Strategies for articulated multibody-based adaptive coarse grain simulation of RNA

Mohammad Poursina[1], **Kishor D. Bhalerao**[2], **Samuel C. Flores** [3], **Kurt S. Anderson**[1], **Alain Laederach**[4,5, §]

[1] Computational Dynamics Lab, Mechanical, Nuclear and Aerospace Engineering Department

Rensselaer Polytechnic Institute, Troy, New York, 12180, USA

e-mail: poursm@rpi.edu, anderk5@rpi.edu

[2] Department of Mechanical Engineering, The University of Melbourne,Victoria 3010 Australia

e-mail: kishorb@unimelb.edu.au

[3] Simbios Center, Bioengineering Department, Stanford University,

Clark Center S231, 318 Campus Drive, Stanford, California 94305-5444, USA

email: samuel.flores@aya.yale.edu

[4] Department of Biomedical Sciences, University at Albany, Albany, New York, 12208, USA

[5] Developmental Genetics and Bioinformatics, Wadsworth Center, Albany, New York, 12208, USA

e-mail: alain@wadsworth.org

[§]To whom correspondence should be addressed. Phone: (518) 486 4103, fax: (518) 474 3181, e-mail alain@wadsworth.org

ABSTRACT

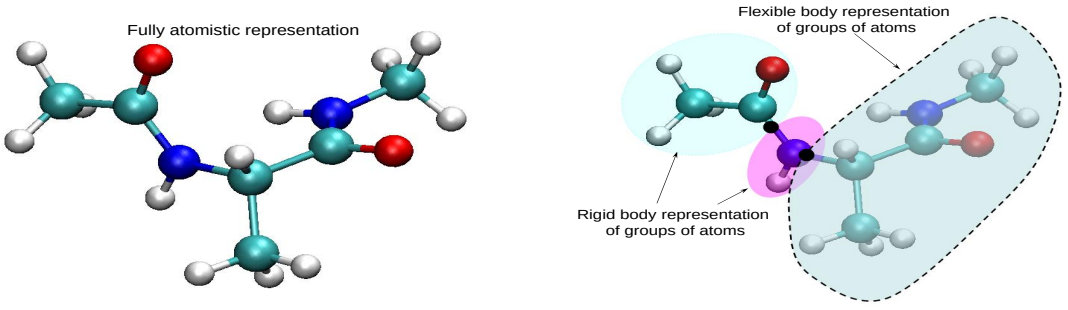
Efficient modeling approaches are necessary to accurately predict large-scale structural behavior of biomolecular systems like RNA (Ribonucleic Acid). Coarse grained approximations of such complex systems can significantly reduce the computational costs of the simulation while maintaining sufficient fidelity to capture the biologically significant motions. However, given the coupling and nonlinearity of RNA systems (and effectively all biopolymers), it is expected that different parameters such as geometric and dynamic boundary conditions, states, and applied forces will affect the system's dynamic behavior. Consequently, static coarse grained models (i.e., models for which the coarse graining is time invariant) are not always able to adequately sample the conformational space of the molecule. We introduce here the concept of adaptive coarse-grained molecular dynamics of RNA, which automatically adapts the coarseness of the model dynamically, in an effort to more optimally increase simulation speed, while maintaining accuracy. Adaptivity requires two basic algorithmic developments; first, a set of integrators that seamlessly allow transitions between higher and lower fidelity models while preserving the laws of motion. Secondly, we propose and validate metrics for determining when and where more or less fidelity needs to be integrated into the model to allow sufficiently accurate dynamics simulation. Given the central role that multibody dynamics plays in the proposed framework, and the nominally large number of dynamic degrees of freedom being considered in these applications, a computationally efficient multibody method which lends itself well to adaptivity is essential to the success of this effort. A suite of Divide-And-Conquer Algorithm (DCA)-based approaches are employed to this end, because these methods offer a good combination of computational efficiency and adaptive structure.

Keywords: adaptive coarse graining, articulated multibody dynamics, divide-and-conquer algorithm, RNA, transition metric

1 INTRODUCTION

Development and application of the efficient techniques to model highly complex biomolecular systems are pursued by engineers and scientists in an effort to predict and understand different structural dynamics of the systems and explain various biological processes (Scheraga *et al.*, 2007; Dill *et al.*, 2008; Parisien and Ma-

jor, 2008; Chen, 2008; Lebrun and Lavery, 1998). Among various types of biomolecular systems, nucleic acids, and in particular, RNA play a central regulatory role in the cell (Grundy and Henkin, 2006; Schroeder *et al.*, 2004; Tucker and Breaker, 2005; Ying and Lin, 2006). Key to RNA function is structure, in particular its ability to fold into a functional molecule capable of gene regulation and catalysis (Guo and Cech, 2002; Woodson, 2002; Zaug *et al.*, 1998). Structural dynamics of biomolecular systems including RNA can be modeled using a variety of different techniques. Conventional molecular dynamics (MD) (Leach, 2001; Haile, 1992) which benefits from the fully atomistic representation of the system as shown schematically in Fig. 1(a) is conceptually the simplest approach, and results from the direct application of Newton's law of motion. As such, these fully atomistic models capture all the dynamics of the system. The formulation and solution of the associated equations is trivial, once the forcing terms (force field calculations) have been determined. The overwhelming majority of the computational cost in MD simulations is in these force field calculations. Additionally, due to the high frequency motion of the atoms, it is necessary to use exceedingly small temporal integration step sizes when explicit integrators are used. High-stability implicit schemes for stiff differential equations, such as implicit-Euler (IE) (Hairer and Wanner, 1991; Schlick and Peskin, 1989; Peskin and Schlick, 1989) and implicit-midpoint (IM) (Mandziuk and Schlick, 1995) are also unsatisfactory for proteins and nucleic acids at atomic resolution at large time steps because of numerical damping (Nyberg and Schlick, 1992; Zhang and Schlick, 1993; Schlick and Peskin, 1995). Consequently, these techniques are not recommended to model the systems with large number of degrees of freedom.



(a) (b)

Figure 1: Different types of modeling of biomolecular systems, a) conventional MD simulations with fully atomistic representation of the system, b) using articulated multibody concepts with different rigid and flexible subdomains connected to each other via kinematic joints.

Alternatively, development of the coarse grained models with the idea of introducing superatoms (beads) (Praprotnik *et al.*, 2005) or using articulated multibody dynamics (Chun *et al.*, 2000; Poursina *et al.*, 2010) shown in Fig. 1(b) reduces the cost of the simulation while still capturing the overall conformational motion. These models can contain multiple resolutions ranging from fine scale atomistic domains, coarse grained macromolecules, to the continuum level system descriptions. Eliminating high frequency modes of motion within certain subdomains of the system allows one to significantly increase the size of the temporal integration steps. Additionally, it is not necessary to solve the equations governing the dynamics of the system for cartesian coordinates of all existing atoms. The dynamics needs only be solved for much smaller number of degrees of freedom of the system which can be more efficient in capturing the overall conformational motion. Furthermore, the structure of these models is such that the force field calculations are performed more efficiently. To roughly estimate and compare the computational costs associated with the force field calculations between MD simulation and the coarse grained approximation, let us consider a system with n number of atoms. MD simulation needs $\frac{n^2}{2}(1-\frac{1}{n})$ force field calculations to find either Van der Waals or electrostatic forces. If the system is modeled by the multibody dynamics approach with s rigid substructures each containing p atoms, the force field calculations (Van der Waals or electrostatic) are limited to the interactions between the atoms not located in the same rigid body (see Fig. 2). As such, the force field calculations reduce to $\frac{n^2}{2}(1-\frac{1}{s})$. This rough estimate of the force field calculations in the coarse grained models demonstrates that significant reduction in the number of the rigid substructures in the modeling (i.e., $s << n$) can result in the significant decrease in the computational cost corresponding to the force field calculations.

