

# Quality meshing of implicit solvation models of biomolecular structures <sup>☆</sup>

Yongjie Zhang <sup>a</sup>, Guoliang Xu <sup>b</sup>, Chandrajit Bajaj <sup>a,\*</sup>

<sup>a</sup> *Computational Visualization Center, Department of Computer Sciences, Institute for Computational Engineering and Sciences,  
The University of Texas at Austin, Austin, TX 78712, USA*

<sup>b</sup> *State Key Laboratory of Scientific and Engineering Computing, Institute of Computational Mathematics,  
Academy of Mathematics and System Sciences, Chinese Academy of Sciences, Beijing, 100080, China*

Received 1 July 2005; received in revised form 18 November 2005; accepted 27 January 2006

Available online 2 May 2006

## 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, a smooth volumetric electron density map is constructed from atomic data using weighted Gaussian isotropic kernel functions and a two-level clustering technique. This enables the selection of a smooth implicit solvation surface approximation to the Lee–Richards molecular surface. Next, a modified dual contouring method is used to extract triangular meshes for the surface, and tetrahedral meshes for the volume inside or outside the molecule within a bounding sphere/box of influence. Finally, geometric flow techniques are used to improve the surface and volume mesh quality. Several examples are presented, including generated meshes for biomolecules that have been successfully used in finite element simulations involving solvation energetics and binding rate constants.

© 2006 Elsevier B.V. All rights reserved.

*Keywords:* Quality mesh; Biomolecule; Implicit solvation model; Finite element simulation

## 1. Introduction

Finite element simulations have become an important tool in the analysis of biomolecular functional models, such as electrophoresis, electrostatics and diffusion influenced reaction rate constants (Song et al., 2004a, 2004b; D. Zhang et al., 2005). For efficient and accurate finite element solutions, adaptive and quality meshes are a necessary first step. The primary difficulty remains in generating satisfactory triangular or tetrahedral meshes for complicated macromolecules such as the Ribosome consisting of proteins and nucleic acids (Nissen et al., 2000) (Fig. 1), or protein structures such as acetylcholinesterases whose active site occurs at the bottom of a narrow gorge (pocket) (Fig. 14).

The protein data bank (<http://www.rcsb.org/pdb>) (Berman et al., 2000) provides PDB format files for protein and nucleic acid structures, with the location of principally all the major atoms making up the molecule (e.g., hydrogen atoms are not discernible via X-ray diffraction and therefore rarely present in the PDB). The summa-

<sup>☆</sup> Visit <http://www.ices.utexas.edu/cvc/meshing/MolMesh>.

\* Corresponding author.

*E-mail address:* [bajaj@cs.utexas.edu](mailto:bajaj@cs.utexas.edu) (C. Bajaj).

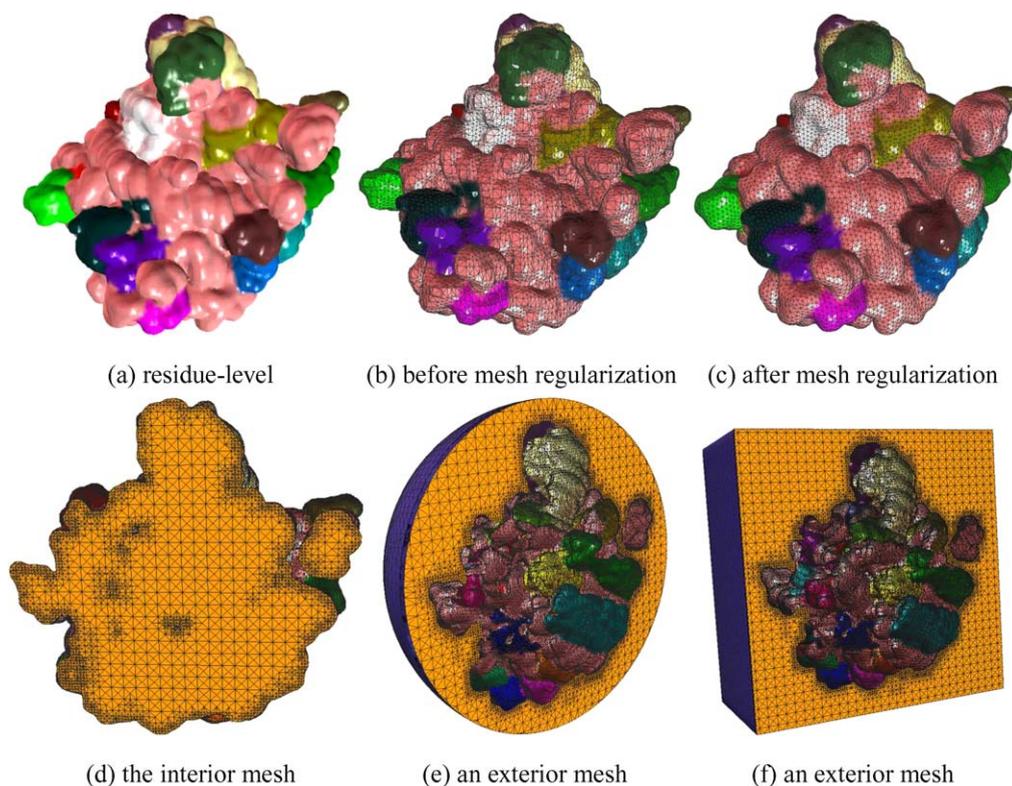


Fig. 1. Implicit solvation models of *Haloarcula Marismortui* large Ribosome 50S (1JJ2) subunit. The light yellow and the pink color show 5S and 23S rRNA respectively, the remaining colors are proteins. (a) The implicit solvation model at the medium resolution level,  $p_1 = 0.0625$ ,  $p_2 = 1.0$ ; (b) and (c) triangular meshes (16700 vertices, 33400 triangles); (d) the interior mesh (230025 vertices, 1141575 tets); (e) an exterior mesh within a sphere (234902 vertices, 1162568 tets); (f) an exterior mesh within a bounding box (260858 vertices, 1315112 tets). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

tion of kernel functions centered at each atom can be used to construct a smooth volumetric electron density map from PDB data (Blinn, 1982; Grant and Pickup, 1995). The volumetric data is often sampled at each rectilinear grid point,  $V = \{F(i, j, k) \mid i, j, k \text{ are indices of } x, y, z \text{ coordinates in a rectilinear grid}\}$ , and the implicit solvation surface is approximated as a level set  $S_F(c) = \{(x, y, z) \mid F(i, j, k) = c\}$ , where  $c$  is a constant (Blinn, 1982; Grant and Pickup, 1995; Lee et al., 2003). 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 (Bajaj and Siddavanahalli, 2005). In this paper, we describe an approach to generate quality triangular/tetrahedral meshes for complicated biomolecular structures directly from the PDB format data, conforming to a good implicit solvation surface approximation. There are three main steps in our mesh generation process:

- (1) **Implicit Solvation Surface Construction**—A smooth implicit solvation model is constructed to approximate the Lee–Richards molecular surface by using weighted Gaussian isotropic atomic kernel functions and a two-level clustering techniques.
- (2) **Mesh Generation**—A modified dual contouring method is used to extract triangular and interior/exterior tetrahedral meshes, conforming to the implicit solvation surface. The dual contouring method (Y. Zhang et al., 2003, 2005) is selected for mesh generation as it tends to yield meshes with better aspect ratio. In order to generate exterior meshes described by biophysical applications (Song et al., 2004a, 2004b; D. Zhang et al., 2005), we add a sphere or box outside the implicit solvation surface, and create an outer boundary. Our extracted tetrahedral mesh is spatially adaptive and attempts to preserve molecular surface features while minimizing the number of elements.

- (3) Quality Improvement—Geometric flows are used to improve the quality of extracted triangular and tetrahedral meshes.

The main contributions of this paper include: a simple and uniform treatment for approximating implicit solvation models, coupled to an adaptive surface and volume mesh extraction scheme combined with geometric flow to yield high quality meshes. The generated tetrahedral meshes of the monomeric and tetrameric mouse acetylcholinesterase (mAChE) (Bourne et al., 1995, 1999) have been successfully used in solving the steady-state Smoluchowski equation using the finite element method (Song et al., 2004a, 2004b; D. Zhang et al., 2005).

The remainder of this paper is organized as follows: Section 2 reviews related previous work. Section 3 introduces how to construct an implicit solvation surface from PDB molecular structural data. Section 4 details our mesh generation scheme. Finally Section 5 describes our mesh quality improvement technique. Section 6 presents several molecular meshing results.

## 2. Previous work

*Molecular surface approximation:* There are three different yet often used molecular interfaces (Sanner et al., 1996), the van der Waals surface (VWS), the solvent-accessible surface (SAS) and the solvent-excluded surface (SES) (Connolly, 1983) or sometimes called the Lee–Richards surface (Lee and Richards, 1971). The SES or the Lee–Richards surface is by far the molecular surface of choice for solvation energy calculations (Holst et al., 2000; Lee et al., 2003), and is the surface for our meshing approximations.

According to the properties of molecular structures, Laug and Borouchaki used a combined advancing front and generalized Delaunay approach to mesh molecular surfaces (Laug and Borouchaki, 2002). Algorithms were developed for sampling and triangulating a smooth surface with correct topology (Akkiraju and Edelsbrunner, 1996). Skin surfaces, introduced by Edelsbrunner in (Edelsbrunner, 1999), have a rich combinational structure and provide a smooth alternative to the Lee–Richard’s surface. Cheng et al. (2001) maintained an approximating triangulation of a deforming skin surface. Bajaj et al. (1997) give NURB approximation of Lee–Richards molecular surfaces as well as present methods to maintain molecular surfaces for varying solvent radii (Bajaj et al., 2003). Compressed volumetric representation of molecular surfaces is given in (Bajaj et al., 2005). Simplex subdivision schemes are used to generate tetrahedral meshes for molecular structures in solving the Poisson–Boltzmann equation (Holst et al., 2000). Gaussian functions have been used to construct density maps (Blinn, 1982; Grant and Pickup, 1995; Mezey, 1993; Agarwal, 1978; Max, 1988), from which implicit solvation models are approximated as an isocontour (Grant and Pickup, 1995; Lee et al., 2003; Gabdoulline and Wade, 1996). However, it still remains a challenge to generate quality and adaptive triangular and tetrahedral meshes for arbitrary molecular structures.

*Mesh generation:* As reviewed in (Owen, 1998; Teng and Wong, 2000), 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 input geometry until the desired resolution is reached (Shephard and Georges, 1991). Advancing front methods start from a boundary and move a meshed front from the boundary towards empty space within the domain (Frey et al., 1996; Lohner et al., 1986). Delaunay refinement is used to refine triangles or tetrahedra locally by inserting new nodes to maintain the Delaunay criterion (‘empty circum-sphere’) (Chew, 1997). Sliver Exudation (Cheng et al., 2000) was used to eliminate slivers (bad aspect ratio). Shewchuk (2002) solves the problem of enforcing boundary conformity by use of constrained Delaunay triangulations (CDT).

The predominant algorithm for isosurface extraction from volume data is Marching Cubes (MC) (Lorenson and Cline, 1987), 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 (Fujishiro et al., 1996). A different and systematic algorithm was proposed for interval volume tetrahedralization (Nielson and Sung, 1997). By combining SurfaceNets (Gibson, 1998) and the extended Marching Cubes algorithm (Kobbelt et al., 2001), octree based dual contouring (Ju et al., 2002) generates adaptive multiresolution isosurfaces with preservation of sharp features. The dual contouring method has also been extended to extract adaptive and quality tetrahedral meshes from volumetric imaging data (Y. Zhang et al., 2003, 2005).

