Segmenting Molecular Surfaces

Vijay Natarajan [a,*] Yusu Wang [b] Peer-Timo Bremer [c]
Valerio Pascucci [d] and Bernd Hamann [a,e]

[a] *Institute for Data Analysis and Visualization, University of California, Davis*

[b] *Department of Computer Science and Engineering, Ohio State University*

[c] *Department of Computer Science, University of Illinois, Urbana-Champaign*

[d] *Center for Applied Scientific Computing, Lawrence Livermore National Laboratory*

[e] *Department of Computer Science, University of California, Davis*

Abstract

This paper presents a new method for segmentation of molecular surfaces. Topological analysis of a scalar function defined on the surface and its associated gradient field reveals the relationship between the features of interest and critical points of the scalar function. The segmentation is obtained by associating segments with local minima/maxima. Controlled simplification of the function merges segments resulting in a hierarchical segmentation of the molecular surface. This segmentation is used to identify rigid components of protein molecules and to study the role of cavities and protrusions in protein-protein interactions.

Key words: Segmentation, Morse theory, Morse-Smale complex, multiresolution data structure, visualization, structural biology, molecular graphics.

* Corresponding author
Email addresses:
`vijayn@ucdavis.edu` (Vijay Natarajan),
`yusu@cse.ohio-state.edu` (Yusu Wang),
`ptbremer@acm.org` (Peer-Timo Bremer),
`pascucci1@llnl.gov` (Valerio Pascucci),
`hamann@cs.ucdavis.edu` (Bernd Hamann).

1 Introduction

Rapid advances in imaging and simulation technologies allow us to study and understand life at the molecular level. In the quest towards understanding how molecules (proteins in particular) function, structural analysis emerges as a fundamental problem, because it is generally believed that the three-dimensional structure of a molecule to a large extent determines its particular functionality [1]. Structural analysis presents one of the major challenges in molecular biology: Detailed structures are extremely hard and time-consuming to obtain using existing imaging technology; protein structures are complex; there is no consensus on the best way to represent them; they are hard to manipulate; and, most important of all, the connection between the geometry (shape) and chemical properties of a molecule is not well-understood. Despite these difficulties, an increasing amount of research is being devoted to annotate and analyze molecular structures, especially in the post-genomic era.

Visualization methods are crucially important for comprehending the vast amount of structural information available to us in multiple online databases [2–5]. Humans are able to interpret and comprehend visual information more easily than numerical or textual information. The field of data visualization capitalizes on this ability and aims at providing the user with a deeper understanding of the data. Better understanding can be achieved by providing a comprehensive display of the data along with annotations. In molecular biology, visualization facilitates the recognition of meaningful relations between bio-chemical information and the shape of a molecule. One question is especially relevant: What aspects of the structure of a molecule determine its functionality and hence are crucial? For example, one can visualize various chemical properties, such as electrostatic potential or hydrophobicity, on molecular surfaces to see whether the binding sites (the locations on the molecule that interact with another molecule) have any inherent characteristics.

We use a surface representation of molecules. A molecule consists of a set of atoms. If each atom is modeled as a ball, then the molecule can be represented as the union of all balls. Various methods have been proposed to obtain a surface from the set of balls. The surface of this union, called the *Van der Waal surface* is easily computable but not continuous [6]. The *molecular surface* refers to the surface traced out by the inward-facing surface of a probing sphere rolled on top of the van der Waal surface [7]. It is continuous in most cases, but can have singularities (self-intersections). The *skin surface* is the envelope of families of an infinite number of evolving spheres [8]. It satisfies many desirable mathematical properties. For example, it is smooth everywhere and, although defined using an infinite number of spheres, it can be described by a finite number of quadric surface patches. We prefer the skin surface because of its above-mentioned properties.

Segmentation of the surface into regions that identify characteristic features is useful for visual analysis of the molecule. Different criteria and methods have been proposed for the segmentation of surfaces in general [9]. We focus on the problem of segmentation of molecular surfaces, in particular. Ideas presented in this paper are based on Morse theory, which allows us to infer the shape characteristics of a surface from the study of a real-valued function, called a Morse function, defined on the surface. For example, when the surface represents a terrain, we would like to capture its *peaks* and *valleys*. Such characterization is obtained by defining a function that evaluates the height of points, making it possible to associate regions with the local maxima and minima of this function. Defining such a function for a more general surface is non-trivial, and will naturally depend on the motivating application. For example, Hilaga et al. [10] define a function whose value at a given point is equal to the average geodesic distance to all points on the surface. Local maxima of this function identifies distinct features in the surface. They use the identified features for shape matching purposes. In certain scientific applications, the function is already defined over the surface and represents a physical quantity: for example, electron density, temperature, or pressure. In these cases, a segmentation of the domain surface helps the scientist understand the characteristics of the function.

We describe an algorithm for segmenting a molecular surface when such a function is given. Our algorithm computes a hierarchical segmentation of a molecular surface, enabling a user to interact with and analyze the surface at multiple levels of detail. We describe two applications to demonstrate the usefulness of our segmentation technique. Existing segmentation methods are either based on bio-chemical information [2] (if detailed atomic information is available), region-growing techniques [11–13], or Morse theory [14–16]. Our method has the advantage that it respects the topology of the associated function as described by its critical points, while those based on growing regions are ad-hoc. Region-growing methods, also called watershed algorithms, construct segments around either minima or maxima. Instead, we are able to simultaneously construct segments around both minima and maxima, and in a hierarchical fashion. Furthermore, we provide various options to compute the boundary of these segments, each appropriate for particular scenarios.

In practice, the input surface is usually represented as a triangle mesh in $\mathbb{R}^3$. The function defined over it is not smooth; it is a piecewise-linear (PL) function. These functions violate some of the assumptions made in the underlying mathematical framework of our methods. We refer to the papers by Edelsbrunner et al. [17] and Bremer et al. [14] for a comprehensive description of how key ideas from Morse theory can be extended to PL functions defined over triangle meshes. Morse theory is discussed in detail by Matsumoto [18] and Milnor [19]. Forman [20] has developed a discrete version of Morse theory for cell complexes, which has also been used to analyze molecular surfaces [15].

2 Flow-based Segmentation

Segmentation is a key ingredient for molecular surface analysis. We segment a given surface by first designing an appropriate scalar function over it and then analyzing its gradient flow field. The *Morse-Smale complex* partitions the surface into monotonic patches. We use these patches to construct a segmentation of the surface into peaks and valleys. Before describing our algorithm, we introduce various ideas from Morse theory and discuss their extension to the PL setting.

2.1 Critical Points

Let $\mathbb{M}$ be a closed surface in $\mathbb{R}^3$ and $f : \mathbb{M} \to \mathbb{R}$ be a real-valued smooth function. A point on the surface is *critical* if the gradient of f at this point is zero. f is a *Morse function* if none of its critical points are degenerate i.e., the Hessian matrix is non-singular for all critical points, and no two critical points have the same function value. The local behavior of a Morse function at critical points is characterized by the following lemma:

Morse Lemma [18]. Let p_0 be a critical point of a Morse function f defined on a surface. Then we can choose appropriate local coordinates (X, Y), with p_0 as the origin, in such a way that the function f expressed with respect to (X, Y) has one of the following three standard forms:

$$\begin{aligned} &\text{(i)}\ f = X^2 + Y^2 + c, \\ &\text{(ii)}\ f = X^2 - Y^2 + c, \quad \text{or} \\ &\text{(iii)}\ f = -X^2 - Y^2 + c, \end{aligned}$$

where $c = f(p_0)$.