In biological systems of nucleic acids, and proteins, which tend to possess many (nominally $O(10^3) - O(10^7)$) degrees of freedom, important physical phenomena can occur at vastly different spatial and temporal scales. The small oscillations of individual tightly bonded atoms with high frequency content provide the sub-femto second $O(10^{-16})$ temporal domain while the conformational motions of interest occur in milliseconds $O(10^{-3})$ or larger time scales. Additionally, RNA molecules are not static, but in fact are highly dynamic (Kent *et al.*, 2000). The dynamic behavior of RNA is complex because of the dominant effect of base-pairing and stacking, which leads to regions of the polymer being highly rigid, connected by flexible loops (Rázga *et al.*, 2004, 2006). Stacking and pairing forces are cumulative, and with time,

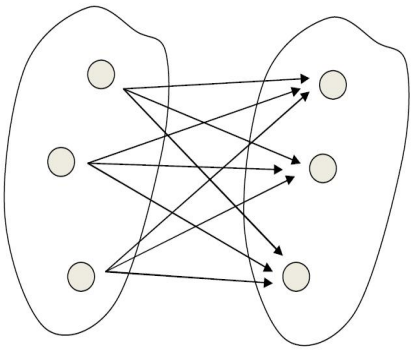


Figure 2: Force field calculation using multi-rigid-body dynamics approach. The interactions between the atoms inside the same rigid bodies are ignored.

different regions of the RNA can become rigid or flexible (Shcherbakova *et al.*, 2008). These important properties indicate that the quasi-static approach (Redon and Lin, 2006; Rossi *et al.*, 2007) is not appropriate to model such systems, and consequently, more challenges are introduced in the development of the adaptive multiscale methods to model the dynamics of RNA.

In the adaptive multiscale framework presented herein, the underlying dynamics formulation used within each system subdomain, as well as the subdomain definitions themselves, need to be intelligently chosen such that the system level solutions are determined with an acceptable accuracy in a timely manner. For instance, Fig. 1(b) provides an example system which is comprised of multiple subdomains, each containing potentially different forms of models. In such a system level model, the system is treated as multi-rigid and/or flexible bodies connected to each other via kinematic joints. In this case, the coarse graining algorithm should as a minimum be capable of adaptively adding/removing degrees of freedom to/from the model, as well as changing the definition of the flexible and rigid regions of the model during the simulation to efficiently provide the appropriate simulation results.

Physics-based, mathematics-based, or knowledge-based internal metrics (Rossi *et al.*, 2007; Poursina *et al.*, 2010) initiate and guide the change in the number and definition of the system subdomains. Additionally, these metrics may be used to assess the model performance and guide the types of models (atomistic, articulated multi-rigid and/or -flexible body) used within those subdomains to obtain the optimal combination of speed and accuracy. Thus, mechanisms need to be put in place for identifying critical locations where

constraints can be added, or removed as needed to enhance the simulation performance. Finally, a multibody formulation needs to be used which both lends itself well to adaptive (on the fly) changes in model and domain definition, while being efficient when applied to large complex systems. In this paper, important features of adaptive coarse grain modeling and simulation built off of articulated multibody dynamics with the application to RNA are addressed.

2 Need for the development of adaptive coarse graining machinery

In this section, we investigate the effects of specific parameters on the behavior of different kinds of RNA's establishing the necessity for developing general and robust adaptive machineries to model complex RNA systems.

2.1 Model description

We study different sequences of nucleotides Adenine (A), Cytosine (C), Guanine (G), Uracil (U), and construct representative single stranded RNA segments, each 18 nucleotides long, as shown in Fig. 3. All the simulations are performed using RNABuilder (Flores *et al.*, 2010). The RNABuilder package is written using the Simbody (Schmidt *et al.*, 2008) internal coordinate mechanics library and its molecular mechanics extension, Molmodel, both available from SimTK.org. Simbody includes variable step size integrators (Crosbie and Heyes, 1976), which continually attempt to maximize the integration step size without exceeding error tolerances, leading to significant time savings as well as stable behavior when dealing with large forces.

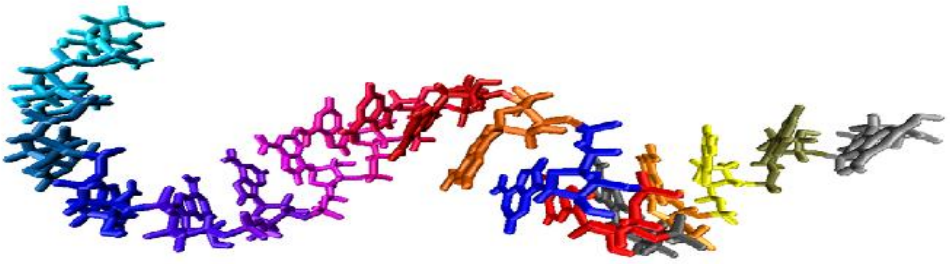


Figure 3: RNA with 18 residues shown in different colors. The prescribed motion is applied to the first residue located at the left side of the RNA. The last residue at the right side of the RNA is fixed in space.

RNABuilder provides a series of *mobilizer* commands which set the flexibility of the molecule (Flores *et al.*, 2010). In this work, we use the "rigid" mobilizer to rigidify each residue in the system. As such, each nucleotide (residue) is treated as a rigid body, being attached to its neighboring nucleotides by kinematic joints producing an articulated multibody system. Defining the *BondMobility* as "torsion", each nucleotide is connected to its parent residue via a revolute joint whose joint axis is coincident with the local bond. Each of the spatial 18 residue segments investigated may be viewed as a portion of a very much longer RNA, and the prescribed motions at the segments boundaries represent the dynamic effect of the rest of the system on the 18 residue RNA segment. For the performed simulations, one boundary (18^{th} residue) is fixed in space using the *constraint* called "Weld". The other boundary of the RNA (first residue) receives a prescribed motion in the form of simple harmonic functions in both x and y directions (in Newtonian frame) as

$$x(t) = 0.05[cos(\omega_1 t) - 1], \tag{1}$$

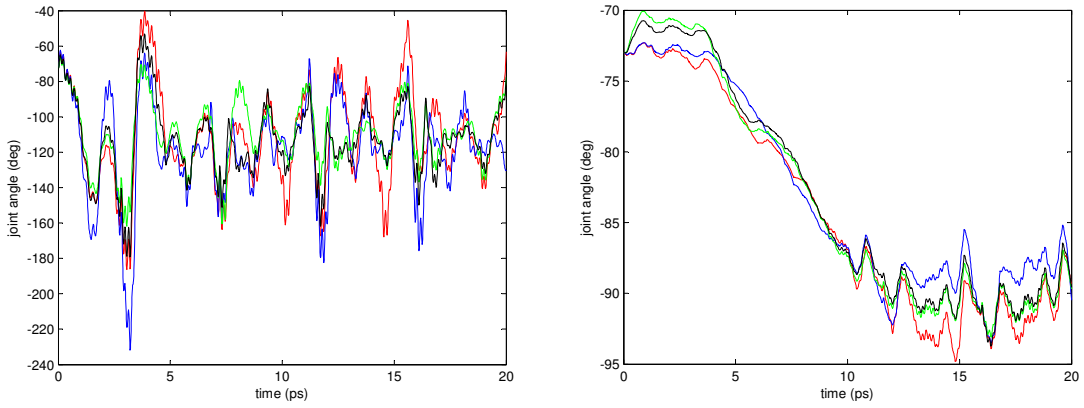
$$y(t) = 0.05[cos(\omega_2 t) - 1], \tag{2}$$

where x and y are measured in nanometers. This task has been performed using the command " Prescribed-Motion", and modifying the class "function" to create the functions indicated in Eqs. (1) and (2).

