Quality Meshing of Implicit Solvation Models of Biomolecular Structures*

Yongjie Zhang
Institute for Computational Engineering and Sciences
Univ. of Texas at Austin
USA
jessica@ices.utexas.edu

Guoliang Xu
Academy of Mathematics and System Sciences
Chinese Academy of Sciences, China
xuguo@lsec.cc.ac.cn

Chandrajit Bajaj
Institute for Computational Engineering and Sciences
Dept. of Computer Sciences
Univ. of Texas at Austin, USA
bajaj@cs.utexas.edu

ABSTRACT

This paper describes a comprehensive approach to construct quality meshes for implicit solvation models of biomolecular structures starting from atomic resolution data in the Protein Data Bank (PDB). First, multi-scale volumetric synthetic electron density maps are constructed from parsed atomic location data of biomolecules in the PDB, using Gaussian isotropic kernels. An appropriate parameter selection is made for constructing an error bounded implicit solvation surface approximation to the Lee-Richards molecular surface. Next, a modified dual contouring method is used to extract triangular meshes for the molecular surface, and tetrahedral meshes for the volume inside or outside the molecule within a bounding sphere/box of influence. Finally, geometric flows are used to improve the mesh quality. Some of our generated meshes have been successfully used in finite element simulations.

Categories and Subject Descriptors: I.3.5 [Computation Geometry and Object Modeling]: Computer Graphics - Computational Geometry and Object Modeling

General Terms: Algorithms

Keywords: Quality mesh, biomolecule, implicit solvation models, multi-scale model, finite element simulations

1. INTRODUCTION

Finite Element Method (FEM) has become an important tool in the analysis of physical properties of complicated biomolecules, such as biomolecular electrophoresis, electrostatics and diffusion influenced reaction rate constants [26], [27], [31]. In efficient and accurate finite element simulations, adaptive and quality meshes are required. Quite often, people have to give up FEM because they can not generate satisfied triangular or tetrahedral meshes to represent the geometric model for large complicated biomolecules such as Ribosome (Fig. 1), or those structures whose active site occurs at the bottom of a narrow pocket (deep gorge) (Fig.3).

The protein data bank (http://www.rcsb.org/pdb) provides PDB format files for protein and RNA structures, with the location of principally all the major atoms (e.g., hydrogen atoms are not discernible via X-ray diffraction and therefore rarely present in the PDB). The summation of kernel functions centered at each atom can be used to construct a smooth volumetric electron density map from PDB data [10] [3]. The volumetric data is often sampled at each rectilinear grid point, $V = \{F(i,j,k)|i,\ j,\ k$ are indices of $x,\ y,\ z$ coordinates in a rectilinear grid$\}$, and the implicit solvation surface is approximated as a level set $S_F(c) = \{(x,y,z)|\ F(i,j,k) = c\}$, where c is a constant [10] [16]. The computation of density maps can be made very efficient with worst case complexity linear in the number of grid points and the number of atoms [2]. In this paper, we describe an approach to generate quality triangular/tetrahedral meshes for complex biomolecular structures from PDB format data, conforming to good implicit solvation surface approximations. There are three main steps in our mesh generation process:

1. Implicit Solvation Surface – A good approximation of the implicit solvation surface is generated from multi-scale volumetric synthetic electron density maps by a careful choice of Gaussian kernel parameters.

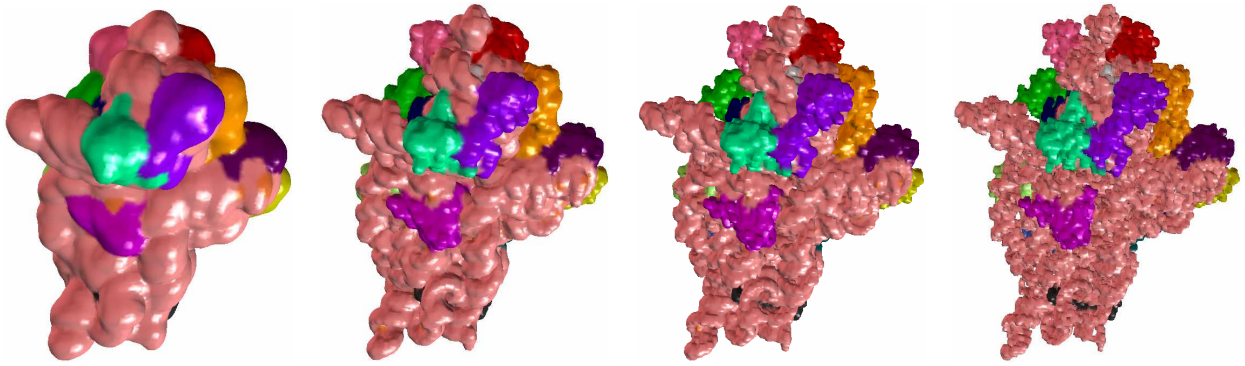
2. Mesh Generation – The modified dual contouring method is used to generate triangular and interior/exterior tetrahedral meshes.

3. Quality Improvement – Geometric flows are used to improve the quality of extracted triangular and tetrahedral meshes.

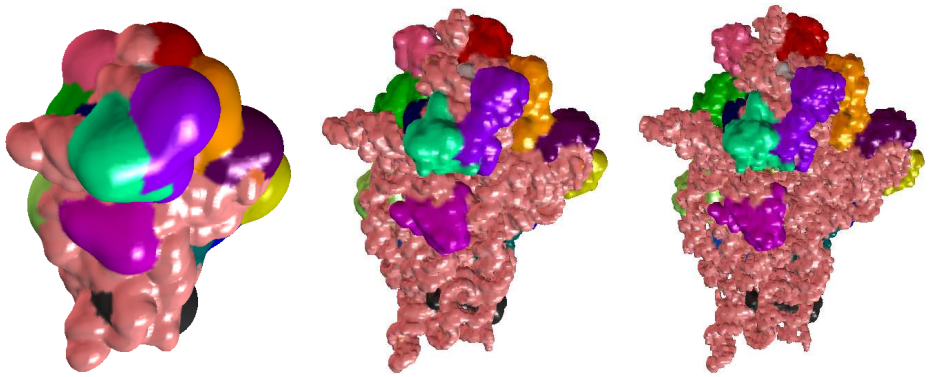
The summation of Gaussian kernel functions is used to construct the density map of a biomolecule and sampled volumetric data. Implicit solvation models can be constructed at multiple scales by choosing various Gaussian kernel parameters. The correspondence between different parameters is addressed in the error analysis.

The dual contouring method [12], [32], [33] is selected for mesh generation as it tends to yield meshes with better aspect ratio. In order to generate exterior meshes, we add a sphere or box outside the biomolecular surface as an outer

*The work at University of Texas was supported in part by NSF grants INT-9987409, ACI-0220037, EIA-0325550 and a grant from the NIH 0P20 RR020647. The work on this paper was done when Prof. Guoliang Xu was visiting Prof. Chandrajit Bajaj at UT-CVC and UT-ICES. His work was partially supported by the aforementioned grants, the J.T. Oden ICES fellowship and in part by NSFC grant 10371130, National 973 Project, 2004CB318006. Please visit http://www.ices.utexas.edu/∼jessica/biomesh



(a) C = -0.03125 (b) C = -0.125 (c) C = -0.25 (d) C = -0.5



(e) B = -0.03125 (f) B = -0.125 (g) B = -0.5

Figure 1: Implicit solvation models of Thermus Thermophilus small Ribosome 30S (1J5E) crystal subunit for various Gaussian kernel parameters. The pink color shows 16S rRNA and the remaining colors are proteins.

boundary. A variant of the dual contouring method is developed to extract interior and exterior meshes. Our tetrahedral mesh is spatially adaptive and attempts to preserve molecular surface features while minimizing the number of elements. An extension step is performed to generate the exterior mesh.

The extracted triangular and tetrahedral meshes can not be directly used for finite element calculations, they need to be modified and improved. Since the isosurface generated from discrete volumetric data suffers from noise, geometric flows are used to smooth the generated surface meshes with feature preserved. The quality of extracted surface and volume meshes is also improved.

