatrix}$$

Matrix $\mathbf{M}^{\mathrm{chir}} = \mathbf{A}^{\mathrm{chir}}\mathbf{Q}$ is as follows:

$$\mathbf{M}^{\mathbf{chir}} = \begin{bmatrix} 3 & 1/2 \pm c & 1/2 \pm c & 1/2 \pm c \\ 0 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 \end{bmatrix}$$

Thus, $g_{1j}^{\mathrm{chir}} = 1/2 \pm c$, $j = 2, 3, 4$, and if $c = 1$, for R- and S-enantiomers $|g_{1j}^{\mathrm{chir}}| = {}^{3}/_{2}$ and $|g_{1j}^{\mathrm{chir}}| = 1/2$, respectively; i.e., these values are not symmetrical relative to a nonchiral value equal to 1/2. This drawback can be improved by a slightly different definition of G_k^{chir} (and J_k^{chir}). Namely, instead of substituting in (29) g_{ij} with $g_{ij} \pm (1/d_{ij}^2)c$, substitute $|g_{ij}|$ with $|g_{ij}| \pm (1/d_{ij}^2)c$. In this case, *the chiral topological charge index for two enantiomers will be always symmetrical relative to the value of this index calculated without taking into account the chirality of the atom defining the enantiomers.*

6.2. Class II Topological Charge Indices. In this case, g_{ij} in (29) must be substituted with $g_{ij} \pm (1/d_{ij}^2)ic$. Since the complex number and its complex conjugate have equal modules, for a pair of enantiomers with one and only one chiral atom, all G_k^{chir} and J_k^{chir} will be equal real numbers; i.e., the topological charge indices defined by (29) are unable to discriminate R- and S-enantiomers. So, another definition of chiral topological charge indices is proposed, namely, in (29) substitute $|g_{ij}|$ with $|g_{ij}| \pm (1/d_{ij}^2)ic$. In this case, *for two enantiomers, the real parts of the chiral topological charge indices will be equal to the corresponding nonchiral descriptors*, i.e.

$$\mathrm{Re}(G_k^{\mathrm{chir}}) = G_k \tag{30a}$$

$$\mathrm{Re}(J_k^{\mathrm{chir}}) = J_k \tag{30b}$$

and the imaginary parts will be symmetrical relative to their values obtained without taking the chirality of the atom

defining the enantiomers into account:

$$\mathrm{Im}(G_k^{\mathrm{chir}}) = \mathrm{Im}(G_k^{\mathrm{chir}*}) \pm i\sum_{j=1}^{N}\frac{c}{d_{ij}^{\,2}} \tag{30c}$$

$$\mathrm{Im}(J_k^{\mathrm{chir}}) = \frac{\mathrm{Im}(G_k^{\mathrm{chir}})}{N-1} \tag{30d}$$

where $\mathrm{Im}(G_k^{\mathrm{chir}*})$ is the part independent of the chirality of atom i.

7. TEST CASE: APPLICATION TO A SERIES OF ECDYSTEROIDS

A goal of these preliminary calculations was to show that QSAR models based on chirality 2D descriptors are useful in QSAR studies and can even outperform the CoMFA models. A series of 78 ecdysteroids,[35] analogues of the steroid hormone ecdyson, which regulate molting in arthropods[36] was selected as a good example for testing the performance of novel chirality descriptors in QSAR studies for the following reasons. (i) All molecules in this dataset contain chiral atoms. (ii) The dataset contains 19 pairs of enantiomers and σ-diastereomers (Table 2a; chemical structures of these compounds are shown in Table 2b). (iii) Bioassay data for binding of these compounds to ecdyson receptor were published recently.[35] (iv) Two CoMFA models were built for a subset of 71 of these compounds.[35] The models were used to predict ED_{50} values for an external test set consisting of the remaining seven of the whole series of 78 compounds.[35] (v) In the absence of chirality descriptors, our previous calculations for this dataset (data not published) had to be limited to only 58 compounds, since we had to eliminate pairs of chiral isomers.

