A New Class of Molecular Shape Descriptors. 1. Theory and Properties

Marc L. Mansfield† and David G. Covell*

Screening Technologies Branch, DCTD, NCI, Frederick, Maryland 21702

Robert L. Jernigan

Laboratory of Experimental and Computational Biology, DBS, NCI, Bethesda, Maryland 20892

Received July 22, 2000

The integrals $V(n_1, n_2, n_3) = \int d\mathbf{r}\, x^{n_1} y^{n_2} z^{n_3}$, where $\int d\mathbf{r}$ represents integration over the volume of a body, such as a molecule, where x, y, and z are Cartesian coordinates of a point in the interior of the body relative to an arbitrary reference frame, and where n_1, n_2, and n_3 are integers greater than or equal to zero, constitute moments of the volume distribution of the body. Considering all such quantities for which $0 \le n_1 + n_2 + n_3 \le 6$ gives a set of 84 independent numbers which characterize the shape of the body and constitute a very useful set of shape descriptors. They also carry information about the absolute orientation and position of the body, and because their behavior under rotations and translations can be calculated quickly, they provide a fast, robust algorithm for the alignment of two similar molecules as well as a qualitative measure of their similarity. This paper reports the performance of the alignment algorithm on a learning set of about 80 different shapes. The algorithm is further tested against a set of small drug-like compounds that have been screened as anticancer agents. In both cases, excellent alignments of "shape-similar" molecules are obtained. Discussions are provided on many basic properties of these moments, e.g., their behavior under translations and rotations of the reference frame and their symmetry properties.

INTRODUCTION

The development and application of molecular shape descriptors is an active area in computational biology. The goal is to develop mathematical descriptors of the shape of an object that can determine, first, whether two molecules have comparable shapes and, second, determine the transformation that optimizes their alignment, assuming initially that they have arbitrary relative orientation. Applications include drug discovery, in which one might scan databases for molecules having the proper shape to fit the active site of an enzyme, and protein structural analysis, where one might wish to find similar folding patterns in different proteins. The need to screen large databases of molecules puts an obvious premium on descriptors that can be calculated or manipulated rapidly.

There are a number of examples in the literature of descriptors based on graph-theoretical or topological concepts.[1−4] Typically, these only depend on the internal structure of the molecule,and, without additional information, can only solve the first of the two problems cited above. They may inform us that two molecules have comparable shapes, but since they carry no information about the absolute orientation or position of a body, they are not useful for computing molecular superpositions. Indeed, any translationally and rotationally invariant descriptor will have this shortcoming. On the other hand, descriptors that are not rotational or translational invariants must be recomputed each time a new position or orientation is considered, and these manipulations can be costly. For example, several groups have considered Fourier or spherical harmonic expansions but were hampered by the necessity to recalculate the expansion each time a new orientation was contemplated.[5−8]

In this work, we sought rotational and translational noninvariants whose behavior under rotations and translations could be computed rapidly. The desire to predict behavior under rotations led us to consider tensor quantities, and we have been able to develop a successful set of molecular shape descriptors based on the tensors described in this paper. The shape descriptors of Grant et al.[9,10] also appear to be noninvariants for which transformations can be calculated rapidly. Indeed, the performance of their approach relative to ours is a topic of future study.

These quantities are also moments of the distribution of molecular volume or surface, and so the problem is equivalent to predicting a distribution from the values of its moments. This paper introduces the moments, explores their properties, and describes a technique for comparing and aligning any two arbitrary shapes. Another paper will discuss the application of these techniques to the alignment of protein molecules.[11]

DEFINITIONS OF MOMENTS

For any given body, we can define a set of shape descriptors as follows. Let r_α, $\alpha \in \{1, 2, 3\}$, represent a Cartesian coordinate of a point relative to some arbitrary reference frame, i.e., $(r_1, r_2, r_3) = (x, y, z)$. (We will follow the convention that lower case Greek letters are indices of Cartesian coordinates.) Then let

$$V^{(K)}_{\alpha\beta\gamma\delta\ldots} = \int d\mathbf{r}\, r_\alpha r_\beta r_\gamma r_\delta \ldots \tag{1}$$

The body may be a volume, a surface, a space curve, a single

* Corresponding author.

† Present address: Department of Chemistry and Chemical Biology, Stevens Institute of Technology, Hoboken, NJ 07030.

10.1021/ci000100o CCC: $22.00
Published on Web 01/26/2002

point, or a collect ion of points. $\int d\mathbf{r}$ represents an integral over the body. In the above, K equals the number of indices in $V_{\alpha\beta\gamma\delta...}$, or equivalently, it equals the number of terms $r_\alpha r_\beta r_\gamma r_\delta ...$ appearing in the integrand. According to this definition, the full set of these quantities for given K forms the components of a rank-K tensor. We let $\mathbf{V}^{(K)}$ represent the complete tensor. $\mathbf{V}^{(K)}$ has units (length)$^{K+T}$, where T is the topological dimension of the body, i.e., $T = 3$ for a volume, 2 for a surface, 1 for a space curve, and 0 for a point or finite collection of points. The superscript (K) is often redundant. For example, in $V^{(K)}_{\alpha\beta\gamma\delta...}$, K can be determined by counting indices, and so in such cases it will usually be omitted. Obviously, the rank-zero tensor $\mathbf{V}^{(0)}$ is either the volume ($T = 3$), the area ($T = 2$), the contour length ($T = 1$), or the number of points ($T = 0$). In what follows, we will usually use the generic term "volume" and the symbol V for $V^{(0)}$ even though T may be less than 3. We also consider tensors normalized by the volume: $\mathbf{T}^{(K)} = \mathbf{V}^{(K)}/V$.

The rank-K tensor has 3^K components, but components are invariant under exchange of indices, $V_{\alpha\beta} = V_{\beta\alpha}$, etc. Because of this, it is also useful to consider a second notation: $V^{(K)}(n_1, n_2, n_3)$, where n_1, n_2, and n_3 represent respectively the total number of 1's, 2's, and 3's appearing in the index list $\alpha\beta\gamma\delta...$. In other words

$$V^{(K)}(n_1, n_2, n_3) = \int d_r\, x^{n_1} y^{n_2} z^{n_3} \tag{2}$$

with $n_1 + n_2 + n_3 = K$: Both notations are useful, and we will switch back and forth as needed. To avoid confusion, a list of K indices appearing as subscripts always denotes the first notation, while a list of three indices appearing in parentheses always denotes the second. Since $K = n_1 + n_2 + n_3$, it is also redundant to specify it with the new notation, and the superscript (K) will again usually be suppressed.

The triple (n_1, n_2, n_3) is a point on the simple cubic lattice, and the condition $K = n_1 + n_2 + n_3$ defines one of the (111) planes of the lattice. Restricting n_1, n_2, and n_3 each to be nonnegative restricts us to lattice points lying within an equilateral triangle on this plane. Therefore, the total number of unique components of the rank-K tensor is equal to one of the "triangle numbers":

$$N_K = (K + 1)(K + 2)/2 \tag{3}$$

As defined, these quantities are moments of the distribution of the volume. It is well-known that Fourier transforms are generating functions for the moments of a distribution. For example, the following may be derived from eq 1 simply by expanding the exponential term by term:

$$G(\mathbf{k}) = \int d\mathbf{r} \exp(i\mathbf{k}\cdot\mathbf{r}) = \sum_{K=0}^{\infty} \frac{i^K}{K!} \sum_{\alpha\beta\gamma...} k_\alpha k_\beta k_\gamma ... V^{(K)}_{\alpha\beta\gamma...} \tag{4}$$

By this definition, $G(\mathbf{k})$ is the Fourier transform of the function that is 1 inside the body and 0 outside. The uniqueness theorem of Fourier transforms indicates therefore that the complete set of all $\mathbf{V}^{(K)}$ uniquely determines the body in question; distinct bodies will always possess distinct moments. However, we present below examples of distinct bodies whose moments match exactly through some arbitrary order.

REFERENCE FRAME TRANSFORMATIONS

Since these are tensors, we know how they transform under arbitrary rotations[12]

$$V'_{\alpha\beta\gamma...} = \sum_{\lambda\mu\nu...} R_{\alpha\lambda} R_{\beta\mu} R_{\gamma\nu} ... V_{\lambda\mu\nu...} \tag{5}$$

where V and V' denote tensor components before and after rotation, respectively, and $R_{\alpha\beta}$ represents one component of the rotation matrix. In terms of Euler angles we have[13]

$$R = \begin{bmatrix} \cos\psi\cos\phi - \cos\theta\sin\phi\sin\psi & -\sin\psi\cos\phi\cos\theta\sin\phi\cos\psi & \sin\theta\sin\phi \\ \cos\psi\sin\phi + \cos\theta\cos\phi\sin\psi & -\sin\psi\sin\phi + \cos\theta\cos\phi\cos\psi & -\sin\theta\cos\phi \\ \sin\theta\sin\psi & \sin\theta\cos\psi & \cos\theta \end{bmatrix} \tag{6}$$

Under translations, the moments transform according to

$$V'(n_1, n_2, n_3) = \sum_{i=0}^{n_1}\sum_{j=0}^{n_2}\sum_{k=0}^{n_3} \binom{n_1}{i}\binom{n_2}{j}\binom{n_3}{k} x_0^{(n_1-i)} y_0^{(n_2-j)} z_0^{(n_3-k)} V(i, j, k) \tag{7}$$

which is a direct result of the binomial theorem, where (x_0, y_0, z_0) is the translation vector and where V and V' denote components before and after translation.

Under a change of scale, $(x, y, z) \rightarrow (\lambda x, \lambda y, \lambda z)$, the moments transform as

$$\mathbf{V}'^{(K)} = \lambda^{K+T} \mathbf{V}^{(K)} \tag{8}$$

for T the topological dimension.

PREFERRED FRAMES OF REFERENCE

These moments have been defined for arbitrary reference frames. However there are frames for which they take on a simpler form.

Consider eq 7 for each of the three components of $\mathbf{V}^{(1)}$:

$$V'(100) = x_0 V + V(100) \tag{9a}$$

$$V'(010) = y_0 V + V(010) \tag{9b}$$

$$V'(001) = z_0 V + V(001) \tag{9c}$$

Eq 9 indicates that any body can be translated to a frame in which all components of $\mathbf{V}^{(1)}$ are zero. If the body has uniform density, then this is any center-of-mass frame. In an arbitrary frame, $\mathbf{V}^{(1)}$ is the total volume times the displacement vector of the center-of-mass. Also in a center-of-mass frame, $\mathbf{V}^{(2)}$ is related to the radius of gyration, R_g. If we define R_g as the rms displacement of all points from the center-of-mass, then obviously

$$R_g^2 = V^{-1} \int d\mathbf{r}\,(x^2 + y^2 + z^2) = V^{-1}(V_{11} + V_{22} + V_{33}) \tag{10}$$

Then also in a center-of-mass frame, and assuming mass units such that the body has unit density, we have a simple

relationship between the moment-of-inertia tensor, **I**, and $\mathbf{V}^{(2)}$

$$\mathbf{I} = R_g^2 V \mathbf{1} - \mathbf{V}^{(2)} \tag{11}$$

where **I** is the moment-of-inertia tensor, R_g is the radius of gyration, and **1** is the unit tensor. Therefore, $\mathbf{V}^{(2)}$ is diagonal in the principal-axis frame of the body, and its eigenvalues are $R_g^2 V - I_\alpha$, for I_α one of the principal moments of inertia.

In summary, $\mathbf{V}^{(0)}$, $\mathbf{V}^{(1)}$, and $\mathbf{V}^{(2)}$ have relationships with more familiar quantities: the volume, center-of-mass displacement vector, moment-of-inertia tensor, and radius of gyration. In the principal axis frame of reference $\mathbf{V}^{(1)} = 0$ and $\mathbf{V}^{(2)}$ is diagonal. In an arbitrary frame, $\mathbf{V}^{(1)}$ equals the total volume times the displacement vector of the center of mass.

SYMMETRY PROPERTIES

In this section we consider some of the properties expected of the moments when the body possesses a given symmetry element. Some symmetry properties can be determined by calculating the moments for *n*-tuples of points. For example, since any body possessing a center of inversion is a superposition of doublets of points (x, y, z) and $(-x, -y, -z)$, its moments can be constructed by superimposing the moments computed from all such doublets. In an Appendix, we prove a number of symmetry properties that can be obtained by examining the sums

$$P(n_1, n_2, n_3) = \sum_{j=1}^{n} x_j^{n_1} y_j^{n_2} z_j^{n_3} \tag{12}$$

computed for appropriate *n*-tuples of points. The following statements summarize the findings of the Appendix.

1. For all odd K, $\mathbf{V}^{(K)} = 0$ if and only if the object possesses a center of inversion at the origin.

2. For all odd n_3, $V(n_1, n_2, n_3) = 0$ if and only if the object possesses a reflection plane at $z = 0$.

3. If the object has a C_n axis aligned with the *z*-axis and if $n_1 + n_2 < n$, then $V(n_1, n_2, n_3) = 0$ unless n_1 and n_2 are both even.

4. If the object has spherical symmetry (point group R_3) about the origin, then $V(n_1, n_2, n_3) = 0$ unless n_1, n_2, and n_3 are all even. (Established in Appendix 2.)

5. If the object has a C_n axis, then $V^{(0)}$, $\mathbf{V}^{(1)}$, $\mathbf{V}^{(2)}$, ..., $\mathbf{V}^{(n-1)}$ are all invariant under arbitrary rotations about the C_n axis. This is true in any reference frame.

6. Given two bodies, one with a C_n axis and one with a C_m axis, with $n < m$, and given that the bodies are aligned so that the two axes as well as their centers of mass coincide, and also given that the bodies have identical longitudinal and radial volume distributions, then $V^{(0)}$, $\mathbf{V}^{(1)}$, $\mathbf{V}^{(2)}$, ..., $\mathbf{V}^{(n-1)}$ are the same for both bodies.

7. If the body has an S_m axis, with m even, oriented so that the axis is along the *z*-axis and with the associated reflection plane at $z = 0$, then $V(n_1, n_2, n_3) = 0$ whenever $n_1 + n_2 < m/2$ and n_3 is odd.

Note that many of the moments of bodies of high symmetry are zero. We see below that this affects the value of the alignment score.

Item 6 above addresses the issue of uniqueness of moments. We know that the complete set of all moments for two bodies will agree if and only if the bodies are identical. However, since we work with finite sets of moments in all numerical work, it is instructive to have examples of unique bodies whose moments all agree through some finite rank. An easy way to construct a pair of such bodies is to begin with two identical bodies having full cylindrical symmetry about an axis, cut $2n$ identical wedges in the first and $2m$ in the second, and then remove every other wedge.

Because of the C_n symmetry properties, alignment algorithms based on these moments cannot be expected to work well on bodies with C_n symmetry unless enough higher order moments are included. Since our algorithm includes moments only through $K = 6$, it can at best only align the principal axes of any bodies of group C_7 or higher.

ROTATIONALLY INVARIANT LINEAR COMBINATIONS AND AVERAGES OVER ORIENTATIONS

We now show that for even K, there exist linear combinations of components that are rotationally invariant. (We have already encountered one example, eq 10.) Furthermore, we show that each of these linear combinations is unique: For any even K, there exists only one linear combination that is a rotational invariant. This uniqueness property then lets us determine the average over orientations of any component.

Define the integrals

$$\Gamma_{2n} = (2n + 1)^{-1} \int d\mathbf{r}\, (x^2 + y^2 + z^2)^n \tag{13}$$

As defined, Γ_{2n} is invariant with respect to rotations about the origin, irrespective of whether the origin coincides with the center-of-mass of the body, and it is also a linear combination of the components of $\mathbf{V}^{(2n)}$. Furthermore, Γ_{2n} is the only linear combination (up to multiplicative constants) of components of $\mathbf{V}^{(2n)}$ that is rotationally invariant, because the integrand of any other would also have to add up to r^{2n}. On the other hand, we do not expect to find any linear combinations of the components of odd-rank moments that are rotational invariants, since these cannot be written as a function of r^2.

Let $\langle ... \rangle$ represent an average over all Euler angles:

$$\langle Z \rangle = (8\pi^2)^{-1} \int_0^{2\pi} \mathrm{d}\psi \int_0^{2\pi} d\phi \int_0^{\pi} \sin\theta \, d\theta \, Z(\theta, \phi, \psi) \tag{14}$$

Again, this average is defined relative to the origin, irrespective of the position of the center-of-mass. Now consider the average over Euler angles of an arbitrary component of $\mathbf{V}^{(K)}$:

$$\langle V_{\alpha\beta\gamma...} \rangle = \sum_{\lambda\mu\nu...} \langle R_{\alpha\lambda} R_{\beta\mu} R_{\gamma\nu} ... \rangle V_{\lambda\mu\nu...} \tag{15}$$

Only the $R_{\alpha\lambda} R_{\beta\mu} R_{\gamma\nu}$ ... in eq 15 are enclosed in $\langle ... \rangle$ brackets because only these terms depend on Euler angles. Equation 15 indicates, therefore, that $\langle V_{\alpha\beta\gamma...} \rangle$ is a linear combination of components of $\mathbf{V}^{(K)}$. But it is also a rotational invariant. Therefore, when K is even, it equals Γ_K times a factor that is independent of the shape of the body

$$\langle V_{\alpha\beta\gamma...} \rangle = k_{\alpha\beta\gamma...} \Gamma_K, \text{ for } K \text{ even} \tag{16}$$

while if K is odd

$$\langle \mathbf{V}^{(K)} \rangle = \mathbf{0}; \text{ for } K \text{ odd} \tag{17}$$

In eq 16, $k_{\alpha\beta\gamma\ldots}$ is independent of the form of the body itself, because it is derived ultimately from $\langle R_{\alpha\lambda} R_{\beta\mu} R_{\gamma\nu} \ldots \rangle$. Therefore, each $k_{\alpha\beta\gamma\ldots}$ can be determined by direct computations on any body with spherical symmetry; results for spheres are found in Appendix 2.

The final result is that $\langle V(n_1, n_2, n_3) \rangle$ is zero whenever n_1, n_2, or n_3 is odd, and that when all are even, it is given by

$$\langle V(n_1, n_2, n_3) \rangle = (4\pi)^{-1} (K+1) S(n_1, n_2) T(\alpha, n_3) \Gamma_K \tag{18}$$

Here $\alpha = (n_1 + n_2)/2$ and the quantities $S(n_1, n_2)$ and $T(\alpha, n_3)$ are defined in Appendix 2. The linear combinations defining Γ_K through $K = 6$ and expressions for each $\langle V(n_1, n_2, n_3) \rangle$ are given in an Appendix.

INTERNAL REPRESENTATIONS OF MOMENTS AND COORDINATE TRANSFORMATIONS

Our calculations employ all seven tensors $\mathbf{V}^{(K)}$ with $K = 0, 1, \ldots, 6$, requiring specification of $N_0 + N_1 + \ldots + N_6 = 84$ different components. It works well to represent the components internally as a single linear array of length 84, the "84-vector". We use look-up tables to map back and forth between the index of each component in the 84-vector and the three indices n_1, n_2, and n_3. It proves possible to bring the expressions for each of the three coordinate transformations into the form

$$V'(n_1, n_2, n_3) = \sum_{n'_1, n'_2, n'_3} \hat{O} \begin{bmatrix} n_1 & n_2 & n_3 \\ n'_1 & n'_2 & n'_3 \end{bmatrix} V(n'_1, n'_2, n'_3) \tag{19}$$

where $\hat{O}$ represents a linear operator. With an operator $\hat{O}$ as in the above equation and with the look-up tables, the action of each transformation can be represented as an 84×84 matrix acting on the 84-vector.

The scale transformation is especially simple, its operator being a diagonal matrix, as in eq 8:

$$V'(n_1, n_2, n_3) = \lambda^{K+T} V(n_1, n_2, n_3) \tag{20}$$

The rotation operator is more difficult to realize, since rotations are more simply represented in the $\alpha\beta\gamma\ldots$ notation. For example, eq 5 for the case V'_{111} reads

$$V'_{111} = \sum_{\lambda\mu\nu} R_{1\lambda} R_{1\mu} R_{1\nu} V_{\lambda\mu\nu} \tag{21}$$

where λ, μ, and ν each take on the three values 1, 2, and 3, for a total of 27 terms. The same equation in the (n_1, n_2, n_3) notation becomes

$$\begin{aligned} V'(300) = {} & R_{11}R_{11}R_{11}V(300) + R_{12}R_{12}R_{12}V(030) + \\ & R_{13}R_{13}R_{13}V(003) + 3R_{11}R_{11}R_{12}V(210) + \\ & 3R_{11}R_{11}R_{13}V(201) + 3R_{12}R_{12}R_{13}V(021) + \\ & 3R_{11}R_{12}R_{12}V(120) + 3R_{11}R_{13}R_{13}V(102) + \\ & 3R_{12}R_{13}R_{13}V(012) + 6R_{11}R_{12}R_{13}V(111) \end{aligned} \tag{22}$$

We worked out (by computer, not by hand) similar equations for each of the other 84 components. These 84 equations define the rotation operator as specified in eq 19. Each of these 84 equations is encoded into a data structure that is saved on the hard drive, read into main memory at the beginning of each computation, and used to construct the full rotation operator each time it is needed. Since rotations only "mix up" components within a given rank, the full rotation operator is block diagonal, and we realize additional savings by multiplying block by block.

Since the formula for translations is given in the (n_1, n_2, n_3) notation, the construction of the translation operator is much more transparent. We can realize savings in its construction by computing the terms $\binom{n_1}{i}\binom{n_2}{j}\binom{n_3}{k}$ only once at the beginning of a run.

USING THE MOMENTS TO ALIGN TWO BODIES

Using these moments, we have developed an algorithm that successfully aligns two similar shapes having initially arbitrary relative position and orientation. It also computes a quantitative measure of their similarity. We assume that the moments $V(n_1, n_2, n_3)$ through $K = 6$ have been calculated in some center-of-mass frame for two different bodies, and we want to use these quantities to align the two bodies. One body will be designated the "test" body and the other the "target" body. The target remains stationary, while the orientation of the test body is adjusted to optimize its fit with the target. We will use superscripts T and $\odot$ to represent the test and the target, respectively. We sometimes refer to this procedure as "parking" the test shape "onto" the target shape.

The alignment score is defined as

$$S = \sum_{K=0}^{6} \frac{1}{N_K M_K^2} \sum_{\{n_1+n_2+n_3=K\}} [V^T(n_1, n_2, n_3) - V^{\odot}(n_1, n_2, n_3)]^2 \tag{23}$$

i.e., a least-squares comparison of moments. The normalization factor M_K^2 is defined such that

$$M_K = \begin{cases} \Gamma^{\odot}_K, & \text{for } K \text{ even}, \\ (\Gamma^{\odot}_{K-1} \Gamma^{\odot}_{K+1})^{1/2} & \text{for } K \text{ odd}, \end{cases} \tag{24}$$

which makes the score scale-invariant and dimensionless. The contribution from each rank K is further normalized by the quantity N_K, the number of independent components in each rank.

The alignment procedure includes two separate stages. The first stage of alignment is obtained by forcing principal axes of the two bodies to align. Any rank-2 symmetric tensor can be represented as an ellipsoid, and the stage-1 alignment can be described as an alignment of the inertia ellipsoids of the two bodies. We diagonalize $\mathbf{T}^{(2)}$, sort the resulting eigenvalues so that $t_1 > t_2 > t_3$, and then take the new x, y, and z axes to be the eigenvectors corresponding to t_1, t_2, and t_3, respectively. However, two precautions must be taken. If $\mathbf{x}$ is an eigenvector of some matrix, then so is $-\mathbf{x}$, and we must confirm that the three eigenvectors form a right-handed set of axes. This can be done by making sure that the new

z axis coincides with the positive cross product of the new *x* and *y* axes. The next precaution is needed because of the symmetry of the inertia ellipsoid. Even after correcting the new eigenvectors for right-handedness, the alignment may improve simply by rotating the body through 180° about any one axis. It is necessary to consider four alignments: one using the eigenvectors as supplied by the algorithm and corrected for right-handedness and three others generated by 180° rotations about each of the three new axes. Then of the four alignments, the one which gives the best alignment score according to eq 23 is selected. Optionally, the alignment includes a scale adjustment. If this scale adjustment option is desired, then the alignment score for each of these four alignments is only computed after the components of $\mathbf{V}^T$ have been scaled, using a scale factor determined from the eigenvalues of both $\mathbf{T}^{(2)T}$ and $\mathbf{T}^{(2)\odot}$:

$$\lambda = \left[\left(\frac{1}{3}\right)\left(\frac{t_1^{\odot}}{t_1^{T}} + \frac{t_2^{\odot}}{t_2^{T}} + \frac{t_3^{\odot}}{t_3^{T}}\right)\right]^{1/2} \tag{25}$$

In summary, we use $V^{(0)}$ and $V^{(2)}$ to calculate the principal axes and the initial scale factor for four tentative alignments, and we use the higher moments to make the selection from among these four.

The second stage of alignment is obtained by letting the test body adjust its position in space to minimize the alignment score, *S*. Usually, we assume that the test body undergoes first a scale transformation of magnitude λ, followed by a rotation through Euler angles (ψ, ϕ, θ), followed by a translation through (x_0, y_0, z_0), so that the final state of the test body, and the values of V^T (n_1, n_2, n_3) in eq 23, are functions of seven transformation variables (λ, ψ, ϕ, θ, x_0, y_0, z_0). If a scale-adjustment is not desired, then there are only six transformation variables. These variables are adjusted through a conjugate gradients search[14] to minimize *S*. The search is usually initiated from the configuration obtained in the stage-one alignment. The final value of *S* gives us a measure of the degree to which the two shapes are comparable.

When both alignment stages depend on λ we are permitting the test body to adjust its size in an attempt to match the target. This is obviously not desireable if we are trying to identify molecules of the same general shape *and* size. However, there may be instances where relative size is not important (e.g., to distinguish only between globular and fibrillar proteins). Therefore, our current algorithm permits us to switch the scale adjustment on or off. Furthermore, since the second alignment depends on (x_0, y_0, z_0) we also allow the test body to translate out of center-of-mass coordinates.

The CPU time for the alignment is about 3 s on a Pentium III machine. However, when running intercomparisons among large sets of shapes, even this may not be fast enough. The current code is designed with a "front-end filter", an initial screening based on only three rapidly computed descriptors. If the three eigenvalues of $\mathbf{V}^{(2)}$ are designated v_1, v_2, v_3, with $v_1 \geq v_2 \geq v_3$, then these three descriptors

$$\rho_1 = v_1/v_2,\ \rho_2 = v_2/v_3,\ \sigma = \frac{(\text{area})}{[36\pi(\text{volume})^2]^{1/3}} \tag{26}$$

are calculated for each shape, and the alignment is not attempted unless these three quantities are within some prespecified threshold, or unless this screening is specifically overridden. This filter screens out alignment attempts when the three descriptors indicate that their shapes are not similar. The quantity σ, often called the "sphericity", is effectively a dimensionless area-to-volume ratio. The factor 36π normalizes the ratio to 1 for a sphere.

The test and target shapes are not placed on an equal footing by this algorithm. Generally, we can expect that both the value of *S* and the precise alignment obtained by parking *A* onto *B* are not exactly the same as would be obtained by parking *B* onto *A*. However, they are generally close, particularly if the two shapes are similar. Finally, the parking procedure is not rotationally invariant. The $V^{\odot}$ terms in eq 23 normally are not rotationally invariant; a change in the orientation of the target changes these numbers and changes the quality of fit between the two bodies. Therefore, the relative alignment of test and target bodies depends on the absolute orientation of the target. However, if both objects are sufficiently similar, we can again anticipate that these differences will be small.

Special concerns arise whenever any of the three eigenvalues t_1, t_2, and t_3 are degenerate or nearly degenerate. Any body possessing a 3-fold or higher rotation axis has at least two degenerate eigenvalues, while bodies of point groups *T*, T_d, T_h, *O*, O_h, *Y*, Y_h, and R_3 possess triply degenerate eigenvalues.[15] And, of course, accidental degeneracies can occur in any point group. A 2-fold degeneracy means that the first-stage alignment is arbitrary with respect to rotations about the nondegenerate eigenvector, while a 3-fold degeneracy means that the first-stage alignment is completely arbitrary.

Near-degeneracies are also problematic. The success of the stage-1 alignment depends on the assumption that if two bodies are nearly identical, then their respective principal axes have nearly identical orientations. However, degenerate perturbation theory teaches us that a perturbation mixes eigenvectors whose eigenvalues lie near one another. In the present context, this means that a minor deformation of a body can lead to relatively large changes in any eigenvectors that are nearly degenerate.

Degeneracies, although problematic, are not fatal. As mentioned above, degeneracies imply that the first-stage alignment becomes arbitrary. We can deal with them by doing a more careful first-stage alignment whenever they are detected. Given a 2-fold degeneracy, we would need to search the space of rotations about the nondegenerate eigenvector, and given a 3-fold degeneracy, we would need to search in the space of all rotations.

EVALUATION OF INTEGRALS

During this proof-of-concept phase of our work, we have not been overly concerned with rapid computation of the integrals of eq 1. Consequently, we have made frequent use of a Monte Carlo algorithm. We imagine enclosing the body in a right rectangular prism or box. A large number, *N*, of points is generated at random within the box. A certain number of these, *M*, also lie within the body. Then the

moments are estimated according to

$$V(n_1, n_2, n_3) = \frac{V_b}{N} \sum_j x_j^{n_1} y_j^{n_2} z_j^{n_3} \qquad (27)$$

with the sum extending over all M points within the body, and where V_b is the volume of the box. Although rather slow, this approach is very conservative. All the shape-specific information on the body is embedded into a plug-in subroutine which reports whether an arbitrary point lies inside the body. In any case, when doing intercomparisons of a large set of $\mathcal{N}$ different bodies, the time required to do the integrals is proportional to $\mathcal{N}$, while the time to run the intercomparisons is proportional to $\mathcal{N}^2$. In such cases, we can afford to be a bit extravagent in doing the integrals.

In the example considered below of molecular alignments, the integrals were computed as discrete sums over atomic centers. This is, of course, a very fast approach for calculating the moments and appears to be adequate.

We are also investigating another approach. It is very common to represent a molecule as a union of overlapping spheres, usually centered on each atom. Assume that we assign the radius of each sphere in such a way that there are never more than pairwise overlaps of spheres. In other words, the spheres representing directly bonded atoms may overlap, but individual spheres are kept small enough to avoid tertiary overlaps. The resulting atomic radius is somewhat smaller than the van der Waals radius, but this still represents the molecule with the same degree of realism as a CPK model. It turns out that the resulting integrals, taken either over the volume or over the surface area, can be computed rigorously and rapidly. This may present a viable way of "fleshing out" the molecule if discrete sums over atomic centers proves to be an oversimplification.

EXAMPLES: SAMPLE SHAPES

The algorithm was developed and tested on approximately 80 sample shapes. Examples for some of these shapes are depicted in Figures 1−7. The range of shapes considered include the following: (1) solid spheres; (2) spherical shells of inner and outer radius R_1 and R_2, respectively, and with R_1/R_2 varying from 0.1 to 0.9; i.e., covering the extreme from very thick to very thin shells; (3) solid spheres of radius R_2 with a cylindrical hole of radius R_1 bored along a diagonal and with R_1/R_2 varying between 0.02 and 0.2; (4) solid spherical caps, i.e., solid spheres with a slice removed [Let d be the diameter of the sphere and let t be the thickness of the cap. Various t/d ratios from 0.1 to 0.9 were included. When $t/d = 0.1$ we have a thin disklike cap, when $t/d = 0.5$ we have a hemispherical cap, and when $t/d = 0.9$, the shape is that of a sphere with a small slice taken out.]; (5) solid right circular cylinders of various diameter to height (d/h) ratios; (6) right rectangular prisms of various length:width: depth (a:b:c) ratios; (7) several right circular cones of diameter-to-height ratio 1/2 and 1/1; and (8) 5-, 6-, 7-, and 8-fold "fans". Starting from a cylinder of diameter-to-height ratio of 10, an n-fold fan is created by slicing the disk into $2n$ identical wedges and removing every other wedge. This produces distinct objects possessing either a C_5, C_6, C_7, or C_8 rotation axis, all of which have identical radial and longitudinal densities. As already noted, we can anticipate problems with the C_7 and C_8 fans.

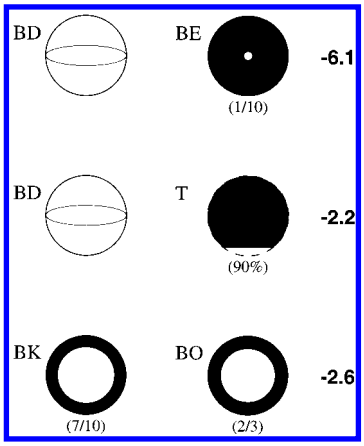

Figure 1. The moments, and therefore the alignments among bodies, are most sensitive to details of the bodies distant from the origin. For example, when aligning a solid sphere (**BD**) and a sphere with a small central cavity (**BE**, inner to outer radii in the ratio 1/10), a much smaller alignment score ($10^{-6.1}$) is obtained than that obtained ($10^{-2.2}$) when aligning the sphere to a sphere with a slice removed from its outer periphery (**T**) or than the score obtained ($10^{-2.6}$) when aligning two spherical shells with only slightly different ratios of inner to outer radii (**BK** and **BO**).

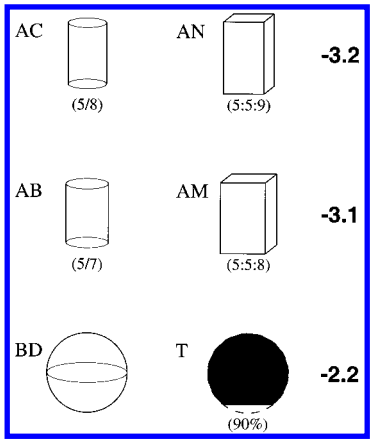

Figure 2. Surprisingly, the alignment score between certain cylinders and prisms is smaller by an order of magnitude that that between a solid sphere and a truncated solid sphere, although most reasonable people would judge that the two spheroidal shapes are more similar than are a prism and a cylinder. However, bodies with complete inversion symmetry are expected to have inherently lower alignment scores. The truncated sphere lacks an inversion center and has a higher score against the sphere.

The moments for each shape were calculated either exactly or by the Monte Carlo technique described above, using $M = 10^5$. For some shapes they were calculated both ways, permitting us to judge the performance of the algorithm when inexact Monte Carlo moments are used.

Alignments were attempted for all pairs of shapes that passed the front-end filter. These alignments also included a scale adjustment. As noted above, parking A onto B is not the same as the parking B onto A. In this section, quoted S values for pairs of shapes are the geometric means of S obtained by performing both alignments.

After alignment, the value of S gives some indication of the degree to which the two shapes resemble one another. In general S below about 10^{-5} indicates identical shapes, with $S \ll 10^{-5}$ when both sets of moments are calculated

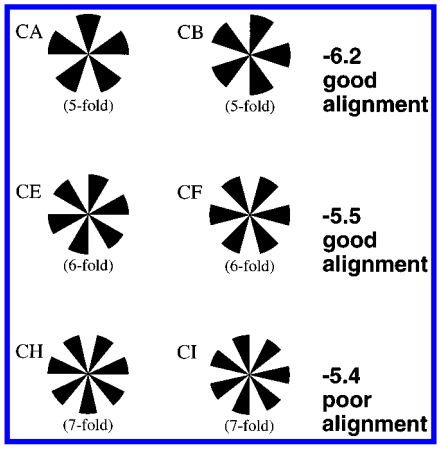

Figure 3. Alignments were attempted between initially nonaligned "fans" of 5-, 6-, and 7-fold symmetry. In all three cases, low alignment scores ($10^{-6.2}$, $10^{-5.5}$, and $10^{-5.4}$, respectively) were achieved, but in fact only the 5- and 6-fold objects were aligned. The algorithm is unable to align 7-fold or higher bodies because it only includes moments through six ranks.

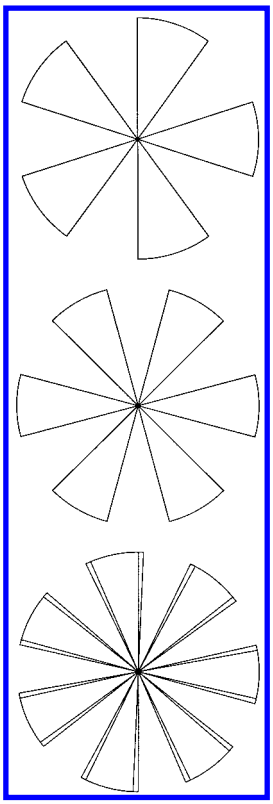

Figure 4. Each of the three diagrams displays the actual alignments obtained from the fans shown in Figure 3. Alignment of 5- and 6-fold fans is perfect to within the resolution of this figure but not the 7-fold fan.

exactly; and with S between about 10^{-6} and 10^{-5} when either set is calculated by Monte Carlo integration. S between about 10^{-5} and 10^{-3} generally indicates that the two shapes are comparable, S between about 10^{-3} and 10^{-2} is a transition zone, and S larger than about 10^{-2} indicates shapes with little or no resemblance. The capacity of the algorithm to align comparable shapes is generally excellent.

To highlight the limitations of the technique, and to indicate ways in which the technique might be improved, we present below examples of alignments that are exceptions to the rules given in the previous paragraph. We emphasize that these are infrequent exceptions, generally the algorithm

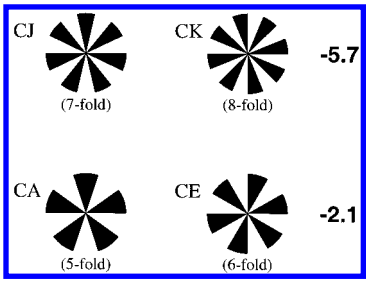

Figure 5. Since the "fans" all have identical radial and longitudinal densities, the 7- and 8-fold fans have identical moments through rank 6. The algorithm is therefore unable to distinguish the two and gives them a very low score, $10^{-5.7}$. The tensors of the 5- and 6-fold fans only agree through rank 4, and their score, $10^{-2.1}$, is much more typical of distinct bodies.

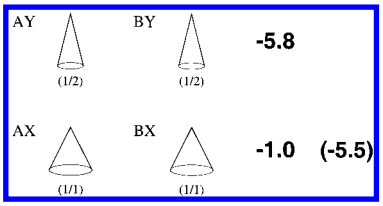

Figure 6. The identical cones **AY** and **BY** were properly aligned by the algorithm and resulted in a score of $10^{-5.8}$. However, identical cones **AX** and **BX**, each with diameter/height ratios of 1, have triply degenerate moments of inertia, which often frustrates the alignment, and these shapes obtained a high score, $10^{-1.0}$. Nevertheless, by performing an initial random search in rotation space, we were able to achieve an optimized alignment of **AX** and **BX** that scored $10^{-5.5}$.

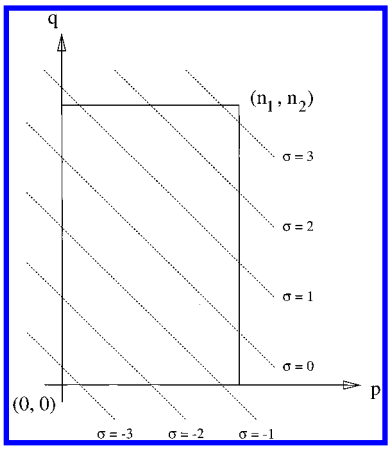

Figure 7. This figure is helpful in analyzing eq 43, which involves sums over integers $p \in \{0; ..., n_1\}$ and $q \in \{0, ..., n_2\}$. Therefore, individual terms in the sum can be represented by points in a $p-q$ space which lie within a rectangular region on the simple square lattice. Equation 43 also involves a Kroneker δ which selects out points lying on diagonals of slope -1. The $\sigma = 0$ diagonal bisects the rectangle. Successive diagonals $\sigma = \pm 1, \pm 2, ...$, are displaced horizontally or vertically by a distance $n/2$ on the lattice, so the total number of σ-diagonals that must be considered is determined by the relative magnitudes of $n_1 + n_2$ and n. For $n_1 + n_2 < n$ only the $\sigma = 0$ diagonal intersects the rectangle and eq 44 eventually ensues.

performs very well, both in aligning comparable shapes and in judging the similarity of distinct shapes.

The moments give much more weight to details of the objects distant from the origin. For example, a solid sphere and a thick spherical shell ($R_1/R_2 = 0.1$), shapes **BD** and **BE** in Figure 1, give $S = 10^{-6.1}$. A solid sphere and a spherical cap that is practically spherical ($t/d = 0.9$), shapes **BD** and **T**, give $S = 10^{-2.2}$. Two spherical shells that are nearly identical ($R_1/R_2 = 0.7$ and $R_1/R_2 = 0.667$), shapes

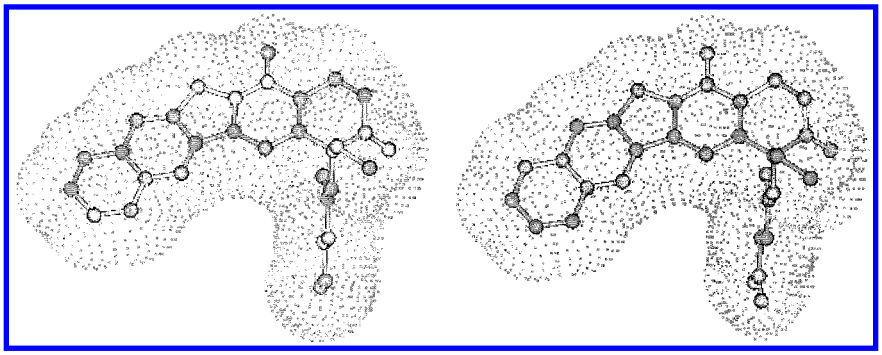

Figure 8. Pairwise alignment for two molecules with only a slightly different atomic composition. A near perfect atomic match is obtained from shape alignment.

BK and **BO**, yield $S = 10^{-2.6}$. The reason for this behavior becomes obvious when we realize that $V^{(0)}$ for shapes **BD** and **BE** agree to one part in 10^3, while the higher $V^{(K)}$ components agree even more closely. Generally speaking, the moments are insensitive to details of the objects near the origin.

Alignments between objects of high symmetry generally yield lower alignment scores. The best agreements obtained between cylinders and prisms were the cylinder **AC**, $d/h = 5/8$, with the prism **AN**, of dimensions 5:5:9, yielding a score of $10^{-3.2}$, and the cylinder **AB**, $d/h = 5/7$, with the prism **AM**, dimensions 5:5:8, yielding a score of $10^{-3.1}$, see Figure 2. Contrast these scores with those already cited for the pair **BD** and **T**, $10^{-2.2}$. Most individuals would agree that the two spheroidal shapes, **BD** and **T**, are more similar than are a cylinder and a prism. However, each cylinder and prism has an inversion center, so that all odd-K moments are uniquely zero in all reference frames and make no contribution to eq 23. **T**, on the other hand, does have nonzero odd-K moments which contribute to eq 23. Generally speaking, objects of high symmetry produce somewhat lower scores. In any case, the principal axes of the cylinder and prism are aligned perfectly by the algorithm, and the algorithm is able to select out prisms and cylinders of comparable aspect ratios.

The following examples are cited to display pitfalls which can occur if the objects have C_n axes with $n > 6$. As already mentioned, the 84-vectors of such objects are invariant to rotations about the C_n axis. (However, since such symmetry elements are very rare, this is not a serious shortcoming and could in principle be avoiding by including higher ranks in the algorithm.) Two 5-fold fans, shapes **CA** and **CB** in Figure 3, have identical shapes but are initially nonaligned. The same is true for the 6-fold fans **CE** and **CF**, and the 7-fold fans **CH** and **CI**. The alignment algorithm perfectly aligned **CA** and **CB**, giving a score of $10^{-5.5}$, and **CE** with **CF**, yielding a score of $10^{-5.5}$. It failed to align **CH** and **CI** (see Figure 4), although a very low score, $10^{-5.4}$, was obtained. Recall that **CA** and **CB** are expected to have identical, rotationally invariant tensors through $\mathbf{V}^{(4)}$, and likewise **CE** and **CF** through $\mathbf{V}^{(5)}$, and **CH** and **CI** through $\mathbf{V}^{(6)}$. Because of the rotational invariance of $\mathbf{V}^{(2)}$, the stage-1 alignment in all three cases is not particularly good. However, the information provided by the higher order tensors of both the 5- and 6-fold fans permitted stage 2 to work properly. On the other hand, the algorithm "thinks" **CH** and **CI** are already aligned since they have identical tensors through $\mathbf{V}^{(6)}$. Furthermore, the algorithm produced a very low score, $10^{-5.7}$, for the two distinct shapes **CJ** and **CK** in Figure 5, having 7- and 8-fold symmetry, respectively. In contrast, the score between **CA** and **CE**, $10^{-2.1}$, is within the range expected for distinct shapes.

The algorithm occasionally fails when moments of inertia are degenerate. The two cones **AY** and **BY** in Figure 6, identical except for the fact that one was integrated exactly and the other by Monte Carlo, were perfectly aligned by the algorithm and produced the score $10^{-5.8}$. In contrast, the two cones **AX** and **BX**, again mutually identical except for integration technique, produced the score $10^{-1.0}$ and were not well aligned. Cones with diameter-to-height ratios of 1 have triply degenerate moments of inertia. The stage-1 alignment is therefore entirely arbitrary, and in this particular case, the stage-2 alignment converged on some local minimum rather than on the global minimum. The cones having diameter/height ratios of 1/2 have doubly degenerate moments of inertia, which means that the stage-1 alignment is only arbitrary with respect to rotations about the C_∞ axis, which in turn means that the stage-1 alignment works appropriately in this particular case. If in place of the stage-1 alignment, we do an initial random search in rotational space prior to stage 2, then we are able to align **AX** and **BX** and achieve a score of $10^{-5.5}$. Cubes are another example of objects with triply degenerate moments of inertia for which stage 1 is arbitrary. However, several alignments among initially nonaligned cubes always produced perfect alignment. In this case, stage 2 was able to perform a perfect alignment despite the failure of stage 1.

The above examples based on fans, cones, and cubes indicate that the algorithm often performs well despite moment-of-inertia degeneracies but that it is not fool-proof. We plan to modify the algorithm so that when it detects near-degeneracies it does a more comprehensive initial search rather than relying on the stage-1 alignment as currently performed.

EXAMPLES: SYNTHETIC COMPOUNDS

We have tested our method by aligning a set of ~400 agents that have been screened by the National Cancer Institute for their antitumor activity (http://www.dtp.nih.gov). These agents represent small synthetic compounds that act putatively by binding either protein or nucleic acid targets and affecting their normal biologic function.[16] Considerable computational effort has been directed at the development of methods useful for examining structurally similar compounds within large libraries of synthetic agents.[17−22] The results presented here will briefly summarize our findings.

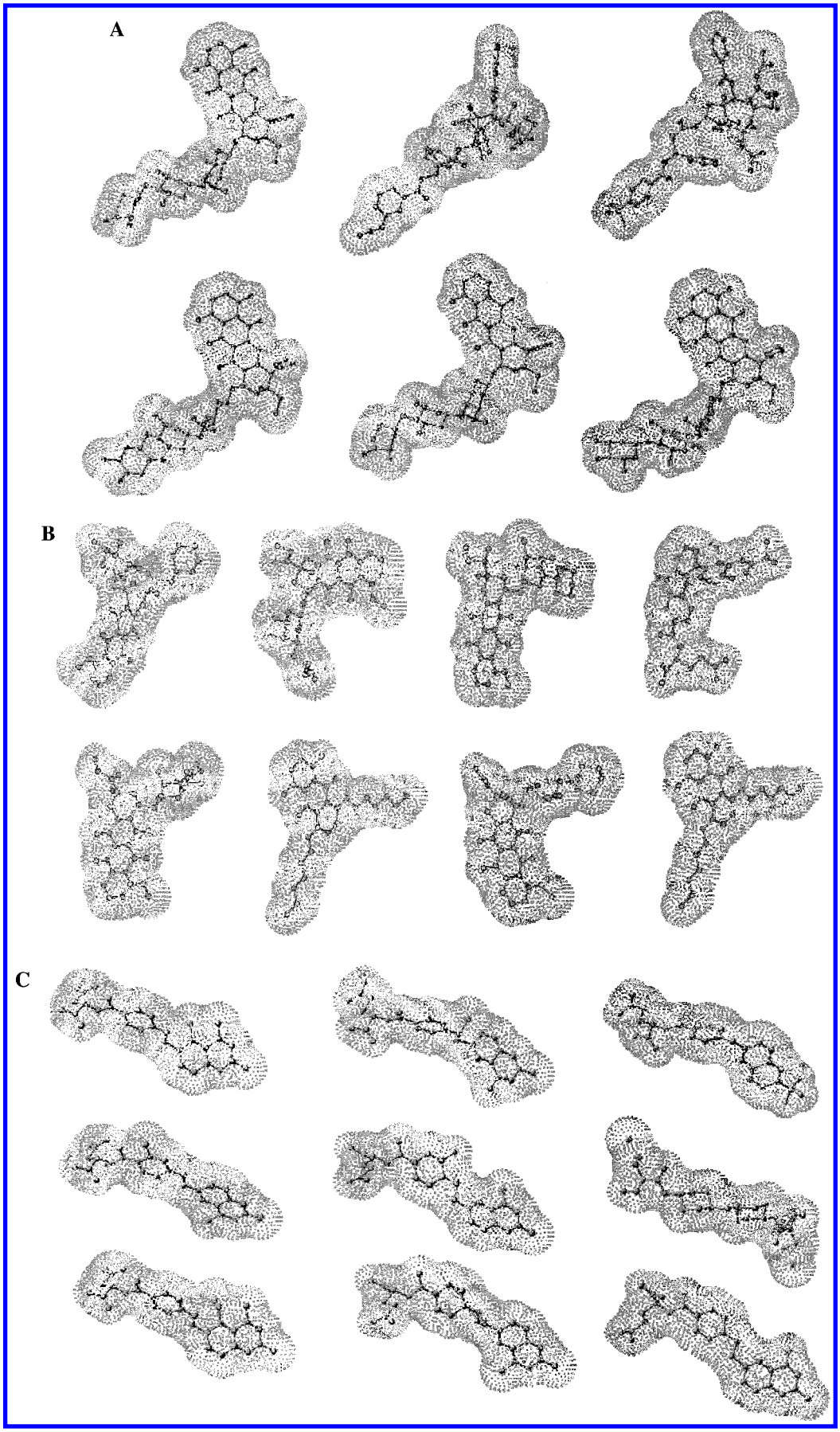

Figure 9. Pairwise alignment scores of 362 anticancer agents were used to cluster molecules into shape-similar groups. Panels A−C represent samples of shape-similar clusters identified for this set of molecules. Each molecule is represented as a ball-and-stick figure. Surrounding each molecule is a semitransparent shell representing its atomic volume. These three examples represent a diverse set of molecular shapes. Additional details can be found in the Examples: Synthetic Compounds section of the text.

Table 1.

K	Γ_K	$n_1n_2n_3$	**V** (**BD**) before	**V** (**BD**) after	**V** (**T**)	contribusion of *S*
0	4.0715	000	4.1888	4.0726	4.0715	7.85×10^{-8}
1	1.8087	001	0	0.10379	0.10179	4.08×10^{-7}
2	0.80344	200	0.83776	0.79940	0.83059	1.22×10^{-3}
		020	0.83776	0.79940	0.83059	
		002	0.83776	0.80205	0.74916	
3	0.52327	201	0	2.0372×10^{-2}	6.1073×10^{-3}	2.44×10^{-4}
		021	0	2.0372×10^{-2}	6.1073×10^{-3}	
		003	0	6.1183×10^{-2}	7.7359×10^{-2}	
4	0.34080	400	0.35904	0.33624	0.35806	1.91×10^{-3}
		220	0.11968	0.11208	0.11935	
		202	0.11968	0.11260	0.11447	
		040	0.35904	0.33624	0.35806	
		022	0.11968	0.11260	0.11447	
		004	0.35904	0.33935	0.29129	
5	0.25300	401	0	8.5687×10^{-3}	8.2448×10^{-4}	3.23×10^{-4}
		221	0	2.8562×10^{-3}	2.7483×10^{-4}	
		203	0	8.5819×10^{-3}	4.4583×10^{-3}	
		041	0	8.5687×10^{-3}	8.2448×10^{-4}	
		023	0	8.5819×10^{-3}	4.4583×10^{-3}	
		005	0	4.2975×10^{-2}	5.9525×10^{-2}	
6	0.18782	600	0.19947	0.18333	0.19929	2.14×10^{-3}
		420	3.9893×10^{-2}	3.6666×10^{-2}	3.9858×10^{-2}	
		402	3.9893×10^{-2}	3.6884×10^{-2}	3.9198×10^{-2}	
		240	3.9893×10^{-2}	3.6666×10^{-2}	3.9858×10^{-2}	
		222	1.3298×10^{-2}	1.2295×10^{-2}	1.3066×10^{-2}	
		204	3.9893×10^{-2}	3.7103×10^{-2}	3.6071×10^{-2}	
		060	0.19947	0.18333	0.18661	
		042	3.9893×10^{-2}	3.6884×10^{-2}	3.9198×10^{-2}	
		024	3.9893×10^{-2}	3.7103×10^{-2}	3.6071×10^{-2}	
		006	0.19947	0.18661	0.14700	
						total 5.84×10^{-3}

A more detailed report will be provided in a future publication. All pairwise alignments were made for this set of compounds to yield ~ 150 000 similarity scores. Based on the vector of alignment scores for each compound against the remaining set, a standard single linkage hierarchical clustering method[23] was used to assign these compounds to "shape-similar" groups. No considerations were made with respect to atomic composition. The goal here is to make assignments only according to shape. Only a single molecular conformation was used for each compound; each obtained by sampling the lowest energy geometries derived from successive molecular dynamics and minimization simulations based on the CVFF force field of Discover95.0 (Biosym/MSI Inc., San Diego, CA).

As a measure of quality control of our method, we first examined congeneric compounds that differed by only a few atoms. These cases represent compounds that, aside from slight conformational differences, can be seen visually to have similar shapes. In all cases where only small atomic differences existed between these compounds, the best alignment scores placed these compounds within the same cluster groupings. An example of such a shape-based alignment is shown in Figure 8 where a near superposition of two camptothecin molecules is obtained, to yield a root-mean-square deviation near but not exactly zero. Clearly a comparable alignment could have been made based on matching the positions of their common atoms; however, our shape-based alignment does not require such atomic parsing. Within our set of ~400 compounds, there are eight cases where nearly identical atomic superpositions are obtained based on shape alignments.

A total of 42 clades were identified in our cluster analysis. The examples described below and shown as Figure 9 panels A−C represent clustered subsets of atomically diverse molecules aligned according to their shape. These three examples indicate that the methodology readily discriminates compounds of diverse molecular shape and size. Panel A represents a chemically similar set of RNA antimetabolites, most of which consist of linked five- and six-membered rings. Although these molecules appear to have a nearly spherical shape, there are sufficient details in this shape to place them into a cluster set based on their alignment scores. Panel B provides an example where the molecules adopt a more cylindrical shape. Inspection of this set finds the fused six-membered rings appearing only at the right of the molecule, in this view, to suggest that their shape is sufficient to place them in the proper orientation, relative to each other. Panel C represents another case where the overall shape in nearly spherical. However, each surface is quite convoluted, with these features being sufficient to place them into a common subgroup. Their shape-based alignments reveal that the underlying atomic substructures within each molecule are located in similar positions. Attempts to find this shape-similar molecular cluster based on root-mean-square distances between selected atoms could also have been used here, but difficulties in selecting those pairwise atomic distances to yield the best match complicate such an approach.

These findings demonstrate that shape-similar alignments of small molecules are possible with our current algorithm. A more detailed assessment of the performance of this method against a larger compound library is currently underway. Based on these preliminary results, our method may provide an efficient means to extract compounds that adopt similar shapes and sizes. Since shape complementary is only a small part of the interaction between a ligand and

Table 2. 84-Vectors for Parking the Prism **AN** onto the Cylinder **AC**[a]

K	Γ_K	$n_1n_2n_3$	**V** (**AN**) before	**V** (**AN**) after	**V** (**AC**)	contribution to *S*
0	157.08	000	225.00	157.26	157.08	1.35×10^{-6}
2	442.88	200	468.75	258.03	245.44	2.72×10^{-4}
		020	468.75	258.03	245.55	
		002	1518.8	836.03	837.76	
4	3064.8	400	1757.8	762.08	766.99	2.69×10^{-4}
		220	976.56	423.38	255.66	
		202	3164.1	1371.8	1309.0	
		040	1757.8	762.08	766.99	
		022	3164.1	1371.8	1309.0	
		004	18453	8000.1	8042.5	
6	29946	600	7847.4	2679.5	2996.1	1.22×10^{-4}
		420	3662.1	1250.4	599.21	
		402	11865	4051.4	4090.6	
		240	3662.1	1250.4	599.21	
		222	6591.8	2250.8	1363.5	
		204	38443	13126	12566	
		060	7847.4	2679.5	2996.1	
		042	11865	4051.4	4090.6	
		024	38443	13126	12566	
		006	2.6691×10^5	91135	91914	
						total 6.65×10^{-4}

[a] Initially, the prism has dimensions $5 \times 5 \times 9$, while the cylinder has diameter 5 and height 8. The long axes of each object coincide with the *z*-axis and each object had its center-of-mass at the origin. Only nonzero components are displayed. The optimum alignment was a change of scale ($\lambda = 0.8875$) without translation or rotation. After this change of scale, all the moments except *V* (220), *V* (420), *V* (240), and *V* (222) agree to about two significant figures or more.

Table 3. Components of the 84-Vectors for the 5-Fold Fans **CA** and **CB**, both before and after Parking **CA** onto **CB**[a]

K	Γ_K	$n_1n_2n_3$	**V** (**CA**) before	**V** (**CA**) after	**V** (**CB**)	contribution to *S*
0	39.2	000	39.3	39.2	39.2	1.57×10^{-7}
1	80.3	001	0.0110	−0.158	−0.261	6.32×10^{-7}
2	164	200	245	245	245	5.38×10^{-7}
		020	246	245	245	
		002	3.28	3.27	3.27	
3	521	300	0.0248	−1.59	−1.83	2.16×10^{-7}
		210	−0.446	−1.36	−1.34	
		030	−1.04	−3.81	−3.79	
4	1650	400	3060	3060	3060	8.14×10^{-7}
		220	1030	1020	1020	
		202	20.5	20.4	20.4	
		040	3070	3070	3060	
		022	20.5	20.5	20.5	
5	6050	500	−3.35	−1100	−1100	5.13×10^{-7}
		410	1390	886	884	
		320	−2.28	1050	1060	
		230	−1400	−920	−912	
		140	0.917	−1070	−1060	
		050	1370	825	837	
6	22200	600	47800	47700	47700	4.42×10^{-7}
		420	9620	9550	9570	
		402	255	254	255	
		222	86	85	85	
		240	9600	9600	9590	
		060	48000	47900	47800	
		042	256	256	255	
						total 3.31×10^{-6}

[a] In this particular case the initial 84-vectors were computed by Monte Carlo integration, are only displayed with three-figure precision, and are not displayed at all unless they have, either before or after alignment, magnitudes greater than about 0.1% of Γ_K. Originally, the two fans were rotated 10° relative to one another. The alignment algorithm compensated for this by applying a rotation of −9:9° (the 0:1° discrepancy is due to the inaccuracy of the Monte Carlo integration.) This table displays the rotational invariance of the components with $K < 5$. Interestingly, the $K = 6$ components are also unchanged. Whether this is true of arbitrary rotations or only of rotations through 10° is not known. In any case, the $K = 5$ components provide enough information to bring the two bodies into excellent alignment. Due to the inaccuracies of the Monte Carlo integration, the parking algorithm applied a very slight change of scale ($\lambda = 0.9997$); this accounts for changes of about one part in a thousand of the components of **CA** before and after parking.

its target,[24] additional considerations that include details of atomic interaction energies must be incorporated into this methodology. Our future efforts will examine the assignment of various weights to individual atoms within a molecule so as to emphasize alignments between selected subregions. (See eq 28 below.)

Table 4. 84-Vectors for the Two Camptothecin Molecules Appearing in Figure 8[a]

K	Γ_K	$n_1n_2n_3$	**V** (**CA**) before	**V** (**CA**) after	**V** (**CB**)	contribution to S
0	33	000	32	32	33	9.18×10^{-4}
1	82.354	100	0	1.238	0	2.11×10^{-4}
		010	0	−0.504	0	
		001	0	−1.585	0	
2	205.52	200	351.06	108.57	119.34	3.39×10^{-3}
		110	90.93	43.39	38.75	
		101	−40.95	−51.42	−67.85	
		020	141.90	378.50	380.28	
		011	35.15	−16.66	−13.44	
		002	90.00	96.03	116.94	
3	829.12	300	−384.35	5.92	6.92	4.88×10^{-4}
		210	−367.30	89.66	88.03	
		201	−20.44	−100.36	−108.50	
		120	−172.63	77.16	55.59	
		111	149.20	−89.06	−101.68	
		102	285.58	236.43	254.43	
		030	−28.72	−800.39	−812.83	
		021	15.96	77.35	102.82	
		012	−106.35	213.94	245.83	
		003	−252.18	−301.17	−323.92	
4	3344.9	400	6517.2	807.5	1022.6	1.03×10^{-3}
		310	2259.6	274.4	278.7	
		301	−500.5	−490.6	−680.8	
		220	1855.0	737.9	740.2	
		211	257.7	−230.1	−247.5	
		202	1026.6	743.5	862.9	
		130	1066.7	150.5	22.6	
		121	−224.5	−157.5	−207.8	
		112	−360.3	525.2	534.0	
		103	−858.2	−987.8	−1072.9	
		040	1451.3	9301.6	9274.4	
		031	141.1	−29.4	96.6	
		022	282.5	548.7	623.2	
		013	577.9	−582.6	−565.5	
		004	1545.4	1810.3	1974.9	
5	16327.3	500	−16804.5	−9.3	285.4	1.18×10^{-4}
		410	−10824.7	964.8	1122.6	
		401	574.0	−1150.0	−1500.5	
		320	−5654.9	408.5	305.8	
		311	2004.3	−877.2	−1076.2	
		302	3417.8	2188.5	2465.7	
		230	−4243.9	−420.4	−578.7	
		221	95.6	−243.7	−181.0	
		212	−1368.2	1603.3	1694.2	
		203	−2849.3	−3120.6	−3271.9	
		140	−2917.5	1711.7	1501.0	
		131	936.7	−353.4	334.4	
		122	840.4	1192.6	1103.5	
		113	1984.6	−1959.6	−1954.3	
		104	5298.1	5480.8	5554.0	
		050	−801.7	−38318.5	−38508.6	
		041	524.4	2297.1	2513.0	
		032	−409.8	1349.9	1544.1	
		023	−1083.4	−1038.8	−837.2	
		014	−2770.6	3838.7	3968.6	
		005	−6977.4	−8783.5	−8789.7	
6	79699	600	157515	7797	12111	3.81×10^{-4}
		510	64921	2075	2813	
		501	−7416	−5105	−8418	
		420	42407	3683	3955	
		411	1135	−2541	−3190	
		402	12836	7132	9117	
		330	25426	1485	1010	
		321	−3942	−1676	−2068	
		312	−4333	4731	4915	
		303	−9515	−10071	−10976	
		240	20509	11622	11593	
		231	64	−584	−391	
		222	3254	3657	3654	
		213	6756	−6237	−5989	
		204	18285	17055	16788	
		150	14816	−5398	−8456	
		141	−1630	239	−418	
		132	−1246	2771	2448	

Table 4 (Continued)

K	Γ_K	$n_1n_2n_3$	**V** (**CA**) before	**V** (**CA**) after	**V** (**CB**)	contribution to S
		123	−3814	−3778	−3465	
		114	−9529	11476	10874	
		105	−23723	−27664	−26077	
		060	19065	313279	311661	
		051	507	−2621	767	
		042	1451	4328	4978	
		033	2362	−1595	−772	
		024	5664	8312	8281	
		015	14408	−17716	16340	
		006	38693	49445	48063	
						total 6.53×10^{-3}

[a] Components were computed as discrete sums over atomic centers, excluding hydrogens. Both molecules were initially situated with centroids at the origin and with arbitrary relative rotational orientation. The change-of-scale adjustment was switched off ($\lambda = 1$). The alignment consisted of a small translation of about 0.06 Å and a rotation that brought the two molecules into excellent alignment.

NUMERICAL EXAMPLES

To demonstrate the behavior of the moments during the parking procedure, we display the 84-vectors obtained from several of the alignments discussed above. Table 1 displays the 84-vectors for the two shapes **BD** and **T** both before and after parking **BD** onto **T**. Only those components that are nonzero for at least one of the three objects (test before alignment, test after alignment, or target) are displayed. Prior to alignment, **BD** is a sphere of radius 1 centered at the origin, and **T** is a spherical cap (see Figure 1), also of radius 1 and with center-of-curvature at the origin. The slice removed from **T** is taken from the "south" pole, so that the center-of-mass of **T** lies at (0, 0, 0.250). The alignment algorithm both shrinks the sphere by approximately 1% ($\lambda = 0.9907$) and translates its center to (0, 0, 0.255), resulting in a near perfect agreement between the volumes and the positions of the centers-of-mass of **T**. Table 1 also gives the values of Γ_K and the contribution to S from each individual rank.

CONCLUSIONS

The 84 moments, defined in eq 1 with $K = 0, 1, ..., 6$, constitute a robust set of molecular shape descriptors that permit alignment of similar shapes and permit assessment of the degree of similarity between distinct shapes. The components are neither translationally nor rotationally invariant, but their behavior under arbitrary translations or rotations can be calculated rapidly. Therefore, they must be calculated initially in some particular reference frame, but this calculation need be done only once for any particular structure. Then the task becomes one of determining the coordinate transformation that best forces agreement between the components of any given pair of shapes. We have also explored some of the symmetry properties of these moments, which helps us anticipate and explain the behavior of the alignment procedure on shapes of high symmetry.

We have developed a two-stage alignment algorithm based on these moments. The first stage is equivalent to aligning the inertia ellipsoids of the two bodies. The second is a conjugate gradient minimization of a least squares sum, eq 23, of the squared differences of the components of the two bodies. In general, the alignment procedure works very well. In fact, we have been able to find only two classes of problems for which the algorithm, as currently conceived, presents any problems. First, since the initial alignment is equivalent to an alignment of the inertia ellipsoid, it becomes arbitrary in cases for which the moments of inertia are degenerate. The second problem arises because the tensors $\mathbf{V}^{(0)}$, ..., $\mathbf{V}^{(n-1)}$ of bodies with an n-fold symmetry axis are all invariant to rotations about the axis. Since we only employ tensors through $\mathbf{V}^{(6)}$, problems arise when considering bodies having a C_7 or higher axis. Neither of these problems are inherently insurmountable. The first can be addressed by designing the algorithm to examine all degenerate alignments of the inertia ellipsoid. The second could be overcome by using even higher tensors, $\mathbf{V}^{(7)}$, $\mathbf{V}^{(8)}$, etc., but since C_7 or higher symmetry elements are very rare, we feel no strong motivation to do so. Note also that both these problems only become acute with bodies of high symmetry. The algorithm as it currently stands is adequate for low symmetry molecules, such as proteins, and has already been applied to protein alignment.[11]

Equation 1 defines these tensors as moments of the scalar function that is 1 inside a body and 0 outside. There is no reason that we could not align molecules based on moments of other functions

$$\int d\mathbf{r}\, \psi(\mathbf{r})\, r_\alpha r_\beta r_\gamma r_\delta ... \tag{28}$$

where $\psi(\mathbf{r})$ could represent any function, e.g., the electron density or some sort of hydrophobicity scale. Therefore, it should be possible, using these same techniques, to design algorithms that align not only based on shape but also on reactivity or other molecular properties. Efforts along these lines are currently underway.

APPENDIX 1

Expressions for Γ_K and $\langle V(n_1, n_2, n_3)\rangle$.

$$\Gamma_0 = V(000) \tag{29}$$

$$\Gamma_2 = \left(\frac{1}{3}\right)[V(200) + V(020) + V(002)] \tag{30}$$

$$\Gamma_4 = \left(\frac{1}{5}\right)\{V(400) + V(040) + V(004) + 2[V(220) + V(202) + V(022)]\} \tag{31}$$

$$\Gamma_6 = \left(\frac{1}{7}\right)\{V(600) + V(060) + V(006) + 3[V(420) + V(402) + V(042) + V(240) + V(204) + V(024)] + 6V(222)\} \quad (32)$$

$\langle V(n_1, n_2, n_3)\rangle$ is zero whenever n_1, n_2, or n_3 is odd. Furthermore, $\langle V(n_1, n_2, n_3)\rangle$ is invariant to exchange of indices, so we need only give the expressions for which the indices are nondescending:

$$\langle V(00K)\rangle = \Gamma_K; \langle V(022)\rangle = \Gamma_4 = 3; \langle V(024)\rangle = \Gamma_6/5; \langle V(222)\rangle = \Gamma_6/15 \quad (33)$$

APPENDIX 2. TWO IMPORTANT INTEGRALS

The components of objects having rotational symmetry about the *z*-axis always involve these quantities:

$$S(n_1, n_2) = \int_0^{2\pi} \cos^{n_1}\phi \sin^{n_2}\phi \, d\phi \quad (34)$$

Write sines and cosines in terms of complex exponentials and expand by the binomial theorem. We obtain $S(n_1, n_2) = 0$ unless both n_1 and n_2 are even. Then we obtain

$$S(n_1, n_2) = 2\pi\left(\frac{1}{2}\right)^{2\alpha} i^n \sum_{j=0}^{n} (-1)^j \binom{n}{j}\binom{N}{\alpha - j} \quad (35)$$

where $\alpha = (n_1 + n_2)/2$, $n = \min(n_1, n_2)$, and $N = \max(n_1, n_2)$. This form implies that $S(n_1, n_2)$ is invariant to interchange of n_1 and n_2.

If the body also has spherical symmetry, then these integrals are required:

$$T(\alpha, n_3) = \int_0^{\pi} \sin\theta \, d\theta \sin^{2\alpha}\theta \cos^{n_3}\theta \quad (36)$$

These can be evaluated with the change of variable $t = \cos\theta$. We obtain $T(\alpha, n_3) = 0$ if n_3 is odd. If n_3 is even

$$T(\alpha, n_3) = 2\sum_{j=0}^{\alpha} (-1)^j \binom{\alpha}{j} (n_3 + 2j + 1)^{-1} \quad (37)$$

For example, the moments for a sphere of radius R are

$$V(n_1, n_2, n_3) = S(n_1, n_2)T(\alpha, n_3)(3 + K)^{-1} R^{(K+3)} \quad (38)$$

APPENDIX 3. ESTABLISHMENT OF SYMMETRY PROPERTIES

The $P(n_1, n_2, n_3)$ sums defined in eq 12 over appropriate *n*-tuples of points permit us to prove general properties of the moments when the body possesses certain symmetry elements.

Center of Inversion. Possession of a center of inversion is a necessary and sufficient condition to force all odd-*K* moments to be zero in any reference frame for which the inversion center lies at the origin. Proof: The *P* sum for the doublet (*a*, *b*, *c*) and (−*a*, −*b*, −*c*) is

$$P(n_1, n_2, n_3) = [1 + (-1)^K]a^{n_1}b^{n_2}c^{n_3} \quad (39)$$

which is zero for *K* odd. Conversely, $G(\mathbf{k})$ for any body for which all odd-*K* moments are zero is real. The symmetry properties of real Fourier transforms indicates that such a body has an inversion center.

Reflection Plane. Existence of a reflection plane coinciding with one of the three planes $x = 0$, $y = 0$, or $z = 0$ is a necessary and sufficient condition to force any component for which the respective index, n_1, n_2, or n_3, is odd to be zero. Proof: The *P* sum for the doublet (*a*, *b*, *c*) and (*a*, *b*, −*c*) is

$$P(n_1, n_2, n_3) = [1 + (-1)^{n_3}]a^{n_1}b^{n_2}c^{n_3} \quad (40)$$

which is zero for n_3 odd. The converse again follows from the symmetry properties of Fourier transforms.

***n*-Fold Rotation Axis.** An arbitrary point (*a*, *b*, *c*) and an additional $n - 1$ image points generated by rotations through $2\pi/n$ about the *z*-axis are considered. Define the complex number $u = a + ib$. Then the *x* and *y* coordinates of each point are the real and imaginary parts, respectively, of $u\Phi^j$, $j = 1, 2, 3, \ldots, n$, where $\Phi = \exp(2\pi i/n)$, while the *z* coordinates of all points are *c*. Then

$$P(n_1, n_2, n_3) = \sum_{j=1}^{n} \left[\frac{u\Phi^j + u^*\Phi^{*j}}{2}\right]^{n_1} \left[\frac{u\Phi^j - u^*\Phi^{*j}}{2i}\right]^{n_2} c^{n_3} \quad (41)$$

where * indicates the complex conjugate. The two terms in brackets can be expanded by the binomial theorem. We also apply the identity

$$\sum_{j=1}^{n} \Phi^{N_j} = \begin{cases} n, & \text{if } N \text{ is a multiple of } n, \\ 0, & \text{otherwise,} \end{cases} \quad (42)$$

which is true for *N* any integer. It follows that we can write

$$P(n_1, n_2, n_3) = n \sum_{\sigma=\sigma_{\min}}^{\sigma_{\max}} c^{n_3}\left(\frac{1}{2}\right)^{n_1}\left(\frac{1}{2i}\right)^{n_2} \times \sum_{p=0}^{n_1}\sum_{q=0}^{n_2} (-1)^{n_2-q}\binom{n_1}{p}\binom{n_2}{q} u^{(p+q)} u^{*(n_1+n_2-p-q)} \delta(2p + 2q - n_1 - n_2, n\sigma) \quad (43)$$

where the σ-sum ranges over all integers between $\sigma_{\min}$ and $\sigma_{\max}$ (to be defined shortly) and where $\delta(\ldots,\ldots)$ is the Kroneker δ. Acceptable values of *p* and *q*, $0 \le p \le n_1$ and $0 \le q \le n_2$, respectively, define a rectangular region on the simple-square lattice. (See Figure 7.) The Kroneker δ restricts us to points such that $p + q = (n\sigma + n_1 + n_2)/2$, i.e., lines of slope −1 that pass diagonally through the lattice. Each integral value of σ generates one such line, and we need only consider those lines that pass through the rectangle generated by acceptable values of *p* and *q*. Therefore $\sigma_{\min}$ and $\sigma_{\max}$ must be defined so that all lines passing through the rectangle are included. For the moment, consider continuous values of σ. The value of σ that produces the $p + q$ line intersecting the point $(p, q) = (0, 0)$ is $\sigma = -(n_1 + n_2)/n$; that producing the line intersecting $(p, q) = (n_1, n_2)$ is $\sigma = +(n_1 + n_2)/n$. The necessary range of σ is all integers such that $|\sigma| < (n_1 + n_2)/n$, which effectively defines $\sigma_{\min}$ and $\sigma_{\max}$. Since the lines obey $p + q = (n\sigma + n_1 + n_2)/2$, we can also omit any values for which $(n\sigma + n_1 + n_2)$ is odd.

The sum in eq 43 is awkward enough that one generally ought to consider direct sums over the n original points in any numerical work. However, an important special case is $n_1 + n_2 < n$, for which only $\sigma = 0$ contributes. Then the sum can eventually be worked into the same form as eq 35, yielding

$$P(n_1, n_2, n_3) = \frac{nc^{n_3}}{2\pi} (a^2 + b^2)^{\alpha} S(n_1, n_2) \qquad (44)$$

where $S(n_1, n_2)$ is given by eq 35.

Whenever $n_1 + n_2 < n$, $P(n_1, n_2, n_3)$ is symmetric with respect to interchange of n_1 and n_2, and it depends on a and b only through $a^2 + b^2$. Therefore, it is invariant not only with respect to rotations through $2\pi/n$ as we would expect but also with respect to arbitrary rotations about the z-axis. Whenever $n_1 + n_2 \geq n$ additional σ terms contribute, and the rotational invariance is lost.

In the most general case, $P(n_1, n_2, n_3)$ is given by eq 43. Whenever $n_1 + n_2 < n$, the simpler result, eq 44, holds. Obviously, $P(n_1, n_2, n_3)$ is always invariant to rotations about the C_n axis through angles of $2\pi/n$, but when $n_1 + n_2 < n$, eq 44 indicates that $P(n_1, n_2, n_3)$ is invariant to arbitrary rotations about the axis. This implies that the moments $\mathbf{V}^{(0)}$, $\mathbf{V}^{(1)}$, $\mathbf{V}^{(2)}$, ..., $\mathbf{V}^{(n-1)}$ are also invariant to arbitrary rotations about the axis. Furthermore, this invariance holds in any reference frame, since the moments in an arbitrary frame can first be transformed to a principal axis frame, rotated about the axis, and then back-transformed. Furthermore, eq 44 indicates that two distinct bodies, one having an n-fold axis and the other having an m-fold axis, with $n < m$, possessing identical radial and longitudinal density distributions, would have precisely the same moments up to $\mathbf{V}^{(n-1)}$. Equation 44 also indicates that in the principal axis frame, all components for which either n_1 or n_2 is odd and for which $n_1 + n_2 < n$ are 0.

***m*-Fold Improper Rotation Axis.** Obviously, we can take m to be even. Then setting $n = m/2$, the analysis developed above for the C_n axis can be applied. We consider n points generated from a point (a, b, c) by rotations through $2\pi/n$ and an additional n points at $z = -c$ and offset from the first set by a rotation through π/n. Therefore, the requisite P-sum is obtained as the expression in eq 43 plus a similar expression, with u replaced with $v = uE$, where $E = \exp(\pi i/n)$, and with c replaced with $-c$. We find that E only appears in this second expression as $E^{n\sigma}$ $(-1)^{\sigma}$. Therefore, the requisite P-sum can be obtained by inserting a factor $[1 + (-1)^{n_3+\sigma}]$ into eq 43. Once again, when $n_1 + n_2 < n = m/2$, the analysis simplifies further, and the requisite P-sum can by obtained by inserting the factor $[1 + (-1)^{n_3}]$ into eq 44. Therefore, $V(n_1, n_2, n_3)$ vanishes whenever $n_1 + n_2 < n = m/2$ and n_3 is odd if the S_m axis is aligned with the z-axis and if the reflection plane is aligned with the $z = 0$ plane.

ACKNOWLEDGMENT

The authors would like to acknowledge the support of the IRSP program at SAIC, NCI-FCRDC under contract NO1-56000. IRSP, SAIC, and NCI-FCRDC are funded in part by NO1-56000.

REFERENCES AND NOTES

(1) Arteca, G. A.; Mezey, P. G. Shape Characterization of Some Molecular Model Surfaces. *J. Comput. Chem.* **1988**, *9*, 554−563.

(2) Balasubramanian, K. Computer Perception of Molecular Symmetry. *J. Chem. Inf. Comput. Sci.* **1995**, *35*, 761−770.

(3) Randic, M.; Razinger, M. Molecular Shapes and Chirality. *J. Chem. Inf. Comput. Sci.* **1996**, *36*, 429−441.

(4) Handschuh, S.; Wagener, M.; Gasteiger, J. Superposition of Three-Dimensional Chemical Structures Allowing for Conformational Flexibility by a Hybrid Method. *J. Chem. Inf. Comput. Sci.* **1998**, *38*, 220−232.

(5) Leicester, S. E.; Finney, J. L.; Bywater, R. P. Description of Molecular Surface Shape Using Fourier Descriptors. *J. Mol. Graphics* **1988**, *6*, 104−108.

(6) Katchalski-Katzir, E.; Shariv, I.; Eisenstein, M.; Friesem, A. A.; Aflalo, C.; Vakser, I. A. Molecular Surface Recognition: Determination of Geometric Fit Between Proteins and their Ligands by Correlation Techniques. *Proc. Natl. Acad. Sci. U.S.A.* **1992**, *89*, 2195−2199.

(7) Leicester, S.; Finney, J.; Bywater, R. A Quantitative Representation of Molecular Surface Shape. I: Theory and Development of the Method. *J. Math. Chem.* **1994**, *16*, 315−341.

(8) Leicester, S.; Finney, J.; Bywater, R. A Quantitative Representation of Molecular Surface Shape. II: Protein Classification Using Fourier Shape Descriptors and Classical Scaling. *J. Math. Chem.* **1994**, *16*, 343−365.

(9) Grant, J. A.; Pickup, B. T. A Gaussian Description of Molecular Shape. *J. Phys. Chem.* **1995**, *99*, 3503−3510.

(10) Grant, J. A.; Gallarldo, M. A.; Pickup, B. T. A Fast Method of Molecular Shape Comparison: A Simple Application of a Gaussian Description of Molecular Shape. *J. Comput. Chem.* **1996**, *17*, 1653−1666.

(11) Mansfield, M. L.; Covell, D.; Jernigan, R. L. A New Class of Molecular Shape Descriptors. II. Application to the Alignment of Protein Molecules. Manuscript in preparation.

(12) Byron, F. W.; Fuller, R. W. *Mathematics of Classical and Quantum Physics*; Addison-Wesley: Reading, MA, 1969; Vol. 1, p 33ff.

(13) Goldstein, H. *Classical Mechanics*; Addison-Wesley: Reading, MA, 1950; p 109.

(14) Press, W. H.; Teukolsky, S. A.; Vetterling, W. T.; Flannery, B. P. *Numerical Recipes in Fortran 77*, 2nd ed.; Cambridge University Press: Cambridge, U.K., 1992; p 413ff.

(15) Aravind, P. K. A Comment on the Moment of Inertia of Symmetrical Solids. *Am. J. Phys.* **1992**, *60*, 754−755.

(16) *Cancer Chemotherapeutic Agents, Prediction of Biochemical mechanism of action from the in vitro antitumor screen of the National Cancer Institute*; Paull, K. D., Hamel, E., Malspeis, L., Foye, W. O., Eds.; American Chemical Society: Washington, DC, 1995; Chapter 2.

(17) Klebe, G. Recent developments in structure-based drug design. *J. Mol. Med.* **2000**, *78(5)*, 269−281.

(18) Meyer, E. F.; Swanson, S. M.; Williams, J. A. *Molecular Modeling Drug Design, Pharmacol. Therapeutics* **2000**, *85*, 113−121.

(19) Maggiora, G.; Johnson, M. A. *Concepts and Applications of Molecular Similarity*; John Wiley: New York, 1990.

(20) M. Randic, On characterization of chemical structure. *J. Chem. Inf. Comput. Sci.* **1997**, *37*, 672−687.

(21) Pearlman, R. S.; Smith, K. M. Novel software tools for chemical diversity. *Perspectives Drug Discovery* **1998**, *9−11*, 339−353.

(22) Martin, Y. C.; Willet, P. *Designing Bioactive Molecules*; American Chemical Society: Washington, DC, 1998.

(23) Sneath, P. H. A.; Sokal, R. R. *Numerical Taxonomy*; W. H. Freeman and Company: San Francisco, 1973.

(24) Wallqvist, A.; Covell, D. G. Docking Enzyme−Inhibitor Complexes using a Preference Based Free-Energy Surface. *Proteins; Struct., Funct., Genet.* **1996**, *25*, 403−419.

CI000100O