The main contributions of this paper include: a simple and uniform treatment for approximating implicit solvation models, a modified adaptive surface and volume mesh extraction scheme combined with geometric flows and therefore yields high quality meshes. The generated meshes of the monomeric and tetrameric mAChE have been successfully used in solving the steady-state Smoluchowski equation using adaptive finite element methods [26], [27], [31].

The remainder of this paper is organized as the following: Section 2 reviews related previous work. Section 3 introduces how to construct implicit solvation surface from PDB format data. Section 4 explains mesh generation schemes. Section 5 discusses surface smoothing and quality improvement. Section 6 presents results and conclusion.

2. PREVIOUS WORK

Molecular Modeling: There are three different approximations of molecular surfaces or interfaces [22], the van der Waals surface (VWS), the solvent-accessible surface (SAS) and the solvent-excluded surface (SES) or the Lee-Richards surface [15]. The VWS is simply the boundary of the union of balls. As introduced in [15], the SAS is an inflated VWS with a probe sphere. The SES is a surface inside of which the probe never intrudes. According to the properties of molecular structures, Laug and Borouchaki used a combined advancing-front and generalized-Delaunay approach to mesh molecular surfaces [14], Algorithms were developed for sampling and triangulating a smooth surface with correct topology [1] [5], and Cheng et. al [4] maintained an approximating triangulation of a deforming skin surface. Simplex subdivision schemes are used to generate tetrahedral meshes for molecular structures in solving the Poisson-Boltzmann equation [11]. However, it still remains a challenging problem to generate triangular and tetrahedral meshes for large complicated molecular structures.

Mesh Generation: As reviewed in [21], [28], octree-based, advancing front based and Delaunay like techniques were used for triangular and tetrahedral mesh generation. The octree technique recursively subdivides the cube containing the geometric model until the desired resolution is reached [24]. Advancing front methods start from a boundary and move a front from the boundary towards empty space within the domain [17]. Delaunay refinement is to refine the triangles or tetrahedra locally by inserting new nodes to maintain the Delaunay criterion [6], [25].

The predominant algorithm for isosurface extraction from volume data is Marching Cubes (MC) [18], which computes a local triangulation within each cube to approximate the isosurface by using a case table of edge intersections. MC was extended to extract tetrahedral meshes between two isosurfaces [7]. A different and systematic algorithm was proposed for interval volume tetrahedralization [20]. By combining SurfaceNets [9] and the extended Marching Cubes

algorithm [13], octree based Dual Contouring [12] generates adaptive multiresolution isosurfaces with good aspect ratio and preserve sharp features. The dual contouring method has already been extended to extract adaptive and quality tetrahedral meshes from volumetric imaging data [32], [33].

Quality Improvement: Algorithms for mesh improvement can be classified into three categories [28], [21]: local coarsening/refinement by inserting/deleting points, local remeshing by face/edge swapping and mesh smoothing by relocating vertices. Laplacian smoothing and optimization techniques are developed to improve the mesh quality. The Laplacian operator was discretized [19], and geometric flows have been used in surface and imaging processing [23], [30].

3. IMPLICIT SOLVATION SURFACE FROM VOLUMETRIC DENSITY MAPS

In this section, we extract the implicit solvation surface (molecular surface) as a level set from volumetric synthetic electron density maps. The implicit solvation surface can approximate different Lee-Richards molecular surfaces [15] with various Gaussian kernel parameters .

As used for Poisson-Boltzmann electrostatics calculations in [11], a characteristic function $f(x)$ is selected to represent an 'inflated' van der Waals-based accessibility

$$f(x) = \begin{cases} 1, & \text{if } \|x - x_i\| < r_i + \sigma \text{ for } i = 1, \ldots, N, \\ 0, & \text{otherwise,} \end{cases} \quad (1)$$

where (x_i, r_i) are the centers and radii of the N atoms in the biomolecule, and σ is the radius of the diffusing species, here we choose $\sigma = 2$ [27]. When $\sigma = 0$, the VWS is constructed. The function $f(x)$ provides a grid-based volumetric data which can be isocontoured at the isovalue 0.5 to represent the SAS. Fig. 3(a) shows one constructed geometric model of mAChE.

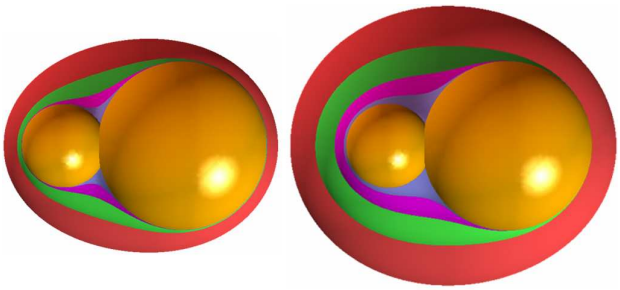


Figure 2: Multi-scale models by choosing various C in (a) and B_i in (b). Yellow balls are two input atoms. The correspondence between C/B_i values and multi-scale models are shown in Table 1.

Table 1: C/B_i and Multi-scale Models in Fig. 2

	Red	Green	Magenta	Blue
Fig.2(a)	C = -0.125	C = -0.25	C = -0.5	C = -1.0
Fig.2(b)	B_i= -0.125	B_i= -0.25	B_i= -0.5	B_i= -1.0

Molecules are often modelled as the union of hard spheres S_i (atoms). The surface, denoted as M_0, of a molecule is therefore described as the boundary of the union of the balls. To have the blurring effect at the intersection parts of atoms, the molecular surface is approximated by the isocontour [3]:

$$M := \{x \in \mathbb{R}^3 : G(x) = 1\} \quad (2)$$

with

$$G(x) = \sum_{i=1}^{N} e^{B_i\left(\frac{\|x-x_i\|^2}{r_i^2} - 1\right)}, \quad (3)$$

where (x_i, r_i) are the center and radius of the ith atom in the biomolecule, and $B_i < 0$ is called 'blobbiness', which controls the blurring effect. Note that B_i must be negative to ensure that the density function goes to zero as $\| x - x_i \|$ goes to infinity. To have the distance between M and M_0 as uniformly as possible, we take

$$C = B_i / r_i^2, \quad (4)$$

where $C < 0$ is a given constant. Now $G(x)$ becomes

$$G(x, C) = \sum_{i=1}^{N} e^{C(\|x-x_i\|^2 - r_i^2)}. \quad (5)$$

The multi-scale presentation $M(C_i) = \{x \in \mathbb{R}^3 : G(x, C_i) = 1\}$ of the molecular surface is therefore achieved by taking $C = C_1, \ldots, C_l$. As shown in Fig. 2, the different effects of C and constant $B_i (= B)$ are studied in a two-sphere system, one is centering at (0, 0, 0) with the radius of 1.0, the other one is at (2.8, 0, 0) with the radius of 2.0. It could be observed that

- C leads to more uniform inflation than B_i.
- Small balls have more inflation than big ones.
- Large error happens around the intersection region, and error reduces gradually away from it.
- Larger C and B_i lead to more inflation. For the same C and B_i value, e.g., -0.125, B_i tends to introduce more inflation.

In order to correspond $M(C_i)$ to the VWS and the SAS by choosing various C_i, we need to relate C with σ. Of course, the inflated radial distance by $M(C)$ is not uniform. It is thick around the intersection part of the atom surfaces, and becomes gradually thinner away from the intersection region. Therefore, we define three metrics to measure the global radial distance between $M(C)$ and M_0:

$$e_{min}(M(C)) = \min_{q \in M_0} \| q - P_{M(C)}(q) \|, \quad (6)$$

$$e_{max}(M(C)) = \max_{q \in M_0} \| q - P_{M(C)}(q) \|, \quad (7)$$

$$e_{mean}(M(C)) = \frac{\int_{\overline{N}_0} \| q - P_{M(C)}(q) \| \, dq}{Area(\overline{N}_0)}, \quad (8)$$

where $P_{M(C)}(q)$ is the projection of $q \in \overline{N}_0 \subset M_0$ on the surface $M(C)$ in the normal direction $n(q)$ (the nearest intersection point of the line $q + n(q)t$ with $M(C)$), $N_0 \subset M_0$ is defined as

$$N_0 = \left\{ q \in M_0 : P_{M(C)}^{-1} \text{ exists at } P_{M(C)}(q) \in M(C) \right\}. \quad (9)$$

To make $P_{M(C)}^{-1}$ meaningful, we first classify the points of $M(C)$ according to the sphere S_i. Let $x \in M(C)$, if

$$\|x - x_i\|^2 - r_i^2 \leq \|x - x_j\|^2 - r_j^2, \quad \forall j \neq i, \quad (10)$$

then we say x belongs to S_i. If the inequality (10) is strict, then we say x belongs to S_i properly. It is easy to see that (10) holds if and only if

$$e^{C(\|x-x_i\|^2 - r_i^2)} \geq e^{C(\|x-x_j\|^2 - r_j^2)}, \quad \forall j \neq i.$$

Hence the point classification is achieved from the Gaussian function.