*Quality improvement:* Algorithms for mesh quality improvement can be classified into three categories (Teng and Wong, 2000; Owen, 1998): local coarsening/refinement by inserting/deleting points, local remeshing by face/edge swapping and mesh smoothing by relocating vertices.

Laplacian smoothing relocates vertex position at the average of the nodes (vertices) incident to it (Field, 1988). Instead of relocating vertices based on a heuristic algorithm, the optimization technique measures the quality of the surrounding elements to a node and attempts to optimize it. The optimization-based smoothing yields better results, nevertheless it is more expensive than Laplacian smoothing. Therefore, a combined Laplacian/Optimization-based approach was recommended (Canann et al., 1998; Freitag, 1997). The Laplacian operator was discretized over triangular meshes (Meyer et al., 2002), and geometric flows have been used in surface and imaging processing (Sapiro, 2001; Xu et al., 2006). Physically-based simulations are used to reposition nodes (Lohner and Parikh, 1988). Anisotropic meshes are obtained from bubble placement and equilibrium (Shimada et al., 1997). Mesh regularization was discussed in (Ohtake et al., 2001; Wood et al., 2000).

### 3. Implicit solvation surface from volumetric density maps

We extract an implicit solvation surface (molecular surface) as a level set (isocontour) of the volumetric electron density maps (Bajaj and Siddavanahalli, 2005). The implicit solvation surface is chosen to be a good approximation of the Lee–Richards molecular surface (Lee and Richards, 1971) by choosing an appropriate weighting parameter of the summation of Gaussian kernel functions.

#### 3.1. Gaussian density map

The molecular surface has been approximated in the past (Blinn, 1982; Gabdoulline and Wade, 1996; Grant and Pickup, 1995) by an isocontour:

$$M := \{x \in \mathbb{R}^3: G(x) = 1\} \quad \text{with } G(x) = \sum_{i=1}^N e^{B_i \left( \frac{\|x-x_i\|^2}{r_i^2} - 1 \right)} \quad (1)$$

where  $(x_i, r_i)$  are the center and radius of the  $i$ th atom in the biomolecule, and  $B_i < 0$  is called the ‘decay rate’, which controls the rate of decay of each atom’s Gaussian kernel. Note that  $B_i$  must be negative to ensure that the density

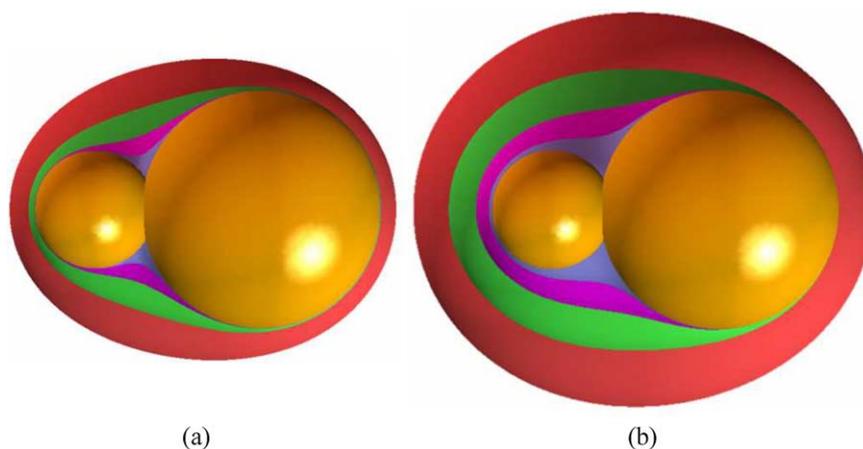


Fig. 2. Implicit solvation models by choosing various  $C$  in (a) and  $B_i$  in (b). Yellow balls are two input atoms. The correspondence between  $C/B_i$  values and these models are shown in Table 1. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 1

$C$  ( $1/\text{Angstrom}^2$ )/ $B_i$  (constant) and implicit solvation models in Fig. 2 (for colors see the web version of the article)

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

function goes to zero as  $\|x - x_i\|$  goes to infinity. In order to make the distance between  $M$  and  $M_0$  as uniform as possible, we take  $C = B_i/r_i^2$ , where  $C < 0$  is a given constant. Now  $G(x)$  becomes

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

In the following for the molecular surface  $M(C_i) = \{x \in \mathbb{R}^3: G(x, C_i) = 1\}$ , we consider  $C = C_1, \dots, C_l$ . As shown in Fig. 2, the different effects of  $C$  and constant  $B_i (= B)$  are shown for a two-sphere system, one is centered at  $(0, 0, 0)$  with radius of 1.0, the other one is at  $(2.8, 0, 0)$  with radius of 2.0. It can be observed that

- $C$  leads to more uniform inflation than  $B_i$ .
- Small balls have more inflation than big ones.
- Large error occurs around the intersection region, and the error reduces gradually away from it.
- Larger  $C$  and  $B_i$  lead to greater inflation. For the same  $C$  and  $B_i$  value, e.g.,  $-0.125$ ,  $B_i$  tends to introduce more inflation.
- Inflation of the molecular surfaces distorts the polar solvation energies and hence to be seriously avoided.

Fig. 3 shows implicit solvation models of Ribosome 30S. Compared with Fig. 3(a), proteins inflate much more seriously in Fig. 3(e). rRNA in Figs. 3(c) and (f) looks similar, but proteins in Fig. 3(f) look a little more inflated than Fig. 3(b). rRNA in Figs. 3(d) and (g) looks similar too, but proteins in Fig. 3(g) are close to proteins in Fig. 3(c).

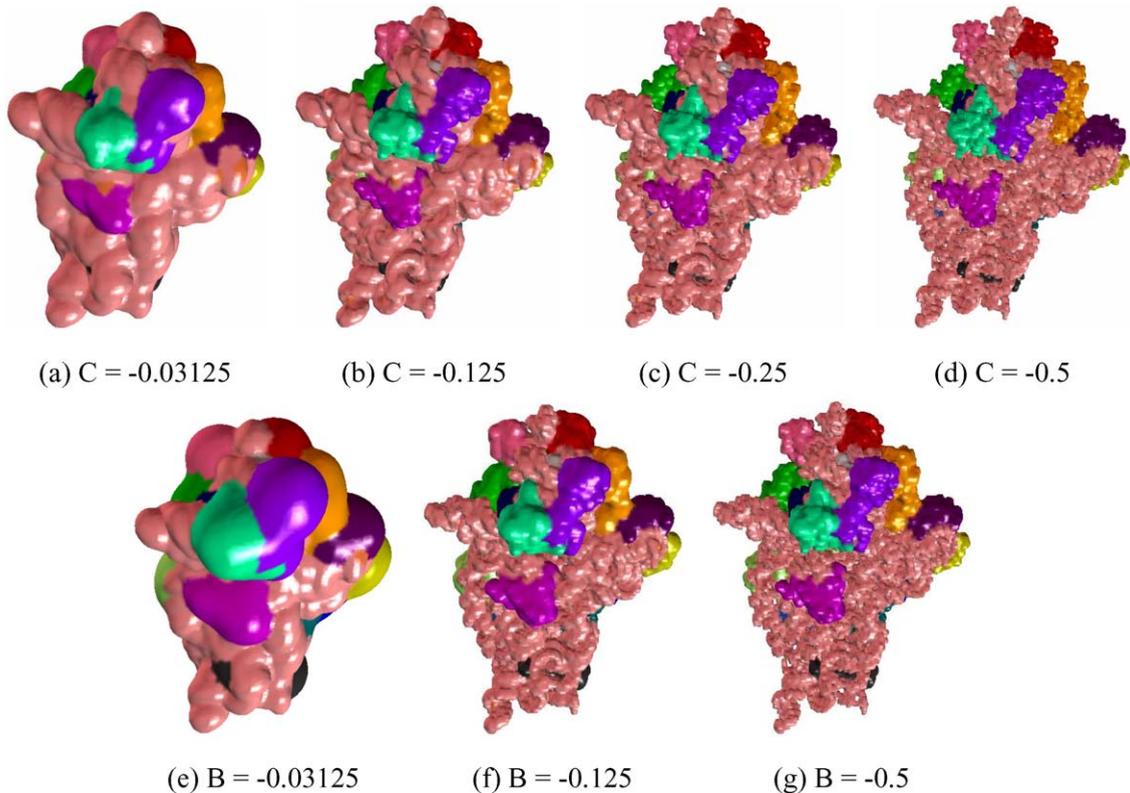


Fig. 3. Implicit solvation models of *Thermus Thermophilus* small Ribosome 30S (1J5E) crystal subunit for various Gaussian kernel parameters. The pink color shows 16S rRNA and the remaining colors are proteins. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

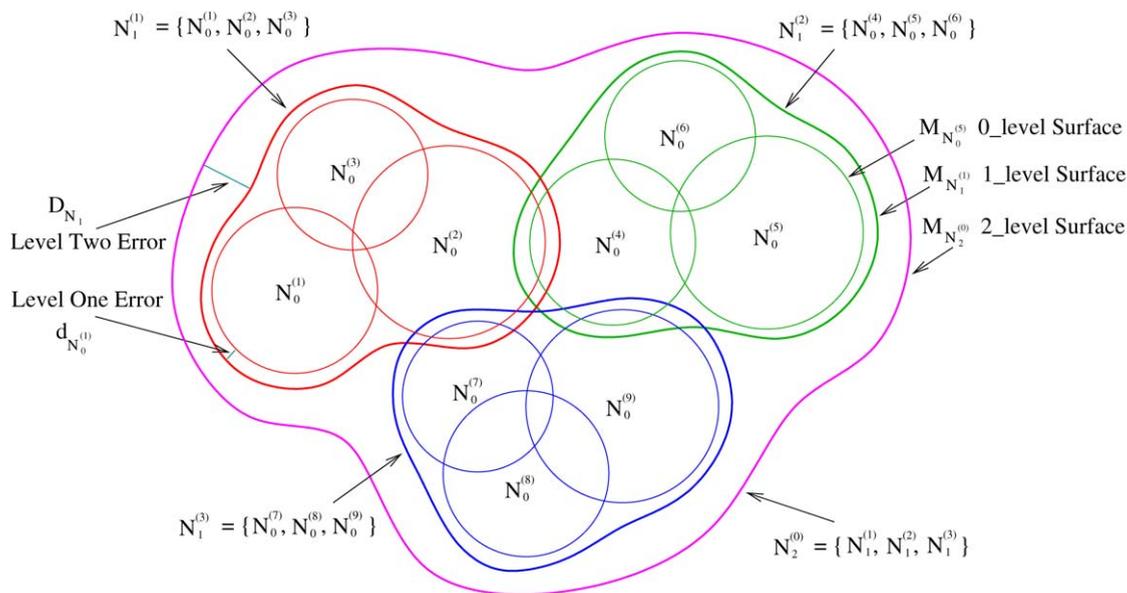


Fig. 4. The definition of multi-level surfaces.

