Atomic Chirality, a Quantitative Measure of the Chirality of the Environment of an Atom

Gilles Moreau[†]

Centre de Recherches Roussel-Uclaf, BP 9, 93235 Romainville Cedex, France

Received February 17, 1997[⊗]

In this article we propose a measure of chirality, more particularly of the chirality of an atom's environment in a molecule. The environment is defined both by the geometry of the closest atoms and by a property of these atoms. The chirality measure depends on the position of the atom with respect to the principal planes of the environment. This measure can account for chirality in most practical cases. Measuring chirality is needed in molecular diversity studies in order to distinguish enantiomers and more generally to have a better assessment of similarity. This measure could also find applications in chemistry.

1. INTRODUCTION

Our interest for the measurement of chirality or asymmetry comes from combinatorial chemistry which is a new and rapidly growing technique for obtaining a large number of various molecules. The pharmaceutical industry is very interested in this new activity to find hits, i.e. molecules which have some activity or affinity in various biological tests. In spite of the miniaturization of the tests and of automation, it is still true that testing a large number (several hundred thousand) of compounds remains somewhat expensive because of the ingredients and of the manpower. Hence, the neccessity of restricting the large capacity of the so-called high throughput screening to *different* molecules, but how different? For instance many companies are interested in having a "microcensus", i.e. a reduced and representative set of their corporate collection, smaller and faster to be tested as a first approach. In our company and with use of the autocorrelation method,[1−3] we were able to reduce the size of the Hoechst corporate collection (about 150 000 compounds on shelves) by a factor of about 50 and to detect hits in such a way that no chemical series was missed. To solve this problem, molecular diversity studies are done: each molecular structure is described by a descriptor which is, most of the time, an *n*-dimensional vector, so that a set of molecules can be visualized as a cloud of points in an *n*-dimensional space (Figure 1a). The reduction of the set is simply accomplished by a judicious sampling. If statistical techniques, such as clustering, play an important role in these studies, it is equally important to describe molecules under all relevant aspects. From this point of view, it is obvious that chirality or asymmetry were never used, until now. Why? The primary reason probably is that chirality is not considered as a measurable property. In fact it is only recently that some papers have been published in which various measures of chirality were proposed, however with various limitations in their practical use. But before going further, why is it necessary to quantify chirality? In fact, when one manipulates 3D structures, as for instance in the docking of a molecule into a protein site, there is no need for chirality measurement, since chirality is naturally accounted for by the various positions of atoms in 3D space. On the contrary,
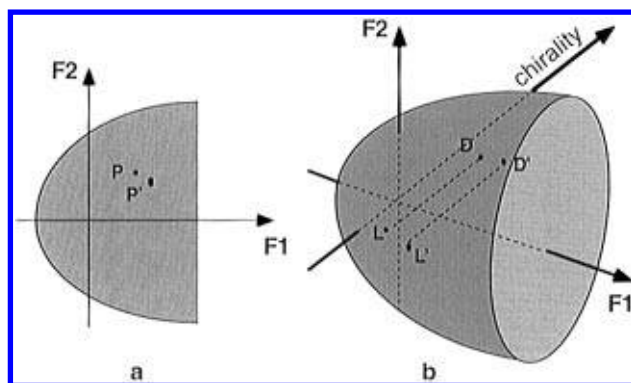
† E-mail address: moreau@mac.rousseluclaf.fr.
⊗ Abstract published in *Advance ACS Abstracts,* August 1, 1997.

Figure 1. Projection on the first principal plane of a set of molecules in autocorrelation. In a chirality is not used; points *P* and *P′* account for two different molecules or conformations *and* their antipodes. In b the use of chirality symbolized by the axis "chirality" should allow one to say which of *L* and *D* (two enantiomers) is similar to *L′* or *D′* (the enantiomers of another structure).

when a molecular structure is described by some sort of descriptor, the exact structure is lost to some extent: for instance, many descriptors are based on distance geometry and classical properties so that they are not affected by a reflection in a mirror plane. With such descriptors two antipodes are represented by the same point (Figure 1a), and we do not know to what extent these two antipodes are similar or dissimilar. For the same reason two different molecules can be represented by two close points, but as long as chirality is not taken into account, we are not sure they are really similar (Figure 1). This defect of the systems based on distance geometry and classical properties tends to decrease their ability to closely cluster really similar molecules.

The object of this article is to define a new atomic property: the quantitative and continuous measure of the chirality of the environment of an atom in a molecule, for any scalar atomic property, easy to use in molecular diversity analyses.

2. METHODS

2.1. Chirality in Chiral and Nonchiral Molecules. Usually one considers that chiral molecules are (at least) those containing one or more chiral centers (i.e. atoms in an
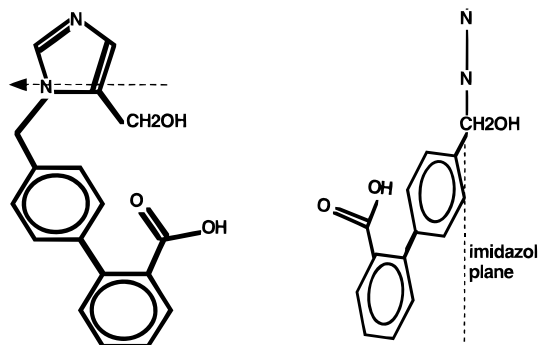
Figure 2. Chiral conformations. If one looks at the low-energy conformation, as indicated by the arrow contained in the plane of the pentagon, the rest of the structure is situated left (as shown) or symmetrically right of this plane. Torsion between the two benzene rings leads also to some chirality.
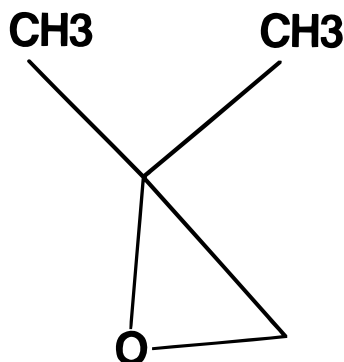

Figure 3. Achiral structure with an achiral conformation. Since the oxygen atom is in the plane of symmetry, the environment it "sees" is symmetric; on the contrary the methyl groups have asymmetric environments. Asymmetry of the environment is by far the most frequent situation for atoms in molecules.

sp³ configuration with different substituents in each direction): so are many natural compounds and their derivatives—steroids, alkaloids, most antibiotics etc.—and some totally synthetic compounds which have been separated in two enantiomers. The presence of such sp³ chiral centers is, however, not indispensable: Hydrogen peroxyde can exist under two (enantiomeric) stable forms with a H—O—O—H dihedral angle of ±90°; similarly the nonstable conformations of ethane in which the dihedral angles H—C—C—H differ from 0 or 60 or 120 or 180° are chiral; ortho-substituted biphenyls are another well-known example of chiral molecules without chiral centers. Pharmaceutical collections contain a large number of compounds with or without chiral centers. In this latter category one finds many molecules whose low-energy conformations are chiral and enantiomeric; an example is shown in Figure 2. Even if the conformational exchange is rapid in solution, it can be blocked if the molecule is docked into a biological receptor or enzyme. A rapid survey of our collection showed that at least 50% of this category possess chiral conformations.

