Proceedings of the
46th IEEE Conference on Decision and Control
New Orleans, LA, USA, Dec. 12-14, 2007

Towards a Mechanistic View of Protein Motion

Filip Jagodzinski and Oliver Brock

Abstract— **Proteins are the fundamental building blocks of all biological systems. To perform their function, proteins generally undergo self-motions that result in changes in their three-dimensional shape. In order to understand the function of a protein and thus to be able to infer how to therapeutically regulate its function, it is necessary to have detailed knowledge of the feasible self-motions of the protein. Such knowledge cannot be obtained by existing experimental methods. In this paper, we present preliminary evidence that accurate and computationally efficient simulation of the self-motions of a protein may be achieved by partitioning the simulation based on the type of self-motions. In support of this view, we present a method and accompanying simulation results that the large-scale motions of a protein can be simulated based entirely on kinematic considerations. The proposed method leverages insights from kinematics and operational space control from robotics. We believe the proposed method to be a first step towards a general, accurate, and efficient method for the simulation of protein motion.**

I. INTRODUCTION

Proteins perform the majority of cellular functions in all living organisms. They do so by binding to other proteins or molecules—or by breaking such a binding. For two molecules to bind to each other, they have to exhibit geometric and physicochemical complementarity at the interaction surface. To achieve this level of complementarity, most proteins change their shape during the binding process. Establishing and breaking bindings among proteins and other biomolecules is responsible for almost all cellular functions, including metabolism, transport of nutrients, synthesis of proteins, signaling, and gene regulation.

Our research is concerned with the internal motions performed by proteins and their bindings with other biomolecules. Fig. 1 shows an example of a protein that undergoes self-motions to perform its function. Because protein motion is so fundamental to many cellular processes, an accurate understanding of protein motion would bring about tremendous scientific advances: It would facilitate the design of drugs to cure or treat many diseases, it would shed light onto some of the cellular processes that remain poorly understood, and it would provide insights into the mechanisms that underlie diseases such as Alzheimer's, Mad Cow disease, and Creutzfeldt-Jakob disease.

Experimental techniques available today cannot observe the motion of proteins directly—they are only able to measure secondary phenomena that arise as a consequence of the motion. For example, it is possible to determine the structure of a protein at two different stages of a motion, but such measurements do not reveal information about the motion that occurs between those two states. Fig. 1 shows two of the known states of HIV-1 protease.

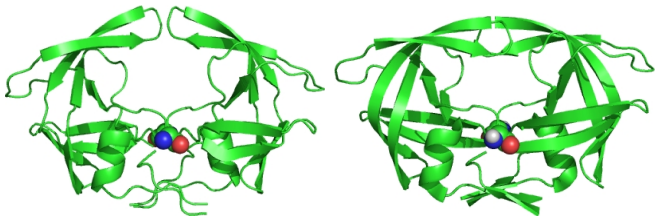

Fig. 1. The protein HIV-1 protease plays a crucial rule in the life cycle of HIV: it cleaves polyproteins into functional subunits that are required for viral assembly. To perform this function, the two flaps of HIV-1 protease (at the top of the protein) fold down from an open position (left) onto the central cavity, resulting in a closed position (right). In order to develop drugs against AIDS, it is necessary to understand the motions of HIV-1 protease.

Computational methods are able to simulate the physical and chemical phenomena that give rise to protein motion. In principle, such methods could simulate the entire motion of proteins, effectively overcoming the limitations of current experimental methods. However, the accuracy of simulated protein motion depends on the fidelity with which the physical and chemical forces are simulated. In Section III, we use the accuracy of simulation to categorize existing computational methods for simulating protein motion. While these existing computational methods can address specific biological questions about protein motion, they do not provide sufficient accuracy and computational efficiency to serve as a general simulation tool for the motion of proteins. In our research, we attempt to overcome this limitation.

The premise of our research is that the motion of proteins at different scales is brought about by different physical and chemical phenomena. To efficiently simulate the overall motion of a protein, it is necessary to partition the simulation of motion according to the motion's scale and thus the relevant phenomena. For example, we believe that large-scale motions mainly depend on a different set of phenomena than detailed, small-scale motions. If this assumption holds, it would be possible to render the simulation of protein motion both accurate and efficient by simulating only the necessary phenomena for different parts of the protein. We postulate that such an approach can lead to a general computational method for the simulation of protein motion.

