Chirality Codes and Molecular Structure

João Aires-de-Sousa,[†] Johann Gasteiger,[‡] Ivan Gutman,*[,§] and Dušica Vidović[‡,§]

REQUIMTE, CQFB, Departamento de Química, Faculdade de Ciências e Tecnologia, Universidade Nova de Lisboa, Quinta da Torre, 2829-516 Caparica, Portugal, Computer-Chemie-Centrum, Institute of Organic Chemistry, University of Erlangen-Nürnberg, Nägelsbachstrasse 25, D-91052 Erlangen, Germany, and Faculty of Science, University of Kragujevac, P.O. Box 60, 34000 Kragujevac, Serbia and Montenegro

Received October 10, 2003

Some time ago a structure-descriptor, named "chirality code", was put forward [*J. Chem. Inf. Comput. Sci.* **2001**, *41*, 369−375], aimed at distinguishing between enantiomers. The chirality code is a sequence of (typically 100) numbers, being equal to the value of a certain "chirality function" at equidistant points within a chosen interval. For molecules of moderate size the chirality function has thousands of peaks (maxima and minima), one for each quartet of atoms. Therefore it looks as if the chirality code cannot provide a faithful representation of the chirality function and thus a faithful representation of the molecular structure. We now show that functional groups present in the molecule result in clusters of near-lying and partially overlapping peaks, whose position in the chirality code is characteristic for the particular functional group. This enables a sound structural interpretation of the chirality code.

INTRODUCTION

In view of the great importance of molecular chirality in chemistry, biochemistry, pharmacology, etc., much effort has been made to design theoretical methods by which enantiomeric species could be distinguished.[1−13] Each of the numerous types of "chirality measures" consist of some molecular-structure-dependent quantity Q whose absolute values are the same for both enantiomers, but whose signs are opposite. Two cases were studied: (a) when Q is a scalar and (b) when Q is a function in one or more variables, referred to as a "chirality function". Especially important is the theory of chirality functions, elaborated by E. Ruch in the 1970s; for details see the review,[1] the book,[2] and the references therein. In two recent works[12,13] an intermediate approach for describing chirality was put forward: (c) Q is a sequence of numbers, equal to the values of some chirality function at certain preselected points from a chosen interval.

The new approach[12,13] was found to be particularly suitable when the predictions of molecular properties were made by means of neural networks. Namely, the sequence Q, consisting of some 100 numbers, contains sufficient information on the structure of a chiral compound, which a neural network can utilize for predicting differences in the behavior of the two enantiomers.

In the works[12,13] two different kinds of chirality codes were designed, named "conformation-independent chirality code" (CICC) and "conformation-dependent chirality code" (CDCC). In what follows we focus our attention on CDCC, although similar conclusions may probably hold in the case of CICC.

As already explained, the CDCC consists of values of a certain chirality function, $f_{\mathrm{CDCC}}(x)$. This function is defined as follows:[13]

The geometry of a particular conformation of the molecule examined needs to be known, i.e., the coordinates of all atoms in this molecule must be known. This is achieved by using the 3D structure-generator CORINA.[14,15] Some property of each atom must be known, usually charge or polarizability; then a_i denotes the respective property of the ith atom. In all examples reported in this paper a_i is the partial atomic charge, computed by a standard method.[16,17]

Let i,j,k,l be a selection of four atoms of the molecule examined. Let

$$e_{ijkl} = \frac{a_i a_j}{r_{ij}} + \frac{a_i a_k}{r_{ik}} + \frac{a_i a_l}{r_{il}} + \frac{a_j a_k}{r_{jk}} + \frac{a_j a_l}{r_{jl}} + \frac{a_k a_l}{r_{kl}} \qquad (1)$$

where r_{ij} stands for the distance between the atoms i and j, computed as the sum of bond lengths between atoms on the path with a minimum number of bond counts.

Assuming that a Cartesian coordinate system is chosen, such that the atom i is at its origin, that the atom j is located on the x-axes, and that the atom k lies in the xy-plane, one defines a quantity c_{ijkl} as

$$c_{ijkl} = \frac{x_j y_k z_l}{x_j y_k + x_j |z_l| + y_k |z_l|} \qquad (2)$$

The sign of c_{ijkl} depends on whether the atom l lies above ($z_l > 0$) or below ($z_l < 0$) the plane formed by the atoms i,j,k. Hence, enantiomeric arrangements of the atoms i,j,k,l have oppositely valued c_{ijkl}'s. If the atom l lies in the plane formed by the atoms i,j,k, then $c_{ijkl} = 0$.