A structure is chiral each time it has no plane of symmetry. When there is a plane of symmetry, the atoms located in this plane have a symmetrical environment, but other atoms do not. Figure 3 shows an achiral structure having an achiral conformation. The oxygen atom has a symmetrical environment since it is in the plane of symmetry, but the methyl groups, which are not in a plane of symmetry, have asymmetrical environments; they are prochiral, and their environments are mirror images of one another. We can
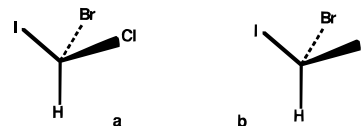

Figure 4. Chirality labeling and chirality measure. In a according to the CIP priority rule the carbon atom is "CIP-R"; if the priority follows the electronegativities of the atoms, it can be said "electronegativity-S". In b the carbon is also CIP-R, but practically there is no chirality.

therefore conclude that asymmetrical environments are not the privilege of atoms in chiral molecules but by far the most frequent situations.

2.2. Chirality of the Environment of an Atom. When a molecule is bound to a biological receptor, various types of interactions are involved: electrostatic interactions, hydrogen bonds between donors and acceptors, hydrophilic or hydrophobic interactions, etc. Each type of interaction is determined by the spatial distribution of electric charges, hydrogens and electron pairs, lipophility and hydrophility, etc., on the atoms in direct contact and probably on some other atoms close to the contact surface. These properties have most often an asymmetric distribution around each atom of the receptor and of the ligand, and recognition and interaction are likely to be affected by the asymmetry relations between both molecules.

The well-known Cahn—Ingold—Prelog rules (CIP rules) have been used by chemists for many years to qualify a chiral tetrahedral atom R or S according to the spatial placement of its substituents and to a priority rule which uses various properties: physical (mass), topological (the graph structure), and pseudostructural (bond orders, pseudoatoms, nomenclature). The result is not a measure but only a labeling of the chirality. For instance the carbon atom of the molecule of chlorobromoiodomethane in Figure 4a is R, we should say "CIP-R"; if we change the priority basis, we obtain what we call "CIP-like" rules.

Let us suppose that the priority follows now the volumes of the substituents, namely, H < Cl < Br < I; the carbon is "volume-R". If the priority follows the electronegativity, H < I < Br < Cl, then the atom can be said to be "electronegativity-S" since the two permutations are not of the same parity. In Figure 4b, the chiral center is CIP-R, yet for most practical properties the magnitude of its chirality should be very small because H and D are very similar.

In the CIP rules (and CIP-like rules) the nature of the central atom is not involved; the labeling R or S is not exactly a property of the atom in its environment but rather a property of its site. In fact, when another atom interacts with the central atom, it does not distinguish between a given property on this atom and the same properties on adjacent atoms. As already mentioned above, this is naturally taken into account if the 3D structures are used, but not if one uses a condensed description. From a quantitative point of view the asymmetry of the distribution of a property around an atom depends also on the property on this atom itself: if the central property is very large in comparison with that of the environment, then the asymmetry is rather small, and conversely if the contribution of the environment is more substantial.

Therefore, there are two slightly different approaches: either the environment is strictly defined as all the atoms except the central atom (option 1) or as all the atoms including the central atom (option 2) which we prefer.
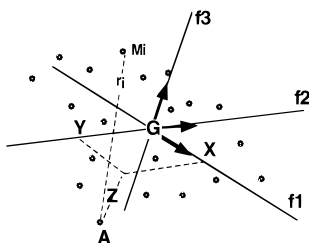
CHIRALITY OF THE ENVIRONMENT OF AN ATOM

J. Chem. Inf. Comput. Sci., Vol. 37, No. 5, 1997 **931**

Figure 5. Atom A surrounded by atoms M_is (property p_i, weight w_i). G is the barycenter of the M_is. The principal axes are Gf_1, Gf_2, and Gf_3. With respect to these axes the coordinates of A are denoted X, Y, and Z. As a first guess the measure of the asymmetry is proportional to XYZ; its sign depends on the octant which contains A.

2.3. General Approach. 2.3.1. Principal Axes and Planes.

We consider a molecular structure in 3D space and a particular atom A. Its environment is made of atoms M_i whose coordinates are x_i, y_i, and z_i ($i = 1, N$); in option 2 atom A is one of the N atoms. The barycenter G (center of mass) of the environment is supposed to be the origin of the coordinates (the calculation of this barycenter takes account of the masses m_i; see below). One can determine the principal axes: they correspond to the three orthogonal directions of extremum extension of the set of points (i.e. directions for which the sum of the squared projections of the GM_is is extremum). The unit vectors on these axes are the eigenvectors associated with the eigenvalues of the matrix **T** whose terms are (the sums $\sum$ are from $i = 1$ to $i = N$):

$$t_{11} = \sum m_i x_i^2 \quad t_{12} = \sum m_i x_i y_i \quad t_{13} = \sum m_i x_i z_i$$
$$t_{21} = \sum m_i x_i y_i \quad t_{22} = \sum m_i y_i^2 \quad t_{23} = \sum m_i y_i z_i$$
$$t_{31} = \sum m_i x_i z_i \quad t_{32} = \sum m_i y_i z_i \quad t_{33} = \sum m_i z_i^2$$

The m_i are the "masses" of the points defined as $m_i = p_i w_i$, where p_i is a property associated with the atoms, supposed to be positive (or negative, but not both) for all atoms. Moreover, some atoms are close to the central atom A, and some others are more remote and less important; it seems therefore justified to weigh their property p_i by a factor w_i which could be an expression like r_0/r_i or $(r_0/r_i)^2$, or $\exp(-(r_i/r_0)^2)$, etc., r_i being the distance AM_i. (In option 2 the weight factor w_a of the central atom A has to be defined specifically if w_i is r_0/r_i or $(r_0/r_i)^2$.

The eigenvalues λ_1, λ_2, and λ_3 of the determinant of the matrix **T** are the solutions (in this sort of problem, they are real and positive) of the characteristic equation below, in which **I** is the unit matrix (diagonal terms = 1; other terms = 0):

$$\text{determinant}(\mathbf{T} - \lambda \cdot \mathbf{I}) = 0 \tag{E1}$$

(Notations: the eigenvalue λ_1 is associated with the eigenvector $\mathbf{v_1}$ which defines the principal axis Gf_1, and so on; the coordinates of A on the principal axes are denoted X, Y, and Z, Figure 5).

2.3.2. Principal Planes and Planes of Symmetry.

Let us suppose the environment—the set of N points—has elements of symmetry such as a plane or an axis; these elements coincide with one of the principal planes or axes, respectively:

The symmetric of a principal plane with respect to the plane of symmetry should also be a principal plane; it can happen only if the plane of symmetry is the same as or perpendicular to a principal plane; therefore it is a principal plane. Similarly the plane obtained by rotating a principal plane around the axis of symmetry should also be a principal plane; being the intersection of two principal planes, this axis is a principal axis.

2.3.3. Measure of Chirality.

Let us suppose now that our eyes are located at point A and look at the environment; what we see is symmetrical if A is in the plane of symmetry. If A is not on this element, what we see is not symmetric: the more distant from the plane, the more dissymmetric it is. Therefore, we can express, as a first guess, the asymmetry (of the environment seen from A) as proportional to the distance from A to the plane. Furthermore if the asymmetry is positive on one side of the plane, it is negative on the other side.

In the general case there is no symmetry for various reasons: the geometry of the set of atom positions has no particular symmetry and even if the structure is geometrically symmetric or almost, this symmetry is broken if the atoms are chemically different. Nevertheless we will consider that the principal planes are the "best approximation" of potential symmetry planes (see Appendix). There are three principal planes, and consequently the asymmetry or chirality AS of the environment could be expressed by the following expression:

$$\text{AS} = C\{XYZ\}S \tag{E2}$$

C is a coefficient; $\{XYZ\}$ is an expression more or less similar to the product XYZ of the coordinates of A with respect to the principal axes; S is +1 or −1, depending on the handedness of $Gf_1 f_2 f_3$. We are now going to discuss these three terms.