In this paper, we present initial evidence in support of our general premise. We hypothesize that protein large-scale motions to a certain extent are encoded in the kinematic

This work is partially supported by NSF grants CCF 0622115, CNS 0551500 and by NIH grant NIGMS 1R01GM076706.

F. Jagodzinski and O. Brock, Robotics and Biology Laboratory, Computational Biology Laboratory, Department of Computer Science, University of Massachusetts Amherst, 140 Governors Drive, Amherst, MA 01003-9264

1-4244-1498-9/07/$25.00 ©2007 IEEE.

structure of the protein, established by the covalent (stable and relatively rigid) bonds. To validate this view, we examine the large-scale motions in the flaps of HIV-1 protease (see Fig. 1). We employ a purely kinematic representation of a protein and use principles from operational space control in robotics. Based on this computationally efficient method, we show that purely kinematic considerations are sufficient to simulate the motions of the flaps. We view these preliminary results as evidence in support of our general premise.

II. BIOLOGICAL BACKGROUND

Proteins are molecular chains, polymerized from elementary building blocks called amino acids. Each amino acid consists of two parts. One part, which is identical for all amino acids, forms the covalently connected chain. This chain is shown in Fig. 2. Collectively, the bars and atoms that constitute the core of the chain are referred to as the backbone of the protein. Molecular branches attached to the backbone are called side chains. The atomic structure of side chains differ among the 20 naturally occurring amino acids. Side chains are represented as octagons in Fig. 2. The planes shown in the figure indicate rigid peptide units. The arrows indicate the degrees of freedom available to the backbone: two adjacent peptide units can rotate about the two covalent bonds with the carbon atom. Our mechanistic perspective of a protein is based on the rigid peptide units and the rotatable joints between them.

Every protein has a distinctive sequence of amino acids along the backbone. This sequence is encoded genetically and is referred to as the primary structure of the protein. The flexible chain of amino acids is subjected to a variety of physical and chemical interactions among the atoms of a protein and between the protein and its environment. These interactions cause the protein to move until it assumes its unique spatial structure. This structure, called tertiary structure, is believed to be entirely determined by the primary structure. The smaller, regular structures that the protein chain assumes and retains after the folding process are called secondary structures, and these include helices and sheets.

So far we have described proteins as branching kinematic chains: the backbone of a protein forms the main kinematic chain and the side chains represent the side-branches. The motion of such a branching kinematic chain can be represented very easily since all joints are holonomic.

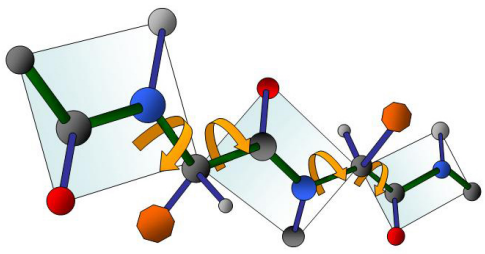

Fig. 2. Schematic illustration of a protein. Individual atoms of a backbone are organized into amino acids. Certain covalent bonds in the backbone can rotate, and so the amino acid chain can assume a variety of conformations.

Unfortunately, a protein is more complex. Up to this point, we have ignored some very important chemical interactions that occur among side chains. The most important of these interactions in the context of this paper is called a hydrogen bond. These bonds occur between two side chains of amino acids that could be anywhere along the backbone. Hydrogen bonds are weaker than covalent bonds but strong enough to play an important role in maintaining the overall structure of the protein. More importantly, they constrain the motion of the protein by forming many closed kinematic chains, called loops or rings [1]. It is the presence of such loops and the difficulties associated with generating their motion that has plagued many existing computational approaches and their simulation of protein motion.

