

Parameterization for molecular Gaussian surface and a comparison study of surface mesh generation

Tiantian Liu¹ · Minxin Chen² · Benzhuo Lu¹

Received: 23 December 2014 / Accepted: 16 March 2015
© Springer-Verlag Berlin Heidelberg 2015

Abstract The molecular Gaussian surface has been frequently used in the field of molecular modeling and simulation. Typically, the Gaussian surface is defined using two controlling parameters; the decay rate and isovalue. Currently, there is a lack of studies in which a systematic approach in the determination of optimal parameterization according to the geometric features has been done. In this paper, surface area, volume enclosed by the surface and Hausdorff distance are used as three criteria for the parameterization to make the Gaussian surface approximate the solvent excluded surface (SES) well. For each of these three criteria, a search of the parameter space is carried out in order to determine the optimal parameter values. The resulted parameters are close to each other and result in similar calculated molecular properties. Approximation of the VDW surface is also done by analyzing the explicit expressions of the Gaussian surface and VDW surface, which analysis and parameters can be similarly applied to the solvent accessible surface (SAS) due to its geometric similarity to the VDW surface. Once the optimal parameters are obtained, we compare the performance of our Gaussian surface generation software TMSmesh with other commonly used software programs, focusing primarily on mesh quality

and fidelity. Additionally, the Poisson-Boltzmann solvation energies based on the surface meshes generated by TMSmesh and those generated by other software programs are calculated and compared for a set of molecules with different sizes. The results of these comparisons validate both the accuracy and the applicability of the parameterized Gaussian surface.

Keywords Molecular Gaussian surface · Parameterization · Mesh quality · Meshing software · Solvation energy

Introduction

Molecular surface plays a very important role in computational structural biology and chemistry, such as in visualization, protein folding and structure prediction, docking and implicit solvent modeling. In the field of implicit solvent modeling and continuum modeling, the molecular surface is used to model the dielectric interface between the low-dielectric solute and the high-dielectric solvent. Considering the utility of knowing the molecular surface, research in accurately representing the molecular surface has been ongoing.

Various definitions of molecular surface exist, including the van der Waals (VDW) surface, the solvent accessible surface (SAS) [1], the solvent excluded surface (SES) [2], the molecular skin surface [3], the minimal molecular surface [4] and the Gaussian surface. The VDW surface is defined as the surface of the union of the spherical atomic surfaces with VDW radius of each atom within the molecule. The VDW surface has the advantage of allowing for the analytical calculation of the corresponding area as well as allowing for the determination of the

✉ Benzhuo Lu
bzlu@lsec.cc.ac.cn

¹ State Key Laboratory of Scientific and Engineering Computing, National Center for Mathematics and Interdisciplinary Sciences, Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing 100190, China

² Center for System Biology, Department of Mathematics, Soochow University, Suzhou 215006, China

normal directions of the surface. However, the VDW surface model contains numerous voids too small for discrete water molecules to occupy. The SAS is defined as the surface traced by the trajectory of the center of a rolling probe on the VDW surface. The size of the solvent molecules, which is typically water, is chosen as the radius for the probe. Geometrically, the SAS is equivalent to the VDW surface of the system whose VDW radius is increased by the size of the radius of the probe. The SES, also referred to as the Lee-Richards surface, is a highly utilized molecular surface model and is represented by the locus of the interboundary of the rolling probe. The SES has less crevices and surface invaginations as compared to the VDW surface. The molecular skin surface, proposed by Edelsbrunner [3], is an implicit molecular surface based on a framework of the Voronoi diagram, Delaunay triangulation and alpha complexes of a finite set of weighted points with a shrink factor $s \in [0, 1]$. The molecular skin surface can be decomposed into a collection of quadratic patches. Each patch is a portion of sphere or hyperboloid clipped within a polyhedron obtained by shrinking the Minkowski sum of the corresponding Voronoi and Delaunay polyhedron. The molecular skin surface is smooth, tangent continuous and free of self-intersections. The minimal molecular surface is defined as the result of the minimization of a type of surface free energy by minimizing mean curvature of a defined hypersurface [4]. The hypersurface function is defined with atomic constraints or obstacles from biomolecular structural information. The minimal molecular surface is probe independent, differentiable, and typically free of singularities.

Different from the previous definitions, the Gaussian surface is defined as a level set of the summation of the Gaussian kernel functions; for which descriptions of the specific forms will be given in the next section. The Gaussian surface has been widely used in molecular geometry calculation, visualization, and biophysics; such as docking problems [5], molecular shape comparisons [6], calculating SAS areas [7], solvation energy calculations based on the generalized Born models [8] and Poisson-Boltzmann model [9], and simulation of ion transport in ion channel [10]. In some PB calculations, a smooth Gaussian-based dielectric function are used [11, 12], which in some sense has as similar influence as using a Gaussian surface because a surface in PB calculation is only used to distinguish the high and low dielectric regions. The Gaussian surface is smooth and provides a realistic representation of the electron density of a molecule as compared to other molecular surface definitions [13]. The Gaussian surface is an implicit surface in which the geometric shapes of the surface are controlled by two parameters (see Eqs. (1)–(2)), the decay rate and the isovalue. The decay rate controls the rate of decay of each atom's Gaussian kernel and the isovalue controls

the volume enclosed by the Gaussian surface. The VDW surface, SAS and SES can all be approximated well by the Gaussian surface given the proper parameters.

A variety of methods have been proposed to compute molecular surface. A method for the analytical computation of the SAS and SES was proposed in 1983 by Connolly [14, 15]. GRASP, proposed by Nicholls, is a popular program used for the visualization of molecular surfaces [16]. The software MSMS was proposed by Sanner et al. in 1996 to mesh the SES and is a widely used program for molecular surface triangulation due to its high efficiency [17]. Both the SIMS [18] and LSMS [19] software were later proposed to compute the SES. EDTsurf, a program used for generating the three major macromolecular surfaces; the VDW surface, SES and SAS was developed based on LSMS in 2009 [20]. A ray-casting-based software, NanoShaper, is proposed to generate SES, skin surface and Gaussian surface in 2013 [21].

Some methods for meshing the Gaussian surface have been proposed. A two-level clustering technique to generate meshes for biomolecular structures was proposed in 2006 [22]. A later tool, GAMer, was developed for both the generation and improvement of Gaussian surface meshes [23]. The software MolSurf was designed to generate and manipulate various molecular surfaces including the Gaussian surface and other molecular surfaces [24, 25]. An efficient mesh generation algorithm accelerated by multi-core CPU and GPU was also recently proposed in 2013 [26]. We recently developed a program called TMSmesh that has the capability to generate manifold surface meshes for arbitrarily large molecular systems [9, 27]. TMSmesh utilizes the trace technique, which is a generalization of predictor corrector technique. The algorithm contains two stages. The first stage is to compute the intersecting points between the molecular Gaussian surface and the lines parallel to x -axis. In this stage, the molecule is placed in a three-dimensional orthogonal grid. The upper and lower bounds of the Gaussian kernel function in each box of the grid is estimated in order to rule out the boxes having no surface points. In the remaining boxes, the intersecting points between the surface and the lines parallel to x -axis are found through root finding algorithms. The second stage is to polygonize the Gaussian surface by connecting the generated surface points. The sampled surface points are connected through technique of adaptive continuation to form loops, and the whole closed manifold surface is decomposed into a collection of patches enclosed by loops on the surface. These patches are often non-single valued along at least one of x, y, z directions, and may contain holes and tunnels. In the step of triangulating the non-single valued patches, these holes and tunnels may be missed, and intersections may occur. To avoid these errors, we finally dissect each patch enclosed by loop into single valued pieces in x, y, z

direction through fold curves and then do the triangulations. In TMSmesh issues such as overlapping, gap filling, and the selection of initial seeds which are present in traditional continuation methods are not found because the Gaussian surface is polygonized by connecting presampled surface points [27]. It is of note that TMSmesh succeeded to generate a surface mesh for biomolecules comprised of more than one million atoms on a PC.