The Morse Lemma states that a Morse function has quadratic behavior within a local neighborhood of every non-degenerate critical point p. This fact leads to a method for identification and characterization of critical points. Figure 1 shows the different types of critical points and how their local neighborhood characterizes the type of criticality. A molecular surface is typically represented by a triangular mesh K. The scalar function [1] f is sampled at the vertices and linearly interpolated within edges and triangles of K. The local neighborhood of a vertex needs to be formally defined in order to characterize critical points of a PL function defined on K. The *star* of a vertex v consists of all triangles

[1] We use f to denote both smooth and PL functions. The context clarifies what type of function we refer to.

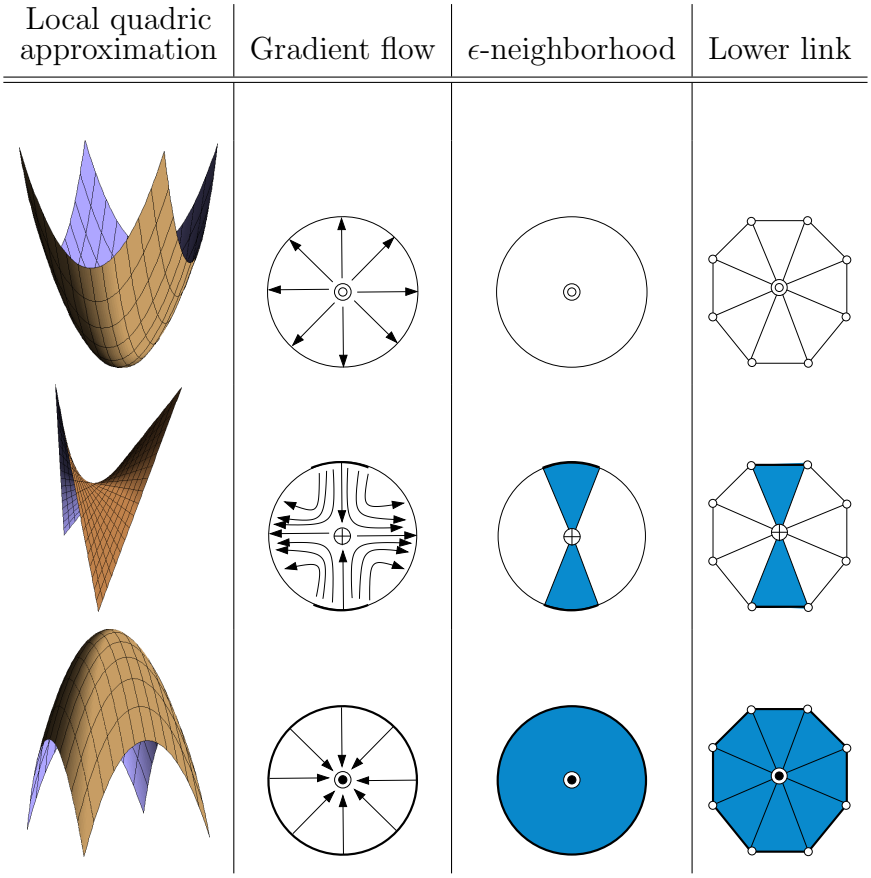
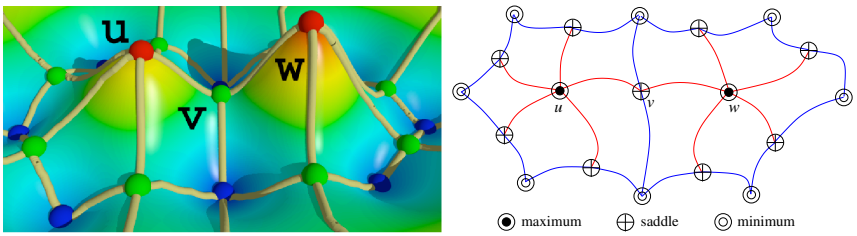


Fig. 1. Local neighborhoods of critical points. Top row: minimum; middle row: saddle; bottom row: maximum. The first column on the left shows the local approximation of the function using a quadric. In the remaining columns, the shaded regions and bold lines/curves have function value lower than the critical point.

and edges containing v. The *link* of v consists of those vertices and edges that are faces of edges and triangles in star(v) and are disjoint from v. The number of connected components of the *lower link* of v , consisting of vertices and edges in the link where f has values lower than $f(v)$, characterizes the critical points of f [21]. The index of a critical point is equal to the number of negative coefficients of its local quadratic approximation, which is equal to the number of negative eigen-values of the Hessian matrix.

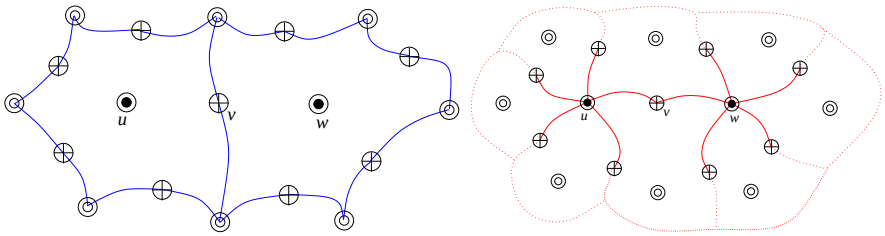
2.2 Morse-Smale Complexes

An integral line of f is a maximal path on the surface $\mathbb{M}$ whose tangent vectors agree with the gradient of f at every point of the path. Integral lines have a natural origin and destination at critical points where the gradient equals zero. Grouping the integral lines based on their origin and destination results in a segmentation of the surface. The Morse-Smale (MS) complex is a topological data structure that stores this segmentation (see Figure 2). Cells of different



(a) (b)

Fig. 2. (a) A simple height function with two maxima surrounded by multiple local minima and its Morse-Smale complex. (b) Combinatorial structure of the Morse-Smale complex in a planar illustration.



(a) (b)

Fig. 3. Morse complex computed for the set of all (a) maxima and (b) minima. The Morse-Smale complex is obtained as an overlay of these two Morse complexes.

dimension in the MS complex are called *nodes, arcs,* and *quads*. A quad clusters integral lines originating at a given minimum and terminating at an associated maximum. An arc consists of a single integral line connecting a saddle and an extremum (minimum/maximum). A node is a trivial integral line whose origin and destination are identical. Every cell of the MS complex is monotonic i.e., it does not contain any critical point in its interior. This characteristic property of the segmentation makes it attractive for various applications including remeshing, parametrization, hierarchical and multiresolution representation. Grouping integral lines based exclusively on either their origin or destination results in two new segmentations called the Morse complex (see Figure 3). The MS complex is also obtained as an overlay of the two Morse complexes.