Two types of calculations were performed. First, all 78 compounds were used to obtain a QSAR model based on the chirality descriptors. The following descriptors of class I and subclass IId as well as the corresponding nonchiral descriptors have been calculated: Zagreb group indices; molecular connectivity indices; extended and overall connectivity indices. After some experimenting, we chose the chirality correction to have the value of 2, to produce the best QSAR model. The maximum order of all descriptors was 7. A series of nonchiral descriptors calculated by MolconnZ[23] have been added to our chirality descriptors. They include valence path, cluster, path/cluster and chain molecular connectivity indices,[28−30] κ molecular shape indices,[37,38] topological[39] and electrotopological[40−43] state indices, differential connectivity indices,[44] the graph's radius and diameter,[45] Wiener[46] and Platt[47] indices, Shannon[48] and Bonchev−Trinajstić [49] information indices, counts of different vertexes,[23] and counts of paths and edges between different types of vertexes.[23] The k-nearest neighbors QSAR (kNN-QSAR) method recently developed in this laboratory[21] was employed. KNN-QSAR is a variable selection approach, which uses the leave-one-out cross-validation procedure and evolution algorithm for optimal descriptor selection. Due to the stochastic nature of this method, as many as 100 models have been generated. The results for the best QSAR model ($q^2 = 0.73$) are presented in Figure 2. Table 2a shows that the model is indeed capable of distinguishing between enantiomers and σ-diastereomers; that is, it produces different values of predicted biological activity for chiral isomers which in most cases agree reasonably well with the experimental values.

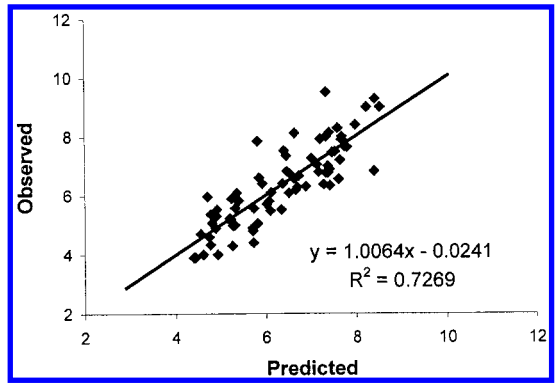

Figure 2. Observed vs predicted ED_{50} values for 78 ecdysteroids.

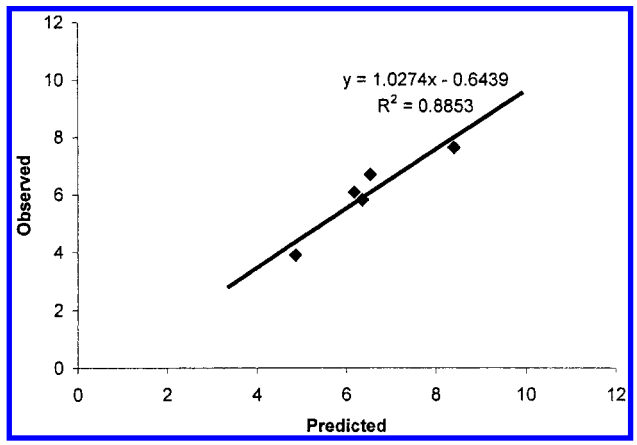

Figure 3. Observed vs predicted ED_{50} values for seven ecdysteroids from the test set from ref 36. The QSAR model was constructed using 71 compounds from the training set from ref 36. Class I chirality descriptors together with a series of nonchiral descriptors were used to derive the model (see section 7).

In addition, following the procedure reported in the CoMFA paper,[35] the 78 compounds have been divided into the training and test sets. Eight different sets of chirality descriptors have been calculated along with conventional descriptors mentioned above. The sets of chirality descriptors included (1) only class I descriptors, (2) only subclass IIb descriptors, (3) only subclass IIc descriptors, (4) only subclass IId descriptors, (5) class 1 with subclass IIb descriptors, (6) class 1 with subclass IIc descriptors; (7) class 1 with subclass IId descriptors, and (8) class I and class II descriptors except for subclass IIa descriptors.

