Mapping Complicated Surfaces onto a Sphere

Sahand Jamal Rahi, Kim Sharp*
Department of Biochemistry and Biophysics, University of Pennsylvania School of Medicine, Philadelphia, Pennsylvania 19104

*To whom correspondence should be addressed at
Dept. of Biochemistry, University of Pennsylvania
3700 Hamilton Walk
Philadelphia, PA
19104
tel: 215-573-3506
fax: 215-898-4217
email: sharpk@mail.med.upenn.edu

Abstract

A method for mapping a closed, triangulated surface of genus 0 onto the unit sphere was developed using a parameterization based on spherical coordinates analogous to latitude and longitude combined with the Mercator scaling. The algorithm was tested on 77 protein and DNA surfaces and produces correct (bijective) mappings. The mappings produce relatively uniform distortion, as judged by angular criteria, and by distribution of octant areas, thus facilitating the graphical analysis of these surfaces.

1. Introduction

Many objects have very complicated shapes. In addition, the surface may be interesting because it has important properties. For example, the solvent-accessible surface (SAS) of a protein or nucleic acid is usually very complicated. Based on the nature of the atoms below the surface, electrostatic charge, potential, hydrophobicity, or other physical properties can be associated with various areas of the surface. Such surfaces with specific distributions of properties on them are crucial for determining other properties of these molecules, such as how they interact with solvent, ligands, and other macromolecules and ultimately for macromolecular function. This work was motivated by the desire to produce for these complicated surfaces something analogous to a planar map of the earth: In everyday use, a true representation of the earth as a spheroid is too cumbersome. A planar map of the earth can capture any feature of the earth's surface, such as borders between countries or height above sea level, in an illustrative manner. However, one must contend with distortion. Applying this idea, one would like to construct a map of a more complicated surface such as that of a protein, but instead of mapping the surface onto the plane, it would be more natural to map it onto the (unit) sphere, onto which one can easily project surface properties such as hydrophobicity or polarity, and visualize them using computer graphics. Given that we are mapping onto a sphere, in what follows we restrict our discussion of complicated surfaces by requiring them to be homeomorphic to a sphere (genus 0), for instance, they should not have "holes" like tori or more than one component such as a buried cavity.

Such a mapping has several potential applications. The most immediate is that it provides in just two orientations (using cartographic terminology say east and west hemisphere views) visual access to the entire surface. As anyone who has used computer graphics programs to study proteins and nucleic acids can attest, just viewing macromolecular surfaces is not trivial. Finding the best orientation to view important parts of the surface such as active site clefts is difficult, and for convoluted surfaces it is impossible without clipping or using partially transparent surfaces. Simultaneously viewing visible but disparate parts of the surface with a single orientation is also often not possible. More quantitative applications suggest themselves, including comparison of patterns of surface properties between structurally dissimilar proteins of similar function or comparison across a series of functionally important surface regions such as active site clefts. Mapping onto a standard surface such as a sphere would remove the confounding variables of protein size and shape. Of course, any mapping would produce distortions in either area, shape (non-conformal), or both, so the types of quantities one could compare must be invariant to such distortions. These would include such 'topological' properties as the number, distribution, and propinquity of regions of surface with specific properties such as hydrophobicity and charge.

The work described here was motivated by our interest in analyzing protein and nucleic acid structures but mapping of a complicated surface onto a sphere has applications in other areas, including re-meshing of complex surface rendering in computer graphics, medical imaging, and analyzing the surfaces of other complex objects, e.g. asteroids. Remeshing, the process of producing coarser meshes by averaging or finer meshes by interpolation (Eck, DeRose et al. 1995), is facilitated by mapping the surface to a sphere or plane since the averaging or interpolation is straightforward there. One then inverts the mapping to produce the required mesh.

Such a mapping to a sphere (also called an embedding or surface parameterization), while simple to describe, is not an easy problem. One difficulty is that the mapping must be a homeomorphism. For example, if the surface is represented by a triangular mesh (a triangulation), the mapping would not be bijective if triangles on the spherical surface were flipped over (the surface is folded), or if the triangles do

not completely cover the surface. For convoluted protein and nucleic acid surfaces this immediately eliminates simple approaches like those based on geometric projection. One procedure specifically for mapping protein surfaces to a sphere has previously been described (Duncan and Olson 1995). This uses a spherical harmonic approximation of the surface, and so it is not strictly a mapping of the original surface. In the field of surface image processing, other genus 0 surface mapping procedures have been described (Brechbuhler, Gerig et al. 1995; Haker, Angenent et al. 2000; Gu and Yau 2002; Gotsman, Gu et al. 2003; Praun and Hoppejournal 2003; Gu, Wang et al. 2004) see for a review (Floater and Hormann 2004). No one approach has been found to be universally suitable. The method of Brechbuhler et al., developed for uniform area quadrilateral element surfaces, produced for example from voxel data, involves a preliminary mapping to a sphere. This is followed by optimization of the mapping by moving points on the sphere surface. The method of Gu et al. requires pre-editing of the surface, which for our application is not desirable. The method of Haker et al. uses a stereographic projection step. This procedure is elegant, but we find that for our type of macromolecular surfaces it is difficult to control the scaling in this method to produce relatively uniform distortion over the entire molecule. The possibility arises that surfaces arising in different disciplines such as medical imaging, computer graphics, structural biology etc. are complex in 'different' ways. Hence, they may be mapped 'better' by one method than another. Combined with the paucity of available methods, it is desirable to develop an additional method that would produce correct mappings for macromolecular surfaces in an essentially automatic way with relatively uniform distortions. However, the method that we developed is quite general in the sense that it does not rely on any specific properties of macromolecular surfaces and so it would work for any genus 0 surface, and it is surprisingly simple.

2. Methods

2.1. Obtaining a Triangulated Solvent-Accessible Protein Surface

This part of our algorithm is specific to molecules such as proteins, the rest of the algorithm is general and only requires a closed, triangulated surface of genus 0 (homeomorphic to a sphere) as defined in Giblin (Giblin 1981). Orientability is guaranteed by surface closure and the genus. Several existing algorithms for computing the solvent-accessible surface (SAS) of a protein just produce a set of points in Cartesian space, i.e. triplets (x, y, z), which lie on the surface. For many purposes an actual surface is essential. One could, in principle, obtain such a surface by connecting these points with edges to produce a tesselation. In general, this is a difficult procedure. Procedures do exist, but they require conditions on the density of points which are difficult to satisfy in actual applications. Our solution to the problem of obtaining a properly triangulated representation of the molecular surface (MS) as defined by Lee and Richards (Lee and Richards 1971) is adapted from the macromolecular graphics program GRASP (Nicholls, Sharp et al. 1991) as follows.

1. The protein's atomic coordinates are obtained from a protein data bank (pdb) file (Berman, Westbrook et al. 2000).
2. The atomic radii are taken from standard values (Bondi 1968).
3. A fine cubic grid is laid over the protein and a small region of the surrounding space.
4. All grid points that lie closer than one atomic radius plus the radius of a solvent probe (representing water, with a value of 1.8Å) are labeled as interior grid points.
5. All interior grid points that lie on this expanded radius surface (i.e. that are labeled as interior and that have at least one exterior neighboring point) are identified.
6. A solvent probe radius sphere is centered on each of these surface grid points, and any grid point that lies within the solvent probe is relabeled as exterior.
7. Interior and exterior grid points are assigned arbitrary but different values, say 1.0 and 0.0, and a three-dimensional contouring routine is used to generate a closed, triangulated surface from the grid values by contouring at an intermediate value, here 0.5.

This algorithm, illustrated in Figure 1, is guaranteed to produce closed surfaces, which can be proved easily. Example of the surface produced by this algorithm are shown in later figures of the paper.

2.2 Preparing the Surface

The algorithm needs a list of vertices and their Cartesian coordinates $\{v_1=(x_1,y_1,z_1), v_2=(x_2,y_2,z_2),\ldots\}$ as well as a list of edges $\{\{v_1,v_2\},\{v_1,v_3\},\{v_2,v_3\},\ldots\}$ indicating how the vertices are connected to form a surface. This is necessary in order to find all vertices that are neighbors of a given vertex. As neighbors of a vertex i we denote vertices $i_1,\ldots,i_n$ that are directly connected to vertex i via an edge. The set of vertices and edges constitute a simple graph of a particular type called a triangulation. At this stage it is straightforward to identify from the connectivity of the vertices the presence of cavities, which are manifest as smaller, disjoint sets of connected vertices (graph components). Currently, these are just discarded so that a single component graph remains but it would be straightforward to apply the same mapping procedure to each component if cavities were of interest. At this stage we also use the Euler equation $F + V - (E + 2) + 2H = 0$ where F, V, E and H are the number of triangles, vertices, edges and handles respectively, to filter out toroidal and handled surfaces.