Gradients, and hence integral lines, are not well-defined for PL functions. However, monotonic curves and surfaces corresponding to arcs and quads of the MS complex can be constructed by simulating a separation between integral lines that merge [14,17]. The MS complex is constructed in a three step process. The two Morse complexes are computed in the first two steps and overlaid in the third step to obtain the quads in the MS complex. Segments in the Morse complex are computed by tracing their boundary curves, which are paths of steepest ascent/descent. In the first step, paths of steepest descent are

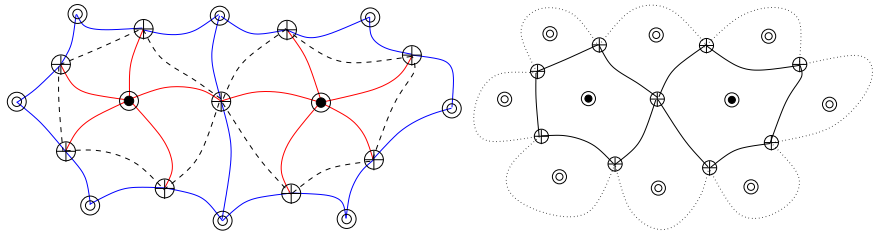
traced beginning at saddles and terminating at minima. If two paths merge geometrically, then they remain together till they terminate at a minimum. However the two paths are identified to be topologically distinct and stored as two separate paths. Paths of steepest ascent are traced in the second step while ensuring that they do not intersect the descending paths. The function f has its critical points at the nodes of this complex and is monotonic within all arcs and quads of the MS complex.

Peaks and valleys are natural features in a terrain. In the case of more general surfaces, the definition of peaks and valleys is non-trivial. Assuming that there exists a function that measures the "height" of a point in $\mathbb{M}$, we can associate peaks and valleys with the local minima, and maxima of this function. We define two such functions and discuss their applications in Section 3. Identifying the local "sea level" that divides two adjacent peaks or valleys is an interesting and challenging problem. The Morse complex and the MS complex provide only a partial solution to this problem. For example, computing the Morse complex for the set of all maxima results in a segmentation where peaks are separated. However, these segments extend all the way to adjoining valleys. The MS complex, on the other hand, refines more than necessary. We now describe how we extend this partial solution to obtain a segmentation that defines both peak and valley regions.

2.3 Peak-valley Decomposition

A quad in the MS complex contains one maximum and one minimum on its boundary. Consider a path within this quad connecting the two saddles. This path bisects the quad: One region contains the maximum and the other contains the minimum. Bisecting all quads that contain a given maximum u, we get all regions that constitute the peak containing u. Similarly, a valley is obtained by bisecting all quads that contain a given minimum. The saddle-saddle path describes the local sea level, forming the boundary between a peak and its adjoining valley. Figure 4 shows how these saddle-saddle paths form the boundaries of peaks and valleys in a surface. For a PL function defined on a triangle mesh, the saddle-saddle paths are restricted to lie along edges of the mesh.

We explore various optimality criteria for the construction of the saddle-saddle paths. Each one has its advantages and disadvantages. We unify all criteria into one framework, where we compute the shortest path between saddles in a weighted graph. The graph is constructed using the vertices and edges of mesh triangles constituting the quad, and edge weights are determined by one of the following optimality criteria:



(a) (b)

Fig. 4. (a) We compute paths connecting saddles within each quad. (b) These paths form the boundaries of a segmentation of the surface into peaks and valleys.

- **Geodesic path.** This is the simplest criterion. The weight of an edge is equal to the Euclidean distance between its endpoint vertices. Under reasonable sampling conditions, the shortest path in the mesh is a good approximation to the geodesic path between two points on the original smooth surface [22].
- **Horizontal path.** Consider sweeping the surface using isocontours of the height function. Two isocontour components that originate from adjacent minima merge at a saddle lying between them. Similarly, an isocontour component that bifurcates at a saddle eventually terminates at adjacent maxima. The height at the saddle, therefore, is a natural candidate for the sea-level. This is one of the reasons why we choose to compute paths between saddles. Restricting the path to lie within the height interval given by the saddle pair while computing the geodesic path results in a good interpolation between the height at the two saddles. Since such a path may not exist, we instead minimize deviation from the height interval by penalizing edges whose end points lie outside this height interval.
- **Monotonic path.** Another possible method for computing an interpolation between the height at saddles is one that attempts to create a path that is monotonic in height i.e., where the function value is non-decreasing (or non-increasing) from one saddle to a neighboring saddle. We implement this criterion by finding a directed path always from the lower saddle to the higher one and penalizing any edge that goes from a higher vertex to a lower one.
- **Interior path.** Paths between saddles can lie partially on the quad boundary. Figure 5 shows an example. This behavior frequently creates disconnected segments because it is very likely that geodesic paths between saddles within the adjacent quad also lies on the boundary. Adding a big penalty to edges lying on the boundary pushes the shortest path toward the interior. The segments are no longer disconnected, but they can be thin and elongated.
- **Medial path.** One way to prevent thin and elongated segments is to penalize nearness to the boundary. This forces the path to lie as far away from the extrema as possible, thereby creating rounded segments. Figure 5(c)

illustrates this idea.

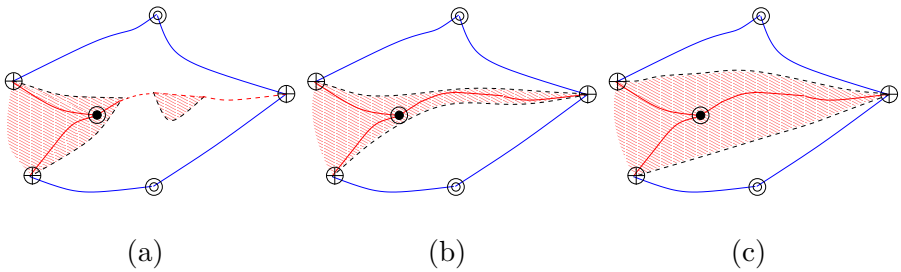


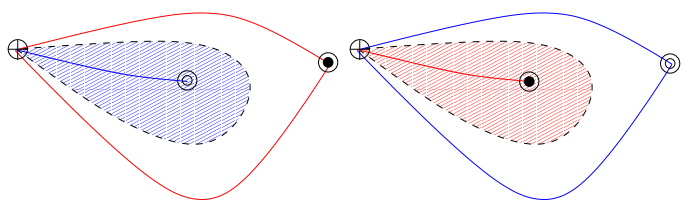
Fig. 5. The geodesic path (dashed) between saddles could lie on the quad boundary thereby creating disconnected segments. (a) The geodesic path within the lower quad touches the boundary twice (red dashed) and has two components that lie in the interior. The two disconnected segments of the maximum are shaded in red. (b) Penalizing boundary edges prevents the creation of disconnected segments but may result in thin and elongated segments. (c) Constructing a medial path that maximizes distance to the boundary results in segments that have a more rounded shape.

Figure 6 tabulates edge weights for the different optimality criteria. We demonstrate the advantages of these optimality criteria in Section 3, where we discuss applications of our segmentation to molecular surfaces.