These software packages based on different surface definitions and meshing methods generate acceptable meshes for visualization purposes, but lead to differences in the performance of numerical calculation. When we develop a meshing software for numerical modeling, the major concerns are robustness, efficiency and mesh quality. Robustness here means that the meshing method is stable and can handle various, even arbitrary sizes of molecular systems within computer power limitations. Efficiency is necessary for simulations and computations that require frequent mesh generation or meshes of large systems. Mesh quality here not only refers to the uniformness but also to the manifoldness and faithfulness. Uniformness means that the mesh should avoid the elements with very sharp angles or very large/small areas. A few properties that are used to measure the uniformness of a triangular mesh include the ratio of the shortest edge over the longest edge in each triangle of the mesh and the distribution of angles. A prerequisite in achieving convergence for finite element method (FEM) in implicit solvent modeling is that the mesh should satisfy certain criteria for uniformness. The manifoldness of a surface means that each point on the surface has a neighborhood which is homeomorphic to a disk in a real plane. Meshing a manifold surface should also produce a manifold mesh, and in turn a manifold mesh means that the surface formed by all of the elements of the mesh is also a manifold. A non-manifold mesh can lead to numerical problems in boundary element method (BEM) and FEM type simulations of biomolecules; for example, a volume mesh generation that conforms to the surface may fail due to a non-manifold surface mesh. Faithfulness here is measured by how accurately the surface mesh preserves the original geometry and topology, such as surface area, volume and curvature of the referenced molecular surface. Faithfulness is a basic requirement when calculating the molecular physical chemical properties.

We have previously shown that TMSmesh, our triangular mesh generation software, is a robust tool for meshing the Gaussian surface for arbitrarily large biomolecules. The meshes generated by TMSmesh are manifold meshes and are applicable to BEM/FEM simulations of biomolecular electrostatics. However, despite the progress that has been made in the development of these mesh generation software programs and considering the potential applications of Gaussian surfaces, there is a lack of detailed

and systematic studies on the issue of parameterization of the Gaussian surface. Specifically, that means how to choose the two parameters in the definition of the Gaussian surface to approximate a specific type of traditionally defined molecular surface in terms of some concerned geometric and biophysical properties. The work in [28] reported some studies on Gaussian surface by choosing different values of one parameter, the decay rate, in the Gaussian kernel. But, to our knowledge, there still lacks of detailed parameterization study of Gaussian surface by searching the full two-parameter space. The focus of this paper is to determine the optimal parameters based on geometric characteristic so that the resulted surface mesh generated by TMSmesh is faithful to the SES and VDW surface (and to the SAS as well which is geometrically similar to the VDW surface). By definition, the VDW surface (and similarly, the SAS) can be easily obtained from a Gaussian surface by setting the decay rate parameter to a sufficiently large value (as shown in this paper, $d = 2.0$ is enough). We then use three global properties, the surface area, the volume enclosed by the surface and the Hausdorff distance as three criteria to determine how well the Gaussian surface approximates the SES. In principle, if the Hausdorff distance is infinitely small, the two surfaces can be considered identical. However, because Gaussian surface and SES are two different types of molecular surfaces, there is no way by choosing parameters to make their Hausdorff distance equal to zero or infinitely small. Therefore, we also choose other two criteria, area and volume, for a more extended exploration. An issue is that a same area or volume value does not correspond to a unique surface shape, which means there exist ambiguities in surface determination through the parameterization. In fact, the local information of each atomic coordinate and the VDW radius has already been contained in the Gaussian surface definition, which can be considered as a constraint and hence largely reduces the ambiguities. Comparisons between the performances of TMSmesh and other commonly used molecular surface meshing software are also done. In addition, the parameterized Gaussian surfaces are also used in PB solvation energy calculations, and the performances are demonstrated as well.

In the following section we define the Gaussian surface and present the methods to compute the surface area, the volume and the Hausdorff distance. The parameter space for each criterion and the results of the parameterization are shown in the section of Results and Discussion. Additionally, using the optimal parameters we were able to assess and compare the performance of surface meshes generated by TMSmesh and other meshing software; including MSMS, Molsurf, NanoShaper, GAMer and EDTsurf. Finally, we discuss the overall conclusions and implications of our results.

Methods

The Gaussian surface is defined as a level set of the summation of the Gaussian kernel functions with two parameters:

$$\{\vec{x} \in R^3, \phi(\vec{x}) = c\}, \quad (1)$$

where

$$\phi(\vec{x}) = \sum_{i=1}^N e^{-d(\|\vec{x}-\vec{x}_i\|^2-r_i^2)}, \quad (2)$$

\vec{x}_i and r_i are the location and radius of the i th atom. d and c are the two parameters that need to be set in the Gaussian surface. d is the decay rate of the Gaussian kernel. As d decreases, the surface becomes smoother and more inflated. c is the isovalue and controls the volume enclosed by the Gaussian surface. In several previous publications, c has been set as 1.0 [22, 29, 30]. In the following parameterization process in selecting optimal values of d and c , the searching scope of d is set at {0.3, 0.4, ..., 0.9, 1.0} and the searching scope of c at {0.8, 0.9, ..., 2.4, 2.5}.

The purpose of the parameterization process is to find the optimal parameters for the Gaussian surface to approximate the SES and VDW surface. TMSmesh is used to mesh the Gaussian surface and MSMS is used to mesh the SES and VDW surface. The area of the surface, the volume enclosed by the surface and the Hausdorff distance are the three criteria used to judge whether two molecular surfaces are close enough. For a triangular surface mesh, the surface area S is determined using the following equation

$$S = \frac{1}{2} \sum_{i=1}^m \left\| \overrightarrow{A_2^i A_1^i} \times \overrightarrow{A_3^i A_1^i} \right\|, \quad (3)$$

where m is the number of triangle elements and A_1^i, A_2^i, A_3^i denote the coordinates of the three vertices for the i th triangle. The volume V enclosed by the surface mesh is determined using the following equation

$$V = \frac{1}{6} \sum_{i=1}^m \overrightarrow{A_2^i A_1^i} \times \overrightarrow{A_3^i A_1^i} \bullet \vec{c}_i, \quad (4)$$

where \vec{c}_i is the vector from the center of the i th triangle to the origin.

The Hausdorff distance between two surface meshes is defined as follows.

$$H(S_1, S_2) = \max \left(\max_{p \in S_1} e(p, S_2), \max_{p \in S_2} e(p, S_1) \right), \quad (5)$$

where

$$e(p, S) = \min_{p' \in S} d(p, p'). \quad (6)$$

S_1 and S_2 are two piecewise surfaces spanned by the two corresponding meshes, and $d(p, p')$ is the Euclidean distance between points p and p' . In our work, we use Metro [31] to compute the Hausdorff distance.

The method used in parameterization for the Gaussian surface approximating SES is described as follows. A set of biomolecules taken from the RCSB Protein Data Bank and some small molecules in our former studies (the benchmark can be downloaded from www.continuummodel.org) with different sizes are chosen as a benchmark set (see Table 1). The SES areas and volumes computed from MSMS meshes are taken as references. For each set of parameters, the corresponding Gaussian surface areas and volumes are computed from the meshes generated by TMSmesh, and then compared with those from MSMS meshes using the two relative errors, which are calculated using the following formulas

$$Area_Error = \sum_{i=1}^{24} \frac{|A_i^{MSMS} - A_i^{TMS}|}{A_i^{MSMS}}, \quad (7)$$

$$Volume_Error = \sum_{i=1}^{24} \frac{|V_i^{MSMS} - V_i^{TMS}|}{V_i^{MSMS}}, \quad (8)$$

where i is the index of biomolecules shown in Table 1, A_i^{MSMS} and A_i^{TMS} denote the corresponding surface areas of meshes generated by MSMS and TMSmesh respectively. V_i^{MSMS} and V_i^{TMS} denote the volumes enclosed by the surface meshes that are generated by MSMS and TMSmesh respectively. Furthermore, for benchmark biomolecules, the average Hausdorff distance is also used to compare the meshes computed by TMSmesh and MSMS:

$$\bar{H} = \frac{1}{24} \sum_{i=1}^{24} H(S_i^{MSMS}, S_i^{TMS}), \quad (9)$$

where i is the index of biomolecules shown in Table 1, S_i^{MSMS} denotes the surface mesh generated by MSMS and S_i^{TMS} denotes the surface mesh generated by TMSmesh for the i th biomolecule. Using the calculated relative errors or average distances, the parameters corresponding to the minimal error or distance are taken as the optimal parameters.

