BUTTERWORTH HEINEMANN

Approximation and visualization of large-scale motion of protein surfaces

Bruce S. Duncan and Arthur J. Olson

The Scripps Research Institute, La Jolla, California 92037

We present a method for the approximation and real-time visualization of large-scale motion of protein surfaces. A molecular surface is represented by an expansion of spherical harmonic functions, and the motion of protein atoms around their equilibrium positions is computed by normal mode analysis. The motion of the surface is approximated by projecting the normal mode vectors of the solvent-accessible atoms to the spherical harmonic representation of the molecular surface. These surface motion vectors are represented by a separate spherical harmonic expansion. Representing the surface geometry and the surface motion vectors by spherical harmonic expansions allows variable-resolution analysis and real-time display of the large-scale surface motion. This technique has been applied to interactive visualization, interactive surface manipulation, and animation.

Keywords: Normal mode analysis, protein surface motion, spherical harmonics.

INTRODUCTION

Computer simulations provide valuable insights into the structure and dynamics of protein molecules. Molecular dynamics simulations provide the most detailed description of the atomic motion of macromolecules.[1–4] One limitation of molecular dynamics, however, is that simulation of large systems is difficult because of the required computational resources. Normal mode analysis determines the small-amplitude displacements of atoms around their equilibrium positions[5–15] and is becoming increasingly popular because of algorithm improvements[16] and increased computer power.

Molecular and solvent accessible surfaces are important for visualization and for the analysis and prediction of protein–protein and protein–ligand interactions.[17–25] Also, it is more efficient to render triangulated molecular surfaces than CPK representations on workstations that lack specialized hardware for sphere rendering. We would like a visualization method that combines the utility of dynamic simulations with efficient triangle rendering. When analyzing the motion of proteins, however, it is computationally expensive to compute an accurate molecular surface for each conformation of a normal mode simulation and display the results in real time. In this article we describe a method to represent the small-amplitude perturbations of protein surfaces computed from normal mode analysis using expansions of spherical harmonic functions. This method allows the real-time display of the approximated surfaces and interactive control of the surface triangulation and surface resolution.

Spherical harmonic expansions of static molecular surfaces are used to represent low-resolution features and to compute shape properties.[26–30] They have also been used in the analysis and visualization of electrostatic interactions,[31] and for the representation of shape constraints during molecular mechanics.[32] We have applied these methods to display surface properties using texture mapping.[33] Here we describe the use of spherical harmonic techniques to represent dynamic information without explicitly recomputing the exact molecular surface. This information is useful in interactive modeling, animation, and the analysis of protein–protein interactions.

METHODS

Spherical harmonic representations of molecular surfaces

A detailed description of our procedure to approximate molecular surfaces by spherical harmonic expansions is presented in Ref. 26 and is summarized below. The molecular surface to be approximated is computed using programs such as the Molecular Surface Package (MSP) by Connolly[19,20,22] or the MSMS program by Sanner et al.[23–25] These programs compute an analytic description of the molecular surface and generate a triangulated surface from this analytic description. The triangulated surface is the input to our approximation procedure. Before spherical harmonic expansion coefficients are computed, it is necessary that there be a one-to-one mapping from the molecular surface

Color plates for this article are on pp. 238–240.

Address reprint requests to Dr. Olson at The Scripps Research Institute, 10666 North Torrey Pines Road, La Jolla, California 92037.
Received 3 February 1995; revised 15 May 1995; accepted 18 May 1995

655 Avenue of the Americas, New York, NY 10010

0263-7855/95/$10.00
SSDI 0263-7855(95)00039-9

to a unit sphere. This mapping is done by smoothly deforming the surface into a spherical shape while retaining the topology and triangulation of the original molecular surface. After the topological mapping, each point on the molecular surface has a unique spherical coordinate (θ,ϕ) on a unit sphere. Conversely, each point on the unit sphere can be assigned a Cartesian coordinate on the molecular surface. Thus, there is a one-to-one mapping from the original molecular surface to the unit sphere representation.

Spherical harmonic functions form an orthonormal basis set on a unit sphere and are evaluated at the spherical coordinate (θ,ϕ). Any sufficiently smooth scalar function defined on a unit sphere, $F(\theta,\phi)$, can be expressed as a spherical harmonic expansion as follows:

$$F(\theta,\phi) = \sum_{l=0}^{\infty} \sum_{m=-l}^{l} c_{lm} Y_l^m(\theta,\phi) \quad (1)$$

where c_{lm} are expansion coefficients, and $Y_l^m(\theta,\phi)$ are spherical harmonics. The integers l and m denote specific harmonic functions. The value of l defines the order of the harmonic, which is greater or equal to zero. There are $2l + 1$ different harmonic functions for each order l and these functions are denoted by the integer m, which varies from $-l$ to l.[30] Spherical harmonics are complex-valued functions, but in this analysis real-valued functions are obtained by forming appropriate linear combinations.

