

Complexity and chirality indices for molecular informatics: differential geometry approach

*A.V.Luzanov, D.Nerukh**

STS "Institute for Single Crystals", Institute for Scintillation Materials,
National Academy of Sciences of Ukraine,
60 Lenin Ave., 61001 Kharkiv, Ukraine

*Unilever Centre for Molecular Informatics, Department of Chemistry,
Cambridge University, Lensfield Road, Cambridge, CB2 1EW, UK

Received February 1, 2005

Novel molecular complexity measures are designed based on the quantum molecular kinematics. The Hamiltonian matrix constructed in a quasi-topological approximation describes the temporal evolution of the modelled electronic system and determines the time derivatives for the dynamic quantities. This allows to define the average quantum kinematic characteristics closely related to the curvatures of the electron paths, particularly, the torsion reflecting the chirality of the dynamic system. A special attention has been given to the computational scheme for this chirality measure. The calculations on realistic molecular systems demonstrate reasonable behaviour of the proposed molecular complexity indices.

Предложены новые индексы сложности молекул, основанные на идеях молекулярной квантовой кинематики. Матрица гамильтониана, построенная в некотором квазитопологическом приближении, определяет временную эволюцию модельной электронной системы и соответствующие производные динамических величин. Подход позволяет задать средние квантовые кинематические характеристики, тесно связанные с кривизнами электронных траекторий, в частности, с кручением, отражающим хиральность динамической системы. Особое внимание уделяется разработке алгоритма вычислений индекса молекулярной хиральности. Расчеты конкретных систем демонстрируют разумное поведение предложенных индексов молекулярной сложности.

Introduction

It is commonly accepted that a consistent approach for treating molecular structures should rest on the quantum-mechanical foundation. In reality, however, molecules function under complicated conditions caused by the environment that has its own complex structure (even if temperature effects, *etc.* are neglected). Molecular systems, especially the ones involved in bio-processes, are tremendously large and entangled. Even a simplified description of them is a very difficult problem and the use of crude models of the chemical structure is inevitable in their applications to the mo-

lecular informatics problems related to the new materials design. Among these models the geometrical and topological approaches are very popular and practically fruitful [1–8] (see also excellent reviews on developments of the molecular chirality problems in [9–13]). The fundamental results reported in [12, 13] are worth noting. There the existence of the chirality is proved for the molecular graphs in the form of rigorous mathematical theorems (the topological chirality theory). Nevertheless, pure topological models give too simplified quantitative descriptions that may result in the so-called isomer degeneracy of the numerical characteristics calculated [1].

It is interesting to note that the first topological description appeared as a result of an extension of the π -electron Hückel MO method to general molecular structures. Since then the internal connection of the topological approach to the quantum-chemical foundation has been lost and numerous new discrimination indices emerge (see, for example [4]).

However, more sophisticated approaches for the large-scale molecular systems are required. In molecular informatics the new schemes of such intermediate level of treatment are worth to be developed in order to include subtle structural chemical effects. The same is true for the so-called molecular complexity problem, which is inherently even more difficult.

Our (as many other researchers') interest in quantifying chemical complexity was stimulated by the works of Bertz [14, 15] (see also his recent paper [16]). Even though there were many discussions of the complexity measures in cybernetics [17], mathematical biology [18], and dynamic chaos theory [19] before, Bertz was the first who considered specifically the complexity of the chemical structures and chemical reactions. He correctly used the conventional Shannon entropy for the relevant molecular graphs in his study. However, such an approach oversimplifies the whole problem as all topological approaches do. For example, the conventional approach does not take into account the molecular chirality and there were attempts to eliminate this drawback using the graph theory methods [4, 20].

Correct quantification of the molecular chirality is too difficult problem that cannot be solved within the schemes that ignore the conformational (real geometry) and electronic levels of treatment. That is why we started *ab ovo* treating the problem at the electron level of description from the start.

Quantum mechanics permits an intuitive visualisation within the so-called Feynman formulation dealing with a set of electronic paths (trajectories). This is also a feature of the Bohmian approach to quantum mechanics (see [21] for a numerical implementation and an illustrative example). Thus, with our problem in mind we should first describe the electron path complexity. For this we have to choose the characteristics that can be considered as usual dynamic quantities because the latter can be easily converted into the quantum-mechanical operators. As we have shown [22] the differential geometry description of the curves,

i.e. the electron kinematics, can be efficiently used for this purpose.