2.3.3.1. Coefficient C.

To determine what it could be, we have to examine some special cases:

(a) Two (or three) eigenvalues can be equal; this means that the set of weighted points has the same extension in any direction of a plane (or in any direction of the space). In such cases, two eigenvectors or principal axes, say Gf_2 and Gf_3, are any orthogonal axes in the principal plane orthogonal to the third eigenvector, Gf_1; therefore a principal plane, say $Gf_1 f_2$, can be chosen so as to contain the point A; consequently the asymmetry of the environment is zero. As the calculation of the eigenvectors does not involve the point A in option 1 and as this point is treated as any other points in option 2, it cannot yield a principal plane containing A; the zero value of the asymmetry has to be accounted for by the coefficient C. The three eigenvalues being supposed in the order $\lambda_1 \geq \lambda_2 \geq \lambda_3 \geq 0$, the expression E3 that follows becomes zero each time two eigenvalues are equal; the denominator makes the coefficient C without dimension (an eigenvalue is the sum of the weighted squared coordinates along the principal axis it refers to).

$$C = (\lambda_1 - \lambda_2)(\lambda_2 - \lambda_3)/(\sum \lambda_i)^2 \tag{E3}$$

(b) Another case that can happen is that an eigenvalue equals zero; this means the set of weighted points is planar. A practical example is a sp^3 nitrogen atom carrying three atoms: the three neighboring atoms form a planar environment (in option 1) with two principal directions defined by the two eigenvectors associated with the nonzero eigenvalues. The coefficient C has to be modified as in E4.

$$C = (\lambda_1 - \lambda_2)(\lambda_2 - \lambda_3)\lambda_3/(\sum \lambda_i)^3 \qquad (E4)$$

(c) Finally, if two eigenvalues equal zero, the environment is linear, and the plane defined by this line and atom A is a symmetry plane; therefore the asymmetry of the environment is zero. The coefficient C is also 0.

We remark about the choice of the matrix **T**: The inertia matrix **M** is related to **T** by $\mathbf{M} = q\mathbf{I} - \mathbf{T}$ with $q = \sum m_i(x_i^2 + y_i^2 + z_i^2) = \lambda_1 + \lambda_2 + \lambda_3$; its eigenvalues are $\mu_i = q - \lambda_i$ with $\mu_1 + \mu_2 + \mu_3 = 2q$. The principal planes given by this matrix are the same as those drawn from **T**; the principal axes are in a reverse order. This matrix does not offer any particular advantage, since only the case of a linear set of points is easily identified by $\mu_3 = 0$.

2.3.3.2. Term {*XYZ*}. This term was introduced, above, to express the fact that asymmetry is measured as the remoteness from the plane of symmetry. Such a term indicates that asymmetry is positive or negative according to the octant in which point A is; this presents some analogy with the empirical octant rule used to predict the sign of circular dichroism of saturated ketones.[4]

The expression *XYZ* is, however, not totally satisfying: *X*, *Y*, and *Z* have to be expressed with respect to the "thicknesses" of the "slab" (Figure 6) which approximates the set of weighted points. The half-thicknesses are

$$e_X = (\lambda_1/\sum m_i)^{1/2} \qquad e_Y = (\lambda_2/\sum m_i)^{1/2}$$

$$e_Z = (\lambda_3/\sum m_i)^{1/2}$$

The expression {*XYZ*} is therefore without dimension:

$$\{XYZ\} = XYZ/(e_X e_Y e_Z) \qquad (E5)$$

If all the points M_i are in a plane, λ_3 is null and e_Z is also zero. However, the complete expression giving the asymmetry becomes 0 (since e_Z is in $(\lambda_3)^{1/2}$).

2.3.3.3. Sign *S* of Expression E2. When calculating the eigenvectors (i.e. unit vectors defining the principal axes), they either form a right-handed referential or not. In order to have an intrinsic measure of the chirality, the simplest solution is to define S as the box-product (which equals +1 or −1) of the three eigenvectors

$$S = (\mathbf{v_1}, \mathbf{v_2}, \mathbf{v_3}) \qquad (E6)$$

2.3.3.4. Final Expression for Chirality. We propose as a measure of the chirality of an atomic environment the following expression E7; the factor 10^3 is only for practical reasons:

AS =

$$10^3(\lambda_1 - \lambda_2)(\lambda_2 - \lambda_3)\lambda_3 XYZ(\mathbf{v_1}, \mathbf{v_2}, \mathbf{v_3})/[e_X e_Y e_Z(\sum \lambda_i)^3]$$
$$(E7)$$

The chirality of a whole molecule is simply defined as the sum of the environment chiralities of all its atoms:

molecular chirality =

$$\sum \text{Atomic Environment Chiralities} \quad (E8)$$

3. RESULTS

3.1. Program. To test the ideas presented above, we have written a short program whose essential features are as follows.
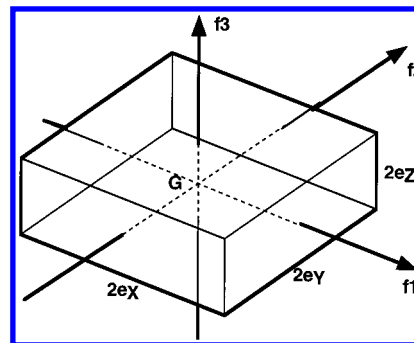

Figure 6. "Slab" thickness. Since λ_1 is the sum of the weighted squared coordinates along f_1, the half-thickness e_X of the slab is defined as $e_X = (\lambda_1/\sum m_i)^{1/2}$, and so on for e_Y and e_Z.
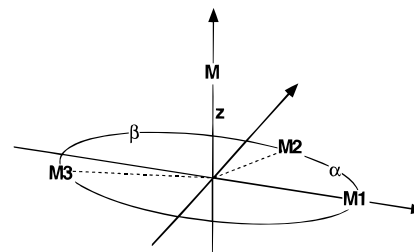

Figure 7. Triangular environment for M formed by M_1 (fixed), M_2, and M_3.

(a) Atoms are given one of the following properties: 1 for any atom, electronegativity according to the Pauling's scale, atomic mass, van der Waals volume.

(b) Their weights w_i can be either chosen as 1 (independent of their distance r_i to the considered atom) or given as r_0/r_i or $(r_0/r_i)^2$, or $\exp(-(r_i/r_0)^2)$.

(c) The eigenvalues of eq E1 are calculated as follows: a first eigenvalue is determined by a dichotomic procedure; the two others are directly calculated as solutions of a second order equation. The eigenvectors are calculated as solutions of the set of equations having the largest determinant; their orthogonality is verified.

(d) The calculation is done in double precision.

3.2. Chiral Triangular Environment. An atom M_1 is placed at coordinates $(1, 0, 0)$, the second atom M_2 is at $(\cos \alpha, \sin \alpha, 0)$, and the third M_3 at $(\cos \beta, \sin \beta, 0)$; they form the environment for a point M located at $(0, 0, z)$ on the z-axis (Figure 7), i.e. at the same distance from M_1, M_2, or M_3. The point M is included in the calculation (option 2). The angles α and β are varied from 0 to 360°, by 1° intervals. The masses are equal to 1 and independent of the distance.

For various values of z, the two angles α and β defining M_2 and M_3 and giving the largest chirality are noted in Table 1. The maximum of chirality is obtained for $z = 0.8845$; the triangle $M_1M_2M_3$ is then almost rectangular. When z approaches 0, the atomic system becomes flat and its chirality vanishes; the limits of α and β, giving the largest chirality, are close to 49 and 157°, respectively. In these conditions the triangle $M_1M_2M_3$ has the edge ratios 1, 0.829, and 0.426: can we compare this triangle with the most chiral triangle calculated by Zabrodsky and Avnir[5] (edge ratios: 1, 0.75, 0.36)? Probably not directly, since their calculation bears on a two-dimensional object, while our calculation is the limit of a three-dimensional environment; in our 3D approach a flat environment or a flat object has no chirality.