The value of the chirality correction was again set to two. The maximum order of all descriptors was seven. 100 kNN-QSAR models were derived for each case and the best models were selected. The results are presented in Table 3. For comparison, statistics of the CoMFA models from ref 35 is also presented. Obviously, the QSAR models based on chirality descriptors compare extremely favorably with the results of CoMFA. Especially good results were obtained in case 1 (see also Figure 3). The statistics for the test set for the CoMFA models developed in ref 35 was significantly worse even after exclusion of the worst outlier (cf. Table 3). Thus, the test case presented in this section of the paper demonstrates that novel chirality descriptors can be success-

Table 2

(a) Predicted vs Observed − log(ED_{50}) for Enantiomeric and σ-Diastereomeric Pairs of Compounds in the Ecdysteroid Dataset

compd	pred activity	obsd activity	compd	pred activity	obsd activity	no. of different chiral centers
1	4.71	5.96	**3**	4.80	5.36	1
1	4.71	5.96	**39**	4.61	4.00	1
1	4.71	5.96	**40**	4.93	4.00	2
1	4.71	5.96	**57**	4.78	5.36	2
2	6.45	7.35	**38**	4.91	5.52	1
3	4.80	5.36	**39**	4.61	4.70	2
3	4.80	5.36	**40**	4.93	4.00	1
3	4.80	5.36	**57**	4.78	5.36	1
7	5.22	5.14	**8**	5.32	5.00	1
11	6.64	8.12	**67**	6.10	5.48	1
13	6.37	6.4	**14**	5.73	4.40	2
16	6.62	6.57	**17**	6.46	6.82	1
18	6.73	6.66	**19**	7.68	7.89	1
30	8.53	9.00	**63**	8.23	9.00	2
39	4.61	4.00	**40**	4.93	4.00	1
39	4.61	4.00	**57**	4.78	5.36	3
40	4.93	4.00	**57**	4.78	5.36	2
64	7.38	6.77	**65**	6.02	5.70	1
66	6.40	7.52	**71**	5.36	6.08	1

(b) Chemical Structures of Ecdysteroid Compounds Considered in *a* (from Ref 35)