In [22, 23] the Hückel-like Hamiltonian was used for generating the electron dynamics and in practice it did not give very reliable description. In this paper we improve our computational scheme using the refined quantum dynamics and introducing new molecular complexity measures. It allows us to present a more general treatment of the geometrical and electronic complexity of the molecular systems. Computations using specific structural classes of molecules demonstrate a reasonable behaviour of the complexity measures proposed in this study.

1. Differential-geometry description of the electron movement

Main differential invariants of a space curve. Any spatial curve can be locally characterized by its arc length s (the so-called natural parameter of the curve) and two differential quantities termed the curvature (the first curvature) κ_1 and the torsion (the second curvature) κ_2 . Let us introduce the moving trihedron $\{\boldsymbol{\tau}, \mathbf{v}, \boldsymbol{\beta}\}$ as a local basis set at a given point $\mathbf{r}=\mathbf{r}(s)$ (see [24]). Here

$$\boldsymbol{\tau} = \mathbf{r}' \quad (1)$$

is a unit tangent vector at point \mathbf{r} , and the unit vector \mathbf{v} is the corresponding normal vector defined as

$$\mathbf{v} = \boldsymbol{\tau}' / |\boldsymbol{\tau}'| \equiv \mathbf{r}'' / |\mathbf{r}''|.$$

With this the binormal vector is determined by

$$\boldsymbol{\beta} = \boldsymbol{\tau} \wedge \mathbf{v},$$

where \wedge symbolizes the vector (cross) product. It is shown in differential geometry that the moving trihedron satisfies the Frenet differential equations:

$$\begin{aligned} \boldsymbol{\tau}' &= \kappa_1 \mathbf{v}, \\ \mathbf{v}' &= -\kappa_1 \boldsymbol{\tau} + \kappa_2 \boldsymbol{\beta}, \\ \boldsymbol{\beta}' &= -\kappa_2 \mathbf{v}. \end{aligned}$$

It is more convenient to assign an arbitrary parameter t to a curve $\mathbf{r} \equiv \mathbf{r}(t)$ that determines the curve. In the mechanical (kinematical) interpretation of the curve the parameter t can be considered as usual time. Then the explicit expressions for the curvatures are of the form

$$\kappa_1 = |\dot{\mathbf{r}} \wedge \ddot{\mathbf{r}}| / |\dot{\mathbf{r}}|^3, \quad (2)$$

$$\kappa_2 = (\dot{\mathbf{r}} \wedge \ddot{\mathbf{r}}) \cdot \dot{\mathbf{r}} / |\dot{\mathbf{r}} \wedge \ddot{\mathbf{r}}|^2, \quad (3)$$

where dots denote differentiation with respect to time.

The following properties of the curvatures are important. The curvature κ_1 determines the deviation of the curve from a tangential straight line at the point \mathbf{r} , and the curvature κ_2 characterises the deviation of the curve from the plane. In other words, κ_1 is a measure of the two-dimensionality whereas κ_2 is a measure of the three-dimensionality of the trajectory. Lastly, a total length L of the trajectory confined between two points $\mathbf{r}_0 \equiv \mathbf{r}(t_0)$ (a starting point) and $\mathbf{r} \equiv \mathbf{r}(t)$ is given by the integral

$$L = \int_{t_0}^t |\dot{\mathbf{r}}(t)| dt. \quad (4)$$

This quantity can be thought of as a measure of "out-of pointness". Thus, we can consider the standard quantities (2)–(4) as differential-geometrical measures of the trajectory complexity at three levels of treatment.

Quantum counterparts of the differential invariants. Let us now turn to the quantum mechanical treatment of the differential geometry invariants considered above. Evidently, the momentum \mathbf{p} as the dynamic quantity for one electron is equivalent to $\dot{\mathbf{r}}$ if atomic units are used (the electron mass is equal to one). Thus, we have the relations

$$\dot{\mathbf{r}} = \mathbf{p}, \quad \ddot{\mathbf{r}} = \dot{\mathbf{p}}, \quad \dddot{\mathbf{r}} = \ddot{\mathbf{p}}. \quad (5)$$

For example, Eq.(4) takes the form

$$L = \int_{t_0}^t |\mathbf{p}| dt \quad (6)$$

and the analogous reformulation is obvious for the other quantities (2), (3). Note that up to a factor the integral (6) is a time-averaged value of the absolute momentum $\langle |\mathbf{p}| \rangle_{time}$. Similar time-averaged quantities can be defined for (2), (3) [23]. It is essential that only the time-averaged quantities $\langle A \rangle_{time}$ are counterparts of the quantum mechanical average values $\langle A \rangle_{\psi}$ where A is a hermitian operator corresponding to a given dynamic quantity A .

