J.American Chem. Soc., vol 117, p. 462-473, 1995

Continuous Symmetry Measures, IV: Chirality

Hagit Zabrodsky[‡], Shmuel Peleg[‡] and David Avnir[*]

Departments of Computer Science[‡] and Organic Chemistry[*]

The Hebrew University of Jerusalem

91904 Jerusalem, Israel

Abstract

We extend the treatment of symmetry as a continuous molecular structural property (*J. Am. Chem. Soc.* 1993, 115, 8278) to chirality. Rather than labeling objects as being either chiral or achiral, we provide an exact quantitative measure of this property, which allows one to distinguish (chiral) molecules from each other by their degree of shape-chirality. The continuous scale is based on the minimal distances that the vertices of a shape must move in order to attain the nearest achiral symmetry point group (in most cases, C_s, reflection symmetry). A detailed description of the methodology and the practical implementation of the Continuous Chirality Measure (CCM) are given. Its generality and versatility are then demonstrated on a wide variety of chirality related issues and in various chirality measurements. These include the identification of the most chiral objects (the most chiral ethane rotamer, the most chiral tetrahedron, etc.), the chirality evaluation of equicontour-representations of molecular orbitals, the calculation of the continuous changes in chirality along racemization pathways (including an all-chiral racemization pathway), the evaluation of chirality of structures with uncertain point locations, the extension of the CCM to diastereomerism (with a comment on prochirality and other stereochemical identifiers), the measurement of the chirality of various phosphates, of a fullerene, of helicenes, of knots, Möbius strips, catenanes and of a large random object (a diffusion limited aggregate), and the calculation of dynamic continuous changes in chirality during fluxional (Walden-type) inversion and in rotating ethane (with a comment on continuous chirality changes along concerted reaction pathways).

Thema: The vast majority of molecules are chiral, not achiral; to realize it, one only needs a sufficiently fine spatial or temporal resolution of measurement.

1 Background

In a recent citation analysis[1] the dominance of chirality as a central stream in modern chemistry seems more solid than ever. This paper is devoted to the quantitative evaluation of geometric chirality as a continuous property of molecular structure.

In previous parts[2–4] we advanced the notion of treating symmetry as a continuous rather than a descrete structural property. Our main argument has been that the static and dynamic structures of the ten million known molecules are so rich and diverse, that much is lost by allowing the assignment of a point-symmetry group to only a small fraction of these molecules and by trying to define correlations between symmetry and various molecular properties only in some strict limited cases. We have proposed that a more natural approach to symmetry issues would be to allow for gradual scaling of this structural property. We have developed this proposition into a working tool which allows one to evaluate quantitatively, on a continuous scale, how much of any symmetry element or symmetry group, exists in any configuration in any dimension. This tool also allows one to identify the symmetry which is nearest to the given configuration; and it allows one to obtain the nearest object with any desired symmetry - all these without reference to a specific ideal shape, but only to a specific symmetry.

The essence of our approach is the general definition of the Continuous Symmetry Measure (CSM) as

$$\mathbf{S}'(G) = \frac{1}{n}\sum_{i=1}^{n} \|P_i - \hat{P}_i\|^2 \qquad (1)$$

where G is a given symmetry group, P_i are the points of the original configuration, $\hat{P}_i$ are the corresponding points in the nearest G-symmetric configuration, and n is the total number of the configuration points. The meaning of Eq. 1 is: Find a set of points $\hat{P}_i$, which possesses the

desired symmetry (G-symmetry), such that the total (normalized) distance from the original shape P_i, is minimal. $\mathbf{S}'$ is bounded between 0 (the object has the desired symmetry) and 1. For convenience, the expanded scale is:

$$\mathbf{S} = 100 * \mathbf{S}' \tag{2}$$

Eq. (1) defines a metric on the space of all sets of n points, satisfying the requirements of being positive, commutative and fulfilling the triangle inequality. The main practical problem, then, is how to find the set of $\hat{P}_i$ which would lead to a minimal $\mathbf{S}(G)$ value. In Part I[2] we solved this problem for symmetry elements, then generalized it in Part II[3] to any symmetry group in any dimension, and extended the approach to contours (orbitals) and to uncertain structural data (e.g. x-ray data) in Part III[4]. The algorithm (the "folding-unfolding" method) is general and easy to implement. It is redescribed below in the context of the chirality problem treated in this report. We have provided rigorous mathematical proofs[3] that the method indeed provides the minimal solution for Eq (1).

A natural outcome of our general approach to symmetry is that $\mathbf{S}$ serves also as a continuous measure of chirality: Since chirality is defined as a lack of certain symmetries (the improper elements)[5], and since the CSM method allows one to evaluate how much of any of these symmetries is lacking in a given chiral configuration, one has to screen over all G_{achiral}'s to find the one that provides the minimal distance to achirality. For a given set of structures, the one with the largest $\mathbf{S}(G_{\text{achiral}})$ value is the most distant from having an improper symmetry element, and hence the most chiral; and vice-versa: as $\mathbf{S}(G_{\text{achiral}})$ approaches zero, the structure under study is minimally or negligibly chiral. In practice, since the minimal requirement for an object to be achiral is that it posseses, either a reflection mirror ($\sigma \equiv S_1$), an inversion center ($i \equiv S_2$) or higher order improper rotation axes S_{2n}, one has to screen $\mathbf{S}$ over the symmetry groups having these elements. In the majority of cases, one finds (below) that the continuous chirality measure is simply $\mathbf{S}(\sigma)$, i.e. the distance of a chiral object from having a reflection mirror. In Section 2 we show how to find the minimal $\mathbf{S}(G_{\text{achiral}})$.

It is in order to emphasize here that, at the moment, we are interested in chirality as a

geometric property, of either a collection of nuclear coordinates, or of equicontours of any molecular property; we return to this point in Section 3.2.

It is interesting to note that chirality as a special case, and not symmetry which is the general encompassing property, attracted most of the attempts to design a scale. This, we believe, is a manifestation of the central role of the former in the very phenomenon of life, and of the consequent major place that assymmetric synthesis[1] has occupied in chemistry over the years[6]. Next, then, we briefly list previous propositions for the quantitative evaluation of geometric chirality.

Perhaps the most succesful attempt has been the chain of papers by Kitaigorodskii[7], Gilat[8], Meyer and Richards[9], and Seri-Levi and Richards[10], which started with a raw idea by the first, and ended with correlations between chirality and chemical properties of real molecules, identified by the last. The idea here is that when left and right enantiomers are maximally overlapped, then the normalized non-overlapping volume is a measure of molecular dissimilarity, and hence also of chirality[7,8]. Gilat indicated the difficulty of performing this calculation [8], but then Meyer et al devised a simple algorithm for an optimal overlap[9], and good correlations between this measure and the pharmacological Pfeiffer rule[11] were shown for various drug molecules[10] (it should be noted that the overlapping procedure suggested in ref. **9,10** is optimal for the specific applications indicated there, but not necessarily maximal). The overlapping idea was adapted by several other groups, e.g. by Buda et al for the analysis of the chirality of triangles[12]. Also based on the overlapping concept, Kuzmin et al designed dissymmetry functions[13] for the evaluation of the difference between two enantiomers, using the tensor of inertia as a descriptor.

Another approach which gained much activity is that of "chirality functions" introduced by Kauzmann et al in 1961[14] and further developed by Ugi[15] by Ruch et al[16] and by King[17]. A basic idea here has been the attachment of chiral ligands to an achiral skeleton, assigning ligand-specific parameters for each such attachment. Derflinger reviewed this approach recently[18], detailing both its achievements and its difficulties. We note here that our approach allows both the analysis of sub-sets of vertices (i.e. various ligands) and the analysis

of the full configuration as a whole.

Mezey et al have contributed to the understanding of the problem of chirality measures by applying a number of different approaches: One is based on the principle of energy weighted fuzzy achirality resemblence[19], which was based on the Syntopy model of Mezey and Maurani[20]. The other applies the principle of resolution based similarity measures, tailored to mimic visual perception of this property[21].

Rassat introduced the evaluation of the smallest Haussdorf distance between chiral objects as a chirality measure[22] and his approach was applied by Buda et al for the tetrahedron[23]. An important source of discussion of chirality issues is Sokolov's book which became available to Western countries recently[6]. Of particular relevance here is his original algebraic analysis of chiral sets[6,24]. Other important contributions to the field of chirality measures are due to Chauvin[25] who introduced a pairing constant of equilibria between enantiomers as an index of topographical chirality for skeletal analogs with different ligands; to Walba[26] and Flapan[27] who introduced a hierarchical topological classification of chirality; to Harary and Mezey who introduced the concept of the degree of Jordan curves[21,28]; to Zimpel[29] who disscused topological vs metric descriptors of chirality (c.f. also Mezey's discussion of this topic[30]); and to Luzanov et al[31] who developed a quantitative measure of molecular dissymilarity, based on a quantum-mechanical approach.

Chirality measures were also developed in sub-atomic physics. Thus Donoghue et al used chiral Lagrangians to provide a measure of chirality of the strong atomic interactions as manifested in kaon decays[32].

Finally we mention earlier propositions of chirality scales made in our research group. One approach was based on a rotational dynamic property of chiral objects[33]: If an achiral object is rotated in a viscous medium, then the force exerted on the object upon clockwise rotation is exactly the same (except for sign) as the force exerted upon counterclockwise rotation. Yet, simple chiral objects will be rotated with greater ease in one direction than in the opposite direction. The difference between left and right rotations is then used as a measure of the degree of chirality. In another attempt we suggested to use the shape

distortion exerted in a reference molecule upon substitution[34]. Despite the fact that shape distortion can preserve achirality, (which is a weakness of that approach), it succeeded in a modefied form to correlate nicely a shape distortion parameter of chiral halogenated alkanes with their optical molecular rotations[35].

The approach and method we report on here, are different in many aspects from previously suggested chirality scales (including our own earlier studies), by offering the following advantages:

A) The chirality measure is an integral part of a most general method of measuring the symmetry content of any configuration in any dimension towards any symmetry group. Thus, a full profile of the symmetry properties of a molecule can be given, including its chirality.

B) The method, as will be seen below, is easily applied to virtually all sorts of known chiral structures: distorted tetrahedra, helicenes, fullerenes, frozen rotamers, knots, equiproperty contours, chiral reaction coordinates, and so on.

C) Chirality is measured without making reference to an ideal specific shape; the reference is only to the nearest σ or S_{2n}. Thus, the chirality of completely different structures can be compared.

D) The shape of the nearest achiral object is obtained, and the method is capable of selecting whether it is σ, i or any other S_{2n}.

E) The scale is well behaved from the point of view that its values can change continuously within the bounds of zero (achiral objects) and one.

2 Continuous Chirality Measure (CCM)

2.1 The General Definition and Approach

We define the Continuous Chirality Measure (CCM) as follows: Given a configuration of points $\{P_i\}_{i=1}^n$, its chirality content is determined by finding the nearest configuration of points $\{\hat{P}_i\}_{i=1}^n$ which has an improper element of symmetry, and by calculating the distance

between the two sets, using Eq. 1. The $\mathbf{S}'(G_{\text{achiral}})$ thus obtained is the minimal chirality measure of the given configuration, on a continuous scale of $0 \leq \mathbf{S}' \leq 1$ or $0 \leq \mathbf{S} \leq 100$.

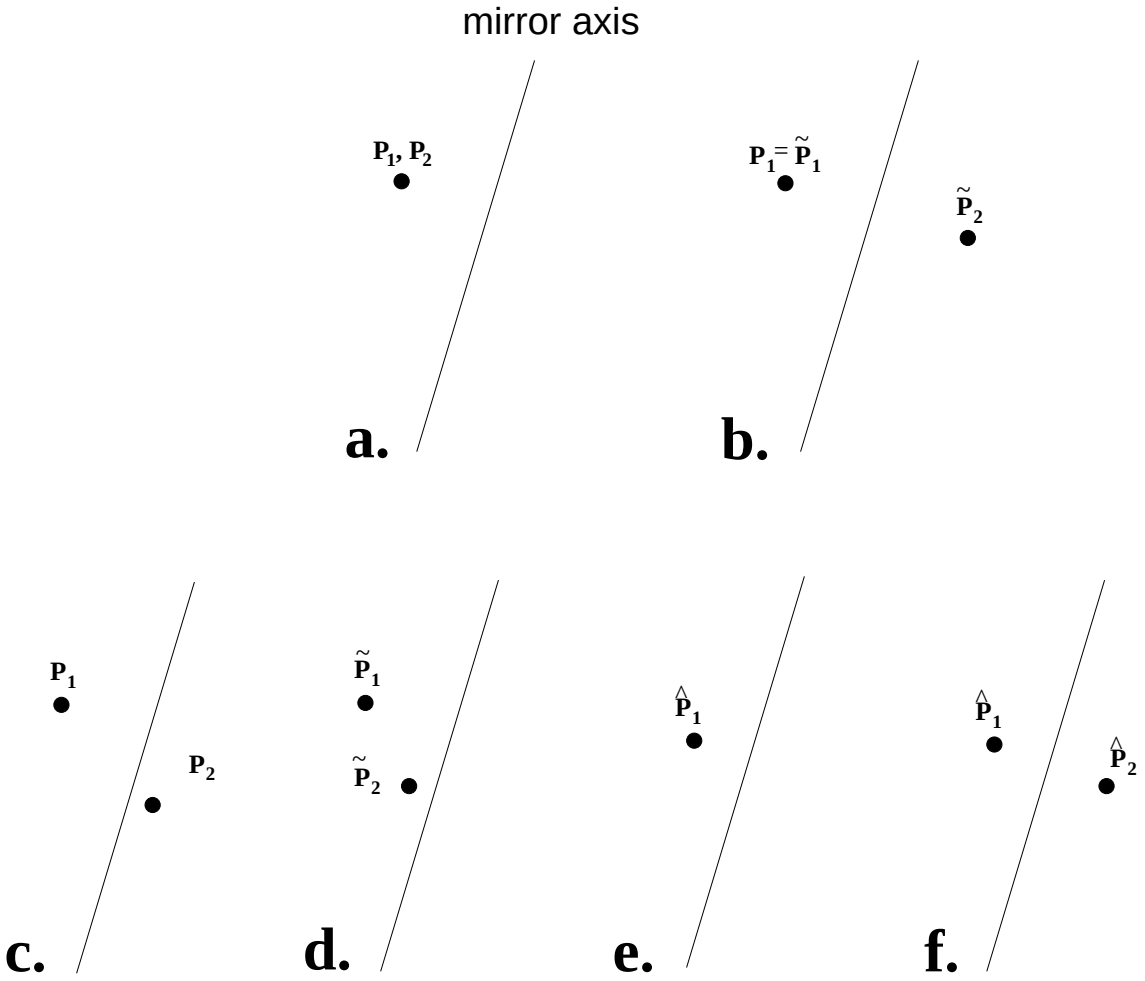

Figure 1: Unfolding and Folding of a pair of points. a) Given a single point, one treats it as a coinciding cluster of two points P_1 and P_2. b) Unfolding the pair of points by applying the identity transformation to P_1 and reflecting P_2 across the mirror plane, a mirror-symmetric pair of points, $\tilde{P}_1$ and $\tilde{P}_2$, is obtained. c) A non mirror-symmetric pair of points. d) Folding the pair of points shown in (c) results in a non-coinciding cluster of two points, $\tilde{P}_1, \tilde{P}_2$. e) The non-coinciding cluster is averaged to $\hat{P}_1$ and f) unfolded to a mirror-symmetric pair $\hat{P}_1, \hat{P}_2$