The list of vertex triplets $\{(v_i,v_j,v_k), (v_l,v_n,v_n) \ldots\}$, each triplet representing a triangle, is sufficient to describe the surface: It is straightforward to generate the list of edges from the list of vertex triples. The orientation of the overall surface and each surface element is encoded in the order of vertices in the triplet through the direction of the outward surface normal of each triangle. While an algorithm such as that described in section **2.1** provides the triplets in the correct ordering, any surface-generating algorithm may be used and then the surface oriented as follows: The two normals of a triangle with vertices (v_i,v_j,v_k) are $\mathbf{n}_1=(v_j-v_i)\times(v_k-v_j)$, and $\mathbf{n}_2=-\mathbf{n}_1$. If $\mathbf{n}_1$ points outward, the order of vertices in the cross product is saved as (v_i,v_j,v_k). If not, two of the entries are switched, (v_i,v_k,v_j). Once the orientation is established for one triangle, the neighboring triangles are oriented by permuting the vertex triplets correctly. If (v_i,v_j,v_k) is the orientation of one triangle then the orientation of a neighboring triangle which shares an edge $\{v_i,v_j\}$ is (v_i,v_l,v_j) because $(v_l-v_i)\times(v_j-v_l)$ points outward. So, first one triangle needs to be oriented by finding its outward normal. The vertex v_m which has the greatest distance from the coordinate origin is selected. At least one of the triangles that v_m is a vertex of has a normal $\mathbf{n}_m$ whose angle with the vector from origin to v_m is strictly smaller than 90°. This normal must point outward from the region enclosed by the surface.

2.3. Mapping a Triangulated Surface onto a Sphere

Given a triangulated surface as specified above the mapping algorithm assigns to each vertex two parameters that determine the vertex position on the unit sphere. We follow Brechbuhler et al. (Brechbuhler, Gerig et al. 1995) and choose the spherical coordinates θ and ϕ. The angle θ denotes the angle with the z-axis and ϕ denotes the azimuthal angle on the unit sphere (Figure 3). θ and ϕ correspond to latitude and longitude in cartography terms, and they are assigned as follows

1. The two vertices with the largest Euclidean distance are assigned to be the North Pole N and the South Pole S. On the sphere, they have $\theta_N = 0$ and $\theta_S = \pi$. These points and their edges are then removed from the graph, which is now topologically equivalent to a cylinder.
2. Neighbor vertices of N and S are assigned fixed arbitrary coordinates $z = +Z$ and $-Z$ respectively We refer to these neighboring vertices as circumpolar vertices. Together with their joining edges the circumpolar vertices form the two circumpolar boundaries.
3. The z-coordinate of each vertex between the circumpolar boundaries is then set to the unweighted average of its neighbors' z-coordinates. This is done by assigning initial values of 0, and using Gauss-Seidel relaxation to converge to numerical precision.
4. To determine the azimuthal ϕ-coordinate a boundary meridian (BM) in the z-direction on the surface is generated by following a path of steepest descent starting at an arbitrary circumpolar vertex with z = + Z until a vertex with z = -Z is reached (Figure 6.i, black vertices). Steepest descent is guaranteed to produce a simple, i.e. non-self-intersecting, path since z does not assume a maximum or minimum except at the circumpolar vertices. The vertices on the boundary meridian are defined to have $\phi(v_i)=0$ if approached from their neighbors from the "East" (Set A Fig. 6.i) and $\phi(v_i)=2\pi$ if approached from their "Western" neighbors (Set B, Fig. 6.i), much as the time immediately on either side of the International Date Line (nominally longitude 180^o) differs

by 24hrs. The orientation of the triangles encoded in the vertex ordering is used to determine the correct direction of increase of f. If this were not available, one could simply choose one direction arbitrarily and then correct the mapping automatically later if found to give the mirror image of the correct mapping.

5. Circumpolar vertices are assigned fixed ϕ values to give a uniform spacing of vertices between 0 and 2π. The ϕ-coordinate of each non-BM, non-circumpolar vertex is then set to the unweighted average of its neighbors' ϕ-coordinates, incorporating the dual BM vertex values described above as appropriate. Assignment of ϕ is performed by initializing values to 0 and iterating to convergence using Gauss-Seidel relaxation
6. To produce the final mapping to the sphere it is necessary to transform z into the latitude coordinate θ. Linear scaling would result in vertices crowded around the equator. A suitable bijective mapping is the Mercator mapping, $z \to q = f^{-1}(z)$ where $f : [0, \pi] \to \Re \cup \{+\infty, -\infty\}$ is the function

$$f(q) = C \ln\left[(\cos q + 1)/\sin q\right] \qquad (1)$$

where C is a constant. The Mercator mapping is suggested by the analogy between harmonic mapping and heat diffusion on the sphere, equation (1) being a solution to the heat equation for the case of a 'hot' North polar and a 'cold' South pole with azimuthal symmetry. Other functions are no doubt possible. The northern circumpolar vertices with $z = +Z$ are mapped to q_0 and the southern circumpolar vertices with $z = -Z$ are mapped to $\pi - q_0$, which is equivalent to setting C to

$$C = Z\left[\ln\left(\left(\cos q_0 + 1\right)/\sin q_0\right)\right]^{-1}. \qquad (2)$$

where θ_o is also a constant. θ_o is chosen to relate the latitudes of the N and S circumpolar vertices to their separation in graph space, or more informally to the density of vertices between the two poles, as follows: The graph distance of the south pole from the north pole is determined by assigning a distance of d = 0 to the North pole. Neighbors of N are assigned a distance of 1. Neighbors of vertices with distance 1 that have not been assigned a distance are assigned a distance or 2. Proceeding iteratively in this manner for d = 2, 3…etc, all unassigned vertices that are neighbors of vertices with distance d are assigned a distance of d+1 until the south pole is reached at graph distance d_s. Then the constant is set to $q_0 = \pi/d_s$. This procedure for determining q_0 results in a suitable scaling term C because d_s is a measure of the north-south extent of the triangulated surface in number of edges.
7. As a final step, the mapped sphere is put into the most similar orientation and position to the original surface by using a standard rigid body superposition algorithm (Kearsley 1989) to minimize the root mean square difference in Cartesian coordinates between the two surfaces.

The Fortran code to perform the spherical mapping is available from the author at crystal.med.upenn.edu.

2.4. Correctness of the mapping

Is the mapping a homeomorphism?

To be useful, a mapping from a complicated surface onto the sphere must at least be a homeomorphism, that is, a bijective continuous mapping whose inverse is also continuous. Since the original surface is homeomorphic to a sphere, we know that a homeomorphism exists but it is necessary to demonstrate that our mapping is one. As long as the final triangulation of the sphere is closed as defined in Giblin (Giblin 1981), the two surfaces are homeomorphic because they also have the same genus. The two 'cuts' that are made around the poles by removing the N, S vertices are clearly repaired as $q_N = 0$ and $q_S = \pi$ and the circumpolar points are mapped to $\theta = \theta_o$ and $\pi - \theta_o$. The mapped triangles and the original triangles are in

bijective correspondence and are neighbors of the same triangles. As long as the mapping is a valid embedding, the original surface and the image agree locally and by closedness, orientability, and genus, they also agree globally.

Is the mapping a valid embedding?
A valid embedding, requires that the mapped triangles are non-degenerate, the intersection of any pair of mapped triangles is either empty, a common vertex, or a common edge. More informally, this requires that the surface not be folded back on itself anywhere evinced by 'flipped' over triangles. To ensure this, our algorithm is designed so that the circumpolar boundaries are at constant z/θ and the boundary meridian is at constant ϕ. With these properties, and by taking z and ϕ as rectilinear coordinates in a plane, our algorithm also maps the surface, excluding the pole vertices, into a region bounded by a rectangle (Figure 6.ii). Two sides of the rectangle are formed by the northern and southern circumpolar vertices at constant z=+/-Z, and uniformly increasing ϕ. The other two sides are formed by the duplicated Boundary Meridian vertices with either $f = 0$ or $f = 2p$, and monotonically descending z. Interior vertices have z-ϕ coordinates which are barycentric averages of their neighbors. If this mapping to a rectangular region of the plane is a valid embedding, it is easy to see that the mapping to a sphere is also a valid embedding by the sequence of (virtual) steps: Transform z into θ using the bijective Mercator function, Eq. 1; 'Roll up' the rectangle, gluing the BM points to themselves to produce a cylinder; Shrink the cylinder radius towards the poles appropriately; Replace the pole points.

We use the results of a generalization of Tutte's barycentric mapping theorem (Tutte 1963) by Floater (Floater 2003), specifically, his theorem 6.1:

> Suppose T is any triangulation and that ξ: $D_T \rightarrow \mathbf{R^2}$ is a convex combination mapping which maps ∂D homeomorphically into the boundary ∂Ω of some convex region $\Omega \subset \mathbf{R^2}$. Then ξ is one-to-one if and only if no dividing edge [v,w] of T is mapped by ξ into ∂Ω.