Now, the chirality function $f_{\mathrm{CDCC}}(x)$ is given by

$$f_{\mathrm{CDCC}}(x) = \sum_{i,j,k,l} c_{ijkl} \exp[-b(x - e_{ijkl})^2] \qquad (3)$$

with the summation including all selections of four atoms

* Corresponding author phone: ++381-34-331-876; fax: ++381-34-335-040; e-mail: gutman@knez.uis.kg.ac.yu.
[†] Universidade Nova de Lisboa.
[‡] University of Erlangen-Nürnberg.
[§] University of Kragujevac.

10.1021/ci030410h CCC: $27.50
Published on Web 02/21/2004

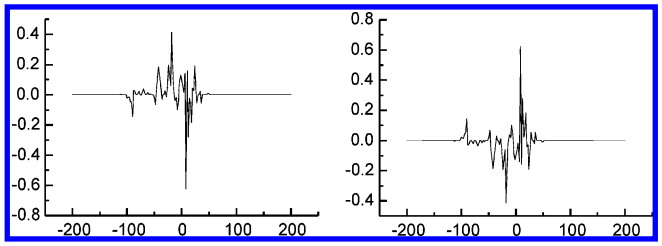

Figure 1. Chirality codes of (R)-2-amino-butane-1-ol (left) and (S)-2-amino-butane-1-ol (right). Details of how the respective calculations have been performed are outlined elsewhere,[12,13] where also additional examples can be found. The values on the abscissa are given in units of 10^{-3} e^2 Å^{-1}; the same applies also to Figures 2−7.

in the molecule examined, and where b is a pertinently chosen "smoothing" parameter.

It is easily seen that, in the general case, if $f_{CDCC}(x)$ is the chirality function of one enantiomer, then $-f_{CDCC}(x)$ is the chirality function of the other enantiomer.

Although the quantities defined via eqs 1−3 do not necessarily have a "physical meaning", it is worth noting that if a_i stands for the charge of the ith atom, then $a_i\, a_j/r_{ij}$ is the potential energy of the electrostatic interaction between the ith and the jth atoms; consequently, e_{ijkl} has the form of a potential-energy term. The quantity c_{ijkl}, defined via eq 2, is a measure of the geometric constellation of the points i, j, k, l; more precisely it is equal to the ratio of the volume and surface of the solid, formed by the points i, j, k, l. The expression f_{CDCC}, defined by eq 3, has a form (but not the "physical meaning") of a radial-distribution function encountered in crystallography. More details on these quantities can be found elsewhere.[12,13]

In eq 1 the distances r_{ij}, r_{ik}, r_{il}, ... depend on the actual conformation of the respective molecule. Therefore the chirality function (3) is also conformation-dependent, as is the chirality code (4) derived from it.

Details explaining why the chirality function $f_{CDCC}(x)$ has been constructed according to eqs 1−3 is found in the works.[12,13]

Let $x_1,x_2,...,x_n$ be some values of the variable x, from a pertinently chosen interval. Then the chirality code associated with the function (3) is the n-tuple:

$$\mathrm{CDCC} = [f_{CDCC}(x_1), f_{CDCC}(x_2),..., f_{CDCC}(x_n)] \qquad (4)$$

Usually (but not necessarily) the numbers $x_1,x_2,...,x_n$ are chosen to be equidistant. The interval to which they belong is chosen so as to embrace all maxima and minima of