The above framework does not hold when the quad contains exactly one saddle on its boundary. Two distinct arcs on the boundary of this quad connect a saddle with a unique extremum. The saddle-saddle path degenerates into a loop, which bisects the quad as shown in Figure 7. The loop passes through the saddle and separates the minimum and maximum. The optimality criteria described above for paths hold for this loop as well. Consider the case when both arcs from the saddle connect it to the same maximum. First, we construct the loop as follows: delete the minimum and its incident edges, creating a hole in the quad. The problem of computing a loop around the minimum reduces to computing a loop that does not shrink to a point. We compute such a loop, which in addition satisfies the optimality criterion, using the minimum weight tree that spans the quad and is rooted at the saddle. The weight of the shortest path from the root to each vertex is stored at the vertex. A loop is constructed as the union of a non-tree edge and paths from its endpoints to the root of the tree. After constructing the tree, we classify non-tree edges depending on whether or not they define a loop around the minimum that separates it from the maximum. The minimum weight loop is computed by simply comparing the weights of loops defined by each separating non-tree edge. The non-tree edges are classified based on an algorithm described by Edelsbrunner et al. [23]. This algorithms performs repeated collapses to classify edges: at each step remove a triangle, two of whose edges are either tree edges or marked, and declare the third edge as marked. This leaves all the separating edges unmarked.

Optimality criterion	Edge weight (w_{ab})
Geodesic	$w_{ab} = d(a,b)$
Horizontal	$w_{ab} = d(a,b) + \begin{cases} 0 & \text{if } f(a) \geq f(s_1), f(b) \leq f(s_2) \\ \alpha(f(s_1) - f(a)) & \text{if } f(a) < f(s_1), f(b) \leq f(s_2) \\ \alpha(f(b) - f(s_2)) & \text{if } f(a) \geq f(s_1), f(b) > f(s_2) \\ \alpha(f(s_1) - f(a)) & \\ +\alpha(f(b) - f(s_2)) & \text{if } f(a) < f(s_1), f(b) > f(s_2) \end{cases}$
Monotonic	$w_{ab} = d(a,b) + \begin{cases} 0 & \text{if } f(a) \geq f(b) \\ \alpha(f(b) - f(a)) & \text{if } f(a) < f(b) \end{cases}$
Interior	$w_{ab} = \begin{cases} \alpha \cdot d(a,b) & \text{if } ab \text{ is a boundary edge} \\ d(a,b) & \text{otherwise} \end{cases}$
Medial	$w_{ab} = \frac{1}{2}(\frac{1}{d_a} + \frac{1}{d_b})$

Fig. 6. Weight of edge ab corresponding to various optimality criteria used to construct paths between saddles that bisect quads. $d(a,b)$ is the length of the edge ab, α is a large constant, and d_a and d_b are the shortest geodesic distances from a and b to the quad boundary.

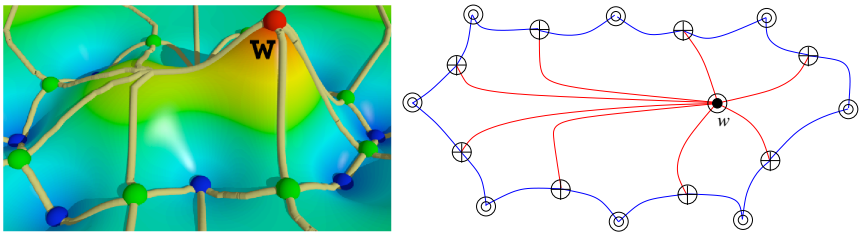


(a) (b)

Fig. 7. For quads that contain exactly one saddle, a loop (dashed) is constructed to bisect the quad. This loop passes through the saddle and separates the minimum and maximum.

2.4 Hierarchical Segmentation

A major advantage of using the MS complex as a starting point for segmentation is that we can segment the surface at multiple levels of detail. A smoother Morse function can be generated from f by repeated cancellation of pairs of critical points that are connected by an arc in the MS complex. Upon cancella-



(a) (b)

Fig. 8. Smoother height functions are created by cancelling pairs of critical points. (a) Cancelling a saddle-maximum pair removes a topological feature, and it is implemented by merging a maximum-saddle-maximum triple into a single maximum w. The function is modified locally by rerouting integral lines to the remaining maximum. (b) The MS complex is modified by removing the two critical points and extending the incident arcs to reconnect the complex.

tion, a saddle-extremum pair is removed from the MS complex along with arcs incident on the critical pair. New arcs are added to reconnect the complex. Figure 8 shows how the MS complex from Figure 2 is modified after cancelling a saddle-maximum pair. Cancellation can be implemented as a purely combinatorial procedure that simplifies the MS complex. A numerical realization of the cancellation requires a local modification of the function and its gradient field. We utilize the combinatorial structure of the simplified MS complex to identify the neighborhood of the critical pair where the function needs to be modified. Repeated application of a Laplacian smoothing operator within each of the quads in this neighborhood yields a function that corresponds to the simplified MS complex. Each cancellation makes the function smoother by removing a topological feature. Critical pair cancellations were originally developed to generate minimal Morse functions, which in turn were used to prove important results like the generalized Poincaré conjecture for higher dimensions [24].

We order critical point pairs based on the notion of *persistence* [25], which quantifies the importance of the associated topological feature. The persistence of a critical point pair is defined as the absolute difference in the values of f between the two points.

The peak-valley decomposition is computed at multiple levels of detail by first cancelling critical point pairs in the MS complex and then constructing saddle-saddle paths within the simplified complex. Cancelling a saddle-maximum pair merges the corresponding peak with a neighboring peak. Similarly, cancelling a saddle-minimum pair merges the corresponding valley with a neighboring valley.

3 Application I: Segmentation of Semi-rigid Molecules

3.1 Motivation

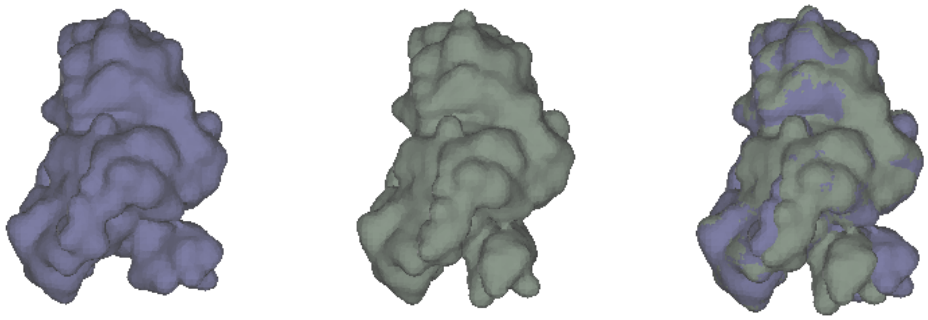
Molecules interact with each other, and the interaction process is highly influenced by their three-dimensional structures. Although, in general, a molecule takes on a unique conformation in space, many of them can change their conformations during important biochemical processes, such as protein-protein interaction. Hence, in order to understand how molecules work, it is essential to understand these deformations. Popular methods to obtain molecular structures, such as X-ray crystallography and nuclear magnetic resonance (NMR) imaging, can only capture snapshots of the conformational states of molecules. In some cases, such as in protein docking, the process happens so rapidly that we can only capture two conformations of the participating proteins: one before and one after the process. The study of the entire deformation is now based on these discrete conformations. To make things worse, high-resolution structures (including information such as type of atoms in the molecule, atom positions, etc) are fairly hard to obtain. Therefore, an increasing number of scientists nowadays study and analyze macromolecules directly using easily obtainable low-resolution structures (such as cryo-electron microscopy data) without any knowledge of atomic information [26–29]. Our goal is to develop visualization tools that support this endeavor.