Unfortunately, the direct use of the dynamic quantities (2), (3) in quantum me-

chanics involves difficulties due to the possible singularities caused by the denominators in (2), (3). For this reason, following the approach taken in [22, 23] we use the numerators reflecting the principal differential properties of the electronic paths. Taking into account Eqs.(5) and (2), (3) we can thus define three main hermitian operators relevant to three complexity levels

$${}^0\hat{K} = |\mathbf{P}|, \quad (7)$$

$${}^1\hat{K} = \frac{1}{2} |\mathbf{P} \wedge \dot{\mathbf{P}} - \dot{\mathbf{P}} \wedge \mathbf{P}|, \quad (8)$$

$${}^2\hat{K} = \frac{1}{4} (\mathbf{P} \wedge \dot{\mathbf{P}} - \dot{\mathbf{P}} \wedge \mathbf{P}) \cdot \dot{\mathbf{P}} + h.c., \quad (9)$$

where \mathbf{P} is a momentum operator for one electron and h.c. denotes the hermitian conjugation terms. In (7), (8) the symbol $||$ signifies a modulus in the vector sense, same as in the operator sense. In particular, ${}^0\hat{K}$ as the operator modulus of the vector-operator $\mathbf{P} = (P_x, P_y, P_z)$ is computed as

$${}^0\hat{K} = (P_x^2 + P_y^2 + P_z^2)^{1/2}, \quad (10)$$

where nonnegative squared root of the operator expression is conventionally defined in linear algebra (see e.g. [25]). Operators ${}^i\hat{K}$ presented in (7)–(9) are termed the curvature operators of the order i .

Let us now define the average values for the curvature operators, that is the quantities

$$\langle {}^i\hat{K} \rangle, \quad i = 0, 1, 2. \quad (11)$$

Here the symbol of the form $\langle \hat{O} \rangle$ denotes the quantum mechanical averaging with respect to a many-electron stationary state of the molecule, specifically

$$\langle \hat{O} \rangle = \text{Tr} \hat{O} \rho \quad (12)$$

with ρ being the usual spinless one-electron density matrix of the considered many-electron state.

What is left for a complete formal treatment is to provide the rules that determine the quantum counterparts of the time derivatives (5). Because all the operators used correspond to the one-electron dynamical quantities, one can apply the Erenfest quantum equations, such as

$$\dot{\mathbf{P}} = -i[h, \mathbf{R}], \quad \dot{\mathbf{P}} = -i[h, \mathbf{P}], \quad \dot{\mathbf{P}} = -i[h, \dot{\mathbf{P}}]. \quad (13)$$

Here \mathbf{R} is the position operator associated with \mathbf{r} , \hbar is the molecular one-electron

Hamiltonian as a generator for the time evolution.

Interpretation of the indices. It is essential that the average (12) possesses the so-called additive separability. That is for a system AB composed of two isolated parts A and B the quantity (12) is additive:

$$\langle \hat{O} \rangle_{AB} = \langle \hat{O} \rangle_A + \langle \hat{O} \rangle_B. \quad (14)$$

In quantum chemistry this property is only realized for the size-consistent electronic models. Therefore, within the differential geometry description of the molecular complexity the correct molecular indices are

$$\langle {}^i\hat{K} \rangle = \text{Tr}^i \hat{K} \rho, \quad i = 0, 1, 2. \quad (15)$$

Here along with the quantities (15) an additional complexity index is introduced:

$$\langle {}^2\hat{K} \rangle = \text{Tr} |{}^2\hat{K}| \rho. \quad (16)$$

In order to understand the difference between $\langle {}^2K \rangle$ and $\langle {}^2\hat{K} \rangle$ the known properties of the triple scalar product should be taken into account. The absolute value of the latter is equal to the volume of the parallelepiped formed by the vectors in the product. Thus, the index (16) can be considered as a measure of the spatial movement of the electrons in the molecule (the deviation of the moving electrons from a plane). Unlike $\langle {}^2K \rangle$ the indices $\langle {}^0K \rangle$, $\langle {}^1K \rangle$ and (16) are nonnegative scalar quantities, invariant under the transformation of the 3-dimensional physical space. As for $\langle {}^2K \rangle$, we can identify this index with a pseudoscalar, which remains invariant only under proper transformations (pure rotations). Due to the change of the sign with respect to improper rotations and inversion, the index $\langle {}^2K \rangle$ is indeed a molecular chirality measure and this fact was used in [22, 23] to design the electron chirality indices. Clearly, this index is nonzero for dissymmetrical molecular systems only, while the nonzero $\langle {}^2\hat{K} \rangle$ values in general exist for all molecular systems, even for atoms since in reality electrons move in the 3-dimensional space. However, for the topological (that is Hueckel-like) and related models this index naturally vanishes for the planar molecules (see the following sections).