III. RELATED WORK

During the motion of a closed kinematic chain, the mechanical constraints imposed by the connected loop must remain satisfied at all times. Any configuration that fulfills this requirement lies on a lower-dimensional manifold embedded in the overall configuration space of the chain. For complex kinematic chains, this manifold is difficult to describe analytically. Several researchers in robotics have devised methods that allow to sample the self-motion manifold of closed kinematic chains in the context of motion planning [2], [3], [4], [5]. Some of these ideas have been applied to problems in molecular biology, such as the loop closure problem [6] or determining large-amplitude motions of flexible molecules [7]. In contrast to these methods, our approach does not rely on a sampling of the self-motion manifold and thus is computationally more efficient.

Molecular biologists have developed a variety of computational tools to simulate molecular motion. Roughly, they fall into three categories. Molecular dynamics [8] uses detailed knowledge about physics and chemistry to perform a highly accurate, yet computationally demanding simulation. Due to the computational complexity of this method, the duration of the protein motion that can be simulated is limited to hundreds of nanoseconds. However, biologically relevant motion lasts on the order of milliseconds and longer. The simulation of such temporally extended motion is thus beyond the capabilities of molecular dynamics, even on the largest supercomputers available today.

A second group of methods, graphics-based methods [9], simply interpolate the atomic coordinates between two known states of the protein. These methods are computationally efficient, but the resulting motion may have no biological accuracy.

A third category of computational methods simplifies the model of a protein to render simulation tractable. Methods in this category represent the protein as a set of masses that are connected by springs [10], [11]. Such a physical system can be simulated efficiently. A normal mode analysis [12] or principle component analysis [13] of the simulated system can reveal characteristics of the motion. For some types of motion, the dominant normal modes correspond to biologically plausible motion. However, many biologically

important motion phenomena cannot be simulated using the simplified mass/spring model.

None of the currently available computational methods can simulate protein motion accurately. They are either too computationally complex or not sufficiently accurate.

We hypothesize that certain aspects of protein motion can be simulated accurately by considering only some of the involved physicochemical phenomena. Our hypothesis is supported by related computational and experimental work. For example, a computational method is able to determine regions of a protein that have been rigidified by covalent bonds and hydrogen bonds. To identify rigid parts, it suffices to examine a graph in which covalent bonds and hydrogen bonds correspond to edges and atoms correspond to vertices [14]. It is thus possible to understand a protein's ability to move by considering only a small subset of all physicochemical phenomena.

Multi-scale techniques for the simulation of molecular systems successfully combine coarse-grained and fine-grained molecular models [15]. These methods simulate molecular systems more efficiently by reducing the simulation accuracy for parts of the molecules, thus demonstrating the benefit of spatially varying simulation accuracy.

Computational methods from the field of computer graphics exclude degrees of freedom from the simulation of motion of branching mechanical systems, when this exclusion does not significantly affect the overall accuracy [16]. This demonstrates that it is possible to achieve accurate motion simulation by only considering a subset of all phenomena; here, entire parts of the protein are excluded from the simulation.

There is also biological evidence for our hypothesis. The experimental study of protein motion during binding interactions has shown that the binding motion is performed by the protein irrespective of the presence of the binding partner [17]. The motion must thus be encoded—at least to some extent—in the protein itself. The binding motion may thus be identifiable without considering the interactions with the binding partner.

IV. MECHANISTIC VIEW OF PROTEIN MOTION

Molecular dynamics computations are considered accurate for the simulation of molecular motion. To achieve such accuracy, they consider a large number of inter-atomic forces and physicochemical effects, such as the stretching, bending, and torsion of covalent bonds, van der Waals forces, electrostatic interactions, as well as the dynamic properties of individual atoms, and sometimes even effects of quantum mechanics. Molecular dynamics methods numerically integrate Newton's law of motion for all atoms based on these forces, effectively simulating our understanding of physics at the atomic level.

The complexity of molecular dynamics stands in stark contrast with the simplicity of graph-theoretic algorithms for the analysis of molecular flexibility [14]. These methods, by merely considering connectivity information from covalent and hydrogen bonds, are able to predict the flexibility of proteins with biological accuracy [14]. They do so without even considering the spatial location of individual atoms, simply by considering a graph in which vertices correspond to atoms and edges correspond to bonds between the atoms. However, such graph-theoretic analysis does not give any insight into the actual motion of the protein, just into the ability of parts of the protein to move.