Spherical harmonic expansions represent functions defined on a unit sphere. The function that we want to approximate is the Cartesian coordinates of the molecular surface given by the independent variables (θ,ϕ). Because spherical harmonic functions represent only scalar functions, three separate expansions are used, one for each of the three Cartesian coordinates. Expansion coefficients are computed by evaluating the surface integral that represents the inner product of the function to be approximated and the individual spherical harmonic functions:

$$c_{lm} = \int_0^{\pi} \int_0^{2\pi} F(\theta,\phi) Y_l^m(\theta,\phi) \sin\theta \, d\theta d\phi \quad (2)$$

Three sets of spherical harmonic expansion coefficients, $(c_{lm}^x, c_{lm}^y, c_{lm}^z)$, are computed. From these coefficients, the Cartesian coordinates of the surface are reconstructed to an order l_{max} as follows:

$$X(\theta,\phi) = \sum_{l=0}^{l_{max}} \sum_{m=-l}^{l} c_{lm}^x Y_{lm}^m(\theta,\phi) \quad (3)$$

$$Y(\theta,\phi) = \sum_{l=0}^{l_{max}} \sum_{m=-l}^{l} c_{lm}^y Y_l^m(\theta,\phi) \quad (4)$$

$$Z(\theta,\phi) = \sum_{l=0}^{l_{max}} \sum_{m=-l}^{l} c_{lm}^z Y_l^m(\theta,\phi) \quad (5)$$

where $X(\theta,\phi)$, $Y(\theta,\phi)$, and $Z(\theta,\phi)$ are the Cartesian coordinates of the molecular surface.

We compute surface shape properties (normal vector, mean curvature, Gaussian curvature, and principal curvature directions) using the expansion coefficients and explicit formulas for the derivatives of the spherical harmonic functions as described in Ref. 26.

This technique, however, has the following limitations. The spherical harmonic method is an approximation technique; it can be difficult to compute a description of a surface geometry that exactly matches some input surfaces because of limited computational resources. This method may require an impractically high l_{max} to accurately represent large, solvent-accessible side chains or invaginations that are deep and narrow.

Because we use a basis set defined on a unit sphere, we can represent only surfaces that are topologically equivalent to a sphere, that is, surfaces with a genus of zero.[34] This method cannot represent toroidal surfaces such as the human immunodeficiency virus (HIV) protease dimer. This method works best when the input surface is approximately spherical and does not require extensive topological mapping. However, we have used this procedure to represent noncompact surfaces such as tRNA and short segments of double-stranded B-DNA. Also, the topological mapping procedure introduces metric distortions because during the mapping some triangles are expanded while others are contracted. This causes nonuniform scaling of the surface representation.

In spite of these limitations, this technique has enough accuracy to model the important geometric features required for the prediction of protein–protein interactions and for many animation and interactive modeling tasks.

Normal mode analysis

Normal mode analysis computes the small-amplitude harmonic motion of a dynamic system about an equilibrium position. The lowest frequency modes are of particular interest because they have the largest amplitudes and the largest spatial extent. A system of N particles has $3N - 6$ normal modes. Each normal mode has a characteristic frequency, ω, and each particle of the system has a direction vector, **k**, for each normal mode. For each mode, the direction vector of a particle is the direction of the harmonic motion. In general, the magnitude and direction of this vector are different for each particle of the system. Normal modes are computed from the eigenvalues and eigenvectors of the mass-weighted second derivative matrix (Hessian) of the energy function evaluated at the equilibrium configuration. Detailed descriptions of normal mode techniques can be found in Refs. 3–5 and 16. Note that normal modes are defined only for an equilibrium configuration; the system must be at a minimum of its energy function.

Normal mode calculations have several limitations. The calculation requires that the molecular conformation is at a minimum of its energy function. Most structures determined from X-ray crystallography or nuclear magnetic resonance (NMR) spectroscopy are not at energy minima and must be minimized before normal modes are computed. Thus the original conformation may not be identical to the conformation used to compute the normal modes. Also, normal mode analysis is valid only for small displacements around the equilibrium configuration; this method cannot

compute large-amplitude displacements or anharmonic motion.

After energy minimization, we compute normal modes with the NMODE module of AMBER.[2] Once the normal modes are computed, the position of each atom in a single normal mode is

$$\mathbf{X}_i(t) = \mathbf{X}_i^\circ + \sqrt{kT}\sin(\omega t + \phi)\mathbf{k}_i \quad (6)$$

where $\mathbf{X}_i(t)$ is the position of atom i, t, $\mathbf{X}_i^\circ$ is the equilibrium position of atom i, k is the Boltzman constant, T is the absolute temperature, ω is the mode frequency, ϕ is the mode phase, and $\mathbf{k}_i$ is the normal mode vector of atom i.

Complex motion is obtained by combining several normal modes. The motion from the combination of N modes is

$$\mathbf{X}_i(t) = \mathbf{X}_i^\circ + \sqrt{kT}\sum_{n=1}^{N}\sin(\omega_n t + \phi_n)\mathbf{k}_{in} \quad (7)$$

where ω_n is the frequency of mode n, ϕ_n is the phase of mode n, N is the maximum number of modes under consideration, and $\mathbf{k}_{in}$ is the normal mode vector of point i for mode n.