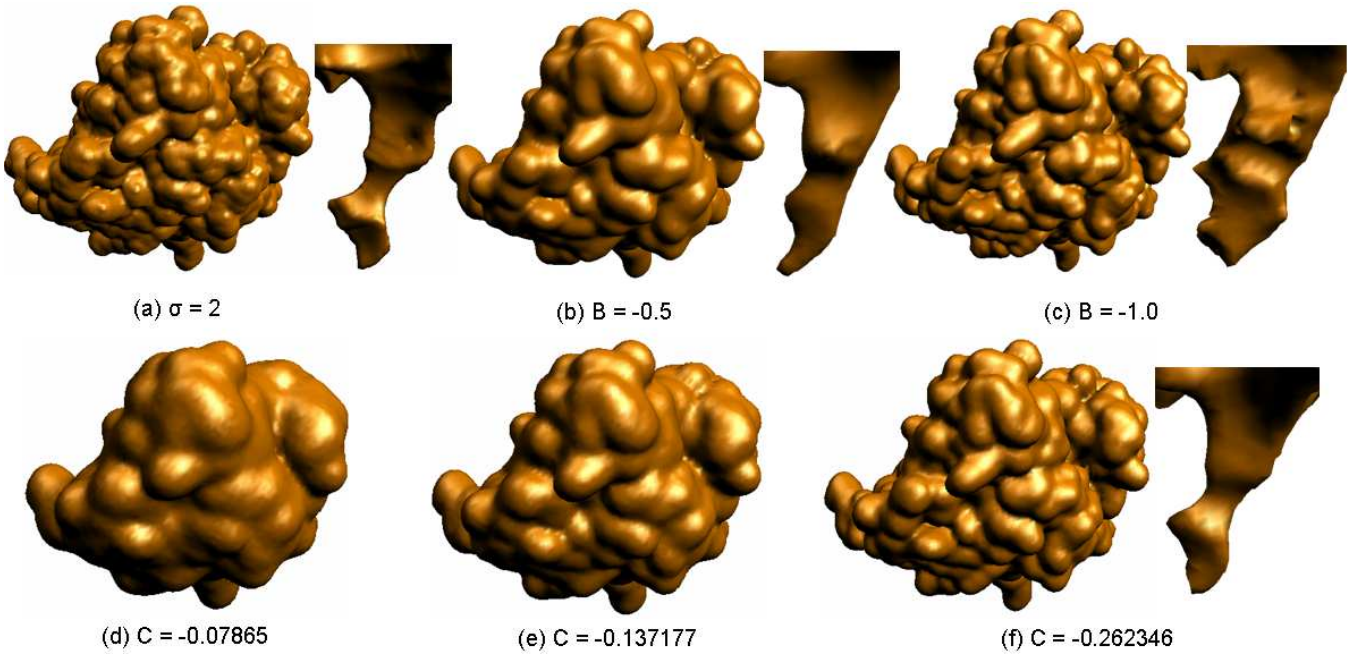
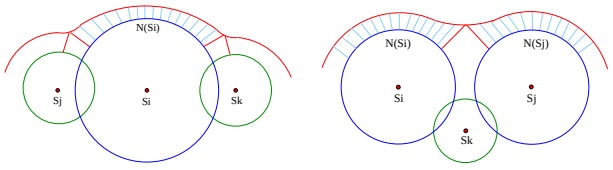


Figure 3: Implicit solvation models of mAChE for σ=2, various Bs and Cs. mAChE has a cavity as shown in (a). Cavity vanishes in (d) and (e). (a), (b), (c) and (f) show multi-scale models and cavities.



(a) (b)

Figure 4: (a) Definition of $N(S_i)$; (b) $N(S_k)$ is an empty set.

If a point $x \in M(C)$ belongs to S_i properly, then $P^{-1}_{M(C)}(x)$ exists. Therefore, N_0 is defined as the union of open subset of sphere S_i:

$$N_0 = \bigcup_{i=1}^{N} N(S_i), \tag{11}$$

where

$$N(S_i) = \left\{ q \in S_i : P_{M(C)}(q) \ \textit{belongs to} \ S_i \ \textit{properly} \right\}. \tag{12}$$

Fig. 4(a) shows the definition of $N(S_i)$ for the sphere S_i. In Fig. 4(b), the sphere S_k has empty $N(S_k)$.

Now our problem becomes: given a $\sigma > 0$, find a constant C such that

$$e(M(C)) = \sigma \tag{13}$$

where e represents one of the three metrics, $e_{min}(M(C))$, $e_{max}(M(C))$ or $e_{mean}(M(C))$. We denote the solutions of Eqn. (13) by C_{min}, C_{max} and C_{mean} respectively. It is easy to see that $M(C_{min})$ is the minimal surface that contains the σ-inflated molecule, $M(C_{max})$ is the maximum surface that is inside the σ-inflated molecule, while $M(C_{mean})$ is between $M(C_{min})$ and $M(C_{max})$. By choosing different Cs in the range $[C_{min}, C_{max}]$, we will achieve different features of the inflated molecule. Basically, smaller C captures global features, and larger C captures local features.

Now we consider the computation of $e(M(C))$. First we assume that each atom (sphere) is uniformly triangulated with m vertices. This triangulation is achieved by translating a triangulated unit sphere to each of the atom center and re-scaling it to the atom size. We obtain the unit sphere triangulation from [29]. For each vertex q on the triangulated atom, $P_{M(C)}(q)$ is computed. The integration in Eqn. (8) is computed as the summation of $\| q - P_{M(C)} \| 4\pi r^2/m$, where r is the radius of the atom considered.

$P_{M(C)}(q)$ is computed as the intersection of the line $q + n(q)t$ with the surface $M(C)$. This is computed by Newton iteration method. Let $P_{M(C)}(q) = q + t^\star n$. Then $t^\star = \| q - P_{M(C)}(q) \|$. The Newton iteration is as follows:

$$t_{k+1} = t_k - \frac{G(q + t_k n(q))}{n^T(q) \nabla G(q + t_k n(q))}. \tag{14}$$

The main cost in this iteration is computing G and ∇G.

It is difficult to find the exact solution of Eqn. (13) for unknown C, here we choose an approximating approach. First, we select a sequence of $C = C_1, \ldots, C_l$ satisfying $C_{k+1} = 2C_k$, for example $C_1 = -0.0625$, $C_2 = -0.125$, $C_3 = -0.25$, ... Then compute $e(M(C_i))$. $e(M(C))$ decreases as $C \to -\infty$. Now given a $\sigma \in [e(M(C_{i+1})), e(M(C_i))]$, we can use the cubic inverse interpolation to compute a C.