Here, the triangulation T is a valid embedding in the plane, D_T is the union of the triangles in T, ∂D refers to the boundary of the triangulated region D_T. We now show that our mapping satisfies each of the requirements in this theorem in turn.

i) To identify our triangulation with the triangulation T in the theorem, ours must be embeddable in the plane with a simple closed polygonal boundary. The original triangulated surface is closed and has genus zero, hence, it is orientable and is homeomorphic to a sphere. The image of the surface on the sphere under this homeomorphism can further be stereographically projected onto the plane showing that the original surface graph is planar, i.e., does not contain Kuratowski subgraphs. By Fary's Theorem there exists a straight-line embedding, which in this case is our triangulation T.

ii) A convex combination mapping ξ is one in which the coordinates of each image vertex (except on the image boundary ∂Ω) are the weighted arithmetic averages of the neighbors' coordinates, all weights are positive and sum to one. Thus, our barycentric mapping is a special case of a convex combination mapping with equal weights of 1/d, where d is the degree of the vertex. Moreover, the numerical procedure for computing the barycentric mapping in z and ϕ converges. Setting z for each vertex to be the average of its neighbors is equivalent to solving the matrix equation **Dz** = **b** where the off-diagonal elements of **D** are zero except when *i* and *j* are neighbors in which case $D_{ij} = -1$, and the diagonal terms are given by $D_{ii} = -\Sigma D_{ij \neq i}$, . Since **D** is diagonal dominant, the Gauss-Seidel relaxation converges (Iserles 1996). The same argument applies to the ϕ averaging.

iii) The boundary ∂D of the original surface is the path delineated as follows: Start at the BM vertex that is also a northern circumpolar point, walk clockwise around those northern circumpolar vertices back to the BM, call this piece of the path A. Then, walk along the BM to the southern circumpolar vertex at its end, call this piece B. Again, proceed along the southern circumpolar vertices in a clockwise direction back to the BM (path C) and back up the BM again to the first vertex we started from (path D). The clockwise

direction is judged from outside the surface. If the surface is cut in this manner, one obtains a closed polygonal boundary for the planar embedding T. This is only self-intersecting at the cut along the BM but Floater's theorem applies because the two sides of the cut can be pulled apart.

iv) ∂D is mapped homeomorphically to $\partial\Omega$ by the algorithm as follows: Paths B and D are mapped to the sides of a rectangle in the z, f-plane at $f = 0$ and $f = 2p$ respectively. Since they have been chosen by the descending z-coordinate condition, they are well-ordered. The paths A and C are placed at $z = -Z$ and $z = +Z$ respectively. The f-coordinates are also well ordered. Clearly all paths connect up properly at the corners (-Z, 0), (+Z, 0), (+Z,π), and (-Z,π). Hence the path follows the sides of the rectangle with no 'doubling' back. Thus the boundary is also closed, encloses a convex region, and is in homeomorphic correspondence with ∂D.

v) A dividing edge is an interior edge that joins two boundary vertices, i.e. an edge that joins two boundary points that are not neighbors along the boundary. In the current context, we are concerned with the triangulation in the z-ϕ plane (Figure 6.ii), where the relevant boundary vertices are the circumpolar and BM vertices.

There are three possible types of dividing edges.

a) Between two BM points. This cannot occur because the vertices are chosen in order of steepest descent: Let *i, j, k…* be a sequence of vertices generated along the Boundary Meridian by moving to each adjacent vertex with the lowest z-value. Let a dividing edge connect *i* with some boundary vertex *k* or further along in the sequence. This vertex must have a z-value less than that of *j*, but if this is the case, this vertex would have been chosen instead of *j* when moving from *i*, so there is no such dividing edge.

b) Between two circumpolar points - This would only happen if a circumpolar vertex v_i had degree 3: One of v_i's neighbors is the N or S pole. The other two neighbors are the circumpolar points v_{i-1} and v_{i+1}. Now v_i, v_{i-1} and v_{i+1} are the vertices of a triangle, so v_{i-1} and v_{i+1} would be connected by an edge, which is a dividing edge. However degree 3 points are rare, and are trivial to eliminate by subdividing two triangles with a shared edge, where one triangle has the offending vertex at the apex (Figure 6.iii). This increases the degree of v_i by one.

c) Between a circumpolar point and a BM point, i.e. 'across the corner'. Since the circumpolar vertices and the BM vertices are mapped to different sides of the rectangle, a dividing edge between them is obviously not mapped into the boundary but into the interior of the rectangle.

Since all the conditions for Floater's theorem hold our mapping from surface to rectangle is valid, subject to numerical error involved in any real implementation on a computer, and so therefore is the mapping to the sphere.

There is, however, one subtlety to note. The barycentric embedding is a straight-line embedding. Our algorithm only projects the vertices and connects them up with geodesics on the sphere. The straight lines connecting the vertices on the original surface are not explicitly mapped. This could lead to apparently flipped triangles if the triangles were large compared to the radius of curvature of the sphere, or had extremely unequal length sides. However, as the fineness of the triangulation increases, the edge lengths compared to the radius of the sphere decreases, the difference between a geodesic and the curves on the sphere, to which the straight lines should be mapped, will tend to zero.

2.5. Evaluating the Results

Any mapping must be judged by two criteria. The most important is correctness. The second is the quality of the mapping, i.e. the extent and distribution of distortion in either area, shape or both. The primary way we judged the correctness of the mapping was examination of the normals of the mapped triangles. If the mapping is injective, no triangle should be flipped. Any invalid mapping would produce an inversion of some triangle orientations.

Given that distortion in mapping is inevitable, criteria for judging the quality of mapping are more subjective, and determined to some extent by the application. One might want a mapping to be conformal (no angular distortion), equal area, or some balance between distortions in both. For our intended purpose of analyzing molecular surfaces, an important criterion is whether the overall mapping is reasonably uniform in terms of area distribution. To quantify this the sphere was subdivided into 8 octants, and the relative areas occupied by the triangles in each octant on the original surface calculated. The results were also examined visually: Each octant was colored differently. Then the pre-image of each colored vertex on the sphere, that is, the original vertex on the original surface, was colored in the same way. 'Good' quality mapping should preserve the overall orientation of the surface, produce boundaries between the octants that are not too convoluted, and the eight areas should not be too different. For example, it would be undesirable for say 10% of the original surface to cover 90% of the sphere and vice versa, or for the map to be twisted along some axis.

3. Results

Protein/DNA structure files were downloaded from the pdb database (Berman, Westbrook et al. 2000). We selected structures with a resolution of 1.2Å or better simply as a means of obtaining a manageable number from the 20,000+ entries without bias regarding size, shape, function or type (protein vs. nucleic acid). A triangulated solvent-accessible surface was computed for each one, and after eliminating handled surfaces using the Euler equation, 77 genus 0 molecular surfaces were obtained. PDB codes for the structures are given in Appendix 1. Visual examination showed these included a very diverse array of shapes. A sample is shown in Figure 7-10 The number of triangles per surface ranged from about 3000 to 6000. The 77 surfaces were then mapped according to the procedure above. The 77 surfaces contained a total of more than 340,000 triangles, of which non were inverted by the mappings so the mappings are assumed to be correct in these cases. For several structures (1ob5, 1ob7, 1p9g, 1r2m), surfaces with higher resolution were generated, each with more than 20,000 triangles. No incorrect mappings occurred in these cases either. Basic statistics on the surfaces, including angular and area distortions are given in Table 1.

Considering the area first, considerable distortion is expected when mapping surfaces of structures such as proteins where there are 'wide' and narrow' regions (e.g. Fig 5). Indeed the average area distortion is about 100%, with a maximum triangle area distortion of about fourfold. The angular distortion is smaller, relatively speaking. Although the uniform weighting harmonic operator is not conformal, the root mean square change in angle is relatively modest at 25^o. For comparison, the method of Gu et al. (Gu, Wang et al. 2004) is conformal in an analytic sense. In imp lementation, however, it still produces a distribution of angle changes with a width of about 10^o, and a significant number of angles are distorted by more than 10^o. We also examined a more conformal variant of our mapping procedure in which the harmonic operator weights are given by the cotangent formula derived from finite element analysis (Iserles 1996; Floater 2003). The weight a neighbor vertex j contributes at vertex i is given by $\frac{1}{2}(\cot\theta_1+\cot\theta_2)$ where θ_1 and θ_2 are the angles subtended by the two triangles opposite the edge v_iv_j they have in common. This produced a more conformal mapping, with a rms angle distortion of 10.6^o (Table 1). Unfortunately, the weights in the cotangent formula can be negative if $\theta_1+\theta_2>\pi$, a frequent occurrence in these surfaces. This results in invalid mappings as judged by inverted triangles. Although the number of flipped triangles is a very small fraction of the total, a significant number of the surfaces (45%) have this error. Another measure of distortion we used was the partitioning of the octant areas. Equal area mapping would produce all octants with 12.5% of the area. Table 1 gives the root mean squared deviation from this ideal value. This is close to 5%, i.e. most octants have area $12.5\pm5\%$. Figure 4a gives the frequency distribution of octant areas. The modal value is in fact close to 12.5%, and most octants occupy between $^1/_{10}$ and ¼ of the surface. These statistics, along with visual analysis of the surfaces (e.g. Figures 5-8) indicate that globally, there is reasonably uniform partitioning of area. Visual examination of the octant boundaries also shows that the majority are relatively smooth. Excessively convoluted boundaries are rare. In figures 6 and 8 we show projections of widely used protein properties onto the mapped and original surfaces to illustrate an application that motivated this work. In figures 6 and 7 the electrostatic potential is mapped onto the surface. Patterns of positive and negative are clearly preserved and recognizable in the mapped surfaces. Analysis of more complex patterns would require 'landmarks' such as amino acid residue numbering, position along the polypeptide chain, amino acid type, etc. This also is easily mapped onto the surface

