



Contents lists available at ScienceDirect

Journal of Computational and Applied Mathematics

journal homepage: www.elsevier.com/locate/cam



Geometric and electrostatic modeling using molecular rigidity functions



Lin Mu^{a,*}, Kelin Xia^b, Guowei Wei^{c,d,e}

^a Computer Science and Mathematics Division, Oak Ridge National Laboratory, Oak Ridge, TN, 37831, USA

^b Division of Mathematical Sciences, School of Physical and Mathematical Sciences, Nanyang Technological University, 21 Nanyang link, Singapore 637371, Singapore

^c Department of Mathematics, Michigan State University, East Lansing, MI 4882, USA

^d Department of Electrical and Computer Engineering, Michigan State University, East Lansing, MI 4882, USA

^e Department of Biochemistry and Molecular Biology, Michigan State University, East Lansing, MI 4882, USA

ARTICLE INFO

Article history:

Received 18 April 2016

Received in revised form 27 June 2016

Keywords:

Flexibility rigidity index

Rigidity function

Molecular surface

Curvature

ABSTRACT

Geometric and electrostatic modeling is an essential component in computational biophysics and molecular biology. Commonly used geometric representations admit geometric singularities such as cusps, tips and self-intersecting facets that lead to computational instabilities in the molecular modeling. The present work explores the use of flexibility and rigidity index (FRI), which has a proved superiority in protein B-factor prediction, for biomolecular geometric representation and associated electrostatic analysis. FRI rigidity surfaces are free of geometric singularities. We proposed a rigidity based Poisson–Boltzmann equation for biomolecular electrostatic analysis. Our approaches to surface and electrostatic modeling are validated by a set of 21 proteins. Our results are compared with those of established methods. Finally, being smooth and analytically differentiable, FRI rigidity functions offer excellent curvature analysis, which characterizes concave and convex regions on protein surfaces. Polarized curvatures constructed by using the product of minimum curvature and electrostatic potential is shown to predict potential protein–ligand binding sites.

© 2016 Elsevier B.V. All rights reserved.

1. Introduction

A major feature of biological sciences in the 21st Century is their transition from phenomenological and descriptive to quantitative and predictive. A fundamental aspect of quantitative biology is the geometric modeling at various scales ranging from molecular, subcellular organelles, cellular, tissue, organ scales to species scale [1–4]. At microscopic scales, geometric modeling provides a vital tool for the conceptualization of biomolecules, namely, protein, DNA and RNA, and the visualization of biological data. At all scales, geometric modeling fills the gap between structural information and mathematical models [5,2]. Indeed, the annotation of biomolecular surfaces together with physical features, such as amino acid residues, functional groups, electrostatics, flexibility, hinges, domains, hydrophilic and hydrophobic regions, drug-binding pockets, and protein–protein interaction hot spots, contributes tremendously to the understanding of molecular recognition in protein design and drug discovery.

* Corresponding author.

E-mail addresses: mul1@ornl.gov (L. Mu), xiakelin@ntu.edu.sg (K. Xia), wei@math.msu.edu (G. Wei).

Geometric modeling consists of a number of important ingredients, including molecular surface definition, surface meshing, molecular visualization, morphological analysis, curvature analysis, surface annotation, pertinent feature extraction, etc. Molecular surface definition is of essential importance in geometric modeling. The visualization of molecular surfaces is also of crucial interest in biophysical modeling. The surface visualization becomes more powerful when electrostatic potential, flexibility information and curvature content are projected on the surface as well. For example, the illustration of electrostatic potentials on protein surfaces [6–8] and the visualization of atomic thermal fluctuations on protein surfaces [9] shed light on protein structure, function and interaction, including protein–DNA specification, protein–protein binding hot spots, ligand–receptor binding sites, and enzymatic mechanism.

Corey–Pauling–Koltun (CPK) model proposed in 1953 is one of the first visualization models [10,11]. It represents atoms as space-filling three-dimensional (3D) spheres with atomic radii and has been widely used in molecular biosciences. Quantum mechanical theory of atoms and molecules offers orbitals based on spherical harmonics. This approach is useful for local bonding visualization and analysis. Backbone model [12], which depicts helices and strands as flat or coiled ribbons, provides another simplified model for macromolecules and often retains the sequence of C_α atoms. This model is frequently employed for qualitative illustration of the biomolecular structural constitution.

In addition to qualitative visualization, biomolecular surface models are often used in quantitative modeling, i.e., multiscale models and implicit solvent models. In this class of applications, the surface is utilized to divide the computational domain into subdomains where different physical approaches, namely, quantum, discrete and continuum descriptions, can be prescribed. A number of molecular surface models have been proposed in the past. One of them is the van der Waals surface (vdWS), which is a union of all the unburied atomic surfaces. Solvent accessible surface (SAS) defined as the trajectory of the center of a probe sphere moving around the van der Waals surface [13] is another commonly used surface representation in biophysics. Both vdWs and SASs are non-smooth and cause computational difficulties. Solvent excluded surface (SES), also known as the Connolly surface [14] or molecular surface, was proposed to remove non-smooth regions in vdWS and SAS [15]. The SES is created by the trace of the inward moving surface of a probe sphere rolling around the vdWS. Connolly divided SES into two major parts, the contact areas formed by the subsets of the vdWS and the re-entrant surfaces, which contain toroidal patches and concave spherical triangles [14]. Analytical expressions have been derived for the re-entrant surface [14]. The vdWS can be reviewed a special case of the SES when the probe radius is set to zero. Qualitatively, SESs are widely used in visualization to analyze protein–protein interactions [16], protein folding [17], protein surface topography [18], DNA binding and bending [19], macromolecular docking [20], enzyme catalysis [21] and drug classification [22]. Quantitatively, SESs are employed in implicit solvent theories [23,24], molecular dynamics [25] and ion channel transport [24,26,27]. Computationally, efficient generations of SESs have been developed using alpha-shapes [28], analytical expressions [29] and marching tetrahedra [30].

An undesirable fact of aforementioned surface models is that they are also non-smooth and admit geometric singularities [14,31,29]. Commonly known geometric singularities include cusps, sharp corners, tips and self-intersection surfaces, which induce instability in numerical simulations and hinder the curvature estimation. Advanced mesh generation algorithms and/or sophisticated interface methods are required to deal with geometric singularities [5]. To avoid the geometric singularities, Gaussian surfaces which represent atoms by Gaussian functions, have been developed to extract a surface from a given level set value [32,5,33,34]. The Gaussian surfaces can be further smoothed by using the PDE transform [35]. Another method is the molecular skinning surface generated by performing surgery operations on molecular surfaces, i.e., replacing singularities with spherical and hyperboloid patches and resulting in continuous normal vector fields on the surface [36,37]. Spherical harmonic expansion has been employed to smooth molecular surfaces by keeping only low order polynomials [38]. The direct use of electron density distributions computed from quantum mechanics gives rise to a more physical representation of biomolecules and related surfaces of molecules are obtained by iso-surface extraction resulting in electron density isosurfaces [39] and molecular faces [40]. These surfaces may be computationally expensive and sensitive to the molecule's interacting environment.

Differential geometry theory of surfaces provides an alternative way to construct biomolecular surfaces. One approach is to utilize geometric partial differential equations (PDEs) to generate biomolecular surfaces with appropriate geometric properties. The first curvature induced biomolecular surface was introduced in 2005 using geometric PDEs [41]. The first variational molecular surface based on minimal surface theory was proposed in 2006 [42,43]. In this approach, the minimization of surface energy leads to the minimal molecular surface (MMS). The mean curvature flow or Laplace–Beltrami equation was devised to generate MMSs for proteins and other biomolecules. This variational approach has been extended to a family of differential geometry based multiscale models where the surface energy plays a crucial role in the nonpolar energy description of biomolecular systems [1,44–48].

In addition to differential geometry based surface generation and multiscale analysis, differential geometry also offers the curvature description of biomolecular surfaces. In a nutshell, the curvature of a surface is a measure of how much it deviates from being flat [49]. For biomolecular surfaces, there are two principle curvatures, namely, minimal curvature and maximum curvature that can be organized into mean curvature and Gaussian curvature. Techniques for extracting biomolecular surface features, namely, electrostatic maps, curvatures, concave regions and convex regions, have been proposed [4]. Curvatures play a role in stereospecificity of biomolecular surfaces [50], protein–protein interaction hot spots, ligand binding pockets [2–4]. This approach is further empowered by coupling to the electrostatic analysis to predict ligand binding sites [4]. In general, Hadwiger integrals, namely, area, volume, mean curvature and Gaussian curvature are all important geometric quantities in the Minkowski functional space and potentially applicable to biomolecular geometric description.

Recently, we have introduced flexibility–rigidity index (FRI) for the analysis of macromolecular flexibility and thermal fluctuations [9]. Intensive validation and comparison using over 400 proteins indicate that our multiscale FRI (mFRI), is about 20% more accurate [51] and our fast FRI (fFRI) [52] is orders of magnitude more efficient than Gaussian network model (GNM) [53,54], which is one of the most popular methods in the field. Our fFRI is able to predict B-factors for α -carbons of an HIV virus capsid (313,236 residues) in less than 30 s on a single processor using only one core, which would take the GNM more than 120 years to compute assuming that the required memory would not be a limiting factor. Recently, we have proposed an FRI based weighted graph Laplacian to generalize and improve GNM [55]. In the FRI theory, the rigidity functions of biomolecules are defined as a summation of monotonically decaying correlation functions. These functions include the Gaussian function as a special case and thus have a potential to be used for the geometric analysis of biomolecules. The utility of the FRI theory for topological analysis has been investigated in our recent work [56,57].

The objective of present work is to explore the utility of the FRI theory for geometric analysis. Specifically, we exploit FRI theory for biomolecular surface generation and associated curvature analysis. We examine FRI surfaces generated by using two types of correlation functions, i.e., exponential kernel and Lorentz kernel functions, for representing biomolecules. Like the Gaussian surfaces, these FRI surfaces are C^∞ smooth and free of geometry singularities. By appropriate selection of the FRI resolution parameter, we generate multiscale surfaces for biomolecules. The associated surface areas and enclosed volumes can be evaluated. We further define a FRI surface based Poisson–Boltzmann model to compute electrostatic potentials for geometric analysis. The combination of electrostatic potentials and curvatures gives rise to polarized curvatures. These quantities are crucial for characterization of drug binding sites.

The rest of this paper is organized as follows: Section 2 is devoted to methods and algorithms. To establish notation and facilitate further discussion, the FRI approach is briefly discussed in Section 2.1. FRI based electrostatic modeling is presented in Section 2.2. We introduce an FRI surface based electrostatic profile for the Poisson–Boltzmann equation, which is robust and stable for electrostatic potential calculations. FRI surface based surface area and volume estimations are given in Section 2.3. In Section 2.4, curvature expressions are discussed. Numerical experiments are presented in Section 3 to validate the proposed methods and algorithms. We start with FRI surface generation and visualization, followed by surface area and volume analysis. Then through the numerical experiment by analytical surface and MSMS surface for proteins from Protein Data Bank (PDB), the reliability and accuracy are shown. The curvature evaluation is confirmed to be the second order numerical scheme. Finally, the definition of polarized curvature is introduced and utilized as a predictive tool for protein binding sites, which can be applied in drug design.

2. Methods and algorithms

2.1. Flexibility rigidity index

Flexibility rigidity index (FRI) was first introduced to analyze the atomic rigidity of biomolecules in 2013 [9]. In this work, FRI is employed as heterogeneous modulus in multiscale elasticity modeling and the resulting model is called continuum elasticity with atomic rigidity (CEWAR). An important aspect of the FRI theory is that the flexibility index can be quantitatively compared with the experimental thermal fluctuation of biomolecules, which provides a way to evaluate the performance of FRI. Consequently, FRI can be designed and improved using experimental data. The key idea of FRI is to construct flexibility index and rigidity index by certain kernel function which reflects the topological connectivity of a macromolecular network or graph consisting of atoms or residues and characterizes the geometric compactness of the macromolecular connectivity network. We construct a multiresolution geometric representation by using the FRI method, which converts the point cloud data into a matrix or a density map. The conversion is modulated by a resolution parameter, which facilitates the multiresolution analysis of complex data.