Representing surface motion

Our strategy is to approximate the motion of the molecular surface by computing a direction vector, **k**, for each point of the surface. The position of the surface at any time t is computed using Eq. (7) with the atom data replaced by surface data. The assignment of surface motion vectors can be done in several ways. In our procedure, we assign to each surface point the motion vector of the atom closest to the surface point. An alternative procedure is to compute surface motion vectors based on a Gaussian-weighted distribution of the nearby atoms. By this procedure, the surface direction vectors incorporate the effect of nearby atoms that are not on the molecular surface.

One difficulty in representing motion by this procedure is that motion vectors for surface points that are close to each other on the surface may point in different directions. Updating the surface using Eq. (7) may result in unacceptable deformations of the surface that do not reveal the underlying regularities of the motion. This problem is partially reduced because we are interested only in the low-frequency normal modes, and these modes have the greatest spatial extent. Thus the normal mode vectors for the low-frequency modes tend to vary smoothly within the molecule and at the surface.

We can further alleviate this problem by representing the surface motion vectors as a separate spherical harmonic expansion. In this procedure, we compute a spherical harmonic representation of the surface motion vectors and reconstruct the motion vectors at a slightly lower spatial resolution. This reveals the low-resolution (large-scale) motion of the protein surface.

Expansion coefficients of the surface motion vectors are computed by a straightforward extension of the method of Duncan and Olson.[26] Instead of representing the Cartesian coordinates, (x,y,z), of the surface for each (θ,ϕ) point on the unit sphere, we represent the components of the motion vectors, (k_x,k_y,k_z). Surface motion vectors for each mode under consideration are represented by a separate spherical harmonic expansion. The motion vectors at a particular length scale (given by l_{max}, the maximum order of the spherical harmonic expansion) are reconstructed from these expansion coefficients. Note that the same mapping of the molecular surface to the spherical coordinates (θ,ϕ) is used for the evaluation of the surface geometry and the surface motion vectors. This relation defines the mapping of all surface properties (Cartesian coordinate, normal mode vector, hydrophobicity, electrostatic potential, etc.) for the spherical harmonic representations.

We control the scale of the representation by altering the number of terms in the expansion. A feature of this representation is that the scale of the surface geometry, surface properties, and surface motion can be controlled independently. We can apply a low-resolution reconstruction of the normal modes to a high-resolution representation of the protein surface.

To view the surface motion, we precompute the surface motion vectors for each mode under consideration. The time-dependent modulation of each surface point is computed using an extension of Eq. (7) in which $\mathbf{X}_i(t)$ is the position of surface point i at time t, and $\mathbf{k}_{in}$ is the spherical harmonic reconstruction of the normal mode vector of surface point i for mode n.

An important property of atomic normal modes is that they form an orthonormal basis set that can be used to represent conformational changes. An arbitrary protein conformation can be projected onto the normal mode coordinates.[9] Conversely, we can use normal mode coordinates to model a deformation of the structure. A deformation with a large spatial extent can be described using only the low-frequency modes. This is important for representing deformations using a small number of independent variables. This principle has been used when applying normal modes to the refinement of X-ray crystal structures.[13–15] We are currently investigating the use of this principle in our protein–protein docking program for the representation of large-scale (small-amplitude) deformations that may occur during the formation of protein complexes.[35] By representing potential deformations as a combination of low-frequency modes, we reduce the dimensionality of the search for configurations with optimum geometric complementarity. After projecting the atomic mode vectors to the surface, the modes are no longer orthogonal. However, they can still be used as independent variables to represent surface deformations. For this application, time is no longer the independent variable. Instead, the independent variables are the magnitudes of the contribution of each mode to the total deformation. In this representation, the deformation of surface point i is given by

$$\mathbf{X}_i(W_1,W_2,\ldots,W_N) = \mathbf{X}_i^0 + \sum_{n=1}^{N} W_n\mathbf{k}_{in} \quad (8)$$

where W_n gives the contribution of mode n to the total deformation. The geometric fit between two surfaces can be optimized by selecting appropriate values for these weights. The number of modes considered, N, is about 20 for each

surface. For the two surfaces, the total number of weights, $2N$, is much less than the number of Cartesian or torsional degrees of freedom for an atomic representation of the complex. Because the low-frequency modes have the largest amplitudes, a plausible strategy is to optimize the low-frequency modes first.

Another application of these techniques is the interactive deformation of spherical harmonic surfaces. Other investigations have also described techniques of interactive surface manipulation.[36] We can manually deform the surface by choosing values for the N weights, W, in Eq. (8). By attaching these parameters to dials, we can visualize the deformations that can occur without a large energy penalty. Conformations obtained from manual selection of the weights can provide initial structures for energy minimization or molecular dynamics refinement. Although we are developing these procedures for our protein–protein docking program, this principle is also applicable to protein–ligand docking. The procedure of manually selecting the weights of the normal modes could be done using an atomic representation, but the surface representation has the advantage of eliminating atoms that do not have direct interprotein interactions. There are also rendering advantages as discussed below (Visualization of Surface Motion). This method, however, can assess only the global conformational changes; local side-chain motion must be analyzed using other methods.