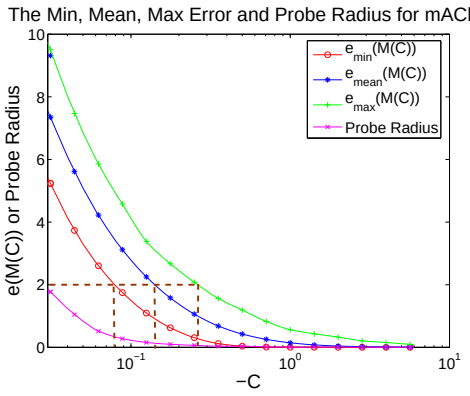
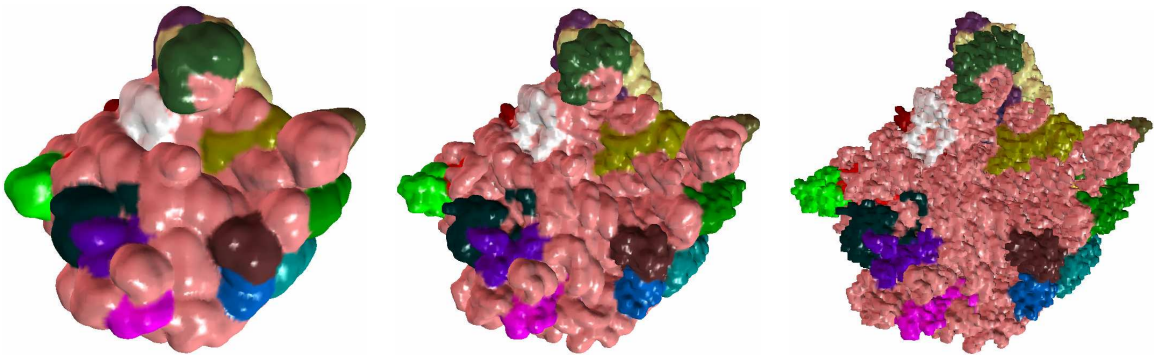


Figure 5: Error and probe radius for mAChE.



(a) C = -0.03125 (b) C = -0.125 (c) C = -0.5

Figure 6: Implicit solvation models of Haloarcula Marismortui large Ribosome 50S (1JJ2) crystal subunit. The light yellow and the pink color show 5S and 23S rRNA respectively, the remaining colors are proteins.

Fig. 5 shows the three error functions for mAChE. When $\sigma = 2$, we can obtain $C_{min} = -0.078650$, $C_{mean} = -0.137177$ and $C_{max} = -0.262346$. Fig. 5 also shows a curve of the probe radius, defined as the maximal radius of sphere which can touch the surface everywhere. Let $H(x) \in \mathbb{R}^3$ be the mean curvature normal at a surface point x. First the principal curvatures $k_1(x)$, $k_2(x)$ are computed with $H^T(x)\nabla G(x) \leq 0$, then the probe radius is $1/\max_x k(x)$, where $k(x) = \max\{|k_1(x)|, |k_2(x)|\}$. From the multi-scale geometric models of mAChE as shown in Fig. 3, we can observe that

- The mAChE structure has an active site, the cavity. It is important to construct a good approximation for the molecule, especially for the cavity.
- Since B leads to non-uniform inflation as shown in the two-sphere system (Fig. 2), no good approximation can be obtained, especially for local features such as the cavity (see Fig. 3(b∼c)).
- When $C = C_{min}$ or $C = C_{mean}$, there is too much inflation and the cavity vanishes. When $C = C_{max}$, the constructed geometric model as shown in Fig. 3(f) provides a good approximation to Fig. 3(a).
- Given a σ or the radius of a probe sphere, we can find a C corresponding to it in Fig. 5. Therefore we can approximate the three surfaces (VWS, SAS and SES).

Fig. 1 shows multi-scale models of Ribosome 30S. Compared with Fig. 1(a), proteins inflates much more seriously in Fig. 1(e). rRNA in Fig. 1(c) and (f) looks similar, but proteins in Fig. 1(f) look a little more inflated than Fig. 1(b). rRNA in Fig. 1(d) and (g) looks similar too, but proteins in Fig. 1(g) are close to proteins in Fig. 1(c).

In our multi-scale modeling, the scale is controlled by the parameter $C \in (-\infty, 0)$ in the Gaussian map. Mathematically, there are infinitely many scales. To make the constructed geometric model correspond to a certain level, such as the residual level, C needs to be selected properly. For a fixed level, the structure at this level should be distinguishable, while the structure at a higher level may not be presented remarkably. For instance, at the residual level, the chain structure of residuals should be observed, while atoms may not be distinguished clearly. In Fig. 6, selected Cs make the constructed models represent the low resolution, residual and atomic levels respectively.

4. MESH GENERATION

There are two main methods for contouring scalar fields, primal contouring [18] and dual contouring [12]. Both of them can be extended to tetrahedral mesh generation. The dual contouring method [32], [33] is often the method of choice as it tends to yield meshes with better aspect ratio.

4.1 Triangular Meshing

Dual contouring [12] uses an octree data structure, and analyzes those edges that have endpoints lying on different sides of the isosurface, called *sign change edges*. The mesh adaptivity is determined during a top-down octree construction. Each sign change edge is shared by either four (uniform case) or three (adaptive case) cells, and one minimizer is calculated for each of them by minimizing a predefined Quadratic Error Function (QEF) [8]:

$$QEF[x] = \sum_i \left[n_i \cdot (x - p_i)\right]^2, \tag{15}$$

where p_i, n_i represent the position and unit normal vectors of the intersection point respectively. For each sign change edge, a quad or triangle is constructed by connecting the minimizers. These quads and triangles provide a 'dual' approximation of the isosurface.

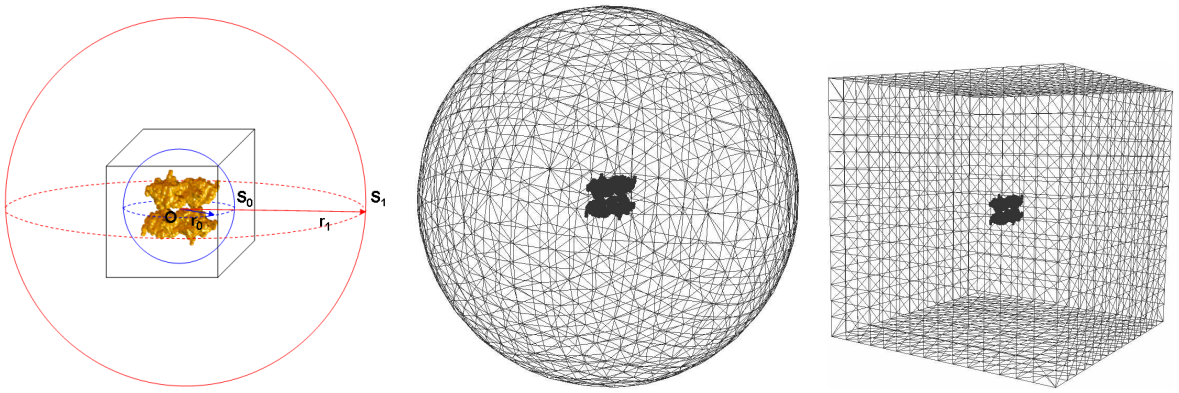
A recursive cell subdivision process was used to preserve the correct topology [33] of the isosurface. During the cell subdivision, the function values at each newly inserted grid point can be exactly calculated since we know the function (Gaussian functions, Eqn. (3)). Additionally, we can generate a more accurate triangular mesh by projecting each generated minimizer onto the isosurface (Eqn. (2)).

4.2 Tetrahedral Meshing

The dual contouring method has already been extended to extract tetrahedral meshes from volumetric scalar fields [32], [33]. The cells containing the isosurface are called boundary cells, and the interior cells are those cells whose eight vertices are inside the isosurface. In the tetrahedral mesh extraction process, all the boundary cells and the interior cells need to be analyzed in the octree data structure. There are two kinds of edges in boundary cells, one is a sign change edge, the other is an interior edge. Interior cells only have interior edges. In [32], [33], interior edges and interior faces in boundary cells are dealt with in a special way, and the volume inside boundary cells is tetrahedralized. For interior cells, we only need to split them into tetrahedra.

4.2.1 Adding an Outer Boundary

In the biological diffusion system, we need to analyze the field which is from infinite faraway to the molecular sur-



(a) (b) (c)

Figure 7: The analysis domain of exterior meshes. (a) - 'O' is the geometric center of the molecule; the circum-sphere of the biomolecule has the radius of r; the box represents the volumetric data; 'S_0' is the maximum sphere inside the box, the radius is $r_0(r_0 > r)$; 'S_1' is an outer sphere with the radius of $r_1(r_1 = (20 \sim 40)r)$. (b) - the diffusion domain is the interval volume between the biomolecular surface and the outer sphere 'S_1', here we choose $r_1 = 5r$ for visualization. (c) - the outer boundary is a cubic box.

face. Assume that the radius of the circum-sphere of a biomolecule is r. The computational model can be approximated by a field from an outer sphere S_1 with the radius of $(20 \sim 40)r$ to the molecular surface. Therefore the exterior mesh is defined as the tetrahedralization of the interval volume between the molecular surface and the outer sphere S_1 (Fig. 7(b)). Sometimes the outer boundary is chosen to be a cubic box as shown in Fig. 7(c).