In our FRI methods, we seek a distance geometry based algorithm to convert protein geometry into protein topological connectivity. To this end, we consider a biomolecule with N particles or superatoms. The choice of particles depends on the representation. In a coarse-grained representation, each superatom can be an amino-acid residue in a protein or a protein in a protein complex, such as a microtubule. The locations of the particles are given by $\{\mathbf{r}_j | \mathbf{r}_j \in \mathbb{R}^3, j = 1, 2, \dots, N\}$. We denote $\|\mathbf{r} - \mathbf{r}_j\|$ the Euclidean space distance between position \mathbf{r} and the j th particle. The distance geometry is utilized to establish the topological connectivity by a correlation function

$$\Phi(\|\mathbf{r} - \mathbf{r}_j\|; \eta_j), \quad (1)$$

where η_j is a characteristic scale for j th particle. As a correlation function is a real-valued monotonically decreasing radial basis function satisfying following admissible conditions

$$\Phi(\|\mathbf{r} - \mathbf{r}_j\|; \eta_j) = 1, \quad \text{as } \|\mathbf{r} - \mathbf{r}_j\| \rightarrow 0 \quad (2)$$

$$\Phi(\|\mathbf{r} - \mathbf{r}_j\|; \eta_j) = 0, \quad \text{as } \|\mathbf{r} - \mathbf{r}_j\| \rightarrow \infty. \quad (3)$$

Commonly used correlation functions include generalized exponential functions

$$\Phi(\|\mathbf{r} - \mathbf{r}_j\|; \eta_j) = e^{-(\|\mathbf{r} - \mathbf{r}_j\|/\eta_j)^v}, \quad v > 0 \quad (4)$$

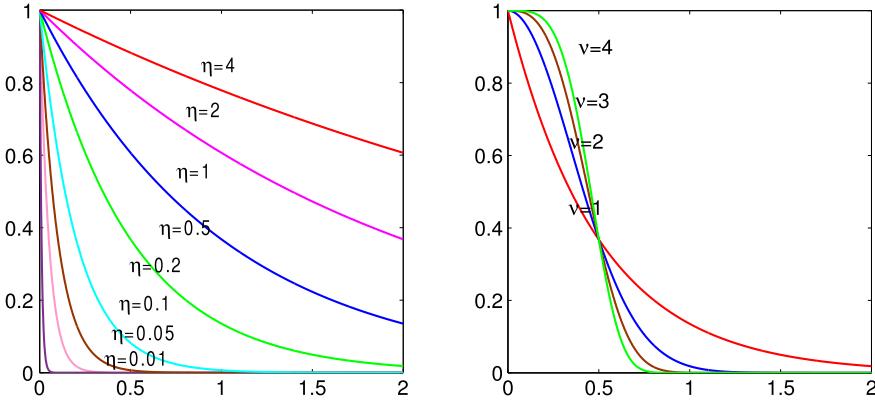


Fig. 1. Illustration of exponential functions. Left chart: $\eta = 0.01 \text{ \AA}, 0.05 \text{ \AA}, 0.1 \text{ \AA}, 0.2 \text{ \AA}, 0.5 \text{ \AA}, 1 \text{ \AA}, 2 \text{ \AA}, 4 \text{ \AA}$ and $\nu = 1$; Right chart: $\nu = 2, 3, 4, 5$ as $\eta = 0.5 \text{ \AA}$.

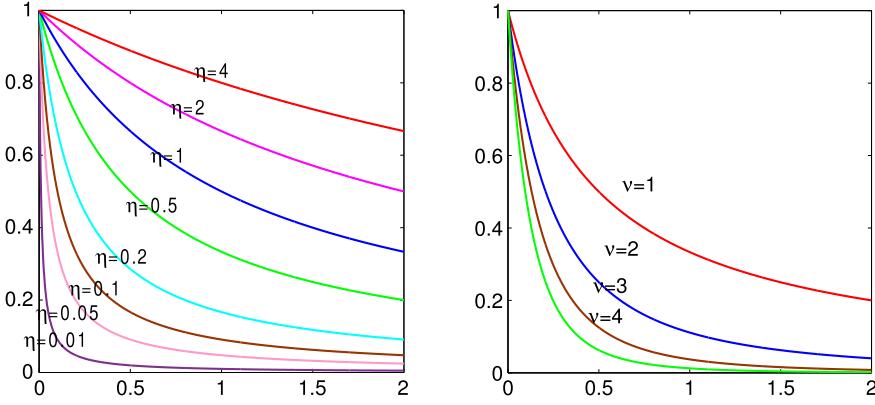


Fig. 2. Illustration of Lorentz functions. Left chart: $\eta = 0.01 \text{ \AA}, 0.05 \text{ \AA}, 0.1 \text{ \AA}, 0.2 \text{ \AA}, 0.5 \text{ \AA}, 1 \text{ \AA}, 2 \text{ \AA}, 4 \text{ \AA}$ as $\nu = 1$; Right chart: $\nu = 2, 3, 4, 5$ and $\eta = 0.5 \text{ \AA}$.

and generalized Lorentz functions

$$\Phi(\|\mathbf{r} - \mathbf{r}_j\|; \eta_j) = \frac{1}{1 + (\|\mathbf{r} - \mathbf{r}_j\|/\eta_j)^\nu}, \quad \nu > 0. \quad (5)$$

Essentially, the correlation between any two particles should have a fast decay. Mathematically, correlation functions are density estimators [58]. Physically, they are low-pass filters.

Figs. 1 and 2 demonstrate the decay property of exponential and Lorentz functions. It is noted that in these functions, a larger η_j value corresponds to a lower resolution and a slower decay. In wavelet theory, η_j is the resolution parameter.

We define the rigidity function or the heterogeneous modulus of a biomolecule as [9]

$$\mu(\mathbf{r}) = \sum_{j=1}^N w_j \Phi(\|\mathbf{r} - \mathbf{r}_j\|; \eta_j), \quad (6)$$

where w_j is a weight function which can be chosen as the atomic weight.

The rigidity function gives rise to the density distribution of a macromolecule and provides an analytical representation of the molecular structure in \mathbb{R}^3 . The morphology of the macromolecule can be visualized at either a common resolution or a set of resolutions. Obviously, the commonly used Gaussian surface is a special case of the present FRI model.

The resolution parameter (η_j) enables us to represent the data at the scale of interest. For simplicity, one may set η_j to a fixed constant η for all atoms or data entries in this work. If atoms in a biomolecular complex are classified into subsets $\{\alpha_k\}$ according to residues, alpha helices, beta strands, domains, or proteins, one can also represent each subset of atoms at a different resolution η_{α_k} .

One can also define a normalized rigidity function

$$\bar{\mu}(\mathbf{r}) = \frac{\mu(\mathbf{r})}{\max(\mu(\mathbf{r}))}, \quad (7)$$

where $\max(\mu(\mathbf{r}))$ is the maximal value of $\mu(\mathbf{r})$.

Furthermore, the molecular rigidity index can be defined as a summation of all the atomic ones

$$\mu_{\text{FRI}} = \int_{\Omega_m} \mu(\mathbf{r}) d\mathbf{r}.$$

The molecular rigidity index of a given macromolecule is a direct measure of its total interaction strength in a general sense.

Similarly, the flexibility function of a biomolecule is defined as

$$f(\mathbf{r}) = \frac{1}{\sum_{j=1}^N w_j \Phi(\|\mathbf{r} - \mathbf{r}_j\|; \eta_j)}. \quad (8)$$

The flexibility function is well-defined inside the biomolecule.

The correlation map is an important quantity for the GNM. FRI has a similar correlation map defined as $\Phi(\|\mathbf{r}_i - \mathbf{r}_j\|; \eta_j)$, $\forall i, j = 1, 2, \dots, N$. The correlation map measures the connectivity in the biomolecule. Additionally, the i th atomic rigidity index μ_i is given by

$$\mu_i = \mu(\mathbf{r}_i), \quad \forall i = 1, 2, \dots, N. \quad (9)$$

Similarly, we define i th atomic flexibility index as

$$f_i = f(\mathbf{r}_i) = \frac{1}{\mu(\mathbf{r}_i)}, \quad \forall i = 1, 2, \dots, N. \quad (10)$$

The FRI method predicts the B-factors of a biomolecule by minimizing the following

$$\min_{a,b} \left\{ \sum_{i=1}^N |af_i + b - B_i^e|^2 \right\}, \quad (11)$$

where B_i^e are the experimental B-factors for the i th particle.

The computational complexity of the original FRI is of $\mathcal{O}(N^2)$. In our fast FRI (ffFRI) algorithm, the summation can be limited to a few particles near \mathbf{r}_i according to the cell lists algorithm [59], which reduces the computational complexity to $\mathcal{O}(N)$ [52]. Anisotropic FRI (aFRI) [52] and multiscale FRI (mFRI) [51] have also been developed.

In the present work, we focus on the use of the rigidity function (6) for describing the molecular shape. Clearly, by setting $v = 2$ in Eq. (4), one has an FRI form of Gaussian surfaces. The utility of FRI rigidity functions is explored in terms of surface representation, electrostatic evaluation, surface area and surface enclosed volume in this work. Additionally, with the analytical rigidity expression, it is easy to investigate its differential geometry and algebraic topology.

Although the distance $\|\mathbf{r}_i - \mathbf{r}_j\|$ is defined as the Euclidean space distance for biomolecular atoms, it can be more generally defined in the present analysis, such as the distance between biological species or other entities. As a result, the rigidity index and flexibility index can be generalized for arbitrary data sets, such as social networks, biological networks, and graphs in the present formulation.

2.2. FRI based electrostatic modeling

FRI based Poisson–Boltzmann models. The normalized rigidity in Eq. (7) can be used as a smooth characteristic function for the molecular domain. As such, we can formulate a rigidity based Poisson–Boltzmann (RBPB) equation:

$$-\nabla \cdot (\epsilon(\bar{\mu}) \nabla \phi) = \bar{\mu} \rho_m + (1 - \bar{\mu}) \sum_{\alpha=1}^{N_c} \rho_{\alpha 0} q_{\alpha} e^{-q_{\alpha} \phi / k_B T} \quad (12)$$

where $\phi(\mathbf{r})$ is the electrostatic potential, and $\epsilon(\bar{\mu}) = \bar{\mu}(\mathbf{r})\epsilon_m + (1 - \bar{\mu}(\mathbf{r}))\epsilon_s$, with ϵ_s and ϵ_m being the dielectric constants of the solvent and solute, respectively. Here $\rho_m(\mathbf{r}) = \sum_j^{N_m} Q_j \delta(\mathbf{r} - \mathbf{r}_j)$ is the density of molecular charges, with N_m being the number of charged particles in the solute and Q_j being the partial charge on a particle located at \mathbf{r}_j , q_{α} is the charge of ion species α , N_c is the number of ion species, k_B is the Boltzmann constant, T is the temperature, and $\rho_{\alpha 0}$ is the bulk concentration of the α th ionic species. The term associated with $\bar{\mu}$ is in the solute domain and that with $(1 - \bar{\mu})$ is in the solvent domain.

The RBPB (12) has the advantage that it avoids the geometric singularity induced numerical instability. Fast computational algorithms such as multigrid methods can be easily applied.

We can turn the smooth characteristic function $\bar{\mu}$ into a sharp characteristic function χ

$$\chi(\mathbf{r}) = \begin{cases} 1, & \bar{\mu}(\mathbf{r}) \geq c, \\ 0, & \bar{\mu}(\mathbf{r}) < c. \end{cases} \quad (13)$$

A sharp FRI interface Γ by taking an isosurface value c

$$\Gamma = \{\mathbf{r} | \mu(\mathbf{r}) = c\}. \quad (14)$$

The computational domain Ω is split into a molecular subdomain Ω_m and a solvent one Ω_s :

$$\Omega = \Omega_m \cup \Gamma \cup \Omega_s. \quad (15)$$

Then, we can propose a sharp interface version of RBPB (12) in the form

$$-\nabla \cdot (\epsilon(\chi) \nabla \phi) = \chi \rho_m + (1 - \chi) \sum_{\alpha=1}^{N_c} \rho_{\alpha 0} q_\alpha e^{-q_\alpha \phi / k_B T}, \quad (16)$$

where $\epsilon(\chi) = \chi(\mathbf{r})\epsilon_m + (1 - \chi(\mathbf{r}))\epsilon_s$. Mathematically, because electrostatic potential $\phi(\mathbf{r})$ is defined on the whole computational domain ($\forall \mathbf{r} \in \Omega$), one has to impose the following interface conditions at the solvent–solute interface Γ in order to make the RBPB equation (16) being well-posed [60–62]

$$[\phi(\mathbf{r})] = 0, \quad \mathbf{r} \in \Gamma; \quad (17)$$

$$[\epsilon(\mathbf{r}) \nabla \phi(\mathbf{r})] \cdot \mathbf{n} = 0, \quad \mathbf{r} \in \Gamma, \quad (18)$$

where $[\cdot]$ denotes the difference of the quantity “.” cross the interface Γ , \mathbf{n} is the interfacial norm vector. Both Eqs. (12) and (16) can be solved subject to the following Debye–Hückel boundary condition

$$\phi(\mathbf{r}) = \frac{1}{4\pi} \sum_{i=1}^{N_m} \frac{Q_i}{\epsilon_s |\mathbf{r} - \mathbf{r}_i|} e^{-\bar{\kappa} |\mathbf{r} - \mathbf{r}_i|}, \quad \mathbf{r} \in \partial\Omega, \quad (19)$$

where $\bar{\kappa}$ is the Debye–Hückel parameter [47].

Electrostatic solvation free energy. The electrostatic solvation free energy in the PB model can be represented as:

$$\Delta G_{\text{el}} = \frac{1}{2} \sum_{i=1}^{N_m} Q_i (\phi(\mathbf{r}_i) - \phi_{\text{homo}}(\mathbf{r}_i)), \quad (20)$$

where $\phi_{\text{homo}}(\mathbf{r}_i)$ is obtained by solving the PBE by switching the dielectric constant ϵ_s in the solvent domain Ω_s to ϵ_m .

In the rest of this paper, for the sake of simplicity, we only consider the case that there is no ion in the solvent. The presence of ions in solvent does not affect our methods.

