Novel *ZE*-Isomerism Descriptors Derived from Molecular Topology and Their Application to QSAR Analysis

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, and Program for Theory of Complex Systems, Texas A&M University, Galveston, Fort Crockett Campus, 5007 Avenue U, Galveston, Texas 77551

We introduce several series of novel *ZE*-isomerism descriptors derived directly from two-dimensional molecular topology. These descriptors make use of a quantity named *ZE*-isomerism correction, which is added to the vertex degrees of atoms connected by double bonds in *Z* and *E* configurations. This approach is similar to the one described previously for topological chirality descriptors (Golbraikh, A., et al. *J. Chem. Inf. Comput. Sci.* **2001**, *41*, 147−158). The *ZE*-isomerism descriptors include modified molecular connectivity indices, overall Zagreb indices, extended connectivity, overall connectivity, and topological charge indices. They can be either real or complex numbers. Mathematical properties of different subgroups of *ZE*-isomerism descriptors are discussed. These descriptors circumvent the inability of conventional topological indices to distinguish between *Z* and *E* isomers. The applicability of *ZE*-isomerism descriptors to QSAR analysis is demonstrated in the studies of a series of 131 anticancer agents inhibiting tubulin polymerization.

1. INTRODUCTION

The first molecular topological descriptor was introduced by Wiener[1] in 1947 in his studies of boiling points of acyclic hydrocarbons. Since then, numerous molecular topological descriptors[2−4] have been proposed and used widely in quantitative structure−property relationship (QSPR) and structure−activity relationship (QSAR) analyses.[5] For instance, Molconn-Z program developed by Kier and Hall calculates several hundreds of molecular topological indices.[6] Another popular software package for studying structure−property and structure−activity relationships, CODESSA,[7] calculates nearly 400 molecular descriptors, and about half of them represent various topological indices. Around 1800 descriptors are included in the "Handbook of Molecular Descriptors" published recently,[3] and a considerable part of them are topological indices. A recent monograph entitled "Topological Indices and Related Descriptors in QSAR and QSPR"[8] also emphasized the popularity and a widespread use of topological indices in QSAR and QSPR. Various topological indices have been successfully employed in database mining and combinatorial library design.[9,10] It has been shown that topological indices selected by a variable selection QSAR procedure can serve as a "descriptor pharmacophore" in the procedures of mining of chemical databases and virtual libraries.[11]

Molecular topological descriptors are invariants of hydrogen-suppressed molecular graphs. Since molecular graphs (or structures) are planar images of molecules, molecular topological descriptors are often referred to as two-dimensional (2D) descriptors, and QSAR methods utilizing topological descriptors are frequently called 2D-QSAR methods. 2D-QSAR methods offer a certain advantage over 3D-QSAR approaches. For instance, as compared to a popular 3D-QSAR method, Comparative Molecular Field Analysis (CoMFA),[12,13] 2D-QSAR approaches are easier to use and allow a higher degree of automation. These methods do not require several time-consuming attributes of 3D-QSAR such as conformational search, unique spatial alignment of molecules (which is impossible, if all molecules are highly flexible and the spatial structure of the receptor binding site is unknown), formal definition of a 3D pharmacophore, etc. Furthermore, 2D-QSAR methods naturally avoid problems of 3D-QSAR methods dealing with the calculations of molecular fields on a grid as discussed in detail in several earlier publications from our[14−16] and other[17] groups. It has been demonstrated that QSAR models built with topological descriptors afford better statistics and higher predictive ability as compared to CoMFA models.[18−20] It has been also shown that 2D descriptors are more efficient than 3D descriptors in the analysis of molecular diversity and database mining.[20]

Despite many successful applications in computer-assisted drug design, topological descriptors generally lack an important feature natural to 3D descriptors, namely, the ability to take into account stereospecific properties of molecules such as atomic chiralities, and thus discriminate between enantiomers and σ-diastereomers. This critical drawback recently motivated us to introduce chirality topological descriptors;[21] other groups have also considered this problem.[22−25] In our work,[21] chirality descriptors were developed using a quantity named chirality correction, which was added to (in the case of R-configuration) or subtracted from (in the case of S-configuration) the vertex degree of asymmetric atoms of hydrogen-depleted molecular graphs.[21]

* Corresponding author phone: (919)966-2955; e-mail: tropsha@email.unc.edu.
† University of North Carolina at Chapel Hill.
‡ Texas A&M University.

Modified vertex degrees were used in standard formulas defining overall Zagreb indices,[26] molecular connectivity indices,[27−30] extended connectivity indices,[31] overall connectivity indices,[32,33] and topological charge indices.[23,34] The chirality descriptors developed by us earlier were used in several successful QSAR studies of different molecular data sets. In all cases, the resulting QSAR models compared favorably with those obtained with different, mainly 3D-QSAR models.[21,35]

In this paper, we consider cis- and trans-isomers defined by double bonds. In the literature, this kind of isomerism is referred to as π-diastereomerism, since in cis−trans-isomers the atoms adjacent to those connected by the double bond are positioned in the plane perpendicular to the main axis of the π-electronic density. 2D descriptors, which are currently used in the majority of QSAR applications, database mining, and design of combinatorial libraries do not take π-diastereomerism into account. An attempt to introduce cis−trans isomerism descriptors was undertaken by G. Lekishvili[36] followed by further developments.[37] The indices introduced in ref 36 were based on a generalized definition of the graph isomorphism, which was used in modification of the adjacency matrix: the author used imaginary number i (square root of −1) or −i instead of zero for the adjacency matrix elements corresponding to two atoms separated by three bonds (where the central bond is a double bond). The signs of these entries depended on whether these atoms were on the same or on different sides of the middle bond. This modified matrix was used to calculate topological indices, which in this case were complex numbers $A+iB$. Since complex numbers cannot be used as descriptors directly, the author used the descriptors in the following general form, $\sqrt{A^2+nB^2} - mB|B|$, where m and n are parameters that could be varied to obtain better models. As an application example, the author showed that for these indices there is an excellent correlation between the modified Randić index and the standard entropies of formation of alkenes.[36]

Herein, we propose a different approach to the description of cis- and trans-isomers, based on Z- and E-designations as they are defined in the IUPAC nomenclature rules. Following the general approach introduced earlier for the chirality descriptors,[21] we introduce a quantity named *ZE-isomerism correction* for the vertex degrees of atoms connected by double bonds in Z- or E-configuration. In the case of a cumulene group with an odd number of double bonds, ZE-isomerism correction is added to the vertex degrees of atoms at the termini of the group. For brevity, we will not formally consider cumulenes, implying that our results are valid for these molecules as well. Similarly to the chirality correction,[21] ZE-isomerism correction can be a real or an imaginary number. To demonstrate practical importance of novel ZE-isomerism descriptors, we consider their application to QSAR analysis of a series of 113 anticancer agents that inhibit tubulin polymerization.[38,39] We anticipate that ZE isomerism descriptors will become useful in QSAR/QSPR analysis, database mining, and molecular diversity studies.

2. METHODOLOGY

2.1. General Properties of the *ZE*-Isomerism Descriptors. We shall consider the *ZE*-isomerism descriptors for pairs of π-diastereomers with only one, several (but not all),

and all double bonds in the opposite configurations. Let ξ be the *ZE*-isomerism correction, $D^* = D^*(\xi; a_1, a_2,..., a_N)$ be a descriptor value, computed without taking into account the *ZE*-isomerism of a double bond, or that of several or all *ZE* bonds [in the latter case, $D^* = D(a_1, a_2,..., a_N)$, where D is the corresponding conventional descriptor]; $a_1, a_2,..., a_N$ are vertex degrees of the hydrogen-suppressed molecular graph, and N is the number of non-hydrogen atoms in a molecule. Let $D^1 = D^1(\xi; a_1, a_2,..., a_N)$ and $D^2 = D^2(\xi; a_1, a_2,..., a_N)$ be the *ZE*-isomerism descriptors for a pair of π-diastereomers with one, several (but not all), or all Z- or E-double bonds in opposite configurations. ξ is considered a real number, and the imaginary isomerism correction will be represented as $i\xi$. We do not impose any particular limit on ξ and a_i, $i = 1,..., N$, i.e. they can have any real value permitted by D, D^*, D^1, and D^2 functions. We suggest that ideally the *ZE*-isomerism descriptors for these pairs of isomers must satisfy the following conditions (cf. ref 24).

(1) D^1 *and* D^2 *are continuous functions of* ξ *such that* D^1-$(0; a_1, a_2,..., a_N) = D^2(0; a_1, a_2,..., a_N) = D$.

(2r) *If a ZE-isomerism descriptor is a real number, its values for isomers with only one or several (but not necessarily all) ZE-double bonds in the opposite configurations must be symmetrical relative to the value of D^*, and for isomers with all ZE-double bonds in the opposite configurations they must be symmetrical relative to the value of D.*

(2i) *If a ZE-isomerism descriptor is a complex number, its values for isomers with one or several (but not all) ZE-double bonds in opposite configurations must also be symmetrical relative to the value of D^* (which in this case is generally a complex number). For isomers with all ZE-double bonds in the opposite configurations these values must be complex conjugates of each other, i.e., their real parts must be equal to each other, their imaginary parts must be opposite numbers and, in addition, the real parts must be equal to D.*

In general, D^1 and D^2 satisfy condition (1) due to the continuity of D and due to the way of defining D^1 and D^2 descriptors as follows (except for topological charge indices, see Section 6 below). Suppose atoms 1, 2,..., $2K$ are incident to all K Z- or E-double bonds, and atoms $2K+1$, $2K+2$,..., N are not. Then, for isomers with all *ZE*-double bonds in the opposite configurations, we define $D^1 = D^1(\xi; a_1, a_2,..., a_{2K}, a_{2K+1},..., a_N) = D(a_1+\xi, a_2+\xi,...,a_{2K}+\xi, a_{2K+1},..., a_N)$ and $D^2 = D^2(\xi; a_1, a_2,..., a_{2K}, a_{2K+1},..., a_N) = D(a_1-\xi, a_2-\xi,..., a_{2K}-\xi,a_{2K+1},...,a_N)$ for the real *ZE*-isomerism correction and $D^1 = D^1(\xi; a_1, a_2,..., a_{2K}, a_{2K+1},..., a_N) = D(a_1+i\xi, a_2+i\xi,...,a_{2K}+i\xi, a_{2k+1},..., a_N)$ and $D^2=D^2(\xi; a_1, a_2,..., a_N) = D(a_1-i\xi, a_2-i\xi,...,a_{2K}-i\xi, a_{2k+1},...,a_N)$ for the imaginary *ZE*-isomerism correction. In these expressions, ξ is positive (negative) for Z- and negative (positive) for E-double bond. Thus, if $\xi=0$, $D^1 = D^2 = D$. Similarly, if atoms 1, 2,..., $2L$ are incident to *ZE*-double bonds, which have the same configuration in both *ZE*-isomers, we can write down for the real and imaginary *ZE*-isomerism correction $D^*= D(a_1+\xi, a_2+\xi,..., a_{2L}+\xi, a_{2L+1},..., a_N)$ and $D^*= D(a_1+i\xi, a_2+i\xi,...,a_{2L}+i\xi, a_{2L+1},..., a_N)$, respectively. In this case, for the real *ZE*-isomerism correction we have $D^1 = D(a_1+\xi, a_2+\xi,...,a_{2L}+\xi, a_{2L+1}+\xi,...,\alpha_{2K}+\xi, a_{2k+1},..., a_N)$ and $D^2 = D(a_1+\xi, a_2+\xi,...,a_{2L}+\xi, a_{2L+1}-\xi,...,\alpha_{2K}-\xi, a_{2K+1},..., a_N)$, whereas for imaginary isomerism correction $D^1 = D(a_1+i\xi, a_2+i\xi,...,a_{2L}+i\xi, a_{2L+1}+i\xi,...,\alpha_{2K}+i\xi, a_{2k+1},...,$

NOVEL ZE-ISOMERISM DESCRIPTORS

J. Chem. Inf. Comput. Sci., Vol. 42, No. 4, 2002 **771**

a_N) and $D^2 = D(a_1+i\xi, a_2+i\xi,...,a_{2L}+i\xi, a_{2L+1}-i\xi,...,\alpha_{2K}-i\xi, a_{2K+1},..., a_N)$ are obtained. As we will see, when the real ZE-isomerism descriptors are nonlinear functions of vertex degrees, condition (2r) is generally not satisfied. Nevertheless, we will show that in this case the following weaker condition is satisfied.

(2r′) *If a ZE-isomerism descriptor is a real number, its values for isomers with only one, several (but not all), or all ZE-double bonds in the opposite configurations must be symmetrical relative to the value* $D+f(|\xi|)$ *for all permissible values of* ξ, *where* $f(|\xi|)$ *is a continuous function.*

$f(\xi)$ is an even function: it depends on the absolute value of the ZE-isomerism correction; thus it has the same value for both ZE-isomers. If it depended on the sign of ξ, $f(\xi)$ would be different from $f(-\xi)$, and $D+f(\xi)$ would be different from $D+f(-\xi)$, and there would be no symmetry of D^1 and D^2 relative to $D+f(\xi)$ for all permissible values of ξ. However, we can always define a function $g(|\xi|) = [f(\xi) + f(-\xi)]/2$, which is independent of the sign of ξ. As we will see, under certain conditions, $g(|\xi|)$ will be, in fact, our $f(|\xi|)$. Function $f(|\xi|)$ must satisfy the condition $f(0) = 0$. Indeed, suppose $f(0) > 0$. Then, due to the continuity of $f(|\xi|)$ and the fulfillment of condition (1), such a vicinity of $\xi = 0$ could always be found, in which $D+f(|\xi|) > D^1$ and $D+f(|\xi|) > D^2$. In other words, $D+f(|\xi|)$ would not be between D^1 and D^2, which would contradict the condition that D^1 and D^2 are symmetrical relative to $D+f(|\xi|)$. The case $f(0) < 0$ can be considered in a similar manner. If conditions (1) and (2r′) are satisfied, it can also be proven that for ZE-isomers with one or several (but not necessarily all) ZE-double bonds in the opposite configurations, D^1 and D^2 must be symmetrical relative to $D^*+f_1(|\xi|)$, where $f_1(|\xi|)$ is a continuous function of ξ, and $f_1(0) = 0$. Indeed, in this case, we can write $D^*+f_1(|\xi|) = D+f(|\xi|)$, and define $f_2(|\xi|) = f(|\xi|)-f_1(|\xi|)$. Then $D^* = D+f_2(|\xi|)$. Function $f_2(|\xi|)$ is continuous of ξ and $f_2(0) = 0$ due to our definition of D^* (see above). Both functions $f(|\xi|)$ and $f_2(|\xi|)$ satisfy these conditions; therefore, $f_1(|\xi|)$ must also satisfy these conditions.

Condition (2i) for the imaginary parts of complex ZE-isomerism descriptors is satisfied for isomers with all ZE-double bonds in opposite configurations only partially, namely, the ZE-isomerism descriptor values for them are complex conjugates. Indeed, if $a_1, a_2,...,a_N$ are fixed, D^1 and D^2 are only functions of ξ, such that $D^1(0)$ and $D^2(0)$ are real numbers and $D^1(i\xi) = D^2(-i\xi)$. Calculations of descriptors include only additions, multiplications, and raising to powers, which in this case for opposite imaginary arguments give complex conjugates (see Sections 3−6). Obviously, if a descriptor is a linear function of ξ, the real parts of its values are equal to D. If a descriptor is a nonlinear function of ξ, this condition is generally not satisfied. Indeed, if it is a nonlinear function of ξ, it can be expanded as a power series in ξ, since it is a regular function at point $\xi = 0$ and continuous anywhere within the area of its definition. When the descriptors are functions of the complex variable ξ, they would have no singularities in the finite points on the complex plane, except for molecular connectivity indices (see Sections 3−6). Molecular connectivity indices have singularities at the isolated points $\xi = -a_k$, $k = 1,...,2K$, where K is the number of Z- and E-double bonds. Nevertheless, we can conclude that molecular connectivity indices can be

expanded as power series in ξ within the range of $0 < \xi < \min(a_k)$. Terms of the series containing even powers of ξ will contribute to the real part of D^1 and D^2, and their contributions into D^1 and D^2 will be the same. Thus, $Re(D^1) = Re(D^2) = D+f_1(|\xi|) \neq D$. In a manner similar to that considered above, it can be shown that $f_1(|\xi|)$ is a continuous function, and $f_1(0) = 0$. If a total number of Z- or E-double bonds in a pair of π-diastereomers is greater than the number of Z- or E-double bonds in the opposite configurations, the condition (2i) will be satisfied provided that the ZE-isomerism descriptor is a linear function of ξ. If the descriptor is a nonlinear function of ξ, it may include multiplications of D* (which in this case is a complex number) by different orders of $i\xi$. For instance, if the descriptor includes the multiplication of D* and $(i\xi)^2 = -\xi^2$, we obtain for both isomers real terms $-\xi^2 Re(D^*)$ and imaginary terms $-\xi^2 Im(D^*)$. Thus, the real and imaginary parts of the descriptor values will not be symmetrical relative to $Re(D^*)$ and $Im(D^*)$. Consequently, we can only assert that the descriptor values will be symmetrical relative to $D^*+g(|\xi|)$, where $g(|\xi|)$ is a continuous complex function, and $g(0) = 0$. We can also assert that the real parts of descriptor values will be symmetrical relative to $D+f_1(|\xi|)$, and imaginary parts will be symmetrical relative to $f_2(|\xi|)$, where both $f_1(|\xi|)$ and $f_2(|\xi|)$ are continuous functions such that $f_1(0) = 0$ and $f_2(0) = 0$. We conclude that complex ZE-isomerism descriptors satisfy the following condition.