either qualitatively (Figure 8), or quantitatively with suitable alphanumeric labeling. Projection of the potentials is designed to facilitate comparison of electrostatic properties among proteins of different size and shape. Coding by position along the polypeptide chain, as in Figure 8, emphasizes the topology and what regions of the polypeptide chain are solvent accessible and what are buried. Note that in the protein structure world topology refers to the overall polypeptide fold, and it is unrelated to its mathematical meaning.

4. Discussion

4.1 General properties of the mapping

The mapping procedure developed here uses a combination of the harmonic mapping operator with the Mercator scaling to map arbitrary closed, genus 0 surfaces on to a sphere. Since all of the 77 surfaces, each containing thousands of surface triangles, were mapped with no inverted triangles our fast and simple mapping method was judged to be correct in principle, and robust in implementation. The scaling factor C, and other mapping options such as triangulation resolution were set automatically by the program, so the method appears to work quite generally for macromolecules with minimal user input. Uniform weighting was chosen since in our implementation it can be proved to give correct (bijective) mappings, as shown in section 2.4. While our mapping is neither equal area nor conformal because we use a uniform weight harmonic operator, examination of the mappings by several criteria show that it gives reasonable results. Angular distortion is not excessive, while judging by octant analysis global partition of area is reasonable, and large scale distortions such as skewing or twisting of the surface are rare. The surface set used to evaluate the method encompasses a variety of initial shapes. For example 1p2g has a long protuberance, while 1ob7 has ovoid regions connected by a narrow waists. In both cases octant areas and octant boundaries are quite balanced. Across the entire 77 molecule set the octant areas are also quite balanced.

4.2 Relationship to previous work

Several methods have been proposed for mapping genus 0 surfaces to a sphere (Brechbuhler, Gerig et al. 1995; Haker, Angenent et al. 2000; Gu and Yau 2002; Gotsman, Gu et al. 2003; Praun and Hoppejournal 2003; Gu, Wang et al. 2004). Our work most closely resembles that of Brechbuhler et al., (BGK) in that both methods choose N, S pole vertices, remove them, and compute spherical coordinates θ and ϕ by harmonic mapping. Our method differs in several important respects.

1) In BGK the harmonic mapping is only a preliminary step. The final mapping is produced by moving vertices on the sphere to minimize some distortion, in their case of area. In principle any preliminary mapping method that was correct would do equally well. Their final mapping, and the properties thereof, are independent of the harmonic mapping step. We assume that their implementation of the first mapping step does not give satisfactory results alone. In contrast, our final mapping is produced by the harmonic mapping in θ/ϕ coordinates.
2) In our method the two types of cut boundaries, the boundary meridian and the circumpolar boundaries are specifically constrained to be either constant ϕ or constant z/θ respectively. In BGK these adopt unconstrained values. The important consequence of this is that our mapping can also be considered as a mapping into a convex (rectangular) region in the plane, followed by a bijective mapping to the sphere. The mapping into the plane is shown to be a correct embedding using the theorem of Floater (Floater 2003). This theorem was probably unavailable to BGK, and we are uncertain whether step one of their method is provably correct using this theorem.
3) Our harmonic mapping step includes a carefully chosen scaling (The Mercator scaling), which is key to producing reasonable mappings, while preserving bijectivity.
4) A less fundamental distinction is that the BGK method was originally implemented for voxel derived data, with equal area quadrilateral elements, and some steps are specific to this kind of surface tessalation. Our method is designed for arbitrary triangulated surfaces. No doubt the BGK could be extended to triangulated surfaces.

4.3 Uniqueness of the mapping

It is interesting to note that with the uniform weighting harmonic operator the Cartesian coordinates of the original surface vertices are only used twice in the algorithm, once in order to determine the outward normal of one triangle and once to find the two most distant vertices of the surface, which become the North and the South Pole. Apart from that, the mapping is entirely based on the connectivity of the surface as a graph. That is, as long as two surface graphs a) have isomorphic vertices and edges and b) the same two vertices are chosen as North and South Pole, the image of the mapping is identical up to mirror symmetry. If in addition, an inward or outward surface normal is provided, the mirror symmetry can be corrected. In other words, once this elementary information is supplied the mapping is independent of the Cartesian coordinates of the original surface vertices. Put another way, generating a different triangulation of the same surface, choosing different pole points, or using different cut boundaries will produce different mappings. The methods of Haker et al and the harmonic mapping step of BGK (Brechbuhler, Gerig et al. 1995; Haker, Angenent et al. 2000) are also non-unique for the same reason that special vertices are chosen to start the mapping off.

Two post-processing steps make non-uniqueness a relatively minor concern in our proposed applications. First, the sphere is put in the best position/orientation with the original surface via a least squares superposition algorithm. This facilitates recognition of corresponding regions by visualization from a common viewpoint. Second, by labeling of landmarks. Unlike typical objects studied in the genus 0 mapping literature such as brain images, or tesselated solid object surfaces, proteins and nucleic acid structures are inherently rich in local information throughout the structure, including at the surface, The information is at various levels of organization. Each atom is uniquely identifiable, its coordinates are known, as are its polarity and hydrophobicity. It is known whether the atom is buried and by how much, or whether it forms part of the surface, and which triangle(s) on the surface. Also atoms can be grouped into amino-acids, polypeptide chains (if there is more than one), enzyme active sites or other functional classes. Providing the overall partition of area is relatively uniform, as in our method, any given mapping can be cogently labeled and analyzed. It need not be unique to be useful.

An alternative to uniform harmonic mapping is to use geometric information about the original surface, such as vertex position, angles, etc. These can be encoded in the harmonic weighting. One conformal-type operator, the cotangent weighting scheme, which uses angle information from the original surface, was examined. It did give a more conformal mapping, reducing the rms angle change from 25^o to 11^o. However it produces negative weights, and inverted triangles. Since even one technically conformal mapping method gives a surprisingly large variation in angle upon implementation (Gu, Wang et al. 2004), and since exact conformality is not essential for our applications, we decided to settle for a provably correct algorithm with moderate angular distortion and other desirable features for our application, rather than try to minimize conformal distortion.

4.4 Dependence on surface tesselation

In any mapping method, the analytical and algorithmic (implementation) aspects must be considered separately. Thus from an analytical point of view, providing the surface is correctly triangulated, single orientable surface of genus 0, and providing suitable N, S pole points are chosen, the mapping scheme described here is valid, guaranteed by the mapping to a convex boundary and the use of Floater's Theorem. In actual implementation, the type, degree and distribution of distortion will depend in part on the initial surface tesselation: The fineness, the connectivity (vertex valence), the range of areas and shapes of the original triangles. Since the triangulation is produced by contouring a function gridded on a uniform cubic grid, this imposes an upper limit on edge length as the longest diagonal of the cubical grid element. In this regard our initial triangulation is more uniform than an arbitrary triangulation. This no doubt contributes to the fact that our method produces well distributed, moderate distortions. With less uniform triangulation, this may not be the case. On the other hand, even with an upper limit on edge length, there is no lower limit, and the triangles in our surfaces have quite a wide distribution of values as indicated in Figure 4b, where a significant percentage of the triangles (indeed the second most common category) are ten-fold smaller than the mean.