As mentioned in Section 1, following this general definition, one has to devise a tool for

locating the set of $\hat{P}_i$'s. Such a tool, termed the *Folding-Unfolding procedure*, was developed for the general case of continuous symmetry[3], and is based on the very method of constructing a symmetric object. This was described in great detail in ref. **3**; here we summarize it using the case of mirror-symmetry, σ:

Suppose we wish to construct a configuration which is symmetric with respect to the mirror-symmetry group $\{E, \sigma\}$ from a given point, P_1, and a given reflection axis σ, as shown in Fig. 1a. Unless the point is on the reflection axis, the minimal number of points needed to obtain a configuration having the required symmetry , is 2 (the number of elements in the symmetry group). Let us therefore treat the given point as a coinciding cluster of two points P_1 and P_2 (Fig. 1a). To obtain a σ-symmetric configuration we unfold the cluster by applying E on P_1 (being the identity element, E leaves P_1 in place, i.e., $\tilde{P}_1 = P_1$) and by applying σ on P_2 obtaining the reflected point $\tilde{P}_2$ (Fig. 1b). A mirror-symmetric configuration has been unfolded from the given point. The symmetric points can undergo a reversed procedure, and can be folded into a cluster of two coinciding points $\{P_1, P_2\}$. This is achieved by applying the inverse operation σ^{-1} on $\tilde{P}_2$ and E^{-1} on $\tilde{P}_1$. Notice that, whereas, folding of 2 mirror-symmetric points results in a coinciding pair of points, the folding of two non-symmetric points (Fig. 1c) results in a non-coinciding cluster where some distance exists between the two folded points (Fig. 1d). If the mirror axis is not pre-determined, then the minimization of this distance through the search of an optimal mirror-alignment is the key step in the evaluation of the minimal $\mathbf{S}(\sigma)$. Once this minimum is found, the coordinates of the folded points are averaged obtaining the coordinates of a single average point $\hat{P}_1$ (Fig. 1e) and the average point is then unfolded into a σ-symmetric configuration (Fig. 1f).

Most objects of interest have, however, more than two points and since in a σ-symmetric object, each point on one side of the mirror axis or plane has a counterpart on the opposite side, the other essential step in the process is to divide the points into pairs, each of which is to be symmetrized around a chosen mirror-axis or mirror plane. For instance, the four non-symmetric points in Fig. 2a can be divided into pairs such as: $\{P_1, P_2\}, \{P_3, P_4\}$ or $\{P_1, P_3\}, \{P_2, P_4\}$ etc. It is possible, however, for a point in a σ-symmetric object to have no

counterpoint, whenever that point lies on the mirror-axis. Therefore, the division of points must allow for sets of pairs and sets of single points. For instance, the four points in Fig. 2a can also be divided into sets: $\{P_1, P_3\}, \{P_2\}, \{P_4\}$. In the symmetrized object, $\hat{P}_2$ and $\hat{P}_4$ must fall on the mirror-axis, whereas $\hat{P}_1$ and $\hat{P}_3$ will be reflections of each other. In this stage of the procedure, all possible divisions into sets must be found. This is a topological problem, the solution of which is detailed in Section 2.3.

Having explained the two key steps - the topological step and the folding/unfolding, we proceed next to show how these steps are practiced in the full procedure.

2.2 The CCM of a Set of Points in 2D with Respect to Reflection

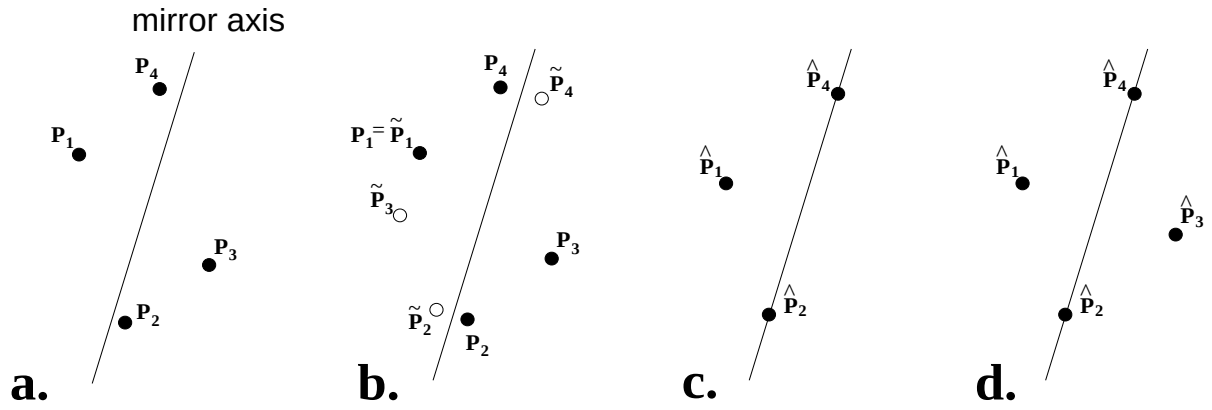

Figure 2: Obtaining the closest mirror symmetric set of points using the folding/unfolding method. a) The original configuration of points $P_1, \ldots, P_4$. The points are divided into sets: $\{P_1, P_3\}, \{P_2, P_2\}$ and $\{P_4, P_4\}$. b) Each pair of points is folded by applying the identity transformation to one point and by reflecting the other point across the mirror plane. The folded points $\{\tilde{P}_i\}$ are obtained. c) Each pair of folded points $\{\tilde{P}_i, \tilde{P}_j\}$ is averaged, obtaining a single averaged point $\hat{P}_i$. d) Each average point $\hat{P}_i$ is unfolded by reflecting back across the mirror plane, obtaining the point $\hat{P}_j$. The points $\{\hat{P}_i\}_{i=1}^4$ are mirror symmetric.

We demonstrate the evaluation of the chirality content on a set of points (Fig. 2a) in 2D, with respect to reflection. The following steps are carried out:

1. **Normalization of the non-symmetric configuration**: a) Determine the centroid of the configuration of points. This is done by averaging the coordinates of the set of points. b) Translate the object so that its centroid coincides with the origin. c) Scale the configuration so that the maximum distance between the centroid and the farthest point is one[36].

2. **Select a symmetry group** and translate it so that all its operations pass through the origin. In our example the relevant groups contain only two elements: the identity (E) and σ. Thus, the reflection plane (line in 2D) should pass through the origin.

3. **Select a reflection plane** or line (passing through the origin) from amongst all possible alignments of this element.

4. **Select a division** of the points into sets of pairs and single points (the topological step - Section 2.3). If a set contains one point, duplicate that point. One possible division, used in the example of Fig 2 is $\{P_1, P_3\}, \{P_2, P_2\}$ and $\{P_4, P_4\}$.

5. **Fold** each set of points $\{P_i, P_j\}$ by applying the identity transformation to one point P_i and by reflecting the other point P_j across the mirror plane. The folded points $\{\tilde{P}_i, \tilde{P}_j\}$ are obtained (Fig. 2b). Applying E and σ on $\{P_2, P_2\}$ results in an unchanged P_2 and in a reflected $\tilde{P}_2$. $\{P_4, P_4\}$ is folded similarly. The pair $\{P_1, P_3\}$ is folded by applying E on P_1 and σ on P_3: P_1 remains in place ($P_1 = \tilde{P}_1$) and a reflected $\tilde{P}_3$ is obtained. (The order of operations in this case is not important. This however, is not the case for other symmetry groups, such as S_{2n} having more than 2 elements[3]).

6. **Average** each pair of folded points $\{\tilde{P}_i, \tilde{P}_j\}$ obtaining a single averaged point $\hat{P}_i$ for each pair (Fig. 2c). In our example, averaging the pair $\{P_2, \tilde{P}_2\}$, point $\hat{P}_2$ is obtained and averaging $\{P_4, \tilde{P}_4\}$, point $\hat{P}_4$ is obtained. Note that both these averaged points must lie on the reflection line σ, by definition. The pair $\{\tilde{P}_1, \tilde{P}_3\}$ averages to $\hat{P}_1$.

7. **Unfold** each averaged point $\hat{P}_i$ by reflecting back accross the mirror plane, obtaining the point $\hat{P}_j$ (Fig. 2d). If the original set $\{P_i, P_j\}$ consists of a single duplicated point, then the two unfolded points $\hat{P}_i$ and $\hat{P}_j$ are at the same location and are considered as a single point $\hat{P}_i$ (Fig. 2d). The points $\{\hat{P}_i\}_{i=1}^n$ are mirror symmetric.

8. **Calculate S**(G_σ) according to Eq.s (1) and (2).

9. **Minimize** the chirality value obtained in Step 8 by repeating Steps 3-7 with all possible division of points into sets and for all possible reflection planes. In practice, the minimization is greatly simplified: in 2D the optimal axis of reflection is found analytically (Appendix A.3 in ref. 3), and in 3D we use a closed form solution which replaces steps 5-7 and is detailed in Section 2.4. The division of points into sets is also greatly simplified when the configuration of points is connected (as is usually the case in chemistry. See Section 2.3).

Rigorous mathematical proof that the procedure outlined here, indeed provides the minimal **S** value, was detailed in ref. **3**. The procedure outlined here is applicable to 3D as well (Sections 3-4). The procedure for symmetry groups having improper axes of rotation (including inversion) is similar and is outlined in Section 2.5.

2.3 Further Comments on the Division of Points into Sets - The Topological Stage

As described above, this stage corresponds to dividing the points in the given configuration into sets, so that for every possible division into sets one finds the closest achiral configuration. Although the coordinates of the points in the set change upon symmetrization, we impose that all other features and characteristics associated with the points (connectivity, mass, atomic number, etc) remain invariant under this transformation. Therefore, the connectivity of the points in the original configuration, namely, the *topology* of the configuration,

determines the division of points into sets. We concentrate in this Section on connectivity and comment on other physical features in Section 3.2.

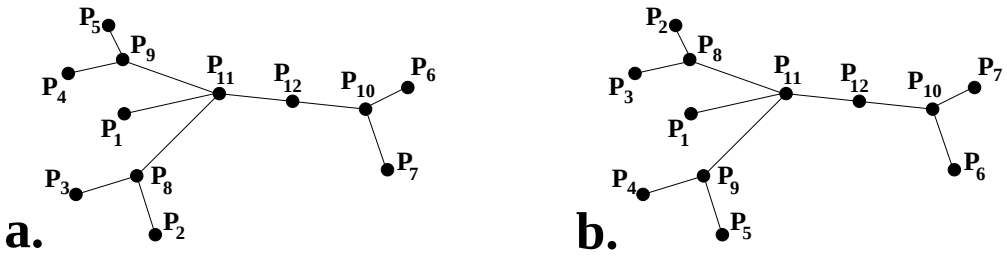

Figure 3: Connected configurations of points. The graph shown in a) is isomorphic to the graph shown in b) (see text).

As an example, let us analyse the "2D branched alkane" skeleton shown in Fig. 3a[26,33]. Points $P_1, \ldots, P_7$ are leaf nodes and can be paired between them. Points $P_8, \ldots, P_{10}$ have the same valency (the number of edges converging at a point) of three and can be paired. Points P_{11} and P_{12} stand alone in their valency of 4 and 2 respectively, and will form single-point pairs (degenerate pairs). Thus a possible division of the points into pairs for measuring mirror-symmetry and for transforming the configuration into a mirror-symmetric configuration, is: $\{P_2, P_5\}, \{P_3, P_4\}, \{P_8, P_9\}, \{P_1\}, \{P_{11}\}, \{P_{10}\}, \{P_6, P_7\}, \{P_{12}\}$. However the valency of a point is insufficient for determining the division into sets. Consider for example points P_8 and P_{10} which have the same valency (3) but obviously cannot be geometrically moved to be mirror-symmetric, because they are not equivalent in their second order connectivity (i.e. in the valency of their neighboring points). Thus point P_8 has two neighbors of valency 1 and one neighbor of valency 4, whereas point P_{10} has two neighbors of valency 1 and one neighbor of valency 2. This reasoning does not stop at the second order connectivity (in Fig. 3a points P_4 and P_6 do not agree in their third order connectivity) but must be taken to the maximal connectivity of the configuration (which is equal to the width of the graph).

Thus, the topological stage of evaluating the chirality measure of a connected configuration of points, lists all possible division of the points into pairs, by taking into account only the topology (the connectivity of the points). When considering a configuration of points as a graph[37], the problem of dividing the points into pairs (proper and degenerate pairs)

reduces to the classical question of listing all graph isomorphisms of order $2^{\mathbf{38}}$. A graph isomorphism is a permutation Π of the graph vertices which leaves the graph topologically equivalent; i.e. given a graph composed of the set of vertices - V and the set of edges - E, $\mathcal{G} = \{V, E\}$, replacing each vertex $i \in V$ with its permutated vertex $\Pi(i)$ results in a graph $\mathcal{G}' = \{V', E'\}$ such that the set of edges E' equals E. Note that if Π is an isomorphism of $\mathcal{G}$ then if $(i, j) \in E$ also $(\Pi(i), \Pi(j)) \in E$. A graph isomorphism of order 2 is an isomorphism where $\Pi(\Pi(i)) = i$ (i.e. either $\Pi(i) = i$, or, $\Pi(i) = j$ and $\Pi(j) = i$). For example, the 12 isomorphisms of order two of the graph $\mathcal{G}$ shown in Figure 3a are listed in Table 2.3. Thus

Table 1: All possible isomorphic configurations of order two of the branched structure shown in Fig. 3a.