We conduct the simulations on RNA's with various sequences described previously with two different prescribed motions for 20 picoseconds. The first prescribed motion is defined with the same frequencies in both x and y directions as $\omega_1 = \omega_2 = 20$ rad/ps, while the second prescribed motion is composed of the frequency content of $\omega_2 = 2\omega_1 = 40$ rad/ps. The values of the frequencies and the amplitudes are selected such that the range of the prescribed motion is in accordance with the real motion. The simulations are performed at 300K, and the AMBER (Parm99) potential field (Wang *et al.*, 2000) is used to determine the bond and non-bond forces. The overall conformation of the RNA is important to identify the structural dynamics of the system. This overall conformation is determined by the values of the generalized coordinates defined at the joints connecting the consecutive residues. Since, in our truth model, all these joints are revolute, and torsional motion is allowed about the axis of the bonds which connect the consecutive residues, the joint angles are monitored during the course of the simulation to assess the overall conformation.

2.2 Inadequacy of the static coarse graining based on simulation results

Some representative simulation results for the systems described previously are provided here to convey the nature of the associated dynamic behavior. Figure 4 illustrates how differently the representative joint angles at various locations of the system and at different instants may behave, when the systems are excited by the prescribed motion characterized by $\omega_2 = 2\omega_1 = 40$ rad/ps. In Fig. 4(a), high amplitude motion is observed in dynamic behavior of the joint angle between residues 1 and 2. Consequently, this joint should not be locked in the coarse graining process. Figure 4(b) shows that after the passage of a couple of picoseconds, the joint angle between residues 8 and 9 experiences a transition from one regime of motion (experiencing little relative motion) to another regime of motion where the motion across the joint is significantly greater. Therefore, the coarse grained model in which this joint is initially locked, though potentially valid at the beginning of the simulation, does not provide a reliable conformation for the entire course of the simulation.



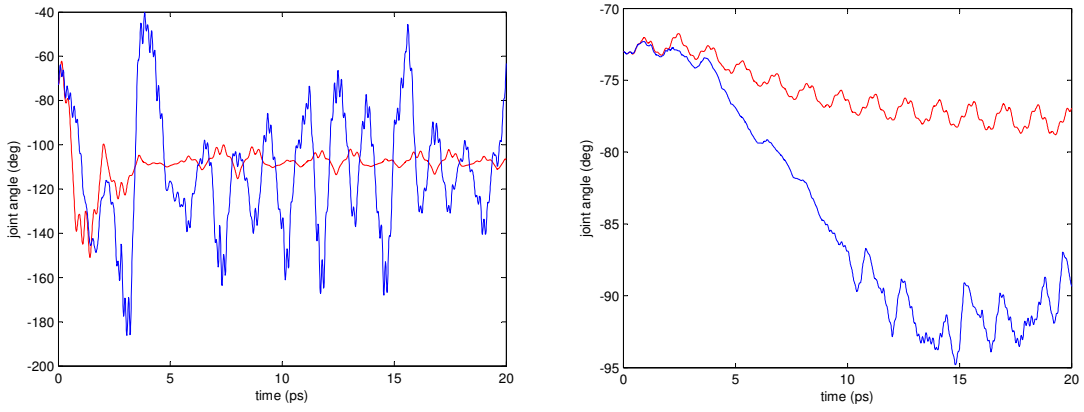
(a) (b)

Figure 4: Illustration of the nonlinear dynamics of RNA's with sequences: 18 A's (red), 18 U's (blue), 9 GC's (green), and 9 AU's (black), when the prescribed motion is characterized by $\omega_2 = 2\omega_1 = 40$ rad/ps, a) high amplitude motion of the joint angle between residues 1 and 2, b) significantly different regimes of motion of the joint angle between residues 8 and 9.

The prescribed motion at the RNA segment boundary represents the effect of the dynamics of the rest of the system as an input to the portion of interest. To study the effects of the changes in the input parameters on the response of the system in more detail, the behavior of the selected joint angles at different locations of

the poly-A RNA is shown in Fig. 5. Due to the subtle change in one of the input frequencies, an interesting behavior is observed which had not been predicted. It is expected that the joint angle between residues 1 and 2 that is located most closely to the prescribed motion would never be a good candidate for being locked within the course of the simulation. However, the results shown in Fig. 5(a) indicate that when the input motions contain the same frequencies as $\omega_1 = \omega_2 = 20$ rad/ps, this joint can be locked after the passage of time, while the same coarsening is inappropriate when the system is driven by the prescribed motions in which the frequency associated with the y component is doubled to 40 rad/ps.

As another example, in Fig. 5(b), the dynamics of the joint angle between residues 8 and 9 is compared when the system is excited by two different prescribed motions. This joint can be locked through the course of the simulation due to its small variations when both frequencies in the prescribed motions are chosen 20 rad/ps. However, the results in Fig. 5(b) indicate that if the value of ω_2 changes to 40 rad/ps, for the same system, this joint angle can not be locked during the whole course of the simulation.



(a) (b)

Figure 5: Effect of changing the input dynamics from $\omega_1 = \omega_2 = 20$ rad/s (red) to $\omega_2 = 2\omega_1 = 40$ rad/s (blue) on behavior of the poly-A RNA with sequence of 18 A's, a) joint angle between residues 1 and 2, the closest one to the excitation point, can be locked after a passage of time when $\omega_2 = 20$ rad/ps; however, it can not be locked when $\omega_2 = 40$ rad/ps, b) joint angle between residues 8 and 9 can be frozen for the whole course of the simulation when $\omega_2 = 20$ rad/ps; however, when $\omega_2 = 40$ rad/ps, the dynamic behavior of this joint angle prevents the model reduction at this location.

The simulation results affirm the nonlinear nature of the systems described previously. They demonstrate

that the dynamic behavior of each joint angle is highly time variant, and is significantly affected by the changes in the dynamics of the rest of the system. Consequently, even for a specific structure, a suggested coarse model may not be valid when the dynamics of the boundaries changes, or the conformation of the structure changes with time. As such, the static (i.e. time invariant) coarse graining may not provide the appropriate results, and the coarse grain modeling of these systems must be implemented in an adaptive framework. These conclusions can be generalized to other biomolecular systems such as DNA's and proteins. In this process, some degrees of freedom of the system are adaptively constrained or released at different instants and different locations of the system. Additionally, based on the behavior of different subdomains of the system, the definition of the rigid and/or flexible regions may change.

3 Metrics to guide transitions in adaptive modeling

Key to this effort is the need to develop metrics that guide the adaptive machinery. These metrics may be knowledge-base (derived empirically), math-based (derived from strictly mathematical relations), and/or physics-based (derived directly from physical laws). Herein, we introduce two different types of metrics for guiding the model transitions to the coarser and finer models.

3.1 Metrics to guide transitions from finer to coarser models

The behavior of the individual degrees of freedom of the model can be used to assess which of them may be removed, while the resulting reduced order model still produces essentially the same behavior, but with less computational effort. As such, the significant (active) degrees of freedom are retained in the coarsened model while the less significant degrees of freedom are identified and removed.