First we add a sphere S_0 with the radius of r_0 (where $r_0 > r$ and $r_0 = 2^n/2 = 2^{n-1}$) outside the molecular surface, and generate meshes between the molecular surface and the outer sphere S_0. Then we extend the tetrahedral meshes from the sphere S_0 to the outer bounding sphere S_1. For each data point inside the molecular surface, we keep the original function value. While for each data point outside the molecular surface, we reset the function value as the smaller one of $f(x) - \alpha$ and the shortest distance from the grid point to the sphere S_0. Eqn. (16) shows the newly constructed function $g(x)$ which provides a grid-based volumetric data containing the biomolecular surface and an outer sphere S_0.

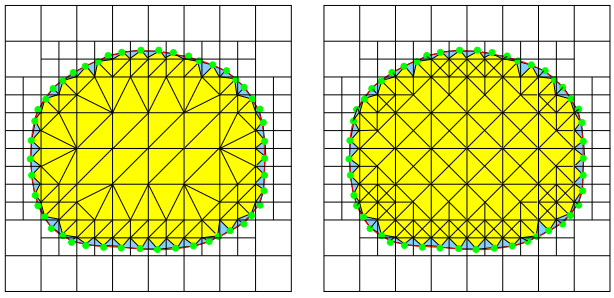
$$g(x)=\begin{cases} \min(\|x-x_0\|-r_0, f(x)-\alpha), & \text{if } f(x)<\alpha, \|x-x_0\|<r_0, \\ \|x-x_0\|-r_0, & \text{if } f(x)<\alpha, \|x-x_0\|\geq r_0, \\ f(x)-\alpha, & \text{if } f(x)\geq\alpha, \end{cases} \tag{16}$$

where x_0 are coordinates of the molecular geometric center. The isovalue $\alpha = 0.5$ for volumetric data generated from the characteristic function, and $\alpha = 1.0$ for volumetric data generated from the summation of Gaussian kernels.

The biomolecular surface and the outer sphere S_0 can be extracted as an isosurface at the isovalue 0, $S_g(0) = \{x|g(x) = 0\}$. All the grid points inside the interval volume $I_g(0) = \{x|g(x) \leq 0\}$ have negative function values, and all the grid points outside it have positive values.

4.2.2 Primal Mesh Extraction

Here we introduce a different algorithm, in which we do not distinguish boundary cells and interior cells when we analyze edges. We only consider two kinds of edges - sign change edges and interior edges. For each boundary cell, we can obtain a minimizer point by minimizing its Quadratic Error Function. For each interior cell, we set the middle point of the cell as its minimizer point. Fig. 8(b) shows a simple 2D example. In 2D, there are two cells sharing each edge, and two minimizer points are obtained. For each sign change edge, the two minimizers and the interior vertex of this edge construct a triangle (blue triangles). For each interior edge, each minimizer point and this edge construct a triangle (yellow triangles). In 3D as shown in Fig. 9, there are three or four cells sharing each edge. Therefore, the three (or four) minimizers and the interior vertex of the sign change edge construct one (or two) tetrahedron, while the three (or four) minimizers and the interior edge construct two (or four) tetrahedra.



(a) (b)

Figure 8: 2D triangulation. (a) Old scheme, (b) New scheme. Blue and yellow triangles are generated for sign change edges and interior edges respectively. The red curve represents the molecular surface, and the green points represent minimizer points.

Compared with the algorithm presented in [32], [33] as shown in Fig. 8(a), Fig. 8(b) generates the same surface meshes, and tends to generate more regular interior meshes with better aspect ratio, but a little more elements for interior cells. Fig. 8(b) can be easily extended to large volume decomposition. For arbitrary large volume data, it is difficult to import all the data into memory at the same time. So we first divide the large volume data into some small subvolumes, then mesh each subvolume separately. For those sign change edges and interior edges lying on the interfaces between subvolumes, we analyze them separately. Finally, the generated meshes are merged together to obtain the desired mesh. The mesh adaptivity is controlled by the structural properties of biomolecules. The extracted tetrahedral mesh is finer around the molecular surface, and gradually gets coarser from the molecular surface out towards the outer sphere, S_0. Furthermore, we generate the finest mesh around the active site, such as the cavity in the monomeric

and tetrameric mAChE as shown in Fig. 14 (a∼b), and a coarse mesh everywhere else.

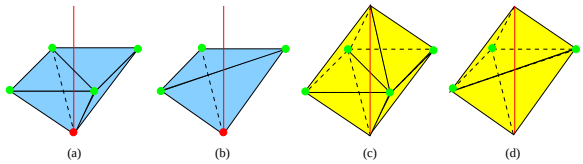


Figure 9: Sign change edges and interior edges are analyzed in 3D tetrahedralization. (a)(b) - sign change edge (the red edge); (c)(d) - interior edge (the red edge). The green solid points represent minimizer points, and the red solid points represent the interior vertex of the sign change edge.

4.2.3 Mesh Extension

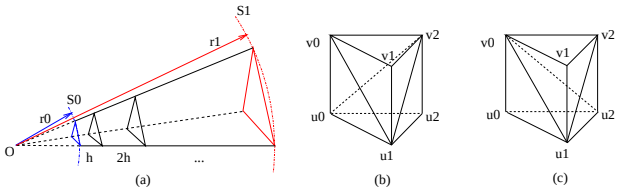


Figure 10: (a) - one triangle in the sphere S_0 (blue) is extended n steps until arriving the sphere S_1 (red); (b) and (c) - a prism is decomposed into three tetrahedra in two different ways.

We have generated meshes between the biomolecular surface and the outer sphere S_0, the next step is to construct tetrahedral meshes gradually from the sphere S_0 to the bounding sphere S_1 (Fig. 7). The sphere S_0 consists of triangles, so we extend each triangle radially as shown in Fig. 10 and a prism is obtained for each extending step. The prism can be divided into three tetrahedra. The extension step length h can be calculated by Eqn. (17). It is better for the sphere S_0 to be triangulated uniformly since the step length is fixed for each extending step.

$$r_0 + h + 2h + \cdots + nh = r_1 \quad \Longrightarrow \quad h = \frac{2(r_1 - r_0)}{n(n+1)} \tag{17}$$

where n is the step number. In Figure 10, suppose $u_0u_1u_2$ is a triangle on sphere S_0, and u_0, u_1, u_2 are the unique index numbers of the three vertices. For one extension step, $u_0u_1u_2$ is extended to $v_0v_1v_2$, and the two triangles construct a prism, which can be decomposed into three tetrahedra. In order to avoid the diagonal conflict problem, a different decomposition method (Fig. 10 (b∼c)) is chosen based on the index number of the three vertices.

Assume there are m triangles on the sphere S_0, which is extended n steps to arrive the sphere S_1. m prisms or $3m$ tetrahedra are generated in each extending step, and totally $3mn$ tetrahedra are constructed in the extension process. Therefore, it is better to keep coarse and uniform triangular mesh on the sphere S_0.

5. QUALITY IMPROVEMENT

In general, the molecular skin generated by iscontouring the Gaussian density function or the characteristic function is bumpy. This is because the volume data could not be sufficiently fine due to the capacity limit of the computer, and is not smooth enough, especially for the data generated from the characteristic function. The error of the isosurface from the characteristic function could be as large as half of the grid size, and could be very large relative to the atom size. Therefore, a post-processing step for the extracted isosurface is necessary. There are three tasks for the mesh quality improvement:

1. Denoising the surface mesh (vertex adjustment in the normal direction).
2. Improving the aspect ratio of the surface mesh (vertex adjustment in the tangent direction).
3. Improving the aspect ratio of the volumetric mesh (vertex adjustment inside the volume).

