

## Research Article

# A Multiscale Model for Virus Capsid Dynamics

Changjun Chen,<sup>1</sup> Rishu Saxena,<sup>1</sup> and Guo-Wei Wei<sup>1,2</sup>

<sup>1</sup>Department of Mathematics, Michigan State University, East Lansing, MI 48824, USA

<sup>2</sup>Department of Electrical and Computer Engineering, Michigan State University, East Lansing, MI 48824, USA

Correspondence should be addressed to Guo-Wei Wei, wei@math.msu.edu

Received 25 September 2009; Accepted 25 November 2009

Academic Editor: Shan Zhao

Copyright © 2010 Changjun Chen et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Viruses are infectious agents that can cause epidemics and pandemics. The understanding of virus formation, evolution, stability, and interaction with host cells is of great importance to the scientific community and public health. Typically, a virus complex in association with its aquatic environment poses a fabulous challenge to theoretical description and prediction. In this work, we propose a differential geometry-based multiscale paradigm to model complex biomolecule systems. In our approach, the differential geometry theory of surfaces and geometric measure theory are employed as a natural means to couple the macroscopic continuum domain of the fluid mechanical description of the aquatic environment from the microscopic discrete domain of the atomistic description of the biomolecule. A multiscale action functional is constructed as a unified framework to derive the governing equations for the dynamics of different scales. We show that the classical Navier-Stokes equation for the fluid dynamics and Newton's equation for the molecular dynamics can be derived from the least action principle. These equations are coupled through the continuum-discrete interface whose dynamics is governed by potential driven geometric flows.

## 1. Introduction

Viruses are omnipresent infectious agents that are about 100 times smaller than bacteria. Unlike bacteria, viruses are not able to grow or reproduce outside a host cell [1–4]. There are more than 5000 types of known viruses. Viruses have a known history of causing epidemics and pandemics. About 70% of native Americans were killed by foreign diseases after the arrival of Columbus in the Americas. The Spanish flu pandemic lasted from 1918 to 1919 and killed about 100 million people, or 5% of the world's population in 1918. AIDS, a disease due to HIV virus, has killed more than 25 million people since it was first recognized on June 5, 1981. There are about 39 million people living with HIV viruses worldwide nowadays. Virus infection processes or virus life cycles differ greatly among species but there are six basic stages: [1–4] (1) selective attachment due to the interaction, binding and/or fusion between viral capsid surface and specific receptors on the host cellular surface, (2) penetration of a virus into a host cell through membrane fusion or receptor-mediated endocytosis, (3) viral genomic nucleic acid releasing in the

host cell due to viral capsid degradation by viral enzymes or host enzymes, (4) virus replication and assembly in the host cell, (5) Posttranslational modification of the viral proteins; and finally, (6) virus releasing from the host cell. For some viruses, such as HIV, the order of stages (5) and (6) is reversed. Body uses two defense mechanisms, innate immune system and cell-mediated immunity to defend host from infection by viruses or other organisms. The innate immune system terminates the virus replication in the host cell by degrading or inhibiting the virus genetic material, DNA or RNA through antibodies or other virus DNA/RNA binding molecules. In the cell-mediated immunity, killer cells known as T cells destroy the infected host cell and its close neighbors by recognizing the viral protein displayed on the cellular surface.

Recent advances in structural biology and microbiology have led to a rapidly growing body of virus structural data [5–7]. A striking feature of virus data is that they are excessively large—a virus complex may involve tens of millions atoms, with detailed information on atomic coordinates, types, and radii. Most virus structural data are collected via X-ray diffraction (X-ray), cryo-electron microscopy (cryoEM) [8],

fiber diffraction, and nuclear magnetic resonance (NMR) techniques. There are a few major virus morphologies: spherical type, helical type, dihedral type, viral envelope type, and complex type. Most animal viruses are of spherical morphology with icosahedral symmetry [6]. Most virus structure information can be obtained from the Protein Data Bank (PDB; <http://www.rcsb.org/pdb/home/home.do>), the Virus Particle Explorer database [7] (VIPERdb; <http://viperdbscripps.edu/>), and the Protein Quaternary Structure server (PQS; <http://pqs.ebi.ac.uk/>).

Currently, the prevention and control of epidemics and pandemics caused by infective viruses, such as H1N1, HIV, SARS, and bird flu are of paramount importance. As an infection starts with the surface attachment between a virus and a host cell, it is important to construct and visualize the surface topology and morphology of viruses in order to understand the surface attachment and further interaction. This information is also crucial to the understanding of the molecular mechanism that gives rise to the assembly of virus capsids and DNA or RNA packaging. Computer-based visualization is able to represent results of explorations in an easy-to-comprehend form and to facilitate convenient information retrieval. Currently, visualization tools are often developed in close conjunction with imaging, data registration, simulation and/or surface construction. Virus visualization plays a unique role in the understanding of virus infection processes, such as, virus attachment of a host cell, binding and fusion between a virus capsid surface and a host cellular surface, and the penetration of a virus into a host cell. However, viruses are not directly visible because their sizes are at the order of tens of nanometers. The virus images are constructed from virus information, which is either collected from modalities described above or generated by computer simulations. Therefore, surface/image construction is a part of the virus visualization. Yu and Bajaj present a computational algorithm to segment asymmetric units of three-dimensional (3D) density maps of icosahedral viruses [9] and a computational approach to structural interpretation from reconstructed 3D electron microscopy (3D-EM) maps of viruses [10]. Some basic biomolecular surface methods are available in visualization software packages Chimera (<http://www.cgl.ucsf.edu/chimera/>) and VMD (<http://www.ks.uiuc.edu/Research/vmd/>).

The difficulty of characterizing a virus complex is not only its massive number of atoms, or data sets, but also its everlasting interactions. Except for envelope type of viruses which typically cover their capsids by envelopes derived from lipids and proteins of their host cell membranes, most viruses use their own capsids to interact with the environment and host cells. A viral capsid usually consists of many identical viral protein subunits that form the capsid by symmetric assembly. There are strong interactions between viral protein subunits so that viral capsids are rigid enough to hold viral genome material and protect its content. Viruses have adapted a number of strategies to maintain the stability and flexibility of viral capsids. For many small viruses, such as one of STMV, their subunit proteins generally only touch each other by their edges. Their capsid stability is achieved by strong nonbonding interactions (i.e., hydrogen

bonding and van der Waals interactions) between edges of subunit proteins. Some large viruses, such as BMV, have developed overlapping strategies to increase the capsid stability. Some viruses even use a few intricately intertwining layers to strengthen their capsids [11]. Virus capsids are further stabilized by their hydrophobic interaction with the aquatic environment. Clearly the boundary profile of the virus complex is determined by the balance of all mechanical forces or equivalently, the energy minimization of the system.