A general issue is worth to be raised in connection with the chirality quantification. When dealing with the extensive molecular characteristics only, a proper extensive chi-

rality measure should be devised. This inevitably leads to the important restriction: such chirality index has to be a pseudoscalar. Indeed, let κ be such quantity satisfying additivity relation (14) for two non-interacting subsystems A and B :

$$\kappa_{AB} = \kappa_A + \kappa_B.$$

If the subsystem B is taken to be a mirror replica, $A^\#$, of the subsystem A then $\kappa_{AA^\#} = 0$ for this "molecular racemate", and $\kappa_{A^\#} = -\kappa_A$ that determines κ as a pseudoscalar. Most of the known geometrical chirality indices (see [4]) being nonnegative do not satisfy the extensivity property that has to do with somewhat artificial way of constructing them from the pure mathematical considerations. Few of the existing chirality measures are pseudoscalar quantities [26–28] and even they are designed in a way that prevents the additivity property.

Our other indices also allow a clear interpretation. The index $\langle {}^1K \rangle$ is nonvanishing if the electron trajectories are not straight lines on average. This index provides a measure of the geometrical nonlinearity of the electron shell in the molecule. It is obvious that $\langle {}^0K \rangle$ describes an effective length of the electron shell. Indeed, in topological approximation this index, up to a prefactor, coincides with the usual geometrical length of the one-dimensional system obtained by arranging its atoms equally spaced on a chosen curve [22]. Thus, we have obtained a description of the electron shell complexity at three levels of the geometrical description, namely as a linear, plane and spatial object.

2. Computation schemes

In order to work out the method outlined above in detail we must specify the matrix representation for the main operators involved in the scheme. Further we will not distinguish operators from the associated matrices. After [22, 23] a very simple approach will be used in this study. We take into consideration only valence atomic orbitals of s type thereby applying M -dimensional basis set of AOs (M is a number of atoms in the molecule under study).

By using the zero differential overlap approximation the coordinate matrix \mathbf{R} can be conveniently computed as a diagonal matrix of the form [22]

$$\mathbf{R} = \|\mathbf{r}_\mu \delta_{\mu\nu}\|, \quad (17)$$

where the set

$$\{\mathbf{r}_\mu\}_{1 \leq \mu \leq M} \quad (18)$$

is the atomic coordinates of the molecule.

The key point is an approximation of the Hamiltonian matrix h . In [22, 23] we treated the molecular electron shell as a Hueckel-like dynamic system for which h is identified with the adjacency matrix B^{top} of the relevant molecular graphs:

$$B_{\mu\nu}^{top} = \begin{cases} 1 & \text{if } \mu \text{ and } \nu \text{ are connected} \\ 0 & \text{otherwise} \end{cases} \quad (19)$$

However, this approximation neglects the long-distance effects important for the molecular chirality and this is a reason why the computational scheme previously suggested was not further developed.

In this work we use a different choice of h by including long-distance as well as short-distance effects described almost in the same manner as in (19). This modified scheme is named here as a quasi-topological approximation. The idea for it is borrowed from the semiempirical theory where many ways exist for computing the resonance integrals matrix, i.e. the matrix B . For our purposes the exponential approximation used in [29] is most appropriate. The adopted method is finally presented by the formula

$$B_{\mu\nu} = (\zeta_{\mu\nu} / |\mathbf{r}_\mu - \mathbf{r}_\nu|)^4, \quad (20)$$

$$\zeta_{\mu\nu} = (\zeta_\mu + \zeta_\nu) / 2.$$

Parameters ζ_μ depend on the atom types only and have dimension of length, so the whole expression (20) is dimensionless, as well as (19). The diagonal elements $B_{\mu\mu}$ are defined to be zero. This scheme

$$h = B \quad (21)$$

works well only for usual organic molecules involving "light" heteroatoms B, N and O. In the case of a system containing heavy heteroatoms (S, P etc) the scheme should be modified in order to produce reasonable values of the resonance integral matrix. In the modified expression a damping factor is introduced for every pair $\{\mu, \nu\}$ containing the heavy heteroatom:

$$B_{\mu\nu} \rightarrow (\text{Min}[\zeta_\mu, \zeta_\nu] / \text{Max}[\zeta_\mu, \zeta_\nu])^2 B_{\mu\nu}. \quad (20')$$

The atomic parameters ζ_μ are the only auxiliary quantities introduced in the com-

putational scheme. At the same time the results of the calculations critically depend on the geometry. Therefore, the method for obtaining the molecular geometry should also be fixed and parameters ζ_μ must be consistent with the molecular geometry used.