4.5 Application to proteins

Regarding application to protein analysis, the lower panels in Figures 6 and 7 give an example of how one surface property, electrostatic potential, appears in the unmapped and mapped surfaces. The electrostatic potential of the molecular surface is indicated by the red/white/blue coloring. With just enough features, the correspondence of the patterns before and after mapping is quite easy to discern. This is aided by ensuring the correct handedness of the mapping and the optimal orientation of the sphere with respect to the original surface. Identifying correspondences with surface property patterns of either lesser or greater complexity becomes more difficult. Here one must use 'landmarks' on the spherical surface. This is illustrated in Figure 8. The tracing of the polypeptide backbone is shown in the left panel, color-coded across the spectrum from blue to red as one moves along the sequence from N-terminus to C-terminus. In the right panel the sphere surface is color-coded the same way. Thus, the color of a surface element indicates the position of the atom(s) in the polypeptide sequence that underlay that surface before mapping. More specifically, at the point at which we generate the surface triangulation, each surface triangle and vertex is labeled with the atom identity, residue type, sequence number, chain designator, and other information contained in the PDB file. Thus, it is possible to label, select, (un)display or analyze every part of the sphere surface according to any of these identifiers. With these features, the immediate application of our mapping algorithm is to simplifying the essential features of a protein surface. This is useful for visualization, for comparing surface property patterns such as electrostatic potential on different protein surfaces, etc.

4.6 Future directions

In addition to the protein and nucleic acid surface analysis outlined here, several obvious developments of our method can be proposed. Other weighting schemes that are more conformal could be explored for applications more sensitive to conformal distortion. It would be desirable to restrict these to convex combination mappings (all weights positive), in order to retain correctness. We note that our method could be easily extended to mapping tori because the intermediate cylinder could be obtained by a suitable circular cut on the toroid, rather than removal of the N/S pole points. The mapped cylinder is then easily transformed into a uniform toroid by transforming z into the circular coordinate, with suitable scaling so that +/-Z map to $0, 2\pi$, thus 'gluing' the cut back together.

5. Acknowledgement

We would like to thank Prof. Dennis DeTurck from the Mathematics Department of the University of Pennsylvania for his many key suggestions, Marcelo Siqueiro and Jean Gallier for helpful discussions. This work was funded by grant MCB02-35440 from the NSF

6. References

Berman, H. M., J. Westbrook, et al. (2000). "The Protein Data Bank." Nucleic Acids Research **28**: 235-242.

Bondi, A. (1968). Molecular Crystals, Liquids and Glasses. New York, John Wiley and Sons.

Brechbuhler, C., G. Gerig, et al. (1995). "Parameterization of closed surfaces for 3-D shape description." Computer Vision and Image Understanding **61**: 154-170.

Duncan, B. and A. J. Olson (1995). "Texture mapping parametric molecular surfaces." J. Mol. Graphics **13**: 258-264.

Eck, M., T. DeRose, et al. (1995). "Multiresolution analysis of arbitrary meshes." Computer Graphics (SIGGRAPH '95 Proc.) **29**(173-182).

Floater, M. (2003). "One-to-one piecewise linear mappings over triangulations." Mathematics of Computation **72**: 685-696.

Floater, M. and K. Hormann (2004). Surface parameterizations: a tutorial and survey. Advances in Multiresolution for Geometric Modelling. N.A. Dodgson, M.S. Floater and M. A. Sabin. Heidelberg, Springer-Verlag**:** 157-186.

Giblin, P. J. (1981). Graphs, surfaces and homology. London, Chapman Hall.

Gotsman, C., X. Gu, et al. (2003). "Fundamentals of spherical parameterization for 3D meshes." ACM Transactions on Graphics **22**: 358--363.

Gu, X., Y. Wang, et al. (2004). "Genus zero surface conformal mapping and its application to brain surface mapping." IEEE Transactions on Medical Imaging **23**: 949-58.

Gu, X. and S.-T. Yau (2002). "Computing conformal structures of surfaces." Communications in Information Systems **2**: 121-146.

Haker, S., S. Angenent, et al. (2000). "Conformal surface parameterization for texture mapping." IEEE Transactions on Computer Graphics **6**: 181-189.

Iserles, A. (1996). A first course in the analysis of differential equations. Cambridge, Cambridge University Press.

Kearsley, S. (1989). "On the orthogonal transformation used for strcuture comparison." Acta Cryst. **A45**: 208-210.

Lee, B. and F. M. Richards (1971). "The interpretation of protein structures: estimation of static accessibility." J. Mol. Biol. **55**: 379-400.

Nicholls, A., K. A. Sharp, et al. (1991). "Protein folding and association: Insights from the interfacial and thermodynamic properties of hydrocarbons." Proteins **11**: 281-296.

Praun, E. and H. Hoppejournal (2003). "Spherical parametrization and remeshing." ACM Transactions on Graphics **22**: 340--349.

Tutte, W. T. (1963). "How to draw a graph." Proc. London Math. Soc. **13**: 743-68.

Table 1

Mapping Protocol	# inverted triangles (% surfaces correct)	Triangle area distortion, Mean (Max)	Triangle angle distortion Mean (Max)	RMS deviation in octant area from ideal of 12.5%
Uniform Weighting,	0 (100%)	104% (470%)	25.0^{o} (29.3^{o})	4.9%
Cotangent Weighting,	204 (55%)	86% (446%)	10.6^{o} (19.6^{o})	4.3%

Figure Legends

Figure 1. Simplified 2D illustration of 3D triangulation algorithm. i) The atoms (indicated by shaded van der waals (vdw) radius spheres, are mapped onto a lattice. Points inside the probe-extended radius (larger unshaded circles) are marked as interior (=1). ii) A solvent probe (white sphere) is placed on each boundary interior lattice points, and all points inside the probe are turned 'out' (=0). iii) The surviving points with a value of 1 approximate the molecular volume. Note that points outside the atoms but inaccessible to the solvent probe are also inside this volume. iv) Contouring the lattice function (1/0) at a contour level of 0.5 produces the molecular surface.

Figure 2. a). Spherical coordinate system used in mapping. b). Assigment of z (Steps 2, 3). c) Assignment of θ from z (Step 6).

Figure 3 a). Assignment of ϕ. b) Equivalent mapping of graph to z-ϕ plane. Boundary Meridian (black vertices), BM-neighboring sets A and B (violet, blue vertices) c). Subdivision of two neighboring triangles increases the order of v_i by one, and removes the direct connection between v_j and v_k.

Figure 4. a) Frequency distribution of octant areas across the 77 mapped structures. b) Frequency distribution of triangle areas in a typical original surface, expressed as a ratio compared to the mean triangle area for that surface.

Figure 5. Mapping of the surface of structure 1ob7. Top. a), b) Two 180^o rotated views colored by the octant of the sphere to which it is mapped. c). close up of original surface triangulation in the region of the narrow waist in the surface (left in panel a). d) triangulation of mapped surface, with the waist region oriented toward the viewer.

Figure 6. Mapping of the surface of protein 1r2m. a), b) Two 180^o rotated views of the molecular surface colored by the octant of the sphere to which it is mapped. The original surface (c) and sphere mapped surface (d) colored according to electrostatic potential, blue: >2kT/e, red: < -2kT/e.

Figure 7. Mapping of the surface of protein 1p2g. a), b) two 180^o rotated views of the molecular surface colored by the octant of the sphere to which it is mapped. The original surface (c) and sphere mapped surface (d) colored according to electrostatic potential, blue: >2kT/e, red: < -2kT/e.

Figure 8. Mapping of surface of protein 1lib. a) Ribbon representation of the protein. b) Sphere mapped surface. The ribbon is color coded according to position along the polypeptide chain, from N-terminus (Blue) to C-terminus (Red). The surface is color coding accordingly, based on the sequence position of the residues which form each part of the original molecular surface.

Appendix 1. List of macromolecular structures used.

1a6m, 1alz, 1brf, 1byz, 1c7k, 1ceq, 1cex, 1d8g, 1d48, 1dcg, 1dj6, 1ea7, 1ejg, 1em0, 1en3, 1en8, 1en9, 1ene, 1enn, 1et1, 1etl, 1etm, 1etn, 1exr, 1f9y, 1f94, 1g4i, 1g6x, 1g66, 1ga6, 1gkm, 1gqv, 1hj9, 1hje, 1i0t, 1i1w, 1ic6, 1ick, 1itt, 1ix9, 1ixh, 1j8g, 1k5c, 1kth, 1lug, 1mso, 1o56, 1oai, 1ob4, 1ob6, 1ob7, 1ok0, 1p9g, 1qyl, 1r2m, 1r6j, 1rb9, 1rtq, 1sfd, 1tqg, 1ucs, 1ug6, 1uzv, 1v0l, 1w0n, 2bf9, 2dcg, 2erl, 3al1, 3lzt, 7a3h, 131d, 292d, 293d, 336d, 485d.