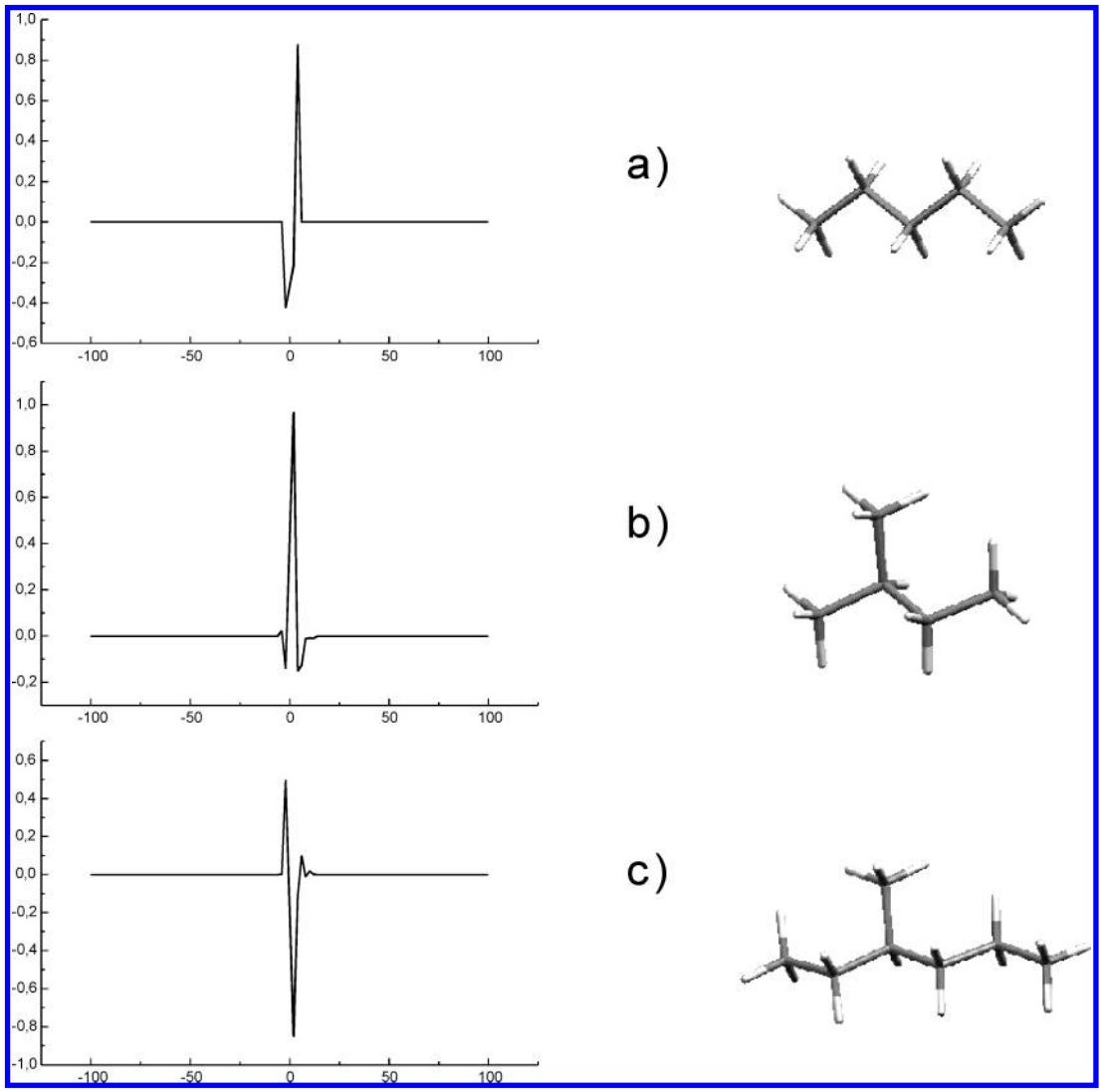

Figure 2. The CDCCs of *n*-pentane (a), 2-methylbutane (b), and (R)-3-methylhexane (c), together with the respective three-dimensional structures of the conformations to which the CDCCs pertain (obtained by means of the structure-generator CORINA[14,15]).