(2i′). *If a ZE- isomerism descriptor is a complex number, its real parts for isomers with one or several (but not necessarily all) ZE-double bonds in the opposite configurations must be symmetrical relative to* $D+f_1(|\xi|)$, *and its imaginary parts must be symmetrical relative to* $f_2(|\xi|)$, *where* $f_1(|\xi|)$ *and* $f_2(|\xi|)$ *are continuous functions such that* $f_1(0) = f_2(0) = 0$. *Descriptor values for ZE-isomerism isomers with all ZE- double bonds in the opposite configurations must be complex conjugates.*

In fact, the continuity of D and our definition of ZE-isomerism descriptors (see above) provide the fulfillment of conditions (2r′) or (2i′) for real or imaginary ZE-isomerism descriptors, respectively. Our chirality descriptors introduced in ref 21 also satisfy the above conditions.

Randić[24] introduced a condition [here it will be referred to as the Randić condition or condition (R)] that chirality descriptors for enantiomers must be opposite numbers, while for achiral compounds they must necessarily be equal to zero. Here we extend this condition to ZE- isomerism descriptors:

(R) *The ZE-isomerism descriptors of isomers with all ZE-double bonds in opposite configurations must be opposite numbers, while for compounds without Z- and E-double bonds they must be equal to zero.*

We believe that this condition is too strong, although as we will show that it can be satisfied in some cases. Using this condition in the process of discrimination between the "true" and "false" chirality or ZE-isomerism descriptors will result in discarding many useful descriptors. Our approach of modifying conventional descriptors using chirality or ZE-isomerism correction makes the descriptors "corrected" by taking into account chirality or ZE-isomerism. These descriptors do not necessarily have to satisfy the Randić condition.

Lekishvili[37] proposed the concept of orthogonality between descriptors describing structural (without taking isomerism into account) and spatial (isomerism) constitution of com-

pounds. According to this author, A = Re(D) and B = Im-(D) describe the structural and spatial information, where D is a complex descriptor, and orthogonality between A and B means r(A,B) = 0, where r is the correlation coefficient between A and B calculated for the training set. Using two training sets, the author showed[37] that his descriptors satisfy this condition automatically and generalized that his A and B descriptors are orthogonal. In fact, it is only an artifact. In both data sets considered in ref 37, for each isomer containing cis- and trans- bonds, there is a corresponding dual isomer in which all trans-bonds are substituted for cis-bonds, and all cis-bonds are substituted for trans-bonds. It can be said that these data sets are "symmetrical" (or invariant) relative to the substitution of all cis-bonds with trans-bonds and simultaneously all trans-bonds with cis-bonds (i.e. the data sets will not change, if one applies this transformation to all molecules). In this case, $\langle B \rangle = 0$, where $\langle B \rangle$ is the average value of B, and for each compound with $B_i \neq 0$ there exists a compound with $B_k = -B_i$. At the same time, for these compounds $A_i = A_k$. This immediately leads to r(A,B) = 0. In the general case, when a data set is not symmetrical in the above sense, r(A,B) will not be equal to zero, and descriptors A and B will not be orthogonal.

2.2. ZE-Isomerism Correction. 2.2.1. Modification of Vertex Degrees. Series of molecular descriptors are calculated using the adjacency matrix for hydrogen-depleted molecular graph. Each non-hydrogen atom of a molecule is a vertex and each bond connecting these atoms is an edge of this graph. Let N be the number of vertices of this graph. The adjacency matrix A of this graph is a square symmetric NxN matrix such that its elements $a_{ij} = 1$, if there is an edge connecting vertices i and j, otherwise $a_{ij} = 0$. Vertex degree a_i is defined as the sum of the *i*th row matrix elements:

$$a_i = \sum_{j=1}^{N} a_{ij} \qquad (1)$$

Thus, a_i is equal to the number of edges incident to the *i*th vertex. Series of descriptors for molecules containing π-diastereomeric groups can be defined by introducing ZE-isomerism correction ξ for each atom incident to the corresponding double bonds. The corresponding vertex degrees a_i are substituted with $(a_i + \xi)$ for Z-configuration and with $(a_i - \xi)$ for E-configuration. This transformation is equivalent to making main diagonal elements a_{ii} of matrix A equal to ξ or $-\xi$, for all atoms of double bonds in Z- or E-configuration, respectively. After taking ZE-isomerism into account, vertex degrees can be calculated using eq 1. In most cases, for atoms incident to a double bond in Z- or E-configuration, $a_i = 2$ or 3.

2.2.2. Two Classes of ZE-Isomerism Descriptors. ZE-isomerism correction ξ can be a real or an imaginary number. In the latter case, we denote the ZE-isomerism correction as $i \xi$, where $i = \sqrt{-1}$, always implying that ξ is a real number. Therefore, if the ZE-isomerism correction is an imaginary number, the corresponding vertex degree will be a complex number equal to $(a_i + i\xi)$ or $(a_i - i\xi)$ for an atom incident to a double bond in Z- or E-configuration, respectively. Subsequently, we introduce two classes of ZE-isomerism descriptors: class I that is based on the real number ZE-isomerism correction and class II that is based on the imaginary number ZE-isomerism correction.

Table 1. Real ZE-Isomerism Descriptors Based on Imaginary ZE-Isomerism Correction and Complex ZE-Isomerism Descriptors $d = \text{Re}(d) + \text{Im}(d)$

subclass	real descriptors
subclass IIb	Re(d) and Im(d)
subclass IIc	Re(d)+Im(d)
subclass IId	Arctan(Re(d),Im(d))

Since class I descriptors are real numbers, no additional transformation is necessary to use them in QSAR analysis or chemical database related applications. It is not the case for class II descriptors, for which several subclasses were defined in this work (as in our earlier publication[21]) as follows.

(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 QSAR or other molecular data processing studies. These complex descriptors will be referred to as ZE-isomerism descriptors of *subclass IIa* of class II. It is the most natural definition of the class II descriptors. Unfortunately, these descriptors cannot be currently applied in the majority of the QSAR and database mining software, which is not adapted yet to employ complex descriptors.

(ii) The real and imaginary part of a complex descriptor as two real descriptors (ZE-isomerism descriptors of *subclass IIb*).

(iii) ZE-isomerism descriptors of *subclass IIc* are defined as

$$D_c = \text{Re}(d) + \text{Im}(d) \qquad (2)$$

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

(iv) ZE-isomerism descriptors of *subclass IId* are defined as follows:

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

These subclasses of descriptors are presented in Table 1. Since descriptors of subclasses IIb, IIc, and IId are real numbers, in the subsequent sections we examine whether they satisfy conditions (2), (2r′), or (R) for real ZE-isomerism descriptors.

2.2.3. Symmetry and Asymmetry of ZE-Isomerism Descriptors for a Pair of Z- and E-Isomers. We have described symmetry and asymmetry of ZE-isomerism descriptors for a pair of ZE-isomers in relation to the corresponding conventional descriptors or some other numbers discussed in Section 2.1. Here we give the necessary definitions. Let D^1 and D^2 ($D^1 > D^2$) be the values of a ZE-isomerism descriptor for a pair of ZE-isomers, and P be some number satisfying the condition $D^1 > P > D^2$. If $D^1 - P =$

$P - D^2$, then the descriptors D^1 and D^2 are symmetrical relative to number P. If $D^1 - P < P - D^2$, the degree of symmetry of D^1 and D^2 relative to number P is defined as $S = (D^1 - P)/(P - D^2)$; if $D^1 - P > P - D^2$, it is defined as $S = (P - D^2)/(D^1 - P)$. The degree of asymmetry Y of D^1 and D^2 relative to number P is defined by the formula

$$Y = 1 - S = \begin{cases} 1 - \dfrac{D^1 - P}{P - D^2}, \text{ if } D^1 - P < P - D^2 \\ \\ 1 - \dfrac{P - D^2}{D^1 - P}, \text{ if } D^1 - P > P - D^2 \end{cases} \quad (4)$$

This definition guarantees that $Y \in [0,1]$. Let D be a descriptor value, calculated without accounting for the Z- and E-double bonds, defining a pair of *ZE*-isomers, and D^1 and D^2 be the values of the corresponding *ZE*-isomerism descriptor. Then if $P = (D^1 + D^2)/2 - D$, $Y = 0$. Since D, D^1, and D^2 are continuous functions of all their variables, P is also a continuous function of these variables, and since $D^1(\xi=0) = D^2(\xi=0) = D$, then $P(\xi=0) = 0$. At last, since $D^1(\xi) = D^2(-\xi)$, then $P(\xi) = P(-\xi)$. Therefore, $P = \mathrm{f}(|\xi|)$ in condition (2r′) (see Section 2.1), i.e., D^1 and D^2 are symmetrical relative to $D+\mathrm{f}(|\xi|)$, where $\mathrm{f}(|\xi|)$ is a continuous function and $\mathrm{f}(0) = 0$. In the case of imaginary *ZE*-isomerism correction, symmetry of real and imaginary parts of a *ZE*-descriptor can be considered separately, and P_1 and P_2 can be introduced to define $\mathrm{f}_1(|\xi|)$, or $\mathrm{f}_2((|\xi|)$ (see condition (2i′) in Section 2.1). In several simple cases, we will obtain the downright expressions for $\mathrm{f}(|\xi|)$, $\mathrm{f}_1(|\xi|)$, and $\mathrm{f}_2((|\xi|)$.

3. OVERALL ZAGREB INDICES $^N M_1$ AND $^N M_2$

The set of overall Zagreb indices $^i M_1$ (i is the order of the index: it is defined as the total number of edges in the corresponding subgraphs) is defined[21,26,40] by the following formulas:

$$^0 M_1 = \sum_{i=1}^{N} a_i^2, \quad ^1 M_1 = \sum_{\text{all edges}} a_{i_1}^2 a_{i_2}^2, ...,$$

$$^{n-1} M_1 = \sum_{\text{all } (n-1)\text{-edge subgraphs}} a_{i_1}^2 a_{i_2}^2 ... a_{i_\nu}^2 \quad (5)$$

Similarly, the set of Zagreb indices $^i M_2$ is defined[21,26,40] as follows

$$^1 M_2 = \sum_{\text{all edges}} a_{i_1} a_{i_2}, \quad ^2 M_2 = \sum_{\text{all 2-edge subgraphs}} a_{i_1} a_{i_2} a_{i_3}, ...,$$

$$^{n-1} M_2 = \sum_{\text{all } (n-1)\text{-edge subgraphs}} a_{i_1} a_{i_2} ... a_{i_\nu} \quad (6)$$

where ν is the number of vertices in $(n-1)$-edge subgraph. (Gutman et al.[26] only define $^0 M_1$ and $^1 M_2$.)

3.1. Class I *ZE*-Isomerism Overall Zagreb Indices. To obtain class I *ZE*-isomerism descriptors, vertex degrees a_j must be replaced by $(a_j+\xi)$ for atoms connected with double bonds in Z-configuration and by $(a_j-\xi)$ for atoms connected with double bonds in E-configuration. Thus, a_j^2 must be replaced by $(a_j \pm \xi)^2 = a_j^2 \pm 2a_j\xi + \xi^2$. (The plus sign in $\pm$ and minus sign in $\mp$ in this and in all subsequent formulas refer to a bond in the Z-configuration, and the minus sign in $\pm$ and the plus sign in $\mp$ refer to a bond in the E-

configuration.) If the total number of Z- and E-double bonds are n_Z and n_E, then

$$^0 M_1^{ZE} =$$

$$^0 M_1 + 2\xi[\sum_{j=1}^{n_Z}(a_{j_1} + a_{j_2}) - \sum_{j=1}^{n_E}(a_{j_1} + a_{j_2})] + 2\sum_{j=1}^{n_Z+n_E} \xi^2$$

$$= {}^0 M_1 + 2\xi[\sum_{j=1}^{n_Z}(a_{j_1} + a_{j_2}) - \sum_{j=1}^{n_E}(a_{j_1} + a_{j_2})] +$$

$$2\xi^2(n_Z + n_E) \quad (7)$$

since each double bond connects two atoms (denoted as j_1 and j_2). Thus, $^0 M_1^{ZE}$ satisfies condition (1) (see Section 2.1), since $^0 M_1^{ZE}$ is a continuous function of ξ, and if $\xi \to 0$, $^0 M_1^{ZE} \to {}^0 M_1$. From formula 7, for a pair of *ZE*-isomers with only one *ZE*-double bond in the opposite configurations we obtain

$$^0 M_1^{ZE} = {}^0 M_1^{ZE*} \pm 2\xi(a_{j_1} + a_{j_2}) + 2\xi^2$$

where $^0 M_1^{ZE*}$ is the value of the descriptor without accounting for the *ZE*-double bond, defining the isomers. Thus, the values of the $^0 M_1^{ZE}$ index are symmetrical with respect to the value of $^0 M_1^{ZE*} + 2\xi^2$ but not to the value of $^0 M_1^{ZE*}$. For a pair of *ZE*-isomers with all *ZE*-double bonds in the opposite configurations, the values of all $^0 M_1^{ZE}$ indices are symmetrical with respect to the value of $^0 M_1 + 2\xi^2(n_Z + n_E)$ but not to the value of the conventional $^0 M_1$ index. Similar conclusion can be made for the *ZE*-isomers with several (but not all) double bonds in the opposite configurations. In this case,

$$^0 M_1^{ZE} = {}^0 M_1^{ZE*} + 2\xi[\sum_{j=1}^{n_Z}(a_{j_1} + a_{j_2}) - \sum_{j=1}^{n_E}(a_{j_1} + a_{j_2})] +$$

$$2\xi^2(n_Z + n_E)$$

where n_Z and n_E denote the number of Z- and E-bonds in the opposite configurations, and $^0 M_1^{ZE*}$ is the descriptor value calculated without taking *ZE*-isomerism of these $n_Z + n_E$ bonds into account. Thus, $^0 M_1^{ZE}$ satisfy condition (2r′) but not the condition (2r) (see Section 2.1).

Consider the following term $a_{i_1}^2 a_{i_2}^2 ... a_{i_\nu}^2$. A sum of such terms defines $^{n-1} M_1$ for $n > 1$. Some of these terms contain only one of the atoms that belong to a Z- or E-bond, and some contain both. Consider these cases separately. In the first case, the term can be overwritten in the form $a_{i_1}^2 a_{i_2}^2 ... a_{i_k}^2 ... a_{i_\nu}^2$, where vertex i_k corresponds to an atom incident to a Z- or E-double bond. In this case, $(a_{i_k} \pm \xi)$ must substitute for a_{i_k}. Let $a_{i_1}^2 a_{i_2}^2 ... a_{i_{k-1}}^2 a_{i_{k+1}}^2 ... a_{i_\nu}^2 = A_s$, where $s = 1,..., S_1$, and S_1 is the total number of $(n-1)$-edge subgraphs containing only one of the vertices i_k incident to the Z- or E-double bond. In the second case, the term can be represented as $a_{i_1}^2 a_{i_2}^2 ... a_{i_k}^2 a_{i_{k+1}}^2 ... a_{i_\nu}^2$, where i_k and i_{k+1} are the atoms connected by Z- or E-double bond. Let $a_{i_1}^2 a_{i_2}^2 ... a_{i_{k-1}}^2 a_{i_{k+2}}^2 ... a_{i_\nu}^2 = B_s$, where $s = 1,..., S_2$, and S_2 is the total number of $(n-1)$-edge subgraphs containing vertices i_k and i_{k+1}. Then for $^{n-1} M_1^{ZE}$ we obtain

$$^{n-1}M_1^{ZE} = {}^{n-1}M_1^{ZE0} + \sum_{s=1}^{S_1} A_s(a_{i_k}^2 \pm 2\xi a_{i_k} + \xi^2) +$$

$$\sum_{s=1}^{S_2} B_s(a_{i_k}^2 \pm 2\xi a_{i_k} + \xi^2)(a_{i_{k+1}}^2 \pm 2\xi a_{i_{k+1}} + \xi^2) =$$

$$^{n-1}M_1^{ZE*} \pm 2\xi[\sum_{s=1}^{S_1} A_s a_{i_k} + \sum_{s=1}^{S_2} B_s a_{i_k} a_{i_{k+1}}(a_{i_k} + a_{i_{k+1}})] +$$

$$\xi^2[\sum_{s=1}^{S_1} A_s + \sum_{s=1}^{S_2} B_s(a_{i_k}^2 + a_{i_{k+1}}^2 + 4a_{i_k} a_{i_{k+1}})] \pm$$

$$2\xi^3 \sum_{s=1}^{S_2} B_s(a_{i_k} + a_{i_{k+1}}) + \xi^4 \sum_{s=1}^{S_2} B_c \quad (8)$$

In (8), $^{n-1}M_1^{ZE0}$ denotes the term independent of i_k and i_{k+1}, and $^{n-1}M_1^{ZE*} = {}^{n-1}M_1^{ZE0} + \sum_{s=1}^{S_1} A_s a_{i_k}^2 + \sum_{s=1}^{S_1} B_s a_{i_k}^2 a_{i_{k+1}}^2$ is the index value, if *ZE*-isomerism correction for atoms i_k and i_{k+1} is not taken into account. We note that $\sum_{s=1}^{S_1} A_s a_{i_k}^2$ includes terms containing one or the other atom of the *Z*- or *E*-double bond but not both atoms. It follows from (**8**) that $^{n-1}M_1^{ZE}$ is continuous by ξ, and if $\xi \to 0$, then $^{n-1}M_1^{ZE} \to {}^{n-1}M_1$ for $n > 1$. It means that $^{n-1}M_1^{ZE}$ satisfy condition (1) from Section 2.1. It follows also that for two *ZE*-isomers, which differ only by the configuration of one *ZE*-double bond, the $^{n-1}M_1^{ZE}$ indices do not satisfy the condition (2r) from Section 2.1. Thus, these indices are not symmetrical relative to the corresponding values of $^{n-1}M_1^{ZE*}$, independent of the *ZE*-isomerism of the double bond in question. Nevertheless, they satisfy condition (2r′) from Section 2.1, i.e. $^{n-1}M_1^{ZE}$ are symmetrical relative to $^{n-1}M_1^{ZE*}$+f($|\xi|$), where

$$f(|\xi|) =$$
$$\xi^2[\sum_{s=1}^{S_1} A_s + \sum_{s=1}^{S_2} B_s(a_{i_k}^2 + a_{i_{k+1}}^2 + 4a_{i_k} a_{i_{k+1}})] + \xi^4 \sum_{s=1}^{S_2} B_c$$

These conclusions also hold true for *ZE*-isomers with several or all *ZE*-double bonds in opposite configurations. Indeed, in this case A_s and B_s will contain additional factors $(a_j \pm \xi)^2 = a_j^2 \pm 2a_j\xi + \xi^2$, and after multiplication, all even powers of ξ will have the same sign for both such *ZE*-isomers, while all odd powers of ξ will have opposite signs. The expression for f($|\xi|$)in this case will be more complicated than the one discussed above. It can be represented in the following general form

$$f(|\xi|) = \sum_{k=1}^{2K} R_k \xi^{2k}$$

where *K* is the total number of *Z*- and *E*-double bonds in the opposite configurations, and R_k are the polynomial functions of the conventional vertex degrees. In the case when all *ZE*-double bonds are in the opposite configurations, $^{n-1}M_1^{ZE*}$ will be identical to the conventional index $^{n-1}M_1$.