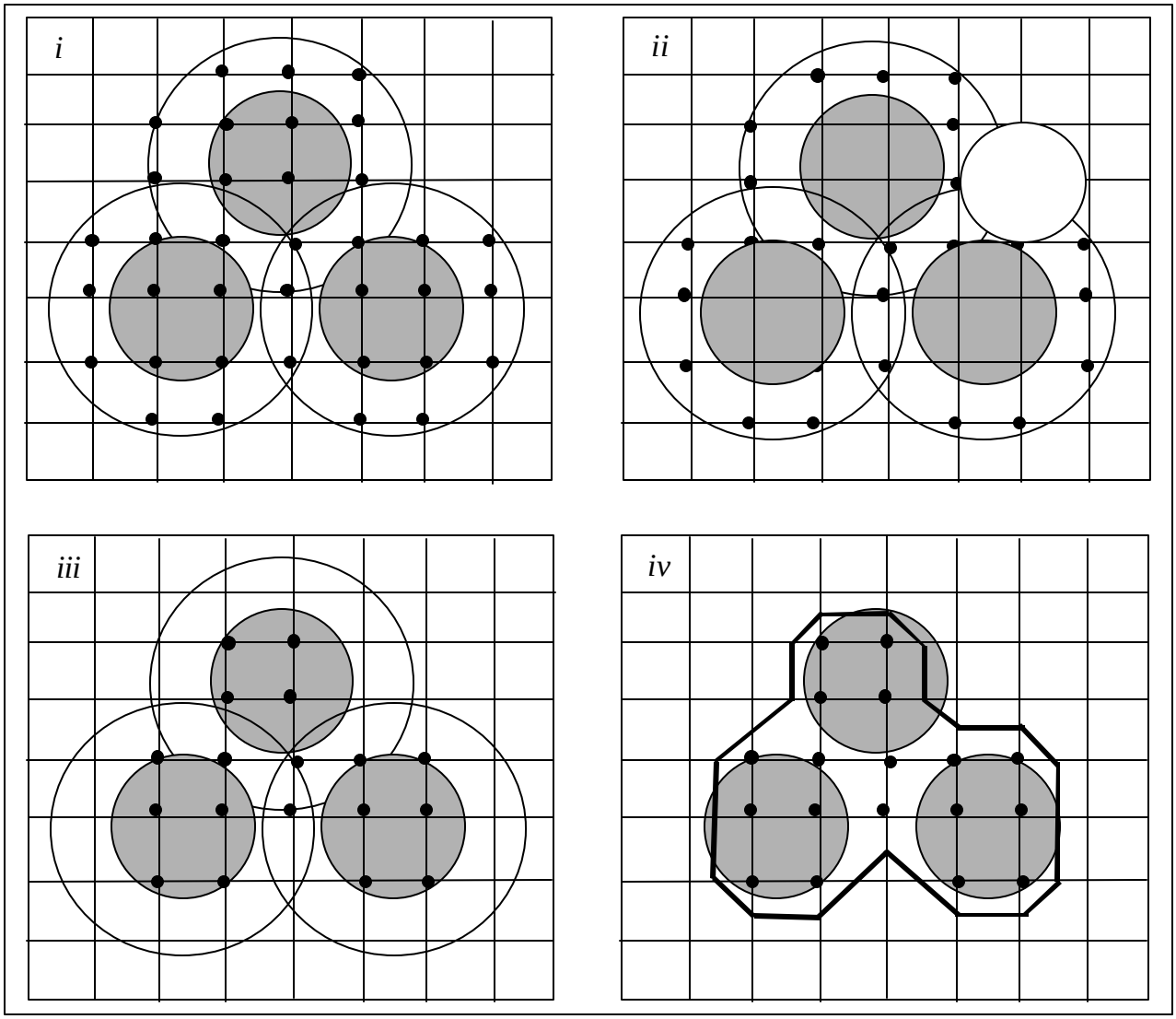


Figure 1

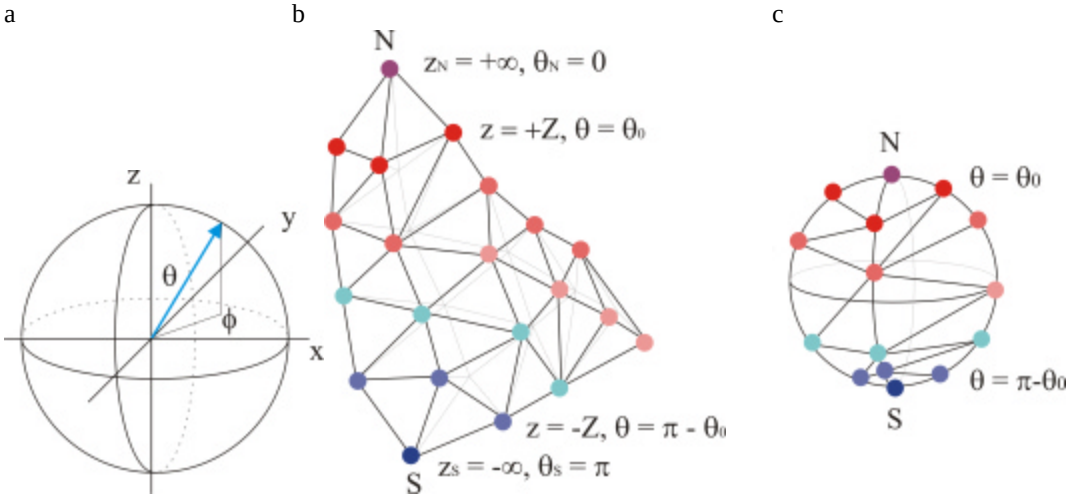

a
b
c
z
y
x
θ
ϕ
N
$z_N = +\infty, \theta_N = 0$
$z = +Z, \theta = \theta_0$
$z = -Z, \theta = \pi - \theta_0$
$z_S = -\infty, \theta_S = \pi$
S
N
$\theta = \theta_0$
$\theta = \pi - \theta_0$
S

Figure 2

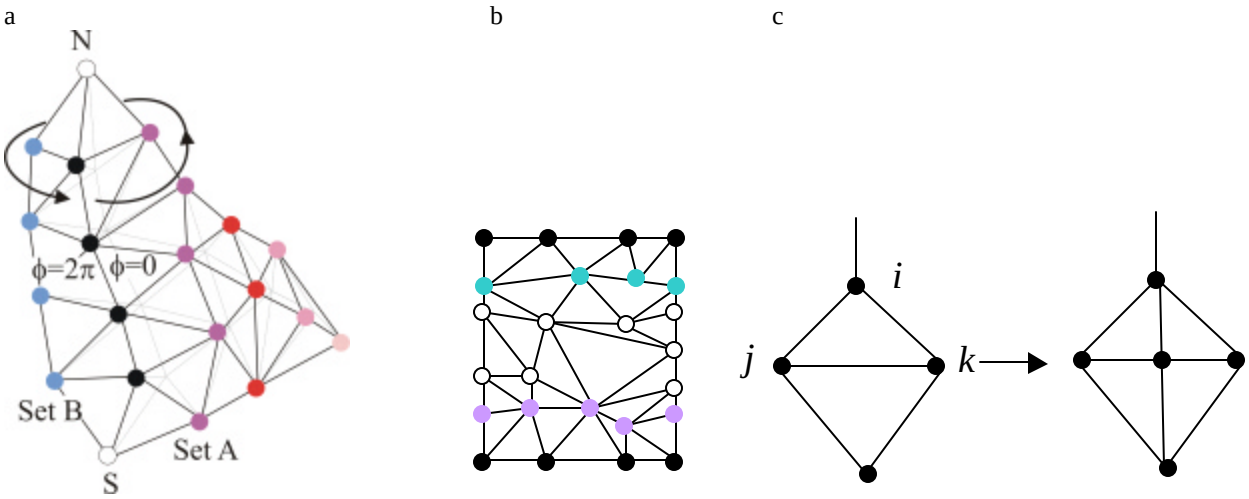
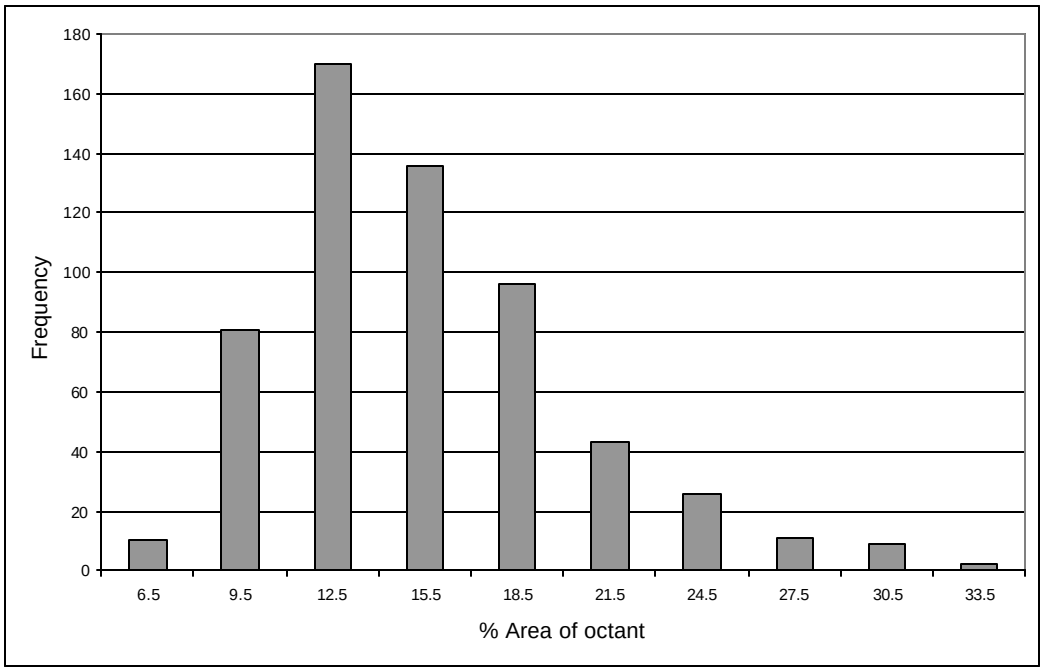