It is known that the majority of macromolecular motions can be described by the so-called *hinge-motion* [30], where a molecule consists of a few semi-rigid components, each moving in a roughly rigid manner (Figure 9 shows an example). Given two conformations of the same molecule undergoing the hinge motion, we would like to segment the input structures into these semi-rigid subcomponents. Input conformations are given as a surface model. The underlying atomic structure is not known, and we are unaware of the correspondences between input structures. Our segmentation is based on the *deformation* function defined on input molecular surfaces, which we describe next.

3.2 Function Definition

Given two triangulated molecular surfaces P and Q together with a small set of *landmarks* $L \subseteq P$, $M \subseteq Q$, where $L = \{l_1, \ldots, l_k\}$, $M = \{m_1, \ldots, m_k\}$, and l_i corresponds to m_i for all $i \in [1, k]$, we want to identify components of the molecule that undergo rigid motion. The constant k is usually small, typically equal to 4 in our experiments. In other words, we know a few reliable corre-

spondences between input structures, and we wish to induce a segmentation from them. These landmarks can either be obtained by attaching biochemical markers to molecules, or by shape analysis methods [31]. Furthermore, since large portions of the input structures are similar, it is not hard to compute a good coarse alignment between P and Q. We use the algorithm from [32] for this purpose. The resulting alignment (i.e., a rigid-body transformation for structure Q) is shown in Figure 9(c). In this example, the original structure, say P, consists of two major components, the main body with one "leg" attached to it, and a small "leg" in the bottom-right corner. The coarse alignment places the two main bodies close together.



(a) (b) (c)

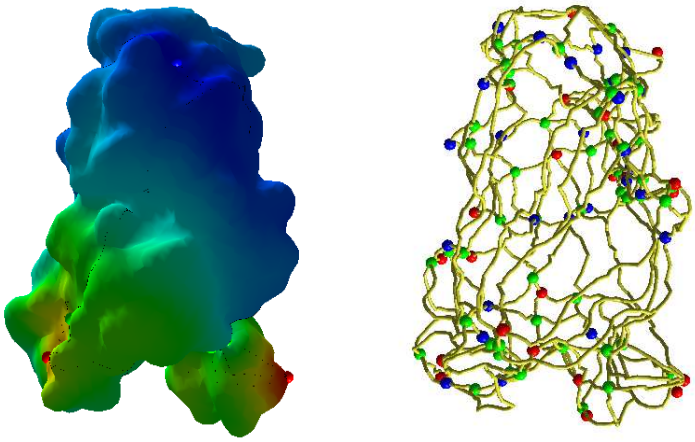
Fig. 9. Two conformations, P and Q, of the same protein (myosin) shown in (a) and (b), respectively. (c) The top-scored alignment between P and Q as computed by [32].

Now, assume that P and Q are already aligned as shown in Figure 9 (c). For every point $p \in P$, let $\mathrm{vc}(p)$ denote the *virtual coordinate of* p based on the landmark set L: $\mathrm{vc}(p) = (\mathrm{gd}(p, l_1), \ldots, \mathrm{gd}(p, l_k))$, where $\mathrm{gd}(p, l_i)$ is the geodesic distance from p to the i'th landmark l_i on surface P. Similarly, define $\mathrm{vc}(q)$ for every $q \in Q$ with respect to landmark set M. We define the distance between $p \in P$ and $q \in Q$ as a weighted sum of the distance between p and q in the Euclidean space and in the virtual coordinate space:

$$\mathrm{dist}(p, q) = \lambda \|p - q\| + (1 - \lambda) \|\mathrm{vc}(p) - \mathrm{vc}(q)\|,$$

where $\| \cdot \|$ denotes the L_2 norm, and $0 \leq \lambda \leq 1$.

For every point $p \in P$, let $\mathrm{NN}(p)$ be the nearest neighbor of p from points in Q using the distance metric introduced above i.e., $\mathrm{NN}(p) = \mathrm{argmin}_{q \in Q} \mathrm{dist}(p, q)$. We define the deformation function, $\mathrm{F} : P \rightarrow \mathbb{R}$, as $\mathrm{F(p)} = \mathrm{dist(p, NN(p))}$. The better the two surfaces are aligned, the smaller the deformation function value is. Figure 10(a) shows a visualization of the function F mapped onto the

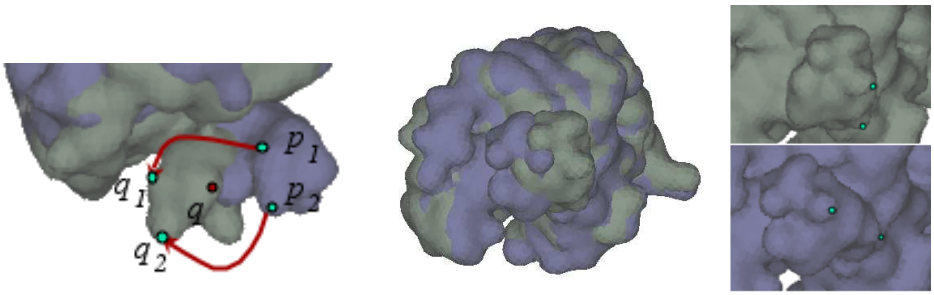


(a) (b)

Fig. 10. (a) Visualization of the deformation function. (b) Corresponding Morse-Smale complex.

surface P. Lighter colors correspond to higher function values.

We apply our segmentation tool to the deformation function F. The intuition is that the main component should be the segment around a minimum (because the main components from the two structures are aligned well and hence the deformation function values are lower), and the moving components should be close to a maximum (points on this component should have high function values).

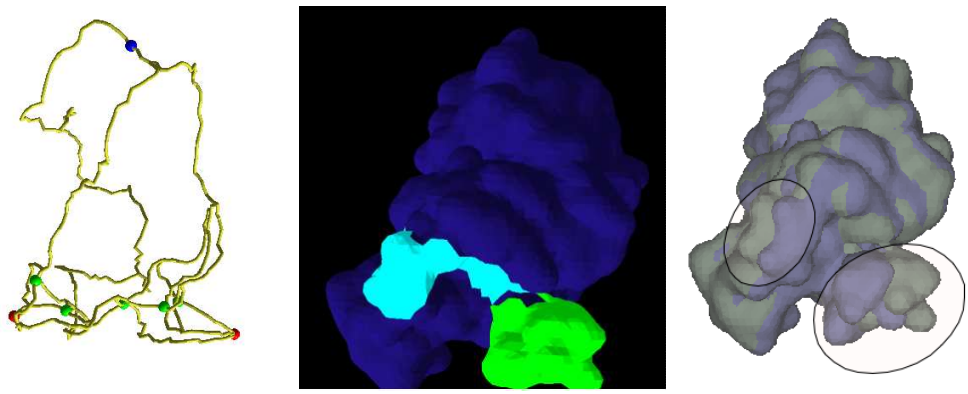


(a) (b) (c)

Fig. 11. Points p_1 and p_2 in (a) should correspond to q_1 and q_2, respectively. The two points will most likely be matched to q if Euclidean distance is used for determining correspondences. The two landmarks chosen for the small components on P and Q are shown in (c). The overall alignment is shown in (b).