### 3.2. Multi-level Gaussian density map

In order to reduce the inflation caused by Gaussian summation as well as to model molecular surfaces with varying resolution on the implicit solvation surface, we introduce a multi-level Gaussian map. First, we introduce some notation as shown in Fig. 4. Let  $N_0 = \{N_0^{(1)}, \dots, N_0^{(n)}\}$  denote the index set of all the atoms with  $N_0^{(i)} = \{i\}$ . Suppose  $N_0$  is grouped into several subsets  $N_1^{(i)}$ ,  $i = 1, 2, \dots, n_1$ , such that

$$\bigcup_{i=1}^{n_1} N_1^{(i)} = N_0, \quad N_1^{(i)} \cap N_1^{(j)} = \phi, \quad 1 \leq i \neq j \leq n_1 \tag{3}$$

The set  $N_1 := \{N_1^{(i)}\}_{i=1}^{n_1}$ , whose elements are also sets, may be further grouped into some subsets  $N_2^{(i)}$ ,  $i = 1, 2, \dots, n_2$ , such that

$$\bigcup_{i=1}^{n_2} N_2^{(i)} = N_1, \quad N_2^{(i)} \cap N_2^{(j)} = \phi, \quad 1 \leq i \neq j \leq n_2 \tag{4}$$

The collection of  $\{N_2^{(i)}\}_{i=1}^{n_2}$  is denoted by  $N_2$ . This hierarchical grouping process could be repeated several times according to the molecular complex considered. In practice, two or three iterations suffice. By using these sets  $N_k^{(i)}$  and a given sequence  $\{p_k\}$  of  $p$  with  $p_k > 0$ , the  $k$ -level Gaussian map are defined recursively as

$$G_{N_k^{(i)}}(x) = \sum_{N \in N_k^{(i)}} [G_N(x)]^{p_k}, \quad N_k^{(i)} \in N_k,$$

where 0-level Gaussian map is defined by Eq. (2) ( $C = 1.0$ ) or

$$G_{N_0^{(i)}}(x) = K(\|x - x_i\|)/K(r_i), \quad K(x) = e^{-x^2}.$$

The atom group format depends on what kind of structure we wish to model and mesh. For a protein, atoms may be grouped by residues, meaning that atoms in the same residue are classified into one group. Then the residues are grouped according to their neighborhood along the protein backbone.

For each  $k$ -level Gaussian Map  $G_{N_k^{(i)}}(x)$ , a  $k$ -level surface is defined by

$$M_{N_k^{(i)}}(x) := \{X \in \mathbb{R}^3: G_{N_k^{(i)}}(x) = 1\}.$$

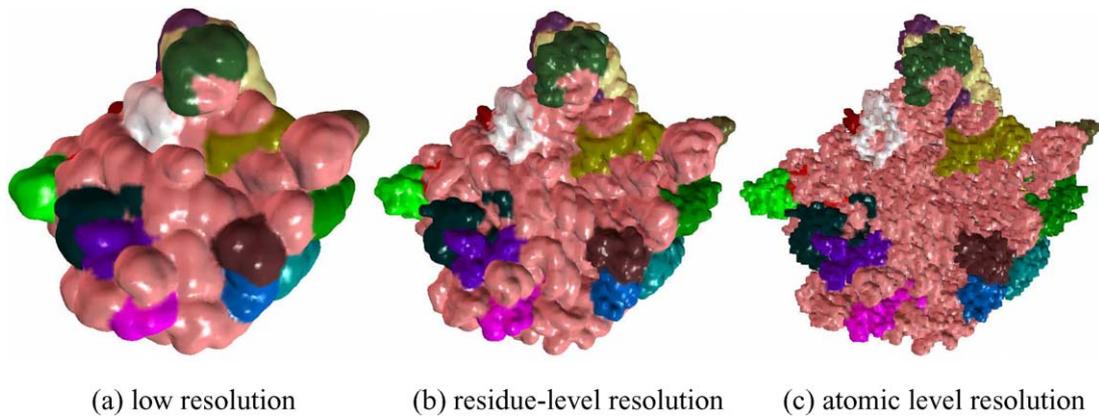


Fig. 5. Implicit solvation models of Haloarcula Marismortui large Ribosome 50S (1JJ2) crystal subunit. (a)  $p_1 = 0.03125$ ; (b)  $p_1 = 0.125$ ; (c)  $p_1 = 0.5$ ,  $p_2 = 1.0$ . The light yellow and the pink color show 5S and 23S rRNA respectively, the remaining colors are proteins. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

This surface encloses the surface  $M_N$  for  $N \in N_k^{(i)}$ . Hence, all these  $N_k^{(i)}$  define a hierarchical surface family. We call the surface  $M_N$  as the child of  $M_{N_k^{(i)}}$ , and  $M_{N_k^{(i)}}$  the parent of  $M_N$ . The enclosing relation of this hierarchical surface family is strict, meaning that the minimal distance from  $M_N$  to  $M_{N_k^{(i)}}$  is greater than zero for any  $N \in N_k^{(i)}$ . We further define the  $B$ -surface of  $M_N$  for all  $N \in N_k^{(i)}$  as

$$S_{N_k^{(i)}} = \text{Bd} \left( \bigcup_{N \in N_k^{(i)}} \{x \in \mathbb{R}^3: G_N(x) \leq 1\} \right),$$

where  $\text{Bd}()$  denotes the boundary of a region in  $\mathbb{R}^3$ . Note that  $S_{N_k^{(i)}}$  is enclosed strictly by  $M_{N_k^{(i)}}$ .

The purpose of introducing a multi-level Gaussian map is to address the structure of molecules at a certain level. For instance, at the residue level of a protein, we dealt with each residue as one unit and therefore the protein is considered at the residue level resolution. The sub-structures of the residue (atoms), are not individually identifiable. Similarly, at the next higher level, a group of residues is dealt as one unit and therefore the protein is considered at an even coarser feature resolution. The goal of addressing certain level structure and un-addressing the higher level ones is achieved by the properly selection of the parameter  $p_k$  in the multi-level Gaussian map. Basically, larger  $p_k$  should be chosen to address the  $k$ -level structure and a smaller  $p_{k-1}$  is used to un-address the  $(k-1)$ -level structures.

Considering three levels of structures, including the atomic, the residue and the next level of grouping, we can construct a three level Gaussian map with given  $p_1$ ,  $p_2$  and  $p_3$ . To address the second level structure, we need to choose  $p_3$  larger and  $p_2$  smaller, while  $p_1$  has less influence than the second level structure. Quite often it also suffices to consider only a two-level Gaussian map instead of three: level one is at the protein residue level, while level two is at a coarser resolution level. Henceforth in this paper, we provide details for only two-level Gaussian maps.

In computing implicit solvation molecular surfaces, various models are constructed by choosing different  $p_1 \in (0, \infty)$  and  $p_2 \in (0, \infty)$  in the Gaussian map. To make the constructed model correspond to a certain level,  $p_1$  and  $p_2$  need to be selected properly. For a fixed level, the structure at this level should be distinguishable. For instance, at the residue level, the individual residues should be observed, while atoms may not be distinguished clearly. Fig. 5 shows constructed models of Ribosome 50S at low resolution, residue and atomic level resolutions.

### 3.3. Approximation computation

In order to obtain a good approximation to the molecular surface from the multi-level Gaussian map, we bound the error at each level. To bound the approximation for the first level, we need to compute the minimal distance from  $M_N$ ,  $N \in N_1^{(i)}$  to its parent surface  $M_{N_1^{(i)}}$ . On the other hand, in order to have an error controlled approximation of the

second level surface, we need to compute the maximal error from  $M_N$ ,  $N \in N_2^{(i)}$  to its parent surface  $M_{N_2^{(i)}}$ . Hence, we need to consider the error computation for both levels of surfaces. The error computations are based on a point projection algorithm.

Given the surface  $M_N$  a point  $q \notin M_N$  and a unit direction  $n$ , the point projection algorithm in the following computes a nearby intersection point  $p$  of the line  $q + tn$  ( $t \in (-\infty, \infty)$ ) with the surface  $M_N$ .

**Algorithm 3.3.1** (Point projection).

- (1) Compute an interval  $[a, b]$  for  $t$ , on which  $G_N(q + tn) - 1$  changes sign once. This is achieved by a linear search step in a certain range  $[A, B]$ . If  $(\nabla G_N(q))^T n [G_N(q) - 1] < 0$ , search in  $n$  direction, otherwise in  $-n$  direction. If such an interval could not be found, the project point does not exist and return a failure signal. After the interval is determined, set  $t_0 = \frac{a+b}{2}$  and  $k = 0$ .
- (2) Compute  $t_{k+1}$  by the Newton iteration method.

$$t_{k+1} = t_k - \frac{G_N(q + t_k n)}{n^T \nabla G_N(q + t_k n)}. \tag{5}$$

If  $t_{k+i} \notin (a, b)$ , replace  $t_{k+1}$  by  $\frac{a+b}{2}$ .

- (3) Replace the interval  $[a, b]$  by  $[a, t_{k+1}]$  if  $G_N(q + tn) - 1$  changes sign over  $[a, t_{k+1}]$ , and replace  $[a, b]$  by  $[t_{k+1}, b]$  otherwise.
- (4) If  $\|b - a\| < \varepsilon$  ( $\varepsilon$  is a given error tolerance, we usually take it to be  $10^{-4}$ ), stop the iteration and  $p = q + t_{k+1}n$  is the projection point, otherwise, set  $k = k + 1$  and go back to step 2.

We specify the searching range  $[A, B]$  in step 1 of the algorithm to be  $[-4, 4]$ , since the atom diameters are around 4. Errors beyond that are not considered here. If the projection exists, then the projection point  $p$  of point  $q$  on the surface  $M_N$  in the direction  $n$  is denoted by  $P_{M_N}(q, n)$ .

3.3.1. Minimal error of level one surface

Now we assume  $k = 1$ , then the child surfaces are atoms. Let  $N = \{j\} \in N_1^{(i)}$ , the minimal error from  $M_N = S_N$  to  $M_{N_1^{(i)}}$  is defined by

$$d_N := \min_{p \in M_{N_1^{(i)}, N}} \|p - x_j\| - r_j, \quad j \in N.$$

Let  $q = x_j + r_j \frac{p - x_j}{\|p - x_j\|}$ , then  $q$  is on the sphere  $S_N$  and  $p$  is the projection of  $q$  to the surface  $M_{N_1^{(i)}}$  in the spherical normal direction  $n(q)$ . That is,  $p = P_{M_{N_1^{(i)}}}(q, n(q))$ . Hence in order to compute  $d_N$ , we need to compute  $P_{M_{N_1^{(i)}}}(q, n(q))$  for  $q \in S_N$ .

Next we consider the computation of the minimal distance from  $M_N$  to  $M_{N_1^{(i)}}$ , where  $N \in N_1^{(i)}$ . First we assume that each atom (sphere) is uniformly sampled with  $m$  vertices. This sampling is achieved by translating a triangulated unit sphere to each of the atomic centers and re-scaling it to the atom's van der Waal radius. We obtain the unit sphere triangulation from (Xu, 2006). For each vertex  $q$  on the triangulated atom surface  $M_N$ ,  $P_{M_{N_1^{(i)}}}(q, n(q))$  is computed using the point projection algorithm, where  $n(q)$  is the spherical normal at  $q$ .

**Algorithm 3.3.2** (Minimal error computation).

Set  $d_N = 4$ .

for each triangle vertex  $q \in S_N \cap S_{N_0}$  do {

compute  $P_{M_{N_1}}(q, n(q))$ , and then compute

$$d_N = \min\{d_N, \|P_{M_{N_1}}(q, n(q)) - x_j\| - r_j\}, \tag{6}$$

if  $P_{M_{N_1}}(q, n(q)) \in M_{N_1, N}$

}

Table 2

Minimal error and maximal error of first level surfaces of Ribosome 30S (1J5E) (Angstrom)

$p_1$	0.25	0.5	1.0	2.0	4.0	8.0	16.0
Min error (atomic)	8.338e-02	2.829e-03	6.287e-06	$< 10^{-6}$	$< 10^{-6}$	$< 10^{-6}$	$< 10^{-6}$
Max error (atomic)	1.634e+00	8.656e-01	4.121e-01	2.038e-01	8.893e-02	3.940e-02	1.842e-02

Table 3

Maximal error of second level surfaces of Ribosome 30S (1J5E) (Angstrom)

$p_2$	0.25	0.5	1.0	2.0	4.0	8.0	16.0
Max error (residue)	3.923e+00	2.124e+00	6.832e-01	3.240e-01	1.550e-01	7.794e-02	3.278e-02
Max error (low)	9.899e+00	7.695e+00	8.045e-01	2.365e-01	1.390e-01	6.113e-02	2.653e-02

Table 2 shows the minimal error of our level one surface for a residue and a chain from Ribosome 30S, where  $e(M)$  is defined as  $e(M) := \max_{N \in N_1^{(i)}} d_N$ . It can be observed that the error decreases as  $p$  increases. The algorithm for computing minimal error can also be used to compute the maximal error by changing the min to max in (6). Maximal errors for Ribosome 30S are also listed in Table 2 for different  $p_1$  (see the second row).