For z = 0.3535, α = 120°, and β = 240°, points M_1, M_2, and M_3 define three sp^3 directions around M. Figures 8 and

CHIRALITY OF THE ENVIRONMENT OF AN ATOM

J. Chem. Inf. Comput. Sci., Vol. 37, No. 5, 1997 **933**

Table 1. Maximum Chirality for a Triangular Environment $M_1M_2M_3$[a]

z	α	β	chirality
0.05	49	157	0.017
0.3535	51	161	0.734
0.8845	63	177	2.02
1	66	180	1.96
3	65	179	0.032

[a] When point M is moved along the z-axis, the chirality of its environment goes from 0 (the system is flat), to a maximum for $z = 0.8845$, and back to 0 for larger z. For $z = 0.3535$, $\alpha = 120°$ and $\beta = 240°$, M_1, M_2, and M_3 form an sp^3 geometry around M.
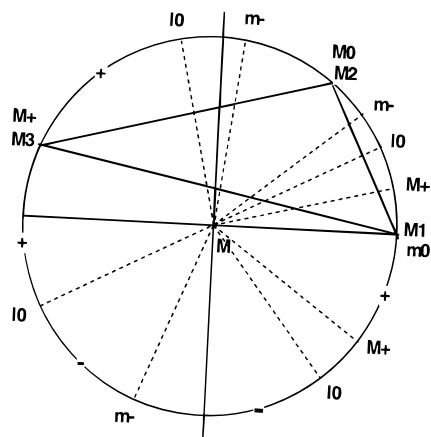

Figure 8. The triangle $M_1M_2M_3$ corresponding to the maximum of chirality for M at coordinate $z = 0.3535$ on the z-axis. When M_1 and M_2 are fixed and M_3 is moved on the circle, the following notations indicate $M+$ is the positive maximum, I_0 is the isosceles triangle $M_1M_2M_3$ and chirality is null, $m-$ negative minimum, M_0 is the null maximum, m_0 is the null minimum.

9 represent the variation of chirality when β goes from 0 to 360°, while $z = 0.3535$ and $\alpha = 51°$. On this curve we can see that each time $M_1M_2M_3$ is isosceles ($\beta = 25.5$, 102, 205.5, 309), the chirality goes through 0 and its sign changes. Chirality is also null when M_2 or M_3 is superimposed with M_1; its sign, however, does not change since the triangles $M_1M_2M_3$, just before and just after the superimposition (see Figure 10) are similar and all three points have the same masses.

3.3. Most Chiral Tetrahedron. The chirality of an object has been defined as the sum of the environment chirality values for all points of this object.

We determined the most chiral tetrahedron in two cases: with masses equal to 1 independent of the distances between the points and with distance dependent masses proportional to $\exp(-M_iM_j^2/R_0^2)$ in which $R_0 = 2$. Option 2 is used (i.e. the calculated point is included in the principal planes determination); expression E7 is used.

We proceeded in the following way: a first point M_1 is placed at coordinates (1, 0, 0). The coordinates of M_2 are (cos α, sin α, 0) with 0° < α < 180°. The coordinates of M_3 are (cos ϕ cos β, cos ϕ sin β, sin ϕ) with 0° < β < 360° and 0° < ϕ < 90°. M_4 is defined as (cos η cos γ, cos η sin γ, sin η) with 0° < γ < 360° and −90° < η < 90°. The maximum of chirality was found by varying the five angles, first by large increments (about 10°) and then by smaller ones, down to 0.1° intervals.

In the case of distance independent masses, the most chiral tetrahedron has three different edge lengths: 1.88, 1.70, 1.25 (ratios: 1.504, 1.361, 1), opposed edges are equal, and there

are three axes of C_2 symmetry joining the middles of opposite edges (Figure 11).

In the case of distance dependent masses, the most chiral tetrahedron seems to be more chiral since the edge lengths are more spread: 1.97, 1.54, 1.54, 1.46, 1.07, 1.07 (ratios: 1.84, 1.44, 1.44, 1.36, 1, 1). There is one axis of C_2 symmetry joining the middles of edges equal to 1.97 and 1.46. This tetrahedron can be compared with the one calculated by Zabrodsky and Avnir[5] (edge ratios: 2.3, 1.6, 1.6, 1, 1, 1), which possesses also one C_2 symmetry axis. The Z and A tetrahedron edge lengths are more spread than in our tetrahedron; moreover two constructing triangles are isosceles in the Z and A cases, while all are scalene in our case.

3.4. Local Chiralities on Some Real Molecules. 3.4.1. Four Hexoses: Allose, Altrose, Glucose, and Idose. The hexoses (and other saccharides and polysaccharides) are a good test for a method aiming at distinguishing diastereoisomeric molecules. The environment chirality for each atom of these four hexoses (see Figure 12) were calculated with the property 1 on each atom (the results are very similar with electronegativity), a weighing proportional to exp(−$M_iM_j^2/R_0^2$) in which $R_0 = 2$ Å was applied.

The results are listed in Table 2 and presented graphically in Figure 12. This simple example shows that these four hexoses can easily be distinguished on the basis of their local chiralities.

3.4.2. Batracotoxine. This complex molecule is represented in Figure 13; it is made of an almost rigid part and of a flexible chain. The histogram of the local chiralities calculated with the property 1 on each atom with the same weighing as above is given in Figure 13. We see that the quaternary atoms C2, C8, C9, C12, and C13, (which are usually known as chiral centers) have almost no local chirality: the reason is that their geometry is very close to sp^3 with the four heaviest neighbors at almost the same distance. The barycenters of these environments are close to the quaternary atoms, and they are almost symmetric. In contrast, atom C22 has the largest chirality (Figure 14 represents the principal axes of the C22 environment), and atom C20, in spite of being a methyl group, has a fairly large chirality: in both cases it is explained by the distance of the barycenters of the systems to C22 or C20 and also by the dissymmetric occupation of the space around these atoms (Figure 14).