Results and discussion

The results of the parameterization of the Gaussian surface in the approximation of SES using three different criteria, area, volume and Hausdorff distance are presented in the first part of this section. A set of parameters is found for each criterion by minimizing the corresponding total relative error or Hausdorff distance. A set of parameters for the approximation of the VDW surface is also given by analyzing the analytical expressions of the Gaussian surface and

Table 1 Number of atoms for 24 test proteins

Index	1	2	3	4	5	6	7	8
Molecule	GLY	diALA	adp	1PLX	3SGS	1OEH	2Y29	2M6C
N_{atom}	8	20	39	75	94	103	116	124
Index	9	10	11	12	13	14	15	16
Molecule	2JP8	1N9V	2B0Y	1ALE	1EDN	2FLY	2Y3L	1B4I
N_{atom}	126	146	285	294	328	355	444	470
Index	17	18	19	20	21	22	23	24
Molecule	1VA3	1BY6	2JM0	1BWX	2O3M	bdna	bpti	fas2
N_{atom}	505	557	589	643	714	778	882	906

the VDW surface. In the second part of this section, the performance of different meshing software; including MSMS, TMSmesh, Molsurf, NanoShaper, GAMer and EDTsurf is studied and compared in mesh quality aspects such as manifoldness, faithfulness and uniformness and in a biophysical application aspect, i.e. solvation calculation.

Parameterization results to approximate SES

The searching scope of the parameter d is set at $\{0.3, 0.4, \dots, 0.9, 1.0\}$ during the process of parameterization and the searching scope of c is set at $\{0.8, 0.9, \dots, 2.4, 2.5\}$. Hence, there are 144 sets of parameters for enumerating. For each set of parameters, we can get the total area relative error, the volume relative error and the Hausdorff distance which are defined in the Method Section. For ease of representation, we convert the relative error plot dimension from 3D to 2D. Therefore, it is necessary to define the abscissa of the new plot; parameter index. The parameter index is defined as:

$$I = 18(10d - 3) + (10c - 8) + 1. \quad (10)$$

The scope of the parameter index is between 1 and 144, where 144 is the number of (d, c) pairs. Within the parameter ranges, each value of the parameter index is corresponding to a (d, c) pair. All the (d, c) pairs form a 8×18 matrix. Each row of the matrix corresponds to a d value and different c values, and each column of the matrix corresponds to a c value and different d values. We map the two-dimensional parameter pair matrix to a one-dimensional index array by sorting the elements of the matrix row-wise and the resulting indices of the (d, c) pairs in the one dimensional array are defined by Eq. (10).

Figure 1a shows the relative errors of area between the SES and the Gaussian surface as a function of the parameter index for the benchmark biomolecules listed in Table 1. The 29th set of parameters, $d = 0.4, c = 1.8$, minimizes the area relative error. The total relative errors of volume between Gaussian surface and SES as a function of parameter index is given in Fig. 1b. The 64th set of parameters,

$d = 0.6, c = 1.7$, minimizes the relative error of the volume and is determined to be the optimal parameter set. In Fig. 1c, the 17th set of parameters, $d = 0.3, c = 2.4$, makes the average Hausdorff distance minimal. However, this set of parameter is not reasonable. For instance, for a Hydrogen atom, if we take $d = 0.3, c = 2.4$ and r as the VDW radius 1.2 \AA , the Gaussian surface is defined as $\{x \in R^3, e^{-0.3(\|\vec{x} - \vec{x}_i\|^2 - 1.2^2)} = 2.4\}$, and this is an empty set, since the maximum of $e^{-0.3(\|\vec{x} - \vec{x}_i\|^2 - 1.2^2)}$ is less than 2.4. Therefore, the best choice is the 23th set of parameter, $d = 0.4, c = 1.2$. Additionally, it is worth mentioning that the 64th set of parameter, $d = 0.6, c = 1.7$, is also a local minimum point in Fig. 1c and makes the Hausdorff distance tiny.

Parameterization results to approximate VDW surface (and the SAS as well)

The explicit formula of the VDW surface is

$$\left\{ \vec{x} \in R^3, \min_{i=1, \dots, N} \frac{\|\vec{x} - \vec{x}_i\|}{r_i} = 1 \right\}, \quad (11)$$

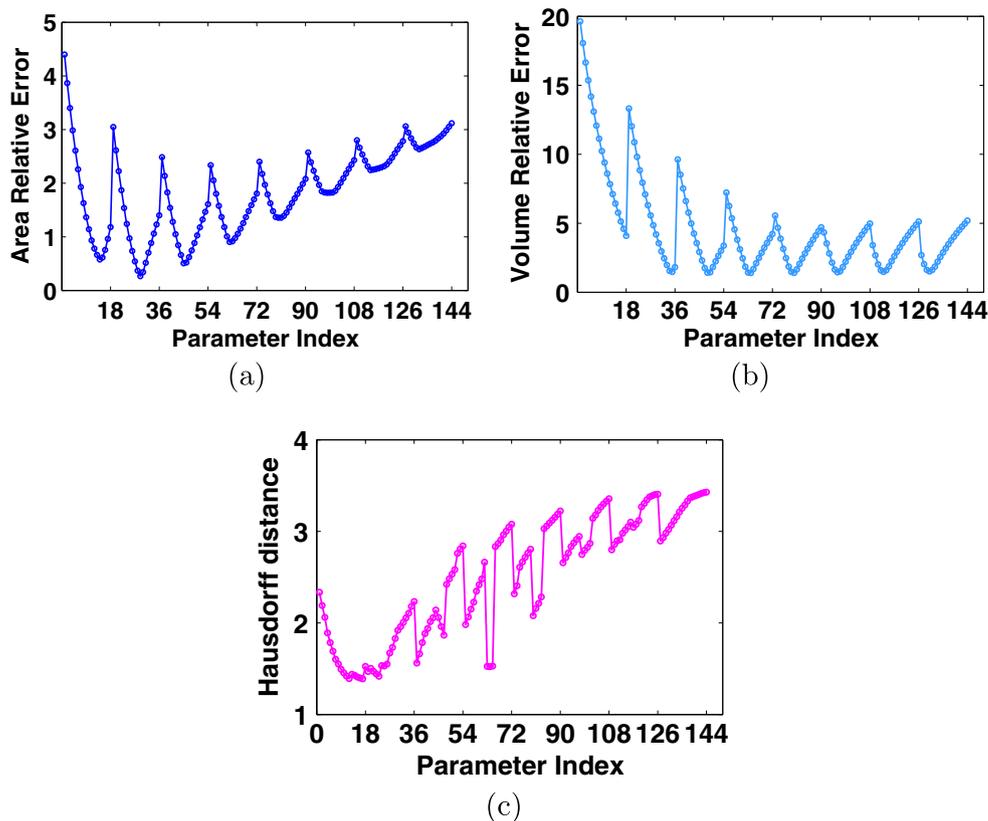
where N is the number of atoms, \vec{x}_i and r_i are the location and radius of the i th atom. From Eqs. (2) and (11), it is easy to see that the Gaussian surface is approaching the VDW surface when $c = 1$ and $d \rightarrow \infty$.

Figure 2 shows the Gaussian surface meshes of the fas2 molecule with different d s when c is 1.0. Increases in d can lead subsequently to a decrease in inflation. To compare the meshes of the Gaussian surface and the VDW surface, the following average distance is defined and determined as

$$D = \frac{1}{M} \sum_{k=1}^M \left(\min_{i=1, \dots, N} \frac{\|\vec{v}_k - \vec{x}_i\|}{r_i} - 1 \right), \quad (12)$$

where M is the number of vertices in the surface mesh, N is the number of atoms of the biomolecule and \vec{v}_k represents the coordinates of the k th vertex in the surface mesh.

Fig. 1 Total relative error of surface area and volume and the average Hausdorff distance between SES and Gaussian surfaces with different parameters for the biomolecules in Table 1. The unit of Hausdorff distance is Å



Four biomolecules are chosen to compute the average distances between Gaussian surface meshes and the VDW surfaces (see Table 2). Table 2 shows that the distances decrease as d increases and the average distances for the four biomolecules with $d = 2.0, c = 1.0$ are all controlled within 5%. This illustrates that $d = 2.0, c = 1.0$

in the Gaussian surface can be used as a set of reasonable parameters to approximate the VDW surface.

Above analysis and parameters can be similarly applied to the SAS, because it is geometrically the same as the VDW surface. The only difference is that the r_i in expression (11) is replaced by the sum of r_i and probe radius.