In biomolecular systems, the high frequency modes of motion that are related to the finer fidelity models provide large instantaneous relative velocities and accelerations. However, such modes do not contribute significantly to the global conformation of the system. Thus, velocity- and acceleration-based metrics are not well-suited for identifying those degrees of freedoms which are more significant in the overall conformational motions. Additionally, biomolecular systems are highly nonlinear and chaotic, and as such, using coarse graining metrics based on the instantaneous values of the states of the system are not expected to

(and have been shown not to) yield improved results. Therefore, we propose to monitor the moving-window statistical properties of the generalized coordinates of the system for math-based metrics to assess and guide the coarse graining process instead of instantaneous velocity- and acceleration-based metrics described in (Redon and Lin, 2006; Rossi *et al.*, 2007). The metric of choice for determining if an existing joint should be kept or removed is the standard deviation of the generalized coordinates defined at the joint collected within the sliding window, as given by

$$S_w = \sqrt{\frac{\sum_{k=1}^{n} (x_k - \bar{x}_w)^2}{n}}. \tag{3}$$

In the above equation, $\bar{x}_w$ is the moving-window average of the sequence of data within the window of the size of n. Each generalized coordinate contributes differently to the overall conformation. In other words, the overall conformation of the system may be more sensitive to some specific generalized coordinates, while this sensitivity varies with time. Consequently, if the weighted (scaled) value of the moving-window standard deviation of any generalized coordinate defined at the joint of the system (i.e., internal coordinate) is less than a predefined threshold, then the associated degree of freedom is considered less significant, and thus eligible to be frozen. In this work, since the torsional motion is the only allowable motion by the kinematic joints of the system, if the moving-window standard deviation of any specific joint angle is insignificant in comparison to the predefined threshold, the entire joint will be locked.

Herein, we pick the Poly-GC 18 nucleotide RNA to perform the coarse graining process based on the results obtained during a one picosecond simulation. Both frequencies in the prescribed motions are considered to be 20 rad/ps. The size of the window to calculate the standard deviation of the joint angles, as well as the value of the threshold on the standard deviation are both system dependent. In this case, the values of the joint angles are sampled every 0.01 picoseconds, and the length of the moving-window is chosen to be 100 sampling times. We also ignore the weights on the standard divinations associated with different joint angles of the system. Examining different values for the threshold, and comparing the results of the associated coarse model to those of the truth model, we lock the joints of the system if the moving-window standard deviation within the first 100 sampling times is less than one degree. Based on this threshold, the joint angles between residues 5 and 6, 7 and 8, 9 and 10, 11 and 12, 13 and 14, 14 and 15, 16 and 17, as well as 17 and 18 are allowed to be frozen.

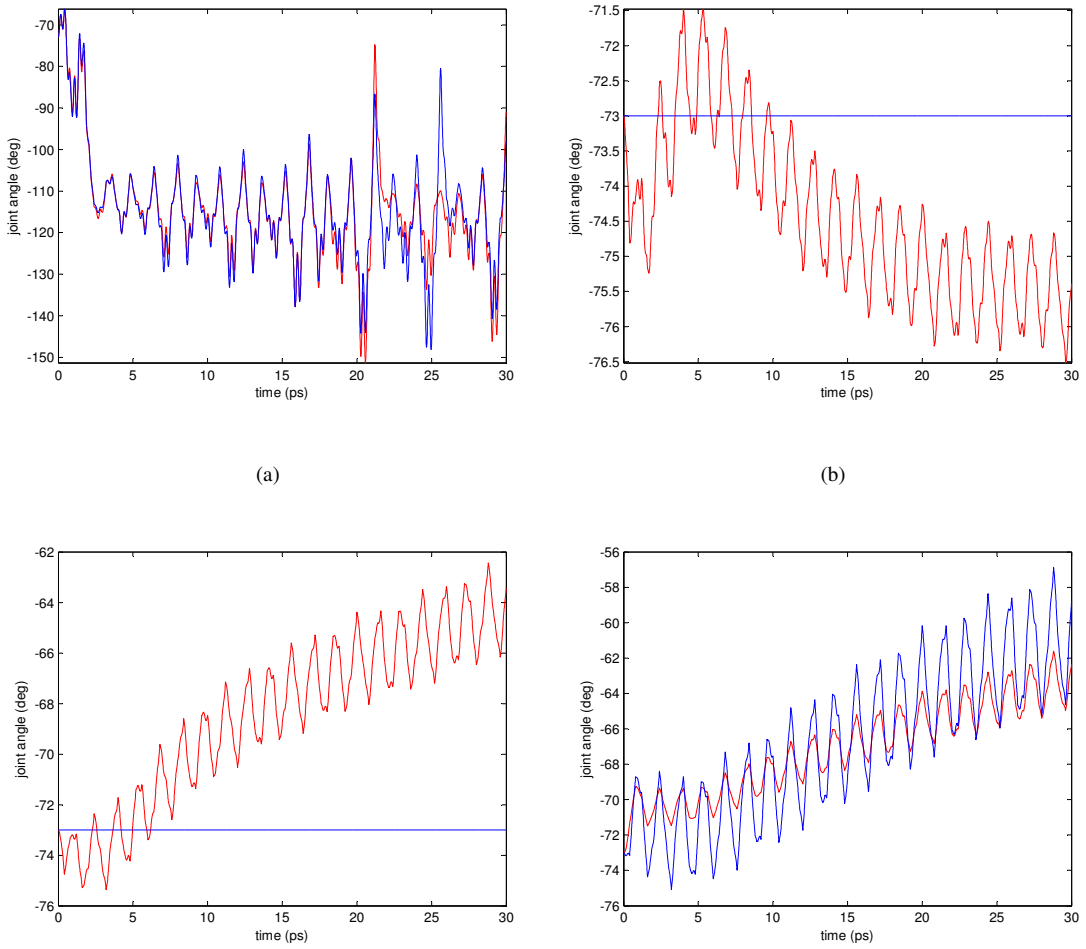


Figure 6: Comparison between the dynamic behavior of the fine system (red) and the suggested coarse model (blue) for poly-GC RNA. a) joint angle between residues 2 and 3, b) joint angle between residues 9 and 10, c) joint angle between residues 11 and 12, d) joint angle between residues 12 and 13.

The adequacy of the proposed coarse graining can be assessed by comparing the results achieved by the coarse grained model with those obtained exclusively from the fine grained "truth model" over the extended simulation periods. The simulation of the suggested coarse model is, therefore, conducted for 30 picoseconds with the same initial conditions and prescribed motions applied to the truth model. Figure 6 shows the behavior of some representative joint angles for both models. The joint angle between residues 2 and 3 in both models experiences the same behavior as shown in Fig. 6(a). Although the joint angle between residues 9 and 10 has been locked in the coarsening process, Fig. 6(b) indicates that the associated behavior

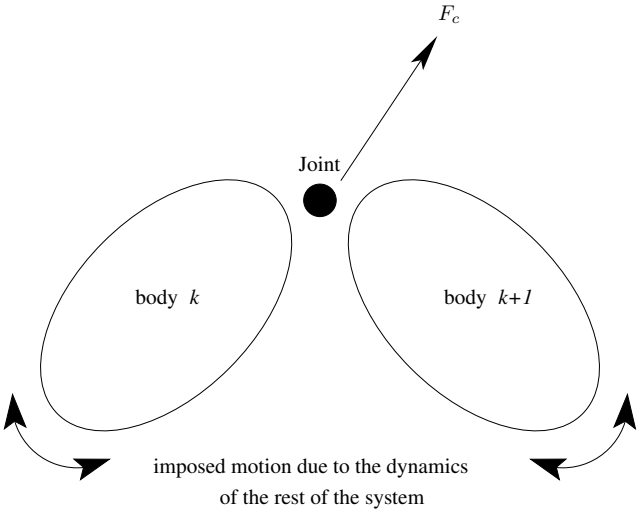


Figure 7: Magnitude of constraint force gives an estimate of errors introduced due to locking the joint.

in the proposed coarse model is very close to that obtained for the finer model. Based on the results shown in Fig. 6(c), the joint angle between residues 11 and 12 in the coarse model deviates from that of the truth model as time passes. According to the required accuracy, it may be needed to unlock this joint angle during the adaptive coarse graining process. Finally, in Fig. 6(d), similar trends are observed in the dynamics of the relative motion between residues 12 and 13 for both models.