The biological accuracy of these graph-theoretic algorithms suggests that the connectivity among atoms encodes important information about the motion behavior of the protein. This motivates us to explore if a more expressive representation of the protein can be used to simulate protein motion. We use a kinematic (or mechanistic) representation that in addition to the connectivity information represented in the aforementioned graph also contains information about the spatial placement, bond angles and lengths between atoms. Such representations have been studied extensively in the field of robotics [18].

Based on the mechanistic view of a protein shown in Fig. 2, we represent the protein as a mechanical linkage or robot. The assumption that covalent bonds can be modeled as rigid links is common in the field of molecular modeling. We now can describe a particular protein structure—or conformation—by specifying a value for every joint angle in our kinematic model. The joint angles, together with the parameters of the kinematic model, uniquely describe the conformation. The space of all possible conformations is called conformation space. Its dimensionality is given by the degrees of freedom present in the mechanism.

The structure of a protein, however, is not exclusively defined by the covalent bonds that give rise to the protein's branching structure. Side chains along the backbone form additional bonds, such as hydrogen bonds or salt bridges [19]. Generally, these bonds remain intact during the lifetime of a protein and constrain the protein's motion. Given the conformation space C of a protein, all conformations that satisfy these additional constraints lie in a lower-dimensional manifold M embedded in C (see Fig. 3).

Our kinematic method for the simulation of protein motions is based on the following simple idea: Instead of explicitly computing the self-motion manifold M, which in

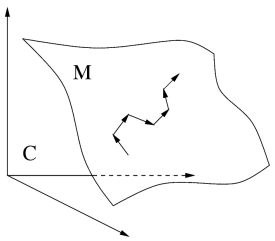

Fig. 3. The conformation space, C, of a branching kinematic chain is the set of coordinates that encompasses all of the possible assignments of joint angles. The self-motion manifold, M, of a branching kinematic chain with closed loops is the set of all configurations of the kinematic chain for which the loop closure constraints are satisfied. The self-motion manifold M is embedded in C. When a kinematic chain with closed loops performs a motion, it traverses the manifold M, as indicated in the figure. When a protein performs a motion, it remains within the self-motion manifold M.

practice is very difficult, we employ methods from operational space control [18], [20] to iteratively explore it in a numerical fashion. We guide this exploration using the biological interactions that cause the protein's motion. A search within the manifold M for the motion is thus replaced with a simple gradient descent in the energy landscape defined by the interaction of two proteins. As we will see, this insight turns the characterization of the self-motion manifold from a computationally very complex task into one with a negligible computational cost. In the remainder of this section, we will provide the details of this method and present its advantages.

The operational space equations of motion [18] of a serial chain manipulator are given by:

$$\Lambda(\mathbf{x})\ddot{\mathbf{x}} + \mu(\mathbf{x}, \dot{\mathbf{x}}) + \mathbf{p}(\mathbf{x}) = \mathbf{F},$$

where $\mathbf{x}$ is the vector of the $6m$ operational coordinates describing the position and orientation of the m effectors, $\Lambda(\mathbf{x})$ is the $6m \times 6m$ kinetic energy matrix associated with the operational space. $\mu(\mathbf{x}, \dot{\mathbf{x}})$, $\mathbf{p}(\mathbf{x})$, and $\mathbf{F}$ are respectively the centrifugal and Coriolis force vector, gravity force vector, and generalized force vector acting in operational space.

After this dynamic decoupling, the motion of a specific operational point on a kinematic chain robot (generally its end-effector) is controlled directly, rather than by explicitly specifying the motion at each of its joints. The behavior of the operational point can be controlled by specifying a force vector $\mathbf{F}$ acting on it. The force "pulls" the operational point through space. The joint torque corresponding to $\mathbf{F}$ is given by

$$\tau = J^T(\mathbf{q})\mathbf{F}, \tag{1}$$

where J is the Jacobian of the kinematic chain for the operational point at which the force is acting and $\mathbf{q}$ represents the configuration of the chain (or conformation of the protein).