	$\Pi(P_1)$	$\Pi(P_2)$	$\Pi(P_3)$	$\Pi(P_4)$	$\Pi(P_5)$	$\Pi(P_6)$	$\Pi(P_7)$	$\Pi(P_8)$	$\Pi(P_9)$	$\Pi(P_{10})$	$\Pi(P_{11})$	$\Pi(P_{12})$
a	P_1	P_2	P_3	P_4	P_5	P_6	P_7	P_8	P_9	P_{10}	P_{11}	P_{12}
b	P_1	P_2	P_3	P_4	P_5	P_7	P_6	P_8	P_9	P_{10}	P_{11}	P_{12}
c	P_1	P_2	P_3	P_5	P_4	P_6	P_7	P_8	P_9	P_{10}	P_{11}	P_{12}
d	P_1	P_2	P_3	P_5	P_4	P_7	P_6	P_8	P_9	P_{10}	P_{11}	P_{12}
e	P_1	P_3	P_2	P_4	P_5	P_6	P_7	P_8	P_9	P_{10}	P_{11}	P_{12}
f	P_1	P_3	P_2	P_4	P_5	P_7	P_6	P_8	P_9	P_{10}	P_{11}	P_{12}
g	P_1	P_3	P_2	P_5	P_4	P_6	P_7	P_8	P_9	P_{10}	P_{11}	P_{12}
h	P_1	P_3	P_2	P_5	P_4	P_7	P_6	P_8	P_9	P_{10}	P_{11}	P_{12}
i	P_1	P_4	P_5	P_2	P_3	P_6	P_7	P_9	P_8	P_{10}	P_{11}	P_{12}
j	P_1	P_4	P_5	P_2	P_3	P_7	P_6	P_9	P_8	P_{10}	P_{11}	P_{12}
k	P_1	P_5	P_4	P_3	P_2	P_6	P_7	P_9	P_8	P_{10}	P_{11}	P_{12}
l	P_1	P_5	P_4	P_3	P_2	P_7	P_6	P_9	P_8	P_{10}	P_{11}	P_{12}

replacing every point with its permuted point of isomorphism (l), for example, we obtain the graph $\mathcal{G}'$ of Figure 3b which is topologically equivalent to graph $\mathcal{G}$.

Graph isomorphisms is a widely studied area (see ref. **37-38** for a review) and has many theoretical results. There are several ways to find all graph isomorphisms of order two[38], but we use a simple recursive algorithm developed by us[39].

2.4 Further Comments on the Folding/Unfolding Procedure

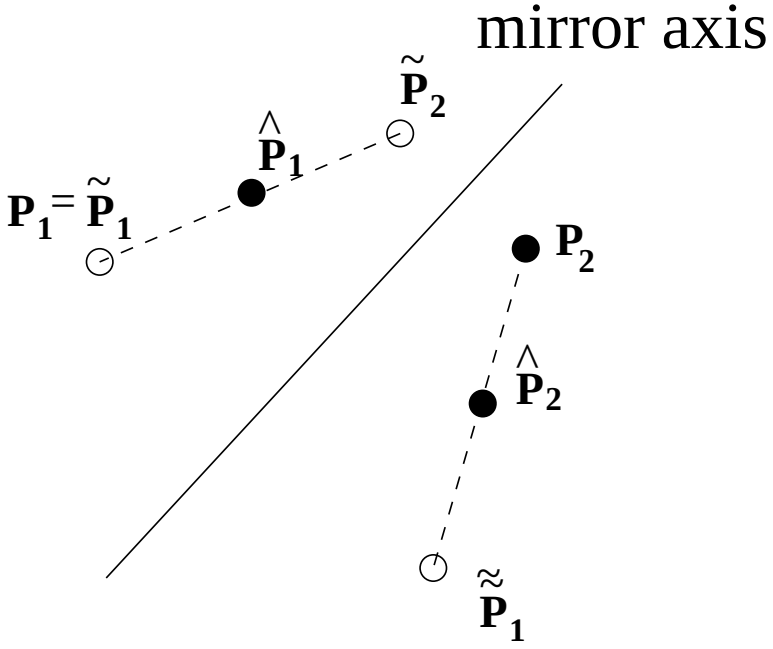

Figure 4: The conversion of $\|P_1 - \hat{P}_1\|^2 + \|P_2 - \hat{P}_2\|^2$ to $\frac{1}{2}\|P_1 - \tilde{P}_2\|^2$ or to $\frac{1}{2}\|\tilde{\hat{P}}_1 - P_2\|^2$. See text (Section 2.4).

We recall that the general definition of **S** (Eq. 1), requires minimization of the $P_i - \hat{P}_i$ distances, and that according to our procedure, the $\hat{P}_i$'s are obtained at the unfolding stage (Section 2.1, Fig. 1c-f). However, for the case of $\mathbf{S}(\sigma)$, the problem can be reformulated so that only the folded points are considered. Let us demonstrate it on the pair of points P_1, P_2 in Fig. 4. We show that $\|P_1 - \hat{P}_1\|^2 + \|P_2 - \hat{P}_2\|^2$, can be expressed in terms of the folded points $\tilde{P}_1, \tilde{P}_2$ (Fig. 4): Since $P_1 = \tilde{P}_1$, the term $P_1 - \hat{P}_1$ can be replaced by $\tilde{P}_1 - \hat{P}_1$; and since $\tilde{P}_2$ is a reflection of P_2 and $\hat{P}_2$ is a reflection of $\hat{P}_1$, the term $\|P_2 - \hat{P}_2\|$ can be replaced by $\|\tilde{P}_2 - \hat{P}_1\|$, obtaining $\|\tilde{P}_1 - \hat{P}_1\|^2 + \|\tilde{P}_2 - \hat{P}_1\|^2$. Next, since $\hat{P}_1$ is the center point (average)

of $\tilde{P}_1, \tilde{P}_2$, we have: $\|\tilde{P}_1 - \hat{P}_1\| = \|\tilde{P}_2 - \hat{P}_1\| = \frac{1}{2}\|\tilde{P}_1 - \tilde{P}_2\|$. Therefore

$$\|P_1 - \hat{P}_1\|^2 + \|P_2 - \hat{P}_2\|^2 = 2(\frac{1}{2}\|\tilde{P}_1 - \tilde{P}_2\|)^2 = \frac{1}{2}\|\tilde{P}_1 - \tilde{P}_2\|^2 = \frac{1}{2}\|P_1 - \tilde{P}_2\|^2 \qquad (3)$$

Denoting by $\tilde{\tilde{P}}_1$, the reflection of P_1 and noting that P_2 is the reflection of $\tilde{P}_2$, we have that the last term of Eq. 3 is equal to: $\frac{1}{2}\|P_1 - \tilde{P}_2\|^2 = \frac{1}{2}\|\tilde{\tilde{P}}_1 - P_2\|^2$

Thus, in order to minimize the distance between P_1 and $\hat{P}_1$ and between P_2 and $\hat{P}_2$ (over all orientations of σ), one may minimize $\frac{1}{2}\|P_1 - \tilde{P}_2\|^2$ or minimize $\frac{1}{2}\|\tilde{\tilde{P}}_1 - P_2\|^2$. Equivqlently, one can minimize

$$\|\tilde{\tilde{P}}_1 - P_2\|^2 + \|\tilde{\tilde{P}}_1 - P_2\|^2 \qquad (4)$$

over all orientations of σ. The meaning of Eq. 4 is that <u>all</u> points are reflected, and the sum of distances between all reflected points and their matching unreflected points is minimized. Suppose, for instance, that the pairing of the four points in Fig. 5a is $\{P_1, P_1\}, \{P_2, P_4\}, \{P_3, P_3\}$. The points are reflected into $\tilde{P}_i$ (Fig. 5b), and the sum $\|P_1 - \tilde{P}_1\|^2 + \|P_2 - \tilde{P}_4\|^2 + \|P_3 - \tilde{P}_3\|^2 + \|P_4 - \tilde{P}_2\|^2$ is minimized over all reflections and rotations of the set $\tilde{P}_i$. the best arrangements' is shown in Fig. 5c.

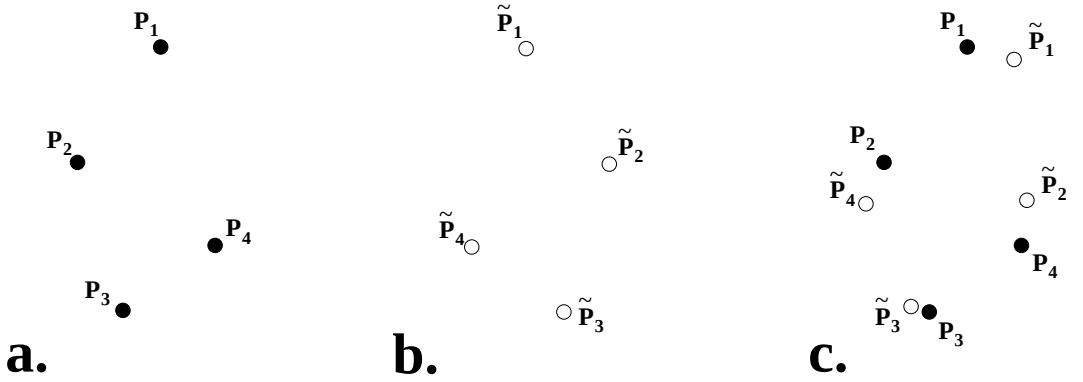

Figure 5: a) The pairing of the four points is, for instance, $\{P_1, P_1\}, \{P_2, P_4\}, \{P_3, P_3\}$. b) The reflected points $\tilde{P}_i$. c) The minimal distance between the original and reflected set of points.

Given two sets of points and given a matching between points of the two sets, the problem of finding the optimal rotation and translation which minimizes the sum of squared distances

between the corresponding points is a classic problem of pose estimation[58]. Several methods have been suggested to solve this problem analytically[59]. We follow the method of Arun et al[59a] which is summarized in the Appendix.

2.5 Improper Axes of Rotation

The above described procedure for $S_1 \equiv \sigma$ can be straightforwardly extended to find the closest S_{2n}-symmetric configuration for any n. Here the points are divided into sets having either $2n$ ordered points, 2 points, or having a single point (in the last two cases the points will lie in the symmetrized object, on the improper rotation axis or at its intersection with the plane, respectively). The folding and unfolding are performed by applying rotation-reflection rather than reflection.

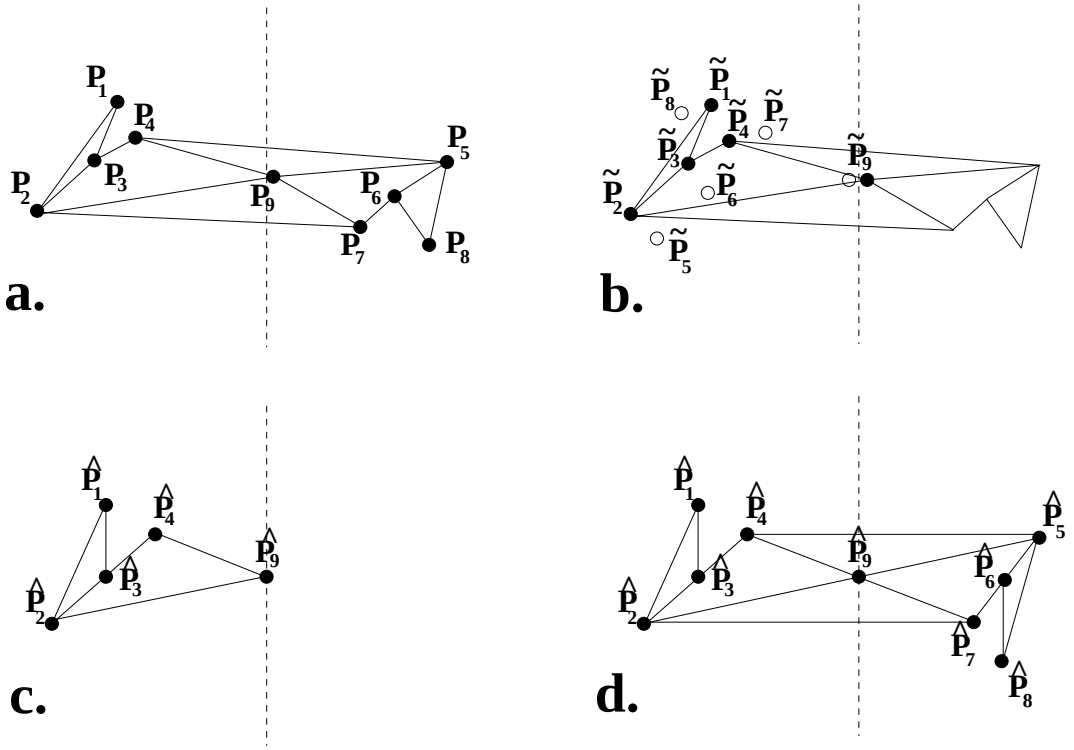

Figure 6: The steps of determining the continuous chirality measure with respect to an improper axis of rotation ($S_2 \equiv i$, in this case). See Section 2.5 for details.

The procedure for S_{2n} is thus as follows (Fig. 6, demonstrated for S_2):

1. **Normalize** the configuration, as in Step 1 in Section 2.2.

2. **Select** an improper rotation axis, passing through the origin (Fig. 6a).

3. **Select a division** of the points into ordered sets (a permutation of the points) where each set contains either $2n$ points, 2 points or a single point (in our example the first two cases are equivalent). If a set contains a single point, that point is multiplied $2n$ times. If the set contains 2 ponits, each of the two points is multiplied n times. For example in Figure 6a a possible division with respect to S_2 is: $\{P_1, P_8\}, \{P_2, P_5\}, \{P_3, P_6\}, \{P_4, P_7\}, \{P_9\}$.