Shape analysis of dynamic surfaces

We have previously described techniques for computing the shape properties of static protein surfaces at multiple resolutions using spherical harmonic expansions.[26] For computing dynamic shape properties, we want to avoid recomputing the expansion coefficients for each instance of the motion. The following procedure computes the expansion coefficients corresponding to the perturbed surface using the coefficients for the static surface, the coefficients for the surface motion vectors, and the mode frequencies and phases. This procedure, however, cannot be used for real-time shape analysis.

Consider a single point on the equilibrium surface defined by expansion coefficients $(c^x_{lm}, c^y_{lm}, c^z_{lm})$, and referenced by spherical coordinates (θ,ϕ). The equilibrium x coordinate of this surface point, x^0, is given by

$$x^0 = \sum_{l=0}^{l_{max}} \sum_{m=-l}^{l} c^x_{lm} Y^m_l(\theta,\phi) \tag{9}$$

The perturbed x coordinate, $x(t)$, of this surface point in normal mode n is given by

$$x(t) = x^0 + \sqrt{kT}\,\sin(\omega_n t + \phi_n) k^x_n \tag{10}$$

where $x(t)$ is the coordinate at time t, k^x_n is the x component of the surface motion vector for normal mode n, ω_n is its frequency, and ϕ_n is its phase.

At this point on the surface, the x component of the surface motion vector of mode n, k^x_n, also has a spherical harmonic representation given by

$$k^x_n = \sum_{l=0}^{l_{max}} \sum_{m=-l}^{l} c^{x_n}_{lm} Y^m_l(\theta,\phi) \tag{11}$$

Equations (9), (10), and (11) are combined to give the following relation between the perturbed x coordinate, $x(t)$, and the spherical harmonic expansion coefficients for the x component of surface geometry and the x component of the surface motion vector:

$$x(t) = \sum_{l=0}^{l_{max}} \sum_{m=-l}^{l} c^x_{lm} Y^m_l(\theta,\phi) + \sqrt{kT}\,\sin(\omega_n t + \phi_n) \sum_{l=0}^{l_{max}} \sum_{m=-l}^{l} c^{x_n}_{lm} Y^m_l(\theta,\phi) \tag{12}$$

This relation is simplified to yield

$$x(t) = \sum_{l=0}^{l_{max}} \sum_{m=-l}^{l} [c^x_{lm} + \sqrt{kT}\,\sin(\omega_n t + \phi_n) c^{x_n}_{lm}] Y^m_l(\theta,\phi) \tag{13}$$

Thus, the perturbed x coordinate, $x(t)$, has a spherical harmonic expansion with time-dependent coefficients given by

$$c^x_{lm}(t) = c^x_{lm} + \sqrt{kT}\,\sin(\omega_n t + \phi_n) c^{x_n}_{lm} \tag{14}$$

This relation is extended to yield the relation for the spherical harmonic expansion coefficients resulting from a combination of N normal modes as follows:

$$c^x_{lm}(t) = c^x_{lm} + \sqrt{kT} \sum_{n=1}^{N} \sin(\omega_n t + \phi_n) c^{x_n}_{lm} \tag{15}$$

where n specifies different normal modes. Coefficients for the y and z coordinates are determined similarly.

Thus to approximate geometric properties for the perturbed surface at time t, it is not necessary to recompute the expansion coefficients for the new geometry. The coefficients for the new geometry are computed directly from the coefficients for the equilibrium surface and the coefficients for the surface motion vectors that describe the displacement of the surface from its equilibrium position.

Note that it is easy to combine an order p representation of the surface geometry with an order q representation of the surface motion by setting all surface geometry coefficients of order $l > p$ to zero and all surface motion coefficients with $l > q$ to zero. By this method, we can evaluate a high- or low-resolution representation of the surface geometry under the influence of only the low-frequency motion.

Although shape properties can be computed for any particular time t, it is computationally expensive to recompute

the shape properties during the real-time display of the surface motion. Computing shape properties requires evaluating not only the sum given by Eq. (3) but also sums for the first and second derivatives of these relations. This technique, however, does allow the generation of the expansion coefficients for a conformation at any particular time and is useful for producing animations of the shape properties.

Visualization of surface motion

There are techniques to animate atom positions during normal mode simulations. For example, the FLEX program[37] displays atomic structures (represented by lines, spheres, or ribbons) under the influence of normal modes. The user can select interactively the display parameters and the normal modes to be considered. Animation of the covalent structure of the protein is an effective and computationally efficient visualization technique. Animating a CPK representation of many atoms, however, is difficult without high-performance graphics hardware. Also, it is computationally impractical to recompute a molecular surface at each instance of a normal mode simulation for real-time display. Consequently, our visualization procedures for molecular surface motion use the spherical harmonic techniques so that real-time performance is achieved. Spherical harmonic surfaces are usually rendered as Gouraud shaded triangles, and line rendering significantly improves the speed of the animations.