The reason for using vc(p) instead of the usual Cartesian coordinates of p and q is the fact that geometric distance is not invariant under the type of deformation that we consider. The correspondence between points in P and Q

might be wrong for many points when using purely Euclidean distance (even if surface normals are known and considered). For example, in Figure 11, points p_1 and p_2 both associated q as their nearest neighbor on the surface Q, while their true correspondences should be q_1 and q_2. Geodesic distances support a more stable behavior and are more appropriate for our semi-rigid-body deformations. Therefore, if the correspondences between the landmarks are reliable, we will find that p_1 and p_2 correspond to q_1 and q_2 in this virtual coordinate system. In our experiments, we choose $\lambda < 0.5$, so that the geodesic distance in the virtual coordinates system has a higher weight.



(a) (b) (c)

Fig. 12. (a) Simplified Morse-Smale complex containing one minimum (blue), two maxima (red), and four saddles (green). (b) Segmentation based on the simplified Morse-Smale complex. (c) Corresponding semi-rigid components.

3.3 Segmentation Results

We are interested in human myosin, which is responsible for moving muscle fibers by moving its two "legs". Figure 9 shows two conformations of myosin during this process. Given two surface models, we use the method from [31] to compute four landmarks on each of them: two from the main body and two from the small leg. Figure 11(b)-(c) shows the two landmarks from the small components; the two surfaces are aligned using the rigid-body transformation computed by the algorithm from [32] (see Figure 9(c)). The two landmarks from the main body components are not shown, because they are close to their correspondences in Euclidean space. Finally, we compute the deformation function F for surface P using the method described above.

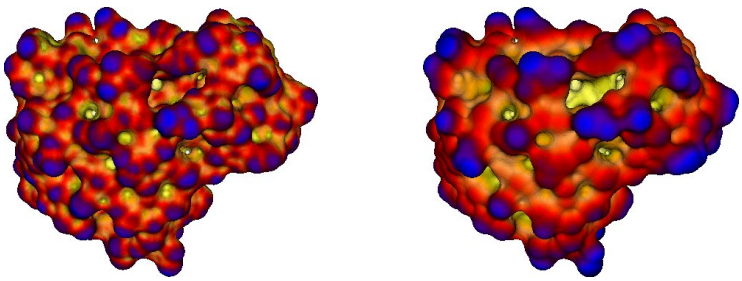
A visualization of F is shown in Figure 10(a). Intuitively, points from the main body have relatively lower function values, while points from the small leg have higher values because they are not well-aligned by a rigid-body transforma-

tion[2]. The MS complex of F is shown in Figure 10(b). We simplify it until there are only two maxima and one minimum remaining, see Figure 12(a). Currently, this is an interactive process, and the user can decide how much simplification to perform. The corresponding segmentation is shown in Figure 12 (b). Our method identifies both the main body and the small leg, as well as another small component that moves substantially, see Figure 12(c).

4 Application II: Cavities and Protrusions

4.1 Motivation

Cavities and protrusions play a crucial role in how molecules function. For example, proteins interact with each other by binding to form a three-dimensional "jigsaw puzzle", either temporarily or permanently. The two structures roughly complement each other in the regions defining the binding sites, and it is generally believed that this phenomenon is to a large extent dominated by features such as protrusions and cavities. In order to study the binding process, it is desirable that the surfaces are segmented into convex and concave patches. Given a triangulated molecular surface, current methods typically use a variant of the popular *Connolly function* [6] to obtain this segmentation, see Figure 13(a). The Connolly function can be thought of as an analog of the



(a) (b)

Fig. 13. Visualization of (a) Connolly function and (b) atomic density function

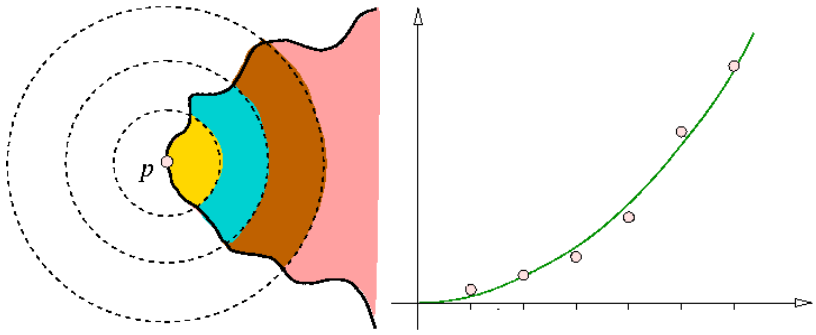
mean curvature within a fixed size neighborhood of each vertex [15]. Hence, a convex patch corresponds to a segment around a minimum, a concave patch

[2] Due to the noise induced by structure obtaining methods, even the main bodies of two conformations seldom have a perfect match. Hence many points may have a low function value, but not strictly zero. In our experiments, we rarely observe large regions of degeneracy (i.e, regions of same function values).

corresponds to a segment around a maximum of the function, and the desired natural segmentation can be obtained using the framework described in Section 2.

4.2 Function Definition

Given a surface model $\mathbb{M} \subseteq \mathbb{R}^3$ of a molecule, let $\overline{\mathbb{M}}$ denote the volume enclosed by $\mathbb{M}$. The particular function we consider, $\mu : M \to \mathbb{R}$, is a variant of the *atomic density* function originally proposed in [33], see Figure 13(b). It improves the Connolly function by taking more surface details around a point into account. This is achieved by considering a series of neighborhoods around each vertex, instead of only one fixed-size neighborhood as considered for the Connolly function. More precisely, for every point $p \in \mathbb{M}$, let $\mathrm{B}(p, r)$ denote



(a) (b)

Fig. 14. (a) Illustration of the atomic density function μ in the plane. We compute $\mathrm{V}(p, r)$ for a sequence of radii. (b) We plot the set of radius-volume pairs in a graph and find the best quadratic approximation (cubic approximation for surfaces in 3D).

the ball centered at p with radius r, and $\mathrm{V}(p, r)$ the volume of $\overline{\mathbb{M}} \cap \mathrm{B}(p, r)$, i.e, the intersection between $\overline{\mathbb{M}}$ and $\mathrm{B}(p, r)$. Assume we compute $y_i = \mathrm{V}(p, r_i)$ for a sequence of radii $r_i = i \cdot r$, $i = 1, \ldots, k$, which defines a set of points (r_i, y_i) in the plane. We now find the best coefficient σ such that the function $y = \sigma x^3$ best approximates these points, and assign $\mu(p) = \sigma$. Figure 14 provides an illustration.