4. **Fold** each of the sets of points by applying an element from the ordered set of elements, $E, S_{2n}, S_{2n}^2 \ldots$, to each point in the ordered set. Except for the case $\{E, S_2\}$, an ordering of elements must be selected here. For the example of Fig. 6b, the identity transformation is applied to points P_1, P_2, P_3, P_4, P_9, and a π/n rotation-inversion is applied to points P_8, P_5, P_6, P_7, P_9. The folded points $\{\check{P}_i\}_{i=1}^{2n}$ are obtained (Fig. 6b).

5. **Average** each set of folded points $\{\check{P}_i\}_{i=1}^{2n}$ obtaining a single averaged point $\hat{P}_1$ for each set (Fig. 6c).

6. **Unfold** each average point $\hat{P}_1$ by applying $E^{-1}, S_{2n}^{-1} \ldots$ as in step 4, on the averaged point. The unfolded points obtained are $\{\hat{P}_i\}_{i=1}^{2n}$. Thus, $\hat{P}_1$ is the identity applied to $\hat{P}_1$, point $\hat{P}_8$ is obtained by applying a π/n backward rotation-inversion to point $\hat{P}_1$, point $\hat{P}_5$ is obtained by applying a backward proper-rotation of $2\pi/n$ to point $\hat{P}_2$ and so on (Fig. 6d). The points $\{\hat{P}_i\}_{i=1}^{2n}$ are S_{2n}-symmetric.

7. **Calculate S**(S_{2n}) according to Eq.s 1 and 2.

8. Minimize the chirality value obtained in Step 7 by repeating Steps 2-6 with all possible division of points into sets and for all possible improper-rotation axes.

As in the mirror-symmetry case, the division of points into sets is greatly simplified when the configuration of points is connected (or partially connected). In the case of S_2, the topological stage as described in Section 2.3 is applicable, since both the mirror-symmetry group and the inversion-symmetry group have two elements and the possible divisions into sets reduces to finding isomorphisms of order two (see Section 2.3). In all other cases of S_{2n}-symmetry, the topological stage as described in Section 2.3 is applicable with slight modifications: the topological stage finds all isomorphisms of order $2n$ of the given graph. This is performed by relieving the restriction that $match(i) = j \iff match(j) = i$ (which is equivalent to the restriction: $match(match(i)) = i$).

3 Further Properties of the Continuous Chirality Measure

3.1 Maximal Chirality Values and the Most Chiral Objects

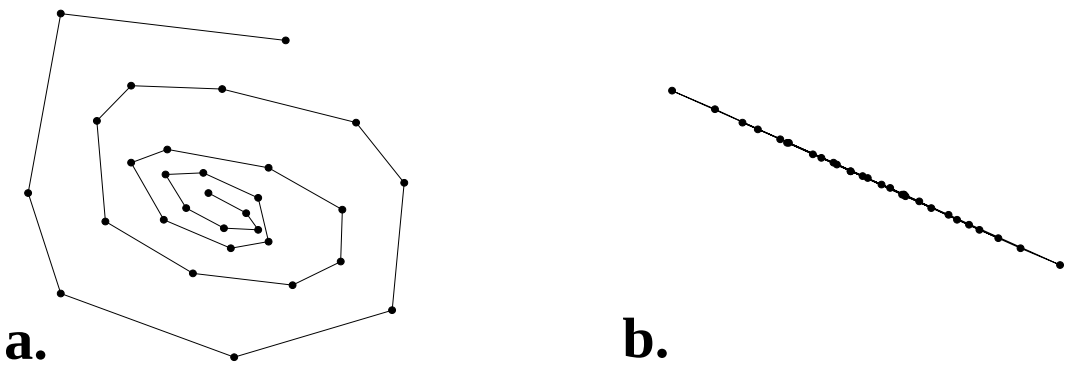

Figure 7: The achiral object nearest to the spiral (a) coincides with the nearest mirror axis (b). The CCM value for this spiral is 8.83.

The upper bound of $\mathbf{S}'$ (Eq. 1), namely 1.0, is attained in cases where the nearest symmetric object requires all of the P_i vertices to move the maximal distance of 1 towards $\hat{P}_i$ (recall the normalization step, Section 2.2). This condition is fulfilled, for instance, if one asks how much C_7-ness exists in a perfect hexagon (not allowing any addition of vertices): Since the

object nearest to a hexagon and having C_7-symmetry, is a single point located at the centroid of the hexagon, one obtaines for the hexagon $\mathbf{S}(C_7) = 1.0$ (for an additional example see Fig. 18 in ref. **2**).

By the same token, if one imposes the determination of the chirality of a perfect hexagon with respect to S_8 symmetry rather than with respect to the obvious $S_1 \equiv \sigma$ element, then again, the nearest S_8-achiral object is the centroid with $\mathbf{S}(S_8) = 1.0$. In the majority of cases, however, the nearest achiral object possesses a σ element, and with such objects the maximal value of 1 is not reached: The maximal distance moved in this case by the set of P_i's is not the collapse to the centroid but the distance to $\hat{P}_i$'s located on the reflection line(2D) or plane (3D). Thus, since $\sum_i \|P_i - \hat{P}_i(\text{centroid})\|^2 < \sum_i \|P_i - \hat{P}_i(\text{plane})\|^2$ one has that $\mathbf{S}'(\sigma) < 1$. Fig. 7 shows one such case where the nearest achiral object is the original configuration colapsed to a reflection line (the CCM value is 8.83).

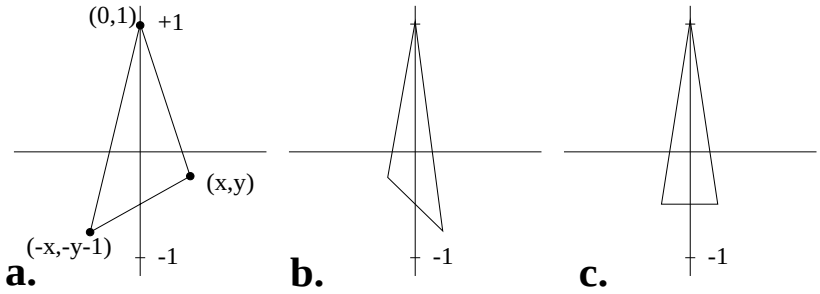

Figure 8: By parametrizing the space of all triangles in 2D (a), the most chiral triangle (b) and the nearest achiral triangle (c) are found.

Next we evaluate the maximal chirality value of the classical Pythagorean *tetraktys*[40], namely, one point, 2 points - a line, 3 points - a triangle and 4 points - a tetrahedron. The chirality of the first two cases is zero for obvious reasons (and we note that this value is obtained by following the protocol of our procedure as well). For the evaluation of the maximal chirality of the triangle we employ a search program as follows: We denote the three vertices as $(0.0, 1.0), (x, y)$ and $(-x, -1.0 - y)$. The sum of coordinates is zero, and thus the centroid is at the origin $(0.0, 0.0)$ (Fig. 8a). Following Step 1 in Section 2.2, the values of x,y are taken so that no vertex is outside the unit circle. A rigorous search is

performed by densely sampling the x and y values in the range $[-1 \ldots 1]$ and verifying that no vertex is outside the unit circle. The CCM is evaluated for every sampling and those values of x and y which maximize the CCM value, are found. We find that the CCM is maximized when $x = 0.220183$ and $y = -0.719058$ representing a triangle with vertex coordinates $(0.0, 1.0), (0.220183, -0.719058)$ and $(-0.220183, -0.280942)$. Thus, the most chiral triangle, a scalene, has an edge (angle) ratio of $1 : 0.75 : 0.36$ and its CCM value is $\mathbf{S}'_{\max}(\sigma) = 0.030259$; it is shown in Fig 8b. Fig 8c shows the closest achiral triangle to the most chiral one (an isosceles triangle). In a previous report[2] we evaluated the $\mathbf{S}'_{\max}(C_3)$ of a triangle and obtained 1/3. The result $\mathbf{S}'_{\max}(\sigma) < S'_{\max}(C_3)$ is expected: it reflects the situation that in order to attain a C_3 configuration, one has to search for near specific points, whereas in order to attain achirality one has to search for the shortest distance to a reflection line; hence also the rather small value of $\mathbf{S}'_{\max}(\sigma)$.

The search for the most chiral tetrahedron (3D simplex) is carried out similarly: We denote the vertices as $(0.0, 0.0, 1.0), (x_1, 0.0, z_1), (x_2, y_2, z_2)$ and $(-x_1 - x_2, -y_2, -1.0 - z_1 - z_2)$, which places the centroid at the origin $(0.0, 0.0, 0.0)$ (Fig. 9a). As in the triangle case, a rigorous search is performed by densely sampling the range $[-1 \ldots 1]$ for the values of x_1, z_1, x_2, y_2 and z_2. For every set of parameter values the points of the tetrahedron are verified to be in the unit sphere and the CCM value is evaluated. We find that the CCM is maximized when $x_1 = 0.394532$, $z_1 = 0.252185$, $x_2 = -0.136945$, $y_2 = -0.298958$ and $z_2 = -0.333343$ representing a tetrahedron with vertex coordinates $(0.0, 0.0, 1.0), (0.395, 0.0, 0.252), (-0.137, 0.299, -0.333)$ and $(-0.258, -0.299, -0.919)$. Thus, the most chiral tetrahedron has an edge ratio of $1 : 1 : 1 : 1.6 : 1.6 : 2.3$ and for each of its 4 constructing triangles the ratios are: $1 : 1 : 1.6$, $1 : 1 : 1.6$, $1 : 1.6 : 2.3$ and $1 : 1.6 : 2.3$. Its CCM value is $\mathbf{S}'_{\max}(\sigma) = 0.040544$, and it is shown in Fig 9b. Very interestingly, the most chiral tetrahedron is, perfectly C_2 symmetric (The C_2 axis bisects the edges 14 and 23). Unlike the case of the triangle, the closest achiral structure is colapsed to a plane (Fig. 9c).

For the maximal chirality of ethane rotamers, see Section 4.2 (c.f. ref. **13c**). See also Gilat and Gordon who recently obtained the theoretical upper bounds for their chiral coefficients

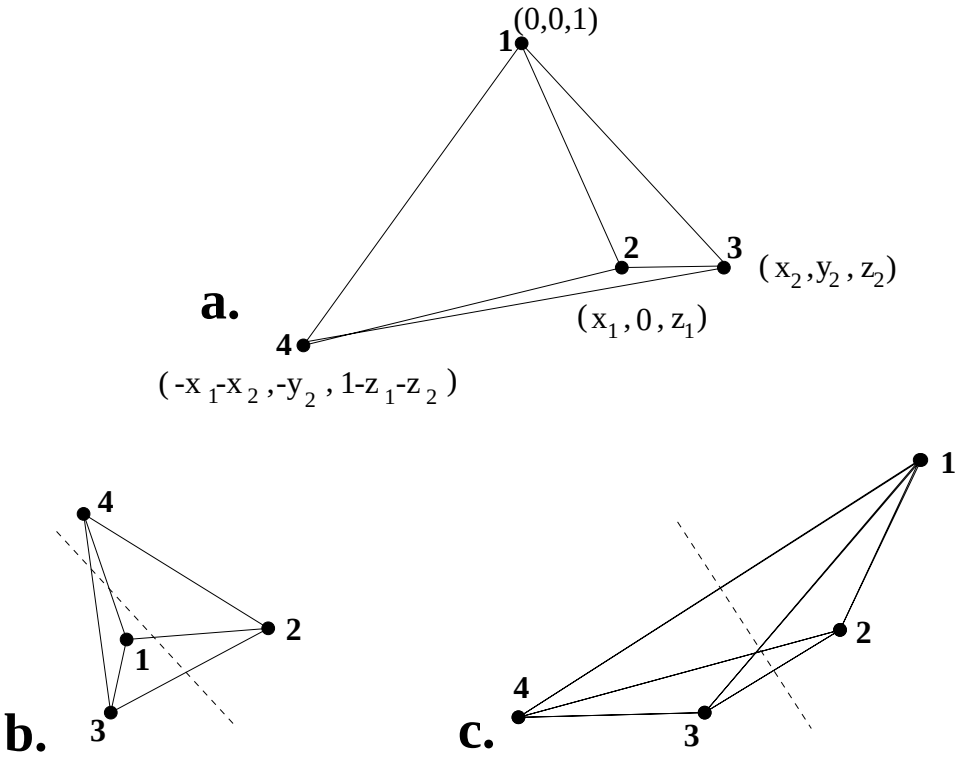

Figure 9: By parametrizing the space of all tetrahedra (a), the most chiral tetrahedra is found. b) The most chiral tetrahedron is a C_2 simplex (the C_2-axis is the dashed line). c) The closest achiral configuration is the tetrahedron colapsed to a plane (shown as coinciding with the plane of this page).

of convex sets[41]: 0.3954 and 0.6977 for 2D and 3D respectively. Buda's result for maximal chirality of a triangle[12a] is paradoxical[12b]. See also ref **23** for an application of Rassat's approach[22] to the question of maximal tetrahedral chirality.

3.2 Some Comments on Chirality Assesment of Physical Properties: The Chirality of Equiproperty Contours

It is in order to re-iterate here that the CCM analysis as described so far has delt purely with shape[42]. Thus, the most chiral tetrahedron obtained above, refers to vertex coordinates only, while this is directly applicable to, say, vibrational distortions of CX_4, the question

arises as to how can one approach chirality issues of molecules, for which chirality is initially linked with different atoms $(CWXYZ)$[13]?

We suggest that since virtually all chemical properties and many of the molecular physical properties are determined by the (frontier) orbitals and by the ensuing molecular charge distributaions, equi-property contours unify the representation of heteronuclear molecules into a homogeneous continuous representation, on which the CCM can then be applied. Thus, although F and Cl cannot be symmetrized, charge density distributions induced by these atoms, can. To implement this solution one has, therefore, to extend the CCM analysis to continuous surfaces or contour lines[4,21]. This is performed by representing the contour as a string of equally spaced points (as dense as one wishes) and then performing the CCM folding-unfolding procedure on the dense polygon as described in Section 2.2. As a preliminary example of how this is done, we evaluate the chirality of the contours of the lone-pair orbital of a distorted water molecule (perhaps a frozen moment of a vibration, or a water molecule in a matrix of amorphous ice, or a water molecule trapped in a micropore) as shown in Figure 10. The ratio of lengths of the two O-H bonds is 0.9 (instead of 1.0) and the H-O-H angle is 104°.[43]. Each of the two shown contours is represented as a string of about 200 points and the CCM with respect to mirror symmetry evaluated. It is seen quantitatively (Fig. 10) that the distortive effects of the unequal bonds, fades away from the inner to the outer contours.

It is also in order to recall here that our method evaluates the CCM by identifying the minimal distances, regardless of whether motion along the shortest-distance line is possible physically. $\mathbf{S}(G)$ values which are <u>not</u> minimal, but which correspond to physical pathways of symmetrization, are of great interest by themselves. We leave for the moment further extensions of these comments and return to the main theme of this report, namely pure shape chirality of collections of connected vertices.

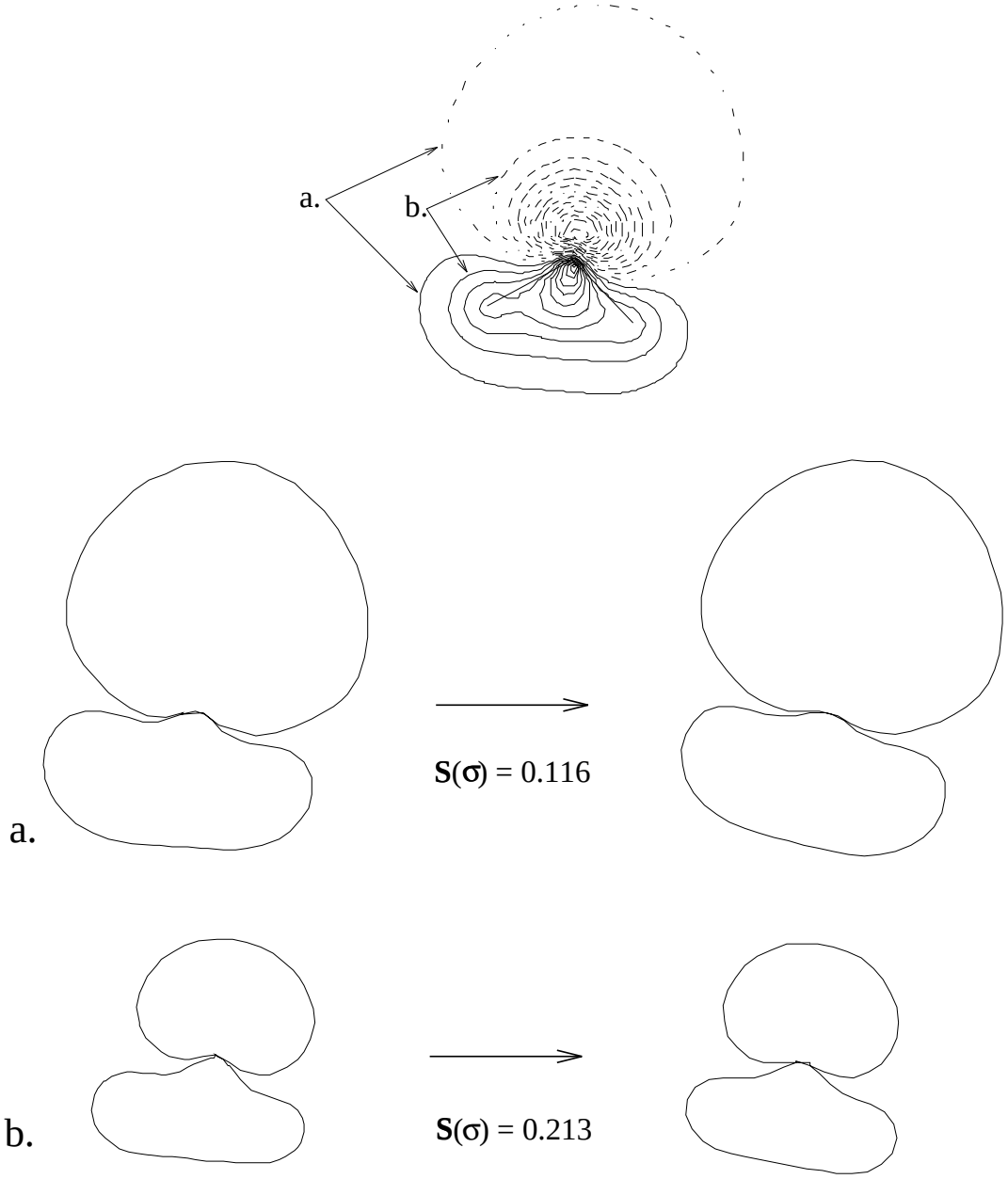

Figure 10: Two equi-amplitude contours of the wave function of the lone-pair orbital of distorted water molecule are shown[43]. The two contours are spaced by 0.05 Bohr$^{-3/2}$, and the value of the outer one is 0.576 Bohr$^{-3/2}$. S(σ) values are indicated in the figure. The CCM value for the next inner contour (not shown) is 0.248.

3.3 The Continuous Change in Chirality along Racemization Pathways

Given a pair of enantiomers, one can racemize from one to the other by various routes. Standard intuition would, perhaps, dictate that at a certain point along the racemization pathway,

an achiral intermediate should be encountered, namely a structure with $\mathbf{S}(G_{\text{achiral}})=0$, where left changes to right. Remarkably, this, by and large, need not be the case: The transition from left to right can take such a pathway that never passes through $\mathrm{S}(G_{\text{achiral}})=0$[6,13c,26]. An example is shown in Fig. 11 for the racemization pathway (1). It is seen in the CCM analysis of this racemization pathway (Fig. 11b) that the **S** value does not reach zero at any point. One must conclude, that somewhere along this pathway there exists a structure that is neither left nor right and yet is chiral. Its location is intimately linked to the specific definition of left and right handedness. A detailed analysis of this most interesing phenomenon and of the tentativeness of the very concept of left/right are treated in a subsequent report. (It is this tentativeness which prompts us not to assign, left-handedness or right-handedness to the chiral objects in this report, whenever such assignment is not needed for the discussion).

3.4 The Chirality of a Set of Vertices With Uncertain Locations

Information obtained from any analytical instrument has a certain degree of uncertainty of both inherent and experimental origin. In x-ray crystallographic analysis, for instance, the uncertainty in the location of atoms as obtained by diffraction is due to crystal imperfections, thermal motion, etc.[16]. We address ourselves now, to this problem[4], focusing on chirality.

Quite often the data is given as a collection of probability distribution functions of point locations. Given points with such uncertain locations, the following questions are of interest:

- What is the most probable closest achiral shape represented by the data?
- What is the probability distribution of the chirality measure values for the given data?

Let us begin our discussion with the first of these questions. Fig. 12a shows a configuration of two points whose locations are given by normal distribution functions. The dot represents the expected location of the point and the rectangle represents the standard deviation where

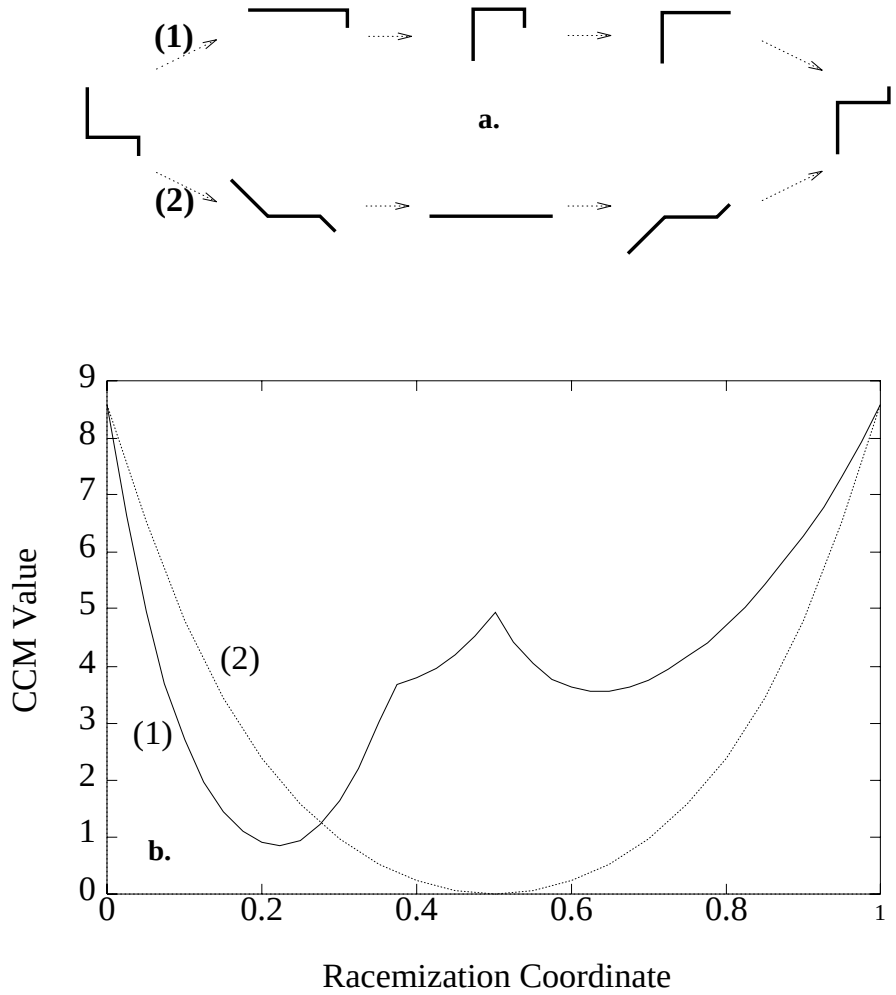

Figure 11: a) A chiral (1) and achiral (2) racemization pathway. b) The Continuous Chirality Measure of the two pathways. The abscissa represents a single cycle of evolution along the pathways. Pathway (1) never drops to $\mathbf{S}(G_{\text{achiral}}) = 0$.

the width and length are proportional to the standard deviation. In evaluating the most probable achiral shape, we apply the Maximum Likelihood criterion[61].

Let us consider, then, a simple case where two measurements Q_1, Q_2 in 2D are given (Fig. 12a), assumming their locations are given as normal probability distributions with expected location P_i and covariance matrix Λ_i, i.e $Q_i \sim \mathcal{N}(P_i, \Lambda_i)$ $i = 1, 2$. We aim at finding the mirror-symmetric configuration of points at locations $\{\hat{P}_i\}_{i=1}^2$ which is optimal under

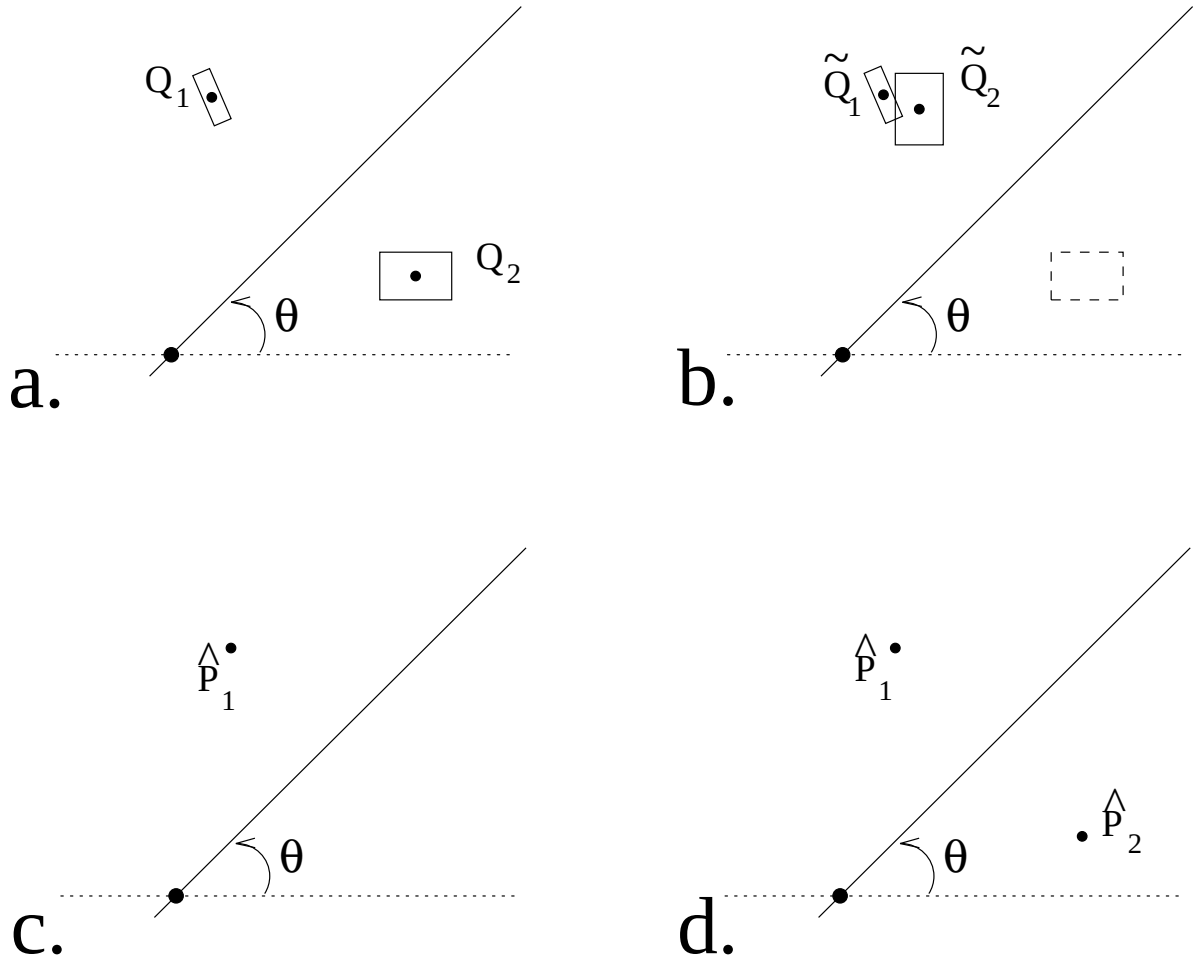

Figure 12: The folding-unfolding procedure, applied on two uncertain locations Q_1, Q_2: a) Original data; b) the folding; c) averaging; and d) unfolding - a σ-symmetric pair is obtained. Compare with Fig.1c-f

the Maximum Likelihood criterion which, maximizes the posteriori probability distribution. Denote by ω the centeroid of the most probable mirror-symmetric set of locations $\hat{P}_i$: $\omega = \frac{1}{2}(\hat{P}_1 + \hat{P}_2)$. The point ω is dependent on the location of the measurements (P_i) and on the probability distribution associated with them (Λ_i). Intuitively, ω is positioned at that point about which the folding (described below) gives the tightest cluster of points with small uncertainty (small standard deviation). We assume for the moment that the centroid ω is given. A method for finding ω is included in the detailed derivation of this method in ref. **4**. Given the angle θ of the reflection axis, we use the following variant of the folding method

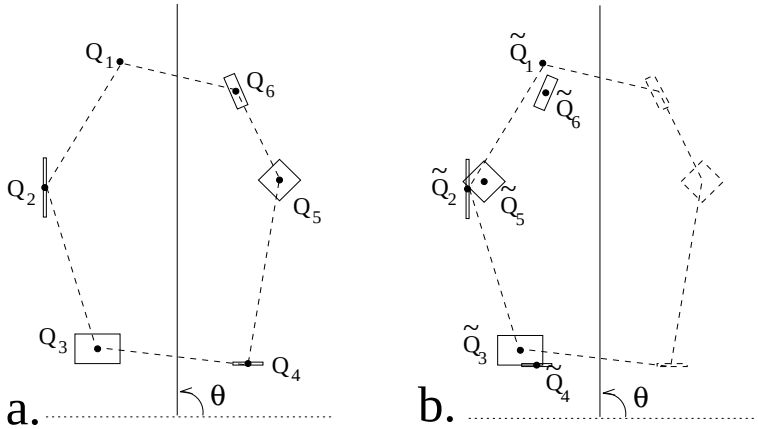

Figure 13: a) A configuration of 6 measurement points $Q_1, \ldots, Q_6$. b) The measurements $\{Q_i\}_{i=1}^6$, are divided into 3 pairs of measurements $(\{Q_1, Q_6\}, \{Q_2, Q_5\}, \{Q_3, Q_4\})$ and the folding is applied seperately to each pair obtaining the measurements $\{\tilde{Q}_i\}_{i=1}^6$. Averaging and unfolding are then carried out as in Fig. 12.

(Section 2.2):

1. The two measurements $Q_i \sim \mathcal{N}(P_i, \Lambda_i)$ are *folded* by reflecting one of the measurements (Q_2) about the reflection axis and leaving the other measurement (Q_1) as is. A new set of measurements $\tilde{Q}_i \sim \mathcal{N}(\tilde{P}_i, \tilde{\Lambda}_i)$ is thus obtained (Fig. 12b).

2. The folded measurements are *averaged* using a weighted average based on the distribution of the measurement, and a single point $\hat{P}_1$ is obtained (Fig. 12c). Averaging is performed by considering the two folded measurements $\tilde{Q}_1, \tilde{Q}_2$ as two measurements of a single point and $\hat{P}_1$ represents the most probable location of that point under the Maximum Likelihood criterion.

$$\hat{P}_1 - \omega = (\tilde{\Lambda}_1^{-1} + \tilde{\Lambda}_2^{-1})^{-1}(\tilde{\Lambda}_1^{-1}(\tilde{P}_1 - \omega) + \tilde{\Lambda}_2^{-1}(\tilde{P}_2 - \omega))$$

3. The average point $\hat{P}_1$ is *unfolded* as described in Section 2.2, obtaining points $\{\hat{P}_i\}_{i=1}^2$ which are perfectly mirror-symmetric with respect to the mirror-axis passing through ω at an angle θ (Fig. 12d).

When $m = 2q$ measurements are given, the m measurements $\{Q_i\}_{i=0}^{m-1}$, are divided into q pairs of measurements, and the folding method as described above is applied seperately to each pair of measurements following the general procedure of Section 2.2 (Fig. 13). Derivations and proof of this case are also found in ref. 4. Several examples are shown in Figure 14, where for a given set of measurements, the most probable mirror-symmetric shapes were found.

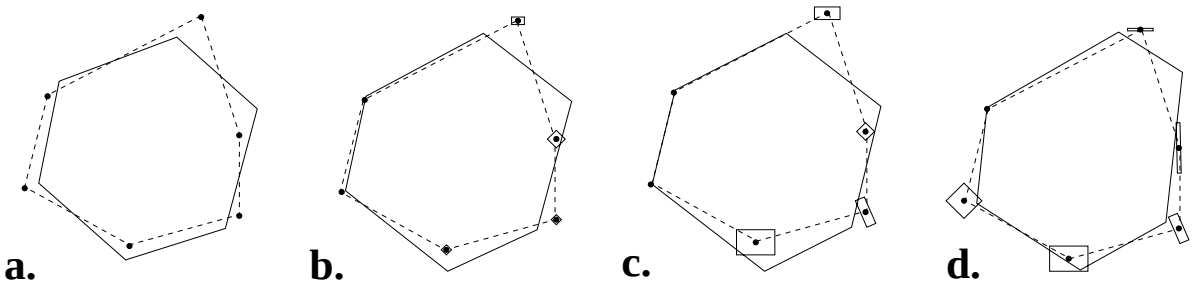

Figure 14: Examples of configurations of six measurements (dashed lines) given by a normal distribution function (marked as rectangles having width and length proportional to the standard deviation) and the most probable σ-symmetric shapes (solid lines).

Next, let us treat the second general question, namely, the probability distribution of chirality values. Consider again the configurations of 2D measurements given in Fig. 14 where each measurement Q_i is a normal probability distribution $Q_i \sim \mathcal{N}(P_i, \Lambda_i)$. The probability distribution of the chirality values of the original measurements is equivalent to the probability distribution of the location of the "average" point given the folded measurements as obtained in Step 1 and Step 2 of the algorithm. It is shown in ref. **4** that this probability distribution is a χ^2 distribution. In Figure 15 we display distributions of the chirality value for the various measurements of Fig. 14. As expected, the distribution of chirality values becomes broader as the uncertainties (the variance of the distribution) of the measurements increase.

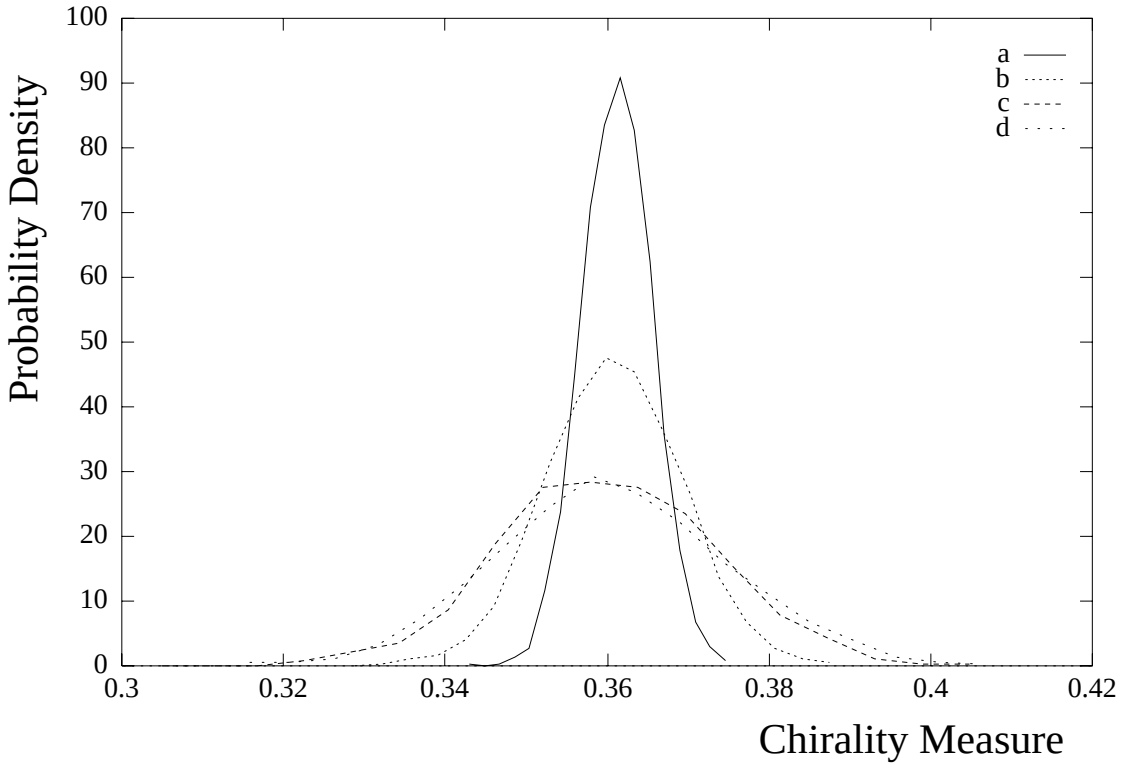

Figure 15: Probability distributions of the chirality measure, given for the sets of measurements in Fig. 14a-d. In this example the reflection line (the y-axis) and the pairing were pre-determined.

3.5 Chirality-Related Stereochemistry: Diastereomerism and Prochirality

The general approach of continuity can be extended to other stereochemical concepts as well. Here we comment on two chirality-related concepts, distereomerism and prochirality[44]. Consider the 2D diastereomeric pairs in Fig. 16 with the two chiral centers at positions 1 and 2: R_1R_2, S_1S_2 (2D-threo) and R_1S_2, R_2S_1 (2D-erythro) (here we follow the notation in ref. **33**).

The chirality content of the threo pair must be different than that of the erythro pair (it is $\mathbf{S}(\sigma) = 2.45$ and 3.16, respectively). We can then define $\mathbf{DE} = \|\mathbf{S}(\sigma)_1 - \mathbf{S}(\sigma)_2\|$ where DE is the diastereometric excess and $\mathbf{S}(\sigma)_1$, $\mathbf{S}(\sigma)_2$ are the CCM values of the diastereomers. Its

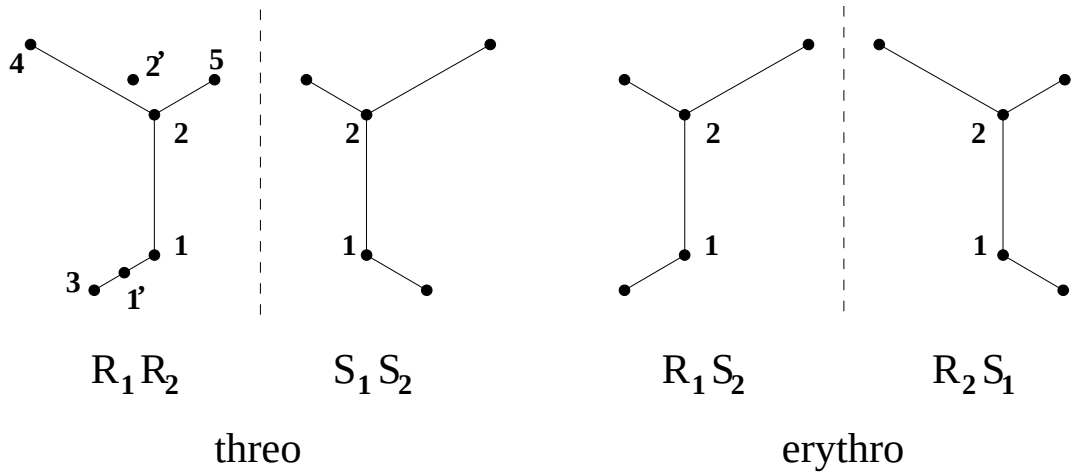

Figure 16: The application of the CCM to diastereomers.

bounds are $0 \leq \mathbf{DE} \leq 100$. For the pair in Fig. 16, $\mathbf{DE} = 0.71$. Note that the distance from one enantiomer to each of the two diastereomeric enantiomers, is not equal: The grounds of comparison between the two pairs is the minimal distance to chirality. Note also that for the case of a meso-RS pair, $\mathbf{DE} = \mathbf{S}$. Similarly, the diaseteromeric interactions between one chiral species and the R and S forms of another one, can be quantified continuously from the point of view of the chirality content. Various possible mutual alignments in such diastereometric interactions will show up in different chirality measures.

For diasterometric pairs, and actually for molecules containing more than two chiral centers, one may wish to analyse the chirality of each of the centers. A natural way to do so would be to replace the subtituents on the chiral center with the centroids of each of the substituents. Thus, the chirality of center 1 in Fig. 16 would be calculated using two substituents: 3, and the centroid of 4,2,5 ($\equiv 2'$); and the chirality of center 2, using the three substituents 4,5, and the centriod of 1,3 ($\equiv 1'$). We obtain for center 1, $\mathbf{S}(\sigma) = 3.37$ and for center 2 $\mathbf{S}(\sigma) = 0.285$, a difference which agrees with intuition.

Finally we comment on prochiral molecules, namely, achiral molecules which carry enantiotopic atoms or groups. This term deals with the potential of an achiral molecule to become chiral. Following the theme of this report, the degree of prochirality can also be analysed as a continuous property[44]. One can evaluate it by either replacing an enantiotopic group

with other groups, or by analysing the shape distortion exerted on the prochiral molecule, when placed in a chiral environment. We likewise notice, that stereotopism and homotopism (exchange by a C_n operation) are open to be analysed on a continuous scale as well, following the CCM and CSM procedures.

4 A Compendium of Chiral Structures and their Chirality Measures

The CCM method can be easily applied to virtually any chiral molecule, structure or process. We regard this property of our approach as an important advantage of it. The purpose of this section is to demonstrate the virsatility of the tool we developed. Needless to say, a next stage in such an investigation is to identify correlations between the CCM values of molecules and measurable physical and chemical properties; indeed our current research focuses on such issues in several of the following examples.

4.1 Static Structures

- Phospates: Following the Thema of this work, we recall the finding that practically all phosphates are not tetrahedral in their crystaline state[45], a problem delt with by Dunitz et al[46]. (The tetrahedricity of one such distorted phosphate was evaluated in ref. **3**). It follows that unless the tetrahedral distortion is symmetric itself (for instance one of the vertices is pulled out into a C_{3v} tetrahedron), phosphates are, by and large, chiral. Table 2 collects the CSM and CCM analysis of a number of phosphates from a compilation of phosphate-coordinates[45]. Measures of tetrahedricity, C_{3v}-ness and chirality are shown. Notice that $\mathbf{S}(T_d)$ values are typically larger than $\mathbf{S}(C_{3v})$ values, which in turn, are larger than the chirality $\mathbf{S}(\sigma)$ values. This is a reflection of the fact that C_s is a subgroup of C_{3v} which is a subgroup of T_d; that is, it is "easier" for the distorted tetrahedron to "find" a near-by reflection plane (in order to attain

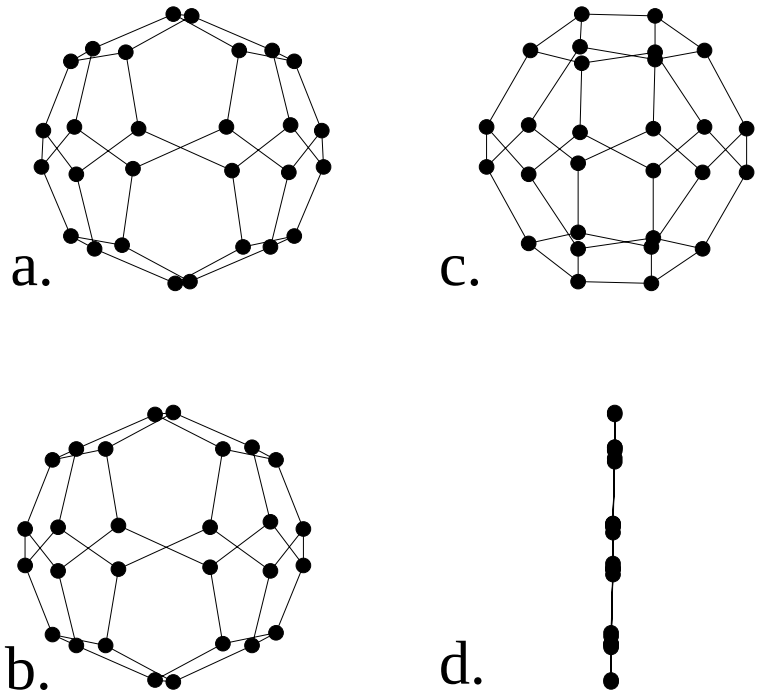

Figure 17: A chiral fullerene having 28 carbon atoms and its closest achiral configuration. a,c) The fullerene from two view points. b,d) The closest chiral configuration from the same view points. Notice that due to the connectivity, the closest achiral structure is the original fullerene collapsed onto the symmetry plane.

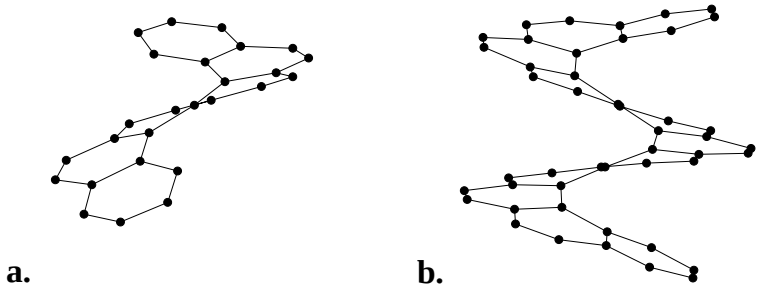

Figure 18: The structure of hexahelicene (a) and undecahelicene (b), as obtained from crystalography data.

achirality), than to shift all vertices to a perfect tetrahedron position. Note that none of the measured $\mathbf{S}(\sigma)$ values exceeds the $\mathbf{S}_{\max}(\sigma)$ of a tetrahedron (4.054) evaluated in Section 3.1.

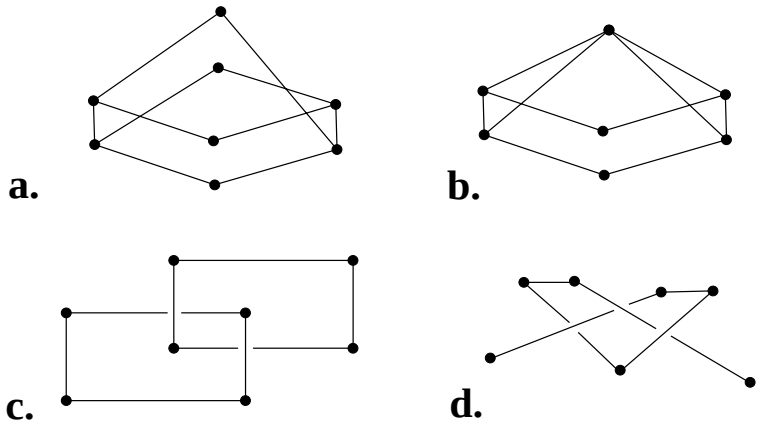

Figure 19: Some topological structures which give rise to chirality: a) Möbius strip and b) its' nearest achiral structure c) tilted catenanes (the two planes are tilted at 25.6° to each other), and c) knots. Their CCM values are respectively: 1.9470, 0.4113 and 0.5299.

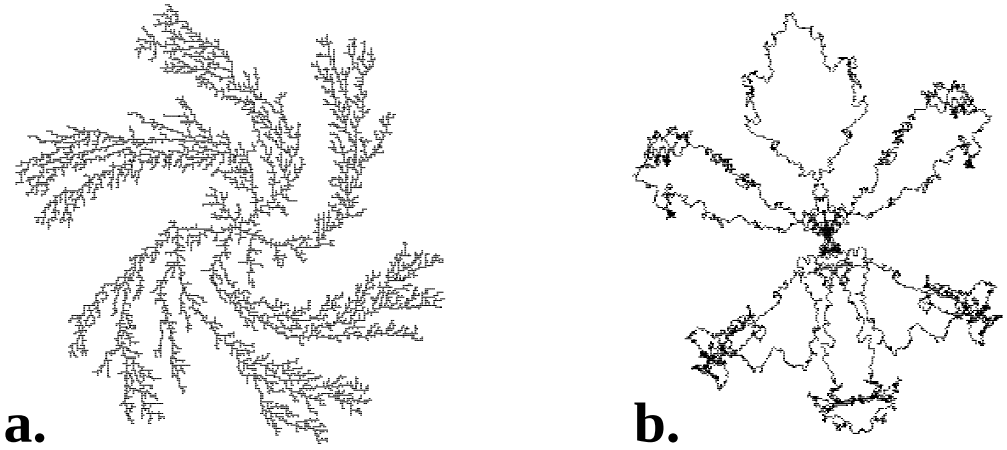

Figure 20: a) A Chiral diffusion limited aggregate (DLA). b) The nearest achiral figure. The chirality measure of the DLA in (a) is 3.40.

- Fullerenes: some members of this fascinating group of molecules, are chiral[47] (and an enantiomeric resolution was acheived recently[48]). Our method is capable of evaluating the degree of chirality of these molecules. As an example, let us take the C_{28}-fullerene shown in Fig. 17a,b. The symmetry of this fullerene is D_2 (achiral point group), and it is one of two topologically distinct C_{28} cages (the other being of T_d symmetry)[49]. The nearest achiral structure of the D_2 isomer, which obeys our crucial restriction of

preservation of topology, is the planar collapsed network shown in Fig. 17c,d. The chirality measure of the D_2-C_{28} fullerene is 24.943.

- Helicenes: Helical compounds from the smallest twisted molecules up to DNA, comprise a major class of chirality in chemistry[50]. A classical group of molecules in this class are the helicenes[51]. The CCM approach is capable of evaluating the chirality content of this type of molecules as well. Fig. 18 shows the structure of hexahelicene and undecahelicene, as obtained from crystalography data[52]. The chirality values of these two molecules are 5.645 and 10.154 respectively: The larger helicene is more chiral than the smaller one[53].

- Knots, Möbius Strips and Catenanes: We continue to demonstrate the versatility of the CCM approach on other topological distinct structures which give rise to chirality,namely, knots, Möbius strips and substituted catenanes[26,27]. Examples of these structures are shown in Figure 19 along with their chirality values. For the knot and the tilted catenane, the nearest achiral structure is planar (or, if one prefers, infinitisimaly close to a plane). For the Möbius strip, the nearest achiral structure colapses the twist to a point (Fig. 19b).

- Large Random Objects: Another family of objects for which chirality analysis is not trivial are the large (random) objects. Fig. 20a shows a chiral diffusion limited aggregate (DLA)[54,55]. We commented already on some basic conceptual difficulties in the application of the standard terms "symmetry" and "chirality" to such objects[3], and will expand on it later. Here, however, we show that the CCM tool is capable of dealing with such complex structures as well. For DLA's, all points are contour points[54], and therefore a contour analysis[2,3] can be applied here, which simplifies the calculations. The chirality measure of the DLA in Figure 20a is 3.40. The nearest chiral structure is shown in Figure 20b. For actual experimental examples of chiral growth phenomena, see for instance, ref. **56**.

4.2 Dynamic Chirality Changes: Fluxional Molecules, Vibrations, Rotations and Concerted Reactions

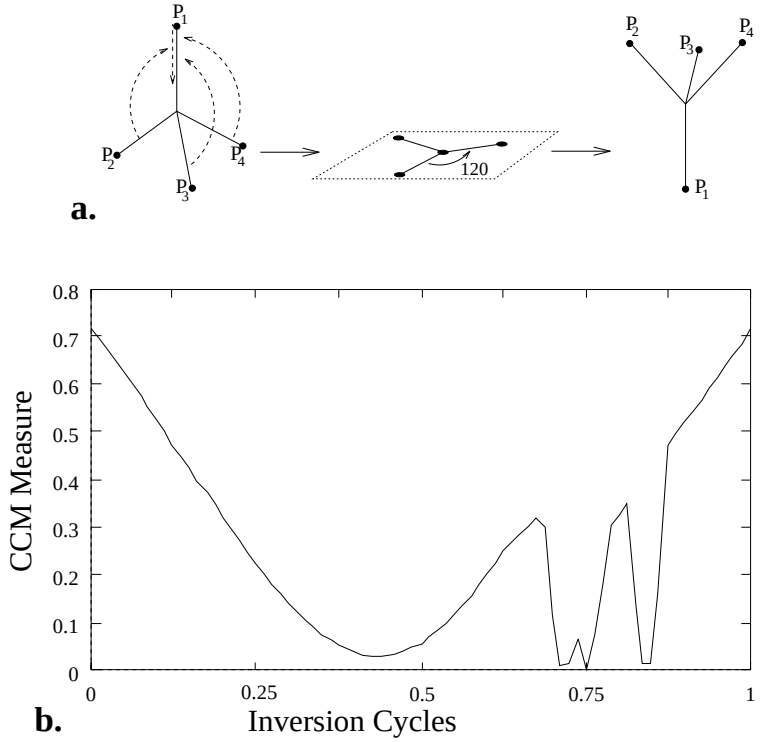

Figure 21: Chirality during the process of a of a Walden inversion. a) In this example, the three arms marked P_2, P_3, P_4 move with random phase shifts (of $2/32, 3/32$ and $6/32$ of the cycle). b) The changes in chirality during the inversion.

Given a sufficiently fast camera, one can follow the continuous changes in the symmetry of any dynamic process. In Parts 2 and 3 we demonstrated it for a fluxional, Walden-type flip-flopping[3], for vibrating CX_4[3] and for roatating ethane[3,4], and analysed the continuous changes in the relevant symmetry groups. If the dynamic process removes an improper element of symmetry from the molecule then it becomes chiral. Actually, as we show now, these systems are chiral during most of the dynamic process. Comming back to some of the examples of Parts 2,3 , we now perform a Walden inversion so that the movements of P_2, P_3, P_4 are phase shifted (see Fig. 21a and details in the caption), and note that the

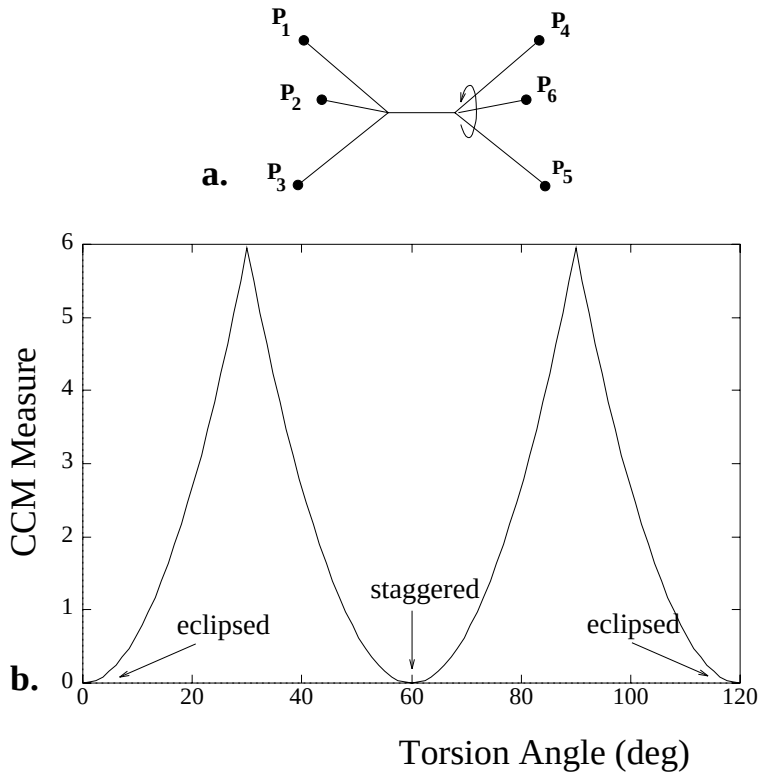

Figure 22: The changes in chirality of a rotating ethane structure (a) are shown in (b). Note that ethane is chiral most of the time.

chirality is retained almost throughout the cycle (Figure 21b); and we note that rotating ethane (Fig 22a)[4] is chiral (D_3) most of the time (Fig 22b)[13c]. Finally, we note that structural changes along the pathway of a concerted reaction, again, may be chiral through most of the process. Consider for instance, the intermolecular approach of the Diels-Alder reaction between, say, 1,3-butadiene and propene; or a [2+2] reaction between two differently substituted ethylenes, not ideally aligned[57]; or a disrotatory ring openning, in which the rotating ends are not exactly in phase, etc. In all of these cases, the reacting system is chiral, its degree of chirality changes throughout the process, and the CCM approach is capable of quantifying it as a novel reaction coordinate. Research in this direction is in progress in our research group.

5 Conclusion

We have presented a versatile general tool for quantifying shape-chirality as a continuous structural property. This report was devoted to a detailed exposition of the method for both simple and more complicated cases. The general behaviour of the chirality measure was analysed for various shapes and structures, for limiting cases, and for continuous structural changes. Returning to the Thema of our work, we hope we were able to convince the reader that chirality should be considered as a universal structural property (existing even in classical "achiral" structures), and that it is up to the detectability limits of analytical tools to realize this universality.

In subsequent reports we expand several topics which we only briefly touched in this introductory paper (especially those outlined in Section 4), emphasizing, both basic issues of the very concept of chirality and correlations between the chirality values and other molecular properties.

Acknowledgements: We thank Joey Edelstein, Shahar Keinan and Omer Katzenelson for for analysis of data described in Section 4, Prof. M. Kaftori for the helicenes data and Prof. P. Fowler for the fullerene data. We thank Mane for helpful discussions. D.A. is a member of the F. Haber Research Center for Molecular Dynamics and of the Farkas Center for Light Energy Conversion.

References

1. Emsley, J. *Science Watch*, July/August 1993, 7.

2. Zabrodsky, H.; Peleg, S.; Avnir, D. *J. Am. Chem. Soc.* 1992, 114, 7843. (For a discussion of the concepts presented in this paper, see Fowler, P.W. *Nature*, 1992, 360,

626).

3. Zabrodsky, H.; Peleg, S.; Avnir, D. *J. Am. Chem. Soc.* 1993, 115, 8278. (Erratum: ibid, 11,656).

4. Zabrodsky, H.; Avnir, D., *Adv. Molec. Struct. Res.* 1994, 1, 1.

5. Heilbronner, E,; Dunitz, J.D. "Reflections on Symmetry in Chemistry and Elsewhere", VCH: Basel, 1993.

6. Sokolov, V.I. "Introduction to Theoretical Stereochemistry", Nauka: Moscow, 1979 (in Russian); English translation: Gordon and Breach, Science Publishers: New York, 1991.

7. Kitaigorodskii, A., "Organic Chemical Crystallography", Consultant Bureau: New York, 1961, p.230.

8. Gilat, G; Schulman, L.S. *Chem. Phys. Lett.*, 1985, 121, 13; Gilat, G. *J. Phys. A: Math Gen.*, 1989, 22, L545.

9. Meyer, A.Y.; Richards, W.G. *J. Computer-Aided Molec. Design*, 1991, 5, 427.

10. Seri-Levy, A.; Richards, W.G., *Tetrahedron: Asymmetry*, 1993, 4, 1917; *J. Medicinal Chemistry*, 1994, in press.

11. Pfeiffer, C.C. *Science* 1956, 124, 29.

12. a) Buda,B.; Auf der Hiede, T.P.E.; Mislow, K. *J. Math. Chem.* 1991, 6, 243. b) As noted by the authors of this reference, their "analysis leads to the paradoxical conclusion that the most chiral triangle is infinitisimally close to an achiral one", namely a line segment. Compare with Section 3.1.

13. a) Kuz'min, V.E.; Stel'mach, I.B. *Zh. Strukt. Khim.* 1987, 8, 45 and 50. b) Kuz'min, V.E.; Stel'mach, I.B.; Bekker, M.B.; Pozigun, D.V. *J. Phys. Org. Chem.* 1992, 5, 295. c) Kuz'min, V.E.; Stel'mach, I.B.; Yudanova, I.V.; Pozigun, D.V; Bekker, M.B.;

ibid, 299. d) Kutulya, L.A; Kuz'min, V.E.; Stel'mach, I.B.; Handrimailova, T.V.; Shtifanyuk, P.P, ibid, 308.

14. Kauzmann, W.; Clough, F.B.; Tobias, I., *Tetrahedron*, 1961, 13, 57.

15. Ugi, Z. *Naturforsch*, 1965, 20b, 405.

16. Ruch, E. Acc. Chem. Res. 1972, 5, 49.

17. King, R.B. in "New Developments in Molecular Chirality", Mezey, P.G., ed., Kluwer: Dordrecht, 1991, p.131.

18. Derflinger, G. in "Chirality, from weak Bosons to the α-Helix", Janoschek, R., ed., Springer: Berlin, 1991, Chapter 3.

19. Mezey, P.G. in ref.17, p.257.

20. Mezey, P.G.; Maruani, *J. Molec. Phys.* 1990, 69, 97; Mezey, P.G. J. Quant. Chem. 1993, 45, 177.

21. Mezey, P.G. *J. Math Chem.* 1992, 11, 27.

22. Rassat, A. *C.R. Acad. Sc. Paris II*, 1984, 299, 53.

23. Buda, A.B.; Mislow, K. *J. Am. Chem. Soc.* 1992, 114, 6006.

24. Sokolov, V.I. *Comput. Math. Appl.* 1986, 12b, 547.

25. Chauvin, R. *J. Phys. Chem.* 1992, 96, 4706.

26. Walba, D.M. in ref.17, p.119.

27. Flapan, E. in ref.17, p.209.

28. Harary, F.; Mezey, P.G., in ref.17, p.241.

29. Zimpel, Z. *J. Math. Chem.* 1993, in press.

30. Mezey, P.G. in "New Theoretical Concepts for Understanding Organic Reactions", Kluwer: Dordrech, 1989, p.77.

31. Luzanov, A.V.; Babich, E.N. *Struct. Chem.* 1992, 3, 175.

32. Donoghue, J.F.; Golowich, E.; Holstein, B.R. *Phys. Rev. D.*, 1984 30, 587.

33. Hel-Or, Y.; Peleg, S.; Avnir, D. *Langmuir* 1990, 6, 1991 (c.f. also ref.26).

34. Avnir, D.; Meyer, Y.M. *J. Molec. Struct. (Theochem)* 1991, 226, 211.

35. Meyer, A.Y.; Avnir, D. *Struct. Chem.* 1991, 2, 475.

36. Other normalization and scaling procedures are possible. These will be compared elsewhere.

37. Even, S. "Graph Algorithms"; Computer Science Press: Potomac, Maryland, 1979.

38. Hoffman, C.M., "Group Theoretic Alg. and Graph Isomorphism"; Springer-Verlag: New York, 1982.

39. The isomorphism algorithm (Section 2.3) which we implemented is a recursive algorithm that uses a depth-first search for permutations. Thus a node is tested for a possible match with every other node in the given graph, and if a match is found then all other nodes are matched. Testing if two nodes can be matched is performed by ecaluating the equality of the valency of the two nodes and te matching of neighbors.

40. Atkins, P.W. "Creaction Revisited"; W.H. Freeman: Oxford, 1992, Chapter 5.

41. Gilat, G.; Gordon, Y. *J. Phys. A*, 1993, submitted.

42. Gilat, G. *J. Math. Chem.* 1993, in press.

43. The contours were computed as Slater-type orbitals, represented by three Gaussian functions (STO-3G) using GAMESS (General Atomic and Molecular Electronic Structure System) program, (Schmidt, M.W.; Baldridge, K.K.; Boaz, J.A.; Jensen, J.H.;

Koseki, S.; Gordon, M.S.; Nguyen, K.A.; Windus, T.L.; Elbert, S.T., *Quant. Chem. Program Exchange Bulletin* 1990, 10,52.), by Dr. David Danovich and Prof. Sasson Shaik.

44. Halevi, E.I. *J. Chem. Res.* (S), 1985, 206.

45. Calvo, C. *Can. J. Chem.* 1969, 47, 3409.

46. Murray-Rust, P.; Bürgi, H.B.; Dunitz, J.D. *Acta Cryst.* 1978, B34, 1787.

47. Manolopoulos, D.E.; Fowler, P.W., *J. Chem. Phys.* 1992, 96, 7603.

48. Hawkins, J.M.; Meyer, A. *Science* 1993, 260, 1918.

49. Fowler, P.W.; Baker, J. *J. Chem. Soc. Perkin Trans 2*, 1992, 1665; Fowler, P.W.; Manolopolous, D.E.; Redmond, D.B.; Ryan, R.P. *Chem. Phys. Lett.* 1993, 202, 371.

50. Meurer, K.P.; Vögtle, F. *Topics in Current Chem.* 1985, 127, 1.

51. Martin, R.H., Angew. Chem. Inter. ed. Eng., 1974, 10, 649.

52. Cambridge Structural Database. For description see, Allen, F.H. et al. *Acta Crystallogr.* 1979, B35, 2331.

53. For a detailed analysis of these structures, along with correlations between S(σ) and physical properties, see H. Zabrodsky; M. Kaftori; J. Edelstein and D. Avnir, manuscript in preparation.

54. Meakin, P. *Heterogeneous Chem. Rev.* 1994, 1, in press; Meakin, P. in "The Fractal Approach to Heterogeneous Chemistry: Surfaces, Colloids, Polymers" Avnir, D., ed., (Wiley: Chichester, 3rd corrected printing, 1992), Chapter 3.1.2.

55. Nagatani, T.; Sagués, F. *J. Phys. Soc. Jap.* 1990 59, 3447.

56. McConnel, M.H. *Ann. Rev. Phys. Chem.* 1991, 42, 171.

57. Halevi, E.A. "Orbital Symmetry and Reaction Mechanism - The OCAMS View" Springer: Berlin 1992, Chapter 6.

58. Horn, B. "Robot Vision"; MIT Press: Cambridge, Mass., 1987.

59. a) Arun, K.S.; Huang,T.S.; Blostein, S.D. *IEEE Pattern Anal. and Mach, Intel.* 1987, 9(5), 698.; b) Horn, B.K.P.; Hilden, H.M. ; Negahdaripour, S. *J. Opt. Soc. Am.* 1988, 5(7), 1127.

60. Press, W.H.;Teukolsky, S.H.; Vetterling, W.T.; Flannery,B.P. "Numerical Recipes in C"; Cambridge Univ. Press: New York, 1992.

61. DeGroot, M.H. "Probability and Statistics"; Addison-Wesley: Reading, Mass., 1975.

Appendix

Pose Estimation Algorithm

Given 2 sets of points $\{P_i\}_{i=1}^n$ and $\{\tilde{P}_i\}_{i=1}^n$ and given a correspondance between them (without loss of generality, we assume point P_i corresponds to point $\tilde{P}_i$):

1. Calculate the centeroids P and $\tilde{P}$ of the two sets.

2. Translate each set so that its centroid aligns with the origin. i.e. $Q_i = P_i - P$ and $\tilde{Q}_i = \tilde{P}_i - \tilde{P}$ for $i = 1 \ldots n$

3. calculate the 3x3 matrix H:
$$H = \sum_{i=1}^{n} Q_i \tilde{Q}_i^t$$

4. Find the Singular Value Decomposition (SVD) of H i.e. find 2 orthonormal 3x3 matrices U and V and find a diagonal 3x3 matrix W such that $H = UWV^t$. (Computational algorithms are readily available - see ref.60 for example).

5. Calculate the rotation matrix R:
$$R = VU^t$$

The translation $T = P - \tilde{P}$ and the rotation matrix R optimally transform points $\{\tilde{P}_i\}$ such that the sum of squared distances between these points and their corresponding point P_i is minimal. Note that the determinant of R should be one. In some extreme cases this is not so and additional steps are required[59a].

List of Figures

List of Tables

Table 2: Symmetry and Chirality Measures of Phosphates.

Compound*	T_d	C_{3v}	σ
$Na_3H(PO_4)_4$ (10)	20.5691	10.3628	0.4784
	25.8591	9.5236	0.4345
	37.3978	21.695	0.0693
	28.0098	0.6014	0.0171
$(NH_4)_2HPO_4$ (35)	13.5883	3.3414	0.1279
$C_7H_{19}N_3 \cdot 3/2H_3PO_4 \cdot 3H_2O$ (19)	34.177	12.3783	0.4417
	21.2617	9.0154	0.0903
	31.2404	11.1761	0.1636
$(NH_2)_2 \cdot CO \cdot H_3PO_4$ (78)	22.6043	2.4808	0.8037
$CH_3COOH \cdot H_3PO_4$ (61)	20.6114	3.4992	1.0853
	18.2685	2.0785	0.1983
$H_3PO_4 \cdot 1/2H_2O$ (16)	17.9484	4.2638	0.2249
	14.4922	2.4886	0.2113
$N_2H_6(H_2PO_4)_2$ (18)	33.959	12.4358	1.7559
$(C_6H_4NO_2)_3PO_4$ (38)	29.6991	4.5375	0.4785
$(NH_4)_3PO_4 \cdot 3H_2O$ (31)	31.7482	6.9858	0.0558
Na_2HPO_4 (53)	23.7603	12.8566	2.1199
$C_{10}H_{26}N_4 \cdot 2H_3PO_4 \cdot 6H_2O$ (84)	28.9243	8.6902	1.9072
$Mn_2(PO_4)Cl$ (33)	19.7312	14.2482	0.9310
$Mn_2(PO_4)F$ (32)	20.0056	11.7915	0.4144
$MgHPO_4 \cdot 3H_2O$ (15)	33.0826	0.0689	0.0051
$CaHPO_4$ (46)	29.3306	3.4538	0.1045
	25.9993	19.2799	0.1945
$Mg_3(PO_4)_2$ (20)	34.3814	15.5892	0.5321
$C_6H_{15}O_2N_4 \cdot 2(HPO_4) \cdot 6H_2O$ (81)	29.4232	2.4762	0.2250
$C_5H_{11}N_3(H_2PO_4)_2 \cdot H_2O$ (59)	27.285	5.2666	0.4227
	30.8529	5.6503	0.2250
$Cd_2P_2O_7$ (58)	17.1107	11.2703	0.4227
	15.3496	7.7854	0.9439

*The numbers in brackets refer to Baur's numbering. More than one entry per compound, refer to distinctly different phosphate moieties in the crystal.