We use geometric partial differential equations (PDEs) to handle the first two problems. Geometric PDEs, such as the mean curvature flow, the surface diffusion flow and Willmore flow, have been intensively used in surface and imaging processing (see [30] for references). Here we choose the surface diffusion flow to smooth the molecular skin.

$$\frac{\partial x}{\partial t} = \Delta H(x)\vec{n}(x), \tag{18}$$

where H is the mean curvature, $\vec{n}$ is the unit surface normal vector, and Δ is the Laplacian-Beltrami operator. This flow is area shrinking and volume preserving. Furthermore, it preserves sphere exactly and torus approximately. Suppose a molecular skin could be ideally represented by the joining of spherical and torus surface patches [14], it is desirable to use the surface diffusion flow to evolve the isosurface. However, this flow could only improve the surface shape, not the mesh regularity. In order to improve the aspect ratio, we need to add a tangent movement in Eqn. (18). Hence the flow becomes

$$\frac{\partial x}{\partial t} = \Delta H(x)\vec{n}(x) + v(x)\vec{T}(x), \tag{19}$$

where $v(x)$ is the velocity in the tangent direction $\vec{T}(x)$.

Eqn. (19) is solved over a triangular mesh with vertices $\{x_i\}$ by discretizing each of its terms. In the temporal space, $\frac{\partial x}{\partial t}$ is approximated by the Euler scheme $\frac{x_i^{n+1} - x_i^n}{\tau}$, where τ is time step-length. x_i^n is the approximating solution at $t = n\tau$, where $x_i^0 = x_i$ serves as the initial value. Discretizing schemes for Δ and H in the spatial space are given in [30], we do not go to detail here. $v(x)\vec{T}(x)$ is approximated by

$$[m(x_i^n) - x_i^n] - \vec{n}(x_i^n)[m(x_i^n) - x_i^n]^T\vec{n}(x_i^n), \tag{20}$$

where $m(x_i^n)$ is defined as the mass center of all the triangles around x_i^n. A mass center P of a region V is defined by finding $p \in V$, such that

$$\int_V \| y - p \|^2 \, d\sigma = min. \tag{21}$$

V could be a piece of surface or a volume in $\mathbb{R}^3$. For our surface mesh case, V consists of triangles around vertex x_i^n. Then from Eqn. (21), we could derive that

$$m(x_i^n) = \frac{1}{3}x_i^n + \frac{1}{3}\sum_{j \in N(i)} x_j^n(\triangle_j + \triangle_{j+1})/A(x_i^m), \tag{22}$$

where $N(i)$ is the index set of the one ring neighbors of x_i^m. $\triangle_j$ is the area of the triangle $[x_i^n x_{j-1}^n x_j^n]$. $A(x_i^n)$ is the total of triangle areas.

Usually, people use the geometric center [30], instead of the mass center. But we found that the mass center works better. The discretization will lead to a linear system. The approximated solution is obtained by solving it.

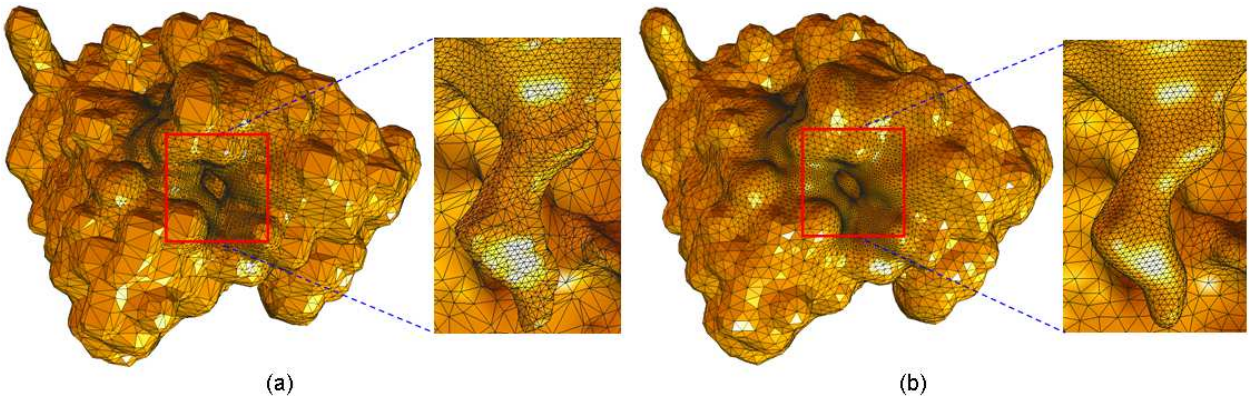


Figure 11: Comparison of mAChE (9308 vertices, 18612 triangles) before and after surface smoothing. (a) - original; (b) - after smoothing.

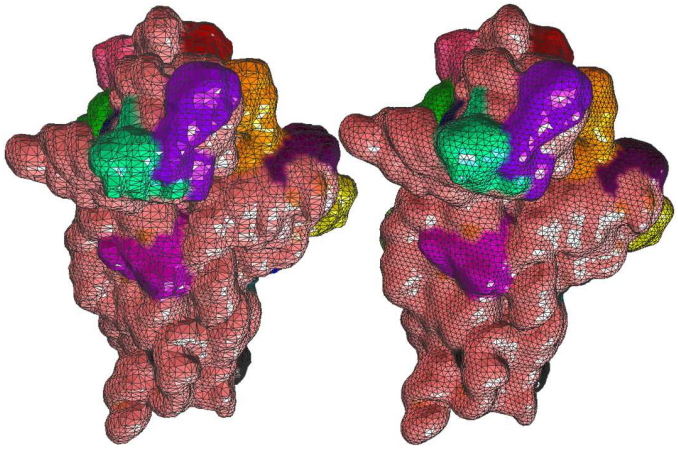


Figure 12: Comparison of Ribosome 30S (13428 vertices, 26852 triangles) before and after surface smoothing. Left - original; Right - after smoothing.

After the molecular skin is smoothed and regularized, the next step is to improve the volumetric mesh by relocating each interior vertex to the mass center of its surrounding tetrahedra. Let p_i be an interior vertex, p_j be one of its neighboring vertices, then the mass center of all tetrahedra around p_i is computed by

$$m(p_i) = \frac{1}{4}p_i + \frac{1}{4V_i}\sum_j V_{ij}p_j, \tag{23}$$

where V_{ij} is the volume summation of all the tetrahedra around the edge $[p_ip_j]$, V_i is the volume summation of the tetrahedra around the vertex p_i.

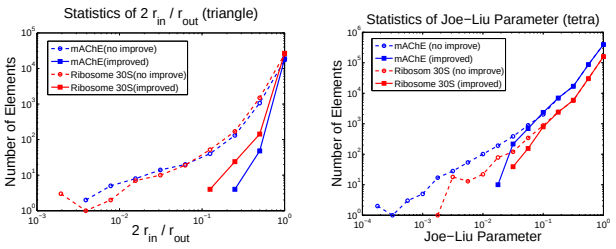


Figure 13: The histogram of the aspect-ratio (left) and Joe-Liu parameter (right).

Here we choose the aspect ratio (twice of the ratio of incircle radius to circumcircle radius) to measure the quality of triangular meshes, and the surface diffusion flow to smooth the surface. The left picture in Fig. 13 shows the improvement of the aspect ratio, and Fig. 11∼12 show the improvement of molecular skins. We can see that noises are removed and features are preserved since the surface diffusion flow preserves volume and spherical geometry. The surface error is restricted within half of the grid size for the binary data from the characteristic function, and almost zero for the data from Gaussian density map since we have projected each boundary vertex onto the isosurface.

In [33], the edge contraction and linear averaging method was used to improve the quality of tetra meshes with the edge-ratio and Joe-Liu parameter as metrics. The goal is to improve the worst parameters in each iteration. Here we still use the same edge contraction scheme, but relocate each interior vertex to its mass center (Eqn.(23)) since it can minimize the energy defined in Eqn.(21). From the right picture in Fig.13, we can see that the worst Joe-Liu parameter is improved significantly after quality improvement. Fig.14(a), (c) show the two interior tetra meshes, mAChE ($C = -0.262346$) and Ribosome 30S ($C = -0.03125$).