4. DISCUSSION

The problem of measuring chirality has been dealt with by several authors;[5−8] most of the time, they consider the totality of an object: for instance, Zabrodsky and Avnir[5] define the chirality as the distance (i.e. without sign) of an object to the nearest symmetrical object (which can be of various symmetry groups, most of the time a reflection in a plane). Kuz'min and others[6] introduced the main axes of inertia (which are the same as the principal axes we use, in a reverse order, since inertia is taken in a mechanical sense rather than in an informational sense). The way they quantify chirality is, however, totally different since they evaluate some sort of distance between a molecular structure and its antipode. A common feature is the fact they use various properties such as the ones we mentioned above for our approach. Another approach, proposed by Rassat,[7] is based
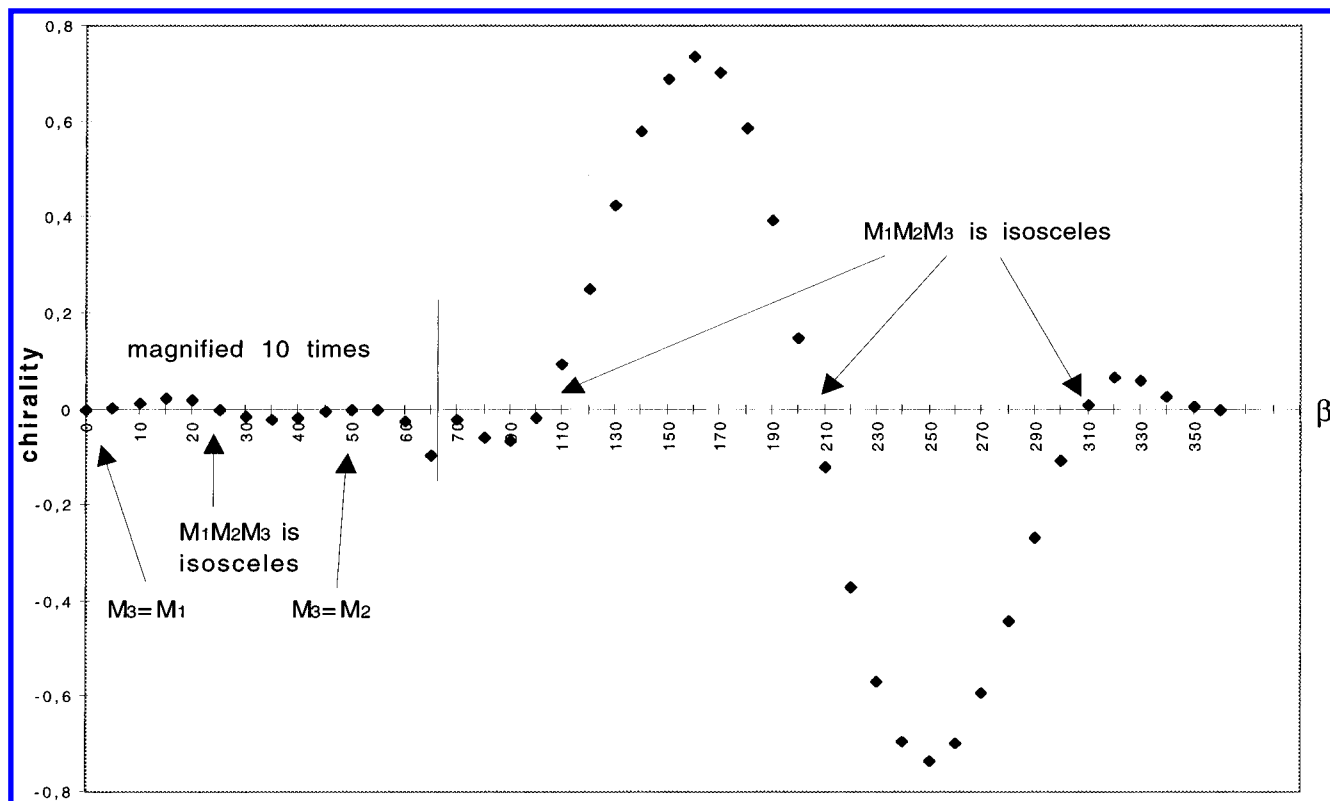
Figure 9. Plot of chirality versus the position of point M_3 (angle β), when M_1 and M_2 are fixed as in Figure 8.
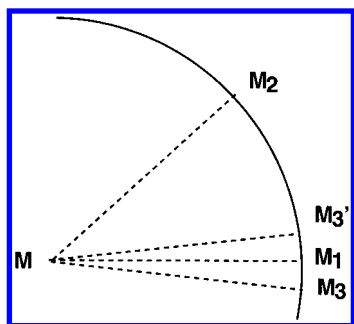

Figure 10. Chirality just before and after superimposition. When the point M_3 is very close to M_1 (positions M_3 and M_3') the triangles $M_1M_2M_3$ and $M_3'M_2M_1$ (correspondence in that order) are almost similar. Since M_1 and M_3 have the same property and weight, they can be exchanged; the environment does not change. This explains why the chirality becomes zero without a sign change when M_3 is superimposed with M_1.
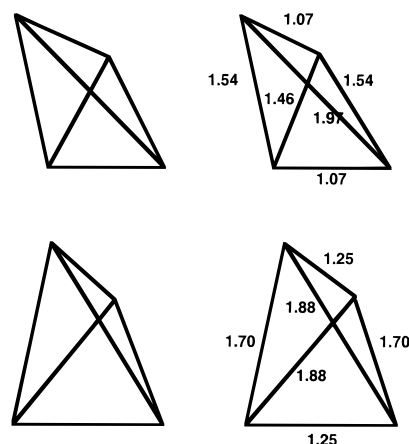

Figure 11. Stereoscopic projection of the most chiral tetrahedron: (bottom) with distance independent weights; (top) with distance dependent weights (see text). This tetrahedron has only one C_2 axis joining the middles of edges 1.97 and 1.46.

on the Hausdorf distance of two sets of points: more exactly, the distances of the chiral object to each of two reference antipode objects are determined, and then the object is said to be of the same handedness as the closest reference antipode. Other developments can be found in a review by Buda et al.[8]

Our approach mainly differs on the following points:

(a) The notion of environment chirality seems to be physically more meaningful than the chirality of an overall structure since interactions usually vanish with distance. The effect of weighing the atoms according to their distance to the central atom appears particularly important, as shown by the determination of the most chiral tetrahedron in comparison with that found by other authors.[5]

(b) In our approach the sign of the chirality of the environment (and also the sign of the molecular chirality

defined as the sum of the atomic environment chiralities) is unambiguously defined.

(c) This approach presents, however, some limitations: if the set of weighted points possesses an axis of symmetry of order $n > 2$, then two eigenvalues are equal and the corresponding principal axes are any orthogonal axes in their plane.