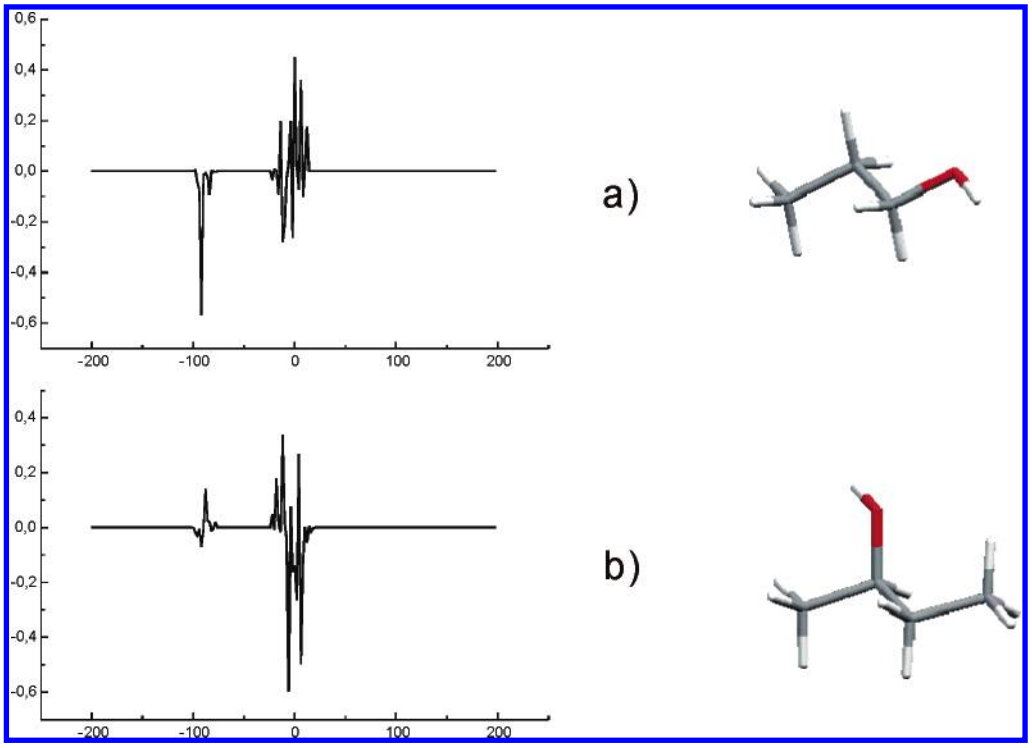

Figure 3. The CDCCs of propane-1-ol (a) and (R)-butane-2-ol (b), together with the three-dimensional structures of the conformations to which the CDCCs pertain.

$f_{CDCC}(x)$. In practical applications the choice $n = 100$ proved to be satisfactory.

In Figure 1 are shown the chirality codes of the two enantiomers of 2-amino-butane-1-ol.

DOES THE CHIRALITY CODE FAITHFULLY REPRESENT THE CHIRALITY FUNCTION?

From eq 3 we see that $f_{CDCC}(x)$ is a linear combination of Gaussian-type functions, each having a maximum (if $c_{ijkl} > 0$) or a minimum (if $c_{ijkl} < 0$) at $x = e_{ijkl}$. The height/depth of such a maximum/minimum is c_{ijkl}. Some of these maxima/minima may coalesce, ultimately forming a complex pattern of "peaks". (Recall that in our case, peaks may be both positive and negative.)

Because the summation in (3) goes over all 4-tuples of atoms, we may expect very many peaks, of the order $\binom{n}{4}$. $\binom{n}{4} = n(n\text{-}1)(n\text{-}2)(n\text{-}3)/24$ is a rapidly increasing function of n. Its values for $n = 12$ (the smallest chiral organic molecules), $n = 25$ (medium-sized chiral organic molecules), and $n = 40$ (typical chiral molecules of pharmacological interest) are 495, 12 650, and 91 390. Thus in practical applications of the CDCCs the chirality function $f_{CDCC}(x)$ may possess many thousands of peaks.

We may now formulate our tantalizing question:

Does a sequence of 100 values of a function with many thousands of peaks provide a faithful representation of this function? or, in other words

Is a chirality code a sound representation of the structure of the underlying molecule? Based on what was said above, the answer seems to be

No.

The only hope to arrive at affirmative answers to the above questions is if there is some systematic grouping of the peaks of the chirality function, that would significantly reduce their number, permitting a short chirality code to reflect all the relevant features.

To check for this possibility we have examined the chirality codes of a number of rather simple organic molecules, in which no or only a single functional group is present. The results obtained are outlined in the subsequent section. However, already at this point we state our answer to the above questions:

Yes, both $f_{CDCC}(x)$ and CDCC consist of clusters of near-lying and partially overlapping peaks, whose position is characteristic for a particular functional group present in the molecule.