3.2 Metrics to guide transitions from coarser to finer models

An inadequate coarse grained model as a result of constraining certain critical degrees of freedom of the system can lead to an incorrect representation of the system behavior and conformation. An important metric to check the validity of the selected coarse grained model is the spatial constraint loads (forces and moments) acting on all the kinematic joints of the system and intermediate locations within the rigid and flexible bodies which represent segments of the articulated molecular system. The system's internal and constraint loads arise from the interactions between the bodies, the imposed boundary conditions, and the kinematic constraints imposed on adjacent body-to-body motions by the connecting joints as shown in Fig. 7. For instance, consider the coarse grained model provided in section 3.1. If the frequency content in the y-direction of the prescribed motion in Eq. (2) changes from 20 rad/ps to 60 rad/ps at $t = 10$ ps, the behavior of the constraint torques change significantly in some joints of the system locked previously in the coarse graining process. For instance, Fig. 8 shows the significant increase in the value of the constraint

torque in the direction of the bond connecting residues 16 and 17 which has been locked in the coarsening process. Thus, the constraint load magnitude gives an indication of the degree to which the body or joint in question is attempting to be deformed at a location. If during the course of the simulation, the value of the spatial constraint load at any location of the system exceeds the nominal load which figuratively causes the mechanical failure, the associated joint is released. This effectively releases (removes) the constraint, and permits the adjacent segments to move relative to one another in a manner permitted by the joint and in response to the system forcing terms. The required mechanisms to change the definition of the joints, and the derivation of the appropriate math for the model transitions remain a contemporary challenge in the field.

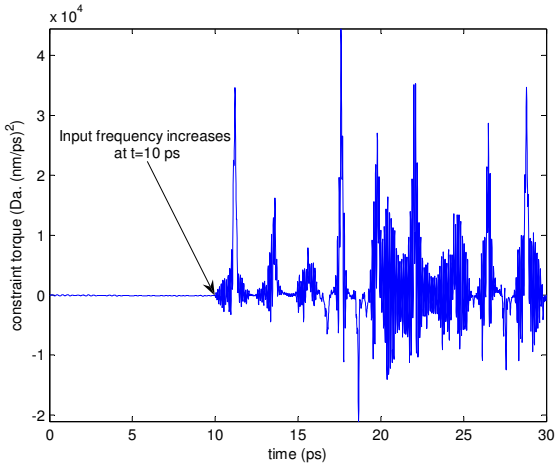


Figure 8: Constraint torque between residues 10 and 11. The constraint torque about the bond direction (locked in the coarsening process) changes significantly when a modest change occurs in the frequency of the prescribed motion at $t = 10$ ps.

Adding degrees of freedom into the system poses additional challenges. In real mechanical systems, since energy is not "created" by unlocking a joint or the failure of a member, the "creation" of a new mode of motion by the removal of a constraint does not pose any problem. Therefore, there is no jump in the system velocity variables. However, in the coarse graining process of biomolecular systems, naturally existing higher modes of motion are ignored in the modeling because the internal metric had previously indicated these modes as less relevant. Therefore, in the transition from the coarse model to the finer model, the current value of the kinetic energy of these ignored modes must be estimated and considered appropriately.

Details of choosing the optimum solution, and the various issues in adding joints in the model have been discussed in detail in (Poursina *et al.*, 2009; Anderson and Poursina, 2009a).

4 Adaptive modeling framework in DCA scheme

Within the adaptive framework, a Divide-And-Conquer Algorithm (DCA) is used to solve the forward dynamics problem for the articulated multibody systems composed of rigid and flexible bodies with any arbitrary configurations such as kinematically open loops and multiple closed loops (Featherstone, 1999a,b; Mukherjee and Anderson, 2007b,c). These DCA-based methods are used in the context of the large-scale adaptive molecular problems because: 1) They are relatively efficient for large scale sequential computer implementation; 2) These formulations have a highly modular structure, which makes their implementation and use within an adaptive framework relatively straightforward; And, 3) These methods are highly parallelizable. The computational complexity of the algorithm is $O(n)$ and $O(log(n))$ in serial and parallel implementations, respectively, where n denotes the number of degrees of freedom of the system. The basic idea of the DCA is to treat a large multibody system by recursively assembling adjacent articulated bodies/subsystems (Fig. 9(a)) into larger encompassing subsystems (Fig. 9(b)). The brief overview of the method is provided here.

Consider the two consecutive articulated bodies k and $k+1$ shown in Fig. 9(a) which are connected to each other via the kinematic joint J^k. The term *handle*, which appears in the paper continually is any selected point on the body used in modeling the interactions of the body with the environment. The handles on the body can correspond to the joint locations, center of mass or any desired reference points. Each body can have any number of handles on it. For the algorithm presented here, the joint locations are chosen as the handles on the body. The two-handle equations of motion for each body are those in which the spatial acceleration of each handle of the body is expressed as a linear combination of the spatial constraint forces applied to the inward and outward handles. These equations for bodies k and $k+1$ are expressed as

$$\mathcal{A}_1^k = \phi_{11}^k \mathcal{F}_{1c}^k + \phi_{12}^k \mathcal{F}_{2c}^k + \phi_{13}^k, \tag{4}$$

$$\mathcal{A}_2^k = \phi_{21}^k \mathcal{F}_{1c}^k + \phi_{22}^k \mathcal{F}_{2c}^k + \phi_{23}^k, \tag{5}$$

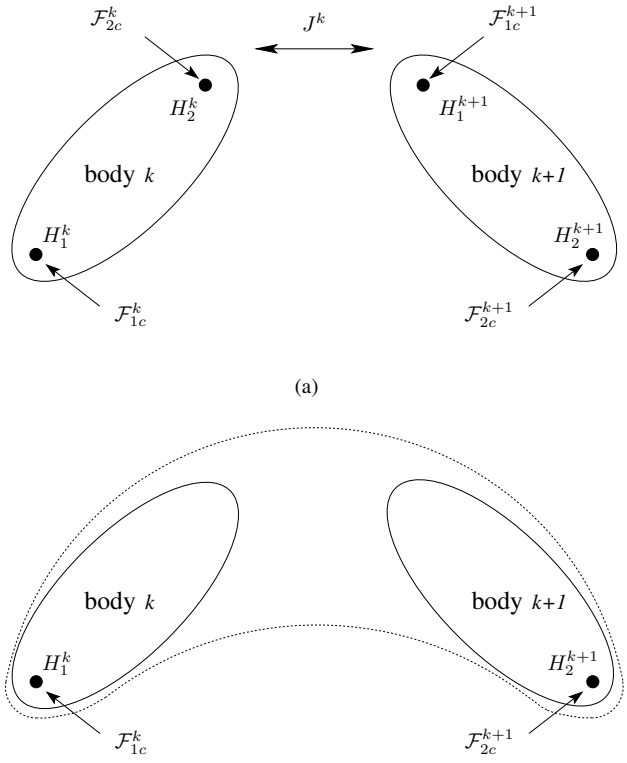


Figure 9: Assembling of two consecutive bodies to form a new body, a) consecutive bodies k and $k+1$ connected to each other via a kinematic joint, b) fictitious body formed from two distinct bodies by eliminating the constraint force and spatial accelerations corresponding to the common joints.