4.3 Segmentation Results

We use the chain D from the protein complex Barnase/Barstar with pdb-id 1BRS for our experiment. We use the software described in [8] to compute a triangulation of the skin surface of the given molecule. Our choice of the skin

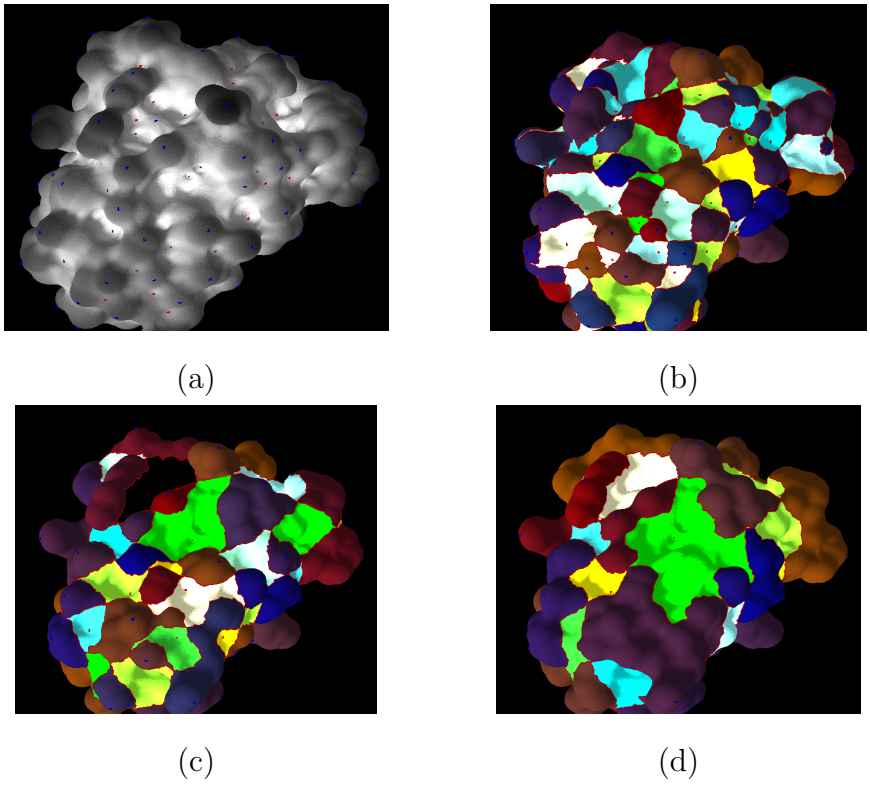
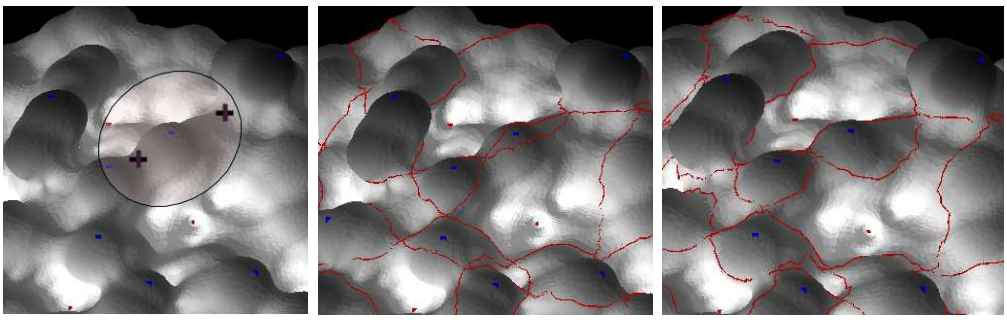


Fig. 15. (a) The atomic density function: Darker regions correspond to protrusions and lighter regions correspond to cavities. Simplified triangulations and their segmentations are shown in (b), (c), and (d), consisting of approximately 300, 160, and 80 segments, respectively.

surface is motivated by the availability of a software tool that can compute a high-resolution triangulation of the molecular skin, and a guarantee that the resulting mesh has no self-crossings and has good quality triangles. We compute the atomic density function over this skin surface. Figure 15(a) shows a visualization of this function. We compute the MS Complex, simplify it to three different resolutions, and compute corresponding segmentations, see Figure 15(b)-(d). As we simplify the function, multiple segments merge into larger segments, capturing cavities/protrusions at progressively coarser levels.

Our framework supports various ways to connect saddles. In most cases, the interior paths produce reasonable results. In particular, segment boundaries constructed using this optimality criterion are usually smooth (due to its similarity to geodesic paths). However, it is more desirable to use other options in certain situations. For example, Figure 16(a) shows part of a surface containing a minimum whose degree in the simplified MS complex is two, i.e., there are two saddle points on the boundary of the unstable manifold around this minimum. The monotonic path criterion produces better results than the geodesic path or interior path. The former criterion respects the function value more and therefore produces a better path, although the resulting paths as shown in Figure 16(c) may not be as smooth as the ones shown in Figure 16(b).

Another case where the influences of various options is more obvious occurs when a quad contains exactly one saddle on its boundary.



(a) (b) (c)

Fig. 16. (a) The minimum (blue) within the ellipse has only two saddles on the boundary of its corresponding segment. Using the interior path criterion to connect saddles gives rise to the results shown in (b), while the monotonic path criterion produces the results shown in (c).

4.4 Time complexity

The input surface is represented as a triangulation of n vertices and $O(n)$ edges. The construction and simplification of MS complex takes $O(n \log n)$ time. The paths connecting saddles are essentially various (weighted) shortest paths based on the graph induced by the triangulation. In the worst case, computing a path within a quad takes $O(n_i \log n_i)$ time, where n_i is the number of vertices in the ith quad. Therefore, the total time taken to compute all saddle-saddle paths equals $O(\sum_{i=1}^{k} n_i \log n_i) = O(n \log n)$, where k is the number of quads. Indeed, current software is very efficient and real-time interactive for data reported in this paper (the triangulation for first and second experiments has 10K and 51K vertices respectively).

5 Conclusions

We describe a novel method for computing a hierarchical segmentation of molecular surfaces. The idea arises naturally from flow-based methods that perform a topological analysis of scalar functions defined on the surface. Our method works for any surface, but we restrict our attention to molecular surfaces in this paper. Our approach supports various options to choose 1) the optimality criterion that drives the segmentation and 2) the resolution at which

to perform the segmentation. We describe two applications of our method to demonstrate it usefulness.

There are several interesting questions that we plan to address in the future. What is a good segmentation when there is exactly one minimum and one maximum? It seems unlikely that there exists a universally good strategy. The above situation arises when we perform extreme topological simplification for a surface with genus zero. One possibility is to stop the simplification when the MS complex has exactly one saddle point and compute the corresponding segmentation. The required segmentation is then obtained by merging segments associated with the two extrema that would merge upon performing the subsequent cancellation. A multiresolution representation of the segmentation can also be obtained by merging segments that were computed from the full resolution MS complex. How does the result of this approach compare with the representation obtained using cancellations on the MS complex?

One problem that we are currently studying is visual analysis of protein-protein interactions using the segmentation induced by the atomic density function. Further study is needed to determine whether our segmentation creates meaningful surface patches that correspond to active sites, and whether there is clear correlation between cavities and protrusions from the two active sites of the interacting proteins. Another important direction for future research is the application of our method to other applications besides structural biology. For example, features in data from neuroscience, material science, and geology can be studied using our segmentation.