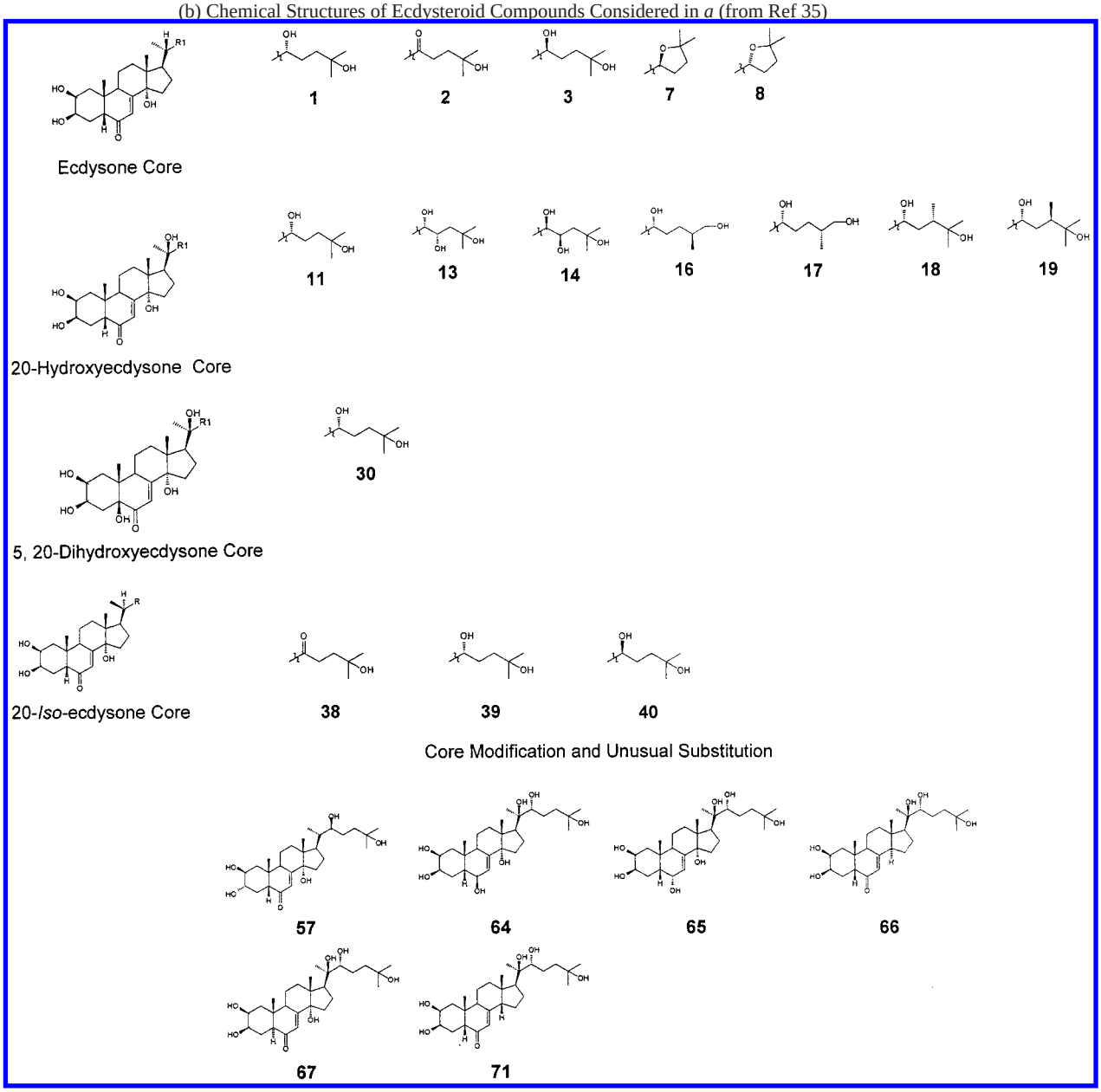

Table 3. Statistics for QSAR Models 1−8 Based on Chirality Dehscriptors and CoMFA Models A and B from Ref 35[a]

model	q^2	pred *R2*	RMS	*F*
1	0.61	0.89	0.18	39
2	0.63	0.82	0.42	22
3	0.55	0.73	0.73	13
4	0.55	0.70	0.34	12
5	0.50	0.78	0.52	18
6	0.57	0.70	0.74	11
7	0.55	0.75	0.58	15
8	0.64	0.79	0.43	18
A	0.63	0.21	1.08	1.4
		(0.52)	(0.79)	(4.4)
B	0.69	0.12	1.08	0.67
		(0.39)	(0.86)	(2.6)

[a] The training and test sets were the same in all calculations. The values of leave-one-out cross-validation R^2 (q^2) for the training set, R^2 for the test set (prediction R^2), RMS (residual mean square) and *F* values are given. The numbers shown in parentheses for models A and B were obtained when the worst outlier was excluded from the test set.[35]

fully used in QSAR studies even in cases when the CoMFA approach fails.

8. CONCLUSIONS AND PROSPECTUS

In sections 2−6, we have introduced several novel chirality descriptors. Table 1 summarizes and compares their most important features. Satisfaction of each of the criteria presented in Table 1 is considered an advantage for a class (or a subclass) of descriptors.

Many of the class I descriptors considered in this paper have two shortcomings. Thus, for a pair of enantiomers with only one asymmetric atom some of these descriptors are nonsymmetrical with respect to the corresponding nonchiral descriptors. A limitation on possible values of chirality correction can be considered as another deficiency. On the other hand, since class I descriptors are real numbers, they can be directly implemented in existing software applications.

The main drawback of the subclass IIa descriptors is that one cannot find an immediate application in the majority of the existing QSAR software since they are complex numbers. In the case of the class IIb desriptors, their imaginary parts for all achiral compounds will be equal to zero. At the same time, for molecules with chiral atoms, both the real and imaginary parts of the complex number descriptors may depend on chirality correction. This is a drawback of this descriptor subclass: it is characterized by an additional correlation between descriptors. Another shortcoming is the (large) number of these descriptors: it is twice the number of the corresponding nonchiral descriptors. On the other hand, descriptors of the IIb subclass have one attractive feature: for a pair of enantiomers with only one chiral atom the real parts of the corresponding complex descriptors are equal, while the values of the imaginary parts of complex descriptors have opposite values. Another advantage of the subclass IIb descriptors is that no additional transformation is necessary prior to their application in current QSAR approaches.

The most attractive features of the subclass IIc descriptors are as follows. For achiral molecules, these descriptors are equal to the corresponding nonchiral descriptors, and for a pair of enantiomers with only one chiral atom they are symmetrical relative to the corresponding real parts. These descriptors can also be used directly by the existing software. In the case of the IId descriptor subclass, all chirality descriptors for compounds containing no chiral atoms are equal to zero, and the descriptors of a pair of enantiomers with only one chiral atom have opposite values.