This defect appears, for instance, in the structures drawn on Figure 15: there is no plane of symmetry but an axis of order 3. The environment of point M on the axis is chiral since structures 15a and 15b are not superimposable. In our approach, the environment of point M, whatever the weights are (distance dependent or not), has always an axis of C_3-symmetry, which is also one of the principal axes; consequently the environment chirality of M is null in structures 15a and in 15b.
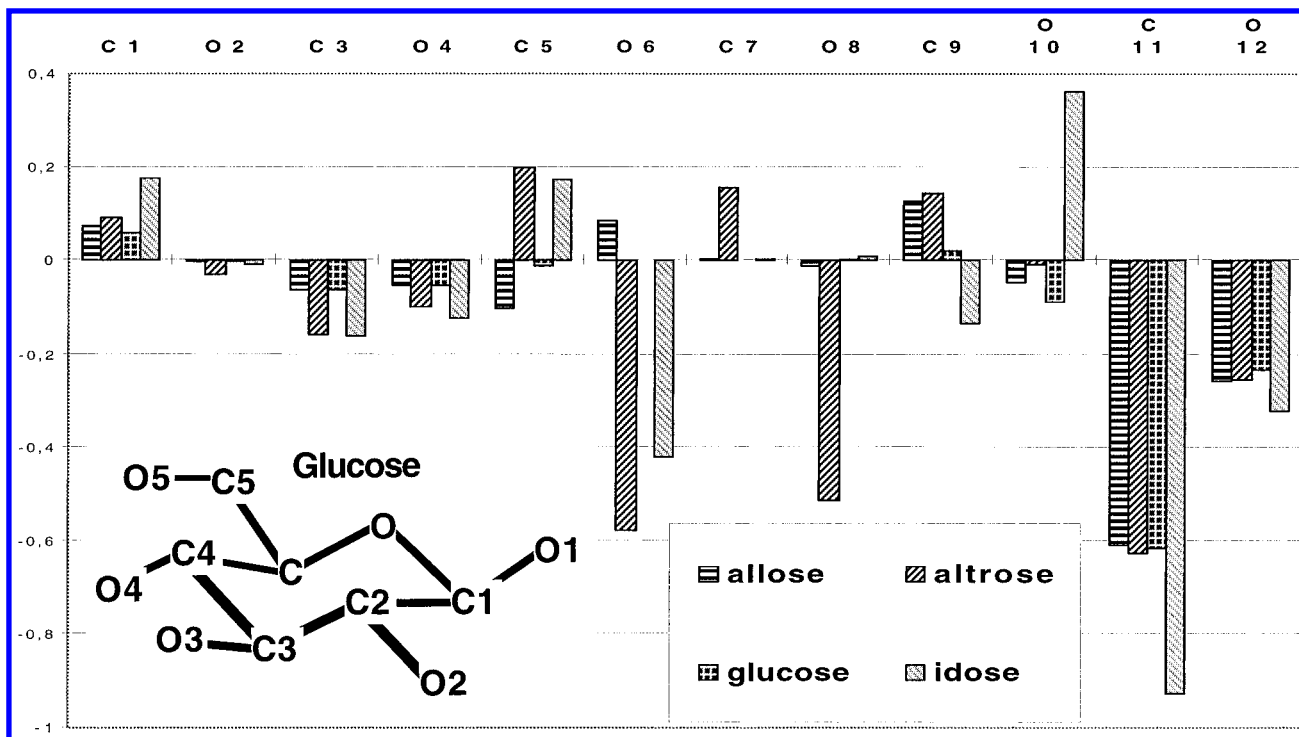
Figure 12. Environmental chiralities. The conformation used for glucose is "all equatorial"; for the other molecules changes were made as follows: in allose O3 is axial, in altrose O2 and O3 are axial, and in idose O2, O3, and O4 are axial.
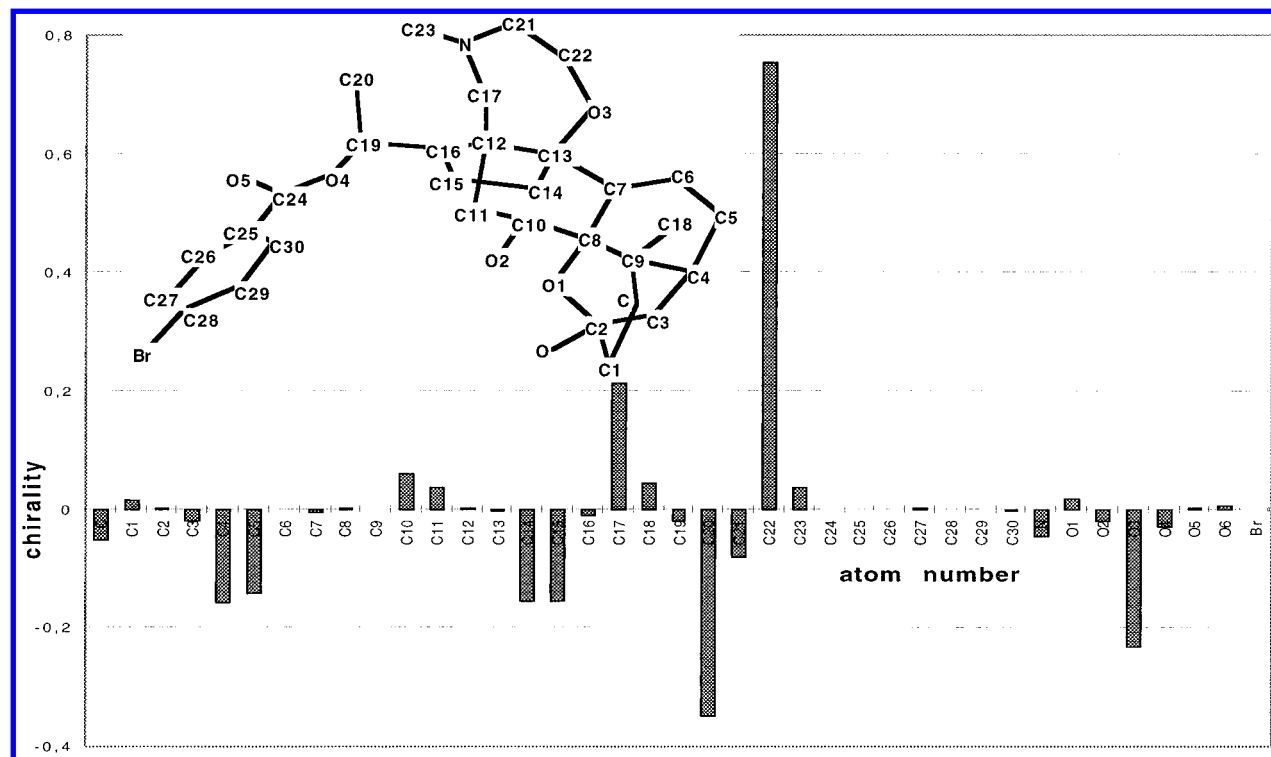

Figure 13. Conformation, atom names, and histogram of the atomic chiralities of batracotoxine, showing the large variation of this property.

For the other points the situation depends on the weighing:

If the weights are independent of the distance, the set of principal axes is the same for any point and for the entire object; because of the axis of symmetry of order 3 (or more), two eigenvalues are equal and the calculated chirality is null.

If the weights are distance dependent, the weighted structure around each point (such as P) is no longer C_3-symmetric and consequently point P has a nonzero calculated chirality.

5. CONCLUSION

We have presented a method for the positive or negative quantification of the chirality of the environment of an atom in a molecule (and for the chirality of a whole object); this method has some theoretical limitations in highly symmetric situations, which should not impair its application to practical problems such as the assessment of the diversity of organic molecules and their antipodes. We hope also that several
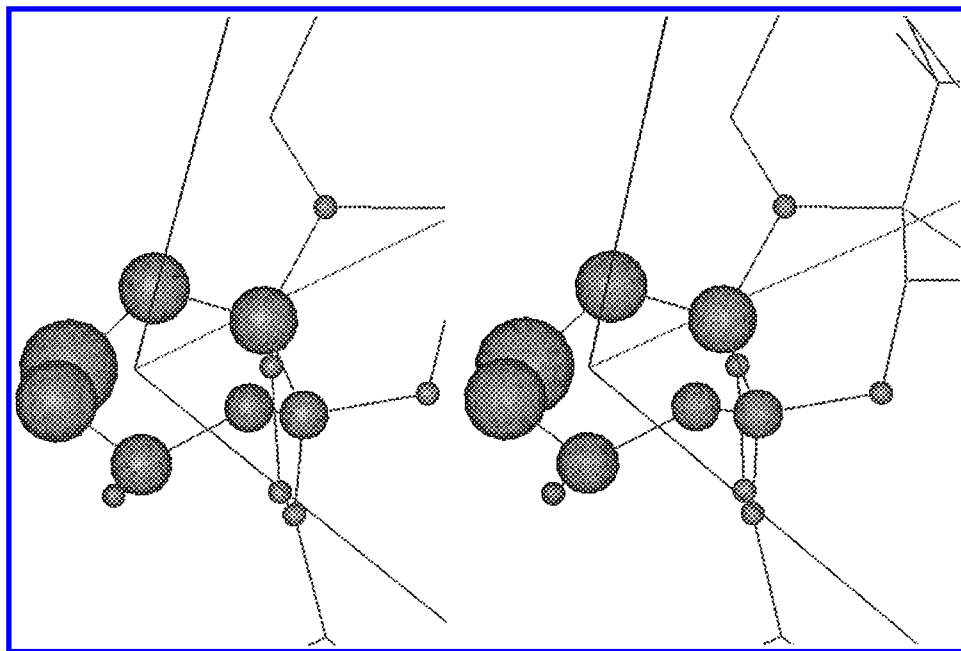
Figure 14. Stereoscopic view of the C22 environment and its principal axes. The larger the ball the closer to C22. This environment looks like an irregular helix.