One of the present authors, Wei, introduced some of the first high-order geometric flow equations for image analysis [12]. These equations have led to many applications [12–16]. Mathematical analysis of Wei's equations has been recently carried out in Sobolev space  $H^1$  by Bertozzi and Greer [17–19], who proved the existence and uniqueness of the solution to a case with  $H^1$  initial data and a regularized operator. Coupled geometric flow equations were introduced by Wei and Jia for image edge detection [13]. An evolution operator based single-step method was proposed by Wei, Wang and their coworkers for image processing [14]. A partial differential equation approach of Connolly surfaces was proposed by Wei and his coworkers [20]. In such an approach, geometric partial differential equation (PDE) is used to describe the solvent density flows. Most biological processes occur in water, which consists of about 70% body mass. Therefore, in general, the biomolecular surface morphology should be determined by the free energy minimization in the aquatic environment. Wei and his coworkers have addressed this question by considering a mean curvature flow model of biomolecular surfaces that minimize the surface-free energy functional [21]. They have also recently introduced stochastic geometric flows to account for the random fluctuation and dissipation in density and pressure near the surface [22]. A general geometric flow structure, the potential driven geometric flows, was introduced [22]. Physical properties, such as free energy minimization (area decreasing) and incompressibility (volume preserving), were realized in new geometric evolution equations [22]. Computational techniques used in this surface analysis are quite similar to the level sets devised by Osher and Sethian [23–25]. Another efficient approach is the Euler-Lagrange formulation of surface variation developed by Chan and others, [26, 27]. Interacting particle systems and point-based approaches have also been proposed for the modeling and animation of surfaces [28].

An unsolved problem in structural virology is the detailed molecular mechanism of the assembly of virus capsids with the right size that is able to accommodate virus genetic material in the subsequent virus DNA/RNA packaging. Additionally, the process of virus attachment on its host cell, the movement of virus fusion with cellular membrane, and the dynamics of virus penetration into its host cell remain unrevealed mysteries. Prerequisites to unveiling these mysteries are efficient computer science and mathematical tools for modeling virus surface construction, evolution, and visualization, and for analyzing the virus interactions with its host cell. A typical virus has millions of atoms, while a large virus may have tens of millions atoms. Huge viral data sets pose severe challenges

to the theoretical understanding and prediction of virus dynamics and interactions. These challenges are considerably exacerbated by the fact that virus behavior and infectivity depend strongly on the physiological environment, where the water molecules are the most common media. This dramatically increases the number of degrees of freedom of a virus system. The real-time dynamic visualization of viral attachment, fusion, and penetration of a host cell in the aquatic environment requires microsecond or even millisecond simulation time and is technically intractable with full-atom models at present [11, 29]. In fact, the elementary operations, that is, the construction of virus surfaces with physical models and real-time visualization of virus morphology present formidable challenges for applied mathematics and computer science.

Recently, one of the present authors, Wei, has developed a differential geometry-based multiscale paradigm to address some of the aforementioned challenges in the nonequilibrium dynamics of viruses, as well as other complex chemical systems, for example, fuel/solar cells, and biological systems, for example, ion channels [30]. In this approach, the differential geometry theory of surfaces and geometric measure theory are employed to couple the macroscopic continuum mechanical description of the aquatic environment with the microscopic discrete atomistic description of the macromolecule. Multiscale action functionals are constructed as a unified framework to derive the governing equations for the dynamics of different scales and different descriptions. The generalized Navier-Stokes equation for the fluid dynamics, the generalized Poisson Boltzmann equation for electrostatic interactions, and Newton's equation for the molecular dynamics were derived by the least action principle. These equations are coupled through the micro-macro boundary whose dynamics is governed by potential driven geometric flows.

The objective of the present work is threefold. First, we apply the differential geometry-based multiscale models to the formation and evolution virus capsids where challenges originated from a large number of atoms and a variety of interactions in a virus system, including the aquatic environment. To dramatically reduce the number of degrees of freedom of a virus system, we treat the water molecules as a macroscopic continuum. However, we maintain atomic description of the virus to allow an optimal access to detailed biomolecular information. Secondly, we propose a new scale, the coarse-grained particles, to improve the earlier multiscale formalism [30]. Our new coarse-grained scale is based on the description of amino acid residues. This additional scale is necessary for excessively large viruses or macromolecules. It efficiently reduces the number of degrees of freedom. Finally, to further reduce computational cost, we utilize virus symmetries to provide an optional reduction in data size. Viruses typically have a few coding genes and they make use of symmetries to reduce their genome size, because capsid genes are repeatedly used. Apparently, viruses also try to make use of symmetry so that they have a high ratio of volume over surface area. As such, virus can maintain the desirable mechanical and chemical stability while without their own cell membranes and complex defense systems.

Some of the proposed ideas are tested by their applications in virus surface formation, evolution, and visualization.

## 2. Theory and Algorithms

In this section, differential geometry theory of surfaces and potential driven geometric flows are utilized to establish a multiscale paradigm for modeling and simulation of virus formation and evolution. Then, a coarse-grained virus model is formulated to further reduce the number of degrees of freedom. Finally, the use of symmetry in virus surface construction is discussed.