One advantage of analytical representations of surfaces is that the polygon mesh of the surface can be specified independent of the surface geometry. We define this mesh so that it has a regular structure and well-defined symmetry properties. Recall that the spherical harmonic representation defines a relation between the unit sphere and the molecular surface. We can arbitrarily choose which points on the unit sphere are used to generate points on the spherical harmonic surface. One such mesh structure uses the spherical coordinates (θ,ϕ) corresponding to the vertices of a regular icosahedron to define which spherical coordinates (θ,ϕ) will be used to generate the Cartesian coordinates of the spherical harmonic surface. This produces a low-resolution triangle mesh composed of 12 vertices and 20 triangles. Smoother surfaces are obtained by subdividing each triangle into four smaller triangles and computing surface coordinates at the three new (θ,ϕ) points on the unit sphere. This process is repeated until the desired level of smoothness is obtained.

Although we typically use triangulated surfaces derived from a regular icosahedron, our protein–protein docking program uses the dual of the triangle mesh. The docking program approximates the interaction between the triangle centers of two protein surfaces, and the dual mesh is defined by the positions and connectivity of these triangle centers. This mesh is composed of hexagons and pentagons derived from a regular dodecahedron. In addition to these meshes, we also use meshes generated by dividing the faces of a cube for rendering surfaces made with square polygons. One limitation, however, is that squares are usually more distorted than the triangles described above because a cube has lower symmetry than an icosahedron. The subdivision mesh sizes are summarized in Table 1.

These mesh representations have some interesting features. Each protein surface, independent of its size or shape, can be constructed from one of these "standard" meshes. If one recomputes the molecular surface with MSMS during a normal mode simulation, each surface would be composed of a different number of vertices and triangles. With the spherical harmonic representation, however, the triangulation mesh is time independent. Only the coordinates of the surface points move during the simulation; the topological relation between points is constant.

The regularity of the triangle mesh allows us to implement a filtering procedure to interactively control the resolution of the surface triangulation independent of the resolution of the surface itself. By this technique, surfaces may be viewed with a coarse triangulation that can be interactively changed to a fine triangulation to obtain a smooth surface representation. This type of filtering is difficult to implement for arbitrarily triangulated surfaces, especially if the triangulation changes during a simulation. Because altering the surface triangulation changes the interactive response time, with this method we can adjust the triangulation to suit the rendering performance of the available hardware. The filter program takes as input a high-resolution surface triangulation and extracts the triangles and vertices for the desired level of detail. This mesh structure is useful for computation as well as visualization. For example, our protein–protein docking program makes use of the subdivision hierarchy to compute interactions using large surface triangles and subdivides the triangles if the estimated error is too large. This procedure is analogous to the one used for hierarchical radiosity calculations[38] as well as other spectral decomposition methods.[39] Figure 1 shows examples of triangle, hexagon, and square meshes.

Table 1. Subdivision mesh sizes

	Icosahedron		Dodecahedron			Cube	
Subdivisions	Vertices	Triangles	Vertices	Pentagons	Hexagons	Vertices	Squares
0	12	20	20	12	0	8	6
1	42	80	80	12	30	26	24
2	162	320	320	12	150	98	96
3	642	1 280	1 280	12	630	386	384
4	2 562	5 120	5 120	12	2 550	1 538	1 536
5	10 242	20 480	20 480	12	10 230	6 146	6 144