6. RESULTS AND CONCLUSION

Monomeric mAChE: For efficient and accurate finite element calculations, adaptive meshes are preferred. Therefore we generated finer meshes around the cavity region since the bottom of the gorge is the active site in mAChE, while coarser meshes in other regions. The extracted tetrahedral meshes of the monomer as shown in Fig. 14(a) have been used as the geometric model in the finite element analysis of the steady-state Smoluchowski equation (SSSE) for rate constant calculations [26], [27]. The calculated rates showed good agreement with experimental results. Our generated surface mesh is being used in calculating the electrostatic potential distribution of biomolecules.

Tetrameric mAChE: We also applied our approach to generate tetrahedral meshes for the tetramer, which has two different arrangement formats from four monomers according to previous crystallographic studies. Each monomer has an active site accessible though a long gorge (20 Angstrom), so there are a total of four gorges. Fig. 14(b) shows the two crystal structures. In the first crystal structure, two gorges are partially blocked, while the other two are completely accessible to solvent. In the second one, all the four gorges are open. Similarly, we generated fine meshes around the region of the four gorges and coarse meshes in other regions for finite element simulations [31].

Ribosome: Ribosomes are macromolecular complexes responsible for the translation of mRNA into protein. These complexes consist of two subunits: the large 50S and the small 30S, both of them are composed of rRNA and protein constituents. Atomic level, residual level and low resolution structures were constructed from density maps as shown in Fig. 1 and 6. Multi-scale structures help to study the machinery of protein production. Fig. 14(c∼d) show interior and exterior meshes of Ribosome 30S/50S.

We have developed a quality molecular meshing approach from PDB data, including implicit solvation surface construction from multi-scale density maps, triangular/tetrahedral mesh generation and quality improvement with surface smoothing. Geometric features are preserved for the molecular surface. Some of our generated meshes have been used or is being used in boundary/finite element simulations.

Acknowledgments

We wish to thank Vinay Siddavanahalli and John Wiggins for all their help with the generation of smooth volumetric electron density maps using CVC's fast summation code for biomolecular structures, and also thank Dr. Yuhua Song, Dr. Deqiang Zhang, Prof. Nathan Baker and Prof. Andrew McCammon for useful discussions.

7. REFERENCES

[1] N. Akkiraju and H. Edelsbrunner. Triangulating the surface of a molecule. *Discr.Appl.Math.*, 71:5–22, 1996.

[2] C. Bajaj, J. Castrillon, V. Siddavanahalli, and W. Zhao. Fast calculations of implicit solvation models of molecules and solvation energies. *Manuscript*, 2004.

[3] J. Blinn. A generalization of algebraic surface drawing. *ACM Transactions on Graphics*, 1(3):235–256, 1982.

[4] H.-L. Cheng, T. Dey, H. Edelsbrunner, and J. Sullivan. Dynamic skin triangulation. *Discr. Comput. Geom.*, 25:525–568, 2001.

[5] S.-W. Cheng, T. Dey, E. Ramos, and T. Ray. Sampling and meshing a surface with guaranteed topology and geometry. In *SoCG*, pages 280–289, 2004.

[6] P. Chew. Guaranteed-quality delaunay meshing in 3d. In *SOCG*, pages 391–393, 1997.

[7] I. Fujishiro, Y. Maeda, H. Sato, and Y. Takeshima. Volumetric data exploration using interval volume. *IEEE Transactions on Visualization and Computer Graphics*, 2(2):144–155, 1996.

[8] M. Garland and P. Heckbert. Simplifying surfaces with color and texture using quadric error metrics. In *IEEE Visualization*, pages 263–270, 1998.

[9] S. Gibson. Using distance maps for accurate surface representation in sampled volumes. In *IEEE sympos. on Volume visualization*, pages 23–30, 1998.

[10] J. Grant and B. Pickup. A gaussian description of molecular shape. *Journal of Phys. Chem.*, 99:3503–3510, 1995.

[11] M. Holst, N. Baker, and F. Wang. Adaptive multilevel finite element solution of the poisson-boltzmann equation i: Algorithms and examples. *J. of Comput. Chem.*, 21:1319–1342, 2000.

[12] T. Ju, F. Losasso, S. Schaefer, and J. Warren. Dual contouring of hermite data. In *SIGGRAPH*, pages 339–346, 2002.

[13] L. Kobbelt, M. Botsch, U. Schwanecke, and H. Seidel. Feature sensitive surface extraction from volume data. In *SIGGRAPH*, pages 57–66, 2001.

[14] P. Laug and H. Borouchaki. Molecular surface modeling and meshing. *Engin. with Computers*, 18:199–210, 2002.

[15] B. Lee and F. Richards. The interpretation of protein structures: estimation of static accessibility. *J. Mol. Biol.*, 55:379–400, 1971.

[16] M. Lee, M. Feig, F. Salsbury, and C. Brooks. New analytic approximation to standard molecular volume definition and its application to generalized born calculation. *J. Comput. Chem.*, 24:1348–1356, 2003.

[17] R. Lohner and P. Parikh. Three dimensional grid generation by the advancing-front method. *Internat. J. for Numer. Methods in Fluids*, 8:1135–1149, 1988.

[18] W. Lorensen and H. Cline. Marching cubes: A high resolution 3d surface construction algorithm. In *SIGGRAPH*, pages 163–169, 1987.

[19] M. Meyer, M. Desbrun, P. Schroder, and A. Burr. Discrete differential-geometry operators for triangulated 2-manifolds. *VisMath'02, Berlin*, 2002.

[20] G. Nielson and J. Sung. Interval volume tetrahedrization. In *IEEE Vis' 97*, pages 221–228.

[21] S. Owen. A survey of unstructured mesh generation technology. In *7th Internat. Meshing Roundtable*, pages 26–28, 1998.

[22] M. Sanner, A. Olson, and J. Spehner. Reduced surface: An efficient way to compute molecular surfaces. *Biopolymers*, 38:305–320, 1996.

[23] G. Sapiro. Geometric partial differential equations and imaging analysis. *Cambridge Univ. Press*, 2001.

[24] M. Shephard and M. Georges. Three-dimensional mesh generation by finite octree technique. *Internat. J. for Numer. Methods in Engin.*, 32:709–749, 1991.

[25] J. Shewchuk. Tetrahedral mesh generation by delaunay refinement. In *SOCG*, pages 86–95, 1998.

[26] Y. Song, Y. Zhang, C. Bajaj, and N. Baker. Continuum diffusion reaction rate calculations of wild type and mutant mouse acetylcholinesterase: Adaptive finite element analysis. *Biophysical J.*, 87(3):1558–1566, 2004.

[27] Y. Song, Y. Zhang, T. Shen, C. Bajaj, J. McCammon, and N. Baker. Finite element solution of the steady-state smoluchowski equation for rate constant calculations. *Biophysical J.*, 86(4):2017–2029, 2004.

[28] S.-H. Teng and C. Wong. Unstructured mesh generation: theory, practice and perspectives. *Int. J. Comput. Geom. Appl.*, 10(3):227–266, 2000.

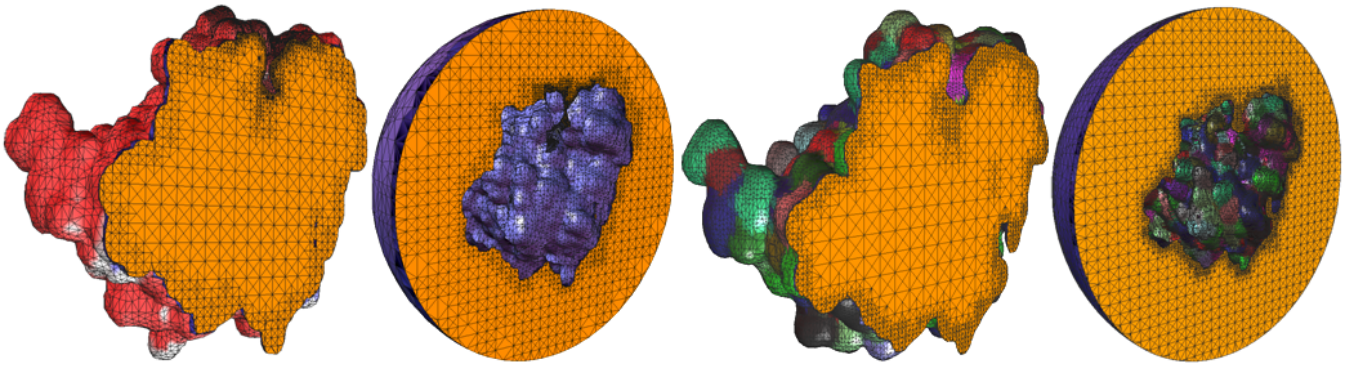
[29] G. Xu. Discrete laplace-beltrami operator on sphere and optimal spherical triangulations. *Research Report, Institute of Comput. Math. and Sci./Engin. Comput., Chinese Academy of Sciences, No. ICM-04-11*, 2004.