2.3. FRI based surface area and volume estimation

Smooth FRI surface. The scale rigidity in Eq. (12) offers a smooth representation of biomolecules. The corresponding surface area and volume can be computed by using geometric measure theory [1]. For example, the surface area can be obtained by

$$\text{Area} = \int_{\Omega} |\nabla \bar{\mu}(\mathbf{r})| d\mathbf{r}. \quad (21)$$

The volume expression is given by

$$\text{Vol} = \int_{\Omega} \bar{\mu}(\mathbf{r}) d\mathbf{r}. \quad (22)$$

Sharp FRI surface. The interface Γ is characterized by a set of intersecting points ($\mathbf{v}_o^l = (x_o^l, y_o^l, z_o^l)$, $\forall l = 1, 2, \dots$), where the interface intersects the mesh lines, and associated norms (\mathbf{N}_o^l , $\forall l = 1, 2, \dots$) at the intersecting points.

If an intersecting point $\mathbf{v}_o = \mathbf{v}(x_o, y_j, z_k) = \mathbf{v}_{o,j,k}$ is located between two grid points (x_i, y_j, z_k) and (x_{i+1}, y_j, z_k) , it can be approximated by

$$\mathbf{v}_o = \left(\frac{\mu(x_o, y_j, z_k) - \mu(x_i, y_j, z_k)}{\mu(x_{i+1}, y_j, z_k) - \mu(x_i, y_j, z_k)} (x_{i+1} - x_i), y_j, z_k \right), \quad (23)$$

where $\mu(x_o, y_j, z_k) = c$ is the recommended isovalue of the surface, $\mu(x_i, y_j, z_k)$ and $\mu(x_{i+1}, y_j, z_k)$ represent the rigidity at nodes (i, j, k) and $(i+1, j, k)$. The corresponding normal vector is interpolated by

$$\mathbf{N}_o = \frac{\mu(x_o, y_j, z_k) - \mu(x_i, y_j, z_k)}{\mu(x_{i+1}, y_j, z_k) - \mu(x_i, y_j, z_k)} (\mathbf{N}_{i+1,j,k} - \mathbf{N}_{i,j,k}) + \mathbf{N}_{i,j,k}, \quad (24)$$

where $\mathbf{N}_{i,j,k}$ is the normal vector at (i, j, k) . And it is calculated by

$$\mathbf{N}_{i,j,k} = \left(\frac{\mu(x_{i+1}, y_j, z_k) - \mu(x_{i-1}, y_j, z_k)}{x_{i+1} - x_{i-1}}, \frac{\mu(x_i, y_{j+1}, z_k) - \mu(x_i, y_{j-1}, z_k)}{y_{j+1} - y_{j-1}}, \frac{\mu(x_i, y_j, z_{k+1}) - \mu(x_i, y_j, z_{k-1})}{z_{k+1} - z_{k-1}} \right). \quad (25)$$

The surface integral of a function $A(x, y, z)$ can be evaluated as

$$\int_{\Gamma} A(x, y, z) dS = \sum_{o \in I} \left(\frac{A(x_o, y_j, z_k) |n_{o,x}|}{h} + \frac{A(x_i, y_o, z_k) |n_{o,y}|}{h} + \frac{A(x_i, y_j, z_o) |n_{o,z}|}{h} \right) h^3, \quad (26)$$

where I is the set of intersection points and h is the grid spacing. Here $n_{o,x}$, $n_{o,y}$ and $n_{o,z}$ are the x , y and z components of the unit normal vector $\mathbf{n}_o = \frac{\mathbf{N}_o}{\|\mathbf{N}_o\|}$ at the intersection point \mathbf{v}_o .

We can obtain a similar volume integral as:

$$\int_{\Omega_m} A(\mathbf{r}) d\mathbf{r} = \frac{1}{2} \left(\sum_{(i,j,k) \in I_1} A(x_i, y_j, z_k) h^3 + \sum_{(i,j,k) \in I_2} A(x_i, y_j, z_k) h^3 \right) \quad (27)$$

where I_1 is the set of grid points inside the interface and $I_2 = I_1 \cup I_0$ is the union of I_1 and I_0 , the set of irregular points defined as all the grid points that have a neighbor grid point from the other side of the interface Γ . Area and volume are obtained by setting $A(x, y, z) = 1$ in Eqs. (26) and (27), respectively.

2.4. FRI based curvature analysis

Embedded surface approach. The surface of interest can be extracted from a given level set with an isovalue μ_0 , i.e., setting $\mu(x, y, z) = \mu_0$. We assume μ to be non-degenerate, i.e., the norm of its gradient is non-zero when it is equal to μ_0 . Without loss of generality, we further assume that its projection onto z is non-zero. Then there exists a function $z = d(x, y)$. Denote the point on surface as $\langle x, y, d(x, y) \rangle$ and it holds $\mu(x, y, d(x, y)) = \mu_0$. Taking the partial differentiation with respect to x and y gives:

$$\mu_x(x, y, d(x, y)) + \mu_z(x, y, d(x, y))d_x = 0, \quad (28)$$

$$\mu_y(x, y, d(x, y)) + \mu_z(x, y, d(x, y))d_y = 0. \quad (29)$$

Furthermore, it can be derived $d_x = -\frac{\mu_x}{\mu_z}$ and $d_y = -\frac{\mu_y}{\mu_z}$.

We define $E(x, y, z)$, $F(x, y, z)$, $G(x, y, z)$, $L(x, y, z)$, $M(x, y, z)$ and $N(x, y, z)$ below to be the coefficients in the first and second fundamental forms. For simplicity, we omit parameter parts. Their values for surface S can be given as:

$$\begin{aligned} E &= \langle \mu_x, \mu_x \rangle, & F &= \langle \mu_x, \mu_y \rangle, & G &= \langle \mu_y, \mu_y \rangle, \\ L &= \langle \mu_{xx}, \mathbf{n} \rangle, & M &= \langle \mu_{xy}, \mathbf{n} \rangle, & N &= \langle \mu_{yy}, \mathbf{n} \rangle, \end{aligned}$$

where $\langle x, y \rangle$ denotes the internal product of x and y .

The Gaussian curvature K can be computed by

$$K = \frac{LN - M^2}{EG - F^2}. \quad (30)$$

Also, we can calculate the mean curvature H by

$$H = \frac{1}{2} \frac{LG - 2MF + NE}{EG - F^2}. \quad (31)$$

Substituting E, F, G, L, M, N into the previous equations, we can obtain the formulas as follows:

$$\begin{aligned} K &= \frac{2\mu_x\mu_y\mu_{xz}\mu_{yz} + 2\mu_x\mu_z\mu_{xy}\mu_{yz} + 2\mu_y\mu_z\mu_{xy}\mu_{xz}}{g^2} - \frac{2\mu_x\mu_z\mu_{xz}\mu_{yy} + 2\mu_y\mu_z\mu_{xx}\mu_{yz} + 2\mu_x\mu_y\mu_{xy}\mu_{zz}}{g^2} \\ &\quad + \frac{\mu_z^2\mu_{xx}\mu_{yy} + \mu_x^2\mu_{yy}\mu_{zz} + \mu_y^2\mu_{xx}\mu_{zz}}{g^2} - \frac{\mu_x^2\mu_{yz}^2 + \mu_y^2\mu_{xz}^2 + \mu_z^2\mu_{xy}^2}{g^2}, \end{aligned}$$

where $g = \mu_x^2 + \mu_y^2 + \mu_z^2$. The mean curvature is the average second derivative with respect to the normal direction,

$$H = -\frac{2\mu_x\mu_y\mu_{xy} + 2\mu_x\mu_z\mu_{xz} + 2\mu_y\mu_z\mu_{yz} - (\mu_y^2 + \mu_z^2)\mu_{xx} - (\mu_x^2 + \mu_z^2)\mu_{yy} - (\mu_x^2 + \mu_y^2)\mu_{zz}}{2g^{3/2}}.$$

Given the Gaussian and mean curvatures K and H , we can easily find the principal curvatures κ_1 (minimal curvature) and κ_2 (maximal curvature), which are the roots of

$$\kappa^2 - 2H\kappa + K = 0,$$

i.e., $\kappa = H \pm \sqrt{H^2 - K}$.

Hessian matrix approach. An alternative algorithm for the curvature extraction from volumetric data is the Hessian matrix method. For volumetric data $\mu(x, y, z)$, we define the surface gradient $\nabla\mu$ and surface norm \mathbf{n} ,

$$\nabla\mu = \langle \mu_x, \mu_y, \mu_z \rangle;$$

$$\mathbf{n} = -\frac{\nabla\mu}{\|\nabla\mu\|}.$$

The Hessian matrix, \mathbf{H} , is given by

$$\mathbf{H} = \begin{pmatrix} \frac{\partial^2 \mu}{\partial x^2} & \frac{\partial^2 \mu}{\partial x \partial y} & \frac{\partial^2 \mu}{\partial x \partial z} \\ \frac{\partial^2 \mu}{\partial x \partial y} & \frac{\partial^2 \mu}{\partial y^2} & \frac{\partial^2 \mu}{\partial y \partial z} \\ \frac{\partial^2 \mu}{\partial x \partial z} & \frac{\partial^2 \mu}{\partial y \partial z} & \frac{\partial^2 \mu}{\partial z^2} \end{pmatrix}. \quad (32)$$

The two principal curvatures can be evaluated by the following procedure.

1. Calculate matrix $\mathbf{P} = \mathbf{I} - \mathbf{n}\mathbf{n}^T$, here \mathbf{I} is the identity matrix and T denotes the transpose;
2. Evaluate matrix $\mathbf{G} = -\mathbf{P}\mathbf{H}\mathbf{P}/\|\nabla\mu\| = (g_{ij})_{(i,j=1,2,3)}$;
3. Calculate the trace t and Frobenius norm s of matrix \mathbf{G} :

$$t = g_{11} + g_{22} + g_{33}; \quad s = \|\mathbf{G}\| = \sqrt{\sum_i \sum_j g_{ij}^2}.$$

4. Compute the principle curvatures:

$$\kappa_1 = \frac{t + \sqrt{2s^2 - t^2}}{2}; \quad \kappa_2 = \frac{t - \sqrt{2s^2 - t^2}}{2}.$$

5. The Gaussian curvature K and mean curvature H can be easily calculated,

$$K = \kappa_1 \kappa_2; \quad H = \frac{\kappa_1 + \kappa_2}{2}.$$

On the basis of the signs of Gaussian and mean curvatures, the neighborhood of a surface point can be roughly classified as one of the eight different shapes, namely, pit, valley, saddle ridge, flat, minimal surface, saddle valley, ridge, and peak. In Table 1, it shows the type of shapes for each possible combination of signs.

3. Numerical experiments

In this section we carry out numerical experiments to demonstrate the utility of rigidity functions for multiscale modeling of molecular surfaces. We also examine the use of molecular rigidity surface as an alternative for surface characterization in Poisson–Boltzmann based electrostatic analysis. Since molecular rigidity surfaces are differential manifolds, it is interesting to analyze the curvatures of molecular rigidity surfaces. Finally, we consider the prediction of protein binding sites based on curvature and electrostatics.

3.1. Molecular rigidity surfaces

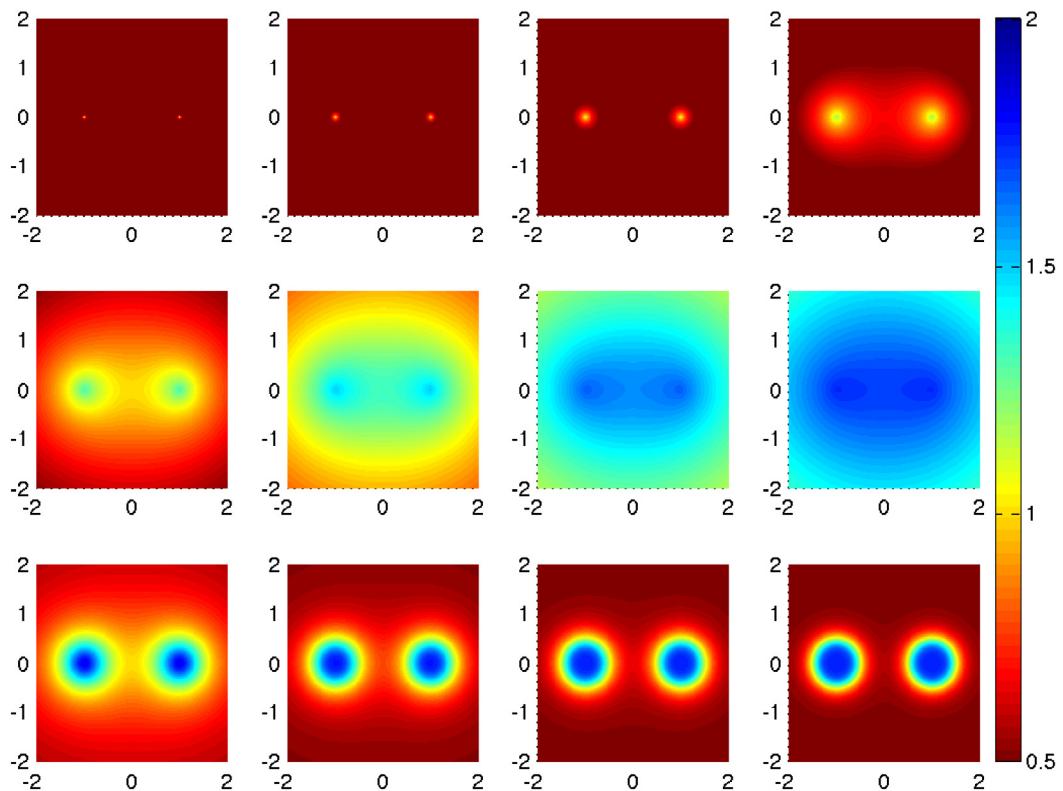
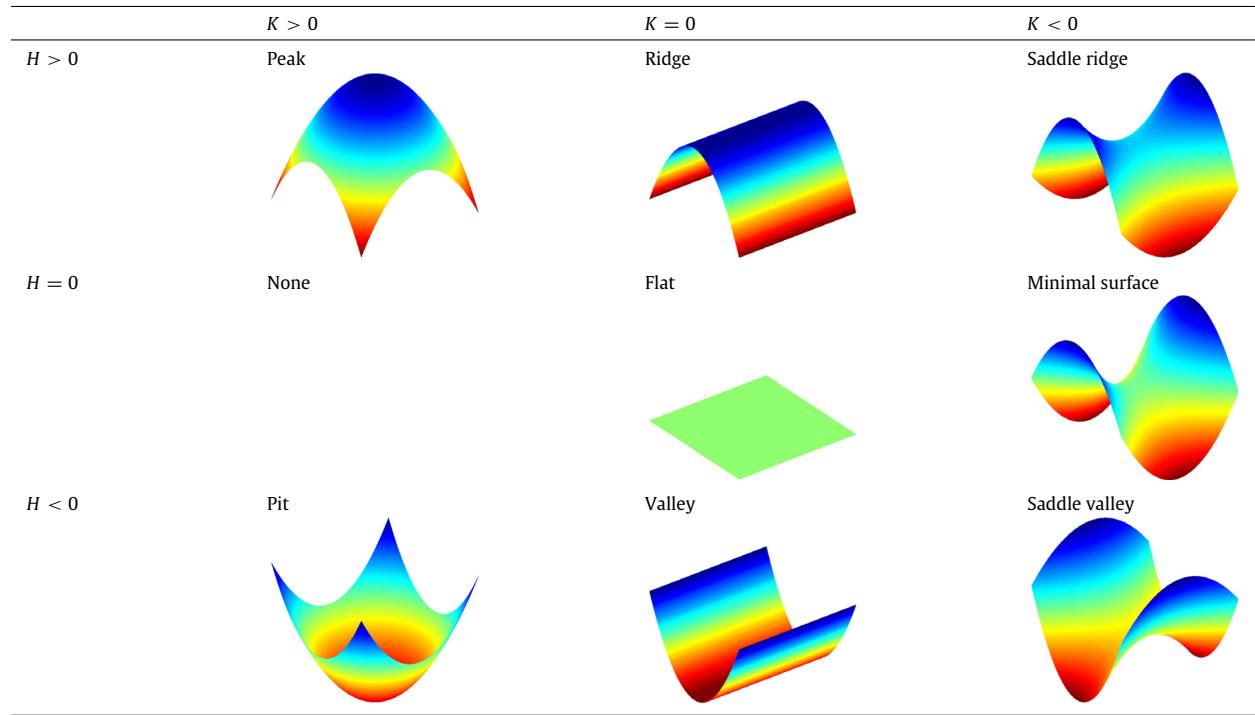
In this section, we illustrate the rigidity function representation of molecular density distributions. Both the exponential kernel (4) and the Lorentz kernel (5) based rigidity functions are studied over a number of different ν and η values. By varying ν values, one can analyze the effect of different kernels. The scale impact can be examined by varying η values.

Density distributions of simple systems. We first study the density distributions generated by using rigidity functions. To examine the behavior of different kernels and the effect of scales, we consider two simple systems. The first one consists of two atoms located at $(-1, 0)$ and $(1, 0)$ in a plane and the other one has three atoms located at $(-1, 0)$, $(1, 0)$ and $(0, \sqrt{3})$ in a plane.

Figs. 3 and 4 plot the density distributions of the two-atom system generated by Lorentz-kernel based rigidity function (5) and exponential-kernel based rigidity function (4), respectively. In both figures, the resolution parameter, η , has been varied over a wide range, i.e., from 0.05 to 6. Meanwhile we have also examined a number of kernel parameter ($\nu = 2, 3, 4$ and 5) to study the impact of the high order kernels. It is seen from Fig. 4 that for a given kernel, a larger η value leads to a more blurred rigidity density distribution. It appears that an η value that is of the order of the distance between atoms (i.e., atomic bond length) does a good job.

Table 1

Surface types according to Gaussian curvature and mean curvature.

**Fig. 3.** Density distributions for a two-atom system generated by using Lorentz kernels. Top row, $v = 1$ and from left to right, $\eta = 0.05 \text{ \AA}$, 0.1 \AA , 0.2 \AA and 0.5 \AA . Middle row, $v = 1$ and from left to right, $\eta = 1 \text{ \AA}$, 2 \AA , 4 \AA and 6 \AA . Bottom row, $\eta = 0.5 \text{ \AA}$ and from left to right, $v = 2, 3, 4$ and 5 .

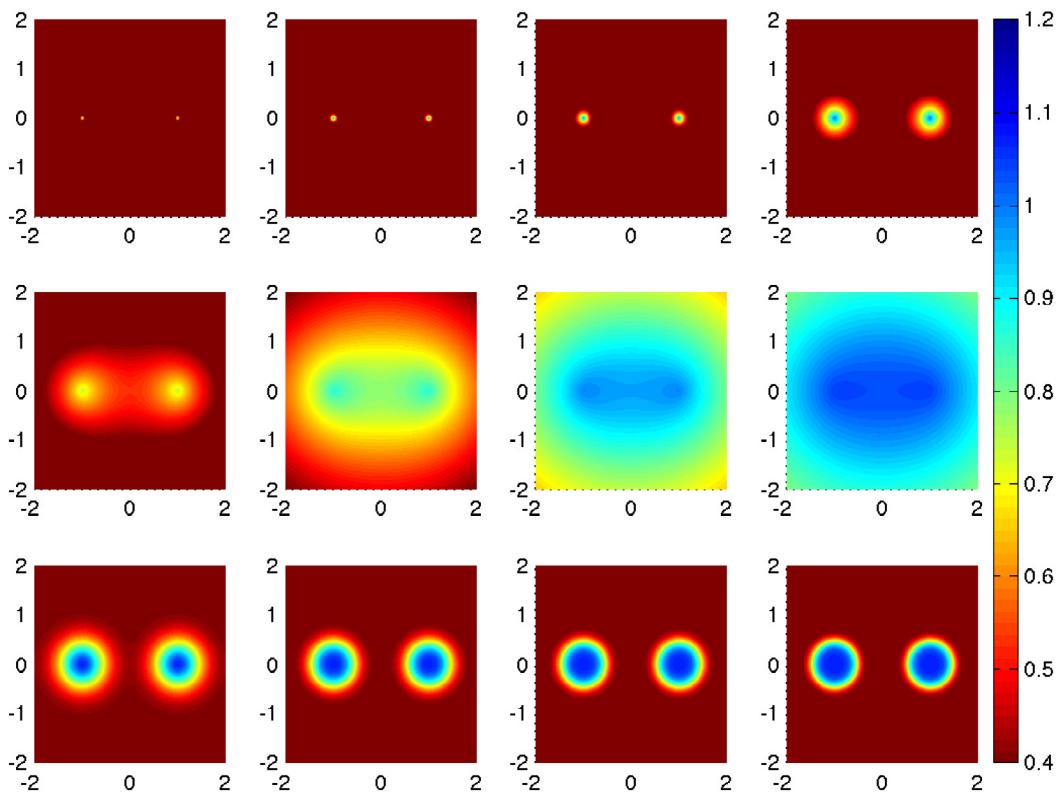


Fig. 4. Density distributions for a two-atom system generated by using exponential kernels. Top row, $v = 1$ and from left to right, $\eta = 0.05 \text{ \AA}$, 0.1 \AA , 0.2 \AA and 0.5 \AA . Middle row, $v = 1$ and from left to right, $\eta = 1 \text{ \AA}$, 2 \AA , 4 \AA and 6 \AA . Bottom row, $\eta = 0.5 \text{ \AA}$ and from left to right, $v = 2, 3, 4$ and 5 .

As shown in Fig. 3, for a given η value, the increase in the power of the Lorentz kernel (v) results in a sharper contrast between particles and their surrounding. In fact, when $v \rightarrow \infty$, the rigidity density profile (μ) behaves like a step function. Therefore, to have a smooth interface profile, one might use small v values, such as 1, 2 and/or 3. Similar behavior is found for exponential kernels shown in Fig. 4. In comparison of exponential kernel based rigidity functions give rise to more localized morphology. Again, an η value that is of the order of the atomic radius is recommended and a relatively small v value, namely, $v = 1, 2$, is preferred.

By taking into the consideration of computational efficiency, an $O(N)$ algorithm is better established with a relatively large v . Therefore, we recommend $v = 2$ in Lorentz-kernel based surface generation.

We further study the three-atom system in Figs. 5 and 6 using Lorentz kernel based and exponential kernel based rigidity functions, respectively. It is seen that there is a similar trend as observed in the two-atom system. These figures indicate that both Lorentz-kernel based and exponential-kernel based rigidity functions can be employed to represent molecular electron densities.

FRI surface of biomolecules. A major purpose of the present work is to examine the suitability of rigidity functions for representing biomolecules. Certainly, different forms of Gaussian surfaces have been used in the past. The FRI theory offers a wide class of rigidity functions that are potentially useful biomolecular representations. In this section, we explore the performance of Lorentz kernel based and exponential kernel based rigidity functions.

To illustrate rigidity functions, we consider a cross-linking of DNA duplexes, PDB ID: 2GJB. In this test, protein surfaces at various scales are generated by MSMS, Lorentz kernel based rigidity, and exponential kernel based rigidity.

Fig. 7 depicts our comparison. Obviously, the solvent excluded surfaces in the top row are non-smooth, which implies the difficulty in solving the Poisson–Boltzmann equation for electrostatic potential and computing electrostatic forces. In contrast, the rigidity functions based on FRI kernels in the second and third rows are smooth and free of geometric singularities. Therefore, they provide good alternatives for biomolecular modeling, electrostatic analysis and visualization.

3.2. Area and volume of FRI surfaces

Having examined the qualitative aspects of rigidity functions for biomolecules, we provide some quantitative analysis. Surface area and surface enclosed volume are two quantities that are often used in biomolecular modeling, such as in solvation and binding analysis.

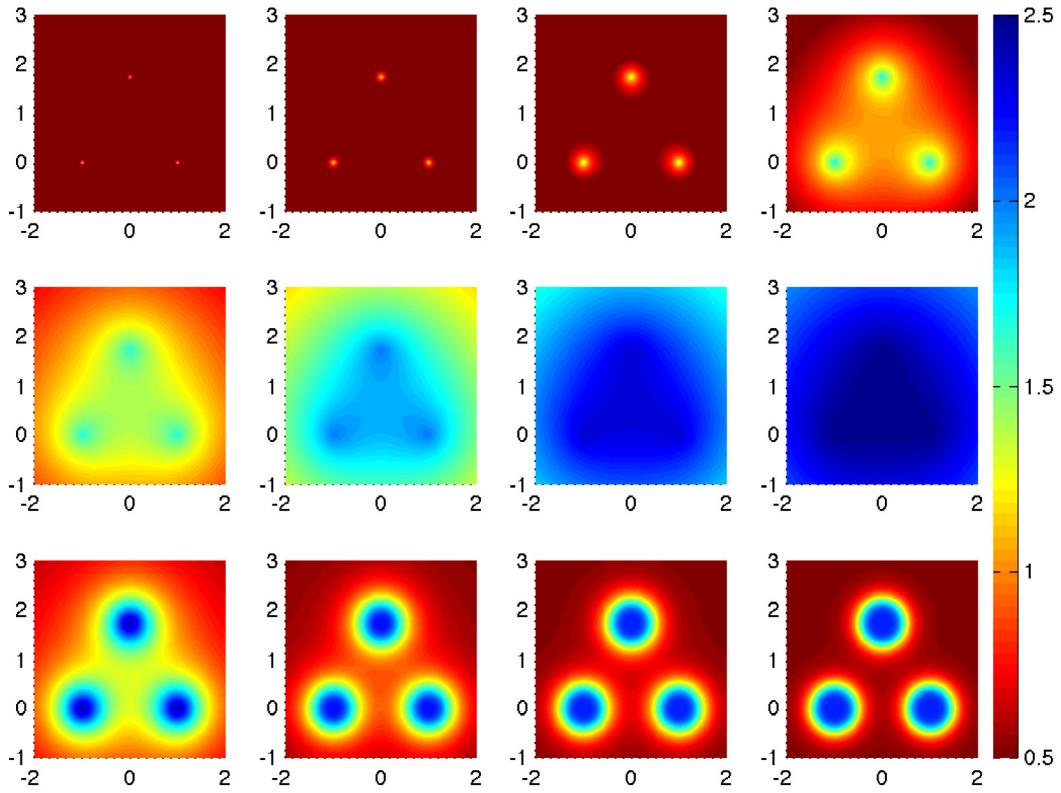


Fig. 5. Density distributions for a three-atom system generated by using Lorentz kernels. Top row, $v = 1$ and from left to right, $\eta = 0.05 \text{ \AA}$, 0.1 \AA , 0.2 \AA and 0.5 \AA . Middle row, $v = 1$ and from left to right, $\eta = 1 \text{ \AA}$, 2 \AA , 4 \AA and 6 \AA . Bottom row, $\eta = 0.5 \text{ \AA}$ and from left to right, $v = 2, 3, 4$ and 5 .