We propose to derive the force $\mathbf{F}$ from the biological interaction among molecules. Equation 1 then lets us determine how the branching kinematic structure of the protein responds to the interaction force. Since the Jacobian J of even very complex branching structures can be computed in $O(n)$ for n degrees of freedom, the motion resulting from the interaction can be computed very efficiently. If the interaction consists of multiple forces, which can be assumed to be the case for proteins, we can generalize Equation 1 to:

$$\tau = \sum_p J_p^T(\mathbf{q})\mathbf{F}_p, \tag{2}$$

so that now the total torque applied to the joints of the branching structure is given by the sum of the torques resulting from forces acting at all points p. Unfortunately, this simple method of generating protein motion based on forces and operational space control does not extend to kinematic structures containing closed loops.

The key insight exploited by our method is that a closed kinematic chain can be decomposed into two open chain branches and a constraint imposed on the ends of the branches, namely that they have to remain in physical contact. This insight allows us to extend the applicability of Equation 2 to the case of kinematic chains with loops—and thus to proteins. In proteins, closed loops are introduced by bonds that form between the side chains of different amino acids. We model these bonds as springs, rather than as kinematic linkages, thus avoiding closed loops in the kinematic structure. The springs exert forces $\mathbf{F}_p$ on the kinematic structure to enforce the same kinematic constraint the link would have enforced. Doing so, we eliminate kinematic loops from our protein representation, and we are able to apply Equation 2 to achieve a computationally efficient and entirely mechanistic simulation of protein motion.

The proposed mechanistic method to simulate protein motion has several important properties. First and foremost, it avoids search in conformation space to determine the self-motion manifold. To our knowledge, this sets our method apart from all existing computational methods that simulate the motion of structures with closed loops. The resulting computational gains are tremendous. Our method is able to simulate motion on the self-motion manifold at interactive rates for proteins of any size. The gain in computational efficiency stems from the fact that the operational space control methods project forces from the physical space into the conformation space of the protein. Based on this projection, we are able to replace search with iterative, numerical constraint satisfaction.

Another advantage of our method is that the exploration of the self-motion manifold is guided by the biologically occurring forces between interacting molecules. This means that we are not only avoiding search in the conformation space, but also in the self-motion manifold itself, yielding further performance improvements.

A third important property of our method is that it can easily be extended to include additional forces. While in this paper we only consider the forces that arise from hydrogen bonds, Equation 2 can accommodate arbitrary forces, as long as they are expressed in the physical space of the protein. This will allow us to scale our method to include additional physicochemical phenomena that affect the motions of the protein without unduly deteriorating the performance.

V. EXPERIMENTAL VALIDATION

Using the computational method described in the previous section, we demonstrate that a mechanistic view of protein motion is sufficient for the simulation of certain large-scale protein motions. To determine biological accuracy, we compare the results of our simulations with experimental data of the motion of HIV-1 protease (see Fig. 1). HIV-1 protease possesses two extruding finger-like regions, also referred to as flaps. These flaps fold in towards the interior of the protein to facilitate an enzymatic reaction. The spatial structures of the open and closed forms are known from X-ray crystallography experiments [21].

We simulated the motions of HIV-1 protease, starting with an open kinematic representation and finishing with a predicted closed form. Our kinematic model of the known open conformation of HIV-1 protease is shown in Fig. 4.

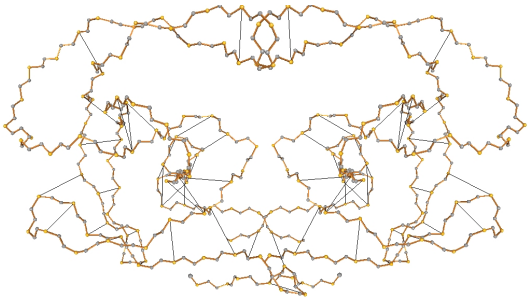

Fig. 4. The kinematic representation of HIV-1 protease.