and

$$\mathcal{A}_1^{k+1} = \phi_{11}^{k+1}\mathcal{F}_{1c}^{k+1} + \phi_{12}^{k+1}\mathcal{F}_{2c}^{k+1} + \phi_{13}^{k+1}, \tag{6}$$

$$\mathcal{A}_2^{k+1} = \phi_{21}^{k+1}\mathcal{F}_{1c}^{k+1} + \phi_{22}^{k+1}\mathcal{F}_{2c}^{k+1} + \phi_{23}^{k+1}, \tag{7}$$

respectively. In the above equations, all the coefficients ϕ_{ij}'s are known quantities at each time step. The terms ϕ_{ij} $(i,j = 1,2)$ are associated with the inertia of the body, and consequently, are constant for each body within the course of the simulation if expressed in body basis. The terms ϕ_{i3} $(i = 1,2)$ are associated with the known applied forces, as well as centripetal and coriolis terms which should be updated at each time step. At the joint J^k, the kinematical constraint at the acceleration level is expressed as:

$$\mathcal{A}_1^{k+1} = \mathcal{A}_2^k + P^{J^k}\dot{u}^{J^k} + \dot{P}^{J^k}u^{J^k}, \tag{8}$$

where P^{J^k} is the known matrix associated with the joint free-modes of motion (Roberson and Schwertassek,

1988), and u^{J^k} represents the known generalized speeds defined at the joint J^k. Using the relations provided above, and introducing the matrix D^{J^k} as the orthogonal complement of the joint free-motion map at J^k, one can arrive at the two-handle equations governing the dynamics of the assembly $k:k+1$ as

$$\mathcal{A}_1^k = \phi_{11}^{k:k+1}\mathcal{F}_{1c}^k + \phi_{12}^{k:k+1}\mathcal{F}_{2c}^{k+1} + \phi_{13}^{k:k+1}, \tag{9}$$

$$\mathcal{A}_2^{k+1} = \phi_{21}^{k:k+1}\mathcal{F}_{1c}^k + \phi_{22}^{k:k+1}\mathcal{F}_{2c}^{k+1} + \phi_{23}^{k:k+1}, \tag{10}$$

where,

$$\phi_{11}^{k:k+1} = [\phi_{11}^k - \mathcal{W}\phi_{21}^k], \tag{11}$$

$$\phi_{12}^{k:k+1} = [\mathcal{W}\phi_{12}^{k+1}], \tag{12}$$

$$\phi_{13}^{k:k+1} = [\phi_{13}^k - \mathcal{W}\mathcal{Y}], \tag{13}$$

$$\phi_{21}^{k:k+1} = [\mathcal{Z}\phi_{21}^k], \tag{14}$$

$$\phi_{22}^{k:k+1} = [\phi_{22}^{k+1} - \mathcal{Z}\phi_{12}^{k+1}], \tag{15}$$

$$\phi_{23}^{k:k+1} = [\phi_{23}^{k+1} + \mathcal{Z}\mathcal{Y}], \tag{16}$$

$$\mathcal{X} = D^{J^k}([D^{J^k}]^T[\phi_{22}^k + \phi_{11}^{k+1}]D^{J^k})^{-1}[D^{J^k}]^T, \tag{17}$$

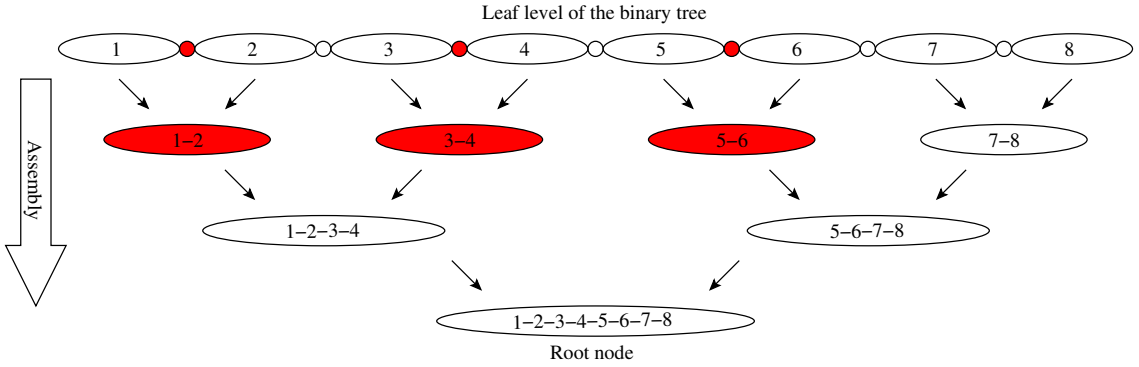
$$\mathcal{Y} = \phi_{23}^k - \phi_{13}^{k+1} + \dot{P}^{J^k}u^{J^k}, \tag{18}$$

$$\mathcal{W} = \phi_{12}^k\mathcal{X}, \tag{19}$$

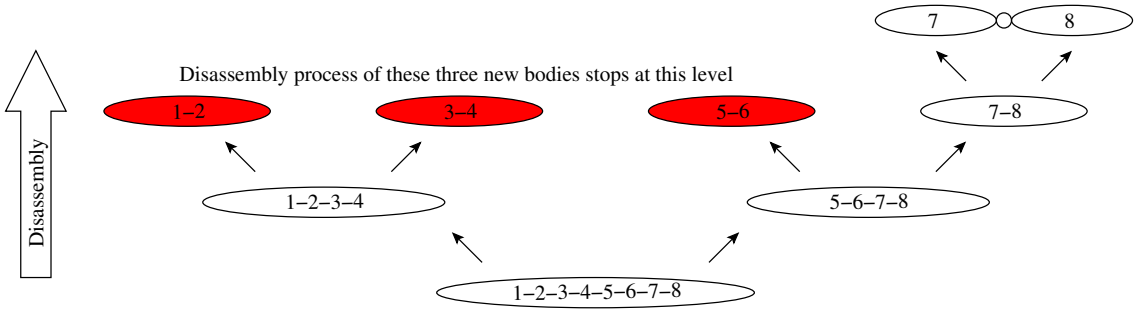
$$\mathcal{Z} = \phi_{21}^{k+1}\mathcal{X}. \tag{20}$$

The divide-and-conquer algorithm is implemented in two main passes: assembly and disassembly as shown in Fig. 10. The assembly process starts at the level of the individual bodies (leaf nodes) by coupling together pairs of adjacent bodies to form the assemblies. Proceeding in this manner, one can recursively eliminate both unknown constraint loads and joint accelerations at the common joints of the consecutive bodies/assemblies to form the two-handle equations of the resulting assemblies. This process works hierarchically exploiting the same structure as that of a binary tree. At the end of the hierarchic assembly process, the whole articulated system may be modeled in terms of the two-handle equations of motion of a single all encompassing assembly. Different types of boundary conditions may be applied to this root node of the binary tree. It is possible for each boundary of the system to be either free floating or constrained by a kinematic joint or prescribed motion. The disassembly process starts by applying the boundary conditions to the

system, and solve for the unknown spatial constraint forces and/or accelerations of the terminal handles of the whole system. Solving this problem is addressed in detail in (Mukherjee and Anderson, 2007c). Then, these known values of the spatial constraint forces and accelerations of the terminal handles are substituted into the two-handle equations of the associated subassemblies (Eqs. (6) and (7)) to find the values of the spatial constraint forces and the accelerations at the common joints of the subassemblies. This process is repeated in a hierarchic disassembly of the binary tree where the known boundary conditions are used to solve the two-handle equations of the subassemblies, until the spatial constraint forces and the generalized speeds of all bodies in the system are determined.