Preliminary QSAR studies of a series of ecdysteroids using several novel chirality descriptors derived from 2D molecular topology demonstrated the utility of this approach as applied to datasets containing chiral compounds. We were able to obtain models for the training sets that were of equal or better quality than those obtained with CoMFA. In addition, models obtained with the chirality descriptors had much better predictive ability than those derived with CoMFA. The success of our studies is encouraging, especially, in comparison with CoMFA, which utilizes true 3D molecular descriptors naturally sensitive to the compounds' stereochemistry. Obviously, the practical utility of all descriptors introduced in this paper could only be assessed experimentally by applying them to different datasets containing chiral compounds. Thus, the following aspects pertinent to further development and implementation of these descriptors remain in the focus of our current research.

Which chiral and nonchiral descriptors could be combined for QSAR, database mining, and other related studies? Will it be useful to introduce chirality descriptors corresponding to all nonchiral descriptors, which are derived from the adjacency matrix?

Which classes and subclasses of chirality descriptors are the most useful? Perhaps, different classes of descriptors should be used together; on the other hand, these descriptors are highly correlated and perhaps the best results can be obtained using only one class or subclass of descriptors.

What is the best value for the chirality correction? Should one apply the same chirality correction to all chiral atoms in a dataset, or should the correction depend on the atom type? In ref 50, similarity measures for mirror images were considered. The chirality correction can also be based on these measures as applied to molecular fragments.

Is the current IUPAC nomenclature for stereoisomers adequate with respect to chirality descriptors?

These and other problems are the subject of the current and future research in this challenging and intriguing area of molecular connectivity theory and its application in computer aided drug design and discovery.

ACKNOWLEDGMENT

This research is supported in part by a research grant from the National Institutes of Health (R03 MH60328-01). The authors are grateful to Dr. Robert Hormann for providing the ecdysteroid dataset and to Dr. Yunde Xiao for his improvements of the kNN QSAR method that made calculations reported in this paper more efficient.

ADDENDUM

A program to calculate most of the chirality descriptors introduced in this paper is available upon request. The program is written in the C programming language; both MS DOS and UNIX versions are available for interested readers. The requests should be sent to the following e-mail address: golbraik@email.unc.edu.