We first investigate the convergence of our formula (26)–(27) in volume and area calculations. To this end, we consider the analytical surface given by a sphere $x^2 + y^2 + z^2 = 1$ in domain $[-2, 2] \times [-2, 2] \times [-2, 2]$. The total area is 4π and volume is $\frac{4}{3}\pi$.

We set the isovalue as $\mu_0 = x^2 + y^2 + z^2 - 1 = 0$ for our test. The errors and convergent orders of our algorithms are given in Table 2 for different mesh sizes. The central finite difference scheme will be used with a three-point stencil for discretization. Through the analysis of error and mesh size, it is observed that our algorithm is essentially of the second order in accuracy.

Table 3 reports a comparison of surface areas and surface-enclosed volumes for 21 proteins generated by the MSMS package and our rigidity representations. In this test, we choose the exponential kernel based rigidity surface (FRI surface 1) with $\eta = 1.85 \text{ \AA}$ and $v = 2$ and then set $\mu_0 = 1.5$. Meanwhile, we set Lorentz kernel based rigidity surface (FRI surface 2) with $\eta = 1.86 \text{ \AA}$, $v = 8$, and $\mu_0 = 1.5$. It is seen from Table 3 that surface areas and surface enclosed volumes generated by these three approaches compare reasonably well, which indicates the proposed FRI is useful in biophysical applications. In fact, a better match in areas and volumes can be achieved. However, we have kept in mind that the present surface models should also be able to deliver consistent results in electrostatic analysis. This aspect is examined in the next section.

3.3. Electrostatic analysis using rigidity surface based Poisson equation

Electrostatic analysis, including solvation free energy estimation, is a major task in computational biophysics. The same sets of parameters as shown in the area and volume calculations are employed to obtain FRI surfaces for defining the Poisson equation. In this work, we employ the sharp surface version in Eq. (16). Therefore, we use our MIBPB method [24,63,61,64] to compute electrostatic solvation free energies of the 21 proteins studied in the last section. All the original data are downloaded from the PDB bank. Water molecules are removed and hydrogen atoms are added in each structure. The AMBER force field was used for the charge assignment.

Table 4 shows the comparison of electrostatic solvation free energies computed from three sets of surfaces for 21 proteins. In general, electrostatic solvation free energies are very sensitive to the surface definition. It was noticed in the literature [34] that it is impossible for the areas, volumes and electrostatic solvation free energies computed from Gaussian surfaces to agree simultaneously with those computed from solvent excluded surfaces. Therefore, the present results are in reasonably good agreement with those of the MSMS surfaces.

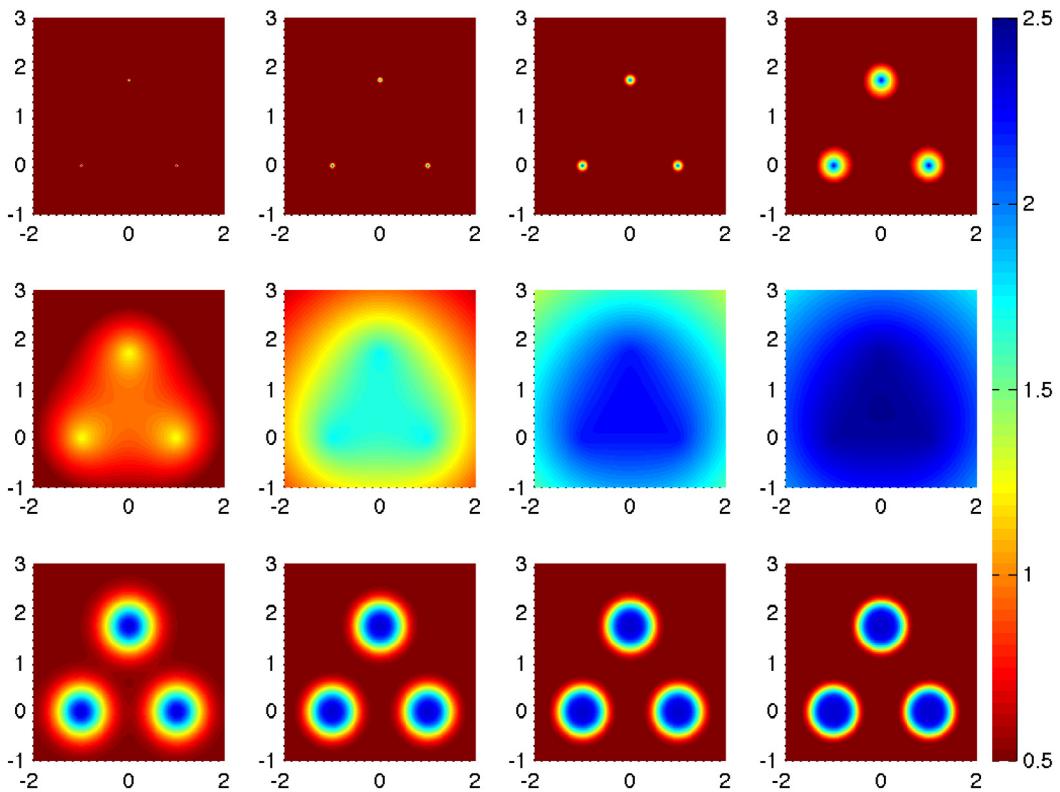


Fig. 6. Density distributions for a three-atom system generated by using exponential kernels. Top row, $v = 1$ and from left to right, $\eta = 0.05 \text{ \AA}$, 0.1 \AA , 0.2 \AA and 0.5 \AA . Middle row, $v = 1$ and from left to right, $\eta = 1 \text{ \AA}$, 2 \AA , 4 \AA and 6 \AA . Bottom row, $\eta = 0.5 \text{ \AA}$ and from left to right, $v = 2, 3, 4$ and 5 .

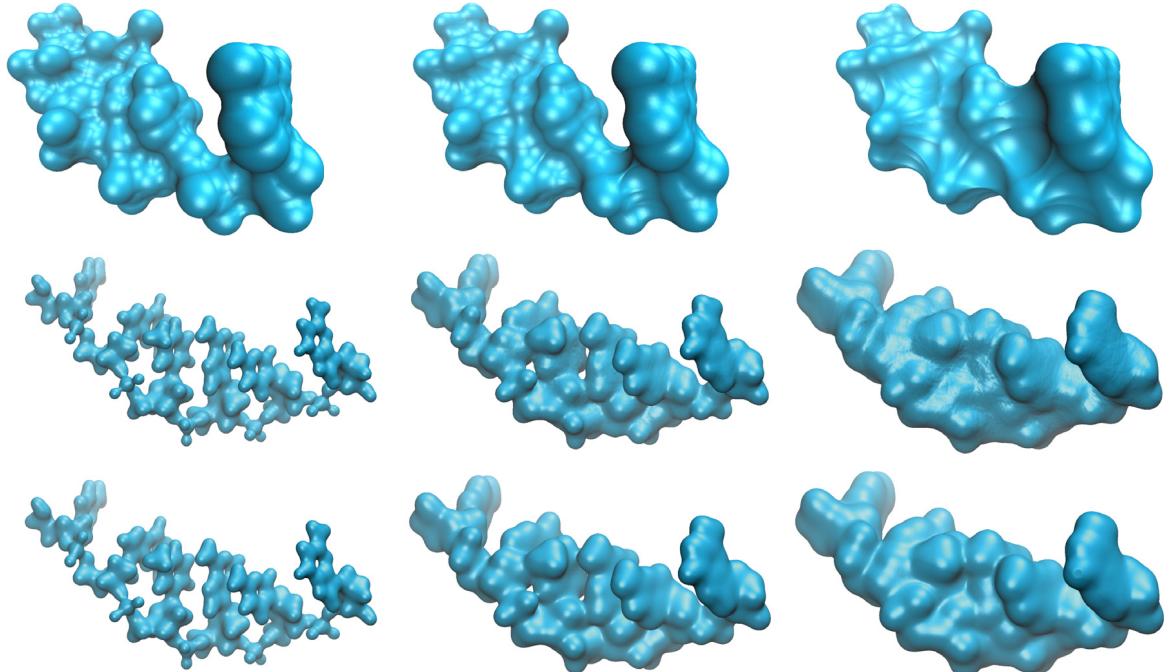


Fig. 7. Crosslinking of DNA duplexes, PDB ID: 2GJB. Top row: MSMS surfaces with probe 1.0, 1.5 and 2.5; Middle row: Lorentz kernel based rigidity surfaces with $\eta = (1.0 \text{ \AA}, 1.5 \text{ \AA}, 2.0 \text{ \AA})$ and $v = 6.0$ from left to right; Bottom row: Exponential kernel based rigidity surfaces with $\eta = (0.5 \text{ \AA}, 0.7 \text{ \AA}, 0.9 \text{ \AA})$ and $v = 1.0$ from left to right.

Table 2

Numerical convergence of the surface area and surface enclosed volume of a sharp interface in the Cartesian representation with $\mu_0 = x^2 + y^2 + z^2 - 1 = 0$.

$n_x \times n_y \times n_z$	Volume		Surface area	
	L^∞ error	Order	L^2 error	Order
10 × 10 × 10	2.5790e−2		1.3887e−01	
20 × 20 × 20	1.4415e−2	0.8392	3.0740e−02	2.1755
40 × 40 × 40	5.1183e−3	1.4938	6.4685e−03	2.2486
80 × 80 × 80	2.0031e−3	1.3534	2.4247e−03	1.4156
160 × 160 × 160	4.7990e−4	2.0614	5.2829e−04	2.1984
320 × 320 × 320	1.2727e−4	1.9148	1.0324e−04	2.3553

Table 3

Comparison of surface areas and surface-enclosed volumes for proteins using surfaces generated by the MSMS package, exponential-kernel based rigidity (FRI surface 1) with $\eta = 1.85 \text{ \AA}$, $v = 2$, $\mu_0 = 1.5$, and Lorentz-kernel based rigidity (FRI surface 2) with $\eta = 1.86 \text{ \AA}$, $v = 8$, and $\mu_0 = 1.5$.

Protein ID	MSMS surface		FRI surface 1		FRI surface 2	
	Volume (\AA^3)	Area (\AA^2)	Volume (\AA^3)	Area (\AA^2)	Volume (\AA^3)	Area (\AA^2)
1ajj	4543.526	2167.107	4779.875	2031.187	4501.297	2255.566
1vii	4996.193	2502.176	5381.172	2310.373	5076.906	2539.835
1bor	6996.808	2915.245	7651.188	2799.509	7194.531	3280.035
451c	10 809.886	4183.758	11 178.141	3808.930	10 520.875	4338.010
1svr	11 886.062	4673.309	13 196.156	4692.398	12 365.438	5489.018
1uxc	6759.569	2832.460	7279.656	2686.962	6923.438	2986.594
1mbg	7838.361	3086.519	8212.016	2892.108	7736.859	3295.856
1ptq	7018.859	2900.185	7277.688	2667.713	6919.391	2922.009
1sh1	6281.299	2746.515	6523.984	2589.967	6101.984	2968.278
2pde	5785.250	2661.122	5904.281	2461.975	5530.797	2718.888
1hpt	7503.161	3278.884	7844.172	2994.043	7426.063	3335.204
1a7m	23 948.055	7771.951	25 752.281	7616.473	24 215.266	9216.854
1neq	10 655.091	4743.671	10 980.969	4466.795	10 253.141	5126.247
1r69	8635.646	3054.674	8932.828	2748.288	8543.766	3053.245
1a2s	11 576.573	4457.460	11 746.406	4064.511	11 047.672	4696.929
2erl	5076.108	2327.632	5282.641	2128.229	4997.328	2388.872
1bbl	5193.353	2607.044	5262.313	2465.358	4914.016	2734.832
1fca	6931.455	2562.609	6735.344	2496.093	6324.453	2827.927
1frd	13 272.642	4376.376	13 542.484	3976.146	12 871.938	4649.500
1bpi	7862.953	3249.312	8316.047	3041.225	7851.125	3462.000
1a63	18 003.830	7026.165	18 964.672	6599.649	17 851.828	7643.768

Table 4

Comparison of electrostatic solvation free energies for proteins using surfaces generated by the MSMS package, exponential kernel based rigidity (FRI surface 1) with $\eta = 1.85 \text{ \AA}$, $v = 2$, and $\mu_0 = 1.5$, and Lorentz kernel based rigidity (FRI surface 2) with $\eta = 1.86 \text{ \AA}$, $v = 8$, and $\mu_0 = 1.5$.