For $^{n-1}M_2^{ZE}$ indices, condition (1) from Section 2.1 is satisfied, but condition (2r) is not, i.e., for a pair of isomers with only one, several or all *ZE*-double bonds in the opposite configurations, the values of all $^{n-1}M_2^{ZE}$ indices are not symmetrical relative to the corresponding values, calculated without taking into account the *ZE*-isomerism of these bonds. At the same time, condition (2r′) is satisfied. Again, consider the $a_{i_1}a_{i_2}...a_{i_k}...a_{i_v}$ product, which includes vertex degrees of only one of the atoms i_k incident to the above-mentioned double bond, and the $a_{i_1}a_{i_2}...a_{i_k}a_{i_{k+1}}...a_{i_v}$ product, which includes vertex degrees of both atoms i_k and i_{k+1} incident to this bond. The $^{n-1}M_2$ index with $n > 1$ is the sum of all these products. In these products, $(a_{i_k} \pm \xi)$ and $(a_{i_{k+1}} \pm \xi)$ must substitute for a_{i_k} and $a_{i_{k+1}}$, respectively. Let $a_{i_1}...a_{i_{k-1}}a_{i_{k+1}}...a_{i_v} = C_s$, where $s = 1,..., S_1$, and S_1 is the total number of (n-1)-edge subgraphs containing only one vertex incident to a *Z* or *E* double bond, and $a_{i_1}...a_{i_{k-1}}a_{i_{k+2}}...a_{i_v} = D_s$, where $s = 1,..., S_2$, and S_2 is the total number of (n-1)-edge subgraphs containing both vertices incident to this bond. Then

$$a_{i_1}a_{i_2}...a_{i_k}...a_{i_v} \to C_s(a_{i_k} \pm \xi) = C_s a_{i_k} \pm C_s \xi$$

and

$$a_{i_1}a_{i_2}...a_{i_k}a_{i_{k+1}}...a_{i_v} \to D_s(a_{i_k} \pm \xi)(a_{i_{k+1}} \pm \xi) =$$
$$D_s a_{i_k} a_{i_{k+1}} \pm D_s \xi(a_{i_k} + a_{i_{k+1}}) + D_s \xi^2$$

where the arrow denotes the substitution. $C_s a_{i_k}$ and $D_s a_{i_k} a_{i_{k+1}}$ are independent of the *ZE*-isomerism correction of atoms i_k and i_{k+1}. The final formula for $^{n-1}M_2^{ZE}$ is

$$^{n-1}M_2^{ZE} = {}^{n-1}M_2^{ZE0} + \sum_{s=1}^{S_1} C_s a_{i_k} \pm \xi \sum_{s=1}^{S_1} C_s +$$

$$\sum_{s=1}^{S_2} D_s a_{i_k} a_{i_{k+1}} \pm \xi \sum_{s=1}^{S_2} D_s(a_{i_k} + a_{i_{k+1}}) + \xi^2 \sum_{s=1}^{S_2} D_s =$$

$$^{n-1}M_2^{ZE*} \pm \xi[\sum_{s=1}^{S_1} C_s + \sum_{s=1}^{S_2} D_s(a_{i_k} + a_{i_{k+1}})] + \xi^2 \sum_{s=1}^{S_2} D_s \quad (9)$$

In (9), $^{n-1}M_2^{ZE0}$ denotes the term independent of atoms i_k and i_{k+1}, and $^{n-1}M_2^{ZE*} = {}^{n-1}M_2^{chir0} + \sum_{s=1}^{S_1} B_s a_{i_k} + \sum_{s=1}^{S_2} D_s a_{i_k} a_{i_{k+1}}$ is the index value, if the *ZE*-isomerism of *Z*- or *E*-double bond under consideration is not taken into account. Again, we note that $\sum_{s=1}^{S_1} C_s a_{i_k}$ includes terms containing one or the other atom of the *Z*- or *E*-double bond but not both of these atoms. Thus, we conclude from (9) that for a pair of *ZE*-isomers with only one *ZE*-double bond in the opposite configurations, the values of all $^{n-1}M_2^{ZE}$ indices are nonsymmetrical relative to the corresponding $^{n-1}M_2^{ZE*}$ values, i.e., part of the index is independent of the *ZE*-isomerism of the *Z*- or *E*-double bond. At the same time, they are symmetrical relative to $^{n-1}M_2^{ZE*} + \xi^2 \sum_{s=1}^{S_2} D_s$. Thus, condition (2r′) is satisfied with f($|\xi|$) $= \xi^2 \sum_{s=1}^{S_2} D_s$. This conclusion also holds true for a pair of *ZE*-isomers with more than one (up to all) *ZE*-double bonds in the opposite configurations. Indeed, in this case C_s and D_s will contain additional factors $(a_j \pm \xi)$, and after multiplication, all even powers of ξ will have the same sign for both such *ZE*-isomers, while all odd powers of ξ will have opposite signs. The general expression for f($|\xi|$) can be represented as $f(|\xi|) = \sum_{k=1}^{K} R_k \xi^{2k}$, where *K* is the total number of *Z*- and *E*-double bonds in the opposite configurations, and R_k are the polynomial functions of the conventional vertex degrees.

In this case, $^{n-1}M_2^{ZE*}$ will be identical to the conventional index $^{n-1}M_2$.

3.2. Class II *ZE*-Isomerism Overall Zagreb Indices. Since real and imaginary parts of the *ZE*-isomerism descriptors have different properties, these parts will be discussed separately. To obtain class II *ZE*-isomerism descriptors, vertex degrees a_i must be substituted with $(a_i + i\xi)$ for an atom incident to *Z*-double bond and with $(a_i - i\xi)$ for an atom incident to *E*-double bond, and a_j^2 must be replaced by $(a_j \pm i\xi)^2 = a_j^2 \pm 2ia_j\xi - \xi^2$. For $^0M_1^{ZE}$ the following formula can be obtained:

$$^0M_1^{ZE} = {}^0M_1 + 2i\xi[\sum_{j=1}^{n_Z}(a_{j_1} + a_{j_2}) - \sum_{j=1}^{n_E}(a_{j_1} + a_{j_2})] - 2\xi^2(n_Z + n_E) \quad (10)$$

Thus, index $^0M_1^{ZE}$ satisfies condition (1) (cf. Section 2.1), since $^0M_1^{ZE}$ is a continuous function of ξ, and if $\xi \to 0$, then $^0M_1^{ZE} \to {}^0M_1$. From (10) it follows also that for two *ZE*-isomers with opposite configurations of all *Z*- and *E*-double bonds, the real parts of $^0M_1^{ZE}$ are equal to each other, but the imaginary parts are opposite numbers. Since $\mathrm{Re}(^0M_1^{ZE}) \neq {}^0M_1$ i.e., $^0M_1^{ZE}$ satisfies condition (2i'). In section 2.3, different subclasses of complex descriptors were defined. From the considerations above, the real part of $^0M_1^{ZE}$ of subclass IIb satisfies condition (2r'), but its imaginary part satisfies the Randić condition (see Section 2.1). The term $^0M_1^{ZE}$ of subclass IIc satisfies condition (2r'), and $^0M_1^{ZE}$ of subclass IId satisfies the Randić condition. For two isomers differing by a configuration of one or several, but not all double bonds, the real parts are also equal to each other, but the imaginary parts are symmetrical relative to the part of the index independent of the *ZE*-isomerism of this (or these) bonds. The formulas for isomers with only one *ZE*-double bond in the opposite configurations are as follows:

$$\mathrm{Re}(^0M_1^{ZE}) = \mathrm{Re}(^0M_1^{ZE*}) - 2\xi^2 \quad (11a)$$

$$\mathrm{Im}(^0M_1^{ZE}) = \mathrm{Im}(^0M_1^{ZE*}) \pm 2\xi(a_{j_1} + a_{j_2}) \quad (11b)$$

In expressions 11a and 11b the terms with the asterisks are independent of the *ZE*-isomerism correction for bonds defining the isomers. In fact, $\mathrm{Im}(^0M_1^{ZE*})$ is the sum of $^0M_1$ and all terms containing ξ^2. $\mathrm{Im}(^0M_1^{ZE*})$ is the sum of the $\pm 2\xi(a_{j_1} + a_{j_2})$ terms for all other *Z*- or *E*-double bonds. All of these terms approach zero when ξ approaches zero. Thus, in this case, $^0M_1^{ZE}$ satisfies condition (2r'). For different subclasses of descriptors based on the imaginary *ZE*-isomerism correction, the real and imaginary parts of $^0M_1^{ZE}$ of subclass IIb and $^0M_1^{ZE}$ of subclasses IIc and IId satisfy condition (2r').

Now, consider terms $a_{i_1}^2 a_{i_2}^2 ... a_{i_k}^2 ... a_{i_v}^2$ and $a_{i_1}^2 a_{i_2}^2 ... a_{i_k}^2 a_{i_{k+1}}^2 ... a_{i_v}^2$, containing vertex degrees of one (i_k) or two atoms (i_k and i_{k+1}) incident to a *Z*- or *E*-double bond, respectively. The sum of all such terms for a molecular graph is equal to $^{n-1}M_1$ with $n > 1$. In this case, $(a_{i_k} \pm i\xi)$ must substitute for a_{i_k}. If the molecule contains more than one *Z*- or *E*-double bond, these terms are complex numbers. Let $a_{i_1}^2 a_{i_2}^2 ... a_{i_{k-1}}^2$

$a_{i_{k+1}}^2 ... a_{i_v}^2 = A_s + iB_s$, where $s = 1,..., S_1$, and S_1 is the total number of (n-1)-edge subgraphs containing only one of the vertices incident to the *Z*- or *E*-double bond i_k. Let $a_{i_1}^2 a_{i_2}^2 ...$ $a_{i_{k-1}}^2 a_{i_{k+2}}^2 ... a_{i_v}^2 = C_s + iD_s$, where $s = 1,..., S_2$, and S_2 is the total number of (n-1)-edge subgraphs containing vertices i_k and i_{k+1}. Then

$$a_{i_1}^2 a_{i_2}^2 ... a_{i_k}^2 ... a_{i_v}^2 \to (A_s + iB_s)[(a_{i_k}^2 - \xi^2) \pm 2i\xi a_{i_k}] =$$
$$[A_s(a_{i_k}^2 - \xi^2) \mp 2B_s\xi a_{i_k}] + i[B_s(a_{i_k}^2 - \xi^2) \pm 2A_s\xi a_{i_k}] \quad (12a)$$

and

$$a_{i_1}^2 a_{i_2}^2 ... a_{i_k}^2 a_{i_{k+1}}^2 ... a_{i_v}^2 \to$$
$$(C_s + iD_s)[(a_{i_k}^2 - \xi^2) \pm 2i\xi a_{i_k}][(a_{i_{k+1}}^2 - \xi^2) \pm 2i\xi a_{i_{k+1}}] =$$
$$[C_s\{(a_{i_k}^2 - \xi^2)(a_{i_{k+1}}^2 - \xi^2) - 4\xi^2 a_{i_k}a_{i_{k+1}}\} \mp$$
$$2D_s\xi\{a_{i_k}(a_{i_{k+1}}^2 - \xi^2) + a_{i_{k+1}}(a_{i_k}^2 - \xi^2)\}] +$$
$$i[D_s\{(a_{i_k}^2 - \xi^2)(a_{i_{k+1}}^2 - \xi^2) - 4\xi^2 a_{i_k}a_{i_{k+1}}\} \pm$$
$$2C_s\xi\{a_{i_k}(a_{i_{k+1}}^2 - \xi^2) + a_{i_{k+1}}(a_{i_k}^2 - \xi^2)\}] \quad (12b)$$

Eventually, the following formulas can be obtained

$$\mathrm{Re}(^{n-1}M_1^{ZE}) = \mathrm{Re}(^{n-1}M_1^{ZE0}) + \sum_{s=1}^{S_1}A_s a_{i_k}^2 - \xi^2\sum_{s=1}^{S_1}A_s \mp$$
$$2\xi\sum_{s=1}^{S_1}B_s a_{i_k} + \sum_{s=1}^{S_2}C_s a_{i_k}^2 a_{i_{k+1}}^2 - \xi^2\sum_{s=1}^{S_2}C_s(a_{i_k}^2 + a_{i_{k+1}}^2 +$$
$$4a_{i_k}^2 a_{i_{k+1}}^2) + \xi^4\sum_{s=1}^{S_2}C_s \mp 2\xi\sum_{s=1}^{S_2}D_s a_{i_k}a_{i_{k+1}}(a_{i_k} + a_{i_{k+1}}) \pm$$
$$2\xi^3\sum_{s=1}^{S_2}D_s(a_{i_k} + a_{i_{k+1}}) = \mathrm{Re}(^{n-1}M_1^{ZE*}) \mp 2\xi[\sum_{s=1}^{S_1}B_s a_{i_k} +$$
$$\sum_{s=1}^{S_2}D_s a_{i_k}a_{i_{k+1}}(a_{i_k} + a_{i_{k+1}})] - \xi^2[\sum_{s=1}^{S_1}A_s + \sum_{s=1}^{S_2}C_s(a_{i_k}^2 + a_{i_{k+1}}^2 +$$
$$4a_{i_k}^2 a_{i_{k+1}}^2)] \pm 2\xi^3\sum_{s=1}^{S_2}D_s(a_{i_k} + a_{i_{k+1}}) + \xi^4\sum_{s=1}^{S_2}C_s \quad (13a)$$

$$\mathrm{Im}(^{n-1}M_1^{ZE}) = \mathrm{Im}(^{n-1}M_1^{ZE0}) + \sum_{s=1}^{S_1}B_s a_{i_k}^2 - \xi^2\sum_{s=1}^{S_1}B_s \pm$$
$$2\xi\sum_{s=1}^{S_1}A_s a_{i_k} + \sum_{s=1}^{S_2}D_s a_{i_k}^2 a_{i_{k+1}}^2 - \xi^2\sum_{s=1}^{S_2}D_s(a_{i_k}^2 + a_{i_{k+1}}^2 +$$
$$4a_{i_k}^2 a_{i_{k+1}}^2) + \xi^4\sum_{s=1}^{S_2}D_s \pm 2\xi\sum_{s=1}^{S_2}C_s a_{i_k}a_{i_{k+1}}(a_{i_k} + a_{i_{k+1}}) \mp$$
$$2\xi^3\sum_{s=1}^{S_2}C_s(a_{i_k} + a_{i_{k+1}}) = \mathrm{Im}(^{n-1}M_1^{ZE*}) \pm 2\xi[\sum_{s=1}^{S_1}A_s a_{i_k} +$$
$$\sum_{s=1}^{S_2}C_s a_{i_k}a_{i_{k+1}}(a_{i_k} + a_{i_{k+1}})] - \xi^2[\sum_{s=1}^{S_1}B_s + \sum_{s=1}^{S_2}D_s(a_{i_k}^2 + a_{i_{k+1}}^2 +$$
$$4a_{i_k}^2 a_{i_{k+1}}^2)] \mp 2\xi^3\sum_{s=1}^{S_2}C_s(a_{i_k} + a_{i_{k+1}}) + \xi^4\sum_{s=1}^{S_2}D_s \quad (13b)$$

where $\mathrm{Re}(^{n-1}M_1^{ZE0})$ and $\mathrm{Im}(^{n-1}M_1^{ZE0})$ denote the real and