### 2.1. Differential Geometry-Based Multiscale Model

*2.1.1. Multiscale Models of Virus Surface Formation and Evolution.* A fundamental issue in biological modeling, and in data analysis, visualization, and dynamical representation is how to deal with a tremendously large number of degrees of freedom resulting from various interaction. Under physiological condition, a virus and its interacting environment may involve tens of millions of protein atoms and water molecules. In principle, the system can be described entirely in the microscopic scale, that is, atomistic description or more detailed description of electrons and nuclei. However, such an approach cannot be productive and does not provide theoretical predictions of physical properties of the virus complex. It is impossible at present, and formidably expensive in near future to describe in full-atomic detail of all the aforementioned interactions for a large virus system. On the other hand, a macroscopic description of the system is incapable of revealing the molecular and atomic information of the virus particle and its dynamics. We plan to reduce the number of degrees of freedom of the virus complex by a differential geometry-based multiscale model. In our multiscale model, we will describe the aquatic environment by a hydrodynamic continuum, that is, a macroscopic description. As such, we are able to dramatically reduce the number of degrees of freedom of millions surrounding water molecules. However, since the biomolecule or the virus is the objective of interest, we will describe the virus in atomic detail, that is, a microscopic, discrete description. Additionally, we carefully consider the solvation process of the virus molecule. The virus surface tension and mechanical work of virus immersion into the solvent are considered in our model, in addition to the possible interaction between virus atoms and the aquatic environment. Finally, the force resulted from virus and solvent interactions is accounted by fluid motion, which is modeled by a viscous fluid.

In our differential geometry-based multiscale model, we use a hypersurface (characteristic) function  $S$  to characterize the boundary of the virus and solvent. As such,  $S = 1$  indicates the virus domain and  $S = 0$  (i.e.,  $1 - S = 1$ ) indicates the aquatic domain. However, at atomic scale, the virus surface, or the flow boundary between the virus particle and aquatic environment cannot behave like the Heaviside function. Instead, it must take a value between zero and one ( $0 \leq S \leq 1$ ). Such a profile characterizes the boundary

between the virus and the aquatic environment. In the rest of this section, we set  $\mathbf{x} \in \mathbb{R}^3$  as the macroscopic variable and  $\mathbf{z} = (\mathbf{z}_1, \mathbf{z}_2, \dots, \mathbf{z}_N) \in \mathbb{R}^{3N}$  as the microscopic variable of  $N$  discrete atoms or particles. The domain of the solvent is denoted as  $\Omega_m : \{\mathbf{x} \mid S(\mathbf{x}) \neq 0\}$  and that of the virus molecule is denoted as  $\Omega_s : \{\mathbf{x} \mid (1 - S(\mathbf{x})) \neq 0\}$ . The whole computational domain is  $\Omega = \Omega_s \cup \Omega_m$ . The solvent-solute boundary is  $\Omega_b = \Omega_s \cap \Omega_m$ .

We consider the total action functional for the virus complex [30]

$$\begin{aligned} \mathcal{S}_{\text{total}}[S, \phi, \mathbf{x}, \mathbf{z}] &= \iiint \left\{ [\gamma \|\nabla S\| + Sp + (1 - S)\rho_s u] + S \left[ \rho_m \phi - \frac{\epsilon_m}{2} |\nabla \phi|^2 \right] \right. \\ &+ (1 - S) \left[ -\frac{\epsilon_s}{2} |\nabla \phi|^2 - k_B T \sum_{j=1}^{N_c} c_j \left( e^{-q_j \phi / k_B T} - 1 \right) \right] \\ &- (1 - S) \left[ \rho_s \frac{\mathbf{v}^2}{2} - p + \frac{\mu_f}{8} \int^t [\nabla \mathbf{v} + (\nabla \mathbf{v})^T]^2 dt' \right] \\ &\left. - S \sum_{j=1}^N \left[ \rho_j \frac{\dot{\mathbf{z}}_j^2}{2} - U(\mathbf{z}) \right] \right\} d\mathbf{x} d\mathbf{z} dt, \end{aligned} \quad (1)$$

where  $\phi \in \Omega$  is the electrostatic potential,  $\gamma$  is the surface tension,  $p$  is the pressure,  $u$  is the interaction potential between the solvent and the solute,  $k_B$  is the Boltzmann constant,  $T$  is the temperature,  $c_j$  is the bulk concentration of  $j$ th ionic species,  $N_c$  is the number of ionic species, and  $\rho_m(\mathbf{x}, \mathbf{z}) = \sum_j Q_j \delta(\mathbf{x} - \mathbf{z}_j)$  is the canonical density of molecular free charges, with  $Q_j$  being partial charges on (discrete) atoms. Here,  $\epsilon_m = \epsilon_0 \epsilon_m$  and  $\epsilon_s = \epsilon_0 \epsilon_s$  are the permittivities of the macromolecule and the solvent, respectively, where  $\epsilon_0$  is the permittivity of vacuum, and  $\epsilon_\alpha$  and  $(\alpha = m, s)$  are relative permittivities. We treat  $\epsilon_\alpha$  as constants. Additionally,  $\rho_s$  and  $\rho_j$  are mass densities of the solvent and virus atom (or coarse-grained particle), respectively. Finally,  $\mathbf{v} \in \Omega_s$  is the fluid velocity,  $\mu_f$  is the viscosity of the fluid, symbol  $T$  in superscript denotes the transpose,  $\dot{\mathbf{z}}_j = d\mathbf{z}_j/dt \in \Omega_m$  is the velocity of the  $j$ th atom, and  $U$  is the interaction potential for atoms.

On the right hand side of (1), the first row is the nonpolar solvation free energy, which includes the surface area effect ( $\gamma \|\nabla S\|$ ), the mechanical work (the volume effect  $Sp$ ), and the solvent-solute interactions ( $(1 - S)\rho_s u$ ). In principle, these interactions take care of important dispersion effects, and other van der Waals effects. Geometric measure theory is used to come up with the expression for the surface area. The second row is the electrostatic polar solvation free energy, which has contributions from the virus particle  $S[\rho_m \phi - \epsilon_m/2 |\nabla \phi|^2]$  and the aquatic solvent  $(1 - S)[-(\epsilon_s/2) |\nabla \phi|^2 - k_B T \sum_{j=1}^{N_c} c_j (e^{-q_j \phi / k_B T} - 1)]$ . Here, the virus particles contribute a set of discrete partial charges while the ion charges in the solvent are treated as a continuous Boltzmann distribution. This is valid as long as the system is near equilibrium. For systems far from