Under the homogeneous dilation ($\mathbf{r}_\mu \rightarrow \zeta \mathbf{r}_\mu$) the complexity indices (13) sharply decrease as

$$\langle \hat{K}^0 \rangle \propto \zeta^{-3}, \quad \langle \hat{K}^1 \rangle \propto \zeta^{-10}, \quad \langle \hat{K}^2 \rangle \propto \zeta^{-21}.$$

This sensitivity of the indices can be accepted as reasonable since this automatically leads to vanishing all the indices for the complete dissociation (atomisation). Note that in the proposed quasi-topological approach atoms are naturally (in the structural chemistry context) endowed with the zero values for all the complexity measures.

It is worth to note that while the matrix h (21) is dimensionless the index $\langle \hat{K}^0 \rangle$ has the dimension of length. Analogously $\langle \hat{K}^1 \rangle$ has the dimension of area, and $\langle \hat{K}^2 \rangle$, $\langle \hat{K}^3 \rangle$, have the dimension of volume. This suggests that along with the primary complexity measures (15), (16) suitable measures all having the dimension of length can be defined:

$$l_i = \text{Tr}^i \hat{L} \rho, \quad i = 0, 1, 2 \quad (22)$$

with

$$\hat{L}^0 = \hat{K}^0, \quad \hat{L}^1 = (\hat{K}^1)^{1/2}, \quad \hat{L}^2 = (\hat{K}^2)^{1/3}. \quad (23)$$

This definition allows the indices to be compared on the same grounds.

The "complexity lengths" (22) are extensive quantities as well as the indices (15), (16), so the total complexity length

$$totl = 0l + 1l + 2l \quad (24)$$

possesses the same additive separability. It is essential that (23) has a nonzero value for all real molecular system starting with diatomics. As for the pseudoscalar $\langle \hat{K}^2 \rangle$ having the volume dimension it is not sensible to make a length quantity from it, therefore, it will be calculated as a "pseudovolume" quantity without any modifications.

An important part of the computational scheme is a choice of the density matrix ρ , which realises the averaging procedure. Previously, [22, 23], a very simple approximation was used:

Table 1. Differential geometry complexity measures $\langle \hat{K} \rangle$ (\AA^2), $\langle |\hat{K}| \rangle$ (\AA^3) and complexity lengths 0l , 1l , 2l and ${}^{tot}l$ (\AA) for some saturated hydrocarbons

No.	Molecule	$\langle \hat{K} \rangle$	$\langle \hat{K} \rangle$	0l	1l	2l	${}^{tot}l$
1	CH ₄ (<i>T_d</i>)	1.3	0.0	6.3	1.9	0.0	8.2
2	C ₂ H ₆ (<i>D_{3h}</i>)	7.7	0.5	11.9	6.0	3.0	20.9
3	C ₃ H ₈ (<i>C_{2v}</i>)	17.7	11.1	16.8	13.4	7.8	38.0
4	C ₄ H ₁₀ (<i>C_{2h}</i>)	30.1	26.0	22.2	20.1	13.2	55.6
5	C ₅ H ₁₂ (<i>C_{2v}</i>)	37.6	32.0	27.6	24.7	13.5	65.9
6	C ₆ H ₁₄ (<i>C_{2h}</i>)	45.2	40.9	33.1	29.4	17.5	78.9
7	2-M-Pentane (<i>C₁</i>)	47.9	55.6	33.1	30.1	25.5	88.7
8	3-M-Pentane (<i>C_s</i>)	48.6	50.5	33.1	30.3	22.3	85.8
9	2,2-MM-Butane (<i>C_s</i>)	50.2	36.3	33.3	30.7	18.7	82.7
10	2,2-MM-Butane (<i>C_{2h}</i>)	49.1	48.2	33.2	30.4	21.1	84.7

$$\rho = I.$$

Here we apply the approximation

$$\rho = I + B \quad (25)$$

that roughly models the bond orders and takes into account the long-range effects.