Acknowledgements

All pictures of segmentation results were generated using a molecular surface viewer developed by Harnit Singh and Michael Maga. We thank Herbert Edelsbrunner, David Cohen-Steiner, and members of the Visualization and Computer Graphics Research Group at the Institute for Data Analysis and Visualization (IDAV), UC Davis, for helpful discussions. This work was partially supported by the National Science Foundation under contract ACI 9624034 (CAREER Award), through the Large Scientific and Software Data Set Visualization (LSSDSV) program under contract ACI 9982251, through the National Partnership for Advanced Computational Infrastructure (NPACI) and a large Information Technology Research (ITR) grant; and the National Institutes of Health under contract P20 MH60975-06A2, funded by the National Institute of Mental Health and the National Science Foundation. This work was performed under the auspices of the U.S. Department of Energy by University of California Lawrence Livermore National Laboratory under contract No. W-7405-Eng-48.

References

[1] W. R. Taylor, A. C. W. May, N. P. Brown, A. Aszodi, Protein structure: geometry, topology and classification, Reports on Progress in Physics 64 (2001) 517–590.

[2] C. Bajaj, P. Djeu, A. Thane, V. Siddavanahalli, TexMol: Interactive visual exploration of large flexible multi-component molecular complexes, in: Proc. IEEE Conf. Visualization, 2004, pp. 243–250.

[3] C. L. Bajaj, V. Pascucci, A. Shamir, R. J. Holt, A. N. Netravali, Dynamic maintenance and visualization of molecular surfaces, Discrete Appl. Math. 127 (1993) 23–51.

[4] A. R. Leach, Molecular modelling: Principles and applications, Pearson Education Limited, 1996.

[5] R. Voorintholt, M. T. Kosters, G. Vegter, G. Vriend, W. G. Hol, A very fast program for visualizing protein surfaces, channels and cavities, Journal of Molecular Graphics 7 (4) (1989) 243–245.

[6] M. L. Connolly, Molecular surface: A review, Network Science.

[7] F. M. Richards, Areas, volumes, packing and protein structure, Ann. Rev. Biophys. Bioeng. 6 (1977) 151–176.

[8] H.-L. Cheng, T. K. Dey, H. Edelsbrunner, J. Sullivan, Dynamic skin triangulation, Discrete Comput. Geom. 25 (2001) 525–568.

[9] A. Shamir, A formulation of boundary mesh segmentation, in: Second International Symposium on 3D Data Processing, Visualization and Transmission (3DPVT'04), 2004, pp. 82–89.

[10] M. Hilaga, Y. Shinagawa, T. Komura, T. L. Kunii, Topology matching for full automatic similarity estimation of 3d shapes, in: Proc. SIGGRAPH, 2001, pp. 203–212.

[11] A. P. Mangan, R. T. Whitaker, Partitioning 3d surface meshes using watershed segmentation, IEEE Transactions on Visualization and Computer Graphics 5 (4) (1999) 308–321.

[12] T. Srinark, C. Kambhamettu, A novel method for 3d surface mesh segmentation, in: Proc. 6th IASTED International Conference on Computers, Graphics, and Imaging, 2003, pp. 212–217.

[13] F. Vivodtzev, L. Linsen, G.-P. Bonneau, B. Hamann, K. I. Joy, B. A. Olshausen, Hierarchical isosurface segmentation based on discrete curvature, in: VISSYM '03: Proceedings of the symposium on Data visualisation, 2003, pp. 249–258.

[14] P.-T. Bremer, H. Edelsbrunner, B. Hamann, V. Pascucci, A topological hierarchy for functions on triangulated surfaces, IEEE Transactions on Visualization and Computer Graphics 10 (4) (2004) 385–396.

[15] F. Cazals, F. Chazal, T. Lewiner, Molecular shape analysis based upon the Morse-Smale complex and the Connolly function, in: Proc. 19th Ann. ACM Sympos. Comput. Geom., 2003, pp. 351–360.

[16] V. Natarajan, V. Pascucci, Volumetric data analysis using Morse-Smale complexes, in: Proc. Intl. Conf. Shape Modeling and Applications, 2005, pp. 320–325.

[17] H. Edelsbrunner, J. Harer, A. Zomorodian, Hierarchical Morse-Smale complexes for piecewise linear 2-manifolds, Discrete and Computational Geometry 30 (1) (2003) 87–107.

[18] Y. Matsumoto, An Introduction to Morse Theory, Amer. Math. Soc., 2002, translated from Japanese by K. Hudson and M. Saito.

[19] J. Milnor, Morse Theory, Princeton Univ. Press, New Jersey, 1963.

[20] R. Forman, Morse theory for cell complexes, Advances in Mathematics 134 (1998) 90–145.

[21] T. F. Banchoff, Critical points and curvature for embedded polyhedral surfaces, American Mathematical Monthly 77 (5) (1970) 475–485.

[22] M. Bernstein, V. de Silva, J. C. Langford, J. B. Tenenbaum, Graph approximations to geodesics on embedded manifolds, Tech. rep., Department of Psychology, Stanford University (2000).

[23] H. Edelsbrunner, J. Harer, V. Natarajan, V. Pascucci, Morse-Smale complexes for piecewise linear 3-manifolds, in: Proc. 19th Ann. Sympos. Comput. Geom., 2003, pp. 361–370.

[24] S. Smale, Generalized Poincaré's conjecture in dimensions greater than four, Ann. of Math. 74 (1961) 391–406.

[25] H. Edelsbrunner, D. Letscher, A. Zomorodian, Topological persistence and simplification, Discrete and Computational Geometry 28 (4) (2002) 511–533.

[26] D. Ming, Y. Kong, M. A. Lambert, Z. Huang, J. Ma, How to describe protein motion without amino acid sequence and atomic coordinates, Proc. Natl. Acad. Sci. USA 99 (13) (2002) 8620–8625.

[27] H. R. Saibil, Conformational changes studied by cryo-electron microscopy, Nat. Struct. Biol. 7 (9) (2000) 711–714.

[28] E. Tama, W. Wriggers, C. L. Brooks III, Exploring global distortions of biological macromolecules and assemblies from low-resolution structural information and elastic network theory, J. Mol. Biol. 321 (2) (2002) 297–305.

[29] E. Tama, O. Miyashita, C. L. Brooks III, Normal mode based flexible fitting of high-resolution structure into low-resolution experimental data from cryo-EM, J. Struct. Biol. 147 (3) (2004) 315–326.

[30] N. Echols, D. Milburn, M. Gerstein, MolMovDB: analysis and visualization of conformational change and structural flexibility, Nucleic Acids Research 31 (1) (2003) 478–482.

[31] L. Guibas, Y. Wang, Segmenting low resolution macro-molecular structures under hinge motion, manuscript (2005).

[32] Y. Wang, P. Agarwal, P. Brown, H. Edelsbrunner, J. Rudolph, Fast geometric algorithm for rigid protein docking, in: Proc. 10th. Pacific Symposium on Biocomputing (PSB), 2005, pp. 64–75.

[33] J. C. Mitchell, R. Kerr, L. F. T. Eyck, Rapid atomic density measures for molecular shape characterization, J. Mol. Graph. Model. 19 (2001) 324–329.