imaginary parts not containing terms pertaining to atoms i_k and i_{k+1}, and $\mathrm{Re}(^{n-1}M_1^{ZE*})$ and $\mathrm{Im}(^{n-1}M_1^{ZE*})$ are independent of Z- or E-isomerism correction of atoms i_k and i_{k+1}. Here, we note that $\sum_{s=1}^{S_1}A_s a_{i_k}$, $\sum_{s=1}^{S_1}A_s a_{i_k}^2$, $\sum_{s=1}^{S_1}B_s a_{i_k}$, and $\sum_{s=1}^{S_1}B_s a_{i_k}^2$ include terms containing either one or another atom of the Z- or E-double bond but not both atoms. If the pair of isomers contains only one Z- or E-double bond, all $B_s = 0$ and $D_s = 0$, and the real parts of $^{n-1}M_1^{ZE}$ are equal, but the imaginary parts are symmetrical relative to $\mathrm{Im}(^{n-1}M_1^{ZE*})$, i.e., to the part of the index independent of the ZE-isomerism correction for the atoms incident to this bond. The stronger condition when the latter assertion is true is that each subgraph of order $(n$-1$)$ includes maximum one Z- or E-double bond or only one atom incident to a Z- or E-double bond. Consider expressions (12) and (13) in detail. If subgraphs of order $(n$-1$)$ contain atoms incident to several Z- and/or E-double bonds, terms $a_{i_1}^2 a_{i_2}^2...a_{i_k}^2...a_{i_v}^2$ and $a_{i_1}^2 a_{i_2}^2...a_{i_k}^2 a_{i_{k+1}}^2...a_{i_v}^2$ contain additional factors of the form $(a_{i_k}^2 - \xi^2) \pm 2i\xi a_{i_k}$. In this case, B_s and D_s are sums of terms containing only odd powers of ξ, starting from one, while A_s and C_s are sums of terms containing only even powers of ξ, starting from zero. It means that if $\xi \rightarrow 0$, then $\sum_{s=1}^{S_1}B_s a_{i_k}^2 \rightarrow 0$ and $\sum_{s=1}^{S_2}D_s a_{i_k}^2 a_{i_{k+1}}^2 \rightarrow 0$. It means also that if $\xi \rightarrow 0$, then $\mathrm{Im}(^{n-1}M_1^{ZE^0}) \rightarrow 0$, because it contains similar terms for atoms incident to other Z- and E-double bonds. Other terms in $\mathrm{Im}(^{n-1}M_1^{ZE})$ also approach 0 with ξ approaching zero. Thus if $\xi \rightarrow 0$, then $\mathrm{Im}(^{n-1}M_1^{ZE}) \rightarrow 0$. At the same time, if $\xi \rightarrow 0$, then $\mathrm{Re}(^{n-1}M_1^{ZE}) \rightarrow {}^{n-1}M_1$, since all other terms contain different natural powers of ξ. From (13a) it follows that for ZE-isomers having all Z- and E-double bonds in opposite configurations, $\mathrm{Re}(^{n-1}M_1^{ZE})$ values are equal to each other, which is one of the general properties of the isomerism descriptors considered in Section 2.1. This analysis shows that the complex term $^{n-1}M_1^{ZE}$ satisfies conditions (1) and (2i′) from Section 2.1. For $^{n-1}M_1^{ZE}$ of ZE-isomers with only one double bond in the opposite configurations we can write the following expressions

$$f_1(|\xi|) =$$
$$-\xi^2[\sum_{s=1}^{S_1}A_s + \sum_{s=1}^{S_2}C_s(a_{i_k}^2 + a_{i_{k+1}}^2 + 4a_{i_k}^2 a_{i_{k+1}}^2)] + \xi^4\sum_{s=1}^{S_2}C_s$$

and

$$f_2(|\xi|) =$$
$$-\xi^2[\sum_{s=1}^{S_1}B_s + \sum_{s=1}^{S_2}D_s(a_{i_k}^2 + a_{i_{k+1}}^2 + 4a_{i_k}^2 a_{i_{k+1}}^2)] + \xi^4\sum_{s=1}^{S_2}D_s$$

Thus, we conclude that the real and imaginary parts of $^{n-1}M_1^{ZE}$ of subclass IIb and $^{n-1}M_1^{ZE}$ of subclasses IIc and IId satisfy condition (2r′). The imaginary parts of $^{n-1}M_1^{ZE}$ of subclass IIb as well as of subclass IId for ZE-isomers with all double bonds in the opposite configurations satisfy the stronger Randić condition (see Section 2.1).

Consider products $a_{i_1}a_{i_2}...a_{i_k}...a_{i_v}$, which include vertex degree of only one of the atoms i_k incident to a ZE-double bond, and $a_{i_1}a_{i_2}...a_{i_k}a_{i_{k+1}}...a_{i_v}$, which include vertex degrees of both atoms i_k and i_{k+1} incident to this bond. The $^{n-1}M_2$ index with $n > 1$ is the sum of all these products. In these

products, a_{i_k} and $a_{i_{k+1}}$ must be substituted with $(a_{i_k} \pm i\xi)$ and $(a_{i_{k+1}} \pm i\xi)$, respectively. Let $a_{i_1}...a_{i_{k-1}}a_{i_{k+1}}...a_{i_v} = A_s + iB_s$, where $s = 1,..., S_1$, and S_1 is the total number of $(n$-1$)$-edge subgraphs containing only one vertex incident to Z- or E-double bond. Let also $a_{i_1}...a_{i_{k-1}}a_{i_{k+2}}...a_{i_v} = C_s + iD_s$, where $s = 1,..., S_2$, and S_2 is the total number of $(n$-1$)$-edge subgraphs containing both vertices incident to this bond. Then

$$a_{i_1}a_{i_2}...a_{i_k}...a_{i_v} \rightarrow (A_s + iB_s)(a_{i_k} \pm i\xi) =$$
$$[A_s a_{i_k} \mp B_s\xi] + i[A_s\xi \pm B_s a_{i_k}], \quad (14a)$$

and

$$a_{i_1}a_{i_2}...a_{i_k}a_{i_{k+1}}...a_{i_v} \rightarrow (C_s + iD_s)(a_{i_k} \pm i\xi)(a_{i_{k+1}} \pm i\xi) =$$
$$[C_s a_{i_k}a_{i_{k+1}} - C_s\xi^2 \mp D_s\xi(a_{i_k} + a_{i_{k+1}})] +$$
$$i[D_s a_{i_k}a_{i_{k+1}} - D_s\xi^2 \pm C_s\xi(a_{i_k} + a_{i_{k+1}})] \quad (14b)$$

where arrows denote the substitutions. The following formulas can be obtained

$$\mathrm{Re}(^{n-1}M_2^{ZE}) = \mathrm{Re}(^{n-1}M_2^{ZE^0}) + \sum_{s=1}^{S_1}(A_s a_{i_k} \mp B_s\xi) +$$
$$\sum_{s=1}^{S_2}[C_s a_{i_k}a_{i_{k+1}} - C_s\xi^2 \mp D_s\xi(a_{i_k} + a_{i_{k+1}})] =$$
$$\mathrm{Re}(^{n-1}M_2^{ZE*}) \mp \xi[\sum_{s=1}^{S_1}B_s + \sum_{s=1}^{S_2}D_s(a_{i_k} + a_{i_{k+1}})] - \xi^2\sum_{s=1}^{S_2}C_s$$
$$\quad (15a)$$

$$\mathrm{Im}(^{n-1}M_2^{ZE}) = \mathrm{Im}(^{n-1}M_2^{ZE^0}) + \sum_{s=1}^{S_1}(B_s a_{i_k} \pm A_s\xi) +$$
$$\sum_{s=1}^{S_2}[D_s a_{i_k}a_{i_{k+1}} - D_s\xi^2 \pm C_s\xi(a_{i_k} + a_{i_{k+1}})] =$$
$$\mathrm{Im}(^{n-1}M_2^{ZE*}) \pm \xi[\sum_{s=1}^{S_1}A_s + \sum_{s=1}^{S_2}C_s(a_{i_k} + a_{i_{k+1}})] - \xi^2\sum_{s=1}^{S_2}D_s$$
$$\quad (15b)$$

where $\mathrm{Re}(^{n-1}M_2^{ZE^0})$ and $\mathrm{Im}(^{n-1}M_2^{ZE^0})$ denote the parts not containing terms pertaining to atoms i_k and i_{k+1}, and $\mathrm{Re}(^{n-1}M_2^{ZE*})$ and $\mathrm{Im}(^{n-1}M_2^{ZE*})$ are the parts independent of the ZE-isomerism correction of these atoms. We must note that the terms in sums $\sum_{s=1}^{S_1}A_s a_{i_k}$ and $\sum_{s=1}^{S_1}B_s a_{i_k}$ include one or the other atom of the Z- or E-double bond but not both of them. If ZE-isomers contain only one Z- or E-double bond, all $B_s = 0$ and $D_s = 0$, and the real parts of $^{n-1}M_2^{ZE}$ are equal for both isomers. The same is true, if each subgraph of order $(n$-1$)$ includes maximum one Z- or E-double bond or only one atom incident to a Z- or E-double bond. On the other hand, under these conditions the imaginary parts of $^{n-1}M_2^{ZE}$ are symmetrical relative to $\mathrm{Im}(^{n-1}M_2^{ZE*})$, i.e., to the part of the index, independent of the ZE correction of the atoms defining the pair of ZE-isomers in question. Consider expressions (14) and (15) in detail. If subgraphs of order $(n$-1$)$ contain atoms incident to several Z- and/or E-double bonds, the $a_{i_1}a_{i_2}...a_{i_k}...a_{i_v}$ and $a_{i_1}a_{i_2}...a_{i_k}a_{i_{k+1}}...a_{i_v}$ terms contain additional factors of the form $(a_{i_k} \pm i\xi)$. As in the case of $^{n-1}M_1^{ZE}$, B_s and D_s are sums of terms containing only odd

powers of ξ, starting from one, while A_s and C_s are sums of terms containing only even powers of ξ, starting from zero. It means that if $\xi \to 0$, then $\sum_{s=1}^{S_1} B_s a_{i_k} \to 0$ and $\sum_{s=1}^{S_2} D_s a_{i_k} a_{i_{k+1}} \to 0$. If $\xi \to 0$, then $\mathrm{Im}(^{n-1}M_1^{ZE_0}) \to 0$, because it contains similar terms for atoms incident to other *Z*- and *E*-double bonds. Other terms in $\mathrm{Im}(^{n-1}M_2^{ZE})$ also approach zero with ξ approaching zero. Thus if $\xi \to 0$, then $\mathrm{Im}(^{n-1}M_2^{ZE}) \to 0$. Also, if $\xi \to 0$, then $\mathrm{Re}(^{n-1}M_2^{ZE}) \to {^{n-1}M_2}$, since all other terms contain different natural powers of ξ. From (15a) it follows that for *ZE*-isomers having all *Z*- and *E*-double bonds in opposite configurations, $\mathrm{Re}(^{n-1}M_2^{ZE})$ values are indeed equal to each other (see the general properties of isomerism descriptors in Section 2.1). This consideration shows that the complex $^{n-1}M_2^{ZE}$ term satisfies conditions (1) and (2i′) from Section 2.1. For $^{n-1}M_2^{ZE}$ of *ZE*-isomers having only one double bond in the opposite configurations we obtain $f_1(|\xi|) = -\xi^2\sum_{s=1}^{S_2}C_s$ and $f_2(|\xi|) = -\xi^2\sum_{s=1}^{S_2}D_s$. All these results show that for *ZE*-isomers with one or several double bonds in the opposite configurations, the real and imaginary parts of $^{n-1}M_2^{ZE}$ of subclass IIb and $^{n-1}M_2^{ZE}$ of IIc and IId satisfy condition (2r′). Again, the imaginary parts of $^{n-1}M_2^{ZE}$ of subclass IIb as well as those of $^{n-1}M_2^{ZE}$ of subclass IId for *ZE*-isomers with all double bonds in the opposite configurations satisfy the stronger Randić condition (see Section 2.1).

4. MOLECULAR CONNECTIVITY INDICES $^n\chi$

Molecular connectivity indices $^n\chi$ are defined as follows[27−30]

$$^0\chi = \sum_{i=1}^{N}(a_i)^{-0.5} \qquad ^1\chi = \sum_{\text{all edges}}(a_{i_1}a_{i_2})^{-0.5},...,$$

$$^{n-1}\chi = \sum_{\text{all }(n-1)\text{-edge subgraphs}}(a_{i_1}a_{i_2}...a_{i_\nu})^{-0.5} \quad (16)$$

where ν is the number of vertices in $(n-1)$-edge subgraph.

As in earlier work,[21] no separate consideration of subtypes (path, cluster, path/cluster, cycle indices) of these indices of order higher than two, is provided here. The formulas obtained below are valid for these subtypes of indices as well; the only difference is that not all but only some special subgraphs (paths, clusters, path/clusters, cycles) of the given order are included in the sums. Thus, we consider only one connectivity index of order 3, one index of order 4, one index of order 5, etc., with all subgraphs of the corresponding orders being taken into account. Furthermore, we will not consider separately valence connectivity indices $^n\chi^v$. These indices are defined by the same formulas (16) with the vertex degrees a_i equal to $(Z_i^v - h_i)/(Z_i - Z_i^v - 1)$,[27,29,30] where Z_i is the atomic number in the Mendeleev table of elements, Z_i^v is the number of valence electrons, and h_i is the number of hydrogen atoms connected to atom i.

4.1. Class I *ZE*-Isomerism Molecular Connectivity Indices. The minimum conventional degree (a_i) of a vertex incident to *Z*- or *E*-double bond is two. Thus, the *ZE*-isomerism correction for class I indices cannot be equal to ± 2 or the other small whole number with the absolute value higher than two, otherwise, some denominators in (16) can be equal to zero. In fact, to avoid having a real number index for one *Z*- or *E*-isomer and the complex number for the other isomer in a pair of *ZE*-isomers, the absolute value of *ZE*-

isomerism correction cannot exceed two (or more generally the lowest conventional vertex degree of an atom a_{i_k} incident to a *Z*- or *E*-double bond). It is a natural limit imposed on the real number *ZE*-isomerism correction. This limit refers only to one of the *Z*- or *E*-double bonds; therefore, this property is not symmetrical relative to *ZE*-isomers. We consider this result as a disadvantage. Further in this section, we assume that $|\xi| < a_{i_k}$.

Consider index $^{n-1}\chi$. It is a sum of S_1 terms $(a_{i_1}a_{i_2}...a_{i_k}...a_{i_\nu})^{-0.5}$ containing one atom i_k and S_2 terms $(a_{i_1}a_{i_2}...a_{i_k}a_{i_{k+1}}...a_{i_\nu})^{-0.5}$ containing two atoms i_k and i_{k+1} incident to a *Z*- or *E*-double bond. Let $(a_{i_1}...a_{i_{k-1}}a_{i_{k+1}}...a_{i_\nu})^{-0.5} = G_j$, $j = 1,...,S_1$ and $(a_{i_1}...a_{i_{k-1}}a_{i_k + 2}...a_{i_\nu})^{-0.5} = H_j$, $j = 1,...,S_2$. Then, for a pair of *ZE*-isomers with only one double bond in the opposite configurations

$$(a_{i_1}a_{i_2}...a_{i_k}...a_{i_\nu})^{-0.5} \to G_j(a_{i_k} \pm \xi)^{-0.5} =$$
$$G_j a_{i_k}^{-0.5}(1 \pm \xi/a_{i_k})^{-0.5} = G_j a_{i_k}^{-0.5}(1 \pm x)^{-0.5}$$

and

$$(a_{i_1}a_{i_2}...a_{i_k}a_{i_{k+1}}...a_{i_\nu})^{-0.5} \to H_j[(a_{i_k} \pm \xi)(a_{i_{k+1}} \pm \xi)]^{-0.5} =$$
$$H_j[a_{i_k}a_{i_{k+1}} \pm \xi(a_{i_k} + a_{i_{k+1}}) + \xi^2]^{-0.5} =$$
$$H_j a_{i_k}^{-0.5}a_{i_{k+1}}^{-0.5}[1 \pm (x + y) + xy]^{-0.5} =$$
$$H_j a_{i_k}^{-0.5}a_{i_{k+1}}^{-0.5}[(1 \pm x)(1 \pm y)]^{-0.5}$$

where $x = \xi/a_{i_k}$ and $y = \xi/a_{i_{k+1}}$. Note, that if $\xi > 0$, then $0 < x < 1$, and $0 < y < 1$. Also, if the *ZE*-isomerism correction of atoms i_k and i_{k+1} is not taken into account, $(a_{i_1}a_{i_2}...a_{i_k}...a_{i_\nu})^{-0.5} = G_j a_{i_k}^{-0.5}$ and $(a_{i_1}a_{i_2}...a_{i_k}a_{i_{k+1}}...a_{i_\nu})^{-0.5} = H_j a_{i_k}^{-0.5}a_{i_{k+1}}^{-0.5}$. Here, we must note that if ξ approaches zero, all $(a_{i_1}a_{i_2}...a_{i_k}...a_{i_\nu})^{-0.5}$ and $(a_{i_1}a_{i_2}...a_{i_k}a_{i_{k+1}}...a_{i_\nu})^{-0.5}$ terms approach their conventional values, thus if $\xi \to 0$, then all $^{n-1}\chi^{ZE} \to {^{n-1}\chi}$, i.e., $^{n-1}\chi^{ZE}$ satisfy condition (1). Let F be the sum of all terms, which do not contain atoms incident to the *Z*- or *E*-double bond. If $\xi > 0$, the positive deviations of $^{n-1}\chi$ from the conventional index value (when the *ZE*-isomerism correction of atoms i_k and i_{k+1} is not taken into account) for the *Z*- and *E*-isomers are