### 3.3.2. Maximal error of level two surface

The maximal error from  $M_N$  to  $M_{N_2^{(i)}}$ ,  $N \in N_2^{(i)}$  is defined as

$$d_N := \max_{q \in M_N, P_{M_{N_2^{(i)}}}(q, n) \in M_{N_2^{(i)}, N}} \|q - P_{M_{N_2^{(i)}}}(q, n)\|,$$

where  $q \in M_N$ ,  $P_{M_{N_2^{(i)}}}(q, n)$  is the normal direction projection of  $q$  to the surface  $M_{N_2^{(i)}}$ . This error is computed as follows. Let  $N_1 \in N_2^{(i)}$ .

#### Algorithm 3.3.3 (Maximal error computation).

Set  $d_{N_1} = 0$ .

for each  $N \in N_1$  do {

  for each triangle vertex  $q \in S_N \cap S_{N_0}$  do {

    compute  $\tilde{q} := P_{M_{N_1}}(q, n(q))$ , and

    compute  $P_{M_{N_2^{(i)}}}(\tilde{q}, n(\tilde{q}))$  if  $q \in M_{N_1, N}$

  and then compute

$$d_{N_1} = \max\{d_{N_1}, \|P_{M_{N_1}}(q, n(q)) - P_{M_{N_2^{(i)}}}(q, n(q))\|$$

$$\text{if } P_{M_{N_2^{(i)}}}(\tilde{q}, n(\tilde{q})) \in M_{N_2^{(i)}, N_1}$$

  }

}

Again, the projection points  $\tilde{q} = P_{M_{N_1}}(q, n(q))$  and  $P_{M_{N_2^{(i)}}}(\tilde{q}, n(\tilde{q}))$  are computed by the point projection algorithm, where the searching range  $[A, B]$  is set to be  $[0, 4]$ , since we know  $M_{N_2^{(i)}}$  enclosing  $M_N$  and we are not interested in the errors that are larger than 4.

The first row of Table 3 shows the maximal errors of the second level (residue level) surfaces for Ribosome 30S, where  $p_1$  is chosen to be 0.5,  $p_2 = 0.25, 0.5, 1.0, \dots, 16$ . The second row lists the maximal errors of the second level (low level) surfaces for the same  $p_1$  and  $p_2$ . The results show that the errors decrease approximately as a linear rate as  $p_2$  increases.

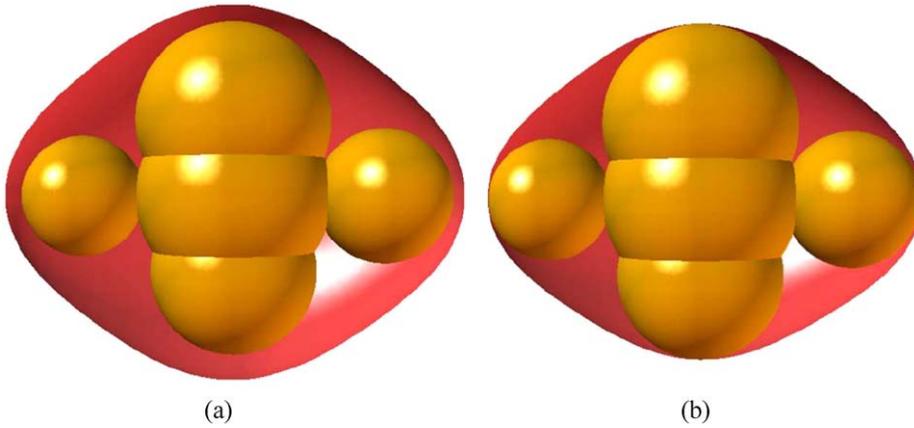


Fig. 6. The left picture shows the inflation effect by the Gaussian map. The right one shows the tight enclosure of atoms. The centers of the five atoms are  $(-2, 0, 0)$ ,  $(2, 0, 0)$ ,  $(0, -1, 0)$ ,  $(0, 1, 0)$  and  $(0, 0, 0)$ . The corresponding radii are 0.8, 0.9, 1.1, 1.3 and 1.3. The parameter  $p$  in the Gaussian map is chosen to be 0.4. The tight approximation on the right figure is obtained by shrinking the five radii into 0.55644, 0.72525, 0.60476, 1.04567 and 0.0 respectively. The unit is Angstrom.

### 3.4. Good approximations of molecular surfaces

We have discussed that it is often sufficient to consider a two-level Gaussian map to approximate molecular surfaces. To address certain structures,  $p_1$  is taken to be a small value to blur the higher level details,  $p_2$  is chosen to be larger to enhance the feature of the current level structure. As we have shown in the last section, a smaller  $p_1$  leads to a larger error for the level one surface, and a larger  $p_2$  leads to a smaller error for the second level surface. Therefore, our strategy for obtaining a tight enclosing surface approximation is to remove the level one error and ignore the error of the second level.

The main idea to obtain a tight level one enclosing surface  $M_{N_1^{(i)}}$  is to reduce the radii of the atoms, such that  $M_{N_1^{(i)}}$  touches the original atoms (see Fig. 6). Suppose  $y \in M_{N_1^{(i)}}$  is the nearest point to the  $j$ th atom,  $j \in N_1^{(i)}$ , and the distance from  $y$  to the atom is  $d_j$ . Then we have

$$\sum_{l \in N_1^{(i)}, l \neq j} [K(\|y - x_l\|) / K(r_l)]^{p_1} + [K(\|y - x_j\|) / K(r_j)]^{p_1} = 1 \tag{7}$$

where  $K(x) = e^{-x^2}$ . Now we adjust the radius  $r_j$  to  $\tilde{r}_j$ , such that the new nearest point  $y$  is on the  $j$ th sphere. Since the dominating part of (7) is the second term of the left hand side, we therefore require  $\tilde{r}_j$  satisfying

$$0 \leq \tilde{r}_j \leq r_j, \tag{8}$$

$$K(r_j + d_j) / K(r_j) = K(r_j) / K(\tilde{r}_j). \tag{9}$$

From this we obtain

$$\tilde{r}_j = \begin{cases} K^{-1} \left[ \frac{K(r_j)^2}{K(r_j + d_j)} \right], & \text{if } \frac{K(r_j)^2}{K(r_j + d_j)} \in \text{Range}(K), \\ 0, & \text{otherwise,} \end{cases}$$

where  $K^{-1}$  denotes the inverse function of  $K(x)$ ,  $\text{Range}(K) := \{y \in \mathbb{R}: y = K(x), x \in (0, \infty)\}$ . Based on this analysis, we build the following iterative algorithm for computing  $\tilde{r}_j$ .

**Algorithm 3.4.1** (Sphere shrinking).

For  $i = 1, 2, \dots, n_i$  do the following steps:

- (1) Set  $l = 0, r_j^l = r_j, d_j^{(l)} = \infty, \forall j \in N_1^{(i)}$ .

Table 4  
Errors  $\epsilon_{\max}^{(l)}$  for 20 amino acids and  $p_1 = 0.4$

$l$	ALA	ARG	ASN	ASP	CYS	GLN	GLU	GLY	HSD	ILE
0	5.13e-01	6.97e-01	5.99e-01	6.23e-01	5.36e-01	6.26e-01	7.06e-01	4.34e-01	7.36e-01	6.00e-01
2	6.22e-02	1.37e-01	2.66e-01	6.75e-02	5.86e-02	1.16e-01	7.78e-02	5.33e-02	7.20e-02	5.62e-02
4	2.80e-03	3.79e-02	5.83e-02	1.50e-03	6.82e-04	1.76e-03	4.57e-04	1.90e-02	1.45e-02	2.73e-03
6	5.76e-04	2.30e-02	1.83e-04	4.93e-04	1.81e-04	4.51e-04	1.38e-04	8.62e-05	5.30e-03	5.60e-04
8	1.30e-04	6.95e-04	6.06e-05	1.64e-04	4.97e-05	1.74e-04	4.26e-05	6.31e-06	2.20e-03	1.25e-04
10	3.14e-05	2.18e-04	2.22e-05	5.59e-05	1.39e-05	7.84e-05	1.32e-05	7.16e-07	9.94e-04	3.11e-05
$l$	LEU	LYS	MET	PHE	PRO	SER	THR	TRP	TYR	VAL
0	8.48e-01	8.62e-01	6.08e-01	6.14e-01	7.98e-01	9.63e-01	1.06e-00	6.01e-01	6.10e-01	7.07e-01
2	6.51e-02	3.96e-01	1.13e-01	8.94e-02	2.06e-03	8.81e-02	3.06e-02	9.17e-02	6.03e-02	2.86e-02
4	5.72e-03	1.54e-03	7.78e-03	6.50e-03	3.62e-04	5.28e-04	6.63e-03	1.49e-02	4.25e-02	5.76e-03
6	1.27e-03	5.18e-04	2.25e-03	1.90e-03	9.12e-05	1.19e-04	1.68e-03	6.42e-03	1.69e-03	1.36e-03
8	3.03e-04	1.77e-04	6.77e-04	7.13e-04	2.35e-05	2.66e-05	4.67e-04	2.90e-03	6.93e-04	3.56e-04
10	7.52e-05	6.23e-05	2.09e-04	3.02e-04	6.26e-06	5.88e-06	1.36e-04	1.56e-03	3.52e-04	9.78e-05

(2) Compute the minimal distance  $d_j^{(l+1)}$ ,  $\forall j \in N_1^{(l)}$  from the  $j$ th atom to the iso-surface defined by the multi-level Gaussian map  $G_{N_1^{(l)}}^{(l)}(x) = \sum_{j \in N_1^{(l)}} [K(\|x - x_j\|)/K(r_j^{(l)})]^{p_1}$ , using Algorithm 4.2.

(3) Compute

$$r_j^{(l+1)} = \begin{cases} K^{-1} \left[ \frac{K(r_j)K(r_j^{(l)})}{K(r_j+d_j^{(l)})} \right], & \text{if } \frac{K(r_j)K(r_j^{(l)})}{K(r_j+d_j^{(l)})} \in \text{Range}(K), \\ 0, & \text{otherwise.} \end{cases}$$

(4) If  $\max_{j \in N_1^{(l)}} \|d_j^{(l)} - d_j^{(l+1)}\| < \epsilon$  (we take  $\epsilon = 10^{-4}$ ), terminate the  $l$  loop and  $r_j^{(l+1)}$  are the required results. Otherwise, set  $l = l + 1$  and go back to step 2.

**Remark.** The condition  $\frac{K(r_j)K(r_j^{(l)})}{K(r_j+d_j^{(l)})} \in \text{Range}(K)$  may lead to some of the atoms located in the interior of the molecule to become untouchable. Fig. 6 shows that the circle at the origin is not touched.

The experiments show the sphere shrinking algorithm converges at a linear rate. Table 4 lists the error  $\epsilon_{\max}^{(l)} = \max_{j \in N_1^{(l)}} \|d_j^{(l)}\|$  for 20 amino acids with  $p_1 = 0.4$ .