Protein ID	MSMS surface energy (kcal/mol)	FRI surface 1 energy (kcal/mol)	FRI surface 2 energy (kcal/mol)
1ajj	−1100.754	−1155.158	−1258.784
1vii	−862.865	−761.195	−846.414
1bor	−927.310	−1021.579	−1140.463
451c	−1003.17	−971.629	−1152.786
1svr	−1582.131	−1530.159	−1700.400
1uxc	−1097.189	−975.583	−1075.070
1mbg	−1340.086	−1329.440	−1412.048
1ptq	−800.130	−765.680	−866.256
1sh1	−729.626	−648.315	−785.820
2pde	−1234.229	−863.702	−990.056
1hpt	−788.626	−768.424	−873.410
1a7m	−2173.814	−2159.535	−2492.651
1neq	−1683.679	−1661.077	−1811.667
1r69	−1115.733	−983.129	−1087.688
1a2s	−1868.827	−2216.464	−2356.554
2erl	−894.960	−1127.617	−1189.821
1bbl	−970.053	−993.386	−1052.617
1fca	−1148.672	−1427.109	−1534.746
1frd	−2691.339	−2935.189	−3106.112
1bpi	−1267.063	−1170.959	−1269.508
1a63	−2291.449	−2233.139	−2495.664

Table 5

Numerical convergence of the Gaussian curvature, mean curvature, and principle curvatures (κ_1, κ_2) the surface given in Eq. (33).

$n_x \times n_y$	Gaussian curvature		Mean curvature		Principle curvature	
	L^∞ error	L^2 -error	L^∞ error	L^2 -error	L^∞ error	L^2 -error
10 × 10	1.8545e-3	3.6541e-3	1.9925e-2	1.2885e-2	1.9450e-3	4.8901e-3
20 × 20	4.8792e-4	9.1656e-4	5.1874e-3	3.2668e-3	5.3034e-4	1.2267e-3
40 × 40	1.2392e-4	2.2957e-4	1.3232e-3	8.2565e-4	1.3590e-4	3.0726e-4
80 × 80	3.1143e-5	5.7434e-5	3.3414e-4	2.0756e-4	3.4231e-5	7.6878e-5
160 × 160	7.8009e-6	1.4363e-5	8.3954e-5	5.2036e-5	8.5793e-6	1.9227e-5
320 × 320	1.9518e-6	3.5914e-6	2.1041e-5	1.3027e-5	2.1469e-6	4.8076e-6
Order	1.9815	1.9983	1.9791	1.9902	1.9702	1.9982

For biophysical modeling, it is not necessary for two different surface definitions to have the same surface area, volume and electrostatic properties. In particular, the well-accepted area, volume and electrostatic potential from the solvent excluded surfaces are by no means more physical than those obtained from a Gaussian surface or FRI rigidity surfaces and vice-versa. In practical applications, such as solvation modeling and analysis, the solvation energy functional typically includes contributions from area, volume, and electrostatic terms. The associated fitting parameters, namely, surface tension and relative pressure, can be adjusted to adapt to different surface definitions.

3.4. Curvature evaluation

Membrane curvature sensing, curvature-sensitive proteins and curvature-mediated interactions between membrane proteins are popular topics in biology and biophysics. Therefore, curvature plays an important role in the biological world. Mathematically, curvatures, including mean, Gaussian, minimum and maximum curvatures are very sensitive measures of local geometric variations. Curvature information leads to a good understanding of geometric shape and provides a prediction of ligand binding potential. Both van der Waals surface and solvent excluded surface are not differentiable and their curvatures are not well defined. Numerically computed curvatures from van der Waals surface and solvent excluded surface are erroneous due to geometric singularities. The FRI surfaces developed in this work provide an excellent platform for both numerical and analytical calculations of molecular curvatures.

In the following, the central finite difference scheme of a three-point stencil will be used in the derivative calculation. Next, we shall validate the convergence rate in the curvatures calculation.

We set domain as $[-10, 10] \times [-10, 10] \times [-10, 10]$, and construct a surface as

$$d(x, y) = \frac{(x^2 - y^2)(x^3 - y^3)}{2000}. \quad (33)$$

The analytical expressions for Gaussian curvature and mean curvature are as follows,

$$K = \frac{d_{xx}d_{yy} - d_{xy}^2}{(1 + d_x^2 + d_y^2)^2}, \quad (34)$$

$$H = \frac{(1 + d_x^2)d_{yy} - 2d_x d_y d_{xy} + (1 + d_y^2)d_{xx}}{2(1 + d_x^2 + d_y^2)^{3/2}}. \quad (35)$$

It should be notice that the mean curvature equation above is different from our previous expression in Eq. (32) by a minus sign. This is due to the different way of looking at the surface, i.e., viewing from top and viewing from bottom will produce mean curvature with different sign. Simply speaking, a pit will be equivalent to a peak if we observe from an opposite direction.

The principle curvature can be calculated by

$$\kappa = H \pm \sqrt{H^2 - K}. \quad (36)$$

The L^∞ -error and L^2 -error for Gaussian curvature, mean curvature, and principle curvatures are reported in Table 5. We found that errors of two principle curvatures behave identically and thus only one set of results is presented. In all cases, errors reduce in a near perfect manner as the grid size increases. The convergence rates for all errors are of order $O(h^2)$. The plots of Gaussian, mean, minimum, and maximum curvatures are provided in Fig. 8.

We consider one more case to test our numerical method. In this case, we set the domain as $[-10, 10] \times [-10, 10] \times [-10, 10]$, and set a surface by

$$d(x, y) = \frac{x^3 - y^3}{30}. \quad (37)$$

The analytical Gaussian, mean, and principle (maximum and minimum) curvatures can be computed by Eqs. (34)–(36).

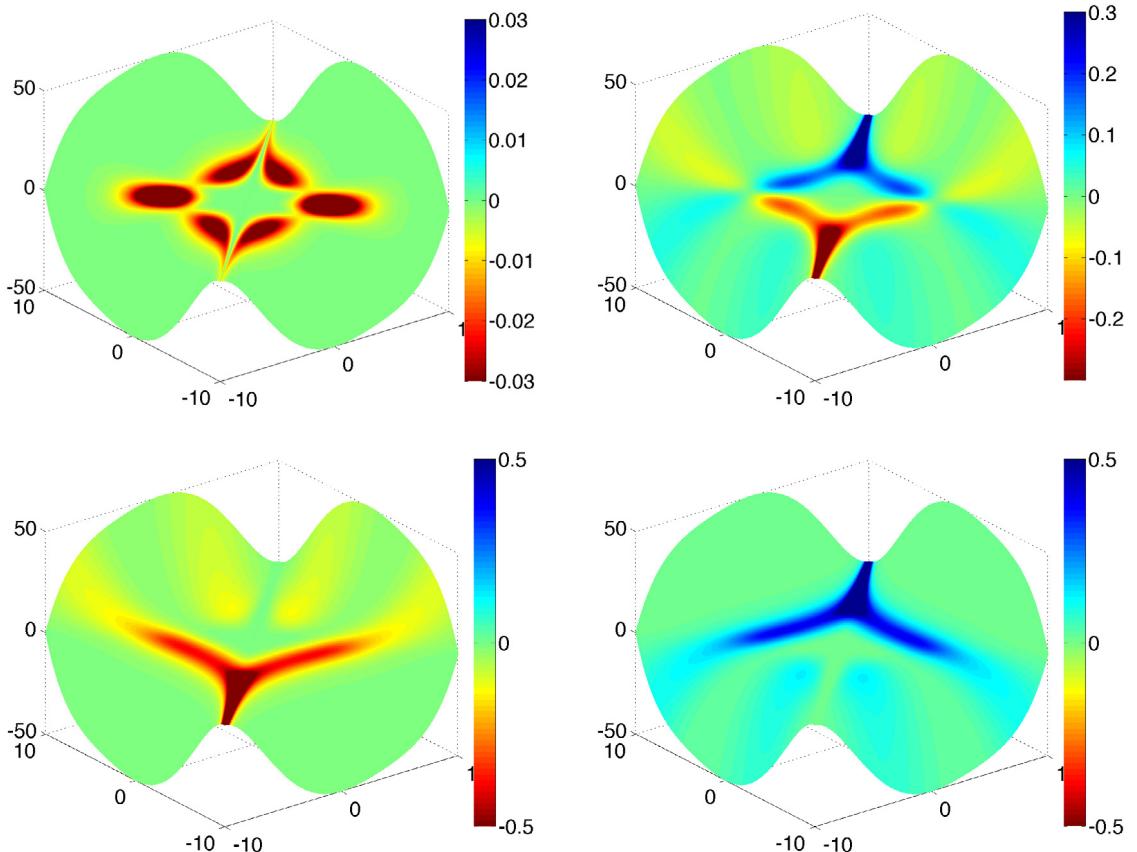


Fig. 8. Illustration of Gaussian curvature K (top left), mean curvature H (top right), minimum curvature κ_1 (bottom left), and maximum curvature κ_2 (bottom right) for the surface given in Eq. (33).

Table 6

Numerical convergence of the Gaussian curvature, mean curvature, and principle curvatures (κ_1 , κ_2) the surface given in Eq. (37).

$n_x \times n_y$	Gaussian curvature		Mean curvature		Principle curvatures	
	L^∞ error	L^2 -error	L^∞ error	L^2 -error	L^∞ error	L^2 -error
10 × 10	1.2172e-4	4.6484e-4	7.8877e-5	2.7691e-4	1.6298e-4	6.8733e-4
20 × 20	3.0445e-5	1.1623e-4	1.9747e-5	6.9241e-5	4.0829e-5	1.7185e-4
40 × 40	7.6134e-6	2.9058e-5	4.9375e-6	1.7311e-5	1.0208e-5	4.2966e-5
80 × 80	1.9034e-6	7.2647e-6	1.2346e-6	4.3278e-6	2.5519e-6	1.0742e-5
160 × 160	4.7585e-7	1.8162e-6	3.0864e-7	1.0820e-6	6.3798e-7	2.6854e-6
320 × 320	1.1897e-7	4.5405e-7	7.7161e-8	2.7049e-7	1.5950e-7	6.7136e-7
Order	1.9998	1.9999	1.9996	1.9999	1.9996	1.9999

The L^∞ -error and L^2 -error for Gaussian curvature, mean curvature, and principle curvatures are reported in Table 6. It shows that the convergence rate for each error is of order $O(h^2)$. Gaussian, mean, minimum, and maximum curvatures are illustrated in Fig. 9.

Having validated the convergence of our numerical methods, we explore the curvatures of protein surfaces generated by our FRI method. We consider protein 1PPL. Due to the geometric singularities, curvatures cannot be well defined for solvent excluded surface and van der Waals surface, as these surfaces are not even smooth. State differently, their first order derivative may not be well-defined on geometric singularities. Direct numerical calculations of curvatures on these surfaces lead to numerous errors. However, our FRI surface definition does not have this problem. Fig. 10 demonstrates the minimal curvature of FRI surfaces with isovalue $C = 2$ and $C = 1$. Obviously, by appropriate selections of the FRI rigidity density values, one can achieve the desirable multiscale representation of protein surfaces. The Left chart shows atomic information, colored by minimal curvature. One can further project other information, i.e., atomic type, atomic bond, atomic electrostatic potential, atomic hydrophobicity or atomic hydrophilicity, on the surface. The Right chart outlines the coarse scale information, such as cavity, ligand binding pocket, etc.

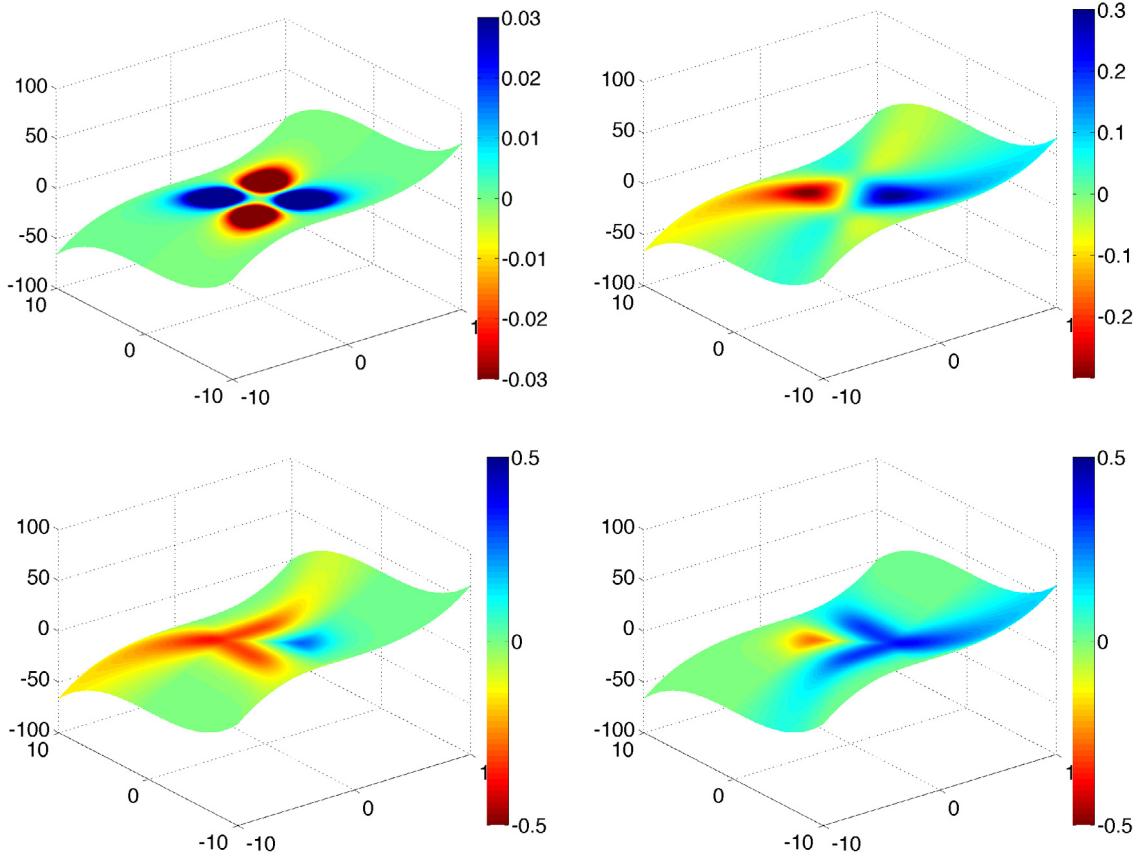


Fig. 9. Case 5: Plots of Gaussian curvature K (top left), mean curvature H (top right), minimum curvature κ_1 (bottom left), and maximum curvature κ_2 (bottom right) the surface given in Eq. (37).

