Protein Simulation Using Fast Volume Preservation

Min Hong[1], David Osguthorpe[2], and Min-Hyung Choi[3]

[1] Division of Computer Science and Engineering, Soonchunhyang University,
646 Eupnae-ri Shinchang-myeon Asan-si, Chungcheongnam-do, 336-745, Korea
Min.Hong@UCHSC.edu

[2] Pharmacology, University of Colorado at Denver and Health Sciences Center,
4200 E. 9th Avenue Campus Box C-245, Denver, CO 80262, USA
David.Osguthorpe@UCHSC.edu

[3] Department of Computer Science and Engineering, University of Colorado at Denver and
Health Sciences Center, Campus Box 109, PO Box 173364, Denver, CO 80217, USA
Min-Hyung.Choi@cudenver.edu

Abstract. Since empirical force fields computation requires a heavy computational cost, the simulation of complex protein structures is a time consuming process for predicting their configuration. To achieve fast but plausible global deformations of protein, we present an efficient and robust global shape based protein dynamics model using an implicit volume preservation method. A triangulated surface of the protein is generated using a marching cube algorithm in pre-processing time. The normal mode analysis based on motion data is used as a reference deformation of protein to estimate the necessary forces for protein movements. Our protein simulator provides a nice test-bed for initial screening of behavioral analysis to simulate various types of protein complexes.

1 Introduction

Since life and the development of all organisms are essentially determined by molecular interactions, the fundamental biological, physical, and chemical understanding of these unsolved detail behaviors of molecules are highly crucial. With the rapid accumulation of 3D (Dimensional) structures of proteins, predicting the motion of protein complexes is becoming of increasing interest. These structural data provide many insights on protein folding, protein-ligand interaction, protein-protein interaction and aid more rational approaches to assist drug development and the treatment of diseases. The analysis of deformation of proteins is essential in establishing structure-function relationships because a structure actually carries out a specific function by movement.

Advanced computer graphics hardware and software can offer real-time, interactive 3D and colorful protein complexes to the screen instead of lifeless chemical formulas. During the last decades, to discover and understand a wide range of biological phenomena, computer simulation experiments using quantum mechanics, molecular mechanics, and molecular dynamics simulation have opened avenues to estimate and predict the molecular level deformation [1, 2, 3]. Several applications such as Amber molecular dynamics [4] and CHARMM [5] have been developed. Quantum mechanics has been widely used for molecular modeling and it calculates the behavior of molecules at the electronic level. Although quantum mechanics provides accurate prediction for molecular simulation, it is limited to small sizes of molecules due to the expensive computational cost. Molecular mechanics calculates the energy of a

V.N. Alexandrov et al. (Eds.): ICCS 2006, Part I, LNCS 3991, pp. 308 – 315, 2006.

molecular system from atoms centered on the nuclear position and an empirical force field. Empirical force fields include bond related forces (bond stretching energy, bond angle bending energy, and torsional energy) and non-bond related forces (electrostatics interaction energy, hydrogen bonding energy, and van der Waals energy) to estimate the energy of the molecular system. Experimental data such as atomic geometry, bond lengths, and bond angles are obtained by X-ray crystal structures or NMR (Nuclear Magnetic Resonance) to set up the values of potential parameters. Molecular mechanics minimizes the given empirical energy functions using initial conditions and finds the minimum energy conformation. However, this method ignores kinetic energy of molecular system.

Although molecular mechanics reduces the computational cost substantially, the computational task is still time-consuming because proteins consist of large numbers of atoms. In addition, the motions of proteins are not static and dynamics are important for protein simulation. In a molecular dynamics simulation, the classical equations of motion for positions, velocities, and accelerations of all atoms are integrated forward in time using the well-known Newton equations of motion. It computes the time dependent deformation of protein structure. However, molecular dynamics simulation also depends on the atomic representation to estimate the evolution of conformations of protein complexes based on the interaction between atoms. Especially both electrostatics interaction forces and van der Waals forces require $O(N^2)$ computational complexity for performing the direct calculation. Thus its computational complexity is expensive for the high number of atoms which are required to predict the deformation of protein structure. The constraint enforcement for bond lengths or bond angles using nonlinear constraint equations at each time step such as SHAKE [6], and RATTLE [7] to eliminate high frequency vibration motions also require heavy computational cost for molecular dynamic simulation.