$$[F + \sum_{j=1}^{S_1}G_j a_{i_k}^{-0.5} + \sum_{j=1}^{S_2}H_j a_{i_k}^{-0.5}a_{i_{k+1}}^{-0.5}] - [F +$$
$$\sum_{j=1}^{S_1^{(1)}}G_j^{(1)}a_{i_k}^{-0.5}(1 + x)^{-0.5} + \sum_{j=1}^{S_1^{(2)}}G_j^{(2)}a_{i_{k+1}}^{-0.5}(1 + y)^{-0.5} +$$
$$\sum_{j=1}^{S_2}H_j a_{i_k}^{-0.5}a_{i_{k+1}}^{-0.5}(1 + x + y + xy)^{-0.5}] =$$
$$a_{i_k}^{-0.5}[1 - (1 + x)^{-0.5}]\sum_{j=1}^{S_1^{(1)}}G_j^{(1)} + a_{i_{k+1}}^{-0.5}[1 -$$
$$(1 + y)^{-0.5}]\sum_{j=1}^{S_1^{(2)}}G_j^{(2)} + a_{i_k}^{-0.5}a_{i_{k+1}}^{-0.5}[1 -$$
$$[(1 + x)(1 + y)]^{-0.5}]\sum_{j=1}^{S_2}H_j$$

and

$$[F + \sum_{j=1}^{S_1^{(1)}} G_j^{(1)} a_{i_k}^{-0.5}(1-x)^{-0.5} + \sum_{j=1}^{S_1^{(2)}} G_j^{(2)} a_{i_{k+1}}^{-0.5}(1-$$

$$y)^{-0.5} + \sum_{j=1}^{S_2} H_j a_{i_k}^{-0.5} a_{i_{k+1}}^{-0.5}(1-x-y+xy)^{-0.5}] -$$

$$[F + \sum_{j=1}^{S_1} G_j a_{i_k}^{-0.5} + \sum_{j=1}^{S_2} H_j a_{i_k}^{-0.5} a_{i_{k+1}}^{-0.5}] =$$

$$a_{i_k}^{-0.5}[(1-x)^{-0.5} - 1]\sum_{j=1}^{S_1^{(1)}} G_j^{(1)} + a_{i_{k+1}}^{-0.5}[(1-x)^{-0.5} - $$

$$1]\sum_{j=1}^{S_1^{(2)}} G_j^{(2)} + a_{i_k}^{-0.5} a_{i_{k+1}}^{-0.5}[[(1-x)(1-y)]^{-0.5} - 1]\sum_{j=1}^{S_2} H_j$$

respectively, since $x + y > xy$ for any $1 > x > 0$ and $1 > y > 0$. In these formulas, all terms in $\sum_{j=1}^{S_1^{(1)}} G_j^{(1)} a_{i_k}^{-0.5}(1-x)^{-0.5}$ contain only one and the same atom of the Z- or E-double bond, and those in $\sum_{j=1}^{S_1^{(1)}} G_j^{(2)} a_{i_{k+1}}^{-0.5}(1-y)^{-0.5}$ contain the other one; $S_1^{(1)} + S_1^{(2)} = S_1$, and $G_j^{(1)}$ and $G_j^{(2)}$ are the corresponding products $(a_{i_1} a_{i_2} ... a_{i_k} ... a_{i_v})^{-0.5}$. The ratio of these two expressions is the degree of symmetry of the corresponding indices for the pair of ZE-isomers relative to the value $F + \sum_{j=1}^{S_1} H_j a_{i_k}^{-0.5} + \sum_{j=1}^{S_2} H_j a_{i_k}^{-0.5} a_{i_{k+1}}^{-0.5}$, obtained when the ZE-isomerism correction for atoms i_k and i_{k+1} is not taken into account. If this ratio is one, the ZE-isomerism descriptors are symmetrical relative to this value. As we will see, it is never the case for these descriptors, except when $\xi = 0$. Let $a_{i_k}^{-0.5}\sum_{j=1}^{S_1^{(1)}} G_j^{(1)} = U_1$, $a_{i_{k+1}}^{-0.5}\sum_{j=1}^{S_1^{(2)}} G_j^{(2)} = U_2$, and $a_{i_k}^{-0.5} a_{i_{k+1}}^{-0.5}\sum_{s=1}^{S_2} H_j = V$. We will consider asymmetry Y instead of symmetry as defined above (see Section 2.3) as a function of x and y:

$$Y(x,y) = 1 - \{U_1[1 - (1+x)^{-0.5}] +$$
$$U_2[1 - (1+y)^{-0.5}] + V[1 - [(1+x)(1+y)]^{-0.5}]\}/$$
$$\{U_1[(1-x)^{-0.5} - 1] + U_2[(1-y)^{-0.5} - 1] +$$
$$V[[(1-x)(1-y)]^{-0.5} - 1]\} \quad (17a)$$

In (17a), x and y belong to interval $(0,1)$. If the degree of asymmetry Y were equal to zero, the ZE-isomerism descriptors would deviate symmetrically from the value $F + \sum_{j=1}^{S_1} H_j a_{i_k}^{-0.5} + \sum_{j=1}^{S_2} H_j a_{i_k}^{-0.5} a_{i_{k+1}}^{-0.5}$, obtained when the ZE-isomerism of the corresponding Z- or E-double bond is not taken into account. We are going to prove that this is impossible, if $\xi \neq 0$. By expansion in the Taylor series, it can be shown that $\lim_{\substack{x\to0 \\ y\to0}} Y = 0$. Indeed, for small x and

$$(1 \pm x)^{-0.5} = 1 \mp 0.5x + O(x^2)$$

$$(1 \pm y)^{-0.5} = 1 \mp 0.5y + O(y^2)$$

and

$$[(1 \pm x)(1 \pm y)]^{-0.5} =$$
$$1 \mp 0.5(x+y) + O(x^2) + O(y^2) + O(xy)$$

thus

$$\lim_{\substack{x\to0 \\ y\to0}} Y = 1 - \{U_1[1 - (1 - 0.5x + O(x^2))] +$$

$$U_2[1 - (1 - 0.5y + O(y^2))] + V[1 - (1 - 0.5(x+y) +$$
$$O(x^2) + O(y^2) + O(xy))]\}/\{U_1[(1 + 0.5x + O(x^2)) -$$
$$1] + U_2[(1 + 0.5y + O(y^2)) - 1] +$$
$$V[(1 + 0.5(x+y) + O(x^2) + O(y^2) + O(xy)) - 1]\} =$$
$$1 - 1 = 0$$

It is also obvious that $\lim_{\substack{x\to1 \text{ or} \\ y\to1}} Y = 1$, (because $\lim_{x\to1}[(1 - x)^{-0.5} - 1] = \infty$ and $\lim_{\substack{x\to1 \text{ or} \\ y\to1}} [[(1 - x)(1 - y)]^{-0.5} - 1] = \infty$).

It means that we can define the following extension for the asymmetry function: $Y(0,0) = 0$, $Y(1,y) = 1$, $Y(x,1) = 1$ and $Y(1,1) = 1$, which satisfies the conditions of continuity of extended $Y(x,y)$. To finish the proof, we must show that $Y > 0$ everywhere within the square with vertices $(0,0)$, $(1,0)$, $(0,1)$, $(1,1)$. In other words, we must show that

$$Y(x,y) = 1 - \{U_1[1 - (1+x)^{-0.5}] +$$
$$U_2[1 - (1+y)^{-0.5}] + V[1 - [(1+x)(1+y)]^{-0.5}]\}/$$
$$\{U_1[(1-x)^{-0.5} - 1] + U_2[(1-y)^{-0.5} - 1] +$$
$$V[[(1-x)(1-y)]^{-0.5} - 1]\} > 0 \quad (18)$$

The last inequality can be rewritten as follows

$$U_1[(1-x)^{-0.5} - 1] + U_2[(1-y)^{-0.5} - 1] +$$
$$V[[(1-x)(1-y)]^{-0.5} - 1] - U_1[1 - (1+x)^{-0.5}] +$$
$$U_2[1 - (1+y)^{-0.5}] + V[1 - [(1+x)(1+y)]^{-0.5}] > 0$$

since the denominator is positive.
Transforming the last expression, we obtain

$$U_1[(1-x)^{-0.5} + (1+x)^{-0.5} - 2] + U_2[(1-y)^{-0.5} +$$
$$(1+y)^{-0.5} - 2] + V[[(1-x)(1-y)]^{-0.5} +$$
$$[(1-x)(1-y)]^{-0.5} - 2] > 0 \quad (19)$$

At last, we prove that each of the three terms of this sum is non-negative. Indeed, U_1, U_2, and V are positive, and

$$(1-x)^{-0.5} + (1+x)^{-0.5} - 2 =$$

$$\frac{\sqrt{1+x} + \sqrt{1-x} - 2\sqrt{1-x^2}}{\sqrt{1-x^2}} \geq$$

$$\frac{2\sqrt[4]{1-x^2} - 2\sqrt{1-x^2}}{\sqrt{1-x^2}} \geq 0$$

since for $a > 0$ and $b > 0$, $a + b \geq 2\sqrt{ab}$, and the equality is reached only if $a = b$. In our case, equality is reached only if $x = 0$. Thus, we proved that the first term in the sum (19) is always positive. The proofs for the second and the last term are similar. Moreover, since the difference $a + b - 2\sqrt{ab}$ increases with the difference $a-b$, the terms in (19) increase when x and y increase. Since the denominator in (18) decreases, asymmetry Y increases with the growth of x or y.

NOVEL *ZE*-ISOMERISM DESCRIPTORS

J. Chem. Inf. Comput. Sci., Vol. 42, No. 4, 2002 **779**

A case when the *ZE*-isomerism correction is negative ($\xi < 0$) can be considered in a similar manner. In this case, asymmetry Y can be defined as

$$Y(x,y) = 1 - \{U_1[1 - (1 - x)^{-0.5}] +$$

$$U_2[1 - (1 - y)^{-0.5}] + V[1 - [(1 - x)(1 - y)]^{-0.5}]\}/$$

$$\{U_1[(1 + x)^{-0.5} - 1] + U_2[(1 + y)^{-0.5} - 1] +$$

$$V[[(1 + x)(1 + y)]^{-0.5} - 1]\} \quad (17b)$$

where x and y belong to the interval $(-1,0)$. Thus, we have proven that the higher the absolute value of the *ZE*-isomerism correction, the higher the degree of asymmetry of molecular connectivity index relative to its conventional value, obtained when the *ZE*-isomerism correction is not taken into account. Nevertheless, it is possible to define a function $f(|\xi|)$, such that $f(0) = 0$, which also makes the $^{n-1}\chi^{ZE}$ term for the two *ZE*-isomers symmetrical relative to $^{n-1}\chi + f(|\xi|)$ (here we follow the general approach considered in Section 2.3). Let $^{n-1}\chi^Z$ and $^{n-1}\chi^E$ be the values of $^{n-1}\chi^{ZE}$ for a pair of *ZE*-isomers. Then define

$$g(\xi) = \frac{^{n-1}\chi^Z + {}^{n-1}\chi^E}{2} - {}^{n-1}\chi$$

First, $g(0)$ is a continuous function and $g(0) = 0$. It follows immediately from the fact that $^{n-1}\chi^Z$ and $^{n-1}\chi^E$ satisfy condition (1). Second, $g(\xi)$ is an even function. It follows from the equality $^{n-1}\chi^Z(\xi) = {}^{n-1}\chi^E(-\xi)$. At last, if $^{n-1}\chi^Z > {}^{n-1}\chi^E$, $Y = 1 - [^{n-1}\chi^Z - ({}^{n-1}\chi + g(\xi))]/[(^{n-1}\chi + g(\xi)) - {}^{n-1}\chi^E] = 0$, and if $^{n-1}\chi^Z < {}^{n-1}\chi^E$, $Y = 1 - [^{n-1}\chi^E - ({}^{n-1}\chi + g(\xi))]/[(^{n-1}\chi + g(\xi)) - {}^{n-1}\chi^Z] = 0$. Thus, we obtained such a function $g(\xi)$, which satisfies all the conditions imposed on $f(|\xi|)$. Therefore, $f(|\xi|) = g(\xi)$, and our $^{n-1}\chi^{ZE}$ descriptors satisfy condition (2r′).

4.2. Class II *ZE*-Isomerism Molecular Connectivity Indices. For calculating the powers of complex numbers, their polar representation is used. Each complex number z $= x + iy$, except for zero, can be represented in the 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, and φ is the argument or the phase of z: $\cos\varphi = x/\rho$, $\sin\varphi = y/\rho$. φ is defined with the exactitude of period 2π, where $\pi = 3.14159...$ We will use the φ values within the interval $[-\pi,\pi]$. The terms in (16) contain square roots. Due to the periodicity of φ, the square root of a complex number $z = \rho e^{i\varphi}$ has two values, one equal to $\rho^{1/2}e^{i\varphi/2}$ and the other equal to $\rho^{1/2}e^{i(\varphi/2+\pi)}$, if $\varphi < 0$, or to $\rho^{1/2}e^{i(\varphi/2-\pi)}$, if $\varphi > 0$. For molecular connectivity indices we will use the first value so that the argument of $z^{1/2}$ will belong to interval $[-\pi/2, \pi/2)$.

Let i be an atom belonging to *Z*- or *E*-double bond. Then $z_i = a_i \pm i\xi$ must substitute for a_i in equations (16). The modulus of z_i is $\rho_i = (a_i^2 + \xi^2)^{1/2}$, and argument $\varphi = \pm$ arctan(ξ/a_i) since $a_i > 0$ by default. After simple transformations, the following formula can be obtained:

$$z_i^{-0.5} = \frac{1}{\sqrt[4]{a_i^2 + \xi^2}} \exp\left(\mp 0.5i\arctan\frac{\xi}{a_i}\right) =$$

$$\frac{1}{\sqrt{a_i^2 + \xi^2}}\left[\sqrt{\frac{\sqrt{a_i^2 + \xi^2} + a_i}{2}} \mp i\sqrt{\frac{\sqrt{a_i^2 + \xi^2} - a_i}{2}}\right] =$$

$$\frac{1}{\rho_i}\left[\sqrt{\frac{\rho_i + a_i}{2}} \mp i\sqrt{\frac{\rho_i - a_i}{2}}\right] \quad (20)$$

For $^0\chi^{ZE}$ the following formulas can be obtained:

$$Re(^0\chi^{ZE}) = \sum_{i=1}^{N}\frac{1}{\rho_i}\sqrt{\frac{\rho_i + a_i}{2}} \quad (21a)$$

$$Im(^0\chi^{ZE}) = -\sum_{i=1}^{2n_Z}\frac{1}{\rho_i}\sqrt{\frac{\rho_i - a_i}{2}} + \sum_{i=1}^{2n_E}\frac{1}{\rho_i}\sqrt{\frac{\rho_i - a_i}{2}} \quad (21b)$$

where N is the total number of vertices, and n_Z and n_E are the number of *Z*- and *E*-double bond, correspondingly. When an atom is not incident to a *Z*- or *E*-double bond, the corresponding term calculated with (20) or (21a) is equal to $(a_i)^{-0.5}$, since for this atom $\xi = 0$, and $\rho_i = a_i$, and its contribution to $Im(^0\chi^{ZE})$, (formula 21b) is zero. Formula 21a can be rewritten as follows

$$Re(^0\chi^{chir}) = \sum_{i=1}^{n_1}(a_i)^{-0.5} + \sum_{i=1}^{n_2}\frac{1}{\rho_i}\sqrt{\frac{\rho_i + a_i}{2}} \quad (21a')$$

where n_1 and n_2 are the number of atoms $n_1 + n_2 = N$ not belonging and belonging to a *Z*- or *E*-double bond, respectively. Let j_1 and j_2 be two atoms connected by a *Z*- or *E*-double bond. Formula 21b can then be rewritten as follows

$$Im(^0\chi^{ZE}) =$$

$$Im(^0\chi^{ZE^0}) \mp \left[\frac{1}{\rho_{j_1}}\sqrt{\frac{\rho_{j_1} - a_{j_1}}{2}} + \frac{1}{\rho_{j_2}}\sqrt{\frac{\rho_{j_2} - a_{j_2}}{2}}\right] \quad (21b')$$

where $Im(^0\chi^{ZE^0})$ is a part that includes all terms in sums (21b) except those for atoms j_1 and j_2. It follows from (21a′) and (21b′) that for a pair of *ZE*-isomers with only one double bond in the opposite configurations, the real parts of $^0\chi^{ZE}$ are equal, and the imaginary parts are symmetrical relative to the $Im(^0\chi^{ZE^0})$ value, i.e., to the part independent of the *ZE*-isomerism correction of atoms connected by this bond. Condition (2i′) is thus satisfied. Moreover, it follows from (21a) and (21b) that, for a pair of *ZE*-isomers with all *ZE*-double bonds in opposite configurations, the real parts of $^0\chi^{ZE}$ are equal to each other, and the imaginary parts are opposite numbers. As discussed in Section 2.1, this condition must always be satisfied for a pair of *ZE*-isomers with all *ZE*-double bonds in the opposite configurations. Thus, the imaginary part of the $^0\chi^{ZE}$ descriptor of subclass IIb, as well as that of the $^0\chi^{ZE}$ descriptor of subclass IId for *ZE*-isomers with all double bonds in the opposite configurations satisfies the Randić condition (see Section 2.1), whereas the real part

of the $^0\chi^{ZE}$ descriptor of subclass IId satisfies condition (2r'). On the other hand, for ZE-isomers with one or several double bonds in the opposite configurations, the real and imaginary parts of $^0\chi^{ZE}$ descriptor of subclass IIb, and $^0\chi^{ZE}$ descriptor of subclasses IIc and IId satisfy condition (2r').