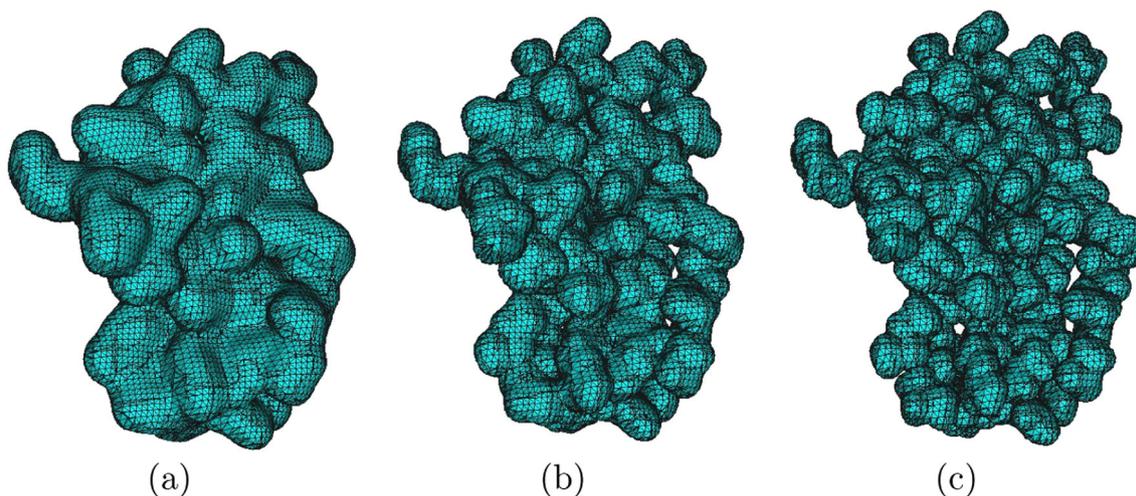


Fig. 2 Gaussian surfaces with different parameters for fas2.pqr. **a** $d = 0.5$ **b** $d = 1.0$ **c** $d = 2.0$

Table 2 The average distances between Gaussian surface meshes and the VDW surfaces for four biomolecules

Molecule	Natoms	Distance		
		$d = 0.5$	$d = 1.0$	$d = 2.0$
2JM0	589	0.263	0.087	0.029
fas2	906	0.364	0.134	0.043
4DUT	4410	0.274	0.105	0.036
ache	8280	0.386	0.148	0.046

Comparison of surface mesh generation software

In this section, the performances of the aforementioned common mesh generation software programs; MSMS, Molsurf, NanoShaper, EDTsurf, TMSmesh and GAMer are compared and discussed.

GAMer and EDTsurf

GAMer is a mesh generation software that is used to generate surface/volume meshes for the Gaussian molecular surface. GAMer has been shown to generate a very smooth and uniform mesh. However, the surface mesh generated by

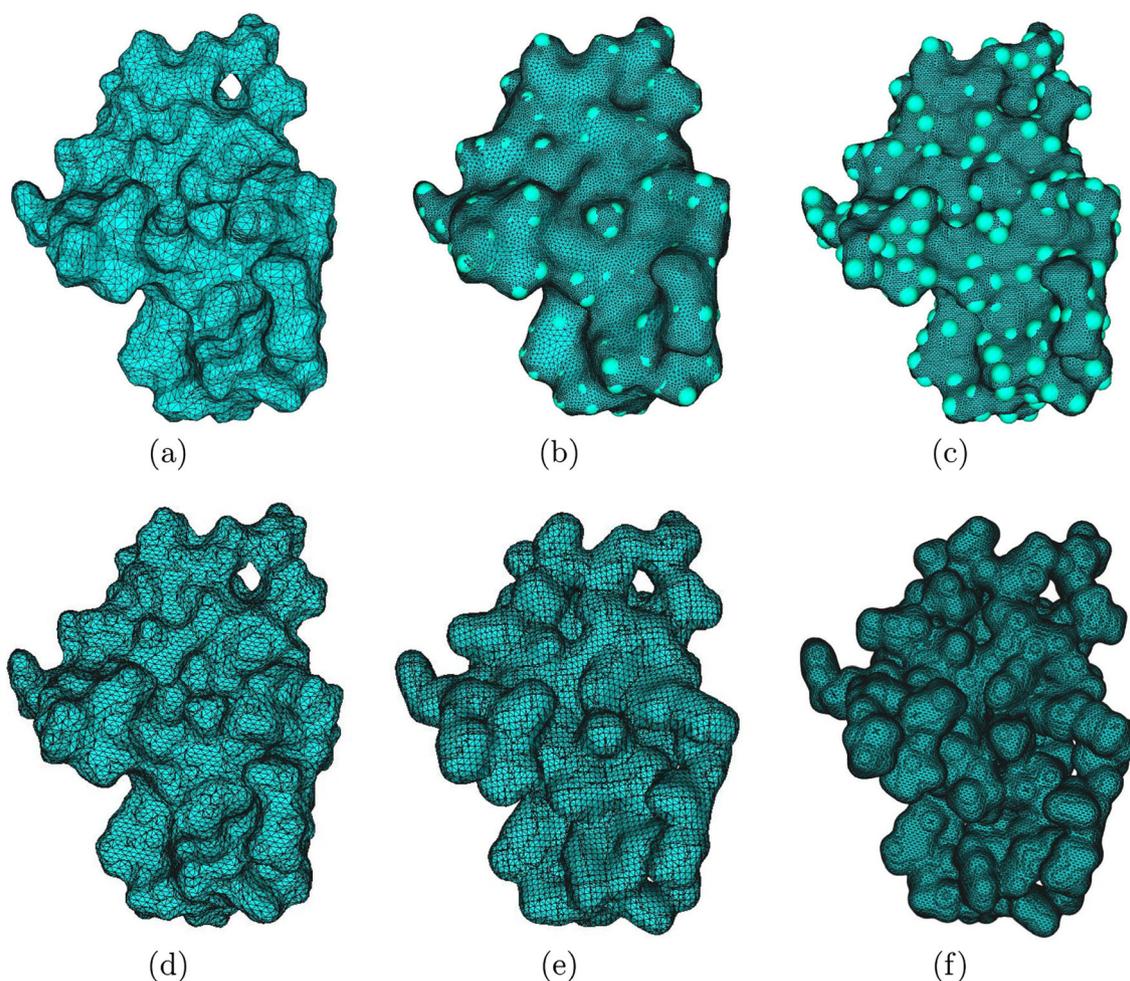


Fig. 3 Surface mesh of fas2. **a** Surface mesh generated by MSMS. **b** Surface mesh generated by GAMer using default parameters. **c** Surface mesh generated by EDTsurf using default parameters. **d** Gaussian molecular surface mesh generated by Molsurf using default

parameters. **e** Surface mesh generated by TMSmesh with $d = 0.6$, $c = 1.7$. **f** Gaussian molecular surface mesh generated by Nanoshaper using default parameters. The green patch outside the mesh is VDW surface

GAMer is not always a manifold mesh and the surface mesh often does not accurately reflect the topology of the original molecular surface. For example, some structures like holes and tunnels are often missed (see Fig. 3 b). Altering the parameters in the Gaussian surface generation could also result in the exclusion of detailed structures of the molecular surface. Additionally, GAMer is unable to efficiently handle macromolecules.

EDTsurf is an improved version of LSMS and was developed for generating three major macromolecular surfaces; the VDWs, SES and SAS, using the technique of fast Euclidean Distance Transform. It is a highly efficient program and can handle macromolecules. The meshes generated by EDTsurf are uniform and smooth. However, some parts of the VDW surface reach the outside of the enclosed volume of the mesh produced by EDTsurf (see Fig. 3 c) and some detailed structures, like holes and tunnels are missed in the produced mesh. Our own observations indicated that the programs Molsurf, NanoShaper and TMSmesh can preserve the topology of the molecular surface relatively well. Therefore, in the following subsections, further performance comparisons are only made among MSMS, Molsurf, NanoShaper and TMSmesh for Gaussian surface mesh generations. Molsurf and NanoShaper can also generate good quality mesh for SES, but it is out of our current discussion.

Performance of MSMS, Molsurf, NanoShaper and TMSmesh

MSMS is one of the most popular software programs used for mesh generation. In MSMS, the SES is computed in two steps. First, the Reduced Surface (which is a sort of skeleton) is built. Then, after every patch of the surface is identified, the triangulation is performed. MSMS has high

efficiency and the resulting mesh is a very close representation of the SES. However, the generated mesh is not always a manifold mesh and is composed of irregular triangles. Molsurf, a module in the software TexMol [32], is another tool for generating various molecular surfaces including SES, SAS and the Gaussian surface. The surface mesh generated by Molsurf plays good in quality and faithfulness in spite of requiring a lot of memory for molecules with large size. NanoShaper, is a new robust and efficient software to triangulate three kinds of molecular surfaces including SES, the skin surface and the Gaussian surface. The surface mesh generated by NanoShaper has high quality. However, the generated mesh based on Gaussian surface sometimes is not manifold.