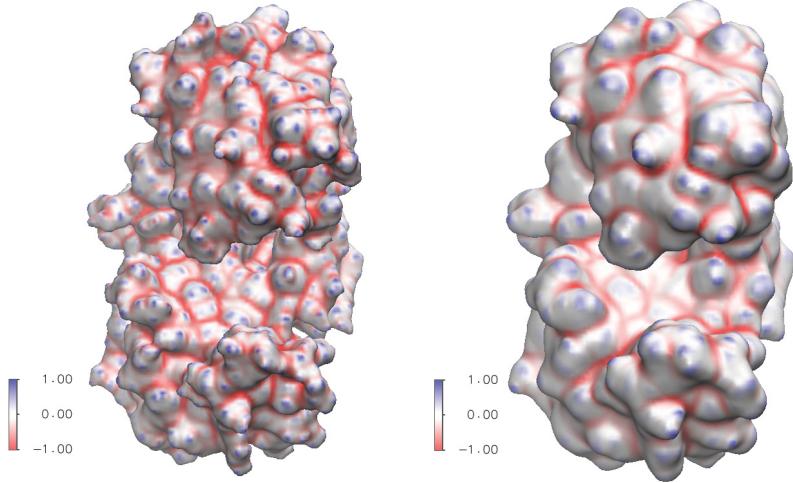


Fig. 10. Minimal curvature of protein 1PPL. Left to right: FRI surface with exponential kernel $v = 1, \eta = 0.5 \text{ \AA}$ for isovalue 2; Right Chart: FRI surface with exponential kernel $v = 1, \eta = 0.5 \text{ \AA}$ for isovalue 1. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

3.5. Polarized curvature and binding site prediction

Surface geometry is of significant importance in drug design and biomolecular modeling. In protein–ligand binding, the geometry of the ligand should match to the geometry of the protein just like the key and lock relation. More specifically, a concave region of the protein is usually the binding site of the ligand. The same principle applies when a substrate binds to

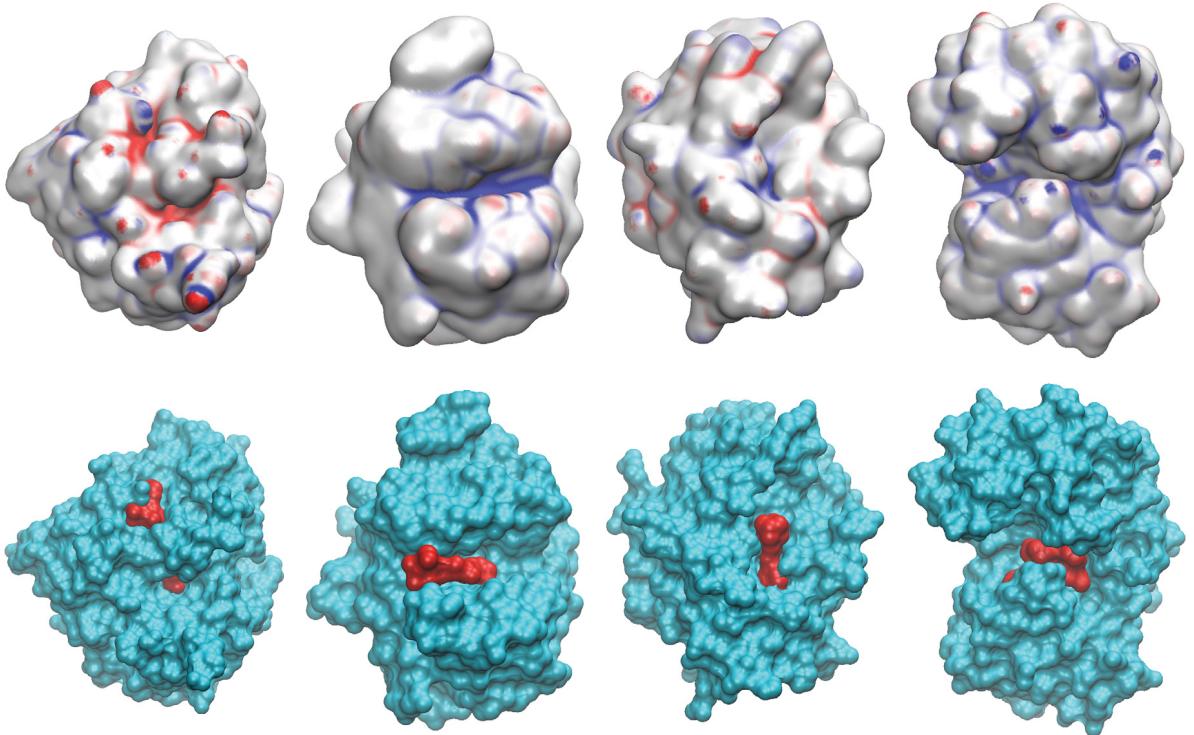


Fig. 11. Protein binding site prediction using polarized curvatures. From left to right, 1ADS, 1BYH, 1EJN, and 2WEB. Top row shows the prediction binding sites. Bottom row depicts the experimental binding sites.

Table 7

Polarized curvatures as binding indicators of protein surfaces. Denote maximal curvature, minimal curvature, positive electrostatic surface potential, and negative electrostatic surface potential as κ_1 , κ_2 , ϕ^+ , and ϕ^- .

$\kappa_1 > 0$	$\kappa_2 < 0$
$\phi^+ > 0$	Site for negatively charged protein
$\phi^- < 0$	Site for positively charged protein

$\kappa_1 > 0$	$\kappa_2 < 0$
	Site for negatively charged small ligand
	Site for positively charged small ligand

the active site of an enzyme or when a protein binds to DNA molecule. Curvatures describe the shape of the biomolecular geometry and have a great potential for quantitative modeling and analysis of geometric impact on biomolecular interaction and specification.

Among four major curvatures, Gaussian curvature is the product of two principle curvatures and characterizes the global topological property of a biomolecular surface. For example, when integrated over the surface, Gaussian curvature gives rise to the information of the genus number, as described by the Gauss–Bonnet theorem. Genus number can be used to describe the open or close states of ion channel proteins. Maximum curvature is good for describing the convex parts of a protein. Convex regions are important for protein–protein interactions. In contrast, concave regions can be best described by minimum curvature. Mean curvature, given by the average of two principle curvature, plays a crucial role in surface area minimization and differential geometry based multiscale modeling [1,45,46].

In addition to geometric shape, surface electrostatics plays an important role protein–drug binding, protein–protein interaction and protein specification. The surface electrostatic property can be easily analyzed by the Poisson–Boltzmann model. In general, charged or polar ligands favor protein polar or charged surface regions. Nonpolar ligands prefer the nonpolar sites of proteins. Therefore, that a potential drug binding site on a protein should be both electrostatically compatible and geometrically compatible. A simple model to combine these compatibilities is to take the products of electrostatic potential and curvatures. We call such products polarized curvatures. To this end, we use maximal curvature (κ_1), minimal curvature (κ_2), positive electrostatic surface potential (ϕ^+) and negative electrostatic surface potential (ϕ^-) to arrive at four different potential binding sites as summarized in Table 7. Specifically, a large amplitude of $\phi^+ \times \kappa_1$ on a protein surface indicates a potential binding site for a negatively charged protein surface, whereas a large amplitude of $\phi^+ \times \kappa_2$ implies a potential binding site for a negatively charged small ligand. Specific descriptions are given to other situations, namely, $\phi^- \times \kappa_2$ and $\phi^- \times \kappa_1$.

Figs. 11 and 12 demonstrate the predictions using our polarized curvatures and experimental binding sites. In Fig. 11, the binding site of protein 1ADS is positively charged, whereas the binding sites of proteins 1BYH, 1EJN, and 2WEB are all

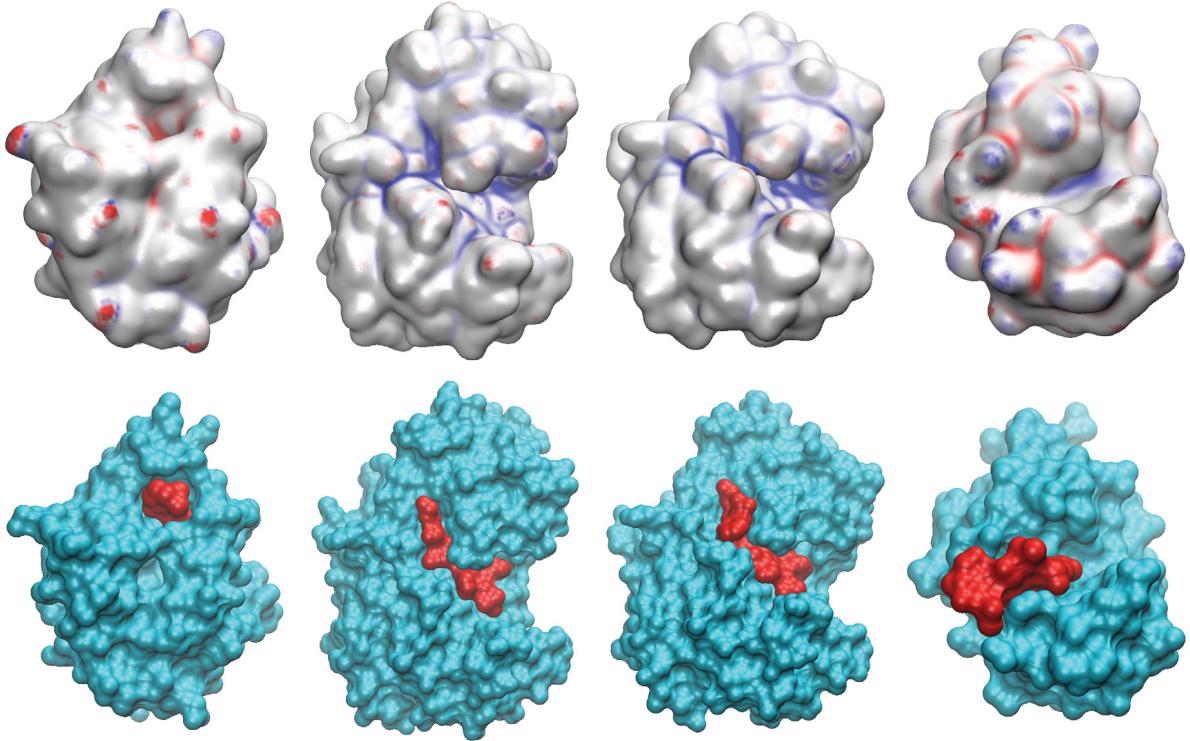


Fig. 12. Protein binding site prediction using polarized curvatures. From left to right, 1CBS, 3ER3, 4ER2, and 1HEW. Top row shows the prediction binding sites. Bottom row depicts the experimental binding sites.

negatively charged. Similarly, in Fig. 12, protein 1CBS is a negatively charged binding site, while proteins 3ER3, 4ER2, and 1HEW have positively charged binding sites. Obviously, there is a good agreement between the current predictions using polarized curvatures and experimental data in these examples.

4. Conclusion

Geometric modeling and electrostatic analysis are central issues in biophysics and quantitative biology. Geometric modeling provides structural representation, visualization and characterization, while electrostatic analysis offers the fundamental understanding of macromolecular hydrophobicity and reactivity. Electrostatic analysis typically depends on geometric modeling. Commonly used geometric models for biomolecules admit various drawbacks, including geometric singularities. There has been a persistent effort in the past decades to develop advanced geometric models for macromolecules. Flexibility and rigidity index (FRI) was proposed for protein flexibility analysis. It was shown to outperform essentially all other existing methods in normal mode analysis and B-factor prediction for over 400 proteins [9,52,51,55]. The present work explores the potential power of the FRI method for geometric modeling and for electrostatic analysis.

We first demonstrate the ability of FRI for surface representation. We show that FRI rigidity functions offer a multiresolution representation of molecular surfaces. These surfaces do not admit geometric singularities. Methods are proposed to generate either smooth surfaces or sharp interfaced surfaces. The corresponding approaches are presented surface area and surface enclosed volume calculations. A Lagrangian representation of FRI surfaces in terms of triangular meshes can be easily generated for finite element simulations via existing triangulation software packages. We also propose an FRI rigidity based Poisson–Boltzmann (RNPB) equation for electrostatic analysis. This method gives rise to smooth dielectric profile, as does our earlier differential geometry based solvation model [1,46].

We use a set of 21 proteins to validate the surface areas and surface enclosed volumes of FRI rigidity surfaces. A comparison is carried out to the solvent excluded surfaces. Good agreement is found for these two types of surface representations. The same set of parameters is also employed to generate electrostatic potentials via the Poisson model. The resulting solvation free energies are comparable to those of the solvent excluded surfaces.