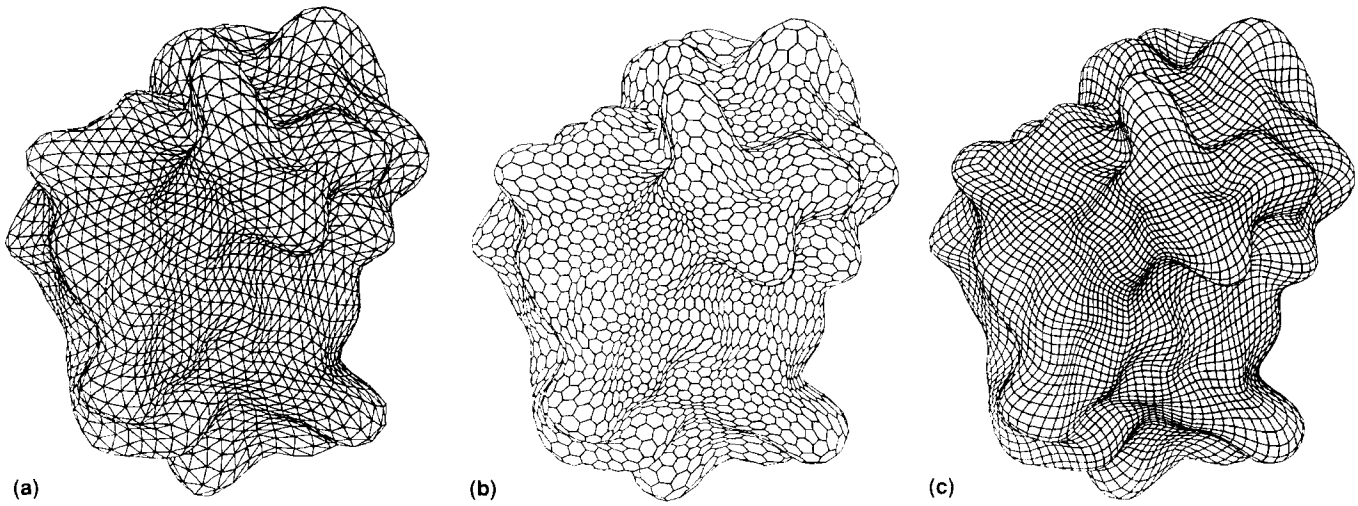

Figure 1. Three wireframe representations of an order 20 spherical harmonic surface of crambin generated from the same expansion coefficients. The triangular mesh (A) has 2 562 vertices and 5 120 triangles. The dual mesh (B) has 5 120 vertices, 12 pentagons, and 2 550 hexagons. The square mesh (C) has 6 146 vertices and 6 144 square polygons. Each mesh is displayed using the "outline gouraud" rendering mode to generate the hidden surface wireframe views.

Another method by which to present dynamic information is to draw simultaneously the surface motion vectors and the spherical harmonic surface. The density of the surface vectors is controlled using the same procedure for filtering the surface and is independent of the surface triangle density. The vectors are color coded to highlight the mobile parts of the structure.

Animation of surface motion is rendered using the Application Visualization System (AVS), a data-flow visualization system that runs on a wide variety of Unix workstations.[40–43] The custom AVS module Combine Modes computes the motion vectors resulting from a combination of precomputed normal modes. This module is used in AVS networks that display surface motion or the motion of individual atoms. This module evaluates the last term of Eq. (7) as a function of time and takes as input the mode frequencies and the mode vectors. The user can control the absolute temperature, the scale of the displacement, and the modes to be combined. The Coordmap module adds this displacement vector to the equilibrium coordinates to produce the new positions. These coordinates are rendered as triangles to show surface motion, or as lines when displaying the atomic excursions. Note that a single module computes the displacement vectors for surfaces and atoms; this is a recurring theme in the AVS style of visualization. The Downsizeico module does the filtering procedure and controls the speed of rendering by adjusting the resolution of the triangulation mesh. Visualization within AVS, however, does incur additional overhead because of data communication between separate modules. Stand-alone applications written directly in XGL, GL, or PHIGS would have less data communication overhead and possibly better performance. We believe that the flexibility of the dataflow architecture more than compensates for this limitation, especially for the development of new visualization techniques. Rendering is done on a DECStation 3000/500 AXP workstation with a Denali graphics accelerator from Kubota Pacific.

Triangulated surfaces can be color coded by any surface property. We often color the surface by the maximum displacement or the magnitude of the instantaneous displacement for a combination of normal modes. This is an effective technique by which to highlight the mobile regions of the surface. When coloring by the maximum displacement, the surface coordinates change during the motion but the surface color is constant. When coloring by the magnitude of the current displacement from the equilibrium position, the surface geometry and the surface color change during the motion.

In addition to visualizing animations of the molecular surfaces, we have used these AVS modules to view simultaneously a wireframe representation of the molecular surface and a line representation of the atomic structure. By this technique, the surface motion highlights the large-scale deformations of the surface and the line representation shows the detailed atomic excursions.

Our visualization strategy adds new flexibility to the display of dynamic information. By allowing interactive control of the resolution of the surface, the triangulation of the surface, and the resolution of the normal modes, the user can easily choose a level of detail that is appropriate for a given analysis.

RESULTS

We present an analysis of the protein crambin as an example of the application of these techniques for the analysis and display of surface motion. The crystal structure (1crn.pdb) was obtained from the Brookhaven Protein Data Bank,[44] and input for the AMBER package was prepared using the LEaP program.[45] The structure was minimized using the conjugate-gradient option of the NMODE module, which also computed the normal mode vectors and mode frequencies. These computations were done on a Convex C240 computer. The program MSMS computed the molecular

surface of the minimized structure.[23–25] Spherical harmonic representation of the surface geometry and surface motion vectors were computed as described above.

To compare the accuracy of the actual molecular surface motion and its spherical harmonic representation, we present two examples of normal mode trajectories. In each example, the amplitude of the oscillation is scaled by 2. Color Plate 1 shows an instance of the simulation of the lowest-frequency normal mode. The atomic motion vectors of the C_α atoms in a tube representation are compared with the surface motion vectors of the closest surface point in the spherical harmonic representation. Note that this comparison is only approximate because the motion of a surface point and that of its nearest C_α atom are not identical. Color Plate 2 shows an instance of the simulation of the five lowest-frequency normal modes. Again, the C_α motion vectors are compared with the spherical harmonic reconstruction. In each case, the surfaces are reconstructed to an order of 5 and the normal modes to an order of 30 [l_{max} of Eq. (3)].