When cyclic urea, a ligand, is placed in the central cavity of HIV-1 protease, two interactions occur, causing the flaps and the core region of the protease to bind with the ligand [22]. We take advantage of the knowledge of these interactions to compute the two forces that the flap and core of the protein would be subjected to as they transition to their known closed orientations. Our simulation's initial conformation with the flap in the open conformation, the ligand in the cleft, and the corresponding interaction forces are shown in Fig. 5.

If we are able to obtain a structure that is similar to the closed form of HIV-1 protease, it would be strong preliminary evidence that the feasible motions of the flaps of HIV-1 protease is governed predominantly by the kinematic constraints imposed by covalent and hydrogen bonds.

To assess the accuracy of our simulations, we measured the difference between the experimentally determined closed conformation and our predicted conformation. The difference between two molecular structures is defined as the root mean square deviation (RMSD) between the spatial coordinates of corresponding atoms in both structures, measured in Ångstrom (10^{-10} meter). To calculate the accuracy of the motion of the flaps, we measured the RMSD between the flaps of the known closed form and those obtained from our simulation. We also evaluate the impact of maintaining the constraints imposed by hydrogen bonds.

It should be noted that for our simulation results, RMSD must be interpreted differently than when comparing two experimentally determined protein structures. In our simulation, molecules are modified incrementally in response to two binding interactions. Thus, we interpret decreasing RMSD values as an indication that one structure is approaching a different structure during this incremental modification. Given the size of the simulated protein and the fact that we consider only covalent and hydrogen bonds, it should be surprising that the two binding interactions do not completely distort the protein, resulting in a vastly increased RMSD.

TABLE I

RMSD CALCULATIONS COMPARING THE KNOWN CLOSED FORM OF HIV-1 PROTEASE WITH OUR PREDICTED CONFORMATION.

Simulation Parameters	Backbone RMSD Å	Flap only RMSD Å
Ligand interactions & 37 hydrogen bonds	0.695	0.945
Ligand interactions without hydrogen bonds	1.346	1.243
Open-form, closed-form comparison	1.083	2.657

Table I shows the RMSD values between the closed conformation predicted by our simulations and the experimentally known closed form of HIV-1 protease; the bottom row gives the RMSD values between the known open and closed conformations of the protease. In the simulations, the flaps assumed a final orientation that is very similar to the conformation of the flaps in the known closed form, as evidenced by the resulting flap RMSD of 0.945Å, down from an initial value of 2.65Å. The backbone of the protein also assumed a shape similar to that of the known closed form. Our results indicate that the simulation of hydrogen bonds contributes to the biological accuracy of our simulation.

We propose several reasons for these residual structural differences. Most importantly, the interaction of the flaps with hydrophobic pockets near the core of HIV-1 protease has been shown to cause the flaps to twist and turn [23], a phenomena that we did not model in our simulation. A visual inspection of Fig. 6 shows that this change in handedness of several amino acids at the tip of the flap contributes significantly to the structural differences observed in our simulation. If we were able to simulate this change in handedness locally, we suspect that the biological accuracy of our simulation would increase—without the requirement of performing full-scale MD simulations for the entire protein.

The biological binding between HIV-1 protease and the

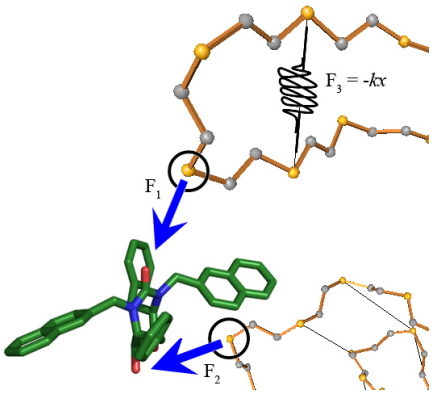

Fig. 5. We model the backbone of HIV-1 protease as a kinematic structure, and we apply forces $\mathbf{F_1}$ and $\mathbf{F_2}$ on the locations of this structure which are known to interact with a ligand. We represent hydrogen bonds as springs, which impart a restoring force $\mathbf{F_3}$ onto the protein.