In this work, for MSMS, the node density and probe radius are set at 10 \AA^2 and 1.4 \AA , respectively. In generating the Gaussian surface mesh using Molsurf, we set the grid size at 256 and the other parameters are set at the default values. For NanoShaper, the grid scale for generating the Gaussian surface is set 3.0 \AA and also take default values of other parameters. The grid space in TMSmesh is set at a value of 0.2 \AA . In the following sections, the performance of Molsurf, NanoShaper and TMSmesh is compared with that of MSMS.

Area, volume and Hausdorff distance The surface areas and the volumes enclosed by the surface for each of the biomolecules listed in Table 1 and another 40 test biomolecules listed in Table 3 were calculated by MSMS, Molsurf, NanoShaper and TMSmesh. The number of atoms in these biomolecules range from hundreds to thousands. These molecules were chosen randomly from RCSB Protein Data Bank, and no particular structure is specified. The Hausdorff distances between the surface meshes generated

Table 3 Number of atoms for another 40 test proteins

Index	1	2	3	4	5	6	7	8
Molecule	2LWC	1GNA	1S4J	1CB3	1V4Z	1BTQ	1ZWF	1AIE
N_{atom}	75	163	182	189	266	310	592	642
Index	9	10	11	12	13	14	15	16
Molecule	adna	4HLI	2IJI	3IM3	3ES0	1COA	2AVP	2ONT
N_{atom}	778	829	929	944	992	1153	1159	1218
Index	17	18	19	20	21	22	23	24
Molecule	375D	3ICB	3SZX	1SM5	3LDE	2YX5	1AYI	1DCW
N_{atom}	1287	1317	1350	1376	1379	1418	1467	1534
Index	25	26	27	28	29	30	31	32
Molecule	112X	4GSG	3LOD	2CEK	3DFG	1TR4	1RMP	11F4
N_{atom}	1663	1715	2315	2660	2927	3423	3549	4264
Index	33	34	35	36	37	38	39	40
Molecule	4DUT	3SQE	1HG8	4DPF	3FR0	2H8H	ache	11L5
N_{atom}	4410	5289	5980	6515	7390	7801	8280	8462

by TMSmesh/NanoShaper/Molsurf and by MSMS for the biomolecules were computed.

Figure 4 shows the surface areas for the 64 biomolecules calculated by MSMS, TMSmesh, Molsurf and NanoShaper. Figure 4a demonstrates that the surface areas computed by TMSmesh with the parameters set at $d = 0.4, c = 1.8$ and $d = 0.4, c = 1.2$ are over all in good agreement with the analytical areas computed by MSMS. Furthermore, the distribution of relative errors in Fig. 4b shows that the parameter at $d = 0.4, c = 1.8$ works better than $d = 0.4, c = 1.2$. The linear regression in Fig. 4a for $d = 0.4, c = 1.8$ gives the correlation coefficient $r = 0.9991$ and the slope of the fitted line is $k = 1.025$. The surface areas computed by both Molsurf, NanoShaper and TMSmesh with the parameters set at $d = 0.6, c = 1.7$ are close to each other and deviate greatly from MSMS's as the size of the molecule increases. This could be due to the fact that larger proteins

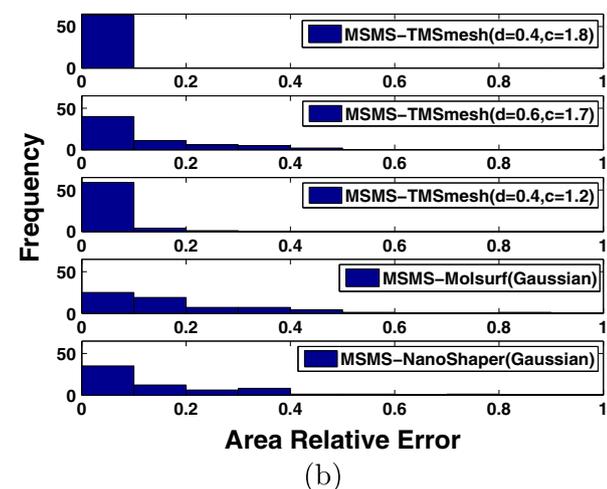
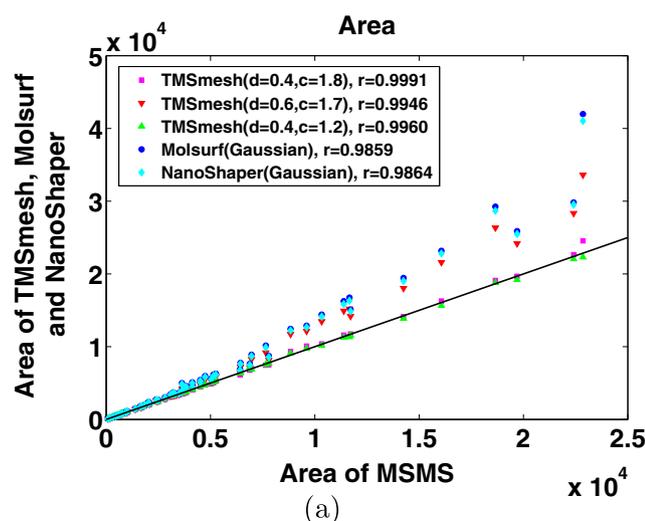


Fig. 4 Comparison of area calculation. The unit of area is \AA^2 . In the legend, r represents the correlation coefficient. **a** Results of area calculations for 64 molecules. **b** The distribution of area relative errors

have the potential to contain more crevices and invaginations.

Figure 5 shows the volume calculation results by MSMS, TMSmesh, NanoShaper and Molsurf for the 64 biomolecules. The results obtained from TMSmesh, Molsurf and NanoShaper are all similar to that of MSMS (see Fig. 5a). The optimal result is obtained using TMSmesh with the parameter $d = 0.6, c = 1.7$ (see Fig. 5b). r for linear correlation of the results by TMSmesh with $d = 0.6, c = 1.7$ is 0.9998. Similarly, Molsurf plays well in the volume calculation and the linear regression analysis gives $r = 0.9989$. The volume calculated by NanoShaper are also in very good agreement with the ones calculated by MSMS. Additionally, the volume calculated by TMSmesh with the parameter $d = 0.4, c = 1.8$ and $d = 0.4, c = 1.2$ are

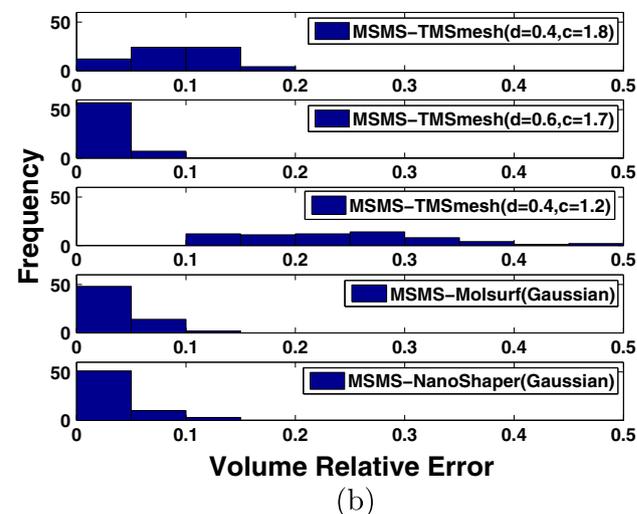
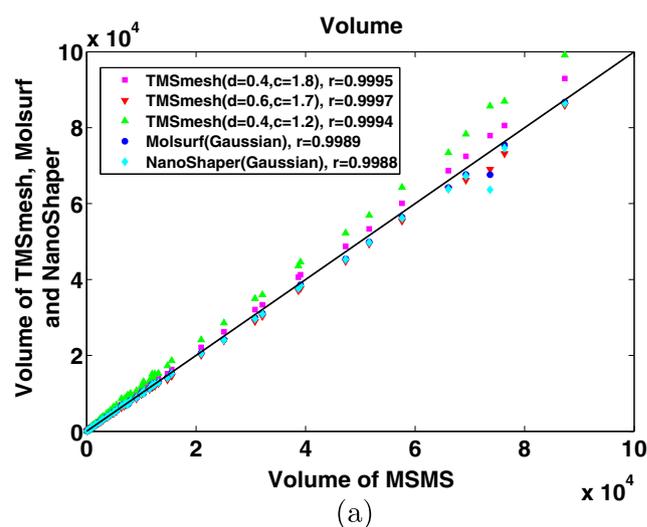


Fig. 5 Comparison of volume calculation. The unit of volume is \AA^3 . In the legend, r represents the correlation coefficient. **a** Results of volume calculations for 64 molecules. **b** The distribution of volume relative errors

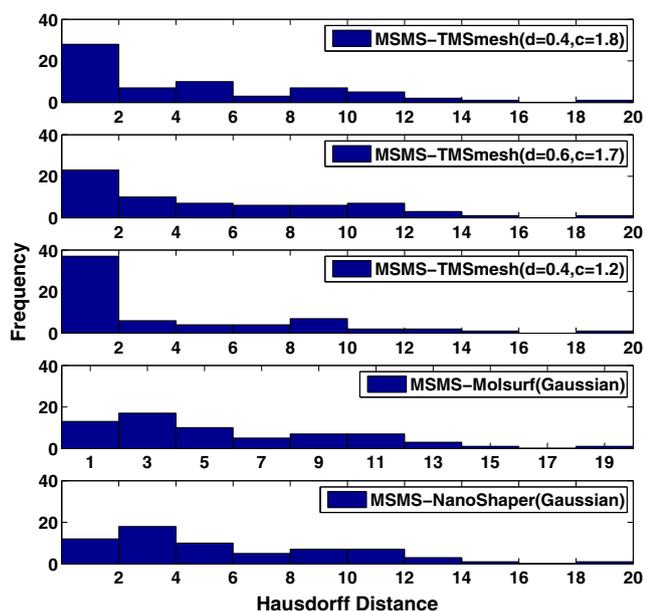


Fig. 6 Distribution of Hausdorff distance for 64 molecules. The unit of Hausdorff distance is Å

slightly larger than the result obtained by MSMS whereas the volume calculation results by TMSmesh with the parameter $d = 0.6$, $c = 1.7$, NanoShaper and Molsurf are slightly smaller than the results obtained by MSMS.

Figure 6 gives the distribution of the Hausdorff distances between the surfaces generated by TMSmesh and MSMS, the distances between NanoShaper and MSMS, as well as the distances between Molsurf and MSMS. Clearly, the meshes computed by TMSmesh with $d = 0.4$, $c = 1.2$ are closest to those generated by MSMS.

Taking the values for the area and the volume obtained by MSMS as references, the relative errors for the area and the volumes calculated via TMSmesh, Molsurf and NanoShaper are shown in Figs. 4b and 5b and the average relative errors are listed in Table 4. The relative errors of area calculated by TMSmesh with three different parameters are both smaller than that of Molsurf and NanoShaper. And the volume results by TMSmesh with $d = 0.6$, $c = 1.7$ give the lowest average relative error 2.98 %. The fourth column in Table 4 shows that the results calculated by TMSmesh with different parameters works a little better than Molsurf and

NanoShaper in terms of Hausdorff distance. These results demonstrate that TMSmesh with the optimal parameters can generate accurate molecular surfaces.

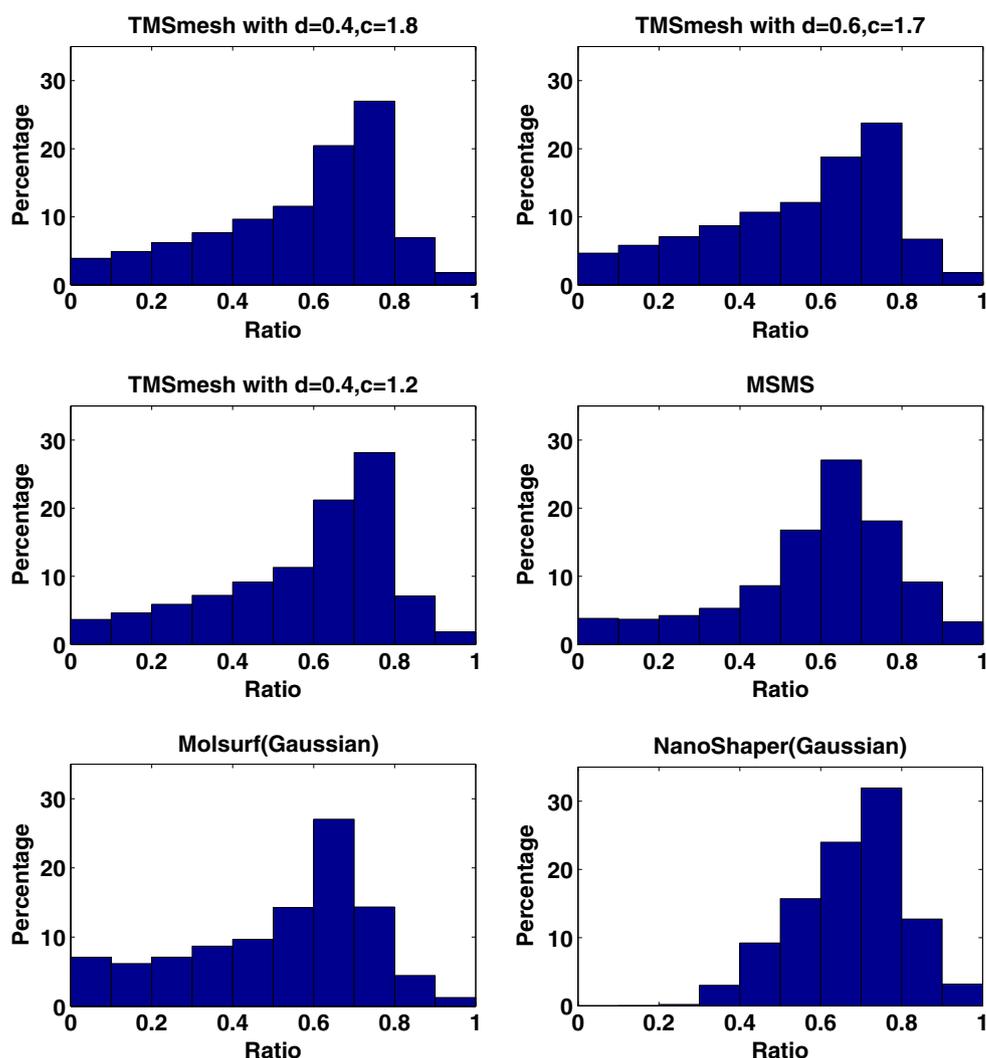
Mesh quality In this paragraph, the mesh qualities, in particular uniformness and manifoldness of meshes produced by MSMS, Molsurf, NanoShaper and TMSmesh are compared. The distribution of the ratios of the shortest edge length to the longest edge length of each triangle and the distribution of the angles of each triangle are used to describe the uniformness of a triangulated surface mesh. A ratio of 1.0 corresponds to an equilateral triangle, and a ratio close to 0 is indicative of a very poor uniformness. That is, the higher the ratio, the better the quality of the triangle. The distribution of ratios and the angles of each triangle produced by TMSmesh, MSMS, Molsurf and NanoShaper for the 64 proteins are shown in Figs. 7 and 8. The results in Fig. 7 show that the ratios of meshes generated by TMSmesh and NanoShaper are clustered around 0.75 and the ratios of the meshes generated by MSMS and Molsurf both center on 0.65. Additionally, there is no triangle whose ratio is between 0 and 0.2 in the meshes from NanoShaper, which means NanoShaper generates few triangles with very poor quality. The distributions of the angles of the triangles in meshes generated by those four softwares are shown in Fig. 8. It signifies high quality that a large fraction of angles center on 60° and fewer are close to 0° or 180° . The distribution of the angles are clustered around 40° to 100° and also shows that the mesh generated by Nanoshaper has the least amount of sharp angles.

In the following, we compare the manifoldness of meshes generated by MSMS, TMSmesh, Molsurf and NanoShaper. A surface is manifold means that each point on the surface has a neighborhood which is homeomorphic to a disk in a real plane. As we mentioned in Ref. [27], there are three necessary conditions for a mesh to be a manifold mesh. a) The first is that each edge should be and only be shared by two faces (a face denotes an element patch) of the mesh. b) The second is that each vertex should have and only have one neighborhood node loop. Neighborhood node of a vertex is a node connecting the vertex through an edge. c) The third is that the mesh has no intersecting face pairs. A high quality mesh has a continuous manifold. Examples

Table 4 Average relative errors of area, volume and Hausdorff distance

Method	Average relative error of area	Average relative error of volume	Average Hausdorff distance
TMSmesh ($d = 0.4$, $c = 1.8$)	1.74 %	9.18 %	4.6437
TMSmesh ($d = 0.6$, $c = 1.7$)	10.35 %	2.98 %	5.1396
TMSmesh ($d = 0.4$, $c = 1.2$)	3.87 %	23.94 %	3.8907
Molsurf (Gaussian)	18.11 %	3.83 %	5.7013
NanoShaper (Gaussian)	15.6 %	3.35 %	5.7170

Fig. 7 Distributions of ratio of the shortest edge length to the longest edge length of each triangle produced by TMSmesh with $d = 0.4, c = 1.8$, TMSmesh with $d = 0.6, c = 1.7$, TMSmesh with $d = 0.4, c = 1.2$, MSMS, Molsurf and NanoShaper

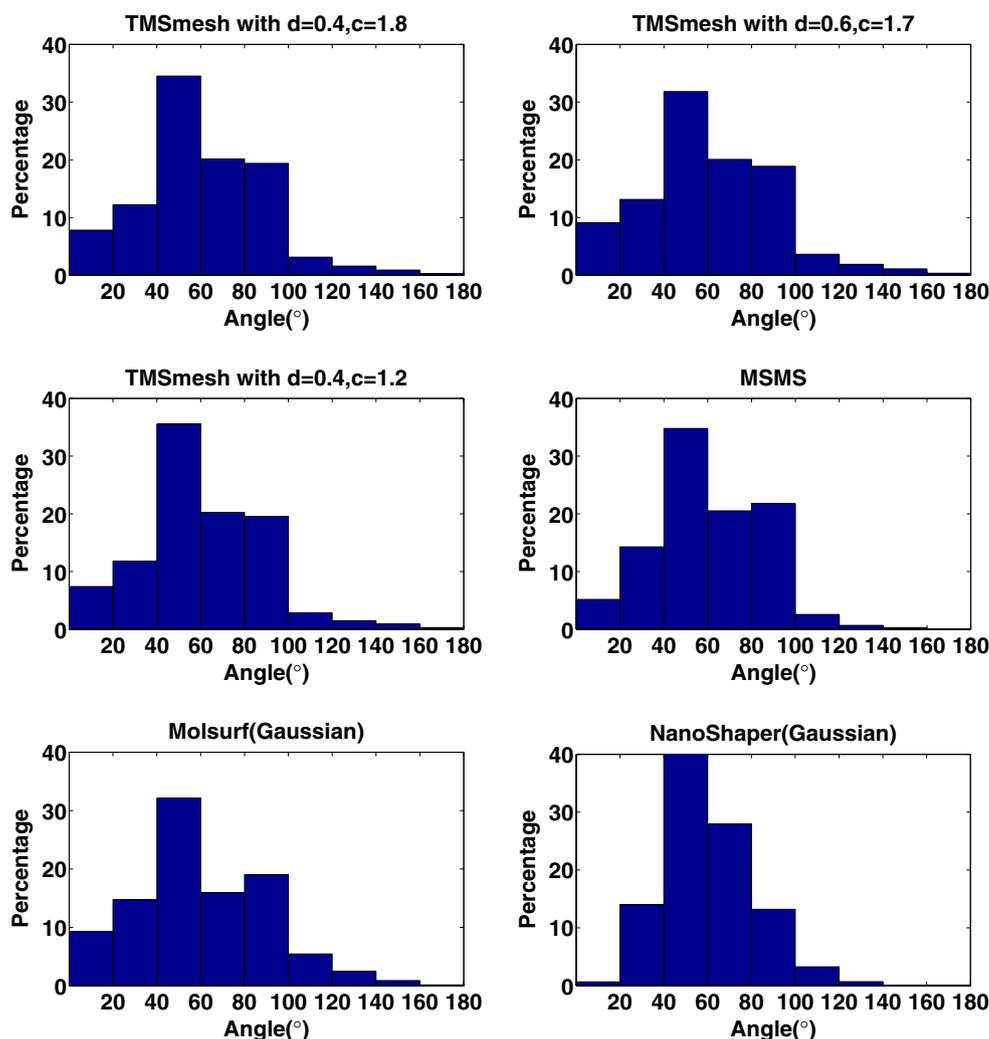


of non-manifold meshes can be those where any case of intersection, isolated nodes/triangles, holes (resulted from missing face) and so on occurs (more details can be found in Ref. [9]). Nonmanifold meshes can lead to numerical problems in BEM and FEM types' simulations. The number of non-manifold defects and the number of intersecting triangle pairs in the meshes produced by TMSmesh, Molsurf, NanoShaper and MSMS are shown in Table 5. For test of the meshing software in this work, we use our previous PQR benchmark [27] that can be found and downloaded from the following website www.continuummodel.org. In this work, we set the node density of MSMS to 10.0 \AA and the grid size in Molsurf to 256 while the grid space in TMSmesh is set at a value of 0.2. The grid scale in NanoShaper is set at 1.0 \AA for 1K4R and 2.0 \AA for other molecules. All computations are run on Dell Optiplex 960 with Intel(R) Core(TM)2 Quad CPU Q9550 @ 2.83GHz and 8GB memory. It is shown in Table 5 that TMSmesh and NanoShaper are robust tools for meshing the Gaussian surface for large biomolecules

and the meshes produced by TMSmesh and Molsurf are manifold meshes, while the meshes of large biomolecules produced by MSMS and NanoShaper are not.

Solvation energy It has been a long existing arguable issue in computational biophysics to ask which type of surface should be used in molecular solvation energy calculations (e.g. see Zhou et al. [33]). The topic is outside the focus of current work. In this subsection, we will show that after the optimal parameters are determined according to geometric criteria, the Gaussian surface can also result in very close solvation energy calculations compared with the SES. We use a boundary element Poisson-Boltzmann solver, AFMPB [34], to compute the corresponding PB electrostatic solvation energies using meshes generated by TMSmesh, MSMS, Molsurf and NanoShaper respectively. The solvation energies by AFMPB computed from the meshes for the 64 biomolecules are shown in Fig. 9. Figure 9 indicates that the results calculated by TMSmesh and MSMS are very close

Fig. 8 Distributions of angles of each triangle produced by TSMesh with $d = 0.4$, $c = 1.8$, TSMesh with $d = 0.6$, $c = 1.7$, TSMesh with $d = 0.4$, $c = 1.2$, MSMS, Molsurf and NanoShaper



(with one exception). For a few biomolecules, the results obtained from Molsurf are shown to greatly deviate from the results obtained by both TSMesh and MSMS, which could be due to sharp topological changes. The results of the NanoShaper showed in Fig. 9 are close to the ones of MSMS. However, AFMPB failed to give the results of

the meshes of 6 molecules including 1IL5, 2CEK, 2H8H, 3DFG, 4DPF and 4DUT generated by NanoShaper because of topological errors, such as self intersections. If the values of the solvation energy determined for MSMS are taken as standard values, the average relative errors of TSMesh with the parameter $d = 0.4$, $c = 1.8$, with the parameter

Table 5 Number of non-manifold errors in meshes produced by TSMesh, Molsurf, NanoShaper and MSMS

Molecule	Number of atoms	Number of non-manifold defects ^a				Number of intersecting triangle pairs			
		TSMesh	Molsurf	NanoShaper	MSMS	TSMesh	Molsurf	NanoShaper	MSMS
FAS2	906	0	0	0	0	0	0	0	0
AChE monomer	8280	0	0	0	0	0	0	10	643
AChE tetramer	36638	0	0	0	30	0	0	17	1428
30S ribosome	88431	0	0	0	–	0	0	25	–
70S ribosome	165337	0	0	0	–	0	0	10	–
1K4R	1082160	0	–	0	–	0	–	947	–

^aNumber of vertices whose neighborhood does not satisfy before mentioned necessary conditions a), b) for a manifold mesh

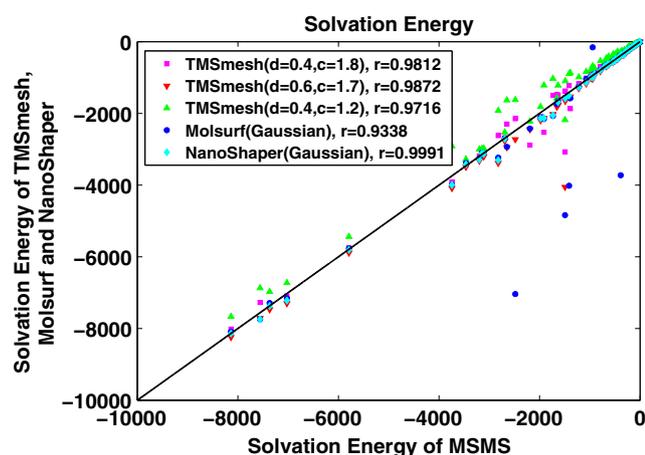


Fig. 9 Comparison of solvation energy

$d = 0.6$, $c = 1.7$, with the parameter $d = 0.4$, $c = 1.2$, Molsurf and NanoShaper are 11.39 %, 7.96 %, 24.49 %, 28.76 % and 6.53 %, respectively.

Conclusions

In this paper, the surface area, the volume enclosed by the surface and the Hausdorff distance are used as three criteria for the parameterization of the Gaussian surface in the approximation of the SES surface. The results of the parameterization indicate that it is not possible to find a unique set of parameters that satisfies all the different criteria. Parameter selection is determined by the criteria and properties that are trying to be fulfilled. For the criterion of area and using SES as a reference surface, the optimal parameter is determined to be $d = 0.4$, $c = 1.8$ and for the criterion of volume, the optimal parameter is determined to be $d = 0.6$, $c = 1.7$, and for the Hausdorff distance, the optimal one is $d = 0.4$, $c = 1.2$. But as also shown by our results, the difference between two optimized sets of parameters is not big, and most of the calculated results do not deviate far from each other when using different set of the optimized parameters. In solvation calculation, we found that the Gaussian surface with the optimal parameters $d = 0.6$, $c = 1.7$ based on volume criterion gives results closest to those obtained with SES, of which the regression coefficient is 0.9872. The Gaussian surface can reproduce the VDW surface reasonably when $d = 2.0$, $c = 1.0$. It worth noted that similar to VDW surface the solvent accessible surface (SAS) can also be approximated by the Gaussian surface with the same parameters and by simply increasing each atomic radius by a probe radius. In this paper, we also compared some mesh generation software; including MSMS, EDTsurf, GAMer, Molsurf, NanoShaper and

TMSmesh in the aspects of robustness and mesh quality. Among these software, EDTsurf and GAMer may result in relatively lower mesh fidelity to the original molecular surface. For example, some holes or/and cavities in a molecule may be missed in their generated meshes, and the VDW surface can apparently reach outside the meshes. Parameterized TMSmesh, NanoShaper and Molsurf can generate Gaussian surface mesh and produce similar and faithful surface properties as SES surface generated by MSMS. NanoShaper is robust and the generated mesh is of high quality. However, the mesh generated by NanoShaper is not always manifold and some topological errors in mesh may cause difficulties in numerical calculations. The mesh generated by Molsurf is manifold and has good quality, but Molsurf seems not so efficient to handle large molecules. TMSmesh is shown as a robust tool for meshing Gaussian surface for large biomolecules and the produced meshes are manifold. However, there exist relatively more triangles with very small angles or short edges in the triangular mesh generated by TMSmesh, this mesh quality may cause trouble or need further improved in numerical simulations.

Acknowledgments We would like to thank Bin Tu for his help. TT and BZ are supported by the State Key Laboratory of Scientific/Engineering Computing, National Center for Mathematics and Interdisciplinary Sciences, the Chinese Academy of Sciences, the China NSF (91230106) and 863 program (2012AA020403). MX is supported by China NSF (NSFC11301368) and the NSF of Jiangsu Province (BK20130278).

References

- Lee B, Richards FM (1971) The interpretation of protein structures: estimation of static accessibility. *J Mol Biol* 55(3):379–400
- Richards FM (1977) Areas, volumes, packing and protein structure. *Annu Rev Biophys Bioeng* 6:151–176
- Edelsbrunner H (1999) Deformable smooth surface design. *Discret Comput Geom* 21:87–115
- Bates PW, Wei GW, Zhao S (2008) Minimal molecular surfaces and their applications. *J Comput Chem* 29(3):380–391
- Mcgann MR, Almond HR, Nicholls A, Grant AJ, Brown FK (2003) Gaussian docking functions. *Biopolymers* 68:76–90
- Grant JA, Gallardo MA, Pickup BT (1996) A fast method of molecular shape comparison: a simple application of a Gaussian description of molecular shape. *J Comput Chem* 17(14):1653–1666
- Weiser J, Shenkin PS, Still WC (1999) Optimization of Gaussian surface calculations and extension to solvent-accessible surface areas. *J Comput Chem* 20(7):688–703
- Yun Z, Matthew PJ, Richard AF (2005) What role do surfaces play in GB models? A new-generation of surface-generalized born model based on a novel Gaussian Surface for biomolecules. *J Comput Chem* 27:72–89
- Chen MX, Lu BZ (2011) TMSmesh: a robust method for molecular surface mesh generation using a trace technique. *J Chem Theory Comput* 7:203–212

10. Bin T, Chen M, Xie Y, Zhang L, Eisenberg B, Lu B (2013) A parallel finite element simulator for ion transport through three-dimensional ion channel systems. *J Comput Chem* 34(24):2065–2078
11. Grant JA, Pickup BT, Nicholls A (2001) A smooth permittivity function for Poisson-Boltzmann solvation methods. *J Comput Chem* 22(6):608–640
12. Li L, Li C, Alexov E (2014) On the modeling of polar component of solvation energy using smooth Gaussian-based dielectric function. *J Theor Comput Chem* 13(03):1440002
13. Duncan BS, Olson AJ (1993) Shape analysis of molecular surfaces. *Biopolymers* 33(2):231–238
14. Connolly ML (1983) Analytical molecular surface calculation. *J Appl Crystallogr* 16(5):548–558
15. Connolly ML (1983) Solvent-accessible surfaces of proteins and nucleic acids. *Science* 221(4612):709–713
16. Nicholls A, Bharadwaj R, Honig B (1995) GRASP: graphical representation and analysis of surface properties. *Biophys J* 64:166–167
17. Sanner M, Olson A, Spehner J (1996) Reduced surface: an efficient way to compute molecular surface properties. *Biopolymers* 38:305–320
18. Vorobjev YN, Hermans J (1997) SIMS: computation of a smooth invariant molecular surface. *Biophys J* 73(2):722–732
19. Ryu J, Park R, Kim D-S (2007) Molecular surfaces on proteins via beta shapes. *Comput Aided Des* 39(12):1042–1057
20. Xu D, Zhang Y (2009) Generating triangulated macromolecular surfaces by Euclidean Distance Transform. *PLoS One* 4(12):e8140
21. Decherchi S, Rocchia W (2013) A general and robust ray casting based algorithm for triangulating surfaces at the nanoscale. *PLoS One* 8(4):e59744
22. Zhang YJ, Xu GL, Bajaj C (2006) Quality meshing of implicit solvation models of biomolecular structures. *Comput Aided Geom Des* 23(6):510–530
23. Yu ZY, Holst MJ, McCammon J (2008) High-fidelity geometric modeling for biomedical applications. *Finite Elem Anal Des* 44(11):715–723
24. Bajaj C, Xu G, Zhang Q (2008) Bio-molecule surfaces construction via a higher-order level-set method. *J Comput Sci Technol* 23(6):1026–1036
25. Zhao W, Xu G, Bajaj C (2011) An algebraic spline model of molecular surfaces for energetic computations. *IEEE/ACM Trans Comput Biol Bioinforma* 8(6):1458–1467
26. Liao T, Zhang YJ, Kekenes-Huskey PM et al (2013) Multi-core CPU or GPU-accelerated multiscale modeling for biomolecular complexes. *Mol Based Math Biol* 1:164–179
27. Chen MX, Tu B, Lu BZ (2012) Triangulated manifold meshing method preserving molecular surface topology. *J Mol Graph Model* 38:411–418
28. Decherchi S, Colmenares J, Catalano CE, Spagnuolo M, Alexov E, Rocchia W (2013) Between algorithm and model: different molecular surface definitions for the Poisson-Boltzmann based electrostatic characterization of biomolecules in solution. *Commun Comput Phys* 13:61–89
29. Blinn JF (1982) A generalization of algebraic surface drawing. *ACM Trans Graph* 1(3):235–256
30. Grant JA, Pickup BT (1995) A Gaussian description of molecular shape. *J Phys Chem* 99:3503–3510
31. Cignoni P, Rocchini C, Scopigno R (1998) Metro: measuring error on simplified surfaces. *Comput Graph Forum* 17(2):176–174
32. Bajaj C, Djeu P, Siddavanahalli V, Thane A (2004) TexMol: interactive visual exploration of large flexible multi-component molecular complexes. In: 15th IEEE visualization 2004, pp 243–250
33. Pang X, Zhou HX (2013) Poisson-Boltzmann calculations: van der Waals or molecular surface? *Commun Comput Phys* 13:1–12
34. Lu BZ, Cheng XL, Huang JF, McCammon AJ (2010) AFMPB: an adaptive fast multipole Poisson-Boltzmann solver for calculating electrostatics in biomolecular systems. *Comput Phys Commun* 181(6):1150–1160