equilibrium, alternative models, such as Poisson-Nernst-Planck (PNP) equations, are required to describe the density of ionic species [30]. The third row is the Lagrangian of the fluid dynamics subsystem with a negative sign. It consists of the kinetic energy ( $\rho_s(\mathbf{v}^2/2)$ ) of the fluid flow and the generalized potential energy. The latter includes pressure ( $p$ ) and stress energy ( $(\mu_f/8) \int^t [\nabla \mathbf{v} + (\nabla \mathbf{v})^T]^2 dt'$ ). The stress energy represents the energy loss due to the interactions among the fluid particles, which are not explicitly described in the present model. The exact expression of the stress tensor for real fluid is usually unknown. Newtonian fluid and NonNewtonian fluid approximations are commonly used, in addition to numerous other approximations. Finally the last row contains the Lagrangian of the virus molecular dynamics subsystem with a negative sign. It describes the kinetic energy  $\rho_j(\dot{\mathbf{z}}_j^2/2)$  and potential energy  $U(\mathbf{z})$ . The latter includes all possible potential interactions among virus atoms or coarse-grained particles. We have chosen negative signs for two Lagrangians so that the potential energies have positive signs and are consistent with other potential energies.

*2.1.2. Governing Equations for Coupled Fluid Dynamics and Molecular Dynamics.* In the present work, we derive four governing equations by employing the principle of the least action to the total action functional ( $\mathcal{S}_{\text{total}}[S, \phi, \mathbf{x}, \mathbf{z}]$ ) in (1) with respect to four variables ( $S, \phi, \mathbf{x}, \mathbf{z}$ )

$$\begin{aligned} \delta \mathcal{S}_{\text{total}}[S, \phi, \mathbf{x}, \mathbf{z}] &= \iiint \left\{ \left[ \nabla \cdot \epsilon(S)\phi + S\rho_m + (1 - S) \sum_j^{N_c} q_j c_j e^{-q_j \phi / k_B T} \right] \delta \phi \right. \\ &+ \left[ -\nabla \cdot \frac{\gamma \nabla S}{\|\nabla S\|} + p - \rho_s u \right] + \left[ \rho_m \phi - \frac{\epsilon_m}{2} |\nabla \phi|^2 \right] \\ &- \left[ -\frac{\epsilon_s}{2} |\nabla \phi|^2 - k_B T \sum_j^{N_c} c_j \left( e^{-q_j \phi / k_B T} - 1 \right) \right] \\ &- \sum_j \left[ \rho_j \frac{\dot{\mathbf{z}}_j^2}{2} - U(\mathbf{z}) \right] \\ &\left. \left[ \rho_s \frac{\mathbf{v}^2}{2} - p + \frac{\mu_f}{8} \int^t [\nabla \mathbf{v} + (\nabla \mathbf{v})^T]^2 dt' \right] \right\} \delta S \\ &+ \left[ S \nabla p + (1 - S) \nabla \rho_s u + S \phi \nabla \rho_m + (1 - S) \rho_s \right. \\ &\quad \left. \left( \frac{\partial \mathbf{v}}{\partial t} + \mathbf{v} \cdot \nabla \mathbf{v} \right) + (1 - S) \nabla p - \nabla \cdot (1 - S) \mathbb{T} \right] \cdot \delta \mathbf{x} \\ &+ \sum_j \left[ (1 - S) \nabla_j \rho_s u + S \phi \nabla_j \rho_m + S \rho_j \dot{\mathbf{z}}_j + S \nabla_j U(\mathbf{z}) \right] \\ &\quad \left. \cdot \delta \mathbf{z}_j \right\} d\mathbf{x} d\mathbf{z} dt = 0. \end{aligned} \quad (2)$$

Here,  $\delta S$ ,  $\delta\phi$ ,  $\delta\mathbf{x}$ , and  $\delta\mathbf{z}$  are four infinitesimally small but nonzero perturbations. In order for the first variation to vanish, the terms associated  $\delta S$ ,  $\delta\phi$ ,  $\delta\mathbf{x}$ , and  $\delta\mathbf{z}$  have to vanish independently. First, the term associated with  $\delta\phi$  gives rise to a generalized Poisson-Boltzmann equation

$$-\nabla \cdot \epsilon(S)\nabla\phi = S\rho_m + (1-S)\sum_j^{N_c} q_j c_j e^{-q_j\phi/k_B T}, \quad (3)$$

where

$$\epsilon(S) = S\epsilon_m + (1-S)\epsilon_s \quad (4)$$

provides a smooth dielectric profile near the interface. This is a new Poisson-Boltzmann equation for overlapping domains. With the sharp interface limit, (3) reduces to the standard Poisson-Boltzmann equation [31–35]

$$\begin{aligned} -\epsilon_m \nabla^2 \phi_m &= \rho_m, \quad \forall \mathbf{x} \in \Omega_m, \\ -\epsilon_s \nabla^2 \phi_s &= \sum_j^{N_c} q_j c_j e^{-q_j\phi_s/k_B T}, \quad \forall \mathbf{x} \in \Omega_s, \end{aligned} \quad (5)$$

and appropriate interface conditions

$$\phi_s = \phi_m, \quad \epsilon_m \nabla\phi_m \cdot \mathbf{n} = \epsilon_s \nabla\phi_s \cdot \mathbf{n}, \quad \forall \mathbf{x} \text{ on } \Gamma, \quad (6)$$

where  $\Omega_m$  and  $\Omega_s$  are, respectively, the virus domain and the solvent domain,  $\Gamma$  is the sharp interface and  $\mathbf{n}$  is the normal vector of the surface.

Additionally, the virus surface evolution equation can be constructed by requiring the term associated with  $\delta S$  in (1) to vanish, followed by the use of the steepest descent scheme

$$\begin{aligned} \frac{\partial S}{\partial t} &= \|\nabla S\| \left\{ \nabla \cdot \frac{\gamma \nabla S}{\|\nabla S\|} - p + \rho_s u - \rho_m \phi + \frac{\epsilon_m}{2} |\nabla\phi|^2 \right. \\ &\quad - \frac{\epsilon_s}{2} |\nabla\phi|^2 - k_B T \sum_j^{N_c} c_j \left( e^{-q_j\phi/k_B T} - 1 \right) \\ &\quad - \left[ \rho_s \frac{\mathbf{v}^2}{2} - p + \frac{\mu_f}{8} \int^t \left[ \nabla\mathbf{v} + (\nabla\mathbf{v})^T \right]^2 dt' \right] \\ &\quad \left. + \sum_j \left[ \rho_j \frac{\mathbf{z}_j^2}{2} - U(\mathbf{z}) \right] \right\}. \end{aligned} \quad (7)$$

The structure of this equation is very similar to the potential driven geometric flows introduced in the earlier work [22, 30]

$$\frac{\partial S}{\partial t} = \|\nabla S\| \left[ \nabla \cdot \frac{\gamma \nabla S}{\|\nabla S\|} + V \right], \quad (8)$$

where  $V$  includes appropriate potential interaction terms. Therefore, (7) can be solved by using the same procedure as that described in the earlier work [22].