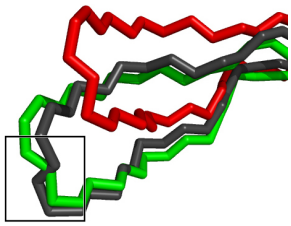

Fig. 6. Schematic showing the open (red), the known closed (gray), and our simulation's predicted (green) closed orientation of the flaps. The square designates the change in handedness of the flap tip that occurs beetween the open and closed conformations.

ligand entails many additional small interaction forces in addition to the two forces that we simulated. This may also account for some of the structural differences between the known closed form and our prediction. Finally, there are physicochemical phenomena that were ignored in our simulation; these phenomena include steric clashes between side chains, which further constrain the motion of a protein.

By considering covalent and hydrogen bonds only, we were able to accurately and efficiently simulate the flap motions in HIV-1 protease. While the accuracy of this simulation is clearly not adequate yet for the evaluation of docking interactions, it permits biological insight into large-scale motion of proteins. We view this as a very encouraging preliminary result in support of our broader hypothesis that general, accurate, and efficient simulation of protein motion can be achieved by simulating only the necessary phenomena for different types of motion and different parts of a protein. In the case of the large-scale motions that occur in HIV-1 protease, we only considered covalent and hydrogen bonds, yet our method was extremely computationally efficient while achieving sufficient accuracy for biological insight.

VI. CONCLUSION

Binding interactions among proteins govern almost all cellular processes in living organisms. A detailed understanding of these binding interactions would enable tremendous advances in cellular biology, and it would facilitate finding cures for many diseases. To obtain an understanding of protein binding, it is necessary to know the self-motions a protein can perform while interacting with another molecule. Unfortunately, existing experimental and computational methods are not able to provide sufficient insight into a protein's self-motions. We have proposed the view that general, accurate, and computationally efficient simulation of protein motion might be achievable when the granularity necessary to capture the distinctive properties of protein motion is matched with the granularity of the simulated physical and chemical phenomena. In support of this view, we presented a control-based, highly computationally efficient method for the kinematic simulation of protein motion. The proposed method overcomes important limitations of existing computational approaches. Furthermore, we demonstrated that the proposed method can simulate large-scale protein motion with sufficient accuracy to provide insight into the biological function of a protein. We believe that this work can be extended into a general, accurate, and computationally efficient approach for the simulation of protein motion.

REFERENCES

[1] N. Gō and H. A. Sheraga, "Ring closure and local conformational deformations of chain molecules," *Macromolecules*, vol. 3, no. 2, pp. 178–197, March-April 1970.

[2] J. Cortés and T. Siméon, "Sampling-based motion planning under kinematic loop-closure constraints," in *Proceedings of the Workshop on the Algorithmic Foundations of Robotics (WAFR)*, Utrecht/Zeist, The Netherlands, 2004.

[3] L. Han, "Hybrid probabilistic roadmap-Monte Carlo motion planning for closed chain systems with spherical joints," in *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, New Orleans, USA, 2004.

[4] S. M. LaValle, J. H. Yakey, and L. E. Kavraki, "A probabilistic roadmap approach for sytems with closed kinematic chains," in *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, Detroit, USA, 1999, pp. 1671–1675.

[5] D. Xie and N. M. Amato, "A kinematics-based probabilistic roadmap method for high DOF closed chain system," in *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, New Orleans, USA, April 2004, pp. 473–478.

[6] A. A. Canutesco and R. L. Dunbrack Jr., "Cyclic coordinate descent: A robotics algoritm for protein loop closure," *Protein Science*, vol. 12, no. 5, pp. 963–972, 2003.

[7] J. Cortés, T. Siméon, V. Ruiz de Angulo, D. Guieysse, M. Remaud-Siméon, and V. Tran, "A path planning approach for computing large-amplitude motions of flexible molecules," in *Proceedings of the International Conference on Intelligent Systems for Molecular Biology (ISMB)*, Detroit, USA, 2005.

[8] A. Jain, N. Vaidehi, and G. Rodriguez, "A fast recursive algorithm for molecular dynamics," *Journal of Computational Physics*, vol. 106, pp. 258–268, 1993.

[9] S. Flores, N. Echols, D. Milburn, B. Hespenheide, K. Keating, J. Lu, S. Wells, E. Yu, M. Thorpe, and M. Gerstein, "The database of macromolecular motions: new features added at the decade mark," *Nucleic Acids Res*, vol. 34, pp. D296–301, 2006.

[10] A. R. Atilgan, S. R. Durrell, R. L. Jernigan, M. C. Demirel, O. Keskin, and I. Bahar, "Anisotropy of fluctuation dynamics of proteins with an elastic network model," *Biophysical Journal*, vol. 80, pp. 505–515, 2001.

[11] M. K. Kim and G. S. Chirikjian, "Modeling macromolecular machines using rigid-cluster networks," in *Proceedings of the Workshop on the Algorithmic Foundations of Robotics (WAFR)*, Utrecht/Zeist, The Netherlands, 2004.

[12] A. J. Rader and I. Bahar, "Folding core predictions from network models of proteins," *Polymer*, vol. 45, pp. 659–668, 2004.

[13] M. L. Teodoro, G. N. Phillips Jr., and L. E. Kavraki, "A dimensionality reduction approach to modeling protein flexibility," in *Proceedings of the International Conference on Research in Computational Molecular Biology (RECOMB)*, 2002, pp. 299–308.

[14] D. J. Jacobs, A. Rader, M. F. Thorpe, and L. A. Kuhn, "Protein flexibilty predictions using graph theory," *Proteins: Structure, Function, and Bioinformatics*, vol. 44, pp. 150–165, 2001.

[15] A. P. Heath, L. E. Kavraki, and C. Clementi, "From coarse-grain to all-atom: Toward multiscale analysis of protein landscapes," *Proteins: Structure, Function, and Bioinformatics*, vol. 68, pp. 646–661, 2007.

[16] R. Rossi, A. Isorce, S. Morin, J. Flocard, K. Arumugam, S. Crouzy, M. Vivaudou, and S. Redon., "Adaptive torsion-angle quasi-statics: a general simulation method with applications to protein structure analysis and design," *Bioinformatics*, vol. 23, no. 13, pp. i408–i417, 2007.

[17] W. Labeikovsky, E. Z. Eisenmesser, D. A. Bosco, and D. Kern, "Structure and dynamics of Pin1 during catalysis by NMR," *Journal of Molecular Biology*, vol. 367, no. 5, pp. 1370–1381, 2007.

[18] O. Khatib, "A unified approach to motion and force control of robot manipulators: The operational space formulation," *International Journal of Robotics and Automation*, vol. 3, no. 1, pp. 43–53, 1987.

[19] C.-I. Branden and J. Tooze, *Introduction to Protein Structure*, 2nd ed. Garland Publishing, 1999.

[20] I. Lotan, H. van den Bedem, A. M. Deacon, and J.-C. Latombe, "Computing protein structures from electron density maps: The missing loop problem," in *Proceedings of the Workshop on the Algorithmic Foundations of Robotics (WAFR)*, 2004.

[21] T. Yamazaki, A. P. Hinck, Y. X. Wang, L. K. Nicholson, D. A. Torchia, P. Wingfield, S. J. Stahl, J. D. Kaufman, C. H. Chang, P. J. Domaille, and P. Y. Lam, "Three-dimensional solution structure of the HIV-1 protease complexed with DMP323, a novel cyclic urea-type inhibitor, determined by nuclear magnetic resonance spectroscopy," *Protein Science*, vol. 5, pp. 495–506, 1996.

[22] P. Y. Lam, P. K. Jadhav, C. J. Eyermann, C. N. Hodge, Y. Ru, L. T. Bacheler, J. L. Meek, M. J. Otto, M. M. rayner, Y. N. Wong, C.-H. Chang, P. C. Weber, D. A. Jackson, T. R. Sharpe, and S. Erickson-Vittanen, "Rational design of potent, bioavailable, nonpeptide cyclic ureas as hiv protease inhibitors," *Science*, vol. 263, pp. 380–384, January 1994.

[23] W. R. Scott and C. A. Schiffer, "Curling of flap tips in hiv-1 protease as a mechanism for substrate entry and tolerance of drug resistance," *Structure*, vol. 8, pp. 1259–1265, 2000.