Coarse-Grained Molecular Dynamics Simulations

Charles Xie
Institute for Future Intelligence, Natick, MA 01760

Interactive molecular simulations have been used in science education [1, 2]. Observing important physical processes, such as diffusion, phase transitions, self-assembly, chemical reactions and so on, emerging from the time evolution of an atomic-scale model can help learners build intuitive, coherent, and predictive mental models of these phenomena.

The increasing power of computers, however, does not automatically remove all the obstacles for building interactive computational models for use in teaching more complicated subjects such as molecular biology, which need more computational resources. The structural complexity of many macromolecules poses a challenge. An accurate atomistic simulation of complex macromolecules based on the all-atom force fields in molecular mechanics [3] requires time-consuming computations. Such simulations may not, however, necessarily result in useful visualizations for students. Many students have difficulties in discerning what is important and what is not from a complicated three-dimensional structural view full of atoms and bonds (e.g., in a ball-and-stick representation). Because of this, coarse-grained "cartoon" views, which show a macromolecule as a single object in a certain shape or a conjunction of a few objects in simple shapes, are commonly used in educational illustrations. These "cartoon" representation omits the internal details of a macromolecule that may not be important to its function. Instead, they highlight the most critical parts (often the active sites or surfaces where functions are performed) in such a clear way that a comprehensible picture of complex systems can therefore be conveyed to learners. Such a "cartoonization" method has been used in molecular graphics software for visualizing secondary and tertiary structures (e.g., the ribbon, tube, Greek key, and surface representations available in most molecular visualization tools). Fig. 1 shows an example of representing polymers with sticks with or without branches.
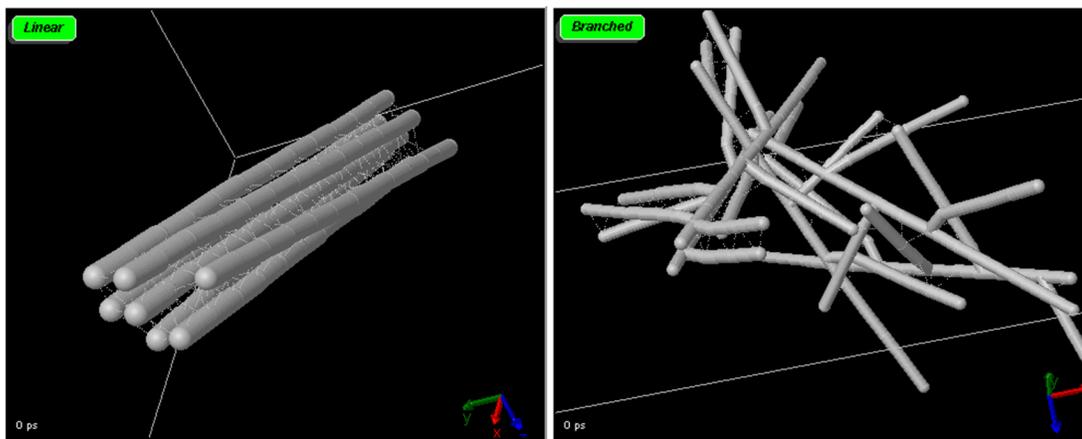
Fig. 1. Linear and branched polymers can be represented by sticks without atomic details.

Because of this difficulty, many animations are created schematically without using rigorous calculations of forces and trajectories to generate the motions, especially at the level of quaternary structures where the number of atoms involved makes the task intractable. These animations usually use "cartoon" objects to represent macromolecules, but do not actually generate the "movie" based on force field calculation of any kind for the objects. This is presumably because computing forces and motions of this kind of "cartoon" objects requires using potential energy models for objects in arbitrary shapes.

To introduce objects in arbitrary shapes to molecular simulations, mathematical models for the interaction potentials among them must be developed. In order to show important biochemical and biophysical processes such as ligand-protein binding, docking and enzymatic reactions, these potentials must be "sticky" in certain conditions to certain objects, so that active sites can be modeled. Furthermore, they must be able to produce as many varieties of structures as possible, to meet the needs of modeling biomolecular systems that have far more complex structural patterns than solids and liquids.

A well-known model for interatomic interactions is the Lennard-Jones potential [4], which, however, applies to only spherical objects without intrinsic rotational degree of freedom. Gay and Berne proposed a potential function that can be applied to elliptic objects [5], which is mainly used in simulating mesogenic systems such as liquid crystals.

This paper first reviews some applications of Lennard-Jones potential in modeling condensed matter, and then introduces our modified Gay-Berne model and its applications. Finally, we present our work about modeling soft objects in arbitrary shapes and discuss about their applications to molecular biology.