One of the advantages of the FRI rigidity surfaces is that their curvatures can be easily computed both numerically and analytically. Curvature properties indicate concave or convex regions, which are likely to be the potential binding sites or active sites. Based on curvature information, we define polarized curvature as the product of curvature and electrostatic potential. In particular, the polarized curvature constructed by using the minimum curvature and electrostatic potential enables us to characterize potential ligand binding sites. It is demonstrated that the proposed polarized curvatures give rise

to a good prediction of protein–ligand binding sites. Our method offers a potentially powerful approach for high throughput screening of protein–ligand binding and protein–protein interaction sites.

Acknowledgments

Lin Mu's research is based upon work supported in part by the U.S. Department of Energy, Office of Science, Office of Advanced Scientific Computing Research, Applied Mathematics program under award number ERKJE45; and by the Laboratory Directed Research and Development program at the Oak Ridge National Laboratory, which is operated by UT-Battelle, LLC., for the U.S. Department of Energy under Contract DE-AC05-00OR22725. Guowei Wei's research was supported in part by National Science Foundation Grant DMS-1160352 and IIS-1302285, NIH Grant No. R01GM-090208.

References

- [1] G.W. Wei, Differential geometry based multiscale models, *Bull. Math. Biol.* 72 (2010) 1562–1622.
- [2] Xin Feng, Kelin Xia, Yiyi Tong, Guo-Wei Wei, Geometric modeling of subcellular structures, organelles and large multiprotein complexes, *Int. J. Numer. Methods Biomed. Eng.* 28 (2012) 1198–1223.
- [3] X. Feng, K.L. Xia, Y.Y. Tong, G.W. Wei, Multiscale geometric modeling of macromolecules II: lagrangian representation, *J. Comput. Chem.* 34 (2013) 2100–2120.
- [4] K.L. Xia, X. Feng, Y.Y. Tong, G.W. Wei, Multiscale geometric modeling of macromolecules i: Cartesian representation, *J. Comput. Phys.* 275 (2014) 912–936.
- [5] Z.Y. Yu, M. Holst, Y. Cheng, J.A. McCammon, Feature-preserving adaptive mesh generation for molecular shape modeling and simulation, *J. Mol. Graphics Model.* 26 (2008) 1370–1380.
- [6] D. Petrey, B. Honig, GRASP2: Visualization, surface properties, and electrostatics of macromolecular structures and sequences, *Methods Enzymol.* 374 (2003) 492–509.
- [7] W. Rocchia, S. Sridharan, A. Nicholls, E. Alexov, A. Chiabrera, B. Honig, Rapid grid-based construction of the molecular surface and the use of induced surface charge to calculate reaction field energies: Applications to the molecular systems and geometric objects, *J. Comput. Chem.* 23 (2002) 128–137.
- [8] S. Decherchi, W. Rocchia, A general and robust ray-casting-based algorithm for triangulating surfaces at the nanoscale, *PLoS One* 8 (2013) e59744.
- [9] K.L. Xia, K. Opron, G.W. Wei, Multiscale multiphysics and multidomain models – Flexibility and rigidity, *J. Chem. Phys.* 139 (2013) 194109.
- [10] R.B. Corey, L. Pauling, Molecular models of amino acids, peptides and proteins, *Rev. Sci. Instrum.* 24 (1953) 621–627.
- [11] W.L. Koltun, Precision space-filling atomic models, *Biopolymers* 3 (1965) 667–679.
- [12] B. Rubin, Macromolecule backbone models, *Methods Enzymol.* 115 (1985) 391–397.
- [13] B. Lee, F.M. Richards, The interpretation of protein structures: estimation of static accessibility, *J. Mol. Biol.* 55 (3) (1971) 379–400.
- [14] M.L. Connolly, Depth buffer algorithms for molecular modeling, *J. Mol. Graphics* 3 (1985) 19–24.
- [15] F.M. Richards, Areas, volumes, packing, and protein structure, *Annu. Rev. Biophys. Bioeng.* 6 (1) (1977) 151–176.
- [16] P.B. Crowley, A. Golovin, Cation-pi interactions in protein–protein interfaces, *Proteins- Struct. Funct. Bioinform.* 59 (2005) 231–239.
- [17] R.S. Spolar, M.T. Record Jr., Coupling of local folding to site-specific binding of proteins to dna, *Science* 263 (1994) 777–784.
- [18] L.A. Kuhn, M.A. Siani, M.E. Pique, C.L. Fisher, E.D. Getzoff, J.A. Tainer, The interdependence of protein surface topography and bound water molecules revealed by surface accessibility and fractal density measures, *J. Mol. Biol.* 228 (1992) 13–22.
- [19] A.I. Dragan, C.M. Read, E.N. Makeyeva, E.I. Milgotina, M.E.A. Churchill, C. Crane-Robinson, P.L. Privalov, Dna binding and bending by hmg boxes: Energetic determinants of specificity, *J. Mol. Biol.* 343 (2004) 371–393.
- [20] R.M. Jackson, M.J. Sternberg, Dna binding and bending by hmg boxes: Energetic determinants of specificity, *J. Mol. Biol.* 250 (1995) 258–275.
- [21] V.J. LiCata, N.M. Alwell, Functionally linked hydration changes in escherichia coli aspartate transcarbamylase and its catalytic subunit, *Biochemistry* 36 (1997) 10161–10167.
- [22] C.A.S. Bergstrom, M. Strafford, L. Lazorova, A. Avdeef, K. Luthman, P. Artursson, Absorption classification of oral drugs based on molecular surface properties, *J. Med. Chem.* 46 (2003) 558–570.
- [23] N.A. Baker, Improving implicit solvent simulations: a Poisson-centric view, *Curr. Opin. Struct. Biol.* 15 (2) (2005) 137–143.
- [24] Duan Chen, Zhan Chen, Changjun Chen, W.H. Geng, G.W. Wei, MIBPB: A software package for electrostatic analysis, *J. Comput. Chem.* 32 (2011) 657–670.
- [25] W. Geng, G.W. Wei, Multiscale molecular dynamics using the matched interface and boundary method, *J. Comput. Phys.* 230 (2) (2011) 435–457.
- [26] Qiong Zheng, Duan Chen, G.W. Wei, Second-order Poisson-Nernst-Planck solver for ion transport, *J. Comput. Phys.* 230 (2011) 5239–5262.
- [27] Qiong Zheng, G.W. Wei, Poisson–Boltzmann–Nernst–Planck model, *J. Chem. Phys.* 134 (2011) 194101.
- [28] Wenyu Chen, Jianmin Zheng, Yiyu Cai, Kernel modeling for molecular surfaces using a uniform solution, *Comput. Aided Des.* 42 (2010) 267–278.
- [29] M.F. Sanner, A.J. Olson, J.C. Spehner, Reduced surface: An efficient way to compute molecular surfaces, *Biopolymers* 38 (1996) 305–320.
- [30] S.L. Chan, E.O. Purisima, Molecular surface generation using marching tetrahedra, *J. Comput. Chem.* 11 (1998) 1268–1277.
- [31] S. Sridharan, A.R. Nicholls, B. Honig, Sims: Computation of a smooth invariant molecular surface, *Biophys. J.* 73 (1997) 722–732.
- [32] J.A. Grant, B.T. Pickup, A gaussian description of molecular shape, *J. Phys. Chem.* 99 (1995) 3503–3510.
- [33] J. Giard, B. Macq, Molecular surface mesh generation by filtering electron density map, *Int. J. Biomed. Imaging* 2010 (923780) (2010) :9.
- [34] Minxin Chen, Benzhou Lu, Tmsmesh: A robust method for molecular surface mesh generation using a trace technique, *J. Chem. Theory Comput.* 7 (2011) 203–212.
- [35] Q. Zheng, S.Y. Yang, G.W. Wei, Molecular surface generation using PDE transform, *Int. J. Numer. Methods Biomed. Eng.* 28 (2012) 291–316.
- [36] T.D.J. Perkins, J.E.J. Mills, P.M. Dean, Molecular surface-volume and property matching to superpose flexible dissimilar molecules, *J. Comput. Aided Mol. Des.* 9 (1998) 479–490.
- [37] Ho-Lun Cheng, Xinwei Shi, Quality mesh generation for molecular skin surfaces using restricted union of balls, *Comput. Geom.* 42 (2009) 196–206.
- [38] N.L. Max, E.D. Getzoff, Spherical harmonic molecular-surfaces, *IEEE Comput. Graph. Appl.* 8 (1988) 42–50.
- [39] W. Heiden, G. Moegel, J. Brickmann, A new approach to analysis and display of local lipophilicity/hydrophilicity mapped on molecular surfaces, *J. Comput.-Aided Mol. Des.* 7 (1993) 503–514.
- [40] L.D. Gong, Z.Z. Yang, Investigation of the molecular surface area and volume: Defined and calculated by the molecular face theory, *J. Comput. Chem.* 31 (2010) 2098–2108.
- [41] G.W. Wei, Y.H. Sun, Y.C. Zhou, M. Feig, Molecular multiresolution surfaces, 2005, pp. 1–11, arXiv:math-ph/0511001v1.
- [42] P.W. Bates, G.W. Wei, S. Zhao, The minimal molecular surface, 2006, arXiv:q-bio/0610038v1 [q-bio.BM].
- [43] P.W. Bates, G.W. Wei, Shan Zhao, Minimal molecular surfaces and their applications, *J. Comput. Chem.* 29 (3) (2008) 380–391.
- [44] Guo-Wei Wei, Qiong Zheng, Zhan Chen, Kelin Xia, Variational multiscale models for charge transport, *SIAM Rev.* 54 (4) (2012) 699–754.
- [45] Guo-Wei Wei, Multiscale, multiphysics and multidomain models I: Basic theory, *J. Theor. Comput. Chem.* 12 (8) (2013) 1341006.
- [46] Z. Chen, N.A. Baker, G.W. Wei, Differential geometry based solvation models I: Eulerian formulation, *J. Comput. Phys.* 229 (2010) 8231–8258.
- [47] Z. Chen, N.A. Baker, G.W. Wei, Differential geometry based solvation models II: Lagrangian formulation, *J. Math. Biol.* 63 (2011) 1139–1200.
- [48] Z. Chen, G.W. Wei, Differential geometry based solvation models III: Quantum formulation, *J. Chem. Phys.* 135 (2011) 194108.

- [49] Jan J. Koenderink, Andrea J. van Doorn, Surface shape and curvature scales, *Image Vis. Comput.* 10 (8) (1992) 557–564.
- [50] G. Cipriano, G.N. Phillips Jr., M. Gleicher, Multi-scale surface descriptors, *IEEE Trans. Vis. Comput. Graphics* 15 (2009) 1201–1208.
- [51] Kristopher Pron, K.L. Xia, G.W. Wei, Communication: Capturing protein multiscale thermal fluctuations, *J. Chem. Phys.* 142 (211101) (2015).
- [52] K. Pron, K.L. Xia, G.W. Wei, Fast and anisotropic flexibility-rigidity index for protein flexibility and fluctuation analysis, *J. Chem. Phys.* 140 (2014) 234105.
- [53] I. Bahar, A.R. Atilgan, B. Erman, Direct evaluation of thermal fluctuations in proteins using a single-parameter harmonic potential, *Fold. Des.* 2 (1997) 173–181.
- [54] I. Bahar, A.R. Atilgan, M.C. Demirel, B. Erman, Vibrational dynamics of proteins: Significance of slow and fast modes in relation to function and stability, *Phys. Rev. Lett.* 80 (1998) 2733–2736.
- [55] K.L. Xia, K. Pron, G.W. Wei, Multiscale gaussian network model (ngnm) and multiscale anisotropic network model (manm), *J. Chem. Phys.* (2015).
- [56] K.L. Xia, Z.X. Zhao, G.W. Wei, Multiresolution topological simplification, *J. Comput. Biol.* 22 (2015) 1–5.
- [57] K.L. Xia, Z.X. Zhao, G.W. Wei, Multiresolution persistent homology for excessively large biomolecular datasets, *J. Chem. Phys.* 143 (2015) 134103.
- [58] G.W. Wei, Wavelets generated by using discrete singular convolution kernels, *J. Phys. A: Math. Gen.* 33 (2000) 8577–8596.
- [59] M.P. Allen, D.J. Tildesley, *Computer Simulation of Liquids*, Clarendon Press, Oxford, 1987.
- [60] Michael J. Holst, The Poisson–Boltzmann equation: Analysis and multilevel numerical solution (Ph.D. thesis), Numerical Computing Group, University of Illinois at Urbana-Champaign, 1994.
- [61] Weihua Geng, Sining Yu, G.W. Wei, Treatment of charge singularities in implicit solvent models, *J. Chem. Phys.* 127 (2007) 114106.
- [62] Sining Yu, G.W. Wei, Three-dimensional matched interface and boundary (mib) method for treating geometric singularities, *J. Comput. Phys.* 227 (1) (2007) 602–632.
- [63] S.N. Yu, W.H. Geng, G.W. Wei, Treatment of geometric singularities in implicit solvent models, *J. Chem. Phys.* 126 (2007) 244108.
- [64] Y.C. Zhou, Shan Zhao, Michael Feig, G.W. Wei, High order matched interface and boundary method for elliptic equations with discontinuous coefficients and singular sources, *J. Comput. Phys.* 213 (1) (2006) 1–30.