Figure 3.

a


180
160
140
120
100
80
60
40
20
0
Frequency
6.5
9.5
12.5
15.5
18.5
21.5
24.5
27.5
30.5
33.5
% Area of octant

b

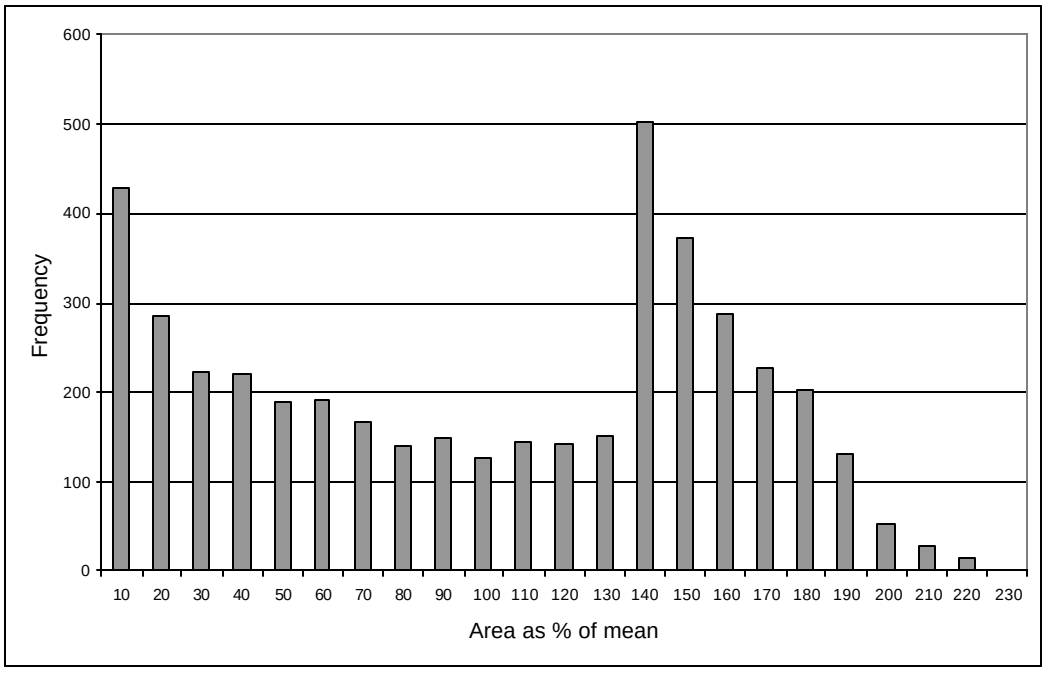

600
500
400
300
200
100
0
Frequency
10
20
30
40
50
60
70
80
90
100
110
120
130
140
150
160
170
180
190
200
210
220
230
Area as % of mean

Figure 4

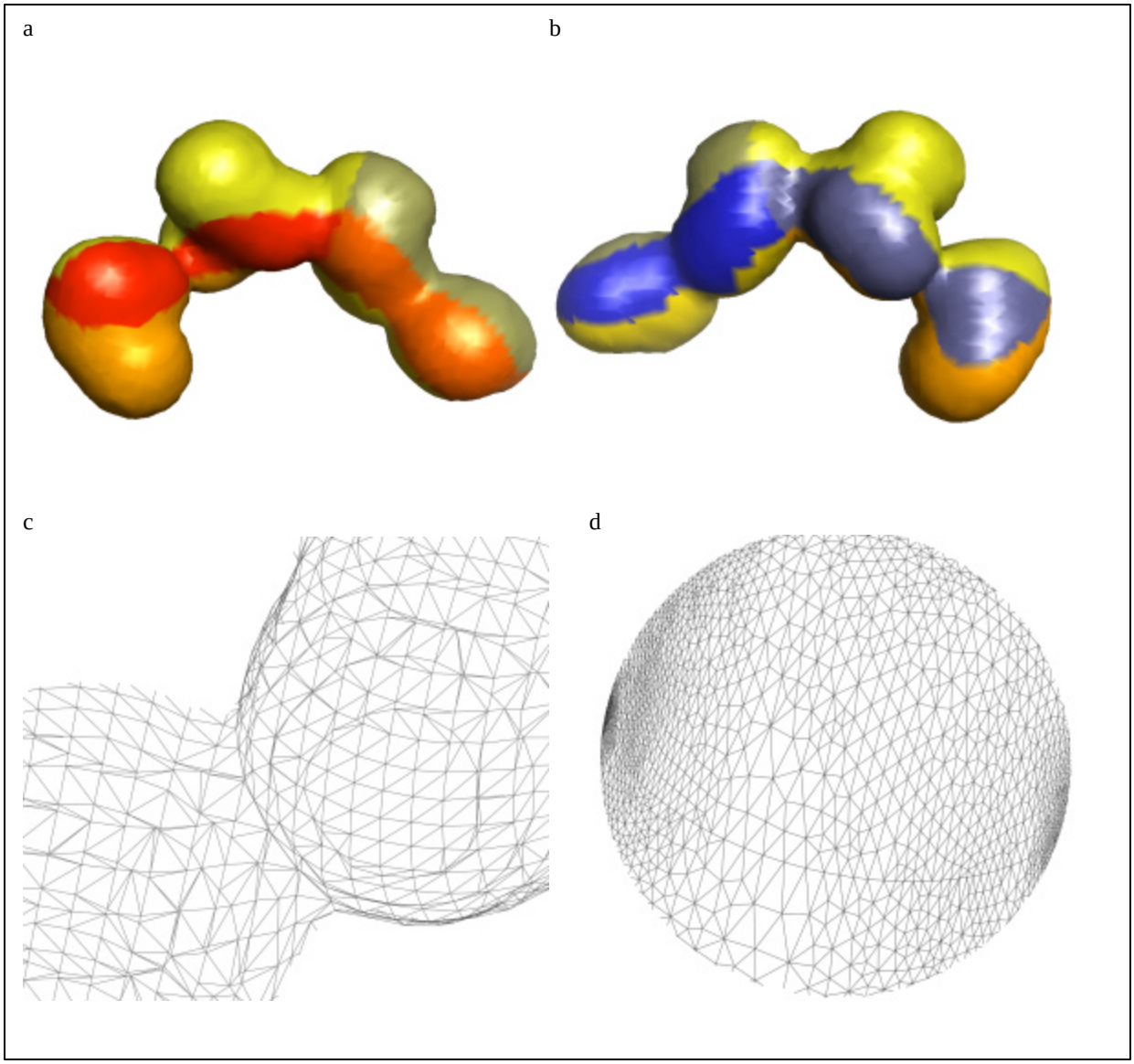


Figure 5.

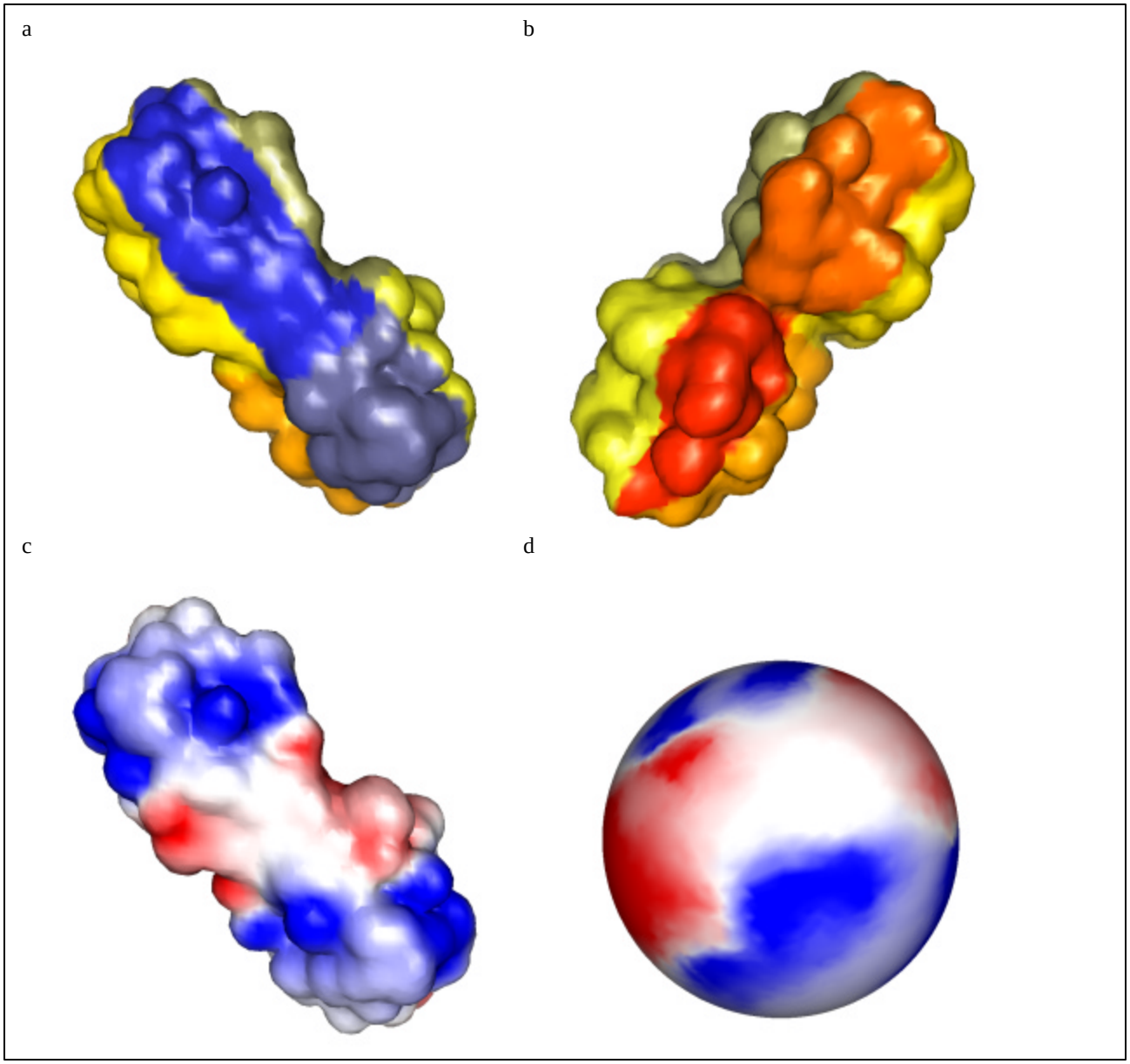


Figure 6.

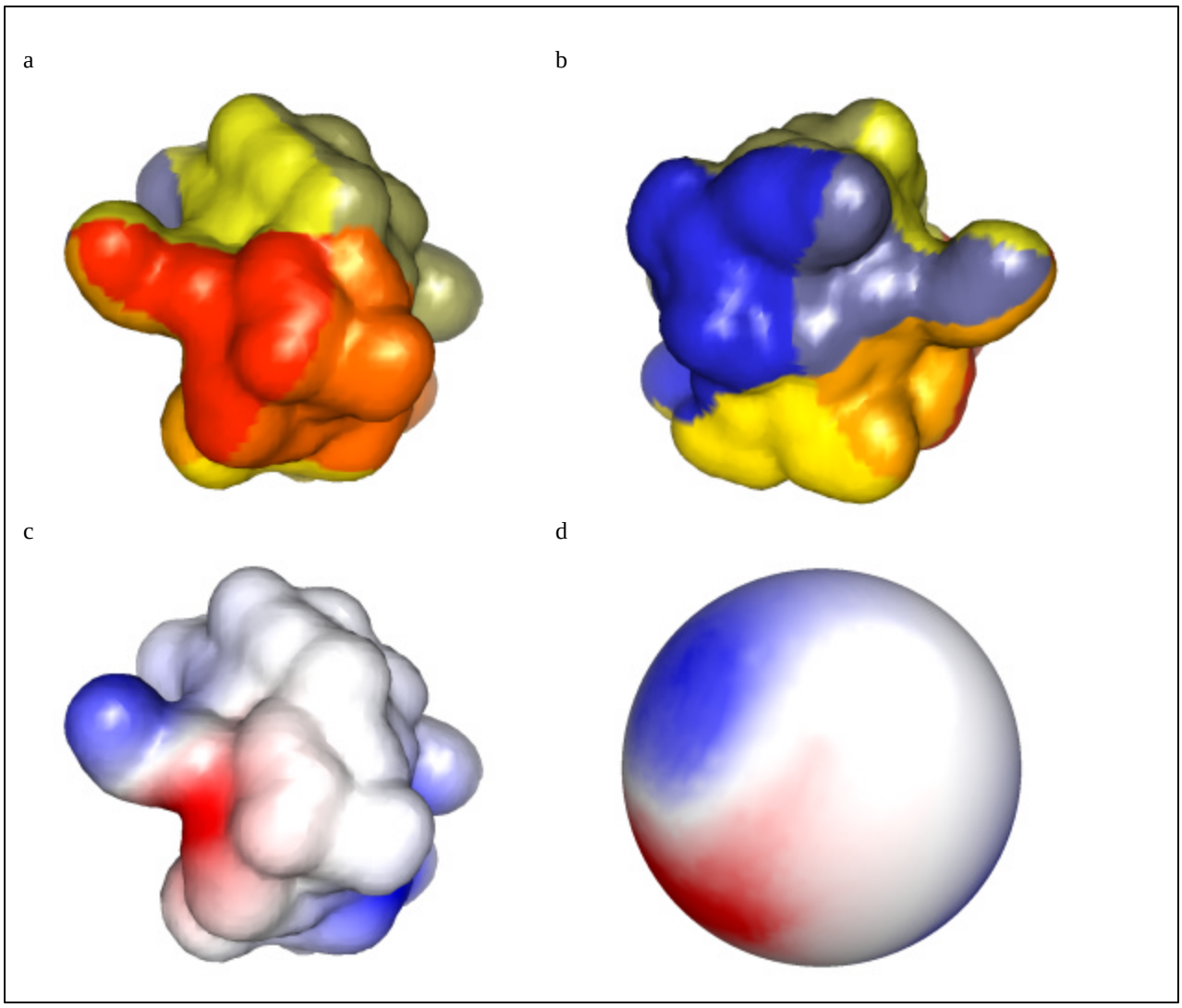


Figure 7.

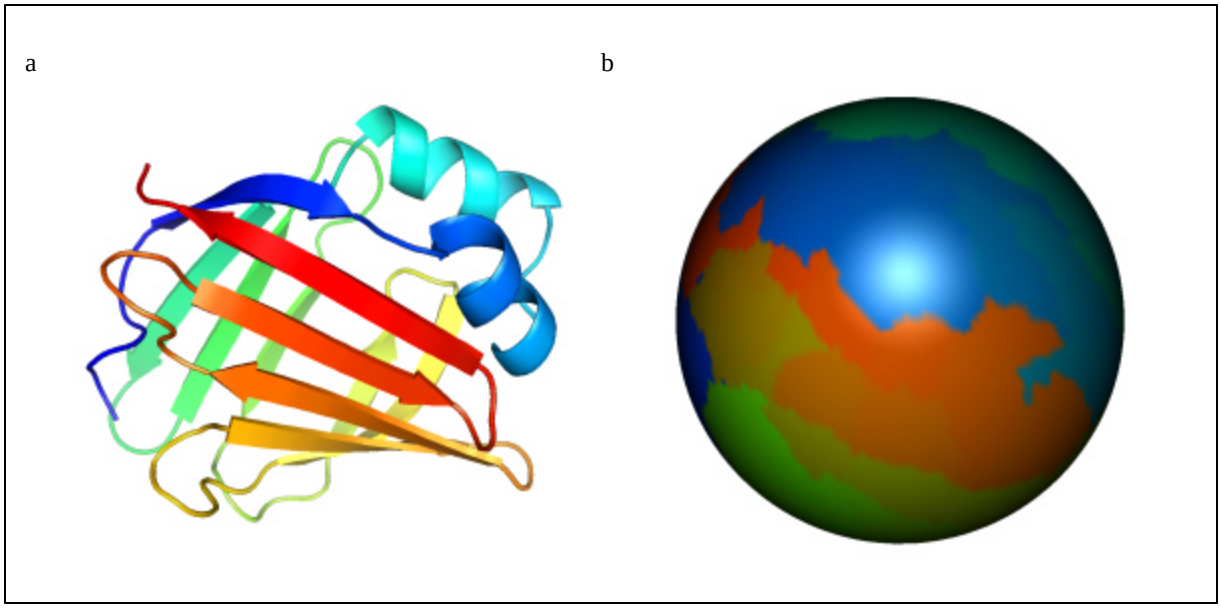


Figure 8.