Summarising, we propose the following working algorithm. Given a molecular geometry in explicit form of the atomic coordinates (18), we calculate the Hamiltonian matrix (20) or (20') and the coordinate matrix (17). Then the momentum and its time derivatives matrices (13) are computed to evaluate the curvature operator matrices (7)–(9), along with the operator modulus matrix $|\hat{K}|$. An iterative scheme for obtaining the operator modulus bypassing a complete solution of the eigenvalue problem is easily formulated by using the standard iterative computation for the squared root. By applying Eq.(24) the main complexity measures $\langle \hat{K} \rangle$, $\langle |\hat{K}| \rangle$, $\langle {}^2K \rangle$, and complexity lengths 0l , 1l , 2l and ${}^{tot}l$ are computed using Eqs.(15), (16) and (22)–(25).

3. Numerical results and discussion

All computations for the specific molecules were carried out using AM1 molecular geometry obtained using the package GAMESS [30]. The atomic parameters ζ_{μ} for typical atoms were chosen as follows. It is naturally suggested that the parameter ζ_{H} should be such that in the case of H₂ the molecule calculated within AM1 scheme the corresponding nondiagonal element (19) has to be $B_{\mu\nu} = 1$. A similar requirement is im-

posed on ζ_{C} with ethane as a reference molecule. It leads to the values

$$\zeta_{\text{H}} = 0.68 \text{ \AA}, \quad \zeta_{\text{C}} = 1.50 \text{ \AA}.$$

For the heteroatoms nitrogen and oxygen we adopted the values

$$\zeta_{\text{N}} = 1.40 \text{ \AA}, \quad \zeta_{\text{O}} = 1.50 \text{ \AA}.$$

Table 1 presents the results of computing the complexity measures and complexity lengths for the first members of saturated hydrocarbons. Even for these simplest organic molecules some interesting effects can be observed. As can be seen from this table, for the first six molecules 0l and, to somewhat lesser extent, indices 1l , $\langle \hat{K} \rangle$ are practically linear with increasing number of the carbon atoms. The other complexity measures reflecting spatial characterisation of the electron shell also increase following, however, a nonlinear law. The last four systems in Table 1 are hexane isomers studied previously in [5, 31] as an interesting example of ranging related systems by the molecular complexity indices. Unlike most of the existing complexity measures the "simplest" iso-hexane (molecule 7, Table 1) is more complex if our total complexity length is used. If one takes into consideration the chirality of this system it seems reasonable to consider it as the most complex one (among the isomeric hexanes). For the molecule 7 the chirality index $\langle {}^2K \rangle = 11.0 \text{ \AA}^3$ can be compared with that of ethane conformation with *D_{3h}* symmetry: $\langle {}^2K \rangle = 0.1 \text{ \AA}^3$.

Our numerical experience shows that usually the behaviour of the complexity measures $\langle \hat{K} \rangle$, $\langle |\hat{K}| \rangle$ are similar to those of the corresponding complexity lengths 1l ,

Table 2. Complexity lengths 0l , 1l , 2l and ${}^{tot}l$ (in Å) and chirality index $\langle \hat{K}^2 \rangle$ (in Å³) for some aromatic hydrocarbons

No.	Molecule	0l	1l	2l	${}^{tot}l$	$\langle \hat{K}^2 \rangle$
1	Benzene (D_{6h})	28.6	18.5	0.0	47.0	
2	Naphthalene (D_{2h})	46.4	34.1	0.0	80.5	
3	Azulene (C_s)	46.6	38.1	1.2	84.0	
4	Biphenyl (D_2)	57.1	41.9	28.2	127.1	32.4
5	Anthracene (D_{2h})	66.1	48.0	1.73	115.9	
6	Phenantrene (C_{2v})	66.2	49.4	0.0	115.6	
7	<i>o</i> -Terphenyl (C_2)	85.7	66.5	53.3	205.5	23.3
8	<i>m</i> -Terphenyl (C_{2v})	85.4	59.0	0.0	144.3	
9	<i>p</i> -Terphenyl (D_{2h})	85.4	56.8	0.0	142.1	
10	1,4'-Binaphthyl (C_2)	94.8	74.6	47.3	216.7	16.1
11	1,2'-Binaphthyl (C_1)	94.7	73.2	43.1	211.0	32.2
12	2,3'-Binaphthyl (C_2)	94.6	70.9	36.9	202.4	39.5
13	[4]-Helicene (C_2)	85.0	65.3	48.0	198.2	119.7
14	[5]-Helicene (C_2)	104.0	80.2	62.2	256.3	146.8
15	[6]-Helicene (C_2)	123.1	94.3	74.7	292.1	162.2
16	[7]-Helicene (C_2)	142.2	108.8	87.9	338.9	212.7
17	[8]-Helicene (C_2)	161.3	123.6	102.5	387.4	269.9
18	[9]-Helicene (C_2)	180.5	139.1	114.3	432.8	323.4
19	[10]-Helicene (C_2)	199.6	152.4	127.8	479.8	381.1
20	[11]-Helicene (C_2)	218.7	167.0	143.3	528.0	439.8
21	[12]-Helicene (C_2)	237.9	181.3	155.0	574.2	496.2

2l while the discriminative ability of the former is somewhat higher. Perhaps, constructed from 0K , 1K "complexity volumes" would lead to more suitable indices, useful for QSAR and related problems but generally these entities have no clear physical meaning, for e.g. the pure linear systems, and they are not considered in this work. In all the following tables only the complexity lengths and nonzero chirality measure will be presented.