Moreover, the requirement of the vanishing of the term associated with  $\delta\mathbf{x}$  gives rise to a generalized Navier-Stokes equation for continuum fluid dynamics [30]

$$\rho_s \left( \frac{\partial \mathbf{v}}{\partial t} + \mathbf{v} \cdot \nabla \mathbf{v} \right) = -\nabla p + \frac{1}{1-S} \nabla \cdot (1-S)\mathbb{T} + \mathbf{F}, \quad (9)$$

where the stress tensor is given by

$$\mathbb{T} = \frac{\mu_f}{2} \left[ \nabla\mathbf{v} + (\nabla\mathbf{v})^T \right]. \quad (10)$$

The Newtonian fluid is assumed in the present work. The force in (9) is given by

$$\mathbf{F} = \frac{S}{1-S} \mathbf{f}. \quad (11)$$

Here, the force includes a few components

$$\mathbf{f} = \mathbf{f}_P + \mathbf{f}_{\text{SSI}} + \mathbf{f}_{\text{RF}}, \quad (12)$$

defined as

$$\begin{aligned} \mathbf{f}_P &= -\nabla p; \\ \mathbf{f}_{\text{SSI}} &= -\frac{(1-S)}{S} \nabla(\rho_s u), \\ \mathbf{f}_{\text{RF}} &= \frac{\rho_m}{S} \nabla(S\phi). \end{aligned} \quad (13)$$

The detailed derivation of the generalized Navier-Stokes equation can be found in [30]. In case of sharp solvent-virus interfaces, the hypersurface function  $S$  becomes a step function, and (9) reduces to the standard Navier-Stokes equation

$$\rho_s \left( \frac{\partial \mathbf{v}}{\partial t} + \mathbf{v} \cdot \nabla \mathbf{v} \right) = -\nabla p + \mu_f \nabla^2 \mathbf{v} + \mathbf{F}, \quad (14)$$

with simplified force expressions.

Finally, the Newton's equation for molecular dynamics of the  $j$ th atom (particle) in the virus is derived from the term associated with  $\delta\mathbf{z}_j$ ,

$$\rho_j \ddot{\mathbf{z}}_j = \mathbf{f}^j. \quad (15)$$

Here the microscopic force associated with the  $j$ th atom is

$$\mathbf{f}^j = \mathbf{f}_{\text{SSI}}^j + \mathbf{f}_{\text{RF}}^j + \mathbf{f}_{\text{PI}}^j. \quad (16)$$

The force components are defined as

$$\begin{aligned} \mathbf{f}_{\text{SSI}}^j &= -\frac{(1-S)}{S} \nabla_j(\rho_s u) \\ \mathbf{f}_{\text{RF}}^j &= \frac{\rho_m}{S} \nabla_j(S\phi) \\ \mathbf{f}_{\text{PI}}^j &= -\nabla_j U(\mathbf{z}), \end{aligned} \quad (17)$$

where  $\mathbf{f}_{\text{SSI}}^j$ ,  $\mathbf{f}_{\text{RF}}^j$ , and  $\mathbf{f}_{\text{PI}}^j$  are, respectively, solvent-solute interaction force, reaction field force, and potential interaction force.

In this multiscale system, all forces are balanced. The fluid dynamics, the molecular dynamics, the electrostatic subsystem, and the hypersurface function are all coupled.

TABLE 1: Coarse-grain radii ( $\text{\AA}$ ) for twenty standard amino acid residues.

Residue	Radius								
GLY	4.20	ALA	4.10	VAL	4.30	LEU	5.70	ILE	5.60
PRO	4.10	PHE	7.00	TYR	8.30	TRP	8.30	SER	4.30
THR	4.40	ASN	5.50	GLN	6.80	CYS	4.50	MET	7.30
ASP	5.50	GLU	6.70	HIS	6.40	LYS	8.20	ARG	9.10

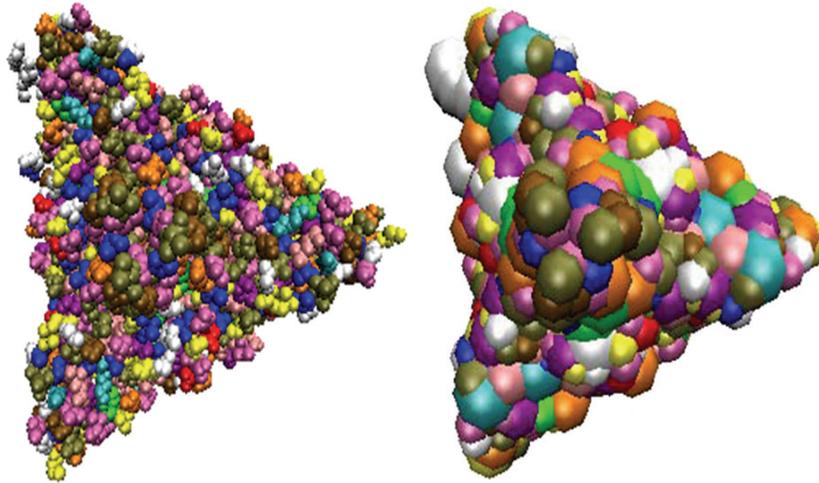


FIGURE 1: Coarse-grained model of a viral protein subunit. Left: the full atomic model of a protein subunit of the Nodamura virus (PDB ID: 1Nov), Right: the coarse-grained model of a protein subunit of the Nodamura virus.