Fig. 7 shows multi-resolution implicit solvation surface approximations of an ASN-THR-TYR peptide with various  $p_1$  and  $p_2$ . Fig. 7(a) shows an atomic level model, Figs. 7(a)–(g) are residue level models. It can be observed that when the same  $p_1$  is selected, smaller  $p_2$  leads to fatter surfaces. Compared with Fig. 7(g), Fig. 7(f) is more tight.

Fig. 8 shows multi-resolution implicit solvation surface approximation of Ribosome 30S. Fig. 8(a) is a low level model, the pink color shows 16S rRNA and the remaining colors are proteins (for colors see the web version of the article). One protein (Chain B) is separated from the whole structure. The residue level model can be constructed by selecting small  $p_1$  and large  $p_2$  as shown in Fig. 8(b), and the atomic level model is constructed by selecting large  $p_1$  and small  $p_2$  as shown in Fig. 8(c).

## 4. Mesh generation

There are two main methods for contouring scalar fields, primal contouring (Lorenson and Cline, 1987) and dual contouring (Ju et al., 2002). Both of them can be extended to tetrahedral mesh generation. The dual contouring method (Y. Zhang et al., 2003, 2005) is often the method of choice as it tends to yield meshes with better aspect ratio.

### 4.1. Triangular meshing

Dual contouring (Ju et al., 2002) 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

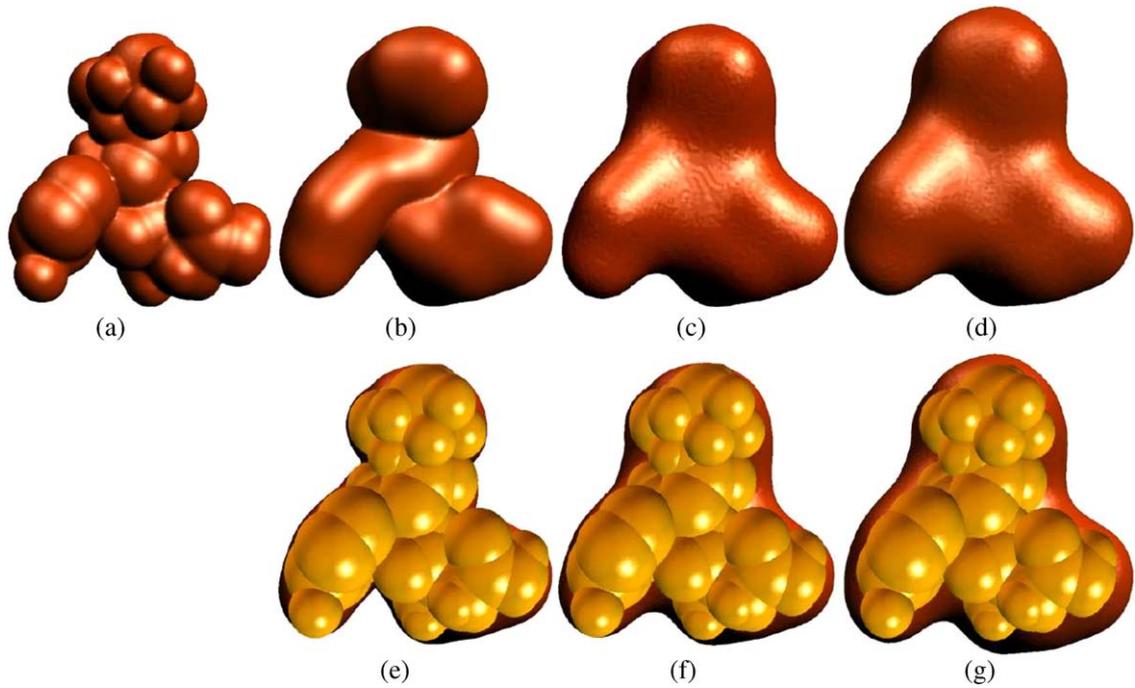


Fig. 7. Different effects of changing  $p_2$  and tight/non-tight approximations for an ASN-THR-TYR peptide which consists of 49 atoms. The surface (b), (c) and (d) are the same as outer surfaces of (e), (f) and (g) respectively. The inner surface of (e), (f) and (g) is the hard sphere model of three residues. (a) shows the atomic level approximation of the hard sphere model, where  $p_1 = 5.0$ ,  $p_2 = 1.0$ ; (b), (e), (c) and (f) show the tight approximation of the residue level with  $p_1 = 0.4$ . But different  $p_2$  are used. We choose  $p_2 = 2.0$  for (b)  $p_2 = 0.5$  for (c). It could be observed that larger  $p_2$  leads to closer approximation. (d) and (g) show non-tight approximations using the same  $p_1$  and  $p_2$  as (c) and (f). Comparing with (f), even larger error is observed in (g).

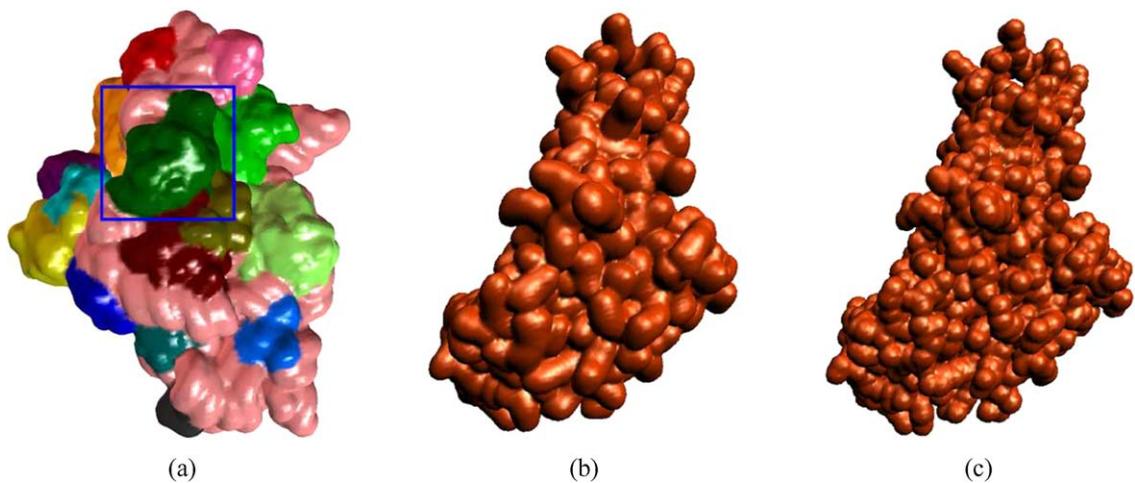


Fig. 8. Multi-resolution models of Ribosome 30S. (a) Ribosome 30S at the low level with  $p_1 = 0.0625$ ,  $p_2 = 1.0$  in multi-level Gaussian map. Ribosome 30S contains 22 chains and each of them is painted in a different color. The pink color shows 16S rRNA and the remaining colors are proteins. The blue box shows one protein (Chain B). (b) Chain B at the residue level with  $p_1 = 0.4$ ,  $p_2 = 5.0$ . It consists of 234 residues. (c) Chain B at the atomic level with  $p_1 = 5.0$ ,  $p_2 = 1.0$ . It consists of 1900 atoms. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

octree construction. Each sign change edge is shared by either four (uniform case) or three (adaptive case) cells, and one minimizer point is calculated for each of them by minimizing a predefined Quadratic Error Function (QEF) (Garland and Heckbert, 1998):

$$QEF[x] = \sum_i [n_i \cdot (x - p_i)]^2, \quad (10)$$

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 trilinear topology (Y. Zhang et al., 2005) of the isosurface. During cell subdivision, the function value at each newly inserted grid point can be exactly calculated since we know the volumetric function (Eq. (2)). Additionally, we can generate a more accurate triangular mesh by projecting each generated minimizer point onto the isosurface (Eq. (1)).

#### 4.2. Tetrahedral meshing

The dual contouring method has already been extended to extract tetrahedral meshes from volumetric scalar fields (Y. Zhang et al., 2003, 2005). 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 (Y. Zhang et al., 2003, 2005), 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 biological diffusion systems, we need to analyze the electrostatic potential field which is faraway from the molecular surface (Holst et al., 2000; Lee et al., 2003). 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. 9(b)). Sometimes the outer boundary is chosen to be a cubic box as shown in Fig. 9(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

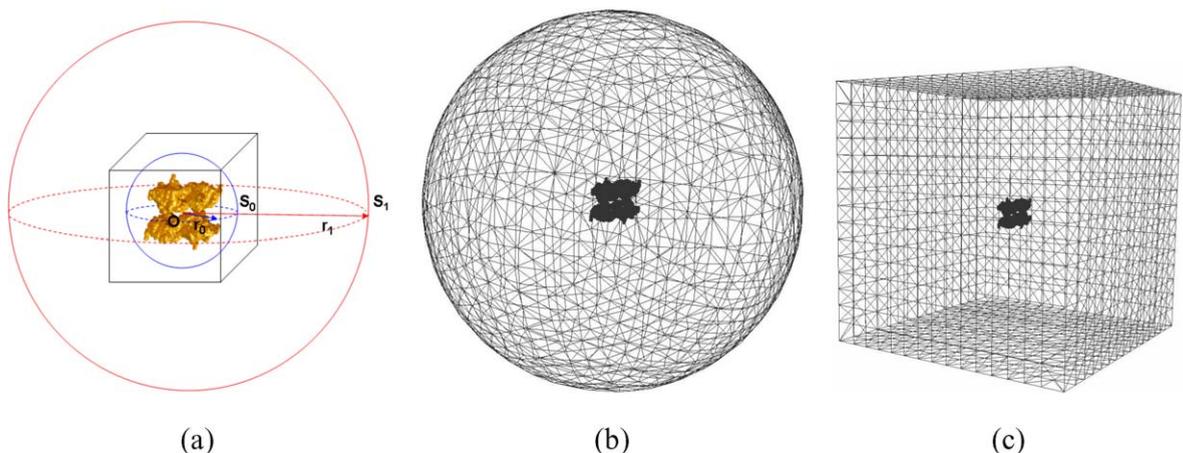


Fig. 9. The analysis domain of exterior meshes. (a) ‘O’ is the geometric center of the molecule, suppose the circum-sphere of the biomolecule has the radius of  $r$ . The box represents the volumetric data, and ‘ $S_0$ ’ is the maximum sphere inside the box, the radius is  $r_0$  ( $r_0 > r$ ). ‘ $S_1$ ’ 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 molecular surface and the outer sphere ‘ $S_1$ ’, here we choose  $r_1 = 5r$  for visualization. (c) The outer boundary is a cubic box.

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  $S_0$ . Eq. (11) shows the newly constructed function  $g(x)$  which provides a grid-based volumetric data containing the biomolecular surface and an outer sphere  $S_0$ .

$$g(x) = \begin{cases} \min(\|x - x_0\| - r_0, f(x) - \alpha), & \text{if } f(x) < \alpha, \|x - x_0\| < r_0, \\ \|x - x_0\| - r_0, & \text{if } f(x) < \alpha, \|x - x_0\| \geq r_0, \\ f(x) - \alpha, & \text{if } f(x) \geq \alpha, \end{cases} \quad (11)$$

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

#### 4.2.2. Mesh extraction

Here we introduce a different scheme from the algorithm presented in (Y. Zhang et al., 2003, 2005), 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. 10(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. 11, 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.