A more precise method of comparing the motion vectors is shown in Color Plate 3. Here, we compare each Cartesian component (X,Y,Z) of the motion vectors for an order 30 spherical harmonic representation with those of the original united-atom CPK model. The spherical harmonic surface is reconstructed to order 10, and each molecule is color coded by the magnitude of a single mode vector component. These results show the level of accuracy that can be obtained by this technique; although atomic-scale features cannot be reconstructed, large-scale features are represented well.

Because the best method of presenting these results is with computer animation, we have prepared several MPEG examples accessible via the World Wide Web at: http://www.scripps.edu/pub/olson-web/doc/harmony.html.

CONCLUSION

The methods presented here are a novel approach for the uniform representation of protein surface geometry, properties, and motion. The advantages of spherical harmonic methods are as follows:

1. Multiresolution analysis: Multiresolution analysis allows the visual representation to match the level of detail appropriate for a given analysis and gives the user control of the interactive response time.

2. Uniform representation: Because surface properties, geometry, and motion are each represented by a set of spherical harmonic expansion coefficients, it is easy to have one surface property modulate another. For example, one can modulate the surface motion by the temperature factors of the solvent-accessible atoms so that the regions of high structural uncertainty have smaller or larger excursions.

3. Efficient visualization: The uniformity of the representation allows a small set of visualization and computation tools to perform many diverse tasks. We have found that the AVS environment is particularly well suited for this type of analysis.

Representing arbitrary motion, as computed from molecular dynamics simulations, is valuable but computationally expensive. Approximate surface motion, as computed by normal mode analysis, is useful for representing large-scale motion. The variable-resolution technique of representing surface geometry and surface motion using expansions of spherical harmonic functions is a technique for efficiently analyzing and visualizing these motions.

ACKNOWLEDGMENT

This work was funded by NIH Grant GM38794 (to A.J.O.).

REFERENCES

1 van Gunsteren, W.F., Weiner, P.K., and Wilkinson, A.J. *Computer Simulations of Biomolecular Systems,* Vol. 2. ESCOM Science Publishers, Leiden, 1993

2 Pearlman, D.A., Case, D.A., Caldwell, J.C., Seibel, G.L., Singh, U.C., Weiner, P., and Kollman, P.A. AMBER 4.0. University of California, San Francisco, California, 1991

3 Levitt, M. Molecular dynamics of native protein. I. Computer simulation of trajectories. *J. Mol. Biol.* 1983, **168,** 595–620

4 Levitt, M. Molecular dynamics of native protein. II. Analysis and nature of motion. *J. Mol. Biol.* 1983, **168,** 621–657

5 Case, D.A. Normal mode analysis of protein dynamics. *Curr. Opin. Struct. Biol.* 1994, **4**(2), 285–290

6 Brooks, B. and Karplus, M. Harmonic dynamics of proteins: Normal modes and fluctuations in bovine pancreatic trypsin inhibitor. *Proc. Natl. Acad. Sci. U.S.A.* 1983, **80,** 6571–6575

7 Seno, Y. and Gō, N. Deoxymyoglobin studied by the conformational normal mode analysis. I. Dynamics of globin and the heme-globin interaction. *J. Mol. Biol.* 1990, **216,** 95–109

8 Seno, Y. and Gō, N. Deoxymyoglobin studied by the conformational normal mode analysis. II. The conformational change upon oxygenation. *J. Mol. Biol.* 1990, **216,** 111–126

9 Horiuchi, T. and Gō, N. Projection of Monte Carlo and molecular dynamics trajectories onto the normal mode axes: Human lysozyme. *Proteins* 1991, **10,** 106–116

10 Ichiye, T. and Karplus, M. Collective motions in proteins: A covariance analysis of atomic fluctuations in molecular dynamics and normal mode simulations. *Proteins* 1991, **11,** 205–217

11 Tidor, B. and Karplus, M. The contribution of cross-links to protein stability: A normal mode analysis of the configuration entropy of the native state. *Proteins* 1993, **15,** 71–79

12 Gō, N., Noguti, T., and Nisikawa, T. Dynamics of a small globular protein in terms of low-frequency vibrational modes. *Proc. Natl. Acad. Sci. U.S.A.* 1983, **80,** 3696–3700

13 Diamond, R. On the use of normal modes in thermal parameter refinement: Theory and application to the bovine pancreatic trypsin inhibitor. *Acta Crystallogr. Sect. A* 1990, **46,** 425–435