In the case of molecular connectivity indices of higher orders, consider terms $a_{i_1}a_{i_2}...a_{i_k}...a_{i_v}$, which include vertex degrees of only one of the atoms i_k incident to a Z- or E-double bond, and $a_{i_1}a_{i_2}...a_{i_k}a_{i_{k+1}}...a_{i_v}$, including vertex degrees of both atoms i_k and i_{k+1} incident to this bond. Let $a_{i_1}...a_{i_{k-1}}a_{i_{k+1}}...a_{i_v} = A_s + iB_s$, where $s = 1,..., S_1$, and S_1 is the total number of $(n-1)$-edge subgraphs containing only one vertex incident to Z- or E-double bond, and $a_{i_1}...a_{i_{k-1}}a_{i_{k+2}}...a_{i_v} = C_s + iD_s$, where $s = 1,..., S_2$, and S_2 is the total number of $(n-1)$-edge subgraphs containing both vertices incident to this bond. The general formulas (14a) and (14b) can be applied here with all designations pertaining to them (see Section 3.2). In this case, terms with only one of the atoms i_k incident to a Z- or E-double bond are

$$\rho_{s,i_k} = [(A_s a_{i_k} \mp B_s \xi)^2 + (B_s a_{i_k} \pm A_s \xi)^2]^{0.5} =$$
$$[(A_s a_{i_k})^2 + (B_s \xi)^2 + (B_s a_{i_k})^2 + (A_s \xi)^2]^{0.5} =$$
$$[(A_s^2 + B_s^s)(\xi^2 + a_{i_k}^2)]^{0.5} \quad (22a)$$

$$\varphi_{s,i_k} =$$
$$\begin{cases} \arctan\dfrac{B_s a_{i_k} \pm A_s \xi}{A_s a_{i_k} \mp B_s \xi}, \text{ if } A_s a_{i_k} \mp B_s \xi > 0 \\ \arctan\dfrac{B_s a_{i_k} \pm A_s \xi}{A_s a_{i_k} \mp B_s \xi} - \pi, \text{ if } A_s a_{i_k} \mp B_s \xi < 0, B_s a_{i_k} \pm A_s \xi < 0 \\ \arctan\dfrac{B_s a_{i_k} \pm A_s \xi}{A_s a_{i_k} \mp B_s \xi} + \pi, \text{ if } A_s a_{i_k} \mp B_s \xi < 0, B_s a_{i_k} \pm A_s \xi > 0 \end{cases}$$
$$(22b)$$

After simple transformations, the following formula is obtained

$$(a_{i_1}^2 a_{i_2}^2...a_{i_k}^2...a_{i_v}^2)^{-0.5} \rightarrow$$
$$\begin{cases} \dfrac{1}{\rho_{s,i_k}}\left[\sqrt{\dfrac{\rho_{s,i_k} + |A_s a_{i_k} \mp B_s \xi|}{2}} - i\sqrt{\dfrac{\rho_{s,i_k} - |A_s a_{i_k} \mp B_s \xi|}{2}}\right], \\ \qquad\qquad\qquad \text{if } B_s a_{i_k} \pm A_s \xi > 0 \\ \dfrac{1}{\rho_{s,i_k}}\left[\sqrt{\dfrac{\rho_{s,i_k} + |A_s a_{i_k} \mp B_s \xi|}{2}} + i\sqrt{\dfrac{\rho_{s,i_k} - |A_s a_{i_k} \mp B_s \xi|}{2}}\right], \\ \qquad\qquad\qquad \text{if } B_s a_{i_k} \pm A_s \xi < 0 \end{cases}$$
$$(23)$$

where the arrow denotes the substitution made.

For terms with both atoms i_k and i_{k+1} incident to Z- or E-double bond

$$\rho_{s,i_k,i_{k+1}} = \{[C_s a_{i_k} a_{i_{k+1}} - C_s \xi^2 \mp D_s \xi(a_{i_k} + a_{i_{k+1}})]^2 + [D_s a_{i_k} a_{i_{k+1}} - D_s \xi^2 \pm C_s \xi(a_{i_k} + a_{i_{k+1}})]^2\}^{0.5} =$$
$$\{(C_s^2 + D_s^2)[\xi^4 + \xi^2(a_{i_k}^2 + a_{i_{k+1}}^2) + (a_{i_k} a_{i_{k+1}})^2]\}^{0.5} =$$
$$\{(C_s^2 + D_s^2)(\xi^2 + a_{i_k}^2)(\xi^2 + a_{i_{k+1}}^2)\}^{0.5} \quad (24a)$$

$$\varphi_{s,i_k i_{k+1}} =$$
$$\begin{cases} \arctan\dfrac{D_s a_{i_k} a_{i_{k+1}} - D_s \xi^2 \pm C_s \xi(a_{i_k} + a_{i_{k+1}})}{C_s a_{i_k} a_{i_{k+1}} - C_s \xi^2 \mp D_s \xi(a_{i_k} + a_{i_{k+1}})}, \\ \qquad \text{if } C_s a_{i_k} a_{i_{k+1}} - C_s \xi^2 \mp D_s \xi(a_{i_k} + a_{i_{k+1}}) > 0 \\ \arctan\dfrac{D_s a_{i_k} a_{i_{k+1}} - D_s \xi^2 \pm C_s \xi(a_{i_k} + a_{i_{k+1}})}{C_s a_{i_k} a_{i_{k+1}} - C_s \xi^2 \mp D_s \xi(a_{i_k} + a_{i_{k+1}})} - \pi, \\ \qquad \text{if } C_s a_{i_k} a_{i_{k+1}} - C_s \xi^2 \mp D_s \xi(a_{i_k} + a_{i_{k+1}}) < 0, \\ \qquad \text{and } D_s a_{i_k} a_{i_{k+1}} - D_s \xi^2 \pm C_s \xi(a_{i_k} + a_{i_{k+1}}) < 0 \\ \arctan\dfrac{D_s a_{i_k} a_{i_{k+1}} - D_s \xi^2 \pm C_s \xi(a_{i_k} + a_{i_{k+1}})}{C_s a_{i_k} a_{i_{k+1}} - C_s \xi^2 \mp D_s \xi(a_{i_k} + a_{i_{k+1}})} + \pi, \\ \qquad \text{if } C_s a_{i_k} a_{i_{k+1}} - C_s \xi^2 \mp D_s \xi(a_{i_k} + a_{i_{k+1}}) < 0, \\ \qquad \text{and } D_s a_{i_k} a_{i_{k+1}} - D_s \xi^2 \pm C_s \xi(a_{i_k} + a_{i_{k+1}}) > 0 \end{cases}$$
$$(24b)$$

After simple transformations, the following formula is obtained:

$$(a_{i_1}a_{i_2}...a_{i_k}a_{i_{k+1}}...a_{i_v})^{-0.5} \rightarrow$$
$$\begin{cases} \dfrac{1}{\rho_{s,i_k j_{k+1}}}\left[\sqrt{\dfrac{\rho_{s,i_k j_{k+1}} + |C_s a_{i_k} a_{i_{k+1}} - C_s \xi^2 \mp D_s \xi(a_{i_k} + a_{i_{k+1}})|}{2}} - \right. \\ \left. i\sqrt{\dfrac{\rho_{s,i_k j_{k+1}} - |C_s a_{i_k} a_{i_{k+1}} - C_s \xi^2 \mp D_s \xi(a_{i_k} + a_{i_{k+1}})|}{2}}\right], \\ \qquad \text{if } D_s a_{i_k} a_{i_{k+1}} - D_s \xi^2 \pm C_s \xi(a_{i_k} + a_{i_{k+1}}) > 0 \\ \dfrac{1}{\rho_{s,i_k j_{k+1}}}\left[\sqrt{\dfrac{\rho_{s,i_k j_{k+1}} + |C_s a_{i_k} a_{i_{k+1}} - C_s \xi^2 \mp D_s \xi(a_{i_k} + a_{i_{k+1}})|}{2}} + \right. \\ \left. i\sqrt{\dfrac{\rho_{s,i_k j_{k+1}} - |C_s a_{i_k} a_{i_{k+1}} - C_s \xi^2 \mp D_s \xi(a_{i_k} + a_{i_{k+1}})|}{2}}\right], \\ \qquad \text{if } D_s a_{i_k} a_{i_{k+1}} - D_s \xi^2 \pm C_s \xi(a_{i_k} + a_{i_{k+1}}) < 0 \quad (25) \end{cases}$$

At last, the following formulas are obtained

$$\text{Re}(^{n-1}\chi^{ZE}) = \text{Re}(^{n-1}\chi^{ZE0}) +$$
$$\sum_{s=1}^{S_1}\dfrac{1}{\rho_{s,i_k}}\sqrt{\dfrac{\rho_{s,i_k} + |A_s a_{i_k} \mp B_s \xi|}{2}} + \sum_{s=1}^{S_2}\dfrac{1}{\rho_{s,i_k i_{k+1}}} \times$$
$$\sqrt{\dfrac{\rho_{s,i_k i_{k+1}} + |C_s a_{i_k} a_{i_{k+1}} - C_s \xi^2 \mp D_s \xi(a_{i_k} + a_{i_{k+1}})|}{2}}$$
$$(26a)$$

$$\text{Im}(^{n-1}\chi^{ZE}) = \text{Im}(^{n-1}\chi^{ZE^0}) +$$

$$Z\sum_{s=0}^{S_1^{(1)}}(-)\frac{1}{\rho_{s,i_k}}\sqrt{\frac{\rho_{s,i_k}-|A_s a_{i_k}\mp B_s\xi|}{2}} +$$

$$\sum_{s=0}^{S_1^{(2)}}(+)\frac{1}{\rho_{s,i_k}}\sqrt{\frac{\rho_{s,i_k}-|A_s a_{i_k}\mp B_s\xi|}{2}} +$$

$$\sum_{s=1}^{S_2^{(1)}}(-)\frac{1}{\rho_{s,i_k,i_{k+1}}}\sqrt{\frac{\rho_{s,i_k,i_{k+1}}-|C_s a_{i_k}a_{i_{k+1}}-C_s\xi^2\mp D_s\xi(a_{i_k}+a_{i_{k+1}})|}{2}} +$$

$$\sum_{s=1}^{S_2^{(2)}}(+)\frac{1}{\rho_{s,i_k,i_{k+1}}}\sqrt{\frac{\rho_{s,i_k,i_{k+1}}+|C_s a_{i_k}a_{i_{k+1}}-C_s\xi^2\mp D_s\xi(a_{i_k}+a_{i_{k+1}})|}{2}}$$

$$(26b)$$

In (26a) and (26b) $\text{Re}(^{n-1}\chi^{ZE^0})$ and $\text{Im}(^{n-1}\chi^{ZE^0})$ are the real and the imaginary parts of the indices independent of atoms i_k and i_{k+1}, respectively. $S_1^{(1)}$ and $S_1^{(2)}$ are the number of terms for which $B_s a_{i_k} \pm A_s\xi > 0$ and $B_s a_{i_k} \pm A_s\xi < 0$, and $S_2^{(1)}$ and $S_2^{(2)}$ are the number of terms for which $D_s a_{i_k}a_{i_{k+1}} - D_s\xi^2 \pm C_s\xi(a_{i_k}+a_{i_{k+1}}) > 0$ and $D_s a_{i_k}a_{i_{k+1}} - D_s\xi^2 \pm C_s\xi(a_{i_k}+a_{i_{k+1}}) < 0$, respectively. When a pair of isomers contain only one *Z*- or *E*-double bond, all $B_s = 0$ and $D_s = 0$, and the real parts of molecular connectivity indices for both isomers are equal. Since in this case, A_s and C_s are positive, for isomers with *Z*- or *E*-double bond, only the first and third, or the second and the fourth sum in (26b) are retained. The first and the third sums for *Z*-isomer are equal to the second and the fourth sums for *E*-isomer, respectively. Therefore, the imaginary parts of $^{n-1}\chi^{ZE}$ are opposite numbers. When the pair of isomers contains more than one *Z*- or *E*-double bonds, 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 $\text{Im}(^{n-1}\chi^{ZE^0})$, which is the part of the index independent of *Z*- or *E*-isomerism correction for atoms i_k and i_{k+1}.

For *ZE*-isomers with one double bond in the opposite configurations, $\text{Re}(^{n-1}\chi^{ZE})$ and $\text{Im}(^{n-1}\chi^{ZE})$ satisfy condition (1) (see Section 2.1). Indeed, they are continuous functions of ξ, and if $\xi \to 0$, then $\text{Re}(^{n-1}\chi^{ZE}) \to {}^{n-1}\chi$ and $\text{Im}(^{n-1}\chi^{ZE}) \to 0$ (since in this case $\rho_{s,i_k} = A_s a_{i_k}$ and $\rho_{s,i_k,i_{k+1}} = C_s a_{i_k}a_{i_{k+1}}$). As in the previous cases, the imaginary parts of $^{n-1}\chi^{ZE}$ of subclass IIb as well as those of $^{n-1}\chi^{ZE}$ of subclass IId for *ZE*-isomers having all *Z*- and *E*-double bonds in the opposite configurations satisfy the Randić condition, while the real parts of $^{n-1}\chi^{ZE}$ of subclass IId satisfy condition (2r′). Moreover, since B_s and D_s are sums of terms containing odd powers of ξ, the real parts of $^{n-1}\chi^{ZE}$ for both *ZE*-isomers with all *Z*- and *E*-double bonds in opposite configurations are equal to each other [see eq 26a]. This result is in agreement with our discussion of the general properties of the isomerism descriptors in Section 2.1. In case of *ZE*-isomers with one or several double bonds in the opposite configurations, the real and imaginary parts of $^{n-1}\chi^{ZE}$ of subclass IIb and the $^{n-1}\chi^{ZE}$ descriptors of subclasses IIc and IId satisfy condition (2r′).

5. EXTENDED AND OVERALL CONNECTIVITY INDICES

First, the definitions of conventional extended and overall connectivity indices are given. Then, the properties of the corresponding *ZE*-isomerism descriptors are considered.

5.1. Extended Connectivity Indices ^nEC.[31] Initially, the sums of the vertex degrees of all vertices connected to each vertex are calculated. If vertex i is connected to j_i vertices, then the following sums are obtained

$$^n a_i = \sum_{k=1}^{j_i} {}^{n-1}a_k \qquad (27a)$$

where n is the order of extended connectivity index ^nEC and $^0 a_k = a_k$, $k = 1,...,N$ (the zero-order vertex degrees are equal to those obtained from the adjacency matrix). The extended connectivity index ^nEC is defined as the sum of $^n a_i$ values over all vertices:

$$^nEC = \sum_{i=1}^{N} {}^n a_i = \sum_{i=1}^{N}\sum_{k=1}^{j_i} {}^{n-1}a_k \qquad (27b)$$

5.2. Overall Connectivity Indices $^nTC_0$ and $^nTC_1$.[32,33] This set of indices of order n is defined as the sum of the vertex degrees a_i in all subgraphs having n edges, where $n = 0, 1, 2, ..., m$, where m is the total number of edges in the molecular graph. Vertex degrees are obtained either by using the adjacency matrix for the molecular graph (for $^nTC_0$) or the adjacency matrices for the subgraphs (for $^nTC_1$). If the total number of subgraphs of order n is j_n, and the total number of vertices of a subgraph i is k_i

$$^nTC_f = \sum_{i=1}^{j_n}\sum_{j=1}^{k_i} {}^n a_j^i \qquad (28)$$

where $^n a_j^i$ is the vertex degree of atom j in subgraph i of order n, and f is zero or one. Evidently, if the *ZE*-isomerism correction is not taken into account, $^0TC_0 = 0$, and $^1TC_0 = {}^0TC_1$, the latter being also equal to the total number of edges of the entire molecular graph multiplied by two. Two overall indices are also defined as the sums of the ^nTC_f ($f = 0,1$) over all orders 0 to m:

$$TC_f = \sum_{n=0}^{m} {}^nTC_f \qquad (29)$$

Since it is computationally difficult to obtain all subgraphs of a large graph,[41,42] m can also be an external parameter, with a value depending on the problem to be solved.

5.3. Class I *ZE*-Isomerism Extended Connectivity and Overall Connectivity Indices. Since all vertex degrees in ^nEC, ^nTC_f, and TC_f are in the first power, for two isomers with only one, or several, or all double bonds in opposite configurations, the corresponding *ZE*-isomerism descriptors are symmetrical relative to their values calculated without taking into account the *ZE*-isomerism correction of the double bond defining these isomers.

$$^nEC^{ZE} = {}^nEC^{ZE*} \pm\xi\sum_{s=1}^{S}({}^n p_{s,1} + {}^n p_{s,2}) \qquad (30a)$$

$$^nTC_f^{ZE} = {}^nTC_f^{ZE*} \pm \xi \sum_{s=1}^{S} ({}^nq_{f,s,1} + {}^nq_{f,s,2}) \quad (30b)$$

$$TC_f^{ZE} = TC_f^{ZE*} \pm \xi \sum_{s=1}^{S} (q_{f,s,1} + q_{f,s,2}) =$$

$$TC_f^* \pm \xi \sum_{s=1}^{S} \sum_{n=0}^{m} ({}^nq_{f,s,1} + {}^nq_{f,s,2}) \quad (30c)$$

where the indices denoted with asterisks are independent of the *ZE*-isomerism correction of atoms incident to the Z- or *E*-double bond defining the isomers, *S* is the number of *ZE*-double bonds in the opposite configurations, and ${}^np_{s,k}$, ${}^nq_{f,s,k}$, and $q_{f,s,k}$ ($k = 1,2$) are the numbers of times the atoms connected by the double bond make a contribution to the corresponding descriptor. These counts depend on the molecular graph. (Obviously, ${}^0q_{1,i} = 1$ and ${}^1q_{0,i} = a_i$.). Evidently, ${}^nEC^{ZE}$, ${}^nTC_f^{ZE}$, and TC_f^{ZE} satisfy both conditions (1) and (2) from Section 2.1. In fact, each of these descriptors can also be represented as a sum of the corresponding conventional descriptor and the terms for *ZE*-double bonds:

$$^nEC^{ZE} = {}^nEC \pm \xi \sum_{k=1}^{2K} {}^np_k \quad (30a')$$

$$^nTC_f^{ZE} = {}^nTC_f \pm \xi \sum_{k=1}^{2K} {}^nq_{f,k} \quad (30b')$$

$$TC_f^{ZE} = TC_f \pm \xi \sum_{k=1}^{2K} \sum_{n=0}^{m} {}^nq_{f,k}, \quad (30c')$$

where *K* is the total number of Z- and *E*-double bonds, and atoms *k* and *k*+1 are connected by a double bond.

5.4. Class II *ZE*-Isomerism Extended Connectivity and Overall Connectivity Indices. For the same reason, as in the previous subsection, for two *ZE*-isomers the real parts of the *ZE*-isomerism descriptors are equal to the corresponding conventional descriptors. The imaginary parts of the *ZE*-isomerism descriptors for two isomers having only one or several double bonds in opposite configurations are symmetrical relative to their values calculated without taking into account the *ZE*-isomerism of this double bond. Thus, one has

$$\text{Re}({}^nEC^{ZE}) = \text{Re}({}^nEC^{ZE*}) = {}^nEC \quad (31a)$$

$$\text{Re}({}^nTC_f^{ZE}) = \text{Re}({}^nTC_f^{ZE*}) = {}^nTC_f \quad (31b)$$

$$\text{Re}(TC_f^{ZE}) = \text{Re}(TC_f^{ZE*}) = TC_f \quad (31c)$$

$$\text{Im}({}^nEC^{ZE}) = \text{Im}({}^nEC^{ZE*}) \pm \xi \sum_{s=1}^{S} ({}^np_{s,1} + {}^np_{s,2}) \quad (31d)$$

$$\text{Im}({}^nTC_f^{ZE}) = \text{Im}({}^nTC_f^{ZE*}) \pm \xi \sum_{s=1}^{S} ({}^nq_{f,s,1} + {}^nq_{f,s,2}) \quad (31e)$$

$$\text{Im}(TC_f^{ZE}) = \text{Im}(TC_f^{ZE*}) \pm \xi \sum_{s=1}^{S} \sum_{n=0}^{m} ({}^nq_{f,s,1} + {}^nq_{f,s,2}) \quad (31f)$$

where indices with asterisks are independent of the *ZE*-

isomerism correction of the atoms connected by the *ZE*-double bond, *S* is the number of *ZE*-double bonds in the opposite configurations, and ${}^np_{s,k}$, and ${}^nq_{f,s,k}$ ($k = 1,2$) are the numbers of times the atoms connected by double bonds make a contribution to the corresponding descriptor. The imaginary parts of *ZE*-isomerism descriptors for two isomers having all *ZE*-double bonds in the opposite configurations are opposite numbers. They can be rewritten as follows:

$$\text{Im}({}^nEC^{ZE}) = \pm \xi \sum_{k=1}^{2K} {}^np_k \quad (31d')$$

$$\text{Im}({}^nTC_f^{ZE}) = \pm \xi \sum_{k=1}^{2K} {}^nq_{f,k} \quad (31e')$$

$$\text{Im}(TC_f^{ZE}) = \pm \xi \sum_{k=1}^{2K} \sum_{n=0}^{m} {}^nq_{f,k}, \quad (31f')$$

Both real and imaginary parts of ${}^nEC^{ZE}$, ${}^nTC_f^{ZE}$, and TC_f^{ZE} satisfy conditions (1) and (2r) from Section 2.1. Thus, the imaginary parts of ${}^nEC^{ZE}$, ${}^nTC_f^{ZE}$, and TC_f^{ZE} of subclass IIb as well as these descriptors of subclass IId for ZE-isomers with all double bonds in the opposite configurations satisfy the Randić condition (see Section 2.1). The real parts of ${}^nEC^{ZE}$, ${}^nTC_f^{ZE}$, and TC_f^{ZE} of subclass IIb and ${}^nEC^{ZE}$, ${}^nTC_f^{ZE}$, and TC_f^{ZE} of subclass IIc satisfy condition (2r). Also, for ZE-isomers having one or several (but not all) double bonds in the opposite configurations, the real and imaginary parts of ${}^nEC^{ZE}$, ${}^nTC_f^{ZE}$, and TC_f^{ZE} of subclass IIb and ${}^nEC^{ZE}$, ${}^nTC_f^{ZE}$, and TC_f^{ZE} of subclass IIc satisfy condition (2r), whereas the ${}^nEC^{ZE}$, ${}^nTC_f^{ZE}$, and TC_f^{ZE} descriptors of subclass IId satisfy condition (2r').

6. TOPOLOGICAL CHARGE INDICES

Definition of the topological charge indices is given in ref 34. In ref 23 chirality topological charge and valence topological charge indices have been introduced. Another definition of chirality topological charge indices was given in ref 21. Here we introduce *ZE*-isomerism topological charge indices in a manner similar to that implemented by Julián-Ortiz et al.[21] First, we define the conventional topological charge indices as in refs 23 and 34. Let **D** be a distance matrix. Its elements d_{ij} are defined as the number of bonds of the minimal path connecting vertices *i* and *j* ($i,j = 1,...,N$, where *N* is the total number of vertices 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, introduce matrix as **M** = **AQ** with matrix elements m_{ij}. Let $g_{ij} = m_{ij} - m_{ji}$. Topological charge indices G_k and J_k are defined as[23,34]

$$G_k = \sum_{i=1}^{N-1} \sum_{j=i}^{N} |g_{ij}| \delta_{k,d_{ij}}, \quad J_k = \frac{G_k}{N-1} \quad (32)$$

where δ is the Kronecker's delta.

Valence topological charge indices G_k^v and J_k^v are defined by the same formulas (32), but matrix **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.[23] G_k^v and J_k^v will not be considered

NOVEL *ZE*-ISOMERISM DESCRIPTORS

J. Chem. Inf. Comput. Sci., Vol. 42, No. 4, 2002 **783**

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 *ZE*-Isomerism Topological Charge Indices. In ref 23, a chirality correction for topological charge indices was considered, namely, for each carbon in R- or S-configuration one was added to or subtracted from the corresponding main diagonal element of matrix **A** or **A**′ (see above). To take *ZE*-isomerism into account, a similar approach can be used. For each atom incident to a *Z*- or *E*-double bond, we substitute the corresponding diagonal elements of matrix **A**, which are zeros, with $+\xi$ or $-\xi$, respectively, or the diagonal elements of matrix **A**′, a'_{ij}, with $(a'_{ii} \pm \xi)$. Then, g_{ij} in (32) is substituted with $g_{ij} \pm (q_{ij}^{EZ_i} - q_{ji}^{EZ_j})\xi$, where $q_{ij}^{EZ_i} = q_{ij}$, if atom i is incident to a *Z*- or *E*-double bond; otherwise $q_{ij}^{EZ_i} = 0$. Thus, if for all atoms j with $d_{ij} = k$, all $|g_{ij}| \geq |(q_{ij}^{EZ_i} - q_{ji}^{EZ_j})\xi|$, the *ZE*-isomerism topological charge index G^k values for two *ZE*-isomers, which differ by a configuration of one, several or all *ZE*-double bonds is symmetrical relative to the index value, calculated without taking *ZE*-isomerism for these bonds into account; otherwise they would not be symmetrical relative to this value. The requirement of symmetry imposes limitations on the *ZE*-isomerism correction ξ. If we introduce this limitation, the *ZE*-isomerism topological charge indices will satisfy condition (2r) from the Section 2.1. Otherwise, this condition will not always be satisfied, but the condition (2r′) will be still satisfied (which can be proven using the general approach from Section 2.3). Evidently, in both cases the *ZE*-isomerism topological charge indices satisfy condition (1) (see Section 2.1). Another way of defining *ZE*-isomerism topological charge indices consists of substituting $|g_{ij}|$ in (32) with $|g_{ij}| \pm (q_{ij}^{EZ_i} - q_{ji}^{EZ_j})\xi$. Evidently, in this case the topological charge index G^k values for two *ZE*-isomers are symmetrical relative to the value of the index when the *ZE*-isomerism of the bond(s) defining the isomers is not taken into account. In this case, topological charge indices satisfy conditions (1) and (2r).

6.2. Class II *ZE*-Isomerism Topological Charge Indices. If in (32) g_{ij} is substituted with $g_{ij} \pm i(q_{ij}^{EZ_i} - q_{ji}^{EZ_j})\xi$, the index values for both isomers will be equal, as will be the moduli of the complex number and its conjugate. Thus, this substitution would not be the right choice. A better choice is to substitute $|g_{ij}|$ in (32) with $|g_{ij}| \pm i(q_{ij}^{EZ_i} - q_{ji}^{EZ_j})\xi$. In this case, for two *ZE*-isomers, the real parts of the *ZE*-isomerism topological charge indices are equal to the corresponding conventional descriptors, i.e.

$$\mathrm{Re}(G_k^{ZE}) = G_k \quad (33a)$$

$$\mathrm{Re}(J_k^{ZE}) = J_k \quad (33b)$$

and the imaginary parts are symmetrical relative to their values obtained without taking the *ZE*-isomerism of the *Z*- or *E*-double bond defining *ZE*-isomers into account

$$\mathrm{Im}(G_k^{ZE}) = \mathrm{Im}(G_k^{ZE*}) \pm i\xi \sum_{j=1}^{N}(q_{ij}^{EZ_i} - q_{ji}^{EZ_j})\delta_{k,dij} \quad (33c)$$

$$\mathrm{Im}(J_k^{ZE}) = \frac{\mathrm{Im}(G_k^{ZE})}{N-1} \quad (33d)$$

where $\mathrm{Im}(G_k^{ZE*})$ is the part independent of the *ZE*-isomerism correction of atoms incident to the *Z*- or *E*-double bond. In fact,

$$\mathrm{Im}(G_k^{ZE}) = \pm \frac{\xi}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}(q_{ij}^{EZ_i} - q_{ji}^{EZ_j})\delta_{k,dij}. \quad (33c')$$

Thus, G_k^{ZE} satisfies conditions (1) and (2i) from Section 2.1. For a pair of isomers with all *ZE*-double bonds in opposite configurations, the imaginary parts of G_k^{ZE} of subclass IIb and G_k^{ZE} of subclass IId satisfy the Randić condition, while G_k^{ZE} of subclass IIc satisfy condition (2r). For a pair of isomers with only one or several (but not all) *ZE*-double bonds in opposite configurations, G_k^{ZE} of subclass IIb and IIc satisfy condition (2r), and G_k^{ZE} of subclass IId satisfy condition (2r′).

7. CASE STUDY: QSAR ANALYSIS OF 131 POTENTIAL ANTICANCER AGENTS INHIBITING TUBULIN POLYMERIZATION

A goal of these calculations was to demonstrate that *ZE*-isomerism topological descriptors introduced in this paper could be useful in practical QSAR studies. A series of 131 stilbene and dihydrostilbene derivatives,[38,39] potential anti-cancer agents inhibiting tubulin polymerization was selected as a representative example to test the performance of novel *ZE*-isomerism descriptors. This data set included 47 *Z*- and 37 *E*-stilbene derivatives and 31 dihydrostilbene derivatives as well as 16 additional compounds with diverse structures with known cytotoxicity with respect to different cancer cell lines[38,39] (see Tables I−IV in ref 38 and Tables 1−6 in ref 39). It included 24 pairs of *ZE*-isomers (see Tables 1 and 2 in ref 38 and Tables 1 and 2 in ref 39). Currently, one of the most potent analogues, combretastatin A-4 is in clinical trials.[43] For QSAR studies we used toxicity ED$_{50}$ (which was expressed in logarithmic units for our calculations) measured for A-549 cancer cell lines.

ZE-izomerism overall Zagreb indices,[26] molecular connectivity indices,[27−30] and extended[31] and overall connectivity indices[32,33] up to order seven were calculated with the *ZE*-isomerism correction 2.0. A series of nonchiral descriptors calculated by MolconnZ[27] have been added to *ZE*-isomerism descriptors. They included simple and valence path, cluster, path/cluster, and chain molecular connectivity indices,[27−30] kappa molecular shape indices,[44,45] topological[46] and electrotopological state indices,[47−50] differential connectivity indices,[51] graph's radius and diameter,[52] Wiener[1] and Platt[53] indices, Shannon[54] and Bonchev-Trinajstić[55] information indices, counts of different vertices,[6] and counts of paths and edges between different kinds of vertices.[6] The *k*-nearest neighbors QSAR (kNN-QSAR) method[18,19] developed in this laboratory[21] was employed. KNN-QSAR is a variable selection approach, which uses leave-one-out cross-validation procedure and evolution algorithm for optimal descriptor selection.

Recently, we suggested that a combination of robust statistical criteria should be used to estimate a predictive power of a QSAR model as follows:[56] (i) leave-one-out cross-validation correlation coefficient q^2; (ii) correlation coefficient R between the predicted and observed activities; (iii) coefficients of determination[57] (predicted versus observed

Table 2. Properties of ZE-Isomerism Topological Descriptors[e]

	ZE-isomers[a]	$^0M_1^{ZE}$	$^{n-1}M_1^{ZE}$ (n>1)	$^{n-1}M_2^{ZE}$ (n>1)	$^0\chi^{ZE}$	$^{n-1}\chi^{ZE}$ (n>1)	$^nE^{ZE}$	$^nTC_{0,1}^{ZE}$	G_k^{ZE}
class I	one or several	(2r′)	(2r′)	(2r′)	(2r′)	(2r′)	(2r)	(2r)	(2r)
	all	(2r′)	(2r′)	(2r′)	(2r′)	(2r′)	(2r)	(2r)	(2r)
subclass IIb	Re[b] one or several	(2r′)	(2r′)	(2′)	(2′)	(2′)	(2)=$^nE^4$	(2)=$^nTC_{0,1}{}^d$	(2)=$G_k{}^d$
	Im[b] one or several	(2r′)	(2r′)	(2′)	(2′)	(2′)	(2)	(2)	(2)
	Re[b] all	(2r′)=[c]	(2r′)=[c]	(2r′)=[c]	(2r′)=[c]	(2r′)=[c]	(2r)=$^nE^d$	(2r)=$^nTC_{0,1}{}^d$	(2r)=$G_k{}^d$
	Im[b] all	(R)	(R)	(R)	(R)	(R)	(R)	(R)	(R)
subclass IIc	one or several	(2r′)	(2r′)	(2r′)	(2r′)	(2r′)	(2r)	(2r)	(2r)
	all	(2r′)	(2r′)	(2r′)	(2r′)	(2r′)	(2r)	(2r)	(2r)
subclass IId	one or several	(2r′)	(2r′)	(2r′)	(2r′)	(2r′)	(2r′)	(2r′)	(2r′)
	all	(R)	(R)	(R)	(R)	(R)	(R)	(R)	(R)

[a] One or several means isomers with one or several Z- and/or E-double bonds in opposite configurations. All means all Z- and E-double bonds in opposite configurations. [b] Re − real part, Im − imaginary part of a complex descriptor. [c] Sign = means that the values of descriptors for both isomers are equal to each other. [d] Descriptor designation after the "=" sign means that the descriptor is equal to the corresponding conventional descriptor. [e] See the meaning of designations (1), (2r), (2r′), and (R) in the text. All descriptors satisfy condition (1).

activities R_0^2 and observed versus predicted activities $R_0'^2$); and (iv) slopes k and k' of regression lines through the origin. Specifically, we consider a QSAR model sufficiently predictive, if the following conditions are satisfied:[56]

$$\text{(i)} \qquad q^2 > 0.5 \qquad\qquad (34)$$

$$\text{(ii)} \qquad R^2 > 0.6 \qquad\qquad (35)$$

$$\text{(iii)} \qquad \frac{(R^2 - R_0^2)}{R^2} < 0.1 \text{ or } \frac{(R^2 - R_0'^2)}{R^2} < 0.1 \qquad (36)$$

$$\text{(iv)} \qquad 0.85 \leq k \leq 1.15 \text{ or } 0.85 \leq k' \leq 1.15 \qquad (37)$$

For the division of a data set into the training and test sets, the sphere-exclusion algorithm[58] was used. Dissimilarity level was set equal to 0.8, 1.0, and 1.2. Due to the stochastic nature of this method, as many as 160 models have been generated for each value of the dissimilarity level. Data set division corresponding to the dissimilarity level of 0.8 yielded the training and test sets containing 120 and 11 compounds, respectively. In total, for these training and test sets, as many as 50 QSAR models satisfied conditions (34)−(37). The best model (Model 1) was characterized by the following statistics: $q^2 = 0.69$, $R^2 = 0.85$, $k = 1.01$, $R_0^2 = 0.83$, $F = 52.6$. This model satisfies conditions (34)−(37). Predictive ability of this model is demonstrated in Figure 1a.

Data set division corresponding to the dissimilarity level of 1.0 produced training and test sets containing 110 and 21 compounds, respectively. The best model was characterized by the following statistics: $q^2 = 0.56$, $R^2 = 0.70$, $k' = 0.91$, $R_0'^2 = 0.60$, $F = 45.1$. This model does not completely satisfy conditions (34)−(37), since $(R^2 - R_0'^2)/R^2 = 0.14$. Finally, the data set division corresponding to the dissimilarity level of 1.2 produced the training and test sets containing 100 and 31 compounds, respectively. In this case, three models satisfied the conditions (34)−(37). The best model (Model 2) had the following statistics: $q^2 = 0.61$, $R^2 = 0.78$, $k = 0.89$, $R_0^2 = 0.77$, $F = 103.8$. This model does satisfy conditions (34)−(37). Predictive ability of this model is demonstrated in Figure 1b.

The power of models 1 and 2 can be demonstrated by their ability to accurately predict relative toxicity of both Z- and E-isomers of a stilbene derivative as well as a corresponding dihydrostilbene analogue (Figure 2). All three compounds were included in the test sets for both models 1
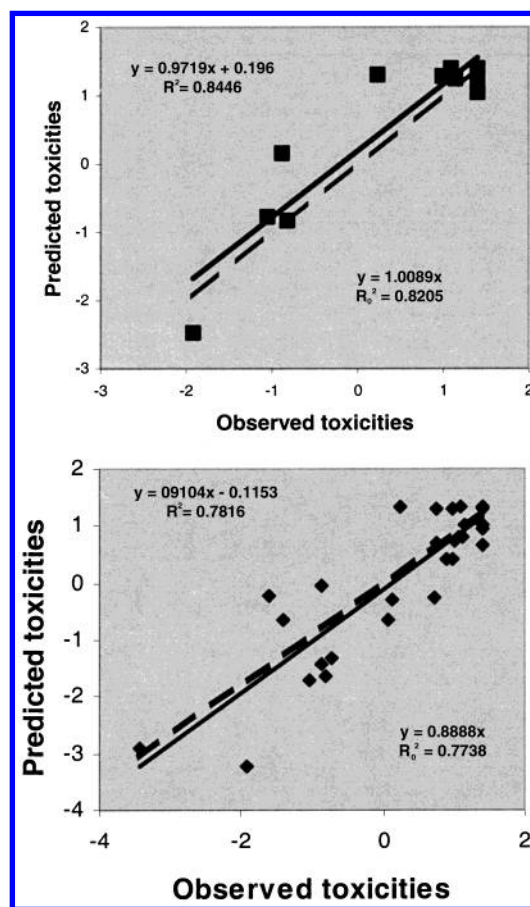

Figure 1. a. Model 1. Observed vs predicted toxicities for 11 stilbene and dihydrostilbene derivatives tested against A-549 cancer cell lines and included in the test set. Training set included 120 compounds. Solid line is the standard linear regression, and dashed line is the regression through the origin. b. Model 2. Observed vs predicted toxicities for 31 compounds tested against A-549 cancer cell lines. Training set included 100 compounds. Solid line is the standard linear regression, and dashed line is the regression through the origin.

and 2. Our results demonstrated that both models 1 and 2 correctly predicted the orders of toxicities of both this pair of Z- and E-isomers and their corresponding dihydrostilbene derivative lacking the respective double bond (Figure 2). The test set used to evaluate Model 2 included three additional pairs of Z- and E-isomers as compared to the test set used to eavluate Model 1. The order of observed and predicted toxicities was the same for two of these pairs. Both

Novel ZE-Isomerism Descriptors

J. Chem. Inf. Comput. Sci., Vol. 42, No. 4, 2002 **785**
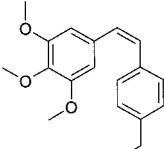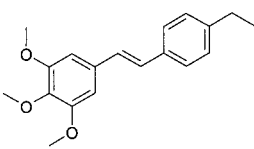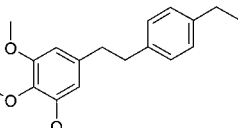

Figure 2. Predicted and observed toxicities for a pair of *cis-* and *trans*-stilbene derivatives and the corresponding dihydrostilbene derivative.

compounds in the third pair had relatively low toxicities, and the predicted values were almost the same.

Calculations were also performed with randomized activities of compounds included in the training sets. All resulting models had q^2 values not exceeding 0.4 and predictive R^2 values not exceeding 0.4 as well, which further proved the robustness and predictive ability of our original models.

In summary, we have shown that *ZE*-isomerism descriptors can be successfully employed in QSAR analyses. Earlier[21,35] we established that for obtaining the best QSAR model using chirality descriptors, extensive calculations with different chirality correction values and different subclasses of descriptors must be used. However, the goal of this example was to obtain just a few statistically robust and predictive QSAR models to prove the practical utility of our *ZE*-isomerism descriptors to QSAR studies.

8. CONCLUSIONS

In this paper, we have introduced several novel *ZE*-isomerism descriptors and analyzed some of their properties. Our descriptors are based on a quantity named *ZE*-isomerism correction, which can be a real or imaginary number. *ZE*-isomerism correction was added to or subtracted from the vertex degrees of atoms connected by *Z*- or *E*-double bonds, respectively. These modified vertex degrees were used in standard formulas of different series of conventional topological descriptors such as Zagreb[26,40] and molecular connectivity indices,[27−30] extended[31] and overall[32,33] connectivity indices, and topological charge indices.[23,34] Since it is currently impossible to use complex descriptors in QSAR analysis, we defined three subclasses of real *ZE*-isomerism descriptors, which use real and imaginary parts of complex descriptors (see Table 1). We note that the same subclasses of descriptors were defined for chirality descriptors introduced earlier.[21]

We have shown that all of our topological descriptors of *ZE*-isomerism satisfy the following conditions.

(1) *ZE*-isomerism descriptors are continuous functions of the *ZE*-isomerism correction and their values coincide with those for respective conventional descriptors, if the *ZE*-isomerism correction is zero.

(2r′) The values of real *ZE*-isomerism descriptors for any pair of *ZE*-isomers are symmetrical relative to the sum of the corresponding descriptor values, which do not account for the underlying *ZE*-double bonds, and a continuous even function of the *ZE*-isomerism correction. This function approaches zero when the *ZE*-isomerism correction approaches zero.

(2i′) The values of a complex *ZE*-isomerism descriptor satisfy the following conditions. For isomers with one or several (but not necessarily all) *ZE*-double bonds in the opposite configurations, the real parts of a *ZE*-isomerism descriptor are symmetrical relative to the sum of the corresponding conventional descriptor and a continuous even function of the *ZE*-isomerism correction. This function approaches zero when the *ZE*-isomerism correction approaches zero. The imaginary parts of the descriptor are symmetrical relative to another continuous even function of the *ZE*-isomerism correction, which also approaches zero when the *ZE*-isomerism correction approaches zero. The descriptor values for *ZE* isomers with all *ZE*-double bonds in the opposite configurations are complex conjugates.

Some of our descriptors satisfy conditions that are stronger than conditions (2r′) or (2i′), as follows:

(2r) The values of some real *ZE*-isomerism descriptor for any pair of *ZE*-isomers are symmetrical relative to the corresponding descriptor value calculated without taking into account *ZE*-double bonds, which define these isomers.

(2i) The values of some complex *ZE*-isomerism descriptors for isomers having one or several (but not necessarily all) *ZE*-double bonds in the opposite configurations are symmetrical relative to the value of the corresponding descriptor value, calculated without taking into account *ZE*-double bonds defining these isomers. For isomers with all *ZE*-double bonds in the opposite configurations these values are complex conjugates with real parts equal to the corresponding conventional descriptor.

The strongest condition some of *ZE*-isomerism descriptors satisfy is the Randić condition (see Section 2.1).

(R) The values of some *ZE*-isomerism descriptors for a pair of *ZE*-isomers having all *Z*- and *E*-double bonds in opposite configurations are opposite numbers. For compounds without *Z*- and *E*-double bonds these values are zeros.

The properties of the *ZE*-isomerism descriptors are summarized in Table 2. We note that our chirality descriptors introduced recently[21] satisfy the same conditions. The only difference is that the values of $^{n-1}M_2{}^{chir}$ of class I satisfy the stronger condition (2r) instead of (2r′).

In general, additional isomerism descriptors can be introduced in the same manner, for instance cis- and trans-isomers of cyclic, polycyclic, and heterocyclic compounds. On the other hand, it is possible to define mixed chirality-*ZE*-isomerism descriptors. If they are defined using the same methodology, as in this paper and in ref 21, they will have the same properties as those presented in Table 2. In general, this approach is not limited to σ- or π-diastereomers and can be extended to polysubstituted cyclic, polycyclic, and heterocyclic compounds.

Using a series of 131 stilbene and dihydrostilbene derivatives with anticancer activity, we have demonstrated that *ZE* isomerism descriptors can be successfully applied in QSAR

studies. We have developed several predictive *k*NN variable selection QSAR models for these agents. We suggest that *ZE*-isomerism descriptors can be used along with conventional and chirality descriptors in QSAR analysis, database mining, and virtual combinatorial library design. A program to calculate both the chirality descriptors[21] and *ZE* isomerism descriptors is available from authors upon request.

ACKNOWLEDGMENT

REFERENCES AND NOTES

(1) Wiener, H. Structural determination of paraffin boiling points. *J. Am. Chem. Soc.* **1947**, *69*, 17.
(2) Randić, M. On characterization of chemical structure. *J. Chem. Inf. Comput. Sci.* **1997**, *37*, 672−687.
(3) Todeschini, R.; Consonni, V. The handbook of molecular descriptors. In *The Series of Methods and Principles in Medicinal Chemistry*; Mannhold, R., Kubinyi, H., Timmerman, H., Eds.; Wiley-VCH: New York, 2000; Vol. 11, p 680.
(4) Kier, L. B.; Hall, H. Molecular structure description: the electrotopological state; Academic Press: 1999.
(5) Karelson, M. *Molecular descriptors in QSAR/QSPR*; Jossey-Bass: A Wiley company, 2000.
(6) http://www.eslc.vabiotech.com/molconn/.
(7) Katritzky, A. R.; Lobanov, V.; Karelson, M. *CODESSA (Comprehensive Descriptors for Structures and Statistical Analysis)*; University of Florida: Gainesville, FL, 1994.
(8) *Topological indices and related descriptors in QSAR and QSPR*; Devillers, J., Balaban, A. T., Eds.; Gordon & Breech: Amsterdam, The Netherlands, 1999.
(9) Zheng, W.; Cho, S. J.; Tropsha, A. Rational combinatorial library design. 1. FOCUS-2D: A new approach to the design of targeted combinatorial chemical libraries. *J. Chem. Inf. Comput. Sci.* **1998**, *38*, 251−258.
(10) Zheng, W.; Cho, S. J.; Waller, C. L. Tropsha, A. Rational combinatorial library design. 3. Simulated annealing guided evaluation (SAGE) of molecular diversity: A novel computational tool for universal library design and database mining, *J. Chem. Inf. Comput. Sci.* **1999**, *39*, 738−746.
(11) Tropsha, A.; Zheng, W. Identification of the descriptor pharmacophores using variable selection QSAR: applications to database mining. *Curr. Pharm. Des.* **2001**, *7*, 599−612.
(12) Cramer III, R. D.; 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.
(13) Marshall, G. R.; Cramer III, R. D. Three-dimensional structure−activity relationships. *Trends Pharmacol. Sci.* **1988**, *9,* 285−289.
(14) Cho, S. J.; Tropsha, A. Cross-validated R2-guided region selection for comparative molecular field analysis (CoMFA): A simple method to achieve consistent results, *J. Med. Chem.* **1995**, *38*, 1060−1066.
(15) 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.
(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) Brusniak, M. Y.; Pearlman, 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.
(18) 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.
(19) 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.
(20) Bures, M. G.; Martin, Y. C. Computational methods in molecular diversity and combinatorial chemistry. *Curr. Opin. Chem. Biol.* **1998**, *2*, 376−80.
(21) Golbraikh, A.; Bonchev, D.; Tropsha, A. Novel chirality descriptors derived from molecular topology. *J. Chem. Inf. Comput. Sci.* **2001**, *41*, 147−158.

(22) 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.
(23) 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 Mod.* **1998**, *16*, 14−18.
(24) Randić, M.; Mezey, P. G. Palindromic perimeter codes and chirality properties of polyhexes. *J. Chem. Inf. Comput. Sci.* **1996**, *36*, 1183−1186.
(25) Randić, M. Graph theoretical descriptors of two-dimensional chirality with possible extension to three-dimensional chirality. *J. Chem. Inf. Comput. Sci.* **2001**, *41*, 639−649.
(26) Gutman I.; Ruscić, B.; Trinajstić, N.; Wilcox, C. F. Jr. Graph theory and molecular orbitals. XII. Acyclic polyenes. *J. Chem. Phys.* **1975**, *62*, 3399.
(27) MolconnZ. http://www.eslc.vabiotech.com/molconn/manuals/.
(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: Amsterdam, The Netherlands, 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.; Salabert, M.; Soler, R. Charge indices: new topological descriptors. *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 520−525.
(35) Golbraikh, A.; Tropsha, A. QSAR modeling using chirality descriptors derived from molecular topology. *J. Chem. Inf. Comput. Sci.* Submitted for publication.
(36) Lekishvili, G. On the characterization of molecular stereostructure: 1. Cis−Trans isomerism. *J. Chem. Inf. Comput. Sci.* **1997**, *37*, 924−928.
(37) Lekishvili, G. On the characterization of molecular stereostructure: 2. The invariants of the two-dimensional graphs. *Match* **2001**, *43*, 135−152.
(38) Cushman, M.; Nagarathnam, D.; Gopal, D.; Chakraborti, A. K.; Lin, C. M.; Hamel, E. Synthesis and evaluation of stilbene and dihydrostilbene derivatives as potential anticancer agents that inhibit tubulin polymerization. *J. Med. Chem.* **1991**, *34*, 2579−2588.
(39) Cushman, M.; Nagarathnam, D.; Gopal, D.; He, H.-M.; Lin, C. M.; Hamel, E. Synthesis and evaluation of analogues of (Z)-1-(4-Methoxyphenyl)-2-(3,4,5-trimethoxyphenyl)ethene as potential cytotoxic and antimitotic agents. *J. Med. Chem.* **1992**, *35*, 2293−2306.
(40) Bonchev, D.; Trinajstić, N. Overall molecular descriptors. 3. Overall Zagreb indices. *SAR QSAR Environ. Res.* **2001**, *12*, 213−235.
(41) Bonchev, D. Overall connectivities /topological complexities: A new powerful tool for QSPR/QSAR, *J. Chem. Inf. Comput. Sci.* **2000**, *40*, 934−941.
(42) Bonchev, D. Overall connectivity − A next generation molecular connectivity, *J. Mol. Graphics Model.* **2001**, *20*, 65−75.
(43) http://www.cpet.ufl.edu/sciexpl/zo14.htm.
(44) Kier, L. B. A shape index from molecular graphs. *Quant. Struct.-Act. Relat.* **1985**, *4*, 109−116.
(45) Kier, L. B. Inclusion of symmetry as a shape attribute in kappa-index analysis. *Quant. Struct-Act. Relat.* **1987**, *6*, 8−12.
(46) 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.
(47) 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.
(48) 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.
(49) Kier, L. B.; Hall, L. H. *Molecular structure description: The electrotopological state*; Academic Press: 1999.
(50) Kellogg, G. E.; Kier, L. B.; Gaillard, P.; Hall, L. H. The E-State Fields. Applications to 3D QSAR. *J. Comput. Aid. Mol. Des.* **1996**, *10*, 513−520.
(51) Kier, L. B.; Hall, L. H. A Differential molecular connectivity index. *Quant. Struct.-Act. Relat.* **1991**, *10*, 134−140.
(52) 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.

NOVEL *ZE*-ISOMERISM DESCRIPTORS

J. Chem. Inf. Comput. Sci., Vol. 42, No. 4, 2002 **787**

(53) Platt, J. R. Prediction of isomeric differences in paraffin properties. *J. Phys. Chem.* **1952**, *56*, 328.

(54) Shannon, C.; Weaver, W. Mathematical theory of communication, University of Illinois, Urbana, IL, 1949.

(55) Bonchev, D.; Mekenyan, O.; Trinajstic, N. Isomer discrimination by topological information approach. *J. Comput. Chem.* **1981**, *2*, 127−148.

(56) Golbraikh, A.; Tropsha, A. Beware of q2! *J. Mol. Graphics Mod.* **2002**, *20*, 269−276.

(57) Sachs, L. *Applied Statistics. A Handbook of Techniques*; Springer-Verlag: 1984.

(58) Golbraikh, A.; Tropsha, A. Rational compound selection for training and test sets in QSAR. In *Advances in molecular diversity and combinatorial chemistry*; Waller, C. L., Ed.; Kluwer: 2002; submitted for publication.