The Lennard-Jones model

To simulate any condensed state, and most phase transitions, attractive forces between atoms are needed. There are many mathematical models for attractive pair potentials. The most widely used model is the Lennard-Jones (LJ) model [4], but many others have a similar shape (e.g., the Morse potential) and are qualitatively equivalent to the LJ model, in the sense that they all generate repulsive forces when particles get too close and attractive forces otherwise.

The LJ particles without charges interact with one another in an isotropic manner. As a result, a system consisting of LJ particles of the same type tend to form the most closed-pack structure, which is a hexagonal lattice in two-dimension (Fig. 2a), when they form a solid. If a system is composed of LJ particles of distinct types (especially with different collision diameters), it usually forms amorphous structures (Fig. 2b) from a liquid phase, provided that the potentials between different types are given by some sort of averages of those between the same types, such as the Lorentz-Berthelot mixing rule. When the two different types of particles in Fig. 1b are given strong positive and negative charges respectively, the system crystallizes into a regular square lattice (Fig. 2c). Because of its capacity in simulating a variety of structures, the LJ potential is good for modeling such systems as crystals, amorphous alloys, intermetallic compounds, and ionic compounds (in reality, researchers use improved pair functional models, such as the embedded-atom method [6] that handles more correctly elastic properties and defect formation energies, to simulate these systems).

Although the LJ potential is mostly used to model the van der Waals interactions between atoms, it is also regarded as a fair approximation to other systems in which the composing elements attract each other but cannot overlap. For example, it is not uncommon to use the LJ potential to simulate fluid dynamics such as thermal convection and obstructed flow [7], despite of the fact that almost no fluid molecules of interest look like a spherical LJ particle. In this case, the LJ particles can be thought of as an approximation in which the atoms in the molecules shrink to their centers of mass, and the polyatomic molecules become point masses; the LJ potentials between them produce qualitatively the dynamical behavior of the fluid molecules, if they keep moving around and the multi-site interactions even out to show no strong orientational preference in the long range. This approximation is sometimes also used in molecular mechanics to model chemical groups of

macromolecules that are not critically important to the simulated processes, for instance a methyl group, and is known by molecular simulators as the united-atom approximation [3].

The LJ potential, however, cannot be used to model more complicated phenomena in which the elements of the system interacts with others in an anisotropic and selective matter. Such interactions are usually responsible for the formation of more complicated structures than close-packed lattices. For example, the LJ potential cannot produce the carbon structures of diamond or graphite, because of the strong orientational dependence of the covalent bonds between carbon atoms. There have been many potential models for addressing the orientational dependence of interatomic interactions in silicon and carbon materials. Most of them involve introducing angular factors and many-body (i.e. three-body and more) terms to the potential functions [8-10].

(a)

(b)

(c)
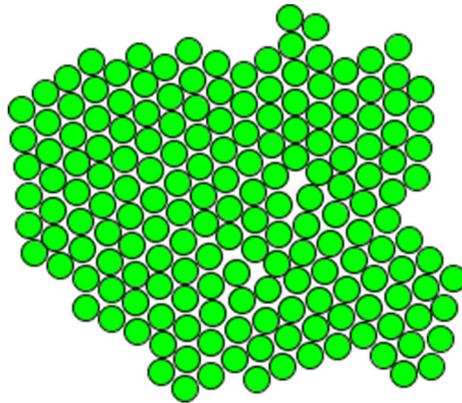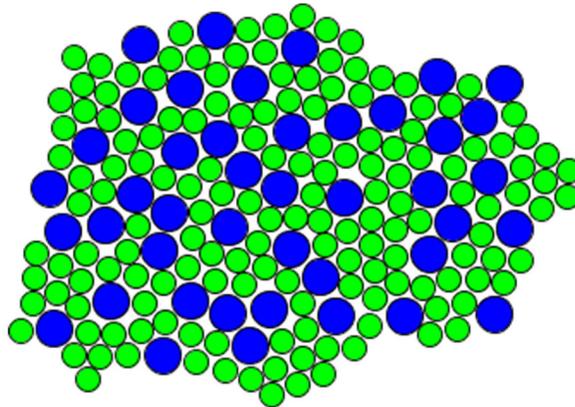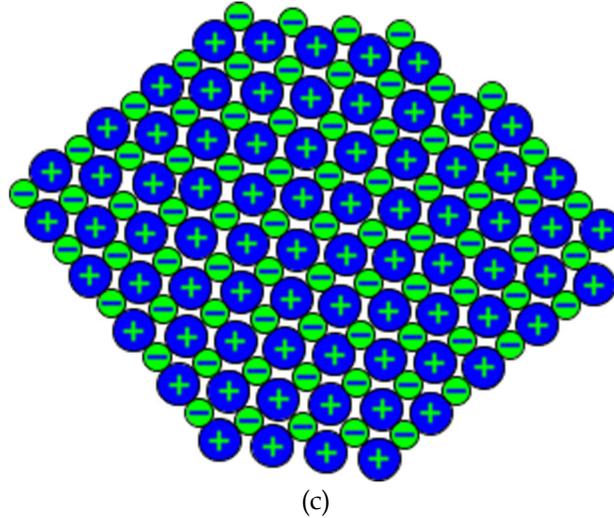
Fig. 2. Three structures formed by the LJ particles produced in molecular dynamics simulations: (a) A hexagonal close-packed lattice (with defects); (b) An amorphous solid; (c) An ionic crystal.