REFERENCES AND NOTES

(1) Moran, L. A.; Scrimgeour, K. G.; Horton, H. R.; Ochs, R. S.; Rawn, J. D. *Biochemistry*; Neil Patterson Publishers, Prentice Hall: Englewood Cliffs, NJ, 1994.
(2) Potapov, V. M. *Stereochemistry*; Khimia: Moscow, 1988.
(3) Mislow K. Fuzzy Restrictions and Inherent Uncertainties in Chirality Studies. In *Fuzzy Logic in Chemistry*; Rouvray, D. H., Ed.; Academic Press: San Diego, 1997; pp 65−90.
(4) Cramer, R. D., III; Patterson, D. E.; Bunce, J. D. Comparative Molecular-Field Analysis (Comfa). 1. Effect of Shape on Binding of Steroids to Carrier Proteins. *J. Am. Chem. Soc.* **1988**, *110*, 5959−5967.
(5) Marshall, G. R.; Cramer, R. D., III. 3-Dimensional Structure Activity Relationships. *Trends Pharmacol. Sci.* **1988**, *9*, 285−289.
(6) *SYBYL. Ligand-Based Design Manual;* Tripos, Inc.: St Louis, MO, 1996.
(7) Wold, S. PLS for Multivariate Linear Modeling. In *Chemometrics Methods in Molecular Design*; Waterbeemd, H. van de, Ed.; VCH: 1995; pp 195−218.
(8) Clementi, S.; Wold, S. How to Choose the Proper Statistical Methodol. In *Chemometrics Methods in Molecular Design*; Waterbeemd, H. van de, Ed.; VCH: Weinheim, Germany, 1995; pp 319−338.
(9) Pérez, C.; Pastor, M.; Ortiz, AR.; Gago, F. Comparative Binding Energy Analysis of HIV-1 Protease Inhibitors: Incorporation of Solvent Effects and Validation as a Powerful Tool in Receptor-Based Drug Design. *J. Med. Chem.* **1998**, *41*, 836−852.
(10) Cho, S. J.; Tropsha, A. Cross-Validated R^2 Guided Region Selection for Comparative Molecular Field Analysis (CoMFA): A Simple Method To Achieve Consistent Results. *J. Med. Chem.* 1995, *38*, 1060−1066.
(11) Klebe, G. Comparative Molecular Similarity Indices Analysis−CoMSIA. In *3D QSAR in Drug Design. Vol. 3. Recent Advances*; Kubinyi, H., Folkers, G., Martin, Y. C., Eds.; Kluwer/ESCOM: Dodrecht, The Netherlands, 1998; pp 87−104.
(12) Kubinyi, H.; Hamprecht, F. A.; Mietzner, T. Three-Dimensional Quantitative Similarity−Activity Relationships (3D QSiAR) from SEAL Similarity Matrices. *J. Med. Chem.* **1998**, *41*, 2553−2564.
(13) Cho, S. J.; Serrano, M. G.; Bier, J.; Tropsha, A. Structure Based Alignment and Comparative Molecular Field Analysis of Acetylcholinesterase Inhibitors. *J. Med. Chem.* **1996**, *39*, 5064−5071.
(14) Pilger, C.; Bartolucci, C.; Lamba, D.; Tropsha, A.; Fels, G. Accurate Prediction of the Bound Conformation of Galanthamine in the Active Site of Torpedo Californica Acetylcholinesterase Using Molecular Docking. *J. Mol. Graphics Modell.*, in press.
(15) Brusniak, M. Y.; Perlman, R. S.; Neve, K. A.; Wilcox, R. E. Comparative Molecular Field Analysis-Based Prediction of Drug Affinities at Recombinant D1A Dopamine Receptors. *J. Med. Chem.* **1996**, *39*, 850−859.
(16) Bucholtz, E. C.; Booth, R. B.; Wyrick, S. D.; Tropsha, A. The Effect of Region Size on CoMFA Analyses. *Med. Chem. Res.* **1999**, *9*, 675−685.
(17) Clementi, S. *GOLPE 3.0*; Multivariate Infometric Analyses (MIA): Perugia, Italy, 1995; SGI.
(18) Kubinyi, H. Variable Selection In Qsar Studies. 1. An Evolutionary Algorithm *Quant. Struct.-Act. Relat.* **1994**, *13*, 285−294.
(19) Kubinyi, H. Variable Selection in Qsar Studies. 2. A Highly Efficient Combination of Systematic Search and Evolution *Quant. Struct.-Act. Relat.* **1994**, *13*, 393−401.
(20) Hoffman, B.; Cho, S. J.; Zheng, W.; Wyrick, S. D.; Nichols, D. E.; Mailman, R. B.; Tropsha, A. Quantitative Structure−Activity Relationship Modeling of Dopamine D-1 Antagonists Using Comparative Molecular Field Analysis, Genetic Algorithms-Partial Least-Squares, and *K* Nearest Neighbor Methods. *J. Med. Chem.* **1999**, *42*, 3217−3226.
(21) Zheng, W.; Tropsha, A. Novel Variable Selection QSAR Approach Based on the *K* Nearest-Neighbor Principle. *J. Chem. Inf. Comput. Sci.* **2000**, *40*, 185−194.
(22) Bures, M. G.; Martin, Y. C. Computational methods in molecular diversity and combinatorial chemistry. *Curr. Opin. Chem. Biol.* **1998**, *2*, 376−80.
(23) MolconnZ. http://www.eslc.vabiotech.com/molconn/manuals/350/.