**2.2. Coarse-Grained Model.** As a part of our multiscale framework, we consider a coarse-grained formalism for viral surface formation and evolution. Coarse-grained models are often used to deal with exceptionally large biological systems. In the present treatment, we consider each amino acid residue as a particle, located at the  $C_\alpha$  position. The radii of twenty standard amino acid residues used in the present work are listed in Table 1. Coarse-grained representations are efficient approaches for data size reduction. Combined with enhanced computer power and efficient computational algorithms, coarse-grained approaches currently enable the simulation of systems of biologically relevant size (sub-micrometric) and timescale (microsecond or millisecond) [29]. Although coarse-grained models cannot be considered as predictive as all-atom ones, they can provide much insight with the use of more rigorous parameterization techniques and efficient algorithms for sampling configurational space. Since the simulation size and timescale of coarse-grained models coincide with those that can be reached with the most advanced spectroscopic techniques, it is possible to directly compare experiment data and simulation predictions. In this work, we will explore the use of coarse-grained models for viral surface formation and evolution. Figure 1 presents an illustration of coarse-graining particles for a viral protein subunit. The original full-atomic subunit of the Nodamura virus has about 10 thousand atoms. In the coarse grained representation, each

amino acid residue is considered as one particle, located at the position of the original  $C_\alpha$  atom. Each type of amino acid residues has a particle radius as shown in Table 1. The discrete-continuum model of viral surface representation discussed above is still applicable to the present coarse-grain-continuum setting. However, to use (7) for viral surface formation and evolution, we need to redefine the Lennard-Jones and Coulomb potential parameters to describe the interaction between amino-acid-residue particles.

### 2.3. Viral Data Size Reduction by Symmetry

**2.3.1. Symmetry in Virus Capsids.** Viral data may involve tens of millions of atomic coordinates and radii, and are enormously large for structural modeling, simulation and visualization. Viral dynamical cycles may last from millisecond to days, and real-time full-atom viral dynamical simulations of viruses are intractable to the present computational capability [11]. However, viruses typically have very small genomes and code a few proteins. In order for a viral capsid to hold all viral genetic material, virus makes use of symmetry in its capsid assembly. Amazingly, most viruses are symmetric, having icosahedral, helical, dihedral, or circular symmetries [6]. As such, an icosahedral virus can self-organize one protein to generate a capsid of 60 symmetry-related subunits (some viruses code hundreds of

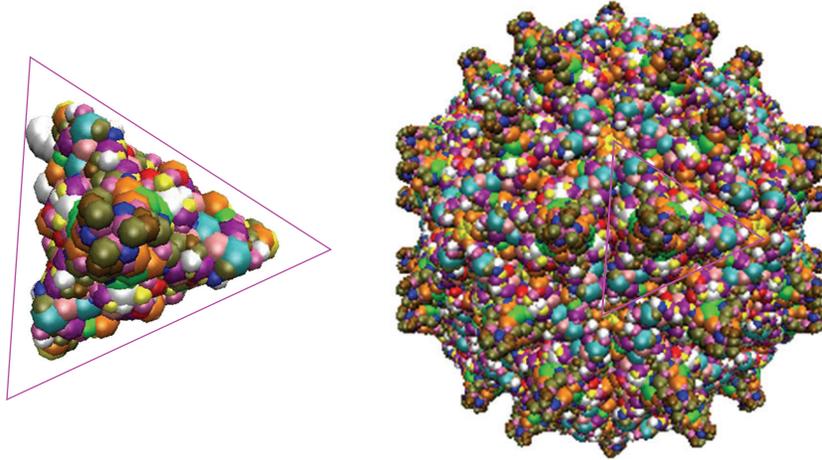


FIGURE 2: Illustration of surface construction from a facet patch by using symmetry. Left: the generating subunit (facet patch) of the Nodamura virus (PDB ID: 1nov), Right: the full surface of the Nodamura virus constructed by symmetric assembly.

proteins). Therefore, it is desirable to take advantage of symmetry in viral data analysis, operation, and management. In particular, we propose to make use of viral symmetries, if they are available in our geometric flow based viral surface formation and evolution. Additionally, we can detect partial and approximate symmetry [36] from viral surfaces, and enforce symmetrization [37]. As such, we will use geometric flows to generate symmetric facets, or patches from viral protein subunits, and construct the whole viral surface by symmetric assembly of viral facets; see Figure 2 for an illustration.

**2.3.2. Virus Symmetry Transformation.** Viruses have adapted five point group symmetries, that is, circular, dihedral, tetrahedral, octahedral, and icosahedral, in their biological assemblies. Mathematically, only three types of symmetric operations, that is, rotation, inversion, and translation are involved. Starting with the basic set of coordinates of a protein subunit, the virus capsid data can be obtained by the transformation

$$\begin{pmatrix} X \\ Y \\ Z \end{pmatrix} = \begin{pmatrix} r_{11} & r_{12} & r_{13} & t_1 \\ r_{21} & r_{22} & r_{23} & t_2 \\ r_{31} & r_{32} & r_{33} & t_3 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \\ 1 \end{pmatrix}, \quad (18)$$

where  $r_{ij}$  are rotational (or inversion) elements, and  $t_i$  are translational elements. The viral data deposited in the Protein Data Bank (PDB) often have problems with missing sets of transformation operations and erroneous coordinate-frame representations. We make corrections by using the Virus Particle Explorer database [7] (VIPERdb; <http://viperdbscripps.edu/>) and/or the Protein Quaternary Structure server (PQS; <http://pqs.ebi.ac.uk/>).

### 3. Numerical Demonstration

Recent advances in structural biology and microbiology have given rise to an increasing body of structural data for over 300 viruses and viral complexes. Quaternary structures of viruses and viral complexes pose many challenges for viral representation, visualization, and the analysis of virus stability and interaction [6]. The proposed multiscale framework can be studied on a wide range of test cases to demonstrate its utility and usefulness to the research community. However, a full-scale demonstration of the proposed multiscale model is a rather computationally challenging task as it involves computational fluid dynamics (CFD), molecular dynamics (MD) of viruses, and surface dynamics of large systems. In this paper, we should primarily focus on the virus surface formation and evolution. The coupling of the surface dynamics to the CFD and MD will be studied in our future work and published elsewhere.

We also test two other proposed ideas in this work, that is, the coarse-grained virus model and the use of symmetry assembly for the virus surface construction. In particular, we are interested in examining the effect of the symmetry assembly on the virus surface visualization. As shown in Figure 3, we consider the coarse-grained model, which is an efficient way to reduce computational cost. Additionally, we test the surface construction by using symmetric assembly. In comparison with surfaces constructed by potential driven geometric flows without using the symmetry (Lower row), geometric flow surfaces constructed by symmetry (Upper row) provide a good representation of the original surfaces. However, one can still see that contact edges in the surfaces constructed by symmetry are not very smooth. Moreover, we expect some impact of symmetric assembly to the MD and fluid dynamics, as the symmetry becomes an additional constraint to virus dynamical motions. The soundness of such a constraint needs to be studied. This aspect as well as many other ideas proposed in this work will be further explored elsewhere.

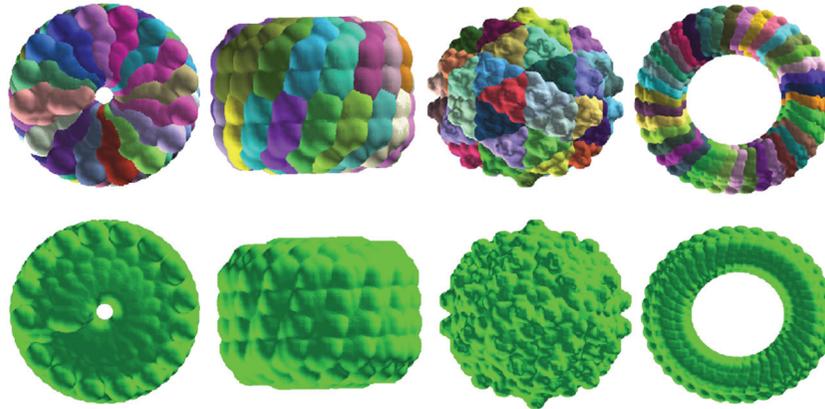


FIGURE 3: Illustration of virus surfaces constructed by using the proposed geometric flow approach in conjunction with the coarse-grained model and the symmetry assembly. Upper row: Surfaces generated from a facet patch by using symmetry assembly. Lower row: Surfaces generated without the use of symmetry. From left to right: Cucumber green mottle mosaic virus (CGMMV) with helical symmetry (1cgm), Tobacco mosaic virus coat protein four-layer aggregate with  $D_{17}$  symmetry (1ei7), Nodamura virus with icosahedral symmetry (1nov), and Viral toxin pneumolysin with  $C_{38}$  circular symmetry (2bk1).

#### 4. Concluding Remarks

The control of infective viruses released by terrorists, and the prevention of viral epidemics and pandemics, such as HIV, SARS, H1N1, and bird flu are of tremendous importance. The understanding of viral surface formation, evolution, viral attachment and penetration of host cells are prerequisites to viral disease prevention and control. This problem, as well as many other similar problems in molecular biology, poses pressing challenges to the theoretical community due to their large number of degrees of freedom. The main purpose of the present work is to introduce a differential geometry-based multiscale framework to handle complex biological systems. The present multiscale model couples macroscopic fluid dynamics, microscopic molecular dynamics, and surface dynamics in a unified framework. The differential geometry theory of surfaces is utilized to put continuum description and discrete description in an equal footing. The present work constructs a generalized action functional to self-consistently couple different scales. Governing equations for the fluid dynamics, that is, the generalized Navier-Stokes equation, and molecular dynamics, that is, the Newton's equation, are derived by minimizing the action functional. Additionally, we make use of viral symmetry to dramatically reduce viral data sizes and improve viral visualization. Finally, some of the proposed approaches are demonstrated by the generation of a few virus surfaces.

The proposed differential geometry-based multiscale model can be easily generalized to complex systems with multiple interfaces or many biomolecules. Additionally, the incorporation of continuum solid description into the present model will be published elsewhere. Finally, the inclusion of a quantum mechanical description can also be pursued in a similar way and will be published elsewhere. Numerical experiments that further demonstrate the proposed ideas are under our consideration.

#### Acknowledgments

This work was supported in part by NSF Grants DMS-0616704 and CCF-0936830, and NIH Grants CA-127189 and GM-090208.

#### References

- [1] T. P. Weber and N. I. Stilianakis, "Inactivation of influenza A viruses in the environment and modes of transmission: a critical review," *Journal of Infection*, vol. 57, no. 5, pp. 361–373, 2008.
- [2] C. M. Fauquet and J. Stanley, "Revising the way we conceive and name viruses below the species level: a review of geminivirus taxonomy calls for new standardized isolate descriptors," *Archives of Virology*, vol. 150, no. 10, pp. 2151–2179, 2005.
- [3] S. S. Rawat, M. Viard, S. A. Gallo, A. Rein, R. Blumenthal, and A. Puri, "Modulation of entry of enveloped viruses by cholesterol and sphingolipids," *Molecular Membrane Biology*, vol. 20, no. 3, pp. 243–254, 2003.
- [4] J. L. McKimm-Breschkin, "Resistance of influenza viruses to neuraminidase inhibitors—a review," *Antiviral Research*, vol. 47, no. 1, pp. 1–17, 2000.
- [5] S. Dutta and H. M. Berman, "Large macromolecular complexes in the protein data bank: a status report," *Structure*, vol. 13, no. 3, pp. 381–388, 2005.
- [6] C. L. Lawson, S. Dutta, J. D. Westbrook, K. Henrick, and H. M. Berman, "Representation of viruses in the remediated PDB archive," *Acta Crystallographica Section D*, vol. 64, no. 8, pp. 874–882, 2008.
- [7] C. M. Shepherd, I. A. Borelli, G. Lander, et al., "VIPERdb: a relational database for structural virology," *Nucleic Acids Research*, vol. 34, pp. D386–D389, 2006.
- [8] T. S. Baker, N. H. Olson, and S. D. Fuller, "Adding the third dimension to virus life cycles: three dimensional reconstruction of icosahedral viruses from cryo-electron micrographs," *Microbiology and Molecular Biology Reviews*, vol. 63, no. 1, pp. 862–922, 1999.