Compared with the algorithm presented in (Y. Zhang et al., 2003, 2005) as shown in Fig. 10(a), Fig. 10(b) generates the same surface meshes, and tends to generate more regular interior meshes with better aspect ratio, but a few more elements for interior cells. Fig. 10(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

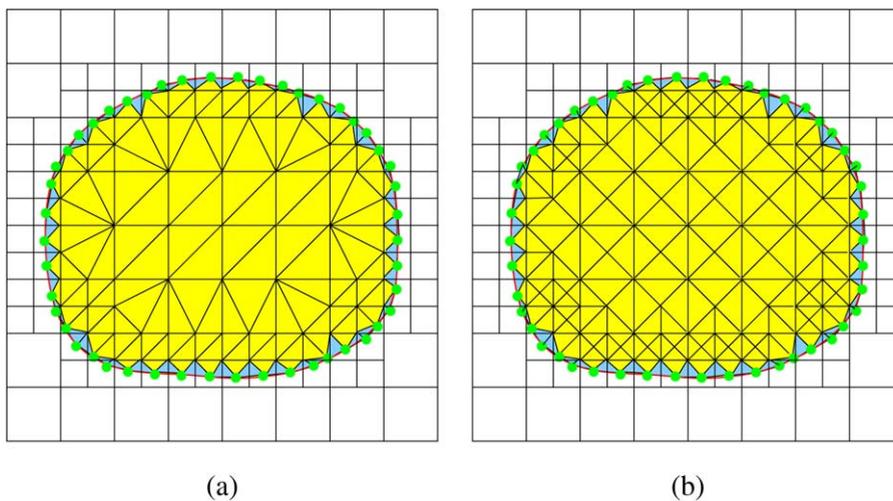


Fig. 10. 2D triangulation. (a) Old scheme, (b) new scheme. Blue and yellow triangles are generated for sign change edges and interior edges respectively. The red curve represents the molecular surface, and the green points represent minimizer points. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

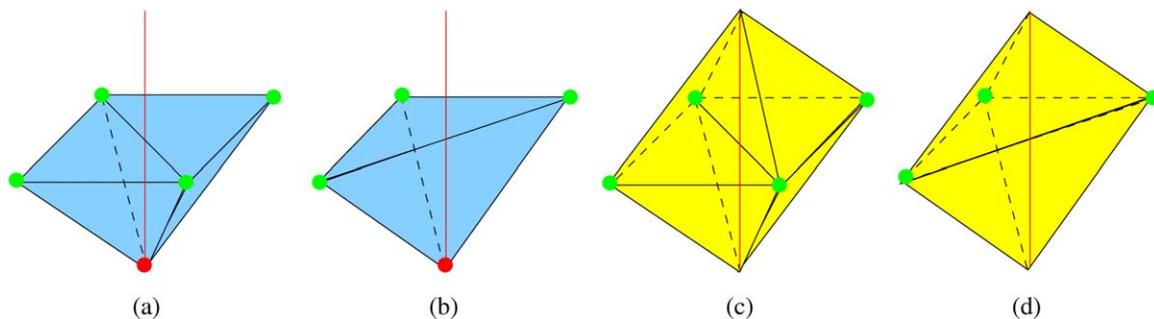


Fig. 11. Sign change edges and interior edges are analyzed in 3D tetrahedralization. (a), (b) Sign change edge (the red edge); (c), (d) interior edge (the red edge). The green solid points represent minimizer points, and the red solid points represent the interior vertex of the sign change edge. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

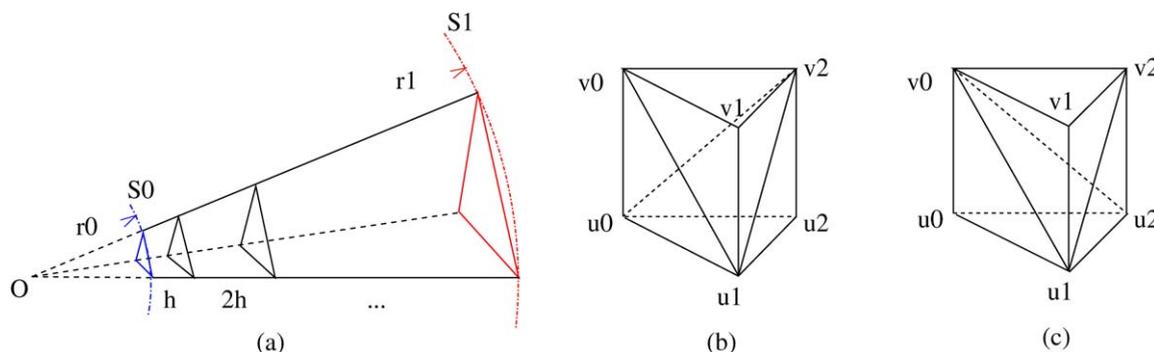


Fig. 12. (a) One triangle in the sphere  $S_0$  (blue) is extended  $n$  steps until arriving at the sphere  $S_1$  (red); (b), (c) a prism is decomposed into three tetrahedra in two different ways. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

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 shown in Figs. 17(a)–(b), and a coarse mesh everywhere else.

#### 4.2.3. Mesh extension

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. 9). The sphere  $S_0$  consists of triangles, so we extend each triangle radially as shown in Fig. 12 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 Eq. (12). It is better for the sphere  $S_0$  to be triangulated uniformly since the step length is fixed for each extension step.

$$r_0 + h + 2h + \dots + nh = r_l \implies h = \frac{2(r_l - r_0)}{n(n+1)} \quad (12)$$

where  $n$  is the step number. In Fig. 12, 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, where  $u_1 < u_0$  and  $u_1 < u_2$ . 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 (Figs. 12(b)–(c)) is chosen based on the index number of the three vertices. If  $u_0 < u_2$ , then we choose Fig. 12(b) to split the prism into three tetrahedra. If  $u_2 < u_0$ , then Fig. 12(c) is selected.

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

## 5. Quality improvement

There are two sub-steps in mesh quality improvement:

- (1) Denoising and improving the aspect ratio of the surface mesh (surface vertex adjustment in the normal and the tangent directions).
- (2) Improving the aspect ratio of the volumetric mesh (vertex adjustment inside the volume).

We use geometric partial differential equations (PDEs) to handle the first step. Geometric PDEs, such as the mean curvature flow, the surface diffusion flow and Willmore flow, have been intensively used in surface and imaging processing (Xu et al., 2006). Here we choose surface diffusion flow to smooth the molecular surface because of its volume preserving, and furthermore it approximates spheres accurately (quadratic precision).

$$\frac{\partial x}{\partial t} = \Delta H(x) \vec{n}(x) + v(x) \vec{T}(x), \quad (13)$$

where  $H$  is the mean curvature,  $\vec{n}$  is the unit surface normal vector,  $v(x)$  is the velocity in the tangent direction  $\vec{T}(x)$ , and  $\Delta$  is the Laplace–Beltrami operator.

Eq. (13) is solved over a triangular mesh with vertices  $\{x_i\}$  by discretizing each of its terms. In temporal space,  $\frac{\partial x}{\partial t}$  is approximated by the Euler scheme  $\frac{x_i^{n+1} - x_i^n}{\tau}$ , where  $\tau$  is time step-length.  $x_i^n$  is the approximating solution at  $t = n\tau$ ,  $x_i^{n+1}$  is the approximating solution at  $t = (n+1)\tau$ , and  $x_i^0 = x_i$  serves as the initial value. Discretizing schemes for  $\Delta$  and  $H$  in the spatial space are given in (Xu et al., 2006), we do not go to detail here. Further  $v(x) \vec{T}(x)$  is approximated by

$$[m(x_i^n) - x_i^n] - \vec{n}(x_i^n)^T [m(x_i^n) - x_i^n] \vec{n}(x_i^n), \quad (14)$$

where  $m(x_i^n)$  is defined as the mass center of all the surface triangles incident to  $x_i^n$ . A mass center  $P$  of a region  $V$  is defined by finding  $p \in V$ , such that  $\int_V \|y - p\|^2 d\sigma = \min$ , where  $V$  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 we could derive that

$$m(x_i^n) = \frac{1}{3} x_i^n + \frac{1}{3} \sum_{j \in N(i)} x_j^n (\Delta_j + \Delta_{j+1}) / A(x_i^n), \quad (15)$$

where  $N(i)$  is the index set of the one ring vertex neighbors of  $x_i^n$ ,  $\Delta_j$  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 (Xu et al., 2006), instead of the mass center, however we observed that the mass center works better for biomolecules. The discretization leads to a positive-definite linear system, and the approximate solution is obtained by solving this linear system.

After the molecular surface is 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  $m(p_i) = \frac{1}{4} p_i + \frac{1}{4V_i} \sum_j V_{ij} p_j$ , where  $V_{ij}$  is the volume summation of all the tetrahedra around the edge  $[p_i p_j]$ ,  $V_i$  is the volume summation of the tetrahedra around the vertex  $p_i$ . This is similar in spirit to the multi-linear centroid smoothing scheme (Bajaj et al., 2002).

Fig. 13 shows the difference of the mesh before and after the quality improvement steps. The left column shows the original iso-surface of an ASN-THR-TYR peptide, and the right column shows the results after mesh regularization. It is obvious that after quality improvement, the surface mesh is more regular and has better aspect ratio (twice the ratio of incircle radius to circumcircle radius).

The left picture in Fig. 16 shows the improvement of the aspect ratio, and Figs. 14 and 15 show the improvement in mesh regularization. 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 (Y. Zhang et al., 2005), the edge contraction and linear averaging method was used to improve the quality of tetra meshes with the edge-ratio (the longest edge length over the shortest edge length) and Joe–Liu parameter

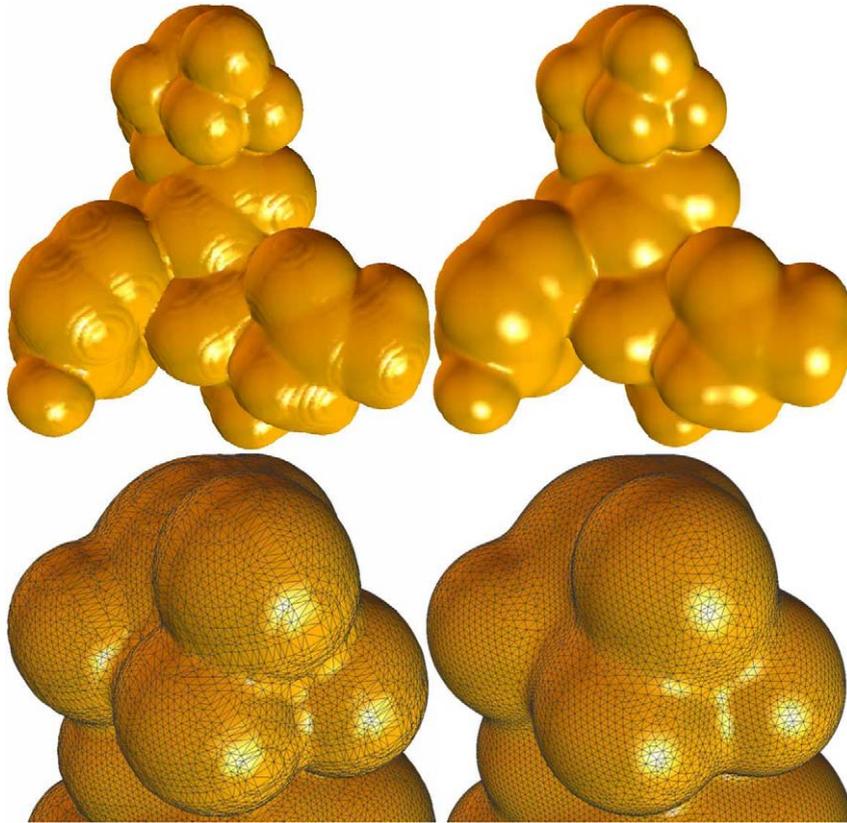


Fig. 13. The surface comparison before/after quality improvement. The left column shows the original surface of an ASN-THR-TYR peptide, and the right column shows the surface after mesh regularization. The top row shows the smooth shading surfaces, and the bottom row shows snapshots of the meshes.