A potential capable of modeling orientational dependence but does not introduce three-body (and higher order) interaction is the Gay-Berne (GB) model, which will be discussed in the next Section.

The modified Gay-Berne model

Because the LJ particles do not have a rotational degree of freedom, they cannot show order-disorder transition of orientation, which is a typical phenomenon observed in liquid crystals. The GB model was proposed to model such phenomena. It can simulate nematic, discotic nematic, and biaxial nematic phases of liquid crystals, and has become popular for modeling mesogenic systems. The commonly used GB model applies only to systems consisting of elliptic particles with identical sizes, which makes it less useful in extending to systems of generic interest. Cleaver, Care, Allen, and Neal (CCAN) generalized the GB model to systems composed of elliptical particles of different sizes [11]. This generalization broadens the scope of systems that the GB model can be used to simulate. There are many processes in which the internal structures of the molecules do not change or play an important role. For example, nanoscale self-assembly of stable chemical compounds is primarily caused by electrostatic interactions rather than covalent bonding. These processes can be modeled based on the generalized GB potential (see Fig. 3 for a schematic illustration).
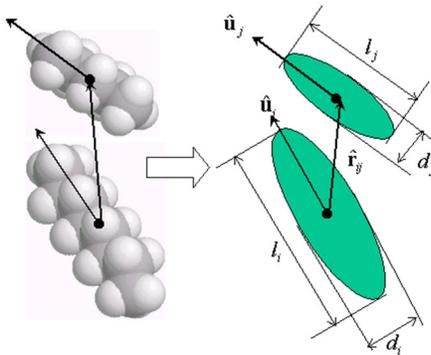
Fig. 3. The Gay-Berne potential can be used to model approximately molecules and their interactions when their internal structures and intramolecular forces do not have a significant contribution to the overall assembly dynamics.

The GB model has a similar 12-6 form to that of the LJ potential but is far more complicated. Suppose the length of an elliptic particle i is l_i, the breadth is d_i, it has a net charge q_i and an electric dipole moment p_i, and its orientation is represented by a unit vector $\hat{\mathbf{u}}_i$ (see Fig. 2). The modified GB potential can be written as:

$$
\begin{aligned}
V_{\mathrm{GB}} = \sum_i \sum_{j>i} 4\varepsilon(\hat{\mathbf{u}}_i,\hat{\mathbf{u}}_j,\hat{\mathbf{r}}_{ij}) &\left\{ \left[\frac{\sigma_0}{r_{ij} - \sigma(\hat{\mathbf{u}}_i,\hat{\mathbf{u}}_j,\hat{\mathbf{r}}_{ij}) + \sigma_0} \right]^{12} - \left[\frac{\sigma_0}{r_{ij} - \sigma(\hat{\mathbf{u}}_i,\hat{\mathbf{u}}_j,\hat{\mathbf{r}}_{ij}) + \sigma_0} \right]^{6} \right\} \\
&+ \sum_i \sum_{j>i} \frac{q_i q_j}{r_{ij}} + \sum_i \sum_{j>i} \frac{1}{r_{rij}^2} \left(q_j p_i \hat{\mathbf{u}}_i \cdot \hat{\mathbf{r}}_{ij} + q_i p_j \hat{\mathbf{u}}_j \cdot \hat{\mathbf{r}}_{ij} \right) \\
&+ \sum_i \sum_{j>i} \frac{p_i p_j}{r_{ij}^3} \left[\hat{\mathbf{u}}_i \cdot \hat{\mathbf{u}}_j - 3(\hat{\mathbf{u}}_i \cdot \hat{\mathbf{r}}_{ij})(\hat{\mathbf{u}}_j \cdot \hat{\mathbf{r}}_{ij}) \right]
\end{aligned}
\tag{1}
$$

where

$$
\sigma(\hat{\mathbf{u}}_i,\hat{\mathbf{u}}_j,\hat{\mathbf{r}}_{ij}) = \sigma_{ij}^0 \left\{ 1 - \frac{\chi}{2} \left[\frac{(\alpha\hat{\mathbf{r}}_{ij} \cdot \hat{\mathbf{u}}_i + \alpha^{-1}\hat{\mathbf{r}}_{ij} \cdot \hat{\mathbf{u}}_j)^2}{1 + \chi\hat{\mathbf{u}}_i \cdot \hat{\mathbf{u}}_j} + \frac{(\alpha\hat{\mathbf{r}}_{ij} \cdot \hat{\mathbf{u}}_i - \alpha^{-1}\hat{\mathbf{r}}_{ij} \cdot \hat{\mathbf{u}}_j)^2}{1 - \chi\hat{\mathbf{u}}_i \cdot \hat{\mathbf{u}}_j} \right] \right\}^{-1/2}
\tag{2}
$$

is the orientation-dependent range parameter derived from the equi-density surface of the electron cloud of a uniaxially stretched Gaussian distribution, $\hat{\mathbf{r}}_{ij}$ is the distance unit vector between the particles i and j, and σ_{ij}^0 is the contact parameter set as the quadratic mean of the breadths of the two interacting particles: $\sqrt{(d_i^2 + d_j^2)/2}$.

χ and α are two parameters, defined as:

$$
\chi = \left[\frac{(l_i^2 - d_i^2)(l_j^2 - d_j^2)}{(l_j^2 + d_i^2)(l_i^2 + d_j^2)} \right]^{1/2}, \quad l_i > d_i \quad \text{and} \quad l_j > d_j
\tag{3}
$$

and

$$
\alpha = \left[\frac{(l_i^2 - d_i^2)(l_j^2 + d_i^2)}{(l_j^2 - d_j^2)(l_i^2 + d_j^2)} \right]^{1/4}, \quad l_i > d_i \quad \text{and} \quad l_j > d_j
\tag{4}
$$

respectively.

$\varepsilon(\hat{\mathbf{u}}_i,\hat{\mathbf{u}}_j,\hat{\mathbf{r}}_{ij})$ is the well depth of the interaction potential:

$$\varepsilon(\hat{\mathbf{u}}_i, \hat{\mathbf{u}}_j, \hat{\mathbf{r}}_{ij}) = \sqrt{\varepsilon_{0,i}\varepsilon_{0,j}}\,\varepsilon_1^\nu(\hat{\mathbf{u}}_i, \hat{\mathbf{u}}_j)\varepsilon_2^\mu(\hat{\mathbf{u}}_i, \hat{\mathbf{u}}_j, \hat{\mathbf{r}}_{ij}) \tag{5}$$

where $\varepsilon_{0,i}, \varepsilon_{0,j}$, μ and ν are adjustable parameters. The first function depends on the relative orientation of the two particles but not the distance between them:

$$\varepsilon_1^\nu(\hat{\mathbf{u}}_i, \hat{\mathbf{u}}_j) = \left[1 - \chi^2\left(\hat{\mathbf{u}}_i \cdot \hat{\mathbf{u}}_j\right)^2\right]^{-1/2} \tag{6}$$

The second function depends on both the orientation and distance:

$$\varepsilon_2^\mu(\hat{\mathbf{u}}_i, \hat{\mathbf{u}}_j, \hat{\mathbf{r}}_{ij}) = 1 - \frac{\chi'_{ij}}{2}\left[\frac{\left(\alpha'_{ij}\hat{\mathbf{r}}_{ij} \cdot \hat{\mathbf{u}}_i + \alpha'^{-1}_{ij}\hat{\mathbf{r}}_{ij} \cdot \hat{\mathbf{u}}_j\right)^2}{1 + \chi'_{ij}\hat{\mathbf{u}}_i \cdot \hat{\mathbf{u}}_j} + \frac{\left(\alpha'_{ij}\hat{\mathbf{r}}_{ij} \cdot \hat{\mathbf{u}}_i - \alpha'^{-1}_{ij}\hat{\mathbf{r}}_{ij} \cdot \hat{\mathbf{u}}_j\right)^2}{1 - \chi'_{ij}\hat{\mathbf{u}}_i \cdot \hat{\mathbf{u}}_j}\right] \tag{7}$$

where the anisotropy parameters of interaction energy for this pair of particles are defined as:

$$\chi'_{ij} = \frac{1 - \left[\left(\varepsilon_E^i/\varepsilon_S^i\right)\left(\varepsilon_E^j/\varepsilon_S^j\right)\right]^{1/(2\mu)}}{1 + \left[\left(\varepsilon_E^i/\varepsilon_S^i\right)\left(\varepsilon_E^j/\varepsilon_S^j\right)\right]^{1/(2\mu)}} \tag{8}$$

and

$$\alpha'_{ij} = \sqrt{\frac{1}{1 + \left[\left(\varepsilon_E^i/\varepsilon_S^i\right)\left(\varepsilon_E^j/\varepsilon_S^j\right)\right]^{1/(2\mu)}}} \tag{9}$$

respectively. In the above equations, ε_S^i and ε_S^j are the well depths corresponding to the side-by-side arrangement for a pair of particles of the same type, and ε_E^i and ε_E^j are those corresponding to the end-to-end arrangement.

The CCAN version of the GB model assumes that all GB particles have the same energetic anisotropy (quantified by $\varepsilon_E^i/\varepsilon_S^i$), which favors the formation of a uniform liquid crystalline phase. In our models, the GB particles can have distinct energetic anisotropies. This permits mesogens to form amorphous phases more easily.

The last three terms in Eq. (1) are charge-charge, charge-dipole and dipole-dipole interactions, respectively. Higher orders of electric moments from quadrupoles on are neglected. These three terms are our addition to the CCAN potential. With these electrostatic interactions, the GB particles can respond to an applied electric field, or form structures that are otherwise not possible (e.g. electrostatic self-assembly).

When the GB particles become spherical (i.e., $l_i = d_i$, $\hat{\mathbf{u}}_i = 0$ and/or $p_i = 0$), Eq. (1) returns to the conventional LJ 12-6 form and Coulombic potential. Obviously, the GB model requires much more calculations than the LJ model. It can be surmised from this that a similar potential for arbitrary shapes may be very computationally expensive, even if it has a closed form that can be analytically derived.

The GB model is able to produce more structure variations than does the LJ one. For example, Fig. 4 shows three typical structures formed by the GB particles of the same shape (but with a different $\varepsilon_E^i / \varepsilon_S^i$ parameter for each case). Fig. 5 shows a snapshot of molecular simulation of GB particles of different shapes, some of them are charged and some polarized. The increasing of the number of different structures that the GB model can generate enables systems with greater structural complexity to be modeled.
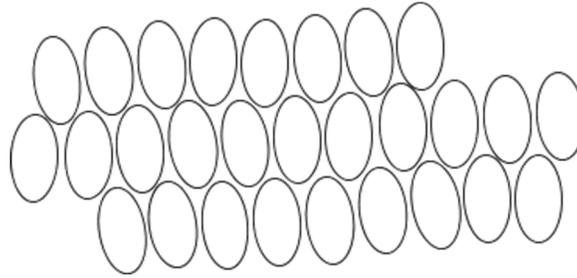
(a) $\varepsilon_E^i / \varepsilon_S^i = 0.1$
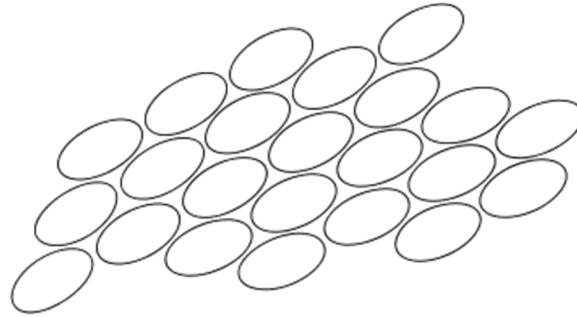
(b) $\varepsilon_E^i / \varepsilon_S^i = 10$
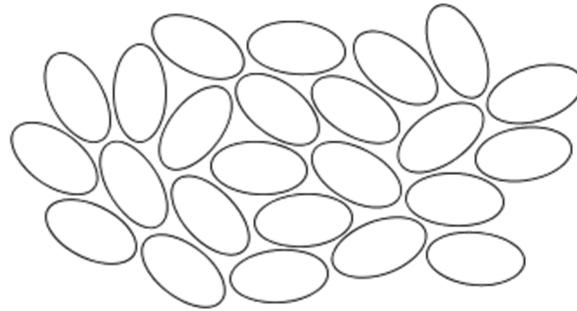
(c) $\varepsilon_E^i / \varepsilon_S^i = 1$

Fig. 4. Three typical structures formed by the GB particles in molecular dynamics simulations: (a) Side-by-side conformation; (b) End-to-end conformation; (c) Disordered conformation.
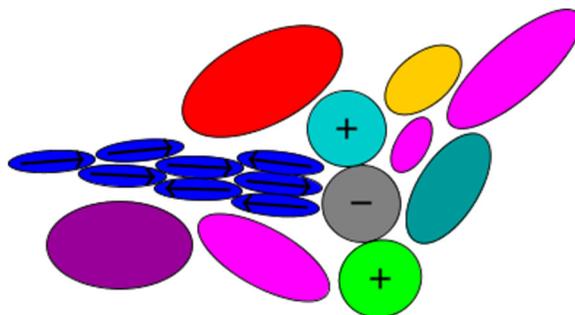
Fig. 5. A snapshot of a molecular simulation for some GB particles of different shapes shows a structural pattern very different from the close-packed one preferred by the LJ particles. The arrows on the GB particles stand for electric dipole moments (pointing from the positive end to the negative end). The plus and minus signs represent positive and negative charges.

Two-dimensional soft bodies in arbitrary shapes

Although the GB potential is capable of simulating non-spherical interacting particles, it is limited to modeling only elliptical particles. To build useful coarse-grained models for molecular biology without having to use full ball-and-stick representations of macromolecules, interacting objects that have arbitrary shapes are needed.

Given the difficulty in deriving an analytical model for interacting particles in arbitrary shapes, we can take a different approach. We can create a drawing tool to permit users to draw arbitrary non-self-intersected open and closed shapes, such as rectangles, ellipses, cubic splines, or free-form lines. When the user finishes drawing a shape, LJ particles with appropriate size are automatically aligned along the line (if the shape is open), or along the border (if the shape is closed), to its full length. After the particle alignment is done, harmonic forces are used to connect the aligned LJ particles in radial and angular directions, to combine them into a single object. Such an object has a van der Waals surface that can attract (and be attracted to) other similar objects or LJ particles. It has also a border that is formed by the repulsive core of the LJ potential to prevent overlaps with another object or a LJ particle. If an atomic probe were used to scan over the border, an enveloping shape would be obtained, because the repulsive cores bounce back the probe. In molecular biology, such an enveloping shape generated by a real or hypothetical probe is called a molecular surface (e.g., electronic density map by a scanning tunneling microscope or solvent accessible surface by a hypothetical solvent molecule). Hence, we call this type of object the *molecular surface* (MS).

Fig. 6 shows the scanning results of equipotential contours of some MS objects using a LJ particle as a probe. The buckling of the contour lines near to the objects reflects the existence of the LJ particles behind the scenes (as shown by the dashed circles in Fig. 6b).
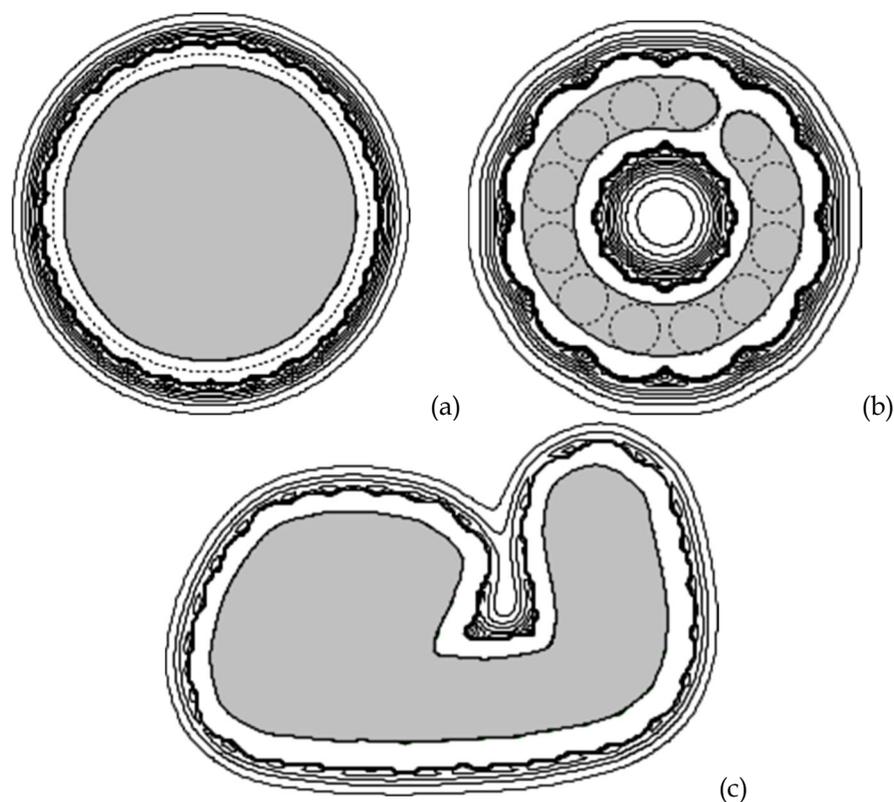
Fig. 6. Equipotential contour plots of some molecular surface objects. The first one can be thought of as a "giant" LJ particle. The second object has a potential energy well that can store other molecules. The third one forms a shallower "pocket" that can bind other objects.

An MS object is not a rigid body. Its shape vibrates, and can be distorted. The harmonic forces used to bind the LJ particles form a "spring chain" that maintains the shape. The rigidity of the shape rests on the strength of the harmonic forces. If the harmonic forces are infinitely strong, it will become a rigid body.

If the whole shape is rigid enough, the translation and rotation of an MS object can still be approximately represented by its mass and inertia, which are given by the total mass of its LJ particles and determined by the distribution of the masses of LJ particles, respectively. For example, in a collision event, the heavier MS object moves less. This is because the effect of a local interaction, e.g. an impact, can be rapidly propagated through the "spring chain" that connects the affected LJ particles with others, which subsequently results in a net effect of the whole object. The net effect becomes less significant if the "spring chain" is more flexible and can fluctuate more to absorb the impact energy. In that case, the inelastic behavior will be more significant. In other words, more kinetic energy will be converted into the internal vibrational energy of the MS object and less to the translational energy of the center of mass.

Although LJ particles and harmonic forces are used to maintain the shape of an MS object in a molecular simulation, an opaque enveloping geometric object is always used to cover them to prevent misunderstanding. To an end user, an MS object is just a soft interacting object that can have an irregular shape and can be charged at different spots. The user should not be confused with the underlying LJ particles, which are used merely for constructing the object and have no real-world meaning.

The MS model offers a much-simplified way to model ensembles of macromolecules. It catches the essential idea that the surface of a macromolecule is generally far more important than the interior in facilitating intermolecular interactions and active site reactions. The complete omission of the inner part of a closed object (i.e., a hollow MS object) not only saves a lot of computing time that would have been spent on calculating the motions of atoms and bonds belonging to that region if they were considered, but also results in a straightforward view of the science being expressed. For example, a molecular simulation for the intermolecular interactions of macromolecules in aqueous solution, in which the macromolecules are represented by MS objects (see Fig. 7), shows clearly the interactions between the charged sites and water molecules around them.
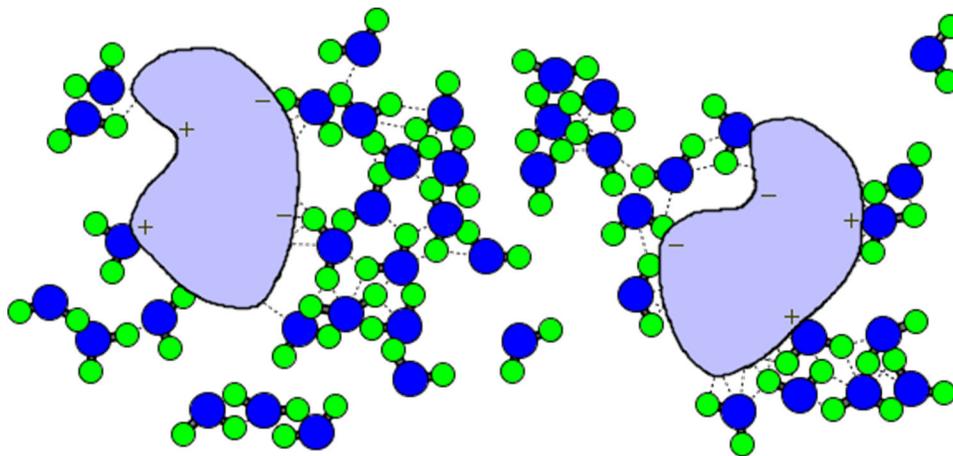
Fig. 7. A screenshot of molecular dynamics simulation of intermolecular interactions in aqueous solution is shown. The smaller circles represent hydrogen atoms, the bigger ones represent oxygen atoms, and the two large objects represent certain macromolecules. The plus and minus signs on the large objects are the charges that make them polar. The dotted lines show the hydrogen bonds. A real-time molecular simulation will show that with this setting, the water molecules will move around, but the two macromolecules are always surrounded by water molecules near the charged sites, and they will finally reach each other and expel much water originally distributed between them (the effective interaction between two solvated molecules is called the potential of mean force, a useful concept used by chemists to describe reactions in solution). Such a simulation may help students understand how water mediates chemical reactions between proteins.

The MS objects can be used to easily create simple models that demonstrate complicated ideas. For example, Fig. 8 illustrates a ligand-protein binding simulation (docking). This simulation shows a successful docking. By varying the model, students can discover the two most important factors in molecular recognition: attractive interactions between sites and shape complementarities. For example, if either one of the charges (or both) are removed, docking may not happen, or happen much more slowly (due to the weak van der Waals interactions between the complementary shapes, they may still have some chance to get together). If the shape compatibility is broken, for instance, the ligand is made much larger than the cavity can accommodate, docking will never happen.

Fig. 8 also shows the difference between using a soft body and using a rigid body. If the ligand and protein were modeled using rigid bodies, docking would probably not occur if the cavity were slightly smaller than the ligand. But because MS objects are flexible, they can yield to the pulling electrostatic force between the charges on them to make docking happen. This advantage makes the MS model a good candidate for building models to illustrate the induced-fit process of enzyme action, in which the enzyme changes its shape slightly to fit and hold the substrates so that the chemical reaction can be catalyzed more easily.
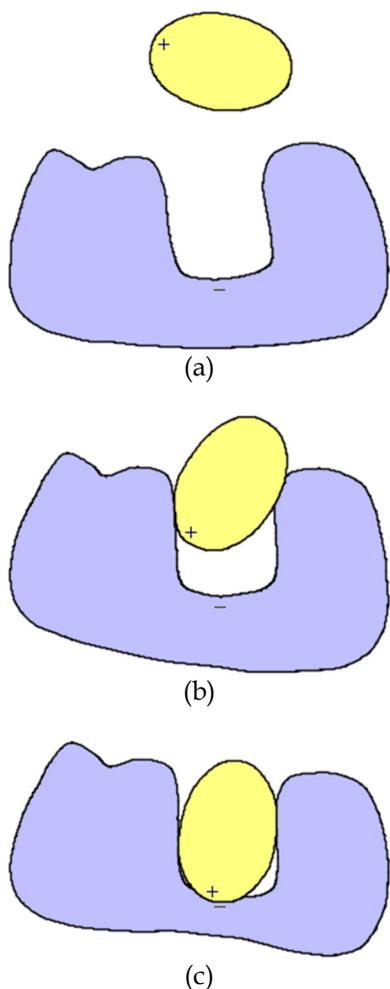
Fig. 8. The molecular surface objects can be used to create simple molecular dynamics simulations that are like cartoon movies (and therefore could be readily understood by students). This figure shows three stages of a docking simulation: (a) approaching; (b) entering; (c) docked. The ligand is represented by an elliptical object with a positive charge on one end of its longer axis. The substrate is represented by an irregular shape with an open curved cavity. There is a negative charge at the bottom of the cavity.

Summary

Designing visual and interactive models that are pedagogically effective and computationally efficient for teaching molecular biology is a challenge to educators who would like to exploit the fast-increasing computer power. Although there has been explosive growth of the number of research papers published every year about molecular simulations in all major science and engineering fields, little attention has been paid to transfer the algorithms and techniques developed by computational scientists to education and/or adapt them to building computer models for classroom use.

The Institute for Future Intelligence hopes to bridge this gap. This paper summarizes our efforts in building effective molecular dynamics models for use in education, based on spherical Lennard-Jones particles, elliptical Gay-Berne particles, and arbitrary-shape objects. Examples are given in each

11

type, with the emphasis gradually switching from simple models for states of matter to more complex models for molecular biology. Such a broad scope of science can be covered by fundamental atomic-scale models with the aid of visual and interactive molecular simulations.

References

[1] Q. Xie and R. Tinker, *J. Chem. Ed.*, **83**, 77-83 (2006).

[2] C. Xie, R. Tinker, B. Tinker, A. Pallant, D. Damelin, and B. Berenfeld, *Science,* **332**, 1516-1517 (2011)

[3] A. R. Leach, *Molecular Modeling: Principles and Applications*, 2nd Edition, Pearson Education, 2001.

[4] M. P. Allen and D. J. Tildesley, *Computer Simulation of Liquids*, Oxford, 1987.

[5] J. G. Gay and B. J. Berne, J. Chem. Phys., **74**, 3316 (1981).

[6] M. S. Daw and M. I. Baskes, Phys. Rev. Lett. **50**, 1245 (1983).

[7] D. C. Rapaport, *The Art of Molecular Dynamics Simulation*, Cambridge University Press, 1997.

[8] F. H. Stillinger and T. A. Weber, Phys. Rev. **B31**: 5262–5271 (1985).

[9] T. Tersoff, Phys. Rev. Lett. **56**, 632 (1986).

[10] D. G. Pettifor, Phys. Rev. Lett. **63**, 2480 (1989)

[11] D. J. Cleaver, C. M. Care, M. P. Allen, and M. P. Neal, Phys. Rev. **E54**, 559 (1996).