Table 2. Chiralities of the Atoms of Some Hexoses

|	allose	altrose	glucose	idose
C	0.073 0	0.092 1	0.057 3	0.174
O	−0.002 32	−0.030 0	−0.001 36	−0.009 61
C1	−0.061 0	−0.156	−0.061 1	−0.159
O1	−0.051 6	−0.098 7	−0.053 0	−0.121
C2	−0.102	0.200	−0.012 2	0.173
O2	0.086 0	−0.578	0.001 03	−0.421
C3	0.002 21	0.155	−0.000 706	0.004 21
O3	−0.011 7	−0.512	0.004 20	0.009 72
C4	0.127	0.144	0.022 1	−0.13 3
O4	−0.046 5	−0.007 11	−0.086 6	0.363
C5	−0.609	−0.627	−0.615	−0.92 8
O5	−0.256	−0.254	−0.234	−0.321
total chirality	−0.852	−1.670	−0.980	−1.37
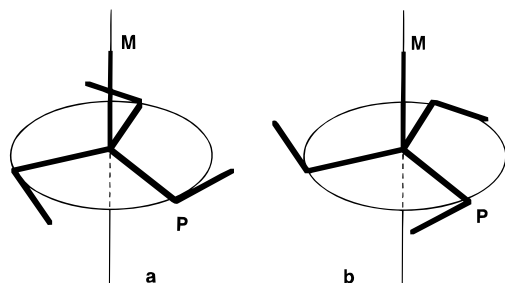
Figure 15. Antipode structures. The environment of *M* has a C_3 axis of symmetry, which is also one of the principal axes; therefore the calculated chirality of *M* is null. Other points, like *P*, have a nonzero chirality provided the weighing is distance dependent.

other applications such as the correlation between physico-chemical properties and calculated chirality could be derived.
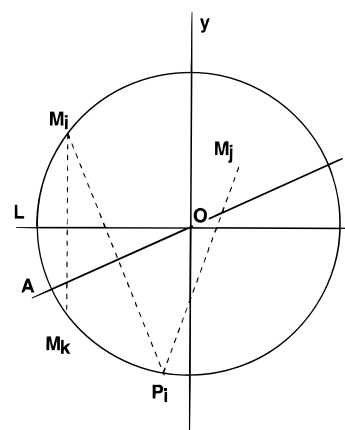
Figure 16.

APPENDIX: TENTATIVE PROOF OF "THE PRINCIPAL PLANES ARE AN APPROXIMATION TO SYMMETRY PLANES"

We consider a set (*M*) of *n* geometrical points M_i in the plane (the extension to 3D is obvious) having an axis of symmetry *L*, Figure 16. The problem we want to solve is, how could we find the symmetry axis (if we do not know it)?

For this purpose let us consider an axis *A* containing O, the barycenter of (*M*) and origin of the coordinates. We designate by (*P*) the set of points P_i symmetric of (*M*) with respect to the axis *A*. Each point P_i will move on a circle when the axis *A* turns around O. During this motion point P_i will reach position M_k symmetric of M_i with respect to *L*; at this moment all the other points of (*P*) will coincide with a point of (*M*). However, for some particular positions of *A*, it could happen that some points of (*P*) coincide with points of (*M*) while *A* is not a symmetry axis (for instance if two points of (*M*) are at the same distance from the barycenter without being symmetric with respect to *L*).

The difficulty comes also from the fact that we do not know which P_i has to be superimposed to which M_k (in other words the points M_i are numbered in any order).
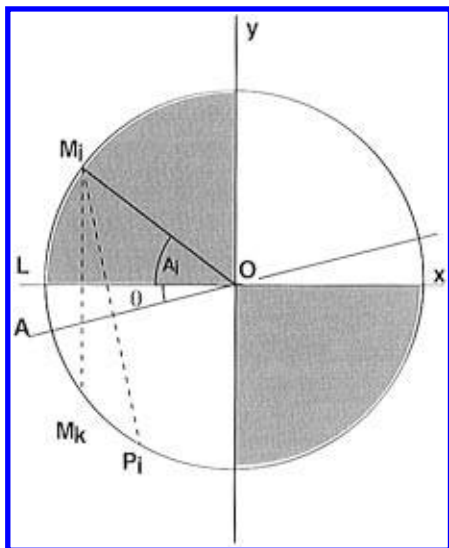
CHIRALITY OF THE ENVIRONMENT OF AN ATOM

J. Chem. Inf. Comput. Sci., Vol. 37, No. 5, 1997 **937**

Figure 17.

To detect the position of axis A which corresponds to a complete one to one superimposition of (P) on (M), we introduce an "overlap function" $R(P_i, M_j)$ of two points P_i and M_j, expression D1 as follows, where P_iM_j is the distance separating P_i and M_j. If $P_iM_j = 0$, the overlap of the two

$$R(P_i, M_j) = \exp(-P_iM_j^2) \qquad (D1)$$

points is maximum and equals 1; the larger P_iM_j is, the smaller the overlap is. We define now the "overlap" between (M) and (P) by the product of all pair overlap functions, expression D2. Expression D2 represents the product of the

$$R((P),(M)) = \prod \exp(-P_iM_j^2)$$
$$[1 \leq i \leq n, 1 \leq j \leq n, i \neq j] \quad (D2)$$

overlaps of each point P_i with all points of (M) except M_i whose P_i is the symmetric with respect to A. The reasons for this are as follows: firstly, we do not know which M_j will be superimposed with a particular P_i; secondly the overlap with M_i is excluded since we do not want to superimpose P_i with M_i.

Now we are going to determine the extrema of $R((P),(M))$, and then we will show that at least one of these extrema corresponds to an axis of symmetry.

The differential of R is $dR = R\sum d(P_iM_j^2)$. The squared distance $P_iM_j^2$ can be calculated as follows ($\mathbf{P_iM_j}$ in bold characters means the vector):

$$\mathbf{P_iM_j} = \mathbf{P_iM_i} + \mathbf{M_iO} + \mathbf{OM_j}$$

$$P_iM_j^2 = \mathbf{P_iM_j \cdot P_iM_j} = \mathbf{P_iM_i \cdot P_iM_i} + \mathbf{M_iO \cdot M_iO} +$$
$$\mathbf{OM_j \cdot OM_j} + 2\mathbf{P_iM_i \cdot M_iO} + 2\cdot\mathbf{P_iM_i \cdot OM_j} +$$
$$2\cdot\mathbf{M_iO \cdot OM_j}$$

Denoting by $\mathbf{N}$ the unit vector perpendicular to the axis A, one gets

$$\mathbf{P_iM_i} = 2\mathbf{N}\cdot(\mathbf{OM_i \cdot N})$$

Finally

$$d(P_iM_j^2) = 8\mathbf{N}\cdot(\mathbf{OM_i \cdot N})(d\mathbf{N}\cdot(\mathbf{OM_i \cdot N}) + \mathbf{N}\cdot$$
$$(\mathbf{OM_i}.d\mathbf{N})) + 4\mathbf{M_iO}\cdot(d\mathbf{N}\cdot(\mathbf{OM_i \cdot N}) + \mathbf{N}\cdot(\mathbf{OM_i \cdot d\mathbf{N}})) +$$
$$4\mathbf{OM_j}\cdot(d\mathbf{N}\cdot(\mathbf{OM_i \cdot N}) + \mathbf{N}\cdot(\mathbf{OM_i \cdot d\mathbf{N}}))$$