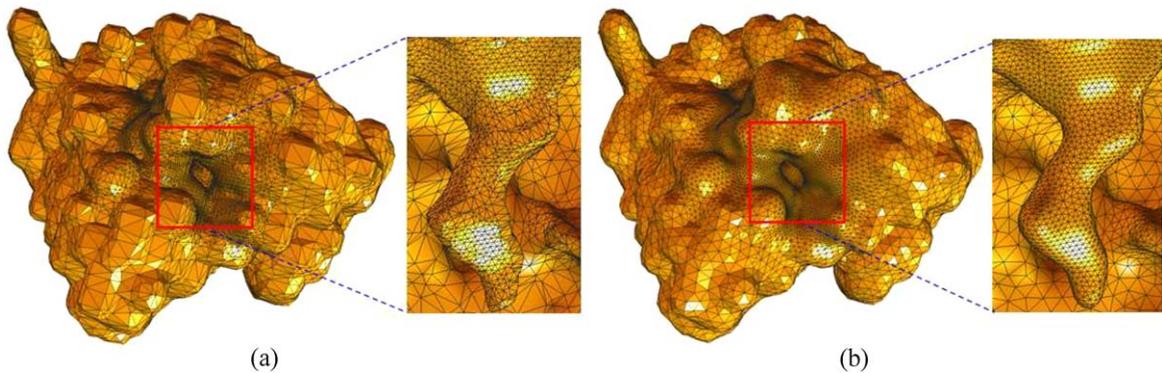


Fig. 14. Comparison of mACHe (9308 vertices, 18612 triangles) before and after surface mesh regularization. (a)—Original; (b)—after mesh regularization.

$(2^{4/3} \times 3 \times (|V|)^{2/3} / \sum_{0 \leq i < j \leq 3} |e_{ij}|^2)$ , where  $|V|$  denotes the volume,  $e_{ij}$  represents the edge connecting vertex  $v_i$  and  $v_j$  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 (Eq. (15)) since it can minimize the energy defined earlier ( $\int_V \|y - p\|^2 d\sigma$ ). From the right picture in Fig. 16, we can see that the worst Joe–Liu parameter is improved significantly after quality improvement. Figs. 17 and 19 show interior tetra meshes of mACHe and Ribosome 30S.

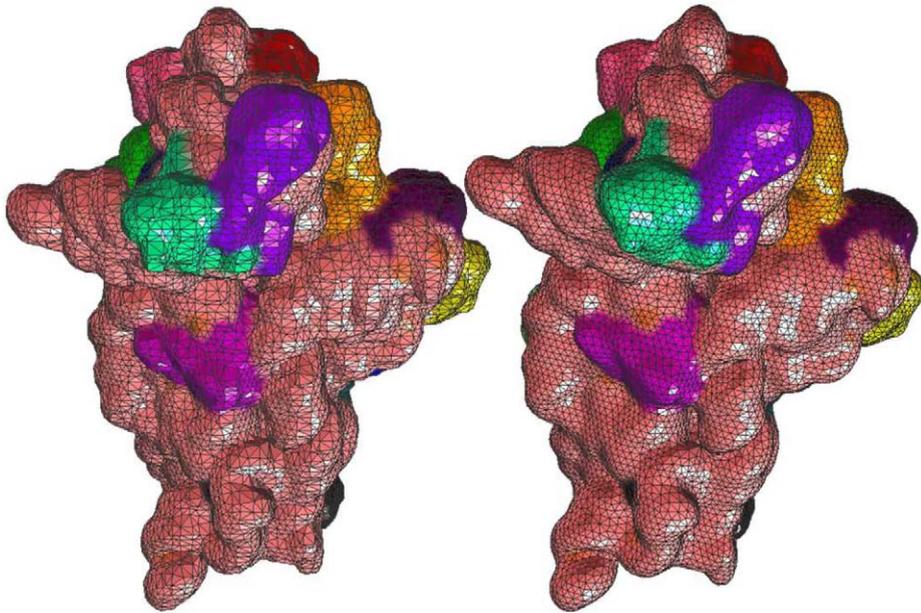


Fig. 15. Comparison of Ribosome 30S (13428 vertices, 26852 triangles) before and after surface mesh regularization. Left—original; right—after mesh regularization.

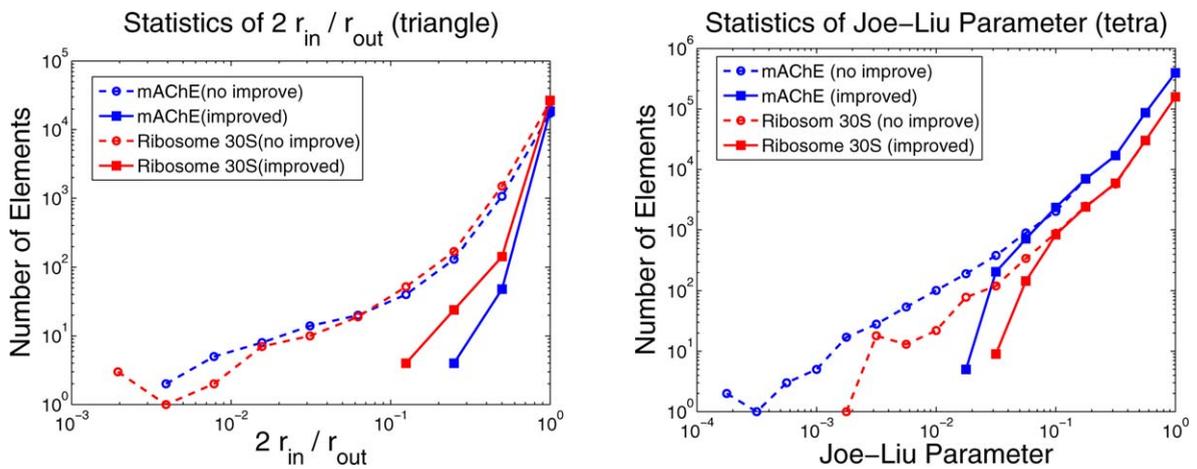


Fig. 16. The histogram of the aspect-ratio and Joe-Liu parameter.

## 6. Results and conclusion

*Monomeric mACHe:* The extracted tetrahedral meshes of the monomer as shown in Fig. 17 have been used in the finite element analysis of the steady-state Smoluchowski equation (SSSE) for diffusion rate constant calculations (Song et al., 2004a, 2004b). Note the adaptive meshes around the narrow gorge region (the active site in mACHe is at the bottom of this gorge).

*Tetrameric mACHe:* We also generated adaptive tetrahedral meshes for the acetylcholinesterase in tetrameric form, with two different arrangement of the monomers. Each monomer has an active site accessible though a long narrow gorge (20 Angstrom), so there are a total of four gorges. Fig. 18 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. Each of the adaptive meshes have finer triangles around the region of the four gorges. These meshes are also used in calculating the diffusion rate constant (D. Zhang et al., 2005).

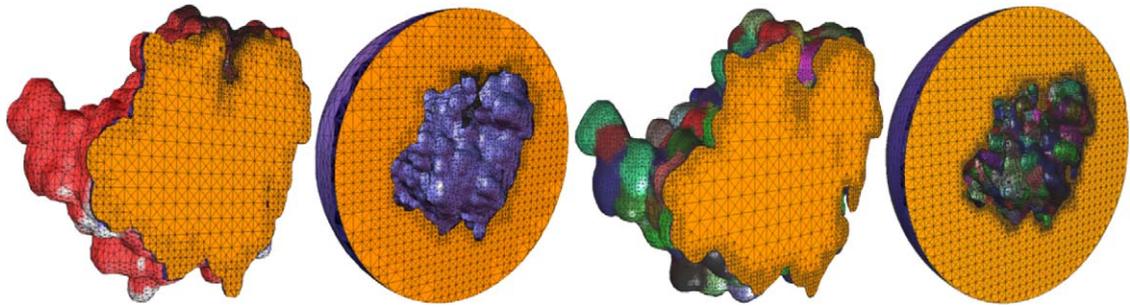


Fig. 17. Interior and exterior tetrahedral meshes of monomeric mAChE. The left two pictures conform to the SAS with inflation  $\sigma = 2$ , and the right two pictures conform to the surface constructed from Gaussian summation with  $p_1 = 0.25$ ,  $p_2 = 1.0$ . From left to right: (65147 vertices, 323442 tets), (121670 vertices, 656823 tets), (103680 vertices, 509597 tets) and (138967 vertices, 707284 tets). The color shows electrostatics potential (the right two). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

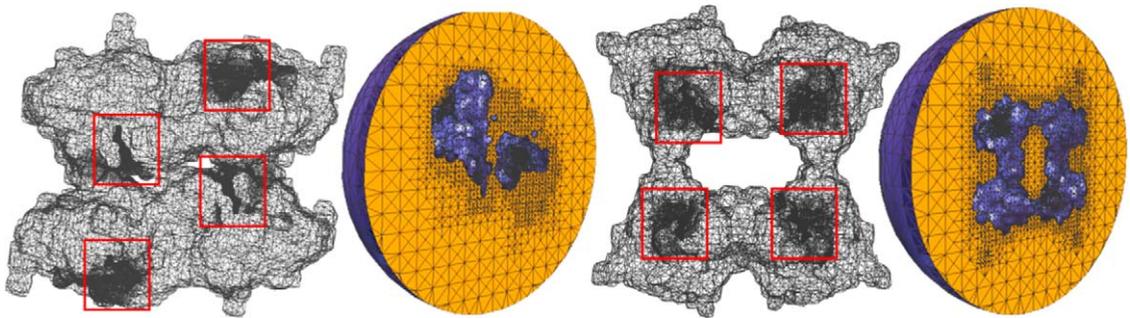


Fig. 18. Interior and exterior tetrahedral meshes of tetrameric mAChE,  $p_1 = 0.5$ ,  $p_2 = 1.0$ . The left two pictures show the 1st crystal structure 1C2O (133078 vertices, 670950 tets), and the right two pictures show the 2nd one 1C2B, (106463 vertices, 551074 tets). Gorges are shown in red boxes. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

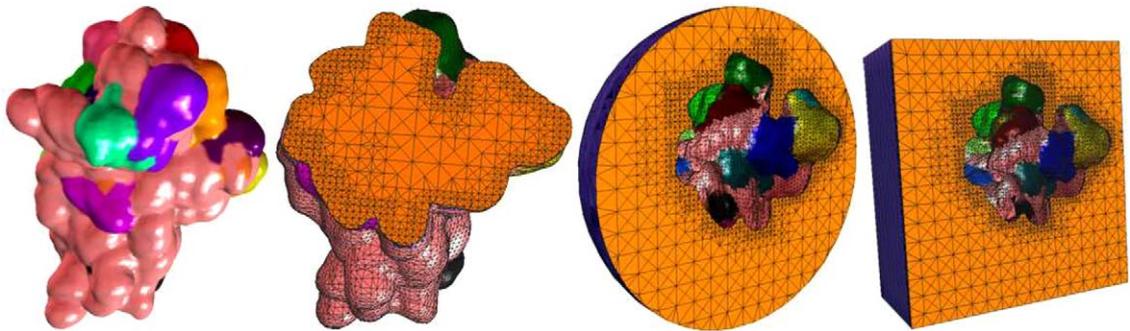


Fig. 19. Interior and exterior tetrahedral meshes of Ribosome 30S, low resolution,  $p_1 = 0.03125$ ,  $p_2 = 1.0$ . From left to right: (33612 vertices, 163327 tets), (37613 vertices, 186496 tets) and (40255 vertices, 201724 tets). The pink color shows 16S rRNA and other colors show proteins. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

**Ribosome:** Ribosomes are macromolecular complexes responsible for the translation of mRNA into proteins. These complexes consist of two subunits: the larger 50S and the smaller 30S, both of the subunits are composed of rRNA and protein constituents. Atomic level, residue-level and low resolution structure models were constructed from density maps as shown in Figs. 3 and 5. The constructed exterior meshes are being used for the finite element solution of the Poisson–Boltzmann equation (Baker et al., 2001). Figs. 19 and 1 show interior and exterior meshes of the Ribosome 30S/50S.