- [9] Z. Yu and C. Bajaj, "Automatic ultrastructure segmentation of reconstructed cryoEM maps of icosahedral viruses," *IEEE Transactions on Image Processing*, vol. 14, no. 9, pp. 1324–1337, 2005.
- [10] Z. Yu and C. Bajaj, "Computational approaches for automatic structural analysis of large biomolecular complexes," *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, vol. 5, no. 4, pp. 568–582, 2008.
- [11] A. Arkhipov, P. L. Freddolino, and K. Schulten, "Stability and dynamics of virus capsids described by coarse-grained modeling," *Structure*, vol. 14, no. 12, pp. 1767–1777, 2006.
- [12] G. W. Wei, "Generalized Perona-Malik equation for image restoration," *IEEE Signal Processing Letters*, vol. 6, no. 7, pp. 165–167, 1999.
- [13] G. W. Wei and Y. Q. Jia, "Synchronization-based image edge detection," *Europhysics Letters*, vol. 59, no. 6, pp. 814–819, 2002.
- [14] Y. Sun, P. Wu, G. W. Wei, and G. Wang, "Evolution-operator-based single-step method for image processing," *International Journal of Biomedical Imaging*, vol. 2006, Article ID 83847, 27 pages, 2006.
- [15] M. Lysaker, A. Lundervold, and X.-C. Tai, "Noise removal using fourth-order partial differential equation with applications to medical magnetic resonance images in space and time," *IEEE Transactions on Image Processing*, vol. 12, no. 12, pp. 1579–1589, 2003.
- [16] G. Gilboa, N. Sochen, and Y. Y. Zeevi, "Image sharpening by flows based on triple well potentials," *Journal of Mathematical Imaging and Vision*, vol. 20, no. 1-2, pp. 121–131, 2004.
- [17] A. L. Bertozzi and J. B. Greer, "Low-curvature image simplifiers: global regularity of smooth solutions and laplacian limiting schemes," *Communications on Pure and Applied Mathematics*, vol. 57, no. 6, pp. 764–790, 2004.
- [18] J. B. Greer and A. L. Bertozzi, "H1 solutions of a class of fourth order nonlinear equations for image processing," *Discrete and Continuous Dynamical Systems*, vol. 10, no. 1-2, pp. 349–366, 2004.
- [19] J. B. Greer and A. L. Bertozzi, "Traveling wave solutions of fourth order PDEs for image processing," *SIAM Journal on Mathematical Analysis*, vol. 36, no. 1, pp. 38–68, 2005.
- [20] G. W. Wei, Y. H. Sun, Y. C. Zhou, and M. Feig, "Molecular multiresolution surfaces," submitted to *Mathematical Physics*.
- [21] P. W. Bates, G. W. Wei, and S. Zhao, "Minimal molecular surfaces and their applications," *Journal of Computational Chemistry*, vol. 29, no. 3, pp. 380–391, 2008.
- [22] P. W. Bates, Z. Chen, Y. Sun, G.-W. Wei, and S. Zhao, "Geometric and potential driving formation and evolution of biomolecular surfaces," *Journal of Mathematical Biology*, vol. 59, no. 2, pp. 193–231, 2009.
- [23] S. Osher and J. A. Sethian, "Fronts propagating with curvature-dependent speed: algorithms based on Hamilton-Jacobi formulations," *Journal of Computational Physics*, vol. 79, no. 1, pp. 12–49, 1988.
- [24] C. Y. Kao, S. Osher, and E. Yablonovitch, "Maximizing band gaps in two-dimensional photonic crystals by using level set methods," *Applied Physics B*, vol. 81, no. 2-3, pp. 235–244, 2005.
- [25] Q. He, Y. Duan, J. Miles, and N. Takahashi, "A context-sensitive active contour for 2D corpus callosum segmentation," *International Journal of Biomedical Imaging*, vol. 2007, Article ID 24826, 8 pages, 2007.
- [26] P. Blomgren and T. F. Chan, "Color TV: total variation methods for restoration of vector-valued images," *IEEE Transactions on Image Processing*, vol. 7, no. 3, pp. 304–309, 1998.
- [27] Y. Wang and H. Zhou, "Total variation wavelet-based medical image denoising," *International Journal of Biomedical Imaging*, vol. 2006, Article ID 89095, 6 pages, 2006.
- [28] R. Szeliski and D. Tonnesen, "Surface modeling with oriented particle systems," *Computer Graphics*, vol. 26, no. 2, pp. 185–194, 1992.
- [29] V. Tozzini, "Coarse-grained models for proteins," *Current Opinion in Structural Biology*, vol. 15, no. 2, pp. 144–150, 2005.
- [30] G. W. Wei, "Differential geometry based multiscale models," *Bulletin of Mathematical Biology*, pp. 1–60, 2010.
- [31] K. A. Sharp and B. Honig, "Electrostatic interactions in macromolecules: theory and applications," *Annual Review of Biophysics and Biophysical Chemistry*, vol. 19, pp. 301–332, 1990.
- [32] M. K. Gilson, M. E. Davis, B. A. Luty, and J. A. McCammon, "Computation of electrostatic forces on solvated molecules using the Poisson-Boltzmann equation," *Journal of Physical Chemistry*, vol. 97, no. 14, pp. 3591–3600, 1993.
- [33] S. Yu, W. Geng, and G. W. Wei, "Treatment of geometric singularities in implicit solvent models," *Journal of Chemical Physics*, vol. 126, no. 24, Article ID 244108, 2007.
- [34] W. Geng, S. Yu, and G. W. Wei, "Treatment of charge singularities in implicit solvent models," *Journal of Chemical Physics*, vol. 127, no. 11, Article ID 114106, 2007.
- [35] Y. C. Zhou, M. Feig, and G. W. Wei, "Highly accurate biomolecular electrostatics in continuum dielectric environments," *Journal of Computational Chemistry*, vol. 29, no. 1, pp. 87–97, 2008.
- [36] N. J. Mitra, L. J. Guibas, and M. Pauly, "Partial and approximate symmetry detection for 3D geometry," *ACM Transactions on Graphics*, vol. 25, no. 3, pp. 560–568, 2006.
- [37] N. J. Mitra, L. J. Guibas, and M. Pauly, "Symmetrization," in *Proceedings of the 34th International Conference on Computer Graphics and Interactive Techniques (SIGGRAPH '07)*, ACM, San Diego, Calif, USA, August 2007.