The next structural class investigated is the aromatics. Typical results for this class are given in Table 2. Naturally, azulene is more complex system comparing to its isomer naphthalene. For the terphenyl isomers the highest total complexity length is for chiral *o*-terphenyl. At the same time this is not true for all chiral binaphthyls, that is the total complexity length is highest for the least chiral isomer. A remarkable class of [*n*]-helicenes as chiral aromatic systems resembling the winding stairs [32] is worth to discuss. In [*n*]-helicenes complexity lengths 0l depends almost linearly on *n*. The other indices are almost linear for large *n*.

The mere fact that the $\langle \hat{K}^2 \rangle$ index gives a systematic increase in the chirality measure is an evidence of the reliability of the proposed computational scheme that improves our previous topological approximation [22, 23].

As an example of the high symmetry chiral systems consider the molecular structures with the C_n and D_n symmetry ($n = 2, 3$) described in [9, 33, 34]. The results presented in Table 3 are not self-evident; some, however, seem to be natural (for example, a low chirality in the hydrogenated fullerenes due to a large distance between the HCHH fragments).

Examining the applicability of the proposed indices to the QSAR problems is out of scope of this paper. Nevertheless, we present an example of the qualitative correspondence of the chirality index and the observed twisting power, β , for the dopants E- and Z-isomers of 2-(4-phenyl-benzylidene)-p-menthane-3-one in MBBA [35]. For these isomers we obtained the chirality indices of 333 Å³ and 252 Å³ respectively that follow the same trend as the observed $|\beta|$ values of 41.9 μm and 8.8 μm do.

Table 3. The spatial extension measure $\langle |\hat{K}|^2 \rangle$ and the chirality index $\langle \hat{K} \rangle$ (both in \AA^3) for some high symmetry chiral systems

Molecule	Symmetry	$\langle \hat{K} ^2 \rangle$	$\langle \hat{K} \rangle$
Cyclohexane (twist)	D_2	64.5	14.5
Twistane	D_2	121.1	30.0
Ditwistane	C_2	213.4	71.1
Tritwistane	D_3	299.3	125.8
Decaline-cis	C_2	129.1	28.0
Perhydroquinacene	C_3	130.9	23.0
Perhydrotriphenylene	D_3	238.6	36.7
3-Layered naphthalenophane	C_2	966.9	31.5
3-Layered anthracenophane	D_2	1208.1	113.4
$C_{60}H_4$	C_2	2456.1	0.6
$C_{60}H_6$	D_3	2539.7	1.6

4. Concluding remarks

In this paper we have introduced several novel complexity measures for discriminating realistic molecular systems. This study demonstrates that using the complexity measures derived from the molecular quantum kinematics [23] it is possible to develop a correct computational technique for a detailed analysis taking into account such nontrivial structural properties as the chirality. In the proposed approach the long-range effects play a fundamental role. Moreover, paying particular attention to these effects allowed us to develop the scheme more efficient than that suggested previously in [22, 23].

In the majority of the publications pursuing the topological and pure geometrical methods the positive semi-definite scalars are considered as acceptable measures of the molecular chirality. Our viewpoint, however, is different. The molecular chirality cannot be studied using the pure mathematical techniques and the important physical restrictions such as the additivity should be imposed on any index quantifying the chemical chirality and this is what requires these measures to be pseudoscalars. Specific calculations confirm that our new computational scheme leads to reasonable results that agree with the expectations in simple cases and, at the same time, provide a useful insight in unclear situations.