We have developed a quality molecular meshing approach directly from PDB molecular structural data, with adaptivity at prescribed active sites on the molecular surface. Our generated meshes continue to be used in several boundary/finite element biophysics simulations (Song et al., 2004a, 2004b; D. Zhang et al., 2005).

## Acknowledgements

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 molecular structures. We also thank Dr. Yuhua Song, Dr. Deqiang Zhang, Prof. Nathan Baker and Prof. Andrew McCammon for using and helping us calibrate our LBIE-Mesher for biomolecules. The work at University of Texas was supported in part by NSF-ITR grants CNS-0540033, EIA-0325550 and grants from the NIH OP20 RR020647 and R01-GM074258, R01-GM073087. A substantial part of the work on this paper was done when Guoliang Xu was visiting Chandrajit Bajaj at UT-CVC. His work was partially supported by the aforementioned grants, the J.T. Oden ICES visitor fellowship and in part by Natural Science Foundation of China (10371130) and National Key Basic Research Project of China (2004CB318000).

## References

- Agarwal, R.C., 1978. A new least-squares refinement technique based on the fast Fourier transform algorithm. *Acta Cryst. A* 34, 791–809.
- Akkiraju, N., Edelsbrunner, H., 1996. Triangulating the surface of a molecule. *Discrete Applied Mathematics* 71, 5–22.
- Bajaj, C., Castrillon-Candas, J., Siddavanahalli, V., Xu, Z., 2005. Compressed representations of macromolecular structures and properties structure. *Discrete Applied Mathematics* 13 (3), 463–471.
- Bajaj, C., Lee, H., Merkert, R., and Pascucci, V., 1997. NURBS based B-rep Models from Macro-molecules and their properties. In: *Proceedings 4th Symposium on Solid Modeling and Applications*, pp. 217–228.
- Bajaj, C., Schaefer, S., Warren, J., Xu, G., 2002. A subdivision scheme for hexahedral meshes. *The Visual Computer* 18 (5–6), 343–356.
- Bajaj, C., Pascucci, V., Shamir, A., Holt, R., Netravali, A., 2003. Dynamic maintenance and visualization of molecular surfaces. *Discrete Applied Mathematics* 127, 23–51.
- Bajaj, C., Siddavanahalli, V., 2005. Fast and adaptive surfaces and derivatives computations for volumetric particle data. CS and ICES Technical Report, the University of Texas at Austin.
- Baker, N., Sept, D., Joseph, S., Hoist, M., McCammon, J., 2001. Electrostatics of nanosystems: application to microtubules and the ribosome. *Proc. Natl. Acad. Sci.* 98, 10037–10041.
- Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., Shindyalov, I.N., Bourne, P.E., 2000. The protein data bank. *Nucleic Acids Research* 28, 235–242.
- Blinn, J., 1982. A generalization of algebraic surface drawing. *ACM Transactions on Graphics* 1 (3), 235–256.
- Bourne, Y., Taylor, P., Marchot, P., 1995. Acetylcholinesterase inhibition by fasciculon: crystal structure of the complex. *Cell* 83, 503.
- Bourne, Y., Grassi, J., Bougis, P.E., Marchot, P., 1999. Conformational flexibility of the acetylcholinesterase tetramer suggested by x-ray crystallography. *Journal of Biological Chemistry* 274 (43), 30370–30376.
- Canann, S., Tristano, J., and Staten, M., 1998. An approach to combined Laplacian and optimization-based smoothing for triangular, quadrilateral and quad-dominant meshes. In: *7th International Meshing Roundtable*, pp. 419–494.
- Cheng, H.-L., Dey, T., Edelsbrunner, H., Sullivan, J., 2001. Dynamic skin triangulation. *Discrete Computational Geometry* 25, 525–568.
- Cheng, S.-W., Dey, T., Edelsbrunner, H., Facello, M., Teng, S., 2000. Sliver exudation. *Proceeding Journal of ACM* 47, 883–904.
- Chew, P., 1997. Guaranteed-quality Delaunay meshing in 3d. In: *ACM Symposium on Computational Geometry*, pp. 391–393.
- Connolly, M.L., 1983. Analytical molecular surface calculation. *J. Appl. Crystallogr.* 16, 548–558.
- Edelsbrunner, H., 1999. Deformable smooth surface design. *Discrete Computational Geometry* 21, 87–115.
- Field, D., 1988. Laplacian smoothing and Delaunay triangulations. *Communications in Applied Numerical Methods* 4, 709–712.
- Freitag, L., 1997. On combining Laplacian and optimization-based mesh smoothing techniques. *Trends in Unstructured Mesh Generation AMD* 220, 37–43.
- Frey, P., Borouchaki, H., George, P.-L., 1996. Delaunay tetrahedralization using an advancing-front approach. In: *5th International Meshing Roundtable*, pp. 31–48.
- Fujishiro, I., Maeda, Y., Sato, H., Takeshima, Y., 1996. Volumetric data exploration using interval volume. *IEEE Transactions on Visualization and Computer Graphics* 2 (2), 144–155.
- Gabdoulline, R.R., Wade, R.C., 1996. Analytically defined surfaces to analyze molecular interaction properties. *Journal of Molecular Graphics* 14 (6), 341–353.
- Garland, M., Heckbert, P., 1998. Simplifying surfaces with color and texture using quadric error metrics. In: *IEEE Visualization*, pp. 263–270.
- Gibson, S., 1998. Using distance maps for accurate surface representation in sampled volumes. In: *IEEE Sympos. on Volume Visualization*, pp. 23–30.
- Grant, J., Pickup, B., 1995. A Gaussian description of molecular shape. *Journal of Physical Chemistry* 99, 3503–3510.
- Holst, M., Baker, N., Wang, F., 2000. Adaptive multilevel finite element solution of the Poisson–Boltzmann equation i: Algorithms and examples. *Journal of Computational Chemistry* 21, 1319–1342.
- Ju, T., Losasso, F., Schaefer, S., Warren, J., 2002. Dual contouring of Hermite data. In: *SIGGRAPH*, pp. 339–346.

- Kobbelt, L., Botsch, M., Schwanecke, U., Seidel, H., 2001. Feature sensitive surface extraction from volume data. In: SIGGRAPH, pp. 57–66.
- Laug, P., Borouchaki, H., 2002. Molecular surface modeling and meshing. *Engineering with Computers* 18, 199–210.
- Lee, B., Richards, F., 1971. The interpretation of protein structures: Estimation of static accessibility. *Journal Molecular Biology* 55, 379–400.
- Lee, M., Feig, M., Salsbury, F., Brooks, C., 2003. New analytic approximation to standard molecular volume definition and its application to generalized born calculation. *Journal of Computational Chemistry* 24, 1348–1356.
- Lohner, R., Parikh, P., 1988. Three dimensional grid generation by the advancing-front method. *International Journal for Numerical Methods in Fluids* 8, 1135–1149.
- Lohner, R., Morgan, K., Zienkiewicz, O., 1986. Adaptive grid refinement for compressible Euler equations. In: Babuska, I., et al. (Eds.), *Accuracy Estimates and Adaptive Refinements in Finite Element Computations*. Wiley, pp. 281–297.
- Lorensen, W., Cline, H., 1987. Marching cubes: A high resolution 3d surface construction algorithm. In: SIGGRAPH, pp. 163–169.
- Max, N.L., 1988. Approximating molecular surfaces by spherical harmonics. *J. Mol. Graphics* 6 (4), 210.
- Meyer, M., Desbrun, M., Schröder, P., Burr, A., 2002. Discrete differential-geometry operators for triangulated 2-manifolds. In: *VisMath'02*, Berlin.
- Mezey, P.G., 1993. *Shape in Chemistry: An Introduction to Molecular Shape and Topology*. VCH Publishers, New York.
- Nielson, G., Sung, J., 1997. Interval volume tetrahedrization. In: *IEEE Visualization*, pp. 221–228.
- Nissen, P., Hansen, J., Ban, N., Moore, P.B., Steitz, T.A., 2000. The structural basis of ribosome activity in peptide bond synthesis. *Science* 289, 920–930.
- Ohtake, Y., Belyaev, A., Bogaevski, I., 2001. Mesh regularization and adaptive smoothing. *Computer Aided Design* 33, 789–800.
- Owen, S., 1998. A survey of unstructured mesh generation technology. In: *7th Internat. Meshing Roundtable*, pp. 26–28.
- Sanner, M., Olson, A., Spehner, J., 1996. Reduced surface: An efficient way to compute molecular surfaces. *Biopolymers* 38, 305–320.
- Sapiro, G., 2001. *Geometric Partial Differential Equations and Imaging Analysis*. Cambridge University Press, Cambridge.
- Shephard, M., Georges, M., 1991. Three-dimensional mesh generation by finite octree technique. *International Journal for Numerical Methods in Engineering* 32, 709–749.
- Shewchuk, J., 2002. Constrained Delaunay tetrahedrizations and provably good boundary recovery. In: *11th International Meshing Roundtable*, pp. 193–204.
- Shimada, K., Yamada, A., Itoh, T., 1997. Anisotropic triangular meshing of parametric surfaces via close packing of ellipsoidal bubbles. In: *6th International Meshing Roundtable*, pp. 375–390.
- Song, Y., Zhang, Y., Bajaj, C., Baker, N., 2004a. Continuum diffusion reaction rate calculations of wild type and mutant mouse acetylcholinesterase: Adaptive finite element analysis. *Biophysical Journal* 87 (3), 1558–1566.
- Song, Y., Zhang, Y., Shen, T., Bajaj, C., McCammon, J., Baker, N., 2004b. Finite element solution of the steady-state Smoluchowski equation for rate constant calculations. *Biophysical Journal* 86 (4), 2017–2029.
- Teng, S.-H., Wong, C., 2000. Unstructured mesh generation: Theory, practice and perspectives. *International Journal of Computational Geometry and Applications* 10 (3), 227–266.
- Wood, Z., Breen, D., Desbrun, M., 2000. Semi-regular mesh extraction from volumes. In: *IEEE Visualization*, pp. 275–282.
- Xu, G., 2006. Discrete Laplace–Beltrami operator on sphere and optimal spherical triangulations. *International Journal of Computational Geometry and Application* 16 (1), 75–93.
- Xu, G., Pan, Q., Bajaj, C., 2006. Discrete surface modelling using partial differential equations. *Computer Aided Geometric Design* 23 (2), 125–145.
- Zhang, D., Suen, J., Zhang, Y., Song, Y., Radic, Z., Taylor, P., Hoist, M., Bajaj, C., Baker, N., McCammon, J., 2005. Tetrameric mouse acetylcholinesterase: Continuum diffusion rate calculations by solving the steady-state Smoluchowski equation using finite element methods. *Biophysical Journal* 88 (3), 1659–1665.
- Zhang, Y., Bajaj, C., Sohn, B.-S., 2003. Adaptive and quality 3d meshing from imaging data. In: *ACM Solid Modeling and Applications*, pp. 286–291.
- Zhang, Y., Bajaj, C., Sohn, B.-S., 2005. 3d finite element meshing from imaging data. *Special issue of Computer Methods in Applied Mechanics and Engineering on Unstructured Mesh Generation* 194 (48–49), 5083–5106.