Since $\mathbf{N \cdot N} = 1$, $\mathbf{N}\, d\mathbf{N} = 0$; after some arrangements we get

$$d(P_iM_j^2) = 4\mathbf{OM_j}\cdot(d\mathbf{N}\cdot(\mathbf{OM_i \cdot N}) + \mathbf{N}\cdot(\mathbf{OM_i \cdot d\mathbf{N}}))$$

In passing we see that $\sum d(P_iM_j^2)$ over all values of i, including $i = j$, is null since $\sum \mathbf{OM_i} = 0$; therefore $\sum(P_iM_j^2)$ is invariant. We can also write

$$\sum d(P_iM_j^2) \, [1 \leq i \leq n, 1 \leq j \leq n, i \neq j] =$$
$$-\sum d(P_iM_i^2) \, [1 \leq i \leq n]$$

This result will be used later. Designating by α, β the components of $\mathbf{N}$, one obtains

$$d(P_iM_j^2) = 4\{(x_j\, d\alpha + y_j\, d\beta)(\alpha x_i + \beta y_i) +$$
$$(\alpha x_j + \beta y_j)(x_i\, d\alpha + y_i\, d\beta)\}$$

$$\sum d(P_iM_j^2) = \sum\{4\, d\alpha(x_j(\alpha x_i + \beta y_i) + x_i(\alpha x_j + \beta y_j)) +$$
$$4\, d\beta(y_j(\alpha x_i + \beta y_i) + y_i(\alpha x_j + \beta y_j))\}$$

The sums above are for $1 \leq i \leq n, 1 \leq j \leq n, i \neq j$. Being given the constraint on $d\alpha$ and $d\beta$, $\alpha\, d\alpha + \beta\, d\beta = 0$, the extrema of $\sum d(P_iM_j^2)$ are obtained when the coefficients of $d\alpha$ and $d\beta$ are proportional to α and β, respectively; i.e. after some arrangement,

$$\alpha\left(2\sum x_jx_i - \mu\right) + \beta\sum(x_jy_i + x_iy_j) = 0$$
$$\alpha\sum(y_jx_i + y_ix_j) + \beta\left(2\sum y_jy_i - \mu\right) = 0$$
$$1 \leq i \leq n, \quad 1 \leq j \leq n, \quad i \neq j$$

This system bears a nontrivial solution only if its determinant is null. Because of relations such as $\sum x_j = 0$ $[1 \leq i \leq n]$, the sum $\sum x_j\,[1 \leq j \leq n, i \neq j] = -x_i$. Therefore we can write the preceding system as

$$\alpha\left(-2\sum x_i^2 - \mu\right) - 2\beta\sum x_iy_i = 0$$
$$2\alpha\sum(-y_ix_i) + \beta\left(-2\sum y_i^2 - \mu\right) = 0$$
$$[1 \leq i \leq n]$$

Writing the determinant is null (and $\mu = -2\lambda$) yields the characteristic equation of the set (M), eventually leading to the principal directions of this set (see expression E1).

Until now we have proved that the extrema of $\sum(P_iM_j^2)$ are obtained when the axis A is any of the two principal directions of the set (M). As this sum includes the overlap functions of pairs which are not in coincidence when all the right pairs are, then we have to prove that the coincidence of sets (M) and (P) is an extremum of $\sum(P_iM_j^2)$.

Let us consider Figure 17 in which L is still the axis of symmetry and A is another axis forming a small angle θ with L. P_i is the symmetric of M_i with respect to A, M_k is the symmetric of M_i with respect to L, and P_k is the symmetric of M_k with respect to A. The notation M_i is for points located in the grey area; the M_k are their symmetric located in the white area. The sum $\sum(P_jM_j^2)$ over all j can be written as

$$\sum (P_j M_j{}^2) = \sum 4\{OM_i{}^2 (\sin^2(A_i + \theta) + OM_k{}^2 \sin^2(A_k - \theta)\}$$

The latter sum is calculated for all M_i's and all their symmetric M_k's. A_i is the angle formed by OM_i and L. M_k being the symmetric of M_i with respect to L, $A_k = A_i$ and $OM_i = OM_k$. One can write without approximation

$$\sum (P_j M_j{}^2) = \sum 4OM_i{}^2\{(\sin A_i \cos \theta + \cos A_i \sin \theta)^2 + (\sin A_i \cos \theta - \cos A_i \sin \theta)^2\}$$

$$\sum (P_j M_j{}^2) = \sum 8OM_i{}^2(\sin^2 A_i \cos^2 \theta + \cos^2 A_i \sin^2 \theta)$$

The sum over the index i is over the grey area in Figure 17. Now with θ being a small angle, we can write

$$\sum (P_j M_j{}^2) \approx \sum 8(y_i{}^2(1 - \theta^2 + \theta^4/3 + ...) + x_i{}^2(\theta^2 - \theta^4/3 + ...))$$

$$\sum (P_j M_j{}^2) \approx \sum \{8y_i{}^2 + 8(x_i{}^2 - y_i{}^2)\theta^2 + ...\}$$

This expression shows that $\sum (P_j M_j{}^2)$ reaches an extremum when the axis A tends to L (this extremum is a maximum if $\sum (x_i{}^2 - y_i{}^2)$ is positive, i.e. if L is the first principal axis). Since $\sum (P_i M_j{}^2)$ (over all i and j) is invariant and

$$\sum (P_i M_j{}^2) [1 \leq i \leq n, 1 \leq j \leq n] = \sum (P_j M_j{}^2) [1 \leq j \leq n] + \sum (P_i M_j{}^2) [1 \leq i \leq n, 1 \leq j \leq n, i \neq j]$$

then when A tends to L, $\sum (P_i M_j{}^2)$ $[1 \leq i \leq n, 1 \leq j \leq n, i \neq j]$ goes through an extremum, and this extremum is therefore one of those detected by the differentiation of $\sum (P_i M_j{}^2)$, as shown above.

As a consequence, in the general case where the set (M) has no particular symmetry, we can therefore take the principal planes as an approximation to symmetry planes.

REFERENCES AND NOTES

(1) Moreau, G.; Turpin, C. Use of similarity analysis to reduce large molecular libraries to smaller sets of representative molecules. *Analusis* **1996**, *24*, 17−21.

(2) Broto, P.; Moreau, G.; Vandycke, C. Molecular structures: Perception, autocorrelation descriptor and SAR studies. *Eur. J. Med. Chem.* **1984**, *19*, 66−70.

(3) Wagener, M.; Sadowski, J.; Gasteiger, J. Autocorrelation of molecular surface properties for modeling *corticosteroid binding globulin* and *cytosolic Ah receptor* activity by neural networks. *J. Am. Chem. Soc.* **1995**, *117*, 7769−7775.

(4) Legrand, M.; Rougier, M. J. In *Stereochemistry*; Kagan, H. B., Ed.; Georg Thieme Publishers: Stuttgart, Germany, 1977; Chaper 2, pp 103−123.

(5) Zabrodsky, H.; Avnir, D. Continuous symmetry measure. 4. Chirality. *J. Am. Chem. Soc.* **1995**, *117*, 462−473.

(6) Kuz'min, V. E.; Stel'makh, I. B.; Bekker, M. B.; Pozigun, D. V. Quantitative aspects of chirality. I. Method of dissymmetry function. *J. Phys. Org. Chem.* **1992**, *5*, 295−298.

(7) Rassat, A. A chiral systems classification based on the Haussdorf distance. *C. R. Acad. Sc., Ser.* **1984**, *299*, 53−55.

(8) Buda, A. B.; Auf der Heyde, T.; Mislow, K. On quantifying chirality. *Angew. Chem., Int. Ed. Engl.* **1992**, *31*, 989−1007.