References

1. D.Bonchev, N.Trinajstic, *Int. J. Quant. Chem.: Quant. Chem. Symposium*, **16**, 463 (1982).
2. Chemical Applications of Topology and Graph Theory, ed. by R.B.King, Elsevier, Amsterdam (1983).
3. L.H.Hall, L.B.Lier, in: Rev. in Computational Chemistry: Advances, ed. by K.B.Lipowitz and D.B.Boyd, VCH Publishers, New York (1991).
4. R.Todeschini, V.Consonni, Handbook of Molecular Descriptors, Wiley-VCH, New York (2000).
5. S.Nicolis, N.Trinajstic, *J. Chem. Inf. Comput. Sci.*, **40**, 920 (2000).
6. M.Randic, *J. Mol. Graphics*, **20**, 19 (2001).
7. G.Gilat, L.S.Shulman, *Chem. Phys. Lett.*, **121**, 13 (1985).
8. V.E.Kuz'min, I.V.Stelmach, *Zh. Strukt. Khim.*, **28**, 45 (1987).
9. V.I.Sokolov, Introduction to Theoretical Stereochemistry, Gordon & Breach Science, New York (1991).
10. A.B.Buda, T.A.D.Heyde, K.Mislow, *Angew. Chem. Intern. Engl.*, **31(8)**, 989 (1992).
11. H.Zabrodsky, D.Avnir, *J. Am. Chem. Soc.*, **117**, 462 (1995).
12. K.Mislow, *Top. Stereochem.*, **22**, 1 (1999).
13. E.Flapan, When Topology Meets Chemistry, Cambridge University Press, Cambridge (2000).
14. S.H.Bertz, *J. Am. Chem. Soc.*, **103**, 3599 (1981).
15. S.H.Bertz, *J. Am. Chem. Soc.*, **104**, 5801 (1982).
16. S.H.Bertz, C.Rucker, *Croat. Chem. Acta*, **77**, 221 (2004).
17. A.N.Kolmogorov, *Probl. Perdachi Inf.*, **5**, 3 (1969).
18. A.N.Rashevsky, *Bull. Math. Biophys.*, **17**, 229 (1955).
19. A.J.Lichtenberg, M.A.Lieberman, Regular and Stochastic Motion, Springer-Verlag, New York (1983).
20. J.V.de Julian-Ortiz, C.de G.Alopont, I.Rios-Santamarina et al., *J. Mol. Graphics*, **16**, 14 (1998).
21. D.Nerukh, J.H.Frederick, *Chem. Phys. Lett.*, **332**, 145 (2000).
22. A.V.Luzanov, E.N.Babich, *Struct. Chem.*, **3**, 175 (1992).
23. A.V.Luzanov, E.N.Babich, *J. Mol. Struct. (Theochem)*, **333**, 279 (1995).
24. J.Thorpe, Elementary Topics in Differential Geometry, Springer-Verlag, New York (1979).
25. R.Horn, C.R.Johnson, Matrix Analysis, Cambridge University Press, Cambridge (1986).
26. G.Moreau, *J. Chem. Inf. Comput. Sci.*, **37**, 929 (1997).
27. A.Ferarini, P.L.Nordio, *J. Chem. Soc. Perkin Trans. 2*, **455**, 3599 (1998).
28. M.Solymosy, R.J.Low, M.Grayson, M.P.Neal, *J. Chem. Phys.*, **119**, 3567 (2003).
29. H.Kon, *Bull. Chem. Soc. Jap.*, **28**, 275 (1955).
30. M.W.Schmidt, K.K.Baldrige, J.A.Boatz et al., *J. Comput. Chem.*, **14**, 1347 (1993).
31. L.B.Kier, L.H.Hall, *J. Mol. Graphics*, **20**, 77 (2001).
32. Comprehensive Organic Chemistry, vol.1, Stereochemistry, Hydrocarbons, Halo Compounds, ed. J.F.Stoddart, Pergamon Press, Oxford (1979).
33. M.Nakazaki, *Top. Stereochem.*, **15**, 199 (1984).
34. C.Thilgen, I.Gosse, F.Diederich, *Top. Stereochem.*, **23**, 1 (2003).
35. L.Kutulya, V.Vashchenko, G.Semenkova et al., *J. Mol. Cryst & Liquid Cryst.*, **361**, 125 (2001).

Індекси складності та хіральності для молекулярної інформатики: диференційно-геометричний підхід

А.В.Лузанов, Д.Нерух

Запропоновано нові індекси складності молекул, що базуються на ідеї молекулярної квантової кінематики. Матриця гамільтоніана, котру побудовано у деякому квазітопологічному наближенні, задає часову еволюцію модельної електронної системи і відповідні похідні динамічних величин. Підхід дозволяє зазначити середні квантові кінематичні характеристики, пов'язані з кривинами електронних траєкторій, зокрема із скрутом, що відображає хіральність динамічної системи. Особливу увагу надано розробці алгоритму обчислень індексу молекулярної хіральності. Розрахунки конкретних систем демонструють розумну поведінку запропонованих індексів молекулярної складності.