In this paper, instead of expensive bond lengths and angles constraints or non-bond related forces computations, we provide a new global shape based volume preservation for protein deformation. The complexity of our volume preservation algorithm is at a constant time and it requires virtually no additional computational burden over the conventional mass-spring dynamics model. Our method incorporates the implicit volume preservation constraint enforcement on a mass-spring system to represent protein complexes, so it provides an efficient platform for user manipulation at an interactive response rate. Therefore, the proposed method is well suitable for initial screening of behavioral analysis for complex protein structures.

2 Related Works

The modeling of complex proteins based on the molecular dynamics simulation has been one of the remarkable research areas. The canonical numerical integrations [8] have been used for molecular dynamics simulations using thermodynamic quantities and transport coefficients since they provide enhanced long-time dynamics [9, 10]. Nose-Hoover [11, 12] molecular dynamic is introduced using a constant-temperature and Martyna et al. [13] improved stiff system problems with Nose-Hoover chain dynamics. Jang and Voth [14] introduced simple reversible molecular dynamics algorithms for Nose-Hoover chain dynamics by extension of the Verlet algorithm. Constrained molecular dynamics simulation also has been used to remove numerical stiffness of bond length or bond angle potentials. In molecular dynamics, the length of the integration time step for numerical integration is dictated by the high frequency motions since hard springs can cause numerical instability. The high frequency motions of molecular dynamic simulation are less important than the low frequency motions which correspond to the global motions.

Ryckaert et al. [6] introduced the SHAKE algorithm based on the Verlet method to allow for bond constraints and is widely used for applying the constraints into molecular dynamic simulations. The SHAKE algorithm is a semi-explicit method. Although the SHAKE iteration algorithm is simple and has a low memory requirement, it requires a sufficiently small integration time step to converge to the solution. The adaptive relaxation algorithm for the SHAKE algorithm as presented by Barth et al. [15] iteratively determines the optimal relaxation factor for enhanced convergence. Andersen [7] proposed the velocity level Verlet algorithm, RATTLE, for a velocity level formulation of constraints. The SHAKE and RATTLE require solving a system of nonlinear constraint equations at each integration time step thus they require substantial computational cost. On the other hand, we applied the implicit volume-preserving constraint into protein simulations with our implicit constraint method to achieve the globally meaningful deformation of proteins.

3 Protein Structure

The protein consists of the spatial arrangement of amino acids which are connected to one another through bonds. Thus the protein is a set of atoms connected by bonds in 3D. Instead of using bonded and non-bonded empirical functions, we represent the bond connectivity by simple mass-spring system. Fig. 1 illuminates the structure of periplasmic lysine, arginine, ornithine binding protein (2LAO). The atoms (mass points) are connected by springs which are colored by green lines to propagate the energy. Although we generated springs for each amino acid, each amino acid is still isolated from other amino acids in the protein. Two or more amino acids are linked together by a dehydration synthesis to form a polypeptide. These characteristic chemical bonds are called peptide bonds. Therefore two amino acids are connected together by the carboxyl group and the amine group. Notice that a water molecule is removed during this process.

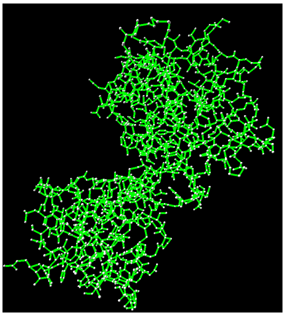

Fig. 1. Simple mass-spring system for periplasmic lysine, arginine, ornithine binding protein (2LAO)

4 Protein Surface

Since proteins interact with one another and other biological compounds through the surface, the protein surface is highly important to understand protein functions and deformations. The concept of a surface for protein involves the geometric properties of molecules, biophysics, and biochemistry. Usually the van der Waals surface is simply computed by overlapping the van der Waals spheres of each atom.

Lee and Richard [16] estimated the solvent accessible surface using a water molecule probe. The solvent accessible surface is determined by a water probe which exactly contacts the van der Waals spheres. Connolly [17] improved the solvent accessible surface using the molecular surface and reentrant surface and has been widely used to represent the surface of proteins. The atoms are idealized as spheres using the van der Waals radius and the Connolly surface is determined by the inward-facing surface of the water probe sphere as it rolls over the protein molecule.