[30] G. Xu. Discrete laplace-beltrami operators and their convergence. *CAGD*, 21:767–784, 2004.

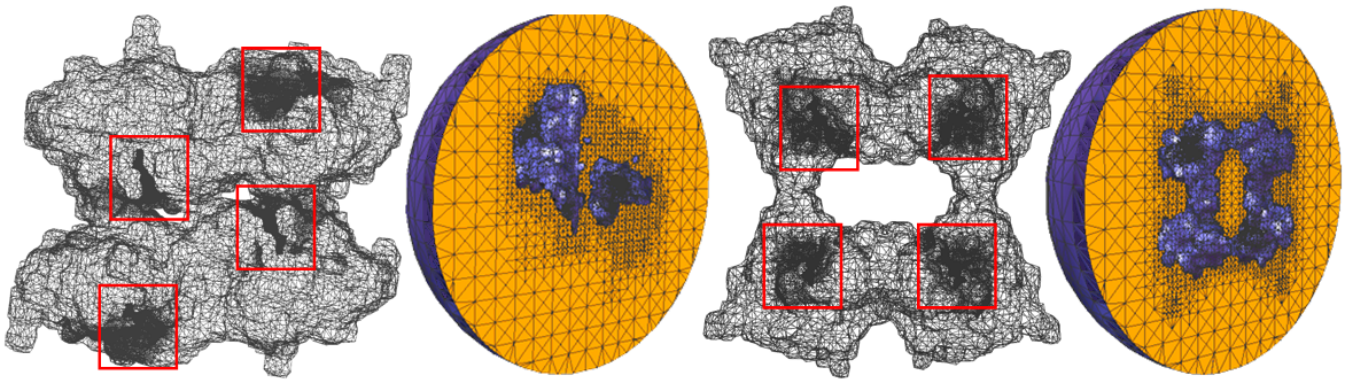
[31] D. Zhang, J. Suen, Y. Zhang, Y. Song, Z. Radic, P. Taylor, M. Holst, C. Bajaj, N. Baker, and J. McCammon. Tetrameric mouse acetylcholinesterase: Continuum diffusion rate calculations by solving the steady-state smoluchowski equation using finite element methods. *Submitted to Biophysical J.*, 2004.

[32] Y. Zhang, C. Bajaj, and B.-S. Sohn. Adaptive and quality 3d meshing from imaging data. In *ACM Solid Modeling and Applications*, pages 286–291, 2003.

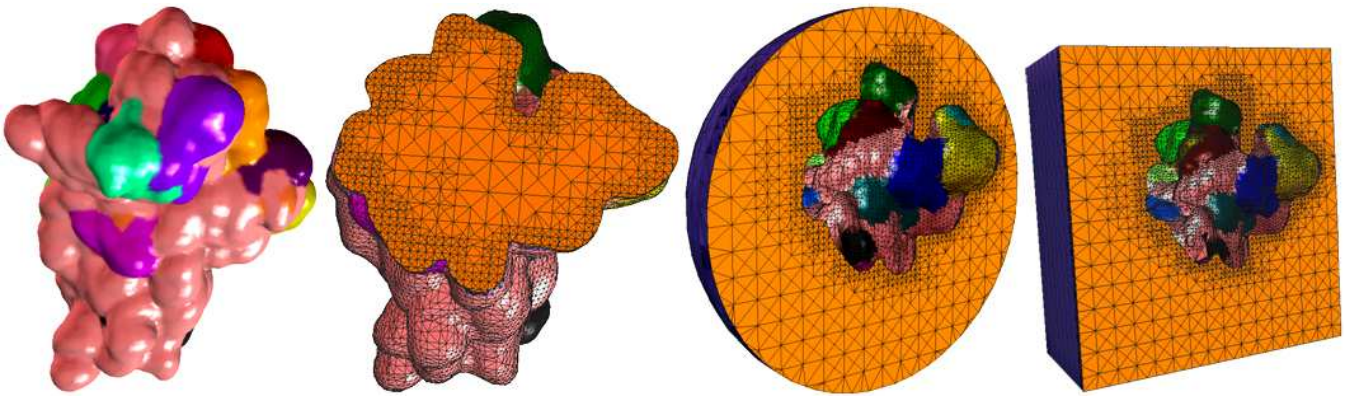
[33] Y. Zhang, C. Bajaj, and B.-S. Sohn. 3d finite element meshing from imaging data. *Special issue of Computer Methods in Applied Mechanics and Engineering on Unstructured Mesh Generation, in press*, 2004.



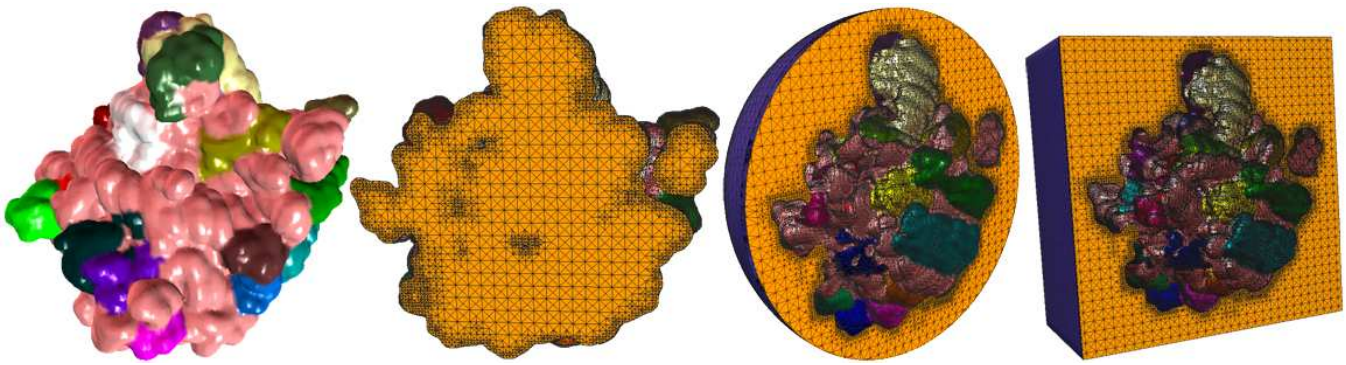
(a) Monomeric mAChE. From left to right: $\sigma = 2$ (65147 vertices, 323442 tetra) and (121670 vertices, 656823 tetra), $C = -0.262346$ (103680 vertices, 509597 tetra) and (138967 vertices, 707284 tetra). The color shows potential (leftmost) or residues (the right two).



(b) Tetrameric mAChE, $\sigma = 2$. The left two pictures show the 1st crystal structure 1C2O (133078 vertices, 670950 tetra), and the right two pictures show the 2nd one 1C2B, (106463 vertices, 551074 tetra). Cavities are shown in red boxes.



(c) Ribosome 30S, low resolution, $C = -0.03125$. From left to right: (33612 vertices, 163327 tetra), (37613 vertices, 186496 tetra) and (40255 vertices, 201724 tetra). The pink color shows 16S rRNA and other colors show proteins.



(d) Ribosome 50S, residual level, $C = -0.0625$. From left to right: (230025 vertices, 1141575 tetra), (234902 vertices, 1162568 tetra), (260858 vertices, 1315112 tetra). The light yellow/pink color show 5S/23S rRNA, other colors show proteins.

Figure 14: Interior and exterior tetra meshes of monomeric/tetrameric mAChE, Ribosome 30S/50S from the volume data generated by the characteristic function ($\sigma = 2$) or the Gaussian summation (various Cs).