(a)



(b)

Figure 10: Adaptive framework in DCA scheme, red joints are locked, and red assemblies are composed of the consecutive bodies whose connecting joint is locked, a) assembly process by coupling together pairs of adjacent bodies/subassemblies to form the new assemblies until reaching the all encompassing node, b) disassembly process to find the spatial constraint forces and the generalized speeds at the common joints.

In performing the adaptive modeling of the dynamics of the system in DCA framework, if based on the values of the moving-window standard deviation of the joint angles, it is desired to lock (remove) a joint

of the system, the associated orthogonal complement of the joint free-motion map becomes the identity matrix. As such, Eq. (17) is simplified as:

$$\mathcal{X} = [\phi_{22}^{k} + \phi_{11}^{k+1}]^{-1}. \tag{21}$$

If based on monitoring the values of the spatial constraint forces, it is deemed necessary to change the definition of the joint within the adaptive modeling, the only change occurs at the leaf level is changing the joint free-motion map P^{J^k} of the associated joint.

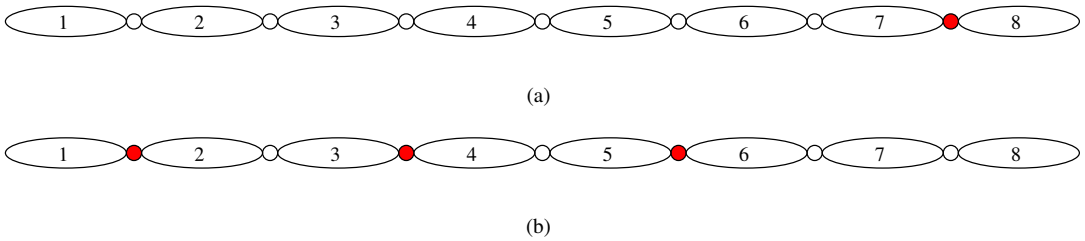


Figure 11: Illustration of transitioning between coarse models, red joints are locked. a) current coarse model, in which a high strain is measured in the eighth joint. b) new coarse model, where the eighth joint is released and the first, third and fifth joints are locked.

For instance, consider the current coarse model shown in Fig. 11(a) composed of eight bodies (residues). All the connecting joints are revolute, except the last one shown in red which is locked. The desired model is formed by releasing the eighth joint due to the corresponding high strain. Additionally, in this model, the first, third and fifth joints are to be locked as shown in Fig. 11(b) because they are determined to be making an insignificant contribution to the overall conformation of the system. In this case, the spatial joint free-motion maps of the locked joints become zero matrices. Consequently, the associated orthogonal complement of the joint free-motion map is replaced by the identity matrix if expressed in the joint coordinate system. For the revolute joints of the system, the spatial joint free-motion map expressed

in joint basis becomes

$$P^{J^k} = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}. \tag{22}$$

As such, the associated orthogonal complement of the joint free-motion map is expressed as

$$D^{J^k} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}. \tag{23}$$

The assembly and disassembly processes are performed as described before using the appropriate matrices characterizing the joints of the desired coarse model. In these processes, subassemblies shown in red in Fig. 10 are those composed of consecutive bodies whose connecting joint is locked. For the new coarse model, the disassembly process of the three new bodies stops at one level prior to the leaf level of the binary tree as shown in Fig. 10(b), since all the associated spatial constraint forces and generalized speeds are available at this level. However, this process continues to the leaf level for the determination of the unknown values of the spatial constraint forces and generalized speeds associated with common joint of bodies 7 and 8.

Any violation in the conservation of the generalized momentum of the system in the transition between different models leads to nonphysical results since the instantaneous switch in the system model definition is incurred without the influence of any external forces. In other words, the system momentum of each differential element projected onto the space of admissible motions permitted by the more restrictive model (whether pre- or post-transition) when integrated over the entire system must be conserved across the model transition (Kane and Levinson, 1985). The jumps in the system partial velocities (free modes of motion

(Roberson and Schwertassek, 1988)) due to the sudden change in the model resolution result in the jumps in the generalized speeds corresponding to the new set of degrees of freedom. Therefore, the adaptive framework should be equipped with the machinery to provide the momenta balance equations. The formation of the impulse-momentum equations within the transitions can also be performed in divide-and-conquer scheme (Mukherjee and Anderson, 2007a). Similar to the DCA, the DCA-based generalized momentum assembly-disassembly procedures are performed on the two-handle impulse-momentum equations of each body as follows (Mukherjee and Anderson, 2007a)

$$\Delta\mathcal{V}_1^k = \Phi_{11}^k \int_{t^-}^{t^+} \mathcal{F}_{1c}^k dt + \Phi_{12}^k \int_{t^-}^{t^+} \mathcal{F}_{2c}^k dt + \Phi_{13}^k, \tag{24}$$

$$\Delta\mathcal{V}_2^k = \Phi_{21}^k \int_{t^-}^{t^+} \mathcal{F}_{1c}^k dt + \Phi_{22}^k \int_{t^-}^{t^+} \mathcal{F}_{2c}^k dt + \Phi_{23}^k, \tag{25}$$

where Φ_{ij}^k is an intermediate known term. The terms $\Delta\mathcal{V}_i^k$ $(i = 1, 2)$ indicate the jumps in the spatial velocities of the i^{th} handle of body k. The spatial impulsive constraint forces acting on body k at its inward (parent) and outward (child) handles within the instantaneous transition time period from t^- to t^+ are represented by $\int_{t^-}^{t^+} \mathcal{F}_{1c}^k dt$ and $\int_{t^-}^{t^+} \mathcal{F}_{2c}^k dt$, respectively. Since the impulses associated with the external loads, as well as the coriolis and centripetal terms can be ignored, the terms Φ_{13}^k and Φ_{23}^k do not appear in the equations governing the impulse-momentum of the individual bodies at the leaf level of the binary tree. Additionally, at all levels of the binary tree, for each body/assembly, the coefficients of the spatial constraint forces, as well as the impulsive constraint forces which are essentially functions of the inertia and the geometry are the same, in other words:

$$\Phi_{ij}^k = \phi_{ij}^k \quad i, j = 1, 2 \tag{26}$$

As such, setting up the generalized momentum balance equations is fast and easy. This algorithm is also capable of efficiently providing the required equations for the knowledge-, math- or physics-based optimization problem in the transitions from the coarse models to the finer models as explained in detail in (Anderson and Poursina, 2009b).