Since fast and accurate computational geometry methods make it possible to compute the topological features of proteins, we applied the marching cube algorithm [18] to determine the surface of the protein. The marching cube algorithm has been widely used in applications of medical imaging to reconstruct 3D volumes of structures which can help medical doctors to understand the human anatomy present in the 2D slices and also applied to the description of the surface of biopolymers [19]. The 3D coordinates of every atom of the protein are obtained from PDB (Protein Data Bank) [20]. Initially, each atom is created using the van der Waals radius. However these standard values of the van der Waals radii create too many spatial cavities and tunnels. Thus we gradually increase the radius of atoms using a threshold until all atoms are overlapped with at least one of the other atoms.

The marching cube algorithm is one of a number of recent algorithms for surface generation with 3D volumetric cube data and it can effectively extract the complex 3D surface of protein. In addition, this algorithm provides the triangle mesh for the surface which is essential for using our volume preservation scheme. This algorithm detects not only the surface of the protein structure but also the cavities or tunnels of the protein. We can readily expand the basic principle of the marching cube algorithm to 3D.

We adopted the marching cube algorithm to generate the complex surface of proteins as a pre-processing stage. The high resolution of the triangulated mesh surface requires more computational cost and the resolution of the surface can be controlled using a threshold value which defines the size of cubes. The Fig. 2 shows the created refined surface of periplasmic lysine, arginine, ornithine binding protein (2LAO) and Adenylate kinase (1AKE). The surface nodes which create the surface of the protein are independently created from atoms. These surface nodes are connected by structure, bending, and sheer springs to preserve the physical properties. To represent the correct surface according to the deformation of the protein, the surface should be re-estimated by the marching cube algorithm at each time step, but it is a computationally expensive and painful task.

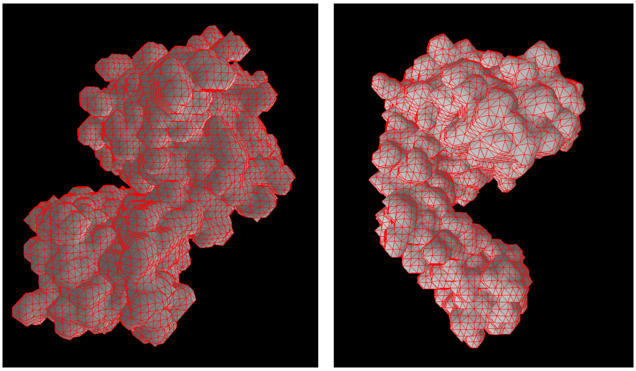

Fig. 2. The surface of proteins using marching cube algorithm: the left figure is periplasmic lysine, arginine, ornithine binding protein (2LAO) and the right figure is Adenylate kinase (1AKE)

5 Protein Volume

Volume changes of proteins on protein folding have been studied [21, 22] and their results show that volume changes are very small (less than 0.5 percentage from original volume) at normal pressure. To overcome the critical inherent drawback of volume loss of a mass-spring system, this paper proposes a real-time volume preservation method. Our volume preservation method maintains the global volume of a closed mesh structure and guarantees the constant volume constraint at every time step of a mass-spring simulation.

The total volume of a protein is estimated by summing the surface triangles of the protein using the Divergence Theorem. The surface of the protein is represented by flat triangular patches with coordinates (x, y, z) varying linearly on these patches. It is convenient to introduce the natural coordinates L_1 , L_2 , and L_3 and express the surface integral as

$$3\mathrm{V}=\int \sum_{\mathrm{i}=1}^{3} \left(x_i L_i N_x + y_i L_i N_y + z_i L_i N_z \right) dA \tag{1}$$

Note that the unit normal vector is constant on the triangular surface patch. The integral is easily evaluated using the following equation for integrating polynomials in L_i

$$\int L_1^{a_1} L_2^{a_2} L_3^{a_3} \, dxdy = 2A \frac{a_1 ! a_2 ! a_3 !}{(a_1 + a_2 + a_3 + 2)!} \tag{2}$$

where a_1 , a_2 , and a_3 are non-negative integers, and A is the area of a triangle. We have the three cases: a_1 , a_2 , and a_3 are $a_1 = 1$ $a_2 = a_3 = 0$, $a_2 = 1$ $a_1 = a_3 = 0$, and $a_3 = 1$ $a_1 = a_2 = 0$. The three integrals we need are given by

$$\int L_1 dxdy = \int L_2 dxdy = \int L_3 dxdy = \frac{A}{3} \tag{3}$$

The total volume V can be obtained by

$$V = \sum_i \frac{A}{3}\left\{n_x(x_1 + x_2 + x_3) + n_y(y_1 + y_2 + y_3) + n_z(z_1 + z_2 + z_3)\right\} \tag{4}$$

where i is the volume contribution of surface triangle i. This volume must remain a constant over the entire simulation, so we cast this condition as a constraint in a dynamic system.

Let $\Phi(q,t)$ be the constraint to representing an algebraic constraint. To preserve the volume of object, the difference between V_0 (original volume) and V (current volume) should be 0 during the simulation.

$$\Phi(q,t) = V_0 - V = 0 \tag{5}$$

We applied the implicit constraint method [23] to maintain the volume constraint.

$$\Phi_q(q,t)M^{-1}\Phi_q^T(q,t)\lambda = \frac{1}{\Delta t^2}\Phi(q,t) + \Phi_q(q,t)(\frac{1}{\Delta t}\dot{q}(t) + M^{-1}F^A(q,t)) \tag{6}$$

where F^A are applied, gravitational and spring forces acting on the discrete masses, M is a diagonal matrix containing discrete nodal masses, λ is a vector containing the Lagrange multipliers and $\Phi_q = \frac{\partial V}{\partial q}$ is the Jacobian matrix. In equation (6), λ can be calculated by a simple division thus our volume preservation uses the implicit constraint method to preserve the protein volume with a mass-spring system at virtually no extra.

6 Force Generation for Protein Movement

Recently NMA (Normal Mode Analysis) [24, 25, 26] has been widely used to infer the dynamic motions of various types of proteins from an equilibrium conformation. NMA is a powerful theoretical method for estimating the possible motions of a given protein. To analyze the necessary external forces for protein movement, we used the Yale Morph Server [26] which applied NMA to achieve the movements of proteins as a reference deformation of the protein. We calculated the necessary forces based on the series of motion data files which are achieved from [26].

$$v_{n+1} = v_n + dt(F/m), \quad q_{n+1} = q_n + dtv_{n+1} \tag{7}$$

Equation (7) is the simple Euler integration to estimate the next status of positions and velocities. Here v is the velocity of atom, q is position of atom, F is the net force of system, m is the mass of atom, and dt is the integration time step. We already know the all position information of atoms for each time step, thus we can calculate the necessary force to move all atoms using equation (8).

$$v_{n+1} = (q_{n+1} - q_n)/dt, \quad F = m(v_{n+1} - v_n)/dt \tag{8}$$

Since the atoms of the protein and surface nodes are not connected to each other, when atoms are moved by estimated external forces in equation (8), the surface of the protein will not move. Instead of creating extra springs to follow the movement of atoms or regenerating the surface using the expensive marching cube algorithm at each time step, the surface nodes of the protein are moved according to the displacement of the closest atom to reduce computational requirements. Unlike the Yale Morph Server which only provides 2D motion of proteins with a fixed view point, our protein simulation provides 3D motion of protein deformation with a controllable view point.

7 Conclusion

Although we are using the current state-of-the-art computing power, it is still not computationally feasible to perform atomic molecular dynamics simulation for huge protein structures at an interactive response rate, partly due to the large conformation space. Instead of expensive and complicated force calculations at each time step to achieve the detail deformation of proteins, we applied the simple and fast mass-spring system. The stable and effective global shape based volume preservation constraint is applied to the protein simulation with the marching cube algorithm. Although our new global shape based volume preservation method sacrifices the micro level of dynamics in the protein movement, the proposed efficient platform of protein simulator can provide the globally correct (low-frequency motion) and biologically meaningful deformation of the protein complexes at an interactive level. Therefore, our simulator can be utilized as a nice test-bed for initial screening of behavioral analysis of protein complexes.

References

1. T. Schlick, Molecular modeling and simulation, Springer, New York, 2002.
2. A. R. Leach, Molecular modeling principles and applications, Addison Wesley Longman Limited, 1996.
3. L. Pauling, and E. B. Wilson, Introduction to quantum mechanics with applications to chemistry, Dover, New York, 1985.
4. Amber Molecular Dynamics, http://amber.scripps.edu/
5. CHARMM Development Project, http://www.charmm.org/
6. J. P. Ryckaert, G. Ciccotti, and H. J. C. Berendsen, Numerical integration of the Cartesian equations of motion of a system with constraints: Molecular dynamics of n-alkanes, Journal of Computational Physics, 23, pp. 327-341, 1977.
7. H. C. Andersen, Rattle: A velocity version of the SHAKE algorithm for molecular dynamics calculations, Journal of Computational Physics, 52, pp. 24-34, 1983.
8. D. Okunbor and R. D. Skeel, Canonical numerical methods for molecular dynamics simulations, Journal of Computational Chemistry, 15(1), pp. 72-79, 1994.
9. P. J. Channell, and J. C. Scovel, Symplectic integration of hamiltonian systems, Nonlinearity, 3, pp. 231-259, 1990.
10. D. Okunbor, Canonical methods for hamiltonian systems: Numerical experiments, Physics D, 60, pp. 314-322, 1992.
11. S. Nose, A molecular dynamics method for simulations in the canonical ensemble, Molecular Physics, 52(2), pp. 255-268, 1984.

12. W. G. Hoover, Canonical dynamics: Equilibrium phase-space distributions, Physical Review A, 31(3), pp. 1695-1697, 1985.
13. G. J. Martyna, M. L. Klein, and M. Tuckerman, Nosé–Hoover chains: The canonical ensemble via continuous dynamics, Journal of Chemical Physics, 97(4), pp. 2635-2643, 1992.
14. S. Jang, and G. A. Voth, Simple reversible molecular dynamics algorithms for Nosé–Hoover chain dynamics, Journal of Chemical Physics, 107(22), pp. 9514-9526, 1997.
15. E. Barth, K. Kuczera, B. Leimkuhler, and R. D. Skeel, Algorithms for constrained molecular dynamics, Journal of Computational Chemistry, 16, pp. 1192-1209, 1995.
16. B. Lee and F. M. Richards, The interpolation of protein structures: Estimation of static accessibility, Journal of Molecular Biology, 55, pp. 379-400, 1971.
17. M. L. Connolly, Solvent-accessible surfaces of proteins and nucleic acids, Science, pp. 709-713, 1983.
18. W. E. Lorensen, and H. E. Cline, Marching Cubes: A high resolution 3D surface construction algorithm, Proceedings of SIGGRAPH 1987, ACM Press / ACM SIGGRAPH, Computer Graphics Proceeding, 21(4), pp. 163-169, 1987.
19. A. H. Juffer, and H. J. Vogel, A flexible triangulation method to describe the solvent-accessible surface of biopolymers, Journal of Computer-Aided Molecular Design, 12(3), pp. 289-299, 1998.
20. The RCSB Protein Data Bank, http://www.rcsb.org/pdb/
21. P. E. Smith, Protein volume changes on cosolvent denaturation, Biophysical Chemistry, 113, pp. 299-302, 2005.
22. M. Gerstein, J. Tsai, and M. Levitt, The volume of atoms on the protein surface: Calculated from simulation, using voronoi polyhedra, Journal of Molecular Biology, 249, pp. 955-966, 1995.
23. M. Hong, M. Choi, S. Jung, S. Welch, and Trapp, J. Effective constrained dynamic simulation using implicit constraint enforcement. In IEEE International Conference on Robotics and Automation, 2005.
24. P. Dauber-Osguthorpe, D.J. Osguthorpe, P.S. Stern, J. Moult, Low-frequency motion in proteins - Comparison of normal mode and molecular dynamics of streptomyces griseus protease A, J. Comp. Phys., 151, pp. 169-189, 1999.
25. A. D. Schuyler, and G. S. Chirikjian, Normal mode analysis of proteins: A comparison of rigid cluster modes with Ca coarse graining, Journal of Molecular Graphics and Modelling, 22(3), pp. 183-193, 2004.
26. Yale Morph Server, http://molmovdb.mbb.yale.edu/