DETAILS OF MOLECULAR STRUCTURE REFLECTED IN THE CHIRALITY CODE

Before we present our computational results, we recall that CDCC is conformation-dependent. This has the consequence that also molecules of nonchiral compounds, existing in chiral conformations, have a nonzero chirality code. Even in the case when the conformation examined is (theoretically) symmetric, small deviations from the ideal geometry cause a nonzero chirality code. For our purposes this happened to be an advantage, rather than a drawback: We could examine the structure-dependency of the chirality code on reasonably small and simple examples.

In alkanes, namely in the absence of any functional group, the CDCC possesses just one group of peaks, around $x = 0$. Three characteristic examples are shown in Figure 2.

Although one must never forget that a peak in the CDCC comes from a quartet of atoms, the feature around $x = 0$ may be safely interpreted as originating from the (saturated) C−C and C−H bonds. A similar viewpoint is applicable to groups of peaks indicating the presence of various functional groups.

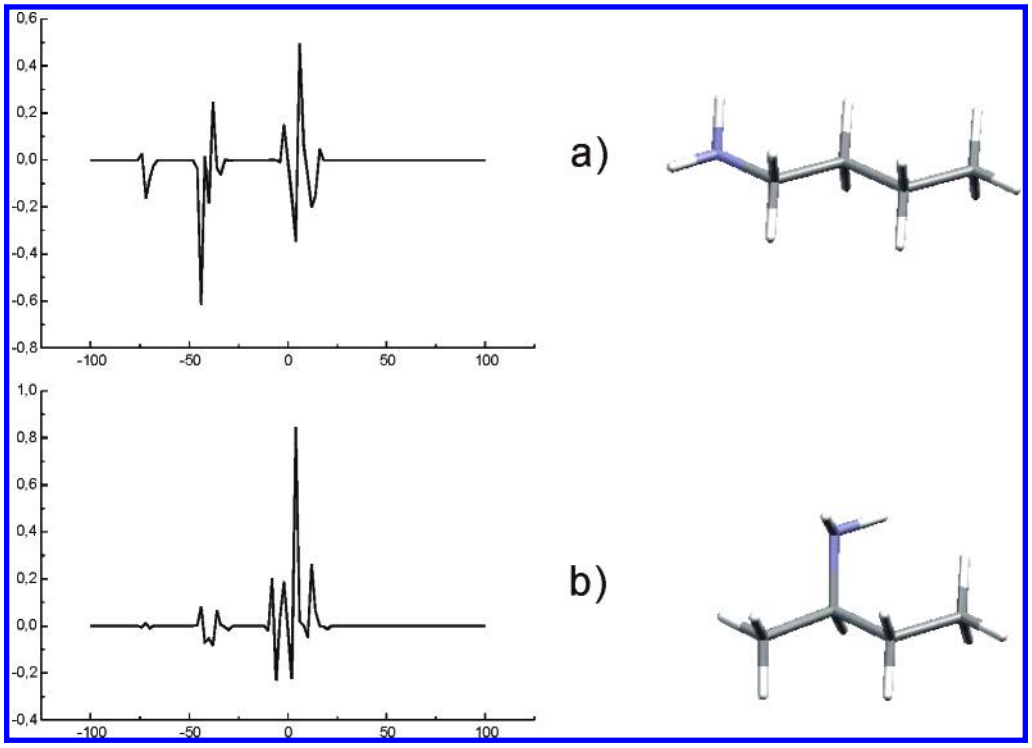

Figure 4. The CDCCs of 1-amino-butane (a) and (R)-2-amino-butane (b), together with the three-dimensional structures of the conformations to which the CDCCs pertain.

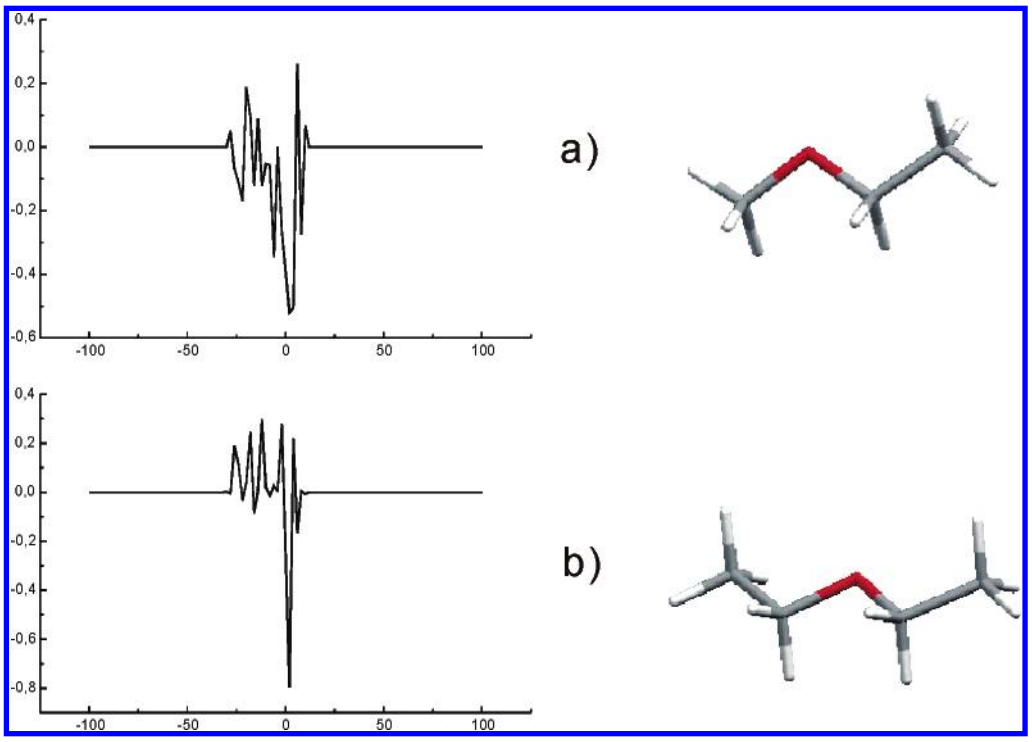

Figure 5. The CDCCs of methyl ethyl ether (a) and diethyl ether (b), together with the three-dimensional structures of the conformations to which the CDCCs pertain.

In Figures 3−7 are shown the CDCCs of two aliphatic alkohols, amines, ethers, ketones, and carboxylic acids, respectively. These examples are typical: in all the studied cases we found that

* alcohols possess a group of peaks around $x = -90$;

* amines possess two groups of peaks around $x = -40$ and $x = -80$;

* ethers possess a group of peaks around $x = -20$;

* ketones possess a group of peaks around $x = -50$;

* carboxylic acids possess two groups of peaks around $x = -50$ (same as in ketones!) and $x = -90$ (same as in alcohols!); another weak peak is observed at $x = -160$, but not in all cases.

The above applies to aliphatic compounds. Their CDCCs also possess a group of peaks around $x = 0$ caused by C−C and C−H groups.

If more than one functional group is present in the molecule, the situation becomes significantly more perplexed.

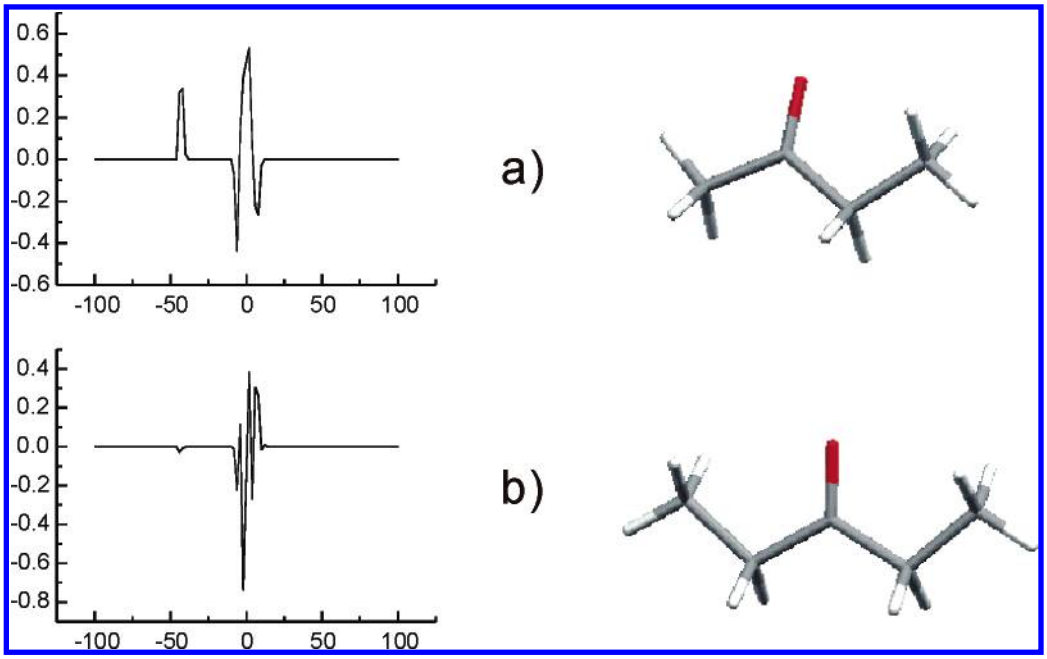

Figure 6. The CDCCs of butane-2-on (a) and pentane-3-on (b), together with the three-dimensional structures of the conformations to which the CDCCs pertain.

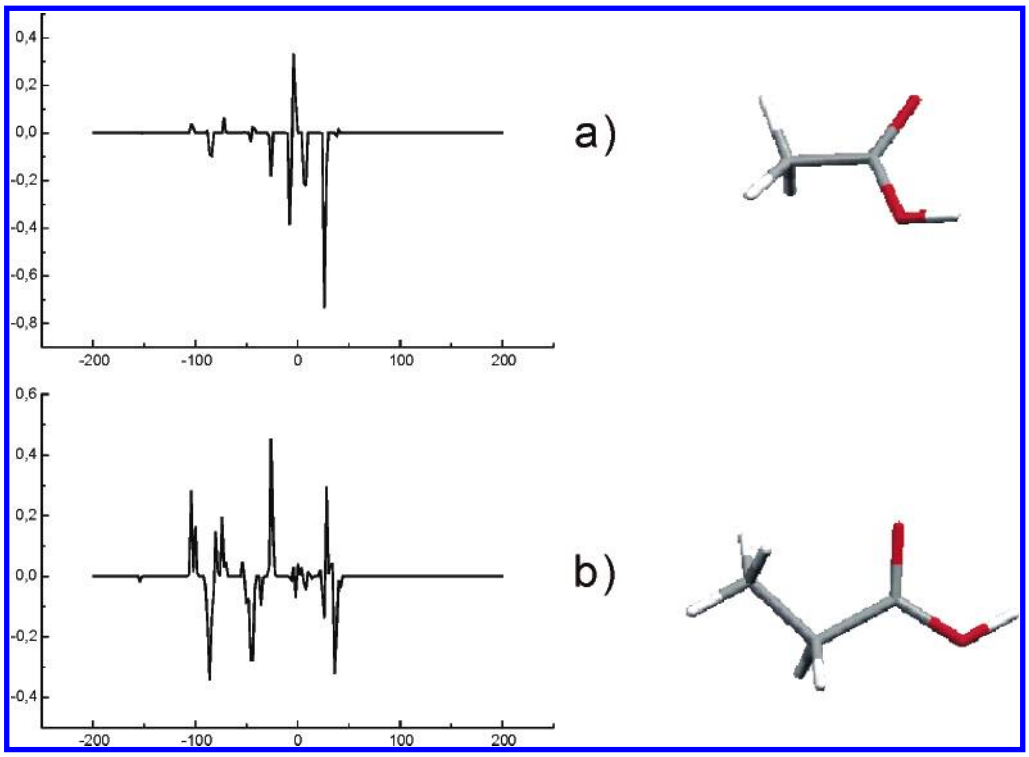

Figure 7. The CDCCs of acetic (a) and propionic acid (b), together with the three-dimensional structures of the conformations to which the CDCCs pertain.

However, as a rule of thumb, the CDCC contains peaks coming from each functional group, positioned nearly at the same place as in the case of monofunctional derivatives; a characteristic example is seen in Figure 1.

CONCLUDING REMARKS

The results outlined in the preceding section clearly point at a rather favorable property of $f_{CDCC}(x)$ and CDCC: their peaks can, in a transparent and chemically sound manner, be associated with the functional groups present in the underlying molecule. Although the conceptual and mathematical basis of the chirality-code-approach is by no means related to any form of molecular spectroscopy, the analogies between CDCCs and molecular (especially IR and NMR) spectra are remarkable.

The regularities noticed by us enable a rational structural interpretation of the features found in the CDCCs. In the simple examples studied by us, these features are easily recognizable. In larger and more complex, but from a point of view of practical applications more interesting, chiral organic molecules obvious and expected complications are encountered. This, however, does not represent a true

difficulty, because—contrary to molecular spectra—we will never be put in the position to have to deduce the structure of a molecule from its CDCC. On the other hand, we now may understand why neural networks can successfully "guess" properties of chiral compounds from the information contained in their CDCCs.

ACKNOWLEDGMENT

I.G. and D.V. thank the Alexander von Humboldt Foundation for financial support, within the "Sonderprogramm Stabilitätspakt für Südosteuropa".

REFERENCES AND NOTES

(1) Ruch, E. Algebraic Aspects of the Chirality Phenomenon in Chemistry. *Acc. Chem. Res.* **1972**, *5*, 49−56.
(2) Mead, C. A. *Symmetry and Chirality*; Springer: Berlin, 1975.
(3) *New Developments in Molecular Chirality;* Mezey, P. G., Ed.; Kluwer: Dordrecht, 1991.
(4) Buda, A. B.; Heide, T.; Mislow, K. On Quantifying Chirality. *Angew. Chem., Int. Ed. Engl.* **1992**, *31*, 989−1007.
(5) Schulz, H. B.; Schulz, E. B.; Schulz, 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.
(6) Zabrodsky, H.; Avnir, D. Continuous Symmetry Measures. 4. Chirality. *J. Am. Chem. Soc.* **1995**, *117*, 462−473.
(7) Moreau, G. Atomic Chirality, a Quantitative Measure of the Chirality of the Environment of an Atom. *J. Chem. Inf. Comput. Sci.* **1997**, *37*, 929−938.
(8) Julián-Ortiz, J. V.; Alapont, C. G.; Ríos-Santamarina, I.; García-Doménech, R.; Gálvez, J. Prediction of Properties of Chiral Compounds by Molecular Topology. *J. Mol. Graphics Model.* **1998**, *16*, 14−18.
(9) Benigni, R.; Cotta-Ramusino, M.; Gallo, G.; Giorgi, F.; Guliani, A.; Vari, M. R. Deriving a Quantitative Chirality Measure from Molecular Similarity Indices. *J. Med. Chem.* **2000**, *43*, 3699−3703.
(10) Golbraikh, A.; Bonchev, D.; Tropsha, A. Novel Chirality Descriptors Derived from Molecular Topology. *J. Chem. Inf. Comput. Sci*. **2001**, *41*, 147−158.
(11) Wildman, S. A.; Crippen, G. M. Validation of DRAPER for 3D QSAR: Conformation Search and Chirality Metric. *J. Chem. Inf. Comput. Sci.* **2003**, *43*, 629−636.
(12) Aires-de-Sousa, J.; Gasteiger, J. New Description of Molecular Chirality and Its Application to the Prediction of the Preferred Enantiomer in Stereoselective Reactions. *J. Chem. Inf. Comput. Sci.* **2001**, *41*, 369−375.
(13) Aires-de-Sousa, J.; Gasteiger, J. Prediction of Enantiomeric Selectivity in Chromatography. Application of Conformation-Dependent and Conformation-Independent Descriptors of Molecular Chirality. *J. Mol. Graphics Model.* **2002**, *20*, 373−388.
(14) Sadowski, J.; Gasteiger, J. From Atoms and Bonds to Three-Dimensional Atomic Coordinates: Automatic Model Builders. *Chem. Rev.* **1993**, *93*, 2567−2581.
(15) Gasteiger, J.; Rudolph, C.; Sadowski, J. Automatic Generation of 3D-Atomic Coordinates for Organic Molecules. *Tetrahedron Comput. Methodology* **1992**, *3*, 537−547.
(16) Gasteiger, J.; Marsili, M. Iterative Partial Equalization of Orbital Electronegativity − A Rapid Access to Atomic Charges. *Tetrahedron* **1980**, *36*, 3219−3228.
(17) The program package PETRA: http://www2.chemie.uni-erlangen.de/software/petra/.

CI030410H