5 Conclusions

We have addressed different issues associated with the adaptive coarse grain modeling of biomolecular systems using articulated multibody dynamics concepts. Based on the results of the simulations of different types of RNA's, it has been demonstrated that the static (time invariant) coarse graining of models are not likely to provide the appropriate information about the structural behavior of the system for the entire course of the simulation due to the nonlinearities embedded in the system's dynamics. Therefore, there is a need for the development of an adaptive molecular modeling and simulation framework. The adaptive algorithm should be capable of identifying the critical locations of the system to remove and add degrees of freedom, or change model types and definitions as necessary. In the paper, we have suggested the moving-window standard deviation of the joint angles, as well as the values of the constraint forces as metrics to guide the transitions to the coarser and finer models, respectively. Since the system is treated as the articulated multibody system, efficient algorithm is necessary for the modeling. Divide-and-conquer algorithm has been proposed as a convenient technique for the efficient modeling of the system. This method is highly modular, and as such, lends itself well to adaptivity, and massive parallelization. The computational complexity of the method is $O(n)$ and $O(log(n))$ in serial and parallel implementations, respectively, where n is the number of the degrees of freedom of the system.

Acknowledgment

This work was supported through the NSF award No. CMMI-0757936 to Kurt Anderson and in part through the US National Institutes of Health grant R00 GM079953 (NIGMS) to Alain Laederach. The authors would like to thank the funding agencies and also show their gratitude to Dr. Russ Altman and Michael Sherman from Simbios group at Stanford University for their help in this effort. The supplementary materials including movies, input and output files and some associated programs are attached to this paper.

REFERENCES

Anderson, K.S., and Poursina, M. (2009a). Energy concern in biomolecular simulations with transition from a coarse to a fine model. In Proceedings of the Seventh International Conference on Multibody Systems,

Nonlinear Dynamics and Control, ASME Design Engineering Technical Conference 2009, (IDETC09), IDETC2009/MSND-87297. San Diego, CA.

Anderson, K.S., and Poursina, M. (2009b). Optimization problem in biomolecular simulations with dca-based modeling of transition from a coarse to a fine fidelity. In Proceedings of the Seventh International Conference on Multibody Systems, Nonlinear Dynamics and Control, ASME Design Engineering Technical Conference 2009, (IDETC09), IDETC2009/MSND-87319. San Diego, CA.

Chen, S.J. (2008). RNA folding: Conformational statistics, folding kinetics, and ion electrostatics. *Annual Review of Biophysics* **37**, 197–214.

Chun, H.M., Padilla, C.E., Chin, D.N., Watenabe, M., Karlov, V.I., Alper, H.E., Soosaar, K., Blair, K.B., Becker, O.M., Caves, L.S.D., Nagle, R., Haney, D.N., and Farmer, B.L. (2000). MBO(N)D: A multibody method for long-time molecular dynamics simulations. *J. Comput. Chem.* **21**, 159–184.

Crosbie, R.E., and Heyes, W. (1976). Variable-step integration methods for simulation applications. *Applied Mathematical Modelling* **1**, 137–140.

Dill, K.A., Ozkan, S.B., Shell, M.S., and Weikl, T.R. (2008). The protein folding problem. *Annual Review of Biophysics* **37**, 289–316.

Featherstone, R. (1999a). A divide-and-conquer articulated body algorithm for parallel $O(\log(n))$ calculation of rigid body dynamics. Part 1: Basic algorithm. *International Journal of Robotics Research* **18**, 867–875.

Featherstone, R. (1999b). A divide-and-conquer articulated body algorithm for parallel $O(\log(n))$ calculation of rigid body dynamics. Part 2: Trees, loops, and accuracy. *International Journal of Robotics Research* **18**, 876–892.

Flores, S., Wan, Y., Russell, R., and Altman, R.B. (2010). Predicting RNA structure by multiple template homology modeling. In Proceedings of the Pacific Symposium on Biocomputing, pages 216–227.

Grundy, F., and Henkin, T. (2006). From ribosome to riboswitch: control of gene expression in bacteria by RNA structural rearrangements. *Critical Reviews in Biochemistry and Molecular Biology* **41**, 329–338.

Guo, F., and Cech, T. (2002). Evolution of tetrahymena ribozyme mutants with increased structural stability. *Nature Structural Biology* **9**, 855–861.

Haile, J. (1992). Molecular Dynamics Simulation : Elementary Methods. Wiley Interscience, New York.

Hairer, E., and Wanner, G. (1991). Solving Ordinary Differential Equations II. Stiff and Differential-Algebraic Problems, volume 14. New York: Springer-Verlag, springer series in comput. math. edition.

Kane, T.R., and Levinson, D.A. (1985). Dynamics: Theory and Application. Mcgraw-Hill, NY.

Kent, O., Chaulk, S., and MacMillan, A. (2000). Kinetic analysis of the M1 RNA folding pathway. *J Mol Biol* **304**, 699–705.

Leach, A.R. (2001). Molecular Modelling Principles and Applications. Prentice Hall, second edition.

Lebrun, A., and Lavery, R. (1998). Modeling the mechanics of a DNA oligomer. *J. Biomol. Struct. Dyn.* **16**, 593–604.

Mandziuk, M., and Schlick, T. (1995). Resonance in the dynamics of chemical systems simulated by the implicit-midpoint scheme. *Chem. Phys. Lett.* **237**, 525–535.

Mukherjee, R.M., and Anderson, K.S. (2007a). Efficient methodology for multibody simulations with discontinuous changes in system definition. *Multibody System Dynamics* **18**, 145–168.

Mukherjee, R.M., and Anderson, K.S. (2007b). A logarithmic complexity divide-and-conquer algorithm for multi-flexible articulated body systems. *Computational and Nonlinear Dynamics* **2**, 10–21.

Mukherjee, R.M., and Anderson, K.S. (2007c). An orthogonal complement based divide-and-conquer algorithm for constrained multibody systems. *Nonlinear Dynamics* **48**, 199–215.

Nyberg, A., and Schlick, T. (1992). Increasing the time step in molecular dynamics. *Chem. Phys. Lett.* **198**, 538–546.

Parisien, M., and Major, F. (2008). The MC-fold and MC-Sym pipeline infers RNA structure from sequence data. *Nature* **452**, 51–55.

Peskin, C.S., and Schlick, T. (1989). Molecular dynamics by the backward Euler's method. *Communications in Pure and Applied Math* **42**, 1001–1031.

Poursina, M., Bhalerao, K.D., and Anderon, K.S. (2010). Divide-and-conquer based coarse grained simulation of RNA. In Proceedings of ASME 2010 First Global Congress on NanoEngineering for Medicine and Biology NEMB2010, NEMB2010-13123. Houston, TX.

Poursina, M., Bhalerao, K.D., and Anderson, K.S. (2009). Energy concern in biomolecular simulations with discontinuous changes in system definition. In Proceedings of the ECCOMAS Thematic Conference - Multibody Systems Dynamics. Warsaw, Poland.

Praprotnik, M., Site, L., and Kremer, K. (2005). Adaptive resolution molecular-dynamics simulation: Changing the degrees of freedom on the fly. *J. Chem. Phys.* **123**, 224106–224114.

Redon, S., and Lin, M.C. (2006). An efficient, error-bounded approximation algorithm for simulating quasi-statics of complex linkages. *Computer-Aided Design* **38**, 300–314.

Roberson, R.E., and Schwertassek, R. (1988). Dynamics of Multibody Systems. Springer-Verlag, Berlin.

Rossi, R., Isorce, M., Morin, S., Flocard, J., Arumugam, K., Crouzy, S., Vivaudou, M., and Redon, S. (2007). Adaptive torsion-angle quasi-statics: a general simulation method with applications to protein structure analysis and design. In ISMB/ECCB (Supplement of Bioinformatics), pages 408–417.

Rázga, F., Spackova, N., Réblova, K., Koca, J., Leontis, N., and Sponer, J. (2004). Ribosