Continuous Symmetry Measures. 2. Symmetry Groups and the Tetrahedron

Hagit Zabrodsky,† Shmuel Peleg,† and David Avnir*

Contribution from the Departments of Computer Science and Organic Chemistry, The Hebrew University of Jerusalem, 91904 Jerusalem, Israel

Received January 20, 1993

Abstract: We treat symmetry as a continuous property rather than a discrete "yes or no" one. Here we generalize the approach developed for symmetry elements (Part 1: *J. Am. Chem. Soc.* **1992**, *114*, 7843–7851) to any symmetry group in two and three dimensions. Using the Continuous Symmetry Measure (CSM) method, it is possible to evaluate quantitatively how much of *any* symmetry exists in a nonsymmetric configuration; what is the nearest symmetry group of any given configuration; and how the symmetrized shapes, with respect to any symmetry group, look. The CSM approach is first presented in a practical easy-to-implement set of rules, which are later proven in a rigorous mathematical layout. Most of our examples concentrate on tetrahedral structures because of their key importance in chemistry. Thus, we show how to evaluate the amount of tetrahedricity (T_d) existing in nonsymmetric tetrahedra; the amount of other symmetries they contain; and the continuous symmetry changes in fluctuating, vibrating, and rotating tetrahedra. The tool we developed bears on any physical or chemical process and property which is either governed by symmetry considerations or which is describable in terms of changes in symmetry.

1. Introduction

Symmetry, as one of the most fruitful working tools of the natural sciences, has delt, by definition, with the exact, the perfect, the ideal, the flawless. Nature, however, in most of its manifestations, cannot be described by these adjectives if symmetry, in its present teaching, is taken as a reference of attitude. Except for man-made objects, exact symmetries rarely exist: they are rare on a molecular scale where vibrations continuously distort "ideal" shapes; they are rare on the macromolecular scale where, e.g., protein chains usually fold in a nonsymmetric arrangement; they are rare at the microscopic range (aggregates, liquid crystals, etc.); and they are rarely found in the largest objects of the universe. To overcome this dichotomy, reality must be brought into the framework of the current language of symmetry by resorting to averaging over long time scales or over large assemblies and by limiting the study to narrow time windows of observation or to shapes of equilibrium state.

We suggest that the opposite approach be attempted: create a symmetry language that is tailored to reality, i.e. to its "imperfections" and to its "nonidealities". In Part 1[1] we presented the notion that a natural new symmetry language that fulfills this requirement should allow shades of gray, in contradistinction to the classical restrictive "black or white" approach. We argue that, rather than asking whether or not a given shape has a desired (element or group of) symmetry, one is better off asking *how much* of that symmetry exists in that shape and that replacing a strict "either-or" language with a simple language that allows "in betweens" is bound, in principle, to enrich the information content of symmetry analysis. Many specific examples in chemistry and material science which may benefit from such an approach were detailed in Part 1. Some previous treatments of similar problems are, e.g., perturbation analysis in spectroscopy,[2] deviation analysis from specific shapes[3] (see Section 4.1), fuzzy-set analysis,[4] and measures on convex sets.[5,6] As will be evident below, our approach to the problem of nonideal symmetry is quite different,[7] being guided by three principles: (1) Nonsymmetric shapes should not be treated as a perturbation of an ideal reference. Such shapes, as well as perfectly symmetric ones, should appear on a single continuous scale with no built-in hierarchy of subjective ideality. (2) Assessing symmetry should be detached from referencing to a specific shape. (3) It should be possible to evaluate the symmetry of a given configuration with respect to *any* symmetry group.

These guidelines are implemented as follows: Given a shape composed of n_p points P_i ($i = 1, ..., n_p$) and a symmetry group G, **the symmetry measure $S(G)$ is a function of the minimal displacement the points P_i of the shape have to undergo in order to acquire G-symmetry**. The tool we developed (Part 1[1] and below) identifies the points $\hat{P}_i$ of the nearest shape having the desired symmetry. Once the nearest $\hat{P}_i$'s are calculated, a continuous symmetry measure is evaluated as

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

(square values are taken so that the function is isotropic, continuous, and differentiable). Prior to evaluation, one normalizes the shape by scaling about its center of shape so that the maximum distance of any P_i to the center is 1. We thus obtain the limits $0 \le S'(G) \le 1$. For convenience, we expand this scale by a factor of 100 and express the symmetry measure as

$$S(G) = 100(S'(G)) \qquad (2)$$

Thus, if a shape has the desired symmetry, $S(G) = 0$. A shape's symmetry measure increases as the shape departs from G-symmetry and it reaches a maximal value (not necessarily 100—see

* Author to whom correspondence should be addressed at the Department of Organic Chemistry.
† Department of Computer Science.

(1) Zabrodsky, H.; Peleg, S.; Avnir, D. *J. Am. Chem. Soc.* **1992**, *114*, 7843.
(2) For example: Orchin, M.; Jaffe, H. H. *Symmetry, Orbitals and Spectra*; Wiley: New York, 1971.
(3) Murray-Rust, P.; Bürgi, H. B.; Dunitz, J. D. *Acta Crystallogr.* **1978**, *B34*, 1787.
(4) Mezey, P. G. In *New Theoretical Concepts for Understanding Organic Reactions*; Bertrán, J., Csizmadia, I. G., Eds.; Kluwer: Dordrecht, The Netherlands, 1989; pp 55 and 77. Mezey, P. G. *J. Quant. Chem.* **1993**, *45*, 177.
(5) Grünbaum, B. Measures of Symmetry for Convex Sets. *Proc. Symp. Pure Math*: *Am. Math. Soc.* **1963**, *7*, 233.
(6) On a qualitative level, the need for a symmetry scale has been expressed in several references. For instance: Hargittai, I.; Hargittai, M. *Symmetry Through the Eyes of a Chemist*; VCH: Weinheim, Germany, 1986. Rosen, J. *A Symmetry Primer for Scientists*; Wiley: New York, 1983; Chapter 5.
(7) Zabrodsky, H.; Avnir, D. *Adv. Mol. Struct. Res.*, in press.

0002-7863/93/1515-8278$04.00/0

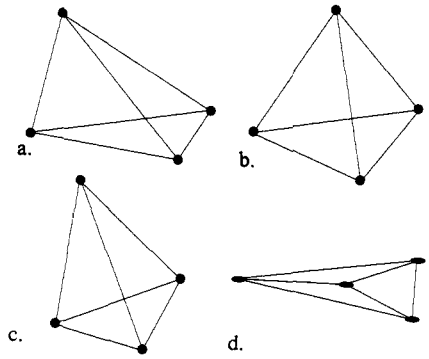

Figure 1. By how much is the distorted tetrahedron in (a) distant from T_d-symmetry? from C_{3v}-symmetry? from D_{3h}-symmetry? (Quantitative answers are detailed in the text.) (b) The perfect tetrahedron closest to a; (c) the C_{3v}-symmetric configuration closest to a; (d) the D_{3h}-symmetric configuration (planar) closest to a.

Section D of the Appendix). Equations 1 and 2 are general and allow one to evaluate the symmetry measure of *any* shape relative to *any* symmetry group or element. No reference shape is assumed at the beginning of the analysis though it is obtained as an end outcome.

In Part 1[1] we solved the problem for the two basic elements, rotation and reflection, mainly for two-dimensional shapes. Here, as mentioned above, we extend our treatment to any symmetry group, sub-group, or class and to three-dimensions with particular reference to the tetrahedron, perhaps the most important three-dimensional structure in chemistry. Finally we note that at the moment we are interested in purely geometrical features. However, the shape of mass distribution, charge distribution, or any other physical property can be used instead.[7] The analysis of such real-molecular properties will follow in subsequent reports.

Our paper is structured as follows: In Section 2 we underline the principles of the continuous symmetry measure using some basic, two-dimensional cases. In Section 3 we extend the treatment to three dimensions and use as examples a variety of tetrahedral structural properties, including the basic tetrahedricity measure. In Section 4 we provide some examples of the implementation of the CSM approach including reanalysis of X-ray data of distorted tetrahedra and modeling of dynamical symmetry behavior of fluxional, vibrating, and rotating tetrahedra. Rigorous mathematical treatment is then provided in the Appendix.

2. Underlying Principles of the Continuous Symmetry Measure (CSM)

2.1. Presentation of the Question. Figure 1a shows a distorted tetrahedron representing, perhaps, different substituents on an sp^3 carbon or a frozen moment in a CX_4 vibration mode, etc. We seek to answer the following: How much symmetry is there in the structure of Figure 1a with respect to *any* symmetry group? In particular, how much tetrahedricity, i.e. T_d-symmetry, does it contain? We emphasize a delicate but important point: we are seeking to quantify the distance of a given object from a symmetry group and *not* the deviation from a specific predetermined object having the desired symmetry. In fact, in most cases, it is quite difficult to guess what is the shape closest to the object and having the desired symmetry. Thus although it is straightforward that the T_d-symmetry object closest to Figure 1a is a perfect tetrahedron (Figure 1b), it is not easy to guess that the closest C_{3v}-symmetric object is the one shown in Figure 1c and that the closest D_{3h} object is that shown in Figure 1d. The continuous symmetry measure (CSM) method is capable of (1) quantifying the symmetry measure, in our example of structure 1a with respect to T_d-ness, C_{3v}-ness, or D_{3h}-ness or with respect to any other symmetry group (Values for Figure 1 are derived below.), and (2) providing the specific closest shapes which have the desired symmetry.

2.2. Evaluating the CSM: The Folding/Unfolding Method. Our approach is based on the very method of constructing a shape which is symmetric with respect to a given symmetry group. Let us recall how this is done, and for simplicity of explanation we use, for the moment, two-dimensional shapes. As an example, we build a two-dimensional (2D) D_3 shape, i.e. a planar structure with one C_3 rotational symmetry element and one reflection symmetry element σ (which is equivalent to C_2 in 3D). In 2D the rotation is about a point in the plane and the reflection is through a line in the plane. The D_3 symmetry group may be of different orientations and positions (thus the rotation can be about any given point in the plane and the reflection about a line of any orientation), but a natural choice would be to consider a D_3 symmetry group where the rotation is about the origin and the reflection is about one of the axes (the y-axis). The D_3 symmetry group is of order 6 with the following elements or operations (Figure 2a):

$g_1 = E$ = the identity

$g_2 = \sigma$ = reflection about the y-axis

$g_3 = C_3$ = rotation about the origin by $2\pi/3$ radians

$g_4 = C_3\sigma = \sigma C_3^2$ = rotation by $4\pi/3$ followed by reflection

$g_5 = C_3^2$ = rotation about the origin by $4\pi/3$ radians

$g_6 = C_3^2\sigma = \sigma C_3$ = rotation by $2\pi/3$ followed by reflection

Given an arbitrary point P_1 in an xy-plane where σ is the y-axis and C_3 rotates about the origin, a two-dimensional D_3-symmetric arrangement of points is obtained by applying the six g_i operations, for instance as follows (Figure 2b): (1) Rotate P_1 by $2\pi/3$ radians: P_3 is obtained. (2) Rotate P_1 by $4\pi/3$ radians: P_5 is obtained. (3) Reflect P_1, P_3, and P_5 about the y-axis: P_2, P_6, and P_4 are obtained, respectively; a D_3-symmetric collection of six points is constructed (Figure 2c). Such a structure can be obtained by many other orderings of the operations and many other orientations and positionings of the symmetry group. However, when *connected* objects, such as molecules, are of interest (in our case a hexagon), it is more natural to select that sequence of operations which follows the desired connectivity. In our D_3 example, we built the hexagon from P_1 by following the order of operations as given above which conserves the order along the hexagon boundary, i.e. $g_1 \rightarrow g_6$ (Figure 2c). In this report we concentrate on envelope (cyclic) connectivities, i.e. polygonal and polyhedral configurations. Branched connectivities are discussed for an introductory case in Section 3.3 and 4.1 and then in detail in Part 4.

We call the procedure of obtaining a symmetric shape by applying a set of operations g_i on a point *unfolding* (in our example, a D_3 object is unfolded from P_1). One is usually interested in unfolding that follows a given connectivity (although the tool we develop here is applicable to disconnected assemblies as well). The inverse procedure of unfolding is *folding*: points P_2, ..., P_6 are folded onto P_1 by applying the inverse operations g_i^{-1}. All points coallesce onto a single point, P_1. Note, however, that if points P_1, ..., P_6 are not perfectly D_3-symmetric, the folding results in a cluster of points rather than a single point. The folding–unfolding procedure is the very basis of our method for evaluating CSM values with respect to symmetry groups, the idea being to minimize the cluster spread.

2.3. The Basic Case ($n_p = n_g$). One can try to match various shapes with various symmetry groups, but let us first explain the folding–unfolding method using the basic case where the number of points (n_p) equals the number of elements in the symmetry group (n_g) and the connectivity is cyclic. Specifically, we shall determine here how much (two-dimensional) D_3-symmetry exists in the distorted hexagon shown in Figure 3a (we recall that in our previous report[1] this question was asked for basic symmetry elements; i.e., the CSM values of a distorted hexagon were evaluated for C_2, C_3, C_6, and σ). The folding–unfolding method is performed as follows: (1) Determine the centroid of the object

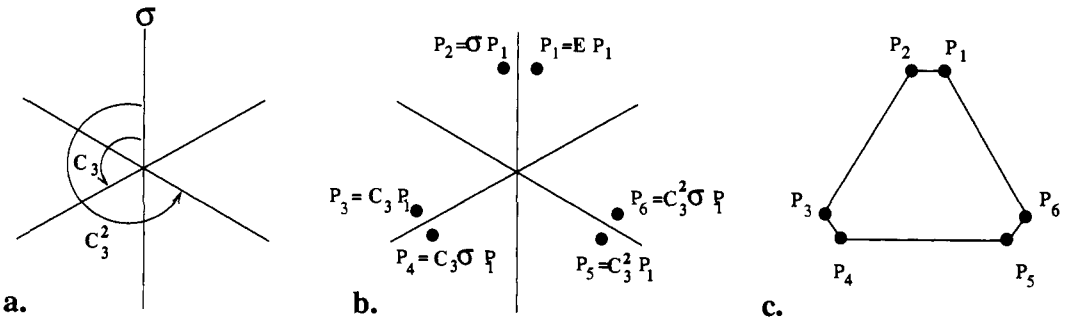

Figure 2. Creating a D_3-symmetric hexagon: (a) the D_3-symmetry group has six elements (see text); (b) applying the six group elements on the point P_1; (c) a D_3-symmetric hexagon of six points is obtained.

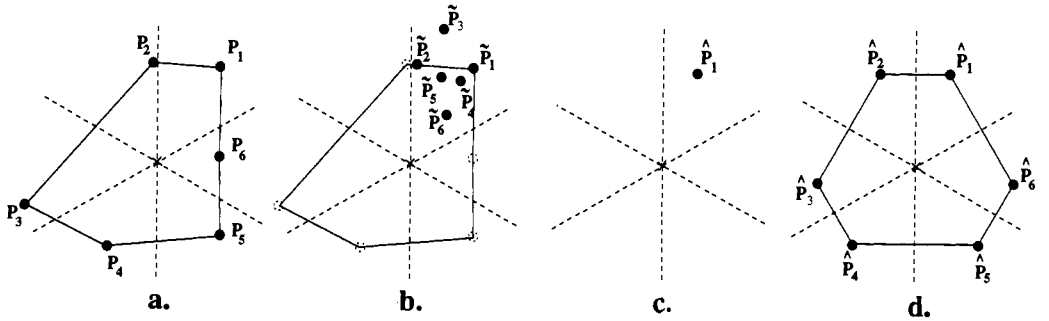

Figure 3. Measuring D_3-symmetry of a distorted hexagon using the folding–unfolding method: (a) the distorted hexagon; (b) folding the hexagon vertices in a—a cluster of points is obtained; (c) averaging the cluster of points in b—a single point is obtained; (d) unfolding the averaged point in c—a D_3-symmetric configuration is obtained.

(hexagon in our case). Translate the object so that it's centroid coincides with the origin, and scale the object so that the maximal distance from the origin to any of the vertices is 1 (Figure 3a). (2) Translate the symmetry group so that all operations are about the origin (i.e. all rotations are about the origin and all reflection lines or planes pass through the origin). (3) Select an ordering of the operations of the desired symmetry group that follows the connectivity of the P_i vertices. In our case, two orderings are possible: $g_1, ..., g_6$ as listed above and the reverse, $g_6, ..., g_1$. We proceed with the first and return to the second in step 8. (4) *Fold* the vertices $P_1, ..., P_6$ by applying to each P_i the symmetry operation g_i^{-1}. A cluster of folded points $\tilde{P}_1, ..., \tilde{P}_6$ is obtained (Figure 3b). (Recall that had the object been a D_3-symmetric one, all $\tilde{P}_i$'s would coincide.) (5) Average the folded points $\tilde{P}_i$, obtaining the average point $\hat{P}_1$ (Figure 3c):

$$\hat{P}_1 = \frac{1}{n_g}\sum_{i=1}^{n_g} g_i^{-1} P_i = \frac{1}{n_g}\sum_{i=1}^{n_g} \tilde{P}_i$$

(6) *Unfold* the average point $\hat{P}_1$ by applying on it each of the g_i operations and obtaining $\hat{P}_i$:

$$\hat{P}_i = g_i \hat{P}_1 \qquad i = 1, ..., n_g$$

The ordering $g_1, ..., g_6$ is followed in order to retrieve the original connectivity and ordering. A D_3-symmetric shape is obtained (Figure 3d). Note that whereas the nonsymmetric object is scaled to 1 (step 1), the symmetrized nearest shape is not necessarily so (see Section 4.2 for an example). (7) Calculate $S(G)$ according to eq 2. (8) *Minimize* the S value by repeating the folding–unfolding procedure (steps 4–7) for all orderings and all orientations of the group elements. This step is equivalent to finding the best cluster of folded points. In the present case, due to the cyclic connectivity, minimization is reduced to the two orderings mentioned in step 3. The optimal orientation is given analytically (see eq 17 in Section A.3) and in our case is simply 0° (i.e. σ is about the y-axis). See Appendix B for further details. The S value obtained by this procedure is the minimal distance to the desired symmetry group (the proof is given in Appendix A). Applying this procedure, we find that the $S(D_3)$ value of the hexagon in Figure 3a is 4.89.

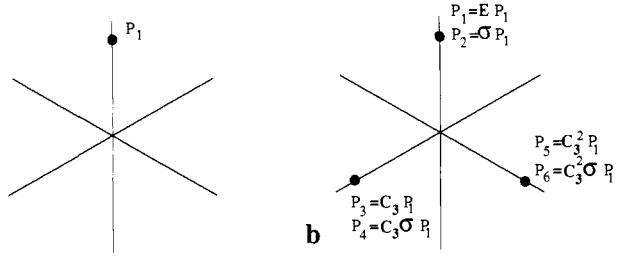

Figure 4. Creating a D_3-symmetric configuration of three points: (a) a single point P_1 is chosen on a symmetry axis; (b) unfolding P_1, one obtains three pairs of coinciding points having D_3-symmetry.

2.4. The Case $n_g = ln_p$ in Two Dimensions. In the previous section we described our method for the basic case in which the number of vertices, n_p, equals n_g, the number of elements in the symmetry group. However, a common situation is that the object under consideration has $n_p < n_g$ and, in particular, $n_g = ln_p$ with $l = 2, 3, ...$ For instance, the tetrahedral group T_d is composed of 24 elements (see below) but, in practice, it is applied in most cases to only four vertices, namely to the (distorted) tetrahedron vertices. In order to deal with $n_g = ln_p$, we first explain the process of creating a G-symmetric configuration of points using the unfolding process for this case. As an example, we continue with D_3 but this time with three vertices ($l = 2$, $n_p = 3$, $n_g = 6$). In general, we regard each of the n_p vertices of an $n_g = ln_p$ object as composed of l coinciding points. Thus, we regard a D_3-triangle as a hexagon in which each two vertices coincide. We construct such an object by following the general unfolding procedure (outlined in Section 2.2 and Figure 2) with one change: the point P_1, from which the shape is unfolded, is not in a general position (as in Figure 2b) but is selected on a symmetry element, σ in our case (Figure 4a). It then follows that the points P_1 and P_2 coincide; therefore, P_3 and P_4 coincide and so do P_5 and P_6—an equilateral triangle is obtained. In general, l is the number of symmetry operations which leave P_1 in place (in our case $l = 2$; E and σ). In effect we divide the elements of group G into n_p sets of l elements each, such that each set G_i contains the elements of the group which bring P_1 to P_i. Thus G_1 is the set of elements which leave P_1 in place, G_2 is the set of elements which, when applied, move

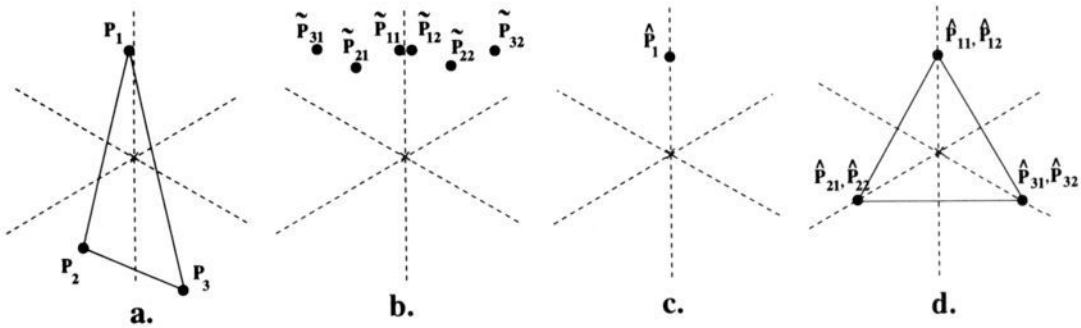

Figure 5. Measuring D_3-symmetry of a distorted triangle: (a) the original triangle; (b) folding the points in a—a cluster of points is obtained; (c) averaging the cluster of points in b—a single point located on a symmetry axis is obtained; (d) unfolding the averaged point in c—a D_3-symmetric triangle is obtained.

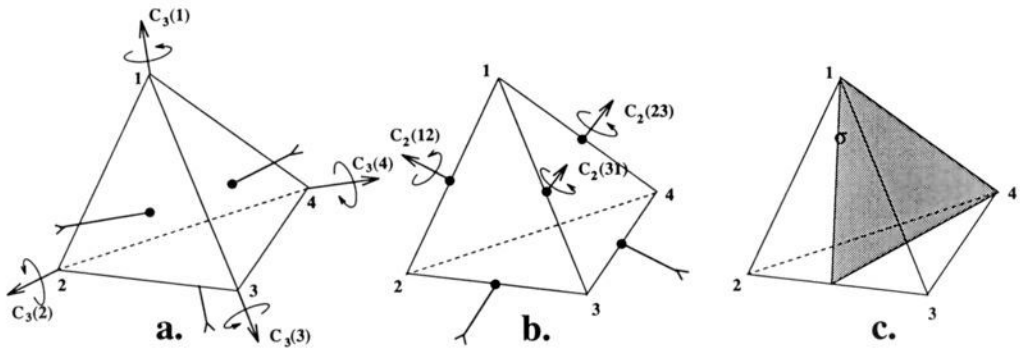

Figure 6. Symmetry elements of the T_d-symmetry group: (a) Four C_3 axes, each passing through the origin and a vertex; (b) three C_2 axes, each passing through the origin and bisecting a pair of opposite edges; (c) six reflection planes, each containing one of the edges and bisecting the opposite edge. A single plane (σ) is shown.

P_1 to P_2, etc. In our example (Figure 4b), $G_1 = \{g_1,g_2\}$, $G_2 = \{g_3,g_4\}$, $G_3 = \{g_5,g_6\}$.

Having detailed how a symmetric $n_g = ln_p$ shape is constructed, it is now clear how the folding–unfolding method is applied for evaluation of S values in such cases. The procedure is basically the same as in Section 2.2 with some modifications, as follows: (1) Translate the shape so that its centroid coincides with the origin (Figure 5a). (2) Translate the symmetry group to the origin (Figure 5a). (3) Select an ordering: (a) Determine l and divide G into n_p sets so that each such G_i contains l elements. (b) Select an ordering of the sets G_i of the symmetry group, that follows the connectivity of the P_i vertices. (In our example the ordering is as mentioned above: G_1, G_2, G_3). (4) Folding: For each vertex P_i, apply the inverse of the l elements of the set G_i, obtaining l folded points $\tilde{P}_{ij}$ with $j = 1, ..., l$. Thus, in our example, g_1^{-1} and g_2^{-1} are applied to P_1, obtaining $\tilde{P}_{11}$ and $\tilde{P}_{12}$ (Figure 5b), g_3^{-1} and g_4^{-1} are applied to P_2, obtaining $\tilde{P}_{21}$ and $\tilde{P}_{22}$, and so on. (5) Averaging: Average the n_g folded points as in Section 2.2. The average point $\tilde{P}_1$ will always lie, in the present case, on one or more symmetry axes or planes such that applying any element of G_1 will leave it in place (for proof, see Appendix A). In our case the averaged point, $\tilde{P}_1$, must rest on the reflection line (see Figure 5c). To understand this, notice that the six group elements can be paired: E, C_3, C_3^2 and their reflections: σE, σC_3, σC_3^2. On each point P_i we applied a single pair so that the obtained points P_{i1} and P_{i2} are related by σ (i.e., they are reflections of each other) and their average lies on the reflection line. (6) Unfolding: Following the unfolding step as described in Section 2.2., we notice that since $\hat{P}_1$ remains in place under the application of elements in G_1, the unfolded points will align in n_p sets of l points each. Thus in order to obtain n_p G-symmetric points, it suffices to unfold $\hat{P}_1$ by applying a single element from each set G_i. In our case $\hat{P}_1$ unfolds into three pairs of coinciding points: $\{\hat{P}_{11},\hat{P}_{12}\}$, $\{\hat{P}_{21},\hat{P}_{22}\}$, $\{\hat{P}_{31},\hat{P}_{32}\}$. By applying, for instance, only g_1, g_3, and g_5, $\hat{P}_1$, $\hat{P}_2$, and $\hat{P}_3$ are obtained (Figure 5d). (7) *Minimize* over the orderings of the sets G_i. (The previous case may be considered a specific example of this case where each group G_i consists of a single element g_i.) (8) Calculate S. $S(D_3)$ of the triangle in Figure 5a is 9.88.

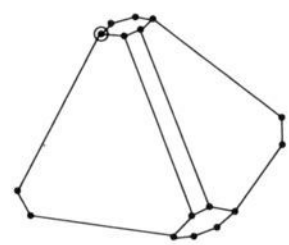

Figure 7. T_d-symmetric 24-polyhedron unfolded from the circled vertex.

3. The Nonsymmetric Tetrahedron

3.1. Measuring Tetrahedricity (T_d). We now apply the method described in Section 2.4 to study the symmetry measure of nonsymmetric tetrahedra, focusing first on tetrahedricity (T_d) and then (Section 3.2) on other symmetries. T_d symmetry includes the following symmetry elements (Figure 6): (1) four C_3 axes passing through the origin and through each of the four vertices (Figure 6a), (2) Three C_2 axes passing through the origin, each bisecting a pair of opposite edges (Figure 6b), (3) Six reflection planes, each containing one of the edges and bisecting the opposite edge. For our purpose only one of these reflections (σ) suffices (Figure 6c). The 24 elements of the T_d symmetry group are thus E (the identity), four C_3 rotations (denoted $C_3(1)$, ..., $C(4)$) and four C_3^2 rotations (denoted $C_3^2(1)$, ..., $C_3^2(4)$), three C_2 rotations (denoted $C_2(12)$, $C_2(23)$, $C_2(31)$), and all the elements obtained by multiplying these elements by σ. As in the D_3 hexagon case discussed in Section 2.3, here too, taking an arbitrary point and applying the 24 elements (in any order) will produce a T_d-symmetric 24-polyhedron (Figure 7). Thus, given 24 points in three-dimensional space, we can evaluate the T_d symmetry following the algorithms in Section 2.3. (In 3D, however, minimization over all orientations of the symmetry group is not analytic and an iterative process is used (see Appendix B).)

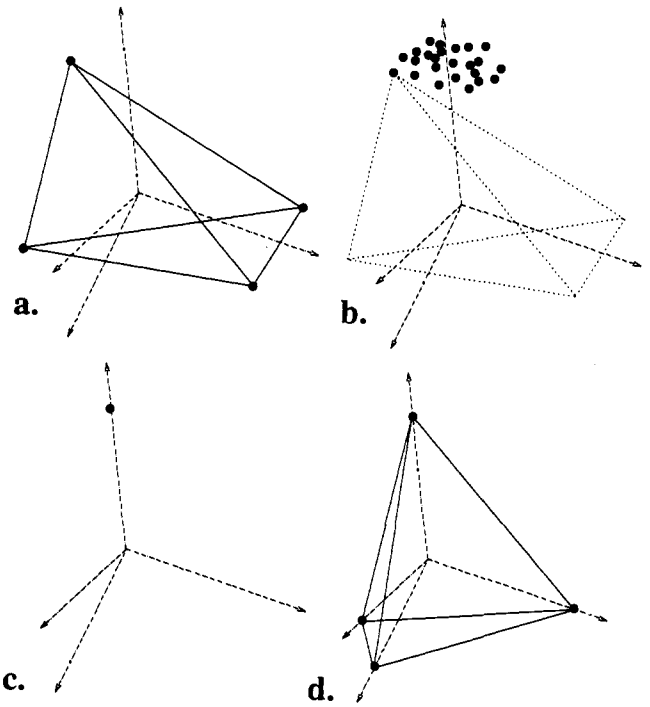

Figure 8. Measuring T_d-symmetry of a distorted tetrahedron: (a) the distorted tetrahedron; (b) folding the tetrahedron in a—a cluster of 24 points is obtained; (c) averaging the cluster of points in b; (d) unfolding the averaged point in c—a T_d-symmetric tetrahedron is obtained.

In most structural analyses, however, the 24-element T_d symmetry group is applied on four vertices in 3D, i.e. on the vertices of a (possibly) nonsymmetric tetrahedron. This case is analogous to the case described in the previous section for a D_3-symmetry of three points. Here one has $n_g = 24$, $n_p = 4$, and $l = 6$. Thus, in order to obtain a T_d-symmetric set of four points (four coinciding clusters of six points) from a single point P_1, it must be positioned so that six symmetry elements of the T_d symmetry group leave it in place. Such a point lies on a C_3 axis and on a σ plane (for example point 1 in Figure 6). The six elements of the group that leave P_1 in place are E, C_3, C_3^2 ($=C_3^{-1}$), and the three compositions σE, σC_3, σC_3^2. When one applies the rest of the elements on P_1, one finds that the four sets of elements are

$$G_1 = \{E, C_3(1), C_3^2(1), \sigma, \sigma C_3(1), \sigma C_3^2(1)\}$$

$$G_2 = \{C_3(3), C_3^2(4), \sigma C_3^2(2), \sigma C_3(4), C_2(12), \sigma C_2(31)\}$$

$$G_3 = \{C_3^2(2), C_3(4), \sigma C_3(3), \sigma C_3^2(4), C_2(31), \sigma C_3(12)\}$$

$$G_4 = \{C_3(2), C_3^2(3), \sigma C_3(2), \sigma C_3^2(3), C_2(23), \sigma C_2(23)\}$$

Thus unfolding P_1 with *any* one of the elements in G_2 will form P_2, with any of the elements in G_3, will form P_3, etc., creating a symmetric tetrahedron. Again, as in the previous cases, we use the construction of a symmetric shape as a guideline for the folding–unfolding procedure for evaluating the T_d-symmetry of any four vertices, as follows:

Figure 8a shows a distorted tetrahedron. Its centroid coincides with the origin, and the maximal distance to a vertex is scaled to 1. The four vertices are denoted $P_1, \ldots, P_4$, and a certain order of the sets G_i is selected, say $G_1, \ldots, G_4$, with a certain orientation of the T_d-symmetry group. Each P_i is then folded by applying the six elements of the group G_i, forming a cluster of 24 points (Figure 8b). These are averaged (Figure 8c) and unfolded by selecting one element from each G_i set. For example applying the group elements E, $C_3(3)$, $C_3^2(2)$, and $C_3(2)$ on $\hat{P}_1$, we obtain, respectively, points $\hat{P}_1$, $\hat{P}_2$, $\hat{P}_3$, and $\hat{P}_4$ (Figure 8d). S is then calculated from eq 2 and minimized over the two G_i orderings: G_1, G_2, G_3, G_4 and G_1, G_2, G_4, G_3 (the reason for needing only two orderings for this minimization is described in Appendix B.3). For each of the two orderings, one minimizes S over all orientations of the symmetry group. (In this case one can reduce the dimensionality of the minimization problem: see Appendix C.) The $S(T_d)$ values of the four tetrahedra in Figure 1 (Figure 1a = Figure 8a) are 16.87, 0.0, 4.73, and 25.0 for 1a, 1b, 1c, and 1d, respectively.[8]

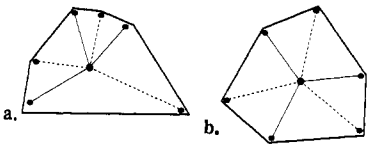

Figure 9. Measuring C_3-symmetry of a distorted hexagon: (a) dividing the points into two sets—two sets of three points each are obtained (one set is marked by the dashed line and the other by the solid line); (b) applying the folding–unfolding method to each set separately—two sets of perfectly C_3-symmetric points (a perfect C_3-hexagon) are obtained.

3.2. The General Case: Symmetry Content of Nonsymmetric Tetrahedra with Respect to Any Symmetry. We return now to the opening question: given any number of vertices in space, i.e. a general nonsymmetric polygon or polyhedron, what is its symmetry measure with respect to any symmetry group, subgroup, or class. The generalized approach (which includes the cases of Sections 2.3 and 2.4) is to divide the given points into n_s sets and to apply the folding–unfolding method separately on each set, while evaluating the S value over all the given points. For example, in the case where n_p is a multiple of n_g (i.e. $n_p = kn_g$), one divides the points into $n_s = k$ sets of n_g points each. On each set one performs the folding–unfolding method as described in Section 2.3. Thus, in Section 2.3 for D_3 we had $n_g = 6$, $n_p = 6$, $l = 1$, and $n_s = 1$; in Section 2.4 (a D_3-triangle) we had $n_g = 6$, $n_p = 3$, $l = 2$, and $n_s = 1$; and for T_d analysis of tetrahedra (Section 3.1) $n_g = 24$, $n_p = 4$, $l = 6$, and $n_s = 1$. The general case, however, will typically require the division of the n_p vertices into $n_s > 1$ vertices subsets, not necessarily of equal size but with (possibly different) l integer values for each of these sets. Once the division is made, each of the n_s subsets is symmetrized with respect to the desired symmetry group either according to the $n_g = n_p$ procedure (Section 2.3) or according to the $n_g = ln_p$ procedure (Section 2.4). Since each of the subsets is symmetrized with respect to the same symmetry group, one obtains a symmetrization of the full set of points. Division into subsets can be performed in various ways, but in order to preserve the cyclic connectivity of the original structure, the subsets must be interlaced (for further details, see Appendix B.1).

To illustrate these points, let us consider the hexagon shown in Figure 9a, for which we wish to evaluate the C_3-symmetry measure. This is done as follows: (1) Divide the points into two sets of three points each (Figure 9a—one set is marked by the dashed line and the other by the solid line). Here, $n_p = 6$, $n_g = 3$ (E, C_3, C_3^2), and $k = n_s = 2$ with $l = 1$ in both subsets. (2) Applying the folding–unfolding method to each set separately, obtaining two sets of perfectly C_3-symmetric points. A perfect C_3-hexagon is obtained (Figure 9b).[9] (3) Evaluate S by considering all six vertices. The symmetry measure of the hexagon in Figure 9a with respect to C_3-symmetry is $S(C_3) = 5.52$. This procedure can also be used for determining $S(\sigma)$, e.g. of a tetrahedron. This is of special interest because it is also a measure of achirality due to the existence of a symmetry plane (Part 4).

(8) Using the same methods and notations as described in this subsection, one can reformulate the [2 + 2] concerted reaction example presented in Part 1.[1] There we evaluated the symmetry of the configuration by evaluating the symmetry class of three reflective perpendicular planes. However, using the notations described here, one can, in fact, measure the D_{2h}-symmetry of the configuration. The D_{2h}-symmetry group consists of eight elements defined by a single C_2 axis and two reflection planes: σ_v passing through the C_2 axis and σ_h perpendicular to the C_2 axis. One has then $n_g = 8$, $n_p = 4$, and $l = 2$, and evaluating the symmetry measure follows the method described in this subsection.

(9) Note that the folding and unfolding performed with the symmetry group at the origin (step 2). In Part 1 we described the analogous case by folding about the centroid of each set G_i rather than about the centroid of all points (the origin). In Appendix A.2 we show the equality of these two cases.

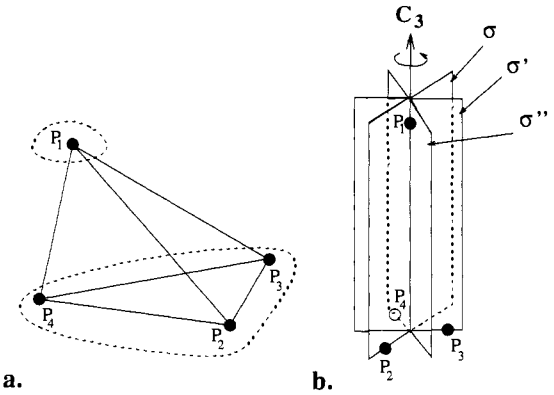

Figure 10. Evaluating C_{3v}-symmetry of a distorted tetrahedron (Figure 1a): (a) the four points are divided into two sets, one having three points (P_2, P_3, and P_4) and one having one point (P_1); (b) using the folding–unfolding method, one obtains two sets, each being C_{3v}-symmetric. The first set (three points) has each point on a reflection plane, and the second set (one point) is located on the C_3 element. Figure 1c is thus obtained with $S(C_{3v}) = 14.49$.

In the above example we divided the points into subsets of n_g points each. However, one may divide the points into sets having less than n_g points, specifically, into sets having a number of points which is a divisor of n_g. In this case we apply on each set the folding–unfolding method as described in Section 2.4 for the case $n_g = ln_p$. Extending this idea further, we need not divide the points into sets of equal size. For example, let us evaluate the $S(C_{3v})$ of the tetrahedron in Figure 1a. The C_{3v}-symmetry group has $n_g = 6$ elements: E, C_3, C_3^2, and all the multiplications of these elements by σ. In order to evaluate the C_{3v}-symmetry of the four points, one divides them into two sets; one having three points (P_2, P_3, and P_4 in Figure 10a) and one having one point (P_1 in Figure 10a). For the first set, one has $n_p = 3$ and $l = 2$ and the folding–unfolding is applied as in Section 2.4. In the second set, one has $n_p = 1$ and $l = 6$ and the folding–unfolding is applied as in Section 2.3. One obtains two sets, each being C_{3v}-symmetric, the first set consisting of three points each on a reflection plane, and the second set consisting of a single point located on the rotation axis and on the reflection planes (Figure 10b). After minimization one obtains $S(C_{3v}) = 14.49$. We use a similar procedure to evaluate $S(D_{3h})$ of Figure 1a. Here $n_g = 12$ and $n_s = 2$ with one subset having $n_p = 1$ and $l = 12$ and the other set having $n_p = 3$ and $l = 4$, resulting in $S(D_{3h}) = 19.7$.

Finally, we comment on the case where the connectivity constrains the division into sets so that no possible division exists. For example a cyclic connected configuration of six points cannot be divided into sets having a divisor of five points (for measuring C_5 symmetry). To overcome these specific cases, one may physically duplicate or eliminate one or more of the given points. Though this has no physical or chemical interpretation, it does give a geometric solution. Thus in order to measure a triangle with respect to C_4-symmetry, one duplicates one of its vertices; and in order to measure C_{3h}-symmetry of four points in space, one duplicates one of the points twice. The symmetrized objects will be with four and six vertices, respectively.

3.3. Tetrahedricity of a Tetrahedron with a Central Atom. Most tetrahedral structures include a central atom. While branched connectivities are not the topic of this report (see Part 4), we do bring here a preliminary case, that of the branched tetrahedron, for the sake of completeness of discussion of that structure. We replace, then, the polyhedral presentation of a tetrahedron with a branched connected set of five points P_1, ..., P_5 as shown in Figure 11a, which models a tetrahedron with a central atom, and applying the CSM folding–unfolding method described in Section 3.2 to evaluate its T_d symmetry. The connectivity constrains the division of points into subsets and restricts the center point (P_5) to be a one-point set. We thus divide the points into $n_s = 2$ subsets; $\{P_1, ..., P_4\}$ and $\{P_5\}$. The

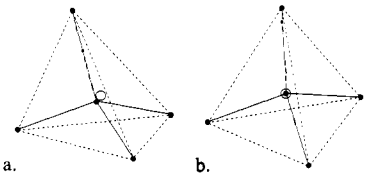

Figure 11. (a) Distorted tetrahedron with a central atom, analyzed as a connected configuration of five points. The open circle marks the centroid of the configuration. (b) Closest T_d-symmetric configuration.

first subset has $n_g = 24$, $n_p = 4$, and $l = 6$, and the second subset has $n_g = 24$, $n_p = 1$, and $l = 24$. The closest symmetric configuration will have point P_5 relocated to a position where all 24 of the T_d symmetry group elements leave it in place. The only such position is at the origin (centroid of the configuration marked as an open circle in Figure 11), where all symmetry planes and axes intersect. Points P_1, ..., P_4 will be relocated to form a perfect T_d-symmetric configuration of four points; i.e., each point will lie on a C_3 rotation axis (see Figure 11b). $S(T_d)$ (or any $S(G)$) is then calculated by considering the full set of $\hat{P}_1$, ..., $\hat{P}_5$.

4. Further Examples

4.1. Distorted Tetrahedra: Reanalysis of Murray-Rust, Bürgi, and Dunitz.[3] Our discussion is not complete without mentioning the study of Murray-Rust et al. on distorted T_d.[3] Their method[3,10] is based on evaluation of the distortion vector that creates the given configuration from an apriori chosen reference structure. The coordinates of the distortion vector are given with respect to a set of basis vectors, denoted the symmetry displacement coordinates. Each such symmetry coordinate is a linear combination of the internal coordinates of the reference structure which transforms according to irreducible representation of the point symmetry group of the reference structure. The symmetry of a distorted configuration is then given by a vector, i.e. a collection of symmetry displacement coordinate values. To intuitively understand the coordinate values and the measure of symmetry thus obtained, one needs to analyze the values with respect to each other and with respect to the symmetry coordinates they correspond to. Thus for the distorted tetrahedral structure of PO_4 found in $Cd_2P_2O_7$ with the reference structure being a perfect tetrahedron with arm lengths of 1.534 Å, the authors obtained a set of 10 symmetry displacement coordinate values.[11]

We now analyze the same distorted phosphate tetrahedron using our method. We first recall that our method evaluates the distance from tetrahedricity and not from a specific tetrahedron and that, rather than reporting the deviation in terms of a table of many coordinates, we provide a single $S(T_d)$ value. To obtain it, the 3D position coordinates of the four oxygens and phosphorus were taken from ref 12 (also used by Bürgi et al., p 1790 in their paper) as:

$$P_1 = (0.0 \quad 0.0 \quad 1.645)$$

$$P_2 = (0.0 \quad 1.518\,860 \quad -0.347\,028)$$

$$P_3 = (-1.286\,385 \quad -0.700\,083 \quad -0.391\,603)$$

$$P_4 = (1.179\,085 \quad -0.755\,461 \quad -0.372\,341)$$

with an additional center point 0.0. We apply the folding method as described in Section 3.3. The symmetry measure obtained in this example is $S(T_d) = 0.17$, and the closest symmetric shape is a regular tetrahedron with arm length 1.537 Å.

(10) See also: Luef, W.; Keese, R.; Bürgi, H. B. *Helv. Chim. Acta* **1987**, *70*, 534 and earlier references cited therein.

(11) Following the notation in Bürgi et al.:[3]

$D_1(A_1) = 0.016$ Å	$D_{2a}(E) = -2.59°$	$D_{4a}(T_2) = -6.84°$
$D_{3a}(T_2) = 0.019$ Å	$D_{2b}(E) = 1.89°$	$D_{4b}(T_2) = -8.37°$
$D_{3b}(T_2) = 0.077$ Å	$D_5(A_1) = -0.66°$	$D_{4c}(T_2) = -5.81°$
$D_{3c}(T_2) = 0.010$ Å		

(12) Calvo, C. *Can. J. Chem.* **1969**, *47*, 3409–3416.

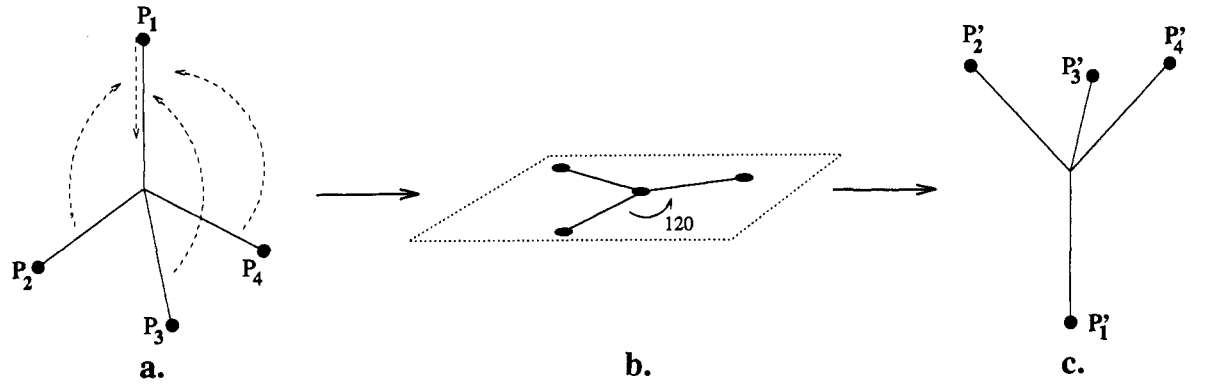

Figure 12. Modeling a Walden-type inversion: (a) a perfect tetrahedral configuration of four points undergoes inversion as shown; (b) the intermediate planar configuration; (c) the inverted tetrahedron.

In a further example, Murray-Rust et al. used the symmetry coordinates to evaluate the threefold axes of 1-methyl-1-silabicycloheptane (Section 5 in ref 3). They found that the distorted SiC_4 structure (Figure 5 in ref 3) is better described with the threefold axis passing through one vertex (point C_1 in their notation) rather than through another (C_2 in their notation). Using the CSM method with respect to C_{3v}-symmetry, we easily support their conclusion as follows: Given the coordinates

$$P_1 = C_1 = (0.0 \quad 0.0 \quad 1.645)$$

$$P_2 = C_2 = (0.0 \quad 0.874\,619\,71 \quad -0.484\,609\,62)$$

$$P_3 = C_3 = (0.751\,286\,05 \quad -0.393\,388\,92 \quad -0.529\,919\,26)$$

$$P_4 = C_4 = (0.751\,286\,05 \quad -0.393\,388\,92 \quad -0.529\,919\,26)$$

with Si at the origin, the $S(C_{3v})$ of the configuration was calculated by the method described in Section 3.2 and found to be $S(C_{3v}) = 0.02$ when the threefold axis passes through point C_1, compared to $S(C_{3v}) = 1.16$ when the threefold axis is constrained to pass through point C_2. Using the folding method, we can also measure the C_{3v}-symmetry of the configuration with the constraint that three of the configuration points are equatorial. In this case the S value increases to 5.26, with the threefold axes passing through point C_2.

4.2. Fluxional Molecules. Our method allows one to follow continuous symmetry changes that fluctuating molecules (amines, phosphines, metal complexes, etc.) experience through the flip-flopping between configurations. As a general case we again take the tetrahedron and follow its symmetry behavior in a Walden-type inversion, leading from one enantiomer to the other through a planar transition configuration. Let us then consider, for the sake of simplicity of demonstration, a perfect tetrahedral configuration of four points with arm length 1 (Figure 12a) which is inverted to the configuration in Figure 12c through the configuration in Figure 12b. The inversion process is modeled by rotating three of the points (P_2, P_3, and P_4) about the origin toward the inverted positions (P_2', P_3', and P_4') and by linearly moving the fourth point (P_1) to its inverted position (P_1'). The movement from the initial tetrahedron to its inverted configuration is considered as one cycle. We evaluate the symmetry measure of the configuration during the inversion process with respect to T_d-symmetry and with respect to σ-symmetry along the plane perpendicular to the vector from the origin to P_1. The results are displayed in Figure 13. The solid line displays the symmetry measure with respect to T_d-symmetry, and as can be seen, the configuration has minimum value ($S(T_d) = 0$) at the start and end of the cycle and a maximum symmetry measure ($S(T_d) = 25.0$) at midcycle when the configuration is restricted to a single plane (Figure 12b). The dashed line in Figure 13 displays the symmetry measure with respect to mirror-symmetry along the prescribed plane. The minimum symmetry measure ($S(\sigma) = 0$) is obtained at midcycle when the configuration is restricted to the symmetry plane (Figure 12b), and the maximum symmetry measure ($S(\sigma) = 33.33$) is obtained at the beginning and end of the cycle when the configuration is perfectly tetrahedral. An interesting observation is that the maximal T_d-symmetry measure is smaller than the maximal mirror-symmetry measure; in other words, the planar structure (Figure 12b) is closer to tetrahedricity (and not necessarily to the original tetrahedron of arm length 1) than the tetrahedron is to a plane. The tetrahedron closest to the shape having the maximal ($S(T_d) = 25.0$) value has an arm length of $1/\sqrt{2}$, and the planar structure representing $S(\sigma) = 33.33$ has an arm length of $2\sqrt{2}/3$ (these values are obtained as described in note 13). We thus demonstrate another important feature of our approach: it can be used not only for evaluating symmetry content of nonsymmetric shapes but also for evaluating distance between perfectly symmetric shapes.

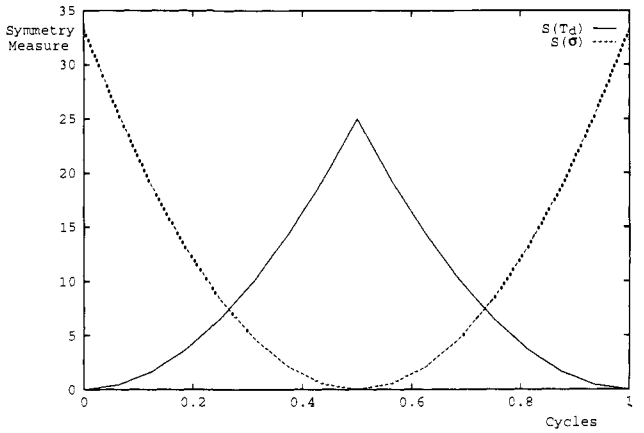

Figure 13. Symmetry measures of the tetrahedron in Figure 12a during a Walden inversion with respect to T_d-symmetry (solid line) and with respect to σ-symmetry along the plane perpendicular to the vector from the origin to P_1 (dashed line).

4.3. The Vibrating Tetrahedron. The CSM method can also follow continuous symmetry changes due to vibrations. We demonstrate it on a vibrating tetrahedron. The model vibration is that of a perfect tetrahedron in which one or more of the vertices moves along the vector connecting the origin to the tetrahedral points (see Figure 14) in a cyclic sinusoidal movement. On this dynamical object we now perform an $S(T_d)$ analysis. For example, Figure 14 shows the results for a vibration that extends two arms

(13) The planar structure representing $S(\sigma) = 33.33$ is obtained by projecting the perfect tetrahedron onto the plane. Thus (Figure 12) point P_1 projects onto the origin, and point P_2, having coordinates $(0, 2\sqrt{2}/3, -1/3)$, projects onto $(0, 2\sqrt{2}/3, 0)$, and one obtains an arm length of $2\sqrt{2}/3$. The tetrahedral structure representing $S(T_d) = 0.25$ is obtained by moving points of the planar configuration (Figure 12b). The point at the origin moves perpendicular to the plane a distance X. The three points each move so that they are at a distance X from the origin and at an angle 109.47° to the perpendicular (or 19.47° to the plane). Using the law of cosines, one has the distance moved by each of the three points equal to $1 + X^2 - 2X\cos(19.47)$ or $1 + X^2 - 4\sqrt{2}X/3$. Since the symmetry measure minimizes the distance squared moved by the points, one must minimize $4X^2 - 8\sqrt{2}X + 3$. Thus one obtains at minimum $X = 1/\sqrt{2}$.

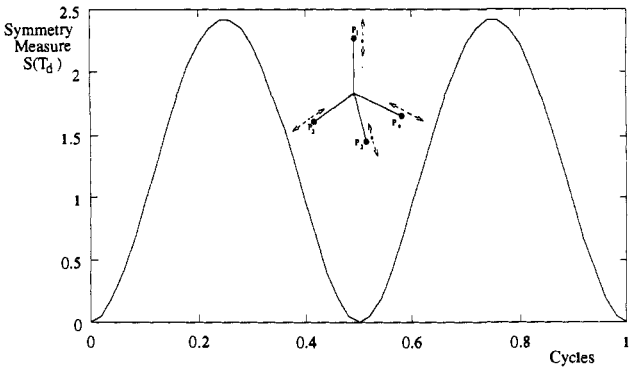

Figure 14. T_d-symmetry measure of a vibrating tetrahedron. In this model the vibration extends two arms to a maximum of 25% of the original arm length and the other two arms to a maximum of 25%, with a phase delay of 180°. Insert: The vibration motion.

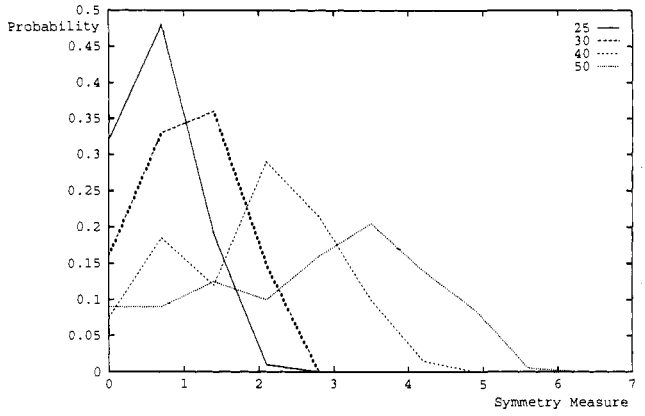

Figure 15. Probability distribution of the symmetry measure $S(T_d)$ of a vibrating tetrahedron with random phase shifts associated with each arm. The distributions are given for several values of the maximum extension of the vibrations: 25, 30, 40, and 50% of original arm length.

to a maximum of 25% of the original arm length and the other two arms to a maximum of 25% at a phase delay of 180°. The CSM behavior indeed follows physical intuition: the symmetry measure reaches a maximum at the two extremes (two arms extended and two arms contracted to the maximum) and returns to a minimum at midcycle when the configuration is the original tetrahedral configuration. The overall shape of the curve of S follows the sinusoidal shape of the vibration.

We now take the vibrating-molecule example to introduce yet another concept: the *probability* of finding a dynamic molecule at a certain $S(G)$ value. Thus, consider, for instance, a tetrahedron with a central heavy atom and with four light atoms anchored to it. Here, the vibration of each of the four arms is only weakly linked to the others. For the sake of argument, let us assume that the vibrations are all of the same frequency but are independent of each other in terms of phase. Then, the vibrating tetrahedron can assume any phase locking, where each arm vibrates with a random phase shift relative to the others. It is then evident that, for such a dynamically random object, one is better off by describing its symmetry in terms of a probability distribution to find it at a certain S value. Given various maxima extensions of the tetrahedral arms during the vibration and fixing the central point, we evaluate $S(T_d)$ for different randomly chosen phase shifts and plot the probability distribution of the symmetry measure. Figure 15 shows the result for several values of the maximum extension of the vibrations (evaluation was performed on 200 samples for each maximum extension value). As can be seen, the distribution of S values is more widespread when the vibration has a greater maximum extension, the probability of finding the molecule at perfect T_d decreases, and the expectation value of S increases.

4.4. Rotating Tetrahedra: Ethane. Another basic mechanism which strongly affects molecular symmetry is intramolecular

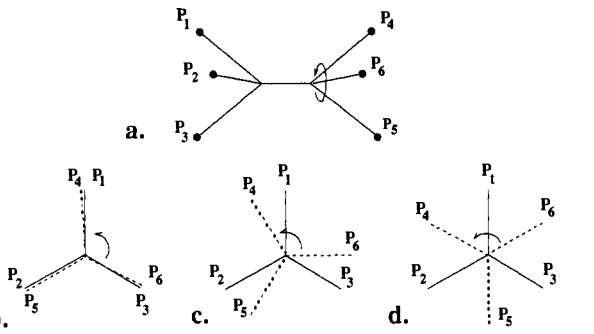

Figure 16. Modeling the C–C rotation in ethane: (a) only the right hand tetrahedron moves; (b) the cycle starts with the eclipsed D_{3h} rotamer (Figure 17); (c) one of the six chiramers (see text); (d) the D_{3d} staggered rotamer.

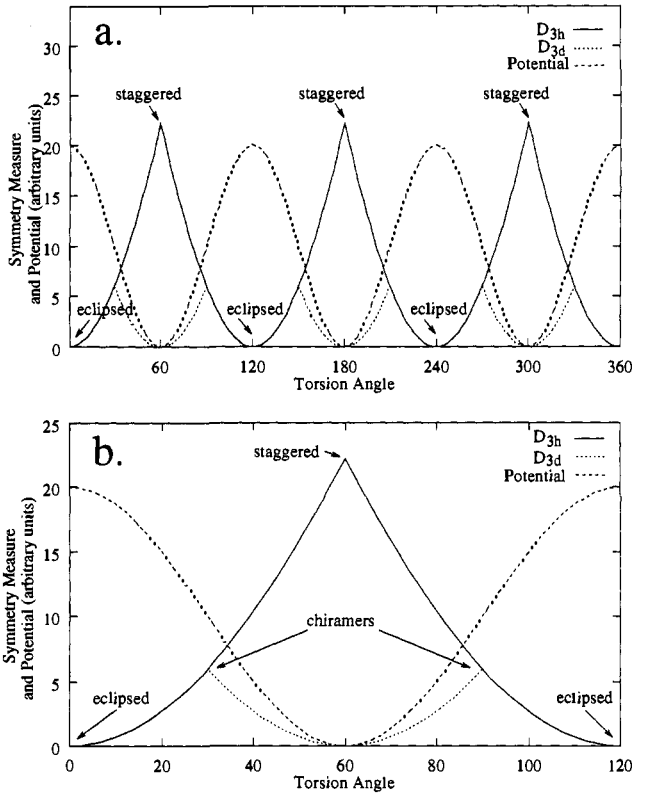

Figure 17. (a) Periodic variations in the symmetry measure of D_{3h} and D_{3d} for rotating ethane. A sinusoidal potential is also shown. (b) Detail of a. Notice the 30° and 90° chiramers.

rotation. Consider, for instance, one of the most basic examples, namely, the rotation of the two ethane tetrahedra around the C–C bond (Figure 16). Current wisdom allows an exceedingly poor description of that process from the symmetry point of view: Ethane is D_{3d} when staggered (Figure 16d), D_{3h} when eclipsed (Figure 16b), and D_3 anywhere in between, including the rotamer which is only 1° away from any of the extremes. Doesn't physical intuition dictate that it is more natural to ask about that 1° rotamer how much D_{3h} or D_{3d} it contains or, for that matter, how much D_{3h} and D_{3d} exists at any point in a full 360° cycle? As has already become evident throughout this paper, the CSM method allows one to select any symmetry group and follow its gradual changes along such a full 360° cycle of rotation. We demonstrate it on two perfect tetrahedral structures connected along one of the tetrahedral arms and rotating with respect to each other around the connecting arm. We model this by stabilizing one of the tetrahedra and rotating the other, beginning the cycle with the two tetrahedra perfectly aligned (eclipsed) and rotating the second tetrahedron anticlockwise. Figure 17 displays the result where the S value is given as a function of the cycle. (Figure 17a shows a full 360° cycle, and Figure 17b shows a detail.) The following

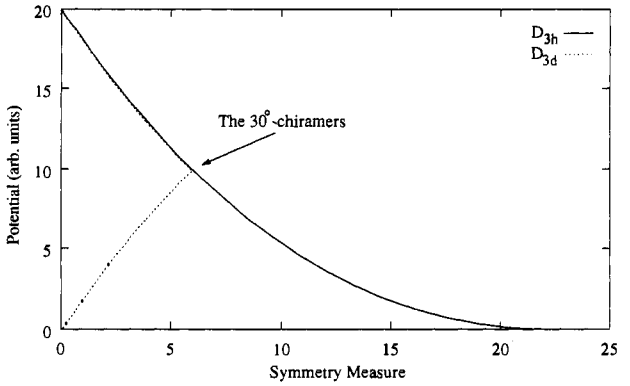

Figure 18. C–C rotation in ethane presented in the plane of potential vs symmetry measure, D_{3d} and D_{3h}.

observations are made: (1) The D_{3h} profile of the rotation reaches a higher S value than the D_{3d} profile. This is a reflection of the fact that the D_{3h}-group contains more elements than the D_{3d}-group. (2) The maximal $S(D_{3h})$ value is at the 60° staggered rotamer which is the farthest away from the perfect D_{3h} eclipsed rotamer (0°, $S(D_{3h}) = 0$). Three such maxima are observed in a full cycle, corresponding to the three staggered rotamers. (3) The D_{3d} behavior is very interesting and points to the importance of a rotamer (Figure 16c) which is neither eclipsed nor staggered, but in between, at 30° + n·60°. The $S(D_{3d})$ value is zero both for the staggered and for the eclipsed rotamers, because D_{3d} is a subgroup of D_{3h}. (The inequality $S(G_1) \leq S(G_2)$ always holds if G_1 is a subgroup of G_2.) Indeed, up to 30°, $S(D_{3h}) = S(D_{3d})$, but beyond that and up to 90° they depart. To understand it, we recall that the CSM method searches for the minimal movement to attain the desired symmetry, which in the case of D_{3d} is the *eclipsed* rotamer up to 30°, in contradistinction with the current convention which attaches D_{3d} only to the staggered rotamer. However, beyond 30° (and up to 90°) the staggered rotamer is indeed the closest to D_{3d}. We term these special chiral (!) rotamers at 30° + n·60° *chiramers* (Figure 16c). There are six of these in a full cycle, compared to three eclipsed and three staggered (which are of course achiral).

Finally, we wish to make a brief preliminary comment on what seems to us an important application of our approach: Many thermodynamic and kinetic quantities vary cyclically with internal rotations. A commonly presented quantity is the (repulsion) potential. It is then interesting to see how this property varies with the *symmetry* rather than with the traditional torsion angle. The results for a model sinusoidal potential (Figure 17) are shown in Figure 18. Let us first detail how the potential follows this new process coordinate: The D_{3h} potential line varies smoothly with S, starting at the eclipsed $S = 0$ value and dropping to zero potential at the staggered $S = 22.22$ value; then it reverses and climbs back up to the maximum potential, completing 120° of the cycle. This drop and rise in potential along the symmetry coordinate is repeated continuously, completing a full 360° cycle. The D_{3d} potential line behaves differently: up to $S = 5.95$ it follows the D_{3h} line, but then as the potential continues to drop, so does the S value, reaching the minimum potential at the staggered $S = 0$ value (the line then climbs back up). Perhaps most notable is that the lines of the two symmetry groups bifurcate at the 30° chiramers. What does such symmetry/potential bifurcation mean? In general, it may mean that for symmetry-governed processes, such a crossing point is where the process may select to proceed one way or the other depending on which symmetry is preferred.

5. Summary

We have shown how to study symmetry as a continuous property. Using the CSM method, it is now possible to evaluate quantitatively the amount of symmetry in a nonsymmetric configuration, to find the closest symmetric shape of a given configuration, and to compare symmetries of different configurations, all of these both in static and dynamic structures. The method is general and can be used to evaluate any symmetry of any configuration. We presented many examples and applications of the CSM method, concentrating on tetrahedral structures and evaluating the amount of tetrahedricity (T_d) and other symmetries in nonsymmetric tetrahedra, and on continuously changing symmetries in fluctuating, vibrating, and rotating tetrahedra. Subsequent reports deal both with further extensions of the CSM approach to other general symmetry issues and with the transition from CSM models to real systems.

Acknowledgment. We thank S. Shaik for illuminating discussions and continuous encouragement.

Appendix A. Mathematical Derivations

A.1. Orbits. We first review some basic definitions required for our proofs and derivations. The *orbit* of x under a group G is the set $\{gx | g \in G\}$. x and y belong to the same orbit if $y = gx$ for some $g \in G$. Given a finite group G and given an ordering of its elements, $g_1, g_2, \ldots, g_m$, the *orbit* under G of a point x in Euclidean space is $x_1, \ldots, x_m$ such that $x_i = g_i x$ for $i = 1, \ldots, m$. If $g_1 = e$ (the identity element of G) then $x_i = g_i x_1$ for $i = 1, \ldots, m$.

Lemma 1. *The centroid of an orbit of finite point-symmetry group G is invariant under G.*

A point $x \in X$ is a *general point* (or is in general position) with respect to G if for all $g \in G$, $g \neq e$ (where e is the identity in G) we have $gx \neq x$.

Lemma 2. *If x is a general point with respect to G, then all points in the orbit of x are general points. Furthermore for* $g_1, g_2 \in G$, $g_1 \neq g_2 \Rightarrow g_1 x \neq g_2 x$.

Thus if x is a general point, its orbit contains $N(G)$ different points ($N(G)$ is the number of elements in group G).

Lemma 3. *If the orbit of x has a point in common with the orbit of y under G, then the two orbits are equal.*

For any $x \in X$ the group $G^x = \{g \in G | gx = x\}$ is called the *isotropy subgroup* of G at x and it contains all elements of group G that leave x invariant. If x is a general point, its isotropy subgroup contains a single element of G—the identity, i.e. $G^x = \{e\}$.

Lemma 4. *If G is finite, the number of different points in the orbit containing x is* $N(G)/N(G^x)$. (Proof is immediate from the 1–1 relationship between points in the orbit of x and the left cosets of G^x. Each left coset of G^x consists of all elements of G that map x to a specific point y.)

A.2. Proof of the Folding Method. As described in Section 1, the CSM of a set of points with respect to a given symmetry group G is evaluated by first finding the set of points which is G-symmetric and which is closest to the given set in terms of the average distance squared. We must thus prove that the folding method indeed finds the closest symmetric set of points.

Given a finite point-symmetry group G centered at the origin and an ordering of its m elements $\{g_1 = e, \ldots, g_m\}$ and given m general points $P_1, \ldots, P_m$, find m points $\hat{P}_1, \ldots, \hat{P}_m$ and find a rotation matrix R and translation vector w such that $\hat{P}_1, \ldots, \hat{P}_m$ form an ordered orbit under G' (where G' is the symmetry group G rotated by R and translated by w) and bring the following expression to a minimum:

$$\sum_{i=1}^{m} \|P_i - \hat{P}_i\|^2 \tag{3}$$

Since G has a fixed point at the origin and G' has the centroid of orbit $\hat{P}_i$ as a fixed point (see Lemma 1), we have that w is the centroid of orbit $\hat{P}_i$:

$$w = \frac{1}{m}\sum_{i=1}^{m} \hat{P}_i \tag{4}$$

(Note that, in the cases where the fixed points of G form an axis or plane, w can be any vector moving the origin to the (rotated)

axis or plane passing through the centroid of orbit $\hat{P}_i$. Thus also in these cases w can be considered the centroid of orbit $\hat{P}_i$.)

The points $\hat{P}_1, ..., \hat{P}_m$ form an orbit of G', thus the following must be satisfied:

$$\hat{P}_i = g_i'\hat{P}_1 = R^t g_i R(\hat{P}_1 - w) + w \quad i = 1, ..., m \tag{5}$$

where g_i' is the matrix representation of the ith symmetry element of G' and is equal to the ith symmetry element g_i of G rotated by R and translated by w.

Using Lagrange multipliers with eqs 3–5, we must minimize the following:

$$\sum_{i=1}^{m}\|P_i - \hat{P}_i\|^2 + \sum_{i=1}^{m}\lambda_i^t(\hat{P}_i - R^t g_j R(\hat{P}_1 - w) - w) + \epsilon\left(w - \frac{1}{m}\sum_{i=1}^{m}\hat{P}_i\right)$$

where ϵ and λ_i for $i = 1, ..., m$ are the Lagrange multipliers.

Equating the derivatives to zero we obtain

$$\sum_{i=1}^{m}(P_i - \hat{P}_i) = 0 \tag{6}$$

and using the last constraint (eq 4) we obtain

$$w = \frac{1}{m}\sum_{j=1}^{m}P_j \tag{7}$$

i.e. the centroid of $P_1, ..., P_m$ coincides with the centroid of $\hat{P}_1, ..., \hat{P}_m$ (in terms of the symmetry measure defined in Section 1, the centroid of a configuration and the centroid of the closest symmetric configuration are the same for any point-symmetry group G).

Noting that $R^t g_i R$ for $i = 1, ..., m$ are isometries and are distance preserving, we have from the derivatives

$$\sum_{i=1}^{m} g_i'^t(P_i - \hat{P}_i) = \sum_{i=1}^{m} R^t g_i^t R(P_i - \hat{P}_i) = 0$$

Expanding using the constraints we obtain

$$m\hat{P}_1 - mw = \sum_{i=1}^{m} R^t g_i^t R P_i - \sum_{i=1}^{m} R^t g_i^t R w$$

or

$$\hat{P}_1 - w = \frac{1}{m}\sum_{i=1}^{m} R^t g_i^t R(P_i - w) \tag{8}$$

The geometric interpretation of eq 8 is the folding method, thus proving that the folding method results in the G-symmetric set of points closest to the given set.

Given $n = qm$ points (i.e. q sets of m points) $\{P_1^j, ..., P_m^j\}$ for $j = 1, ..., q$, we obtain the result given in eq 8 for each set of m points separately, i.e. for $j = 1, ..., q$:

$$\hat{P}_1^j - w = \frac{1}{m}\sum_{i=1}^{m} R^t g_i^t R(P_i^j - w) \tag{9}$$

where $w = 1/n\sum_{j=1}^{q}\sum_{i=1}^{m}P_i^j$ is the centroid of all n points.

The folding method for the cases described in Section 2.4, where the number of points is a divisor of the number of elements in the symmetry group ($n_g = ln_p$), is similarly derived: Given n points $P_1, ..., P_n$, a finite point-symmetry group G centered at the origin and having $m = ln$ elements, and a subgroup G_1 of G having l elements and an ordering $\{G_1, ..., G_n\}$ of the left cosets of G_1, we would like to find n points $\hat{P}_1, ..., \hat{P}_n$ and find a rotation matrix R and translation vector w such that $\hat{P}_1, ..., \hat{P}_n$ form an ordered orbit under G' (where G' is the symmetry group G rotated by R and translated by w) such that G_1 leaves $\hat{P}_1$ invariant and the following expression is brought to a minimum:

$$\sum_{i=1}^{n}\|P_i - \hat{P}_i\|^2 \tag{10}$$

As before we have

$$w = \frac{1}{n}\sum_{i=1}^{n}\hat{P}_i \tag{11}$$

Additionally the elements of G_1 (denoted $g_{11}, ..., g_{1l}$) leave $\hat{P}_1$ in place; thus we have

$$\hat{P}_1 = R^t g_{1j} R(\hat{P}_1 - w) + w \quad \text{for } j = 1, ..., l \tag{12}$$

and since $\hat{P}_1, ..., \hat{P}_n$ form an orbit of G', we have for each coset G_i

$$\hat{P}_i = R^t g_{ij} R(\hat{P}_1 - w) + w \quad \text{for } i = 1, ..., n, j = 1, ..., l \tag{13}$$

Using Lagrange multipliers with eqs 11–13, we must minimize the following:

$$\sum_{i=1}^{n}\|P_i - \hat{P}_i\|^2 + \sum_{i=1}^{n}\sum_{j=1}^{l}\lambda_{ij}^t(\hat{P}_i - R^t g_{ij} R(\hat{P}_1 - w) - w) + \epsilon\left(w - \frac{1}{n}\sum_{i=1}^{n}\hat{P}_i\right)$$

where ϵ, λ_{ij} for $i = 1, ..., n$ and $j = 1, ..., l$ are the Lagrange multipliers. Solving as before we obtain

$$w = \frac{1}{n}\sum_{i=1}^{n}P_i = \frac{1}{m}\sum_{i=1}^{n}lP_i \tag{14}$$

and

$$\hat{P}_1 - w = \frac{1}{m}\sum_{i=1}^{n}\sum_{j=1}^{l} R^t g_{ij}^t R(P_i - w) \tag{15}$$

The geometric interpretation of eqs 14 and 15 is the folding method for the case described in Section 2.4 where each point P_i for $i = 1, ..., n$ is duplicated times. Thus we proved that the folding method results in the G-symmetric set of points closest to the given set.

A.3. Finding the Optimal Orientation in 2D. Following the derivation in Appendix A.2 we derive, here, an analytic solution to finding the orientation (rotation matrix R) which minimizes eq 3 under the constraints given in eqs 4 and 5. In Part 1[1] (Appendix A.2) we gave the derivation for the specific case of the D_1 group having the two elements $\{E, \sigma\}$.

In 2D there are two classes of point-symmetry groups: the class C_n having rotational symmetry of order n and the class D_n having rotational symmetry of order n and n reflection axes. The problem of finding the minimizing orientation is irrelevant for the C_n-symmetry groups, and R is usually taken as I (the identity matrix). We derive here a solution for the orientation in the case where G is a D_n-symmetry group.

The $2n$ elements of the D_n-symmetry group $(g_1, ..., g_{2n})$ are the n elements $E, C_n^1, C_n^2, ..., C_n^{n-1}$ $(g_1, ..., g_n$, respectively) and the n elements obtained by applying a reflection σ on each of these elements: $\sigma, \sigma C_n^1, \sigma C_n^2, ..., \sigma C_n^{n-1}$ $(g_{n+1}, ..., g_{2n}$, respectively). We denote the orientation of the symmetry group as the angle θ between the reflection axis and the y-axis. Thus

$$R = \begin{pmatrix}\cos\theta & \sin\theta \\ -\sin\theta & \cos\theta\end{pmatrix}$$

Without loss of generality we assume the centroid (w) is at the origin. The matrix representation of the rotational elements of D_n is then $g_i' = R^t g_i R = g_i$ for $i = 1, ..., n$. The matrix representation of the operation σ is given by

$$R_f = R^t g_{n+1} R = \begin{pmatrix}\cos\theta & -\sin\theta\\ \sin\theta & \cos\theta\end{pmatrix}\begin{pmatrix}-1 & 0\\ 0 & 1\end{pmatrix}\begin{pmatrix}\cos\theta & \sin\theta\\ -\sin\theta & \cos\theta\end{pmatrix} = \begin{pmatrix}-\cos 2\theta & -\sin 2\theta\\ -\sin 2\theta & \cos 2\theta\end{pmatrix}$$

and $g_i' = R_f g_{i-n}$ for $i = n+1, \ldots, 2n$. Thus from Appendix A.2 we must minimize the following over θ:

$$\sum_{i=1}^{2n} \|P_i - \hat{P}_i\|^2 = \sum_{i=1}^{2n} \|g_i'\,{}^tP_i - \hat{P}_1\|^2$$

$$= \sum_{i=1}^{2n} \|g_i'\,{}^tP_i - \frac{1}{2n}\sum_{j=1}^{2n} g_j'\,{}^tP_j\|^2$$

$$= \sum_{i=1}^{2n} \|\sum_{i=1}^{n} g_j{}^tP_j + \sum_{j=n+1}^{2n} R_f g_{j-n}\,{}^tP_j - 2n g_i'\,{}^tP_i\|^2 \qquad (16)$$

Denoting by x_i, y_i the coordinates of the point $g_i'\,{}^tP_i$ and taking the derivative of eq 16 with respect to θ we obtain

$$\tan 2\theta = \frac{\sum_{i=1}^{n}\sum_{j=n+1}^{2n}(x_i y_j + x_j y_i)}{\sum_{i=1}^{n}\sum_{j=n+1}^{2n}(x_i x_j - y_i y_j)} \qquad (17)$$

which is an analytic solution for the 2D case of orientation. However in higher dimensions a minimization procedure is used (see Appendix B.4).

B. Minimization

In the general case, the folding–unfolding method depends on minization of the S value over (1) all divisions of the given points into sets n_s (see Section 3.2), (2) all divisions of the symmetry-group elements into sets (denoted G_i in the text—see Section 2.4), (3) all orderings of the sets G_i, and (4) all positions and orientations of the symmetry group. The minimization problem seems very forbidding; however, a closer inspection shows that the complexity of the minimization problem can be greatly reduced and, in some cases, even an analytic expression can be given.

B.1. Divisions of the Given Points into Sets. Dividing the points into sets allows one to apply the folding method on each set independently; i.e., each set is transformed into a G-symmetric configuration of points. Thus the points of a set must be "similar"; i.e., if the points were colored, weighted by mass, or associated with any other varying feature, we would expect points of each set to have the same feature. In this paper we restrict ourselves to geometric features and the only "similarity" between points is their connectivity. Thus, dividing the points into sets is constrained by the connectivity of the points: each set contains points having the same connectivity. We delay the general discussion of the division rules to the next report and concentrate here on the special case of cyclic connectivity, which is of relevance to most examples in this report. In this case all points have the same connectivity and seem to impose no constraints on the division of points into sets. However, considering symmetric configurations of cyclic connectivity, we find strict restrictions on possible divisions of points into sets: In order to preserve the original connectivity in the final symmetrized object, the sets must be interlaced for C_n-symmetries and inversely interlaced for D_n-symmetries. For example Figure 19a shows a C_3-symmetric configuration of 12 points. The sets of 3 points marked as ●, ○, and □ are C_3-symmetric, and all three sets are interlaced. In the case of D_4-symmetry shown in Figure 19b, the interlacing is inversed in order to account for the reflection symmetry. Thus, instead of finding the cyclic order ● ○ □ ● ○□, ..., ● ○□ ● ○□, every other run is inverted: ● ○ □□ ○ ●, ..., ● ○□□ ○ ●. Thus given a nonsymmetric cyclic configuration of m points, we greatly reduce the possible divisions into sets: for C_n-symmetry only one division is possible, and for D_n-symmetry only $m/2n$ divisions are possible.

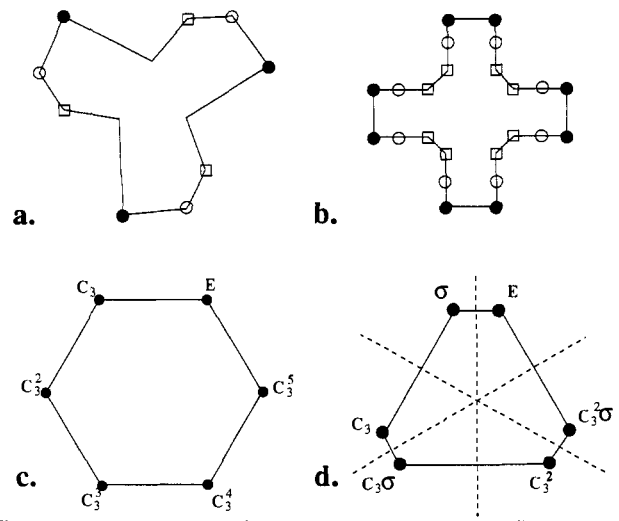

Figure 19. Divisions and orderings: (a) C_3-symmetric configuration of cyclic connected points. Each set of differently marked points (●, ○, and □) is C_3-symmetric. Each set is interlaced within the other sets. (b) D_4-symmetric configuration of cyclic connected points. Each set of differently marked points (●, ○, and □) is D_4-symmetric. Each set is inversely interlaced in the other sets. (c) C_6-symmetric configuration of cyclic connected points and the associated symmetry element. (d) D_4-symmetric configuration of cyclic connected points and the associated symmetry element.

B.2. Divisions of the Symmetry-Group Elements into Sets. This step of the minimization should be performed for each set of points separately; however in practice the division is usually the same for all sets having the same number of elements.

Given a symmetry group of n_g elements and given a set of n_p points, the elements of the group must be divided into n_p sets, each having a divisor of $l = n_g/n_p$ elements (Step 3 in Section 2.4). As described in Section 2.4 one of the sets (G_1 in the text) contains all group elements that leave a point in place. All other sets are constructed and dependent on that set. Thus we must actually minimize over all possible selections of l elements that leave a point in place. Considering the symmetry groups, we find that this greatly restricts the possible divisions of symmetry elements into sets. For example in the case of four points and the T_d-symmetry group, there is only one possible division of the elements into sets (the division given in Section 3.1).

B.3. Ordering the Sets of Elements. Given sets of elements (G_i), we must minimize over all possible orderings of the sets. Following the ordering of the sets, the folding method associates each set with a point. Thus, as in Section B.1, here too the connectivity of the points constrains the possible orderings of the element sets. For example consider the C_6-symmetric configuration of cyclic connected points in Figure 19c. The connectivity constrains the ordering of the elements of the C_6 group to be e, C_3^1, ..., C_3^5 or any cyclic permutation of the sequence or any reflections of the sequence (where e is the identity element and C_3^i is the rotation element of $2\pi i/6$ radians counterclockwise). The connectivity of the points in Figure 19d constrains the ordering of the six elements of the D_3 group to be e, σ, C_3, $C_3\sigma$, C_3^2, $C_3^2\sigma$ or any cyclic permutation of the sequence or any reflections of the sequence (where C_3^i is the rotation element of $2\pi i/3$ radians counterclockwise). Further simplification is obtained when one notices that the folding method applied under any of the cyclic orderings gives the same result. Thus only one of the cyclic orderings and its reflected ordering need to be considered. Thus, for cyclic configurations the possible orderings of the element sets are reduced to two.

B.4. Positions and Orientations of the Symmetry Group. A priori it seems that the most difficult stage in the minimization process is the sreening of all possible positions and orientations of the symmetry group. However even this stage can be simplified as follows: We have proven that the minimum is obtained when the symmetry group is positioned at the center of mass of the

points to be measured (i.e. all lines and points of rotation and all lines and planes of reflection must pass through the centroid of the configuration) (Appendix A.2). Thus, steps 1 and 2 in the folding method (Section 2.3) eliminate the need to minimize over all positions of the symmetry group, so that we only need to minimize on the orientation. Considering the orientation of symmetry groups in 2D, we find that the orientation of C_n-symmetry groups (having only rotational elements about a point) is irrelevant. For those symmetry groups in 2D having reflections, the minimization over all orientations is reduced to an analytic equation (see Appendix A.3). In 3D, minimizing over all orientations of a symmetry group is a three-dimensional problem. However, in some cases one can reduce the complexity. For example in the T_d-symmetry case one can reduce the complexity to two free parameters (see Appendix C). Furthermore, one should notice that minimization need not be performed over *all* orientations, since, being a symmetry group, some orientations leave the group invariant (up to order). Thus, in practice, one needs to minimize only over a cone of possible orientations. To minimize, the gradient descent method is used: One initially selects a single possible orientation as the current orientation. At each step one compares between the S value for the current orientation and that for neighboring orientations. The current orientation is changed to be that orientation which gives the lowest S value. If the current orientation does not change, the neighborhood size is decreased. The process is terminated when neighborhood size reaches a predefined minimum size. The current orientation at the end of the process is taken as the orientation which minimizes the S value.

C. Tetrahedral Minimization

As described in Section 3.1, the evaluation of $S(T_d)$-symmetry requires minimization over all orientations of the T_d-symmetry group. In general, orientation is a three-parameter problem (pitch, roll, and yaw). However, in the case of T_d-symmetry of a tetrahedron, one can reduce the problem to a two-parameter problem: given the direction of one of the C_3 axes of the symmetry group, one can determine the direction of the other axes analytically. This determination is a two-parameter problem: Given four points in 3D, P_1, P_2, P_3, and P_4, and given a unit vector w_4 representing the direction of the C_3 axis, find four points $\hat{P}_1$, $\hat{P}_2$, $\hat{P}_3$, and $\hat{P}_4$ and three unit vectors w_1, w_2, and w_3 such that $\hat{P}_i$ lies on the direction vector w_i (i = 1, ..., 4). The points $\hat{P}_1$, ..., $\hat{P}_4$ form a T_d-symmetric configuration, and the following is minimized:

$$\sum_{i=1}^{4} \|P_i - \hat{P}_i\|^2$$

Denoting by R the rotation matrix of $2\pi/3$ radians about w_3 and noting that $Rw_2 = w_1$ and $R^2w_3 = w_1$, one uses the Lagrange multipliers and minimizes the following:

$$\|P_4 - hw_4\|^2 + \|P_1 - hw_1\|^2 + \|RP_2 - hw_1\|^2 + \|R^2P_3 - hw_1\|^2 + \lambda_1(\langle w_1, w_1\rangle - 1) + \lambda_2(\langle w_1, w_3\rangle - {}^1/_3)$$

where λ_1 and λ_2 are the Lagrange multipliers and h is a constant such that $\hat{P}_i = hw_i$. The second constraint stems from the fact that the w_j are C_3 axes of the T_d-symmetry group. Deriving and equating to zero, one obtains

$$\lambda_2 = 6h\langle \bar{P}, w_4\rangle + \frac{6h^2 + 2\lambda_1}{3}$$

$$\lambda_1 = -3h^2 \pm \frac{9h}{2\sqrt{2}}(\langle \bar{P}, \bar{P}\rangle - \langle \bar{P}, w_4\rangle^2)^{1/2}$$

$$w_1 = \frac{6h\bar{P} - \lambda_2 w_4}{6h^2 + 2\lambda_1} \tag{18}$$

and

$$w_1 = \frac{2\sqrt{2}}{3}\frac{\bar{P} - \langle \bar{P}, w_4\rangle w_4}{\|\bar{P} - \langle \bar{P}, w_4\rangle w_4\|} - \frac{w_4}{3} \tag{19}$$

where $\bar{P}$ is the average of the rotated vectors, i.e. $\bar{P} = {}^1/_3(P_1 + RP_2 + R^2P_3)$. Notice that the direction of w_1 is independent of h. The latter is evaluated as

$$h = \frac{1}{4}(3\langle \bar{P}, w_1\rangle + \langle P_4, w_4\rangle)$$

The geometric interpretation of this result is as follows. Given a w_4 direction, perform the following steps: (1) fold points P_1, P_2, and P_3 about the w_4 axis; (2) average the folded points, obtaining a single point $\bar{P}$; (3) project $\bar{P}$ onto the cone of possible w_1 vectors (i.e. a cone of vectors x such that $\langle x, w_4\rangle/\|x\| = {}^1/_3$), obtaining point w_1', and assign $w_1 = w_1'\|w_1'\|$ (This step is the geometric interpretation of eq 19.); (4) project P_4 onto w_4, obtaining w_4'; (5) fold point w_1' toward w_4'; (6) average the folded w_1' and w_4' using a weighted average (point w_1' is itself a folded average of three points—step 2), obtaining a single averaged point $\hat{P}_4$, $\hat{P}_4 = (3w_1' + w_4')/4$; (7) unfolded point $\hat{P}_4$ toward w_1', obtaining $\hat{P}_1$; (8) unfold point $\hat{P}_1$ about the w_4 axis, obtaining $\hat{P}_2$ and $\hat{P}_3$. A perfect T_d-symmetric set of four points is obtained.

D. The Bounds of S Values

Following the definition of the CSM in eqs 1 and 2 in Section 1, the S values are limited to the range 0, ..., 100 or S' values are limited to the range 0, ..., 1. The lower bound of S' is obvious from the fact that the average of the square of the distances moved by the object points is necessarily non-negative. The upper bound of the average is limited to 1, since the object is previously normalized to a maximum distance of 1, and by translation of all vertex points to the center of mass, a symmetric shape is obtained.

The upper bound on S can be tightened for specific cases. For instance in 2D one can show that the maximum S' value for a triangle, with respect to C_3, is ${}^1/_3$: Consider the three vertices of a normalized triangle P_1, P_2, and P_3 in 2D (the centroid is at the origin). Without loss of generality assume $P_1 = (0, 1)$ and that P_2 has a positive x-coordinate and denote by (x, y) the coordinates of P_2. Given the constraint that the centroid is at the origin, one has $P_3 = (-x, -1, -y)$. In fact P_2 is limited to a circle sector due to the centroid constraint and the normalization constraint (limiting all P_i's to be in the unit circle). Given these notations, we have that the S' value of the triangle with respect to C_3-symmetry is given by

$$\frac{1}{3}(1 + y^2 + y - \sqrt{3}x + x^2)$$

Considering the limited range of the P_2 coordinates, the maximum value is obtained when $P_2 = (0, 0)$ or $P_2 = (0, -1)$ (which are equivalent cases) and the maximum S' value is ${}^1/_3$.

The maximum S' value is actually obtained for extreme cases such as a polygon of m vertices ($m = qn$) whose contour outline a regular q-gon (i.e. every qth vertex of the m-gon concides with a vertex of a regular q-gon). For details, see the Appendix in Part 1.[1]

Note Added in Proof. Two relevant reports appeared this year: Mezey has proposed in *J. Math. Chem.* (**1992**, *11*, 27) symmetry deficiency measures, applicable to both discrete and continuous problems, and Cammi et al. used symmetry coordinates to analyze tetrahedral and octahedral distortions (Cammi, R.; Cavalli, E. *Acta Crystallogr.* **1992**, *B48*, 245). Chirality measures have continued to attract attention this year: Cauvin, R. *J. Phys. Chem.* **1992**, *96*, 4706. Kuz'min, V. E.; et al. *J. Phys. Org. Chem.* **1992**, *5*, 295. Part 4 in our series is devoted to chirality as an integral part of the generalized treatment of symmetry described in this report.