14 Kidera, A. and Gō, N. Normal mode refinement: Crystallographic refinement of protein dynamic structure. I.

Theory and test by simulated diffraction data. *J. Mol. Biol.* 1992, **225,** 457–475
15 Kidera, A. and Gō, N. Normal mode refinement: Crystallographic refinement of protein dynamic structure. II. Application to human lysozyme. *J. Mol. Biol.* 1992, **225,** 477–486
16 Hao, M. and Harvey, S.C. Analyzing the normal mode dynamics of macromolecules by the component synthesis method. *Biopolymers* 1992, **32,** 1393–1405
17 Richards, F.M. Areas, volumes, packing and protein structure. *Annu. Rev. Biophys. Bioeng.* 1977, **6,** 151–176
18 Connolly, M.L. Solvent-accessible surfaces of proteins and nucleic acids. *Science* 1983, **221,** 709–713
19 Connolly, M.L. Analytical molecular surface calculation. *J. Appl. Crystallogr.* 1983, **16,** 548–558
20 Connolly, M.L. Molecular surface triangulation. *J. Appl. Crystallogr.* 1985, **18,** 499–505
21 Connolly, M.L. Shape complementarity at the hemoglobin $\alpha_1\beta_1$ subunit interface. *Biopolymers* 1986, **25,** 1229–1247
22 Connolly, M.L. The Molecular Surface Package. *J. Mol. Graphics* 1993, **11**(2), 139–143
23 Sanner, M.F. Modeling and Applications of Molecular Surfaces. Ph.D. Dissertation, Université de Haute-Alsace, France, 1992
24 Sanner, M.F., Olson, A.J., and Spehner, J.-C. Fast and robust computation of molecular surfaces. In: *Proc. 11th ACM Symp. Comp. Geom.*, 1995
25 Sanner, M.F., Olson, A.J., and Spehner, J.-C. Reduced surface: An efficient way to compute molecular surfaces (in press)
26 Duncan, B.S. and Olson, A.J. Approximation and characterization of molecular surfaces. *Biopolymers* 1993, **33,** 219–229
27 Max, N.L. and Getzoff, E.D. Spherical harmonic molecular surfaces. *IEEE Comput. Graphics Appl.* 1988, **8**(4), 42–50
28 Max, N.L. Approximating molecular surfaces by spherical harmonics. *J. Mol. Graphics* 1988, **6**(4), 210
29 Leicester, S.E., Finney, J.L., and Bywater, R.P. Description of molecular surface shape using Fourier descriptors. *J. Mol. Graphics* 1988, **6**(2), 104–108
30 Hobson, E.W. *The Theory of Spherical and Ellipsoidal Harmonics*. Chelsea Publishing Co., New York, 1955
31 Fisher, C.L., Tainer, J.A., Pique, M.E., and Getzoff, E.D. Visualization of molecular flexibility and its effects on electrostatic recognition. *J. Mol. Graphics* 1990, **8**(3), 125–132
32 Malhotra, A., Tan, R.K.-Z., and Harvey, S.C. Utilization of shape data in molecular mechanics using a potential based on spherical harmonic surfaces. *J. Comput. Chem.* 1994, **15**(2), 190–199
33 Duncan, B.S. and Olson, A.J. Texture mapping parametric molecular surfaces. *J. Mol. Graphics* 1995 (this issue)
34 Guggenheimer, H.W. *Differential Geometry*. Dover, New York, 1977
35 Duncan, B.S. and Olson, A.J. Protein-protein docking using parametric surfaces. 1995 (in preparation)
36 Klein, T.E., Huang, C.C., Petterson, E.F., Couch, G.S., Ferrin, T.E., and Langridge, R. A real-time malleable molecular surface. *J. Mol. Graphics* 1990, **8**(1), 16–24
37 Pique, M.E., Macke, T.J., and Arvai, A.S. Flex: A light-weight molecular display program. *J. Mol. Graphics* 1991, **9,** 40–41
38 Hanrahan, P., Salzman, D., and Aupperle, L. A rapid hierarchical radiosity algorithm. *Comput. Graphics* 1991, **25**(4), 197–206
39 Greengard, L. Fast algorithms for classical physics. *Science* 1994, **265**(5174), 909–914
40 Upson, C., Faulhaber, T.J., Kamins, D., Laidlaw, D., Schlegel, D., Vroom, J., Gurwitz, R., and van Dam, A. The Application Visualization System: A computational environment for scientific visualization. *IEEE Comput. Graphics Appl.* 1989, **9**(4), 30–42.
41 Duncan, B.S., Pique, M., and Olson, A.J. AVS for molecular modeling. In: *Proc. AVS '93 Conf.*, Int. AVS Center, Research Triangle Park, North Carolina, 1993
42 Duncan, B.S. and Olson, A.J. AVSTool: An interface to the AVS CLI. In: *Proc. AVS '94 Conf.*, Int. AVS Center, Research Triangle Park, North Carolina, 1994
43 Bowie, J.E. and Olson, A.J. (eds.). *Data Visualization in Molecular Science*. Addison-Wesley, New York, 1995
44 Bernstein, F.C., Koetzle, T.F., Williams, G.J.B., Meyer, E.F.J., Brice, M.D., Rodgers, J.R., Kennard, O., Shimanouchi, T., and Tasumi, M. The Protein Data Bank: A computer-based archival file for macromolecular structures. *J. Mol. Biol.* 1977, **112,** 535–542
45 Schafmeister, C.E.A.F. LEaP (Version 1.0). Department of Pharmaceutical Chemistry and Biophysics Program, University of California, San Francisco, California, 1993