(24) Carhart, R. E.; Smith, D. H.; Venkataraghavan, R. Atom Pairs as Molecular-Features in Structure Activity Studies−Definition and Applications. *J. Chem. Inf. Comput. Sci.* **1985**, *25*, 64−73.
(25) Schultz, H. P.; Schultz, E. B.; Schultz, T. P. Topological Organic Chemistry. 9. Graph Theory and Molecular Topological Indices of Stereoisomeric Organic Compounds. *J. Chem. Inf. Comput. Sci.* **1995**, *35*, 864−870.
(26) Julián-Ortiz, J. V. de; Alapont, C. de G.; Ríos-Santamarina, I.; García-Doménech, R.; Gálvez, J. Prediction of properties of chiral compounds by molecular topology. *J. Mol. Graphics Modell.* **1998**, *16*, 14−18.
(27) Gutman, I.; Ruscić, B.; Trinajstić, N.; Wilcox, C. F., Jr. Graph theory and molecular orbitals. XII. Acyclic polyenes. *J. Chem. Phys.* **1975**, *62*, 3399−3405.
(28) Randić, M. On Characterization on Molecular Branching. *J. Am. Chem. Soc.* **1975**, *97*, 6609−6615.
(29) Kier, L. B.; Hall, L. H. *Molecular Connectivity in Chemistry and Drug Research*; Academic Press: New York, 1976.
(30) Kier, L. B.; Hall, L. H. *Molecular Connectivity in Structure−Activity Analysis*; Wiley: New York, 1986.
(31) Rücker, G.; Rücker, C. Counts of All Walks as Atomic and Molecular Descriptors. *J. Chem. Inf. Comput. Sci.* **1993**, *33*, 683−695.
(32) Bonchev, D. Overall Connectivity and Molecular Complexity. In *Topological Indices and Related Descriptors*; Devillers, J., Balaban, A. T., Eds.; Gordon and Breach: Reading, U.K., 1999; pp. 361−401.
(33) Bonchev, D. Novel Indices for the Topological Complexity of Molecules. *SAR/QSAR Environ. Res.* **1997**, *7*, 23−43.
(34) Gálvez, J.; García-Domenech, R.; Julian-Ortiz, J. V. de; Soler, R. Topological approach to drug design. *J. Chem. Inf. Comput. Sci.* **1995**, *35*, 272−284.
(35) Dinan, L.; Hormann, R. E.; Fujimoto T. An extensive ecdysteroid CoMFA. *J. Comput.-Aided Mol. Des.* **1999**, *13* (Mar), 185−207.
(36) *Ecdysone: From Chemistry to Mode of Action*; Koolman, J., Ed.; Thieme: Leipzig, Germany, 1989.
(37) Kier, L. B. A shape index from molecular graphs. *Quant. Struct.-Act. Relat.* **1985**, *4*, 109−116.
(38) Kier, L. B. Inclusion of symmetry as a shape attribute in kappa-index analysis. *Quant. Struct-Act. Relat.* **1987**, *6*, 8−12.
(39) Hall, L. H.; Kier, L. B. Determination of topological equivalence in molecular graphs from the topological state. *Quant. Struct.-Act. Relat* **1990**, *9*, 115−131.
(40) Hall, L. H.; Mohney, B. K.; Kier, L. B. The Electrotopological State: An Atom Index for QSAR. *Quant. Struct.-Act. Relat.* **1991**, *10*, 43−51.
(41) Hall, L. H.; Mohney, B. K.; Kier, L. B. The Electrotopological State: Structure Information at the Atomic Level for Molecular Graphs. *J. Chem. Inf. Comput. Sci.* **1991**, *31*, 76−82.
(42) Kier, L. B.; Hall, L. H. *Molecular Structure Description: The Electrotopological State*; Academic Press: New York, 1999.
(43) Kellogg, G. E.; Kier, L. B.; Gaillard, P.; Hall, L. H. The E-State Fields. Applications to 3D QSAR. *J. Comput.-Aided Mol. Des.* **1996**, *10*, 513−520.
(44) Kier, L. B.; Hall, L. H. A Differential Molecular Connectivity Index. *Quant. Struct.-Act. Relat.* **1991**, *10*, 134−140.
(45) Petitjean, M. Applications of the Radius-Diameter Diagram to the Classification of Topological and Geometrical Shapes of Chemical Compounds. *J. Chem. Inf. Comput. Sci.* **1992**, *32*, 331−337.
(46) Wiener, H. Structural Determination of Paraffin Boiling Points. *J. Am. Chem. Soc.* **1947**, *69*, 17.
(47) Platt, J. R. Prediction of Isomeric Differences in Paraffin Properties. *J. Phys. Chem.* **1952**, *56*, 328.
(48) Shannon, C.; Weaver, W. *Mathematical theory of Communication*; University of Illinois: Urbana, 1949.
(49) Bonchev, D.; Mekenyan, O.; Trinajstić, N. Isomer discrimination by topological information approach. *J Comput Chem* **1981**, *2*, 127−148.
(50) Mezey, P. G. Fuzzy Measures of Molecular Shape and Size. In *Fuzzy Logic in Chemistry*; Rouvray, D. H., Ed.; Academic Press: San Diego, 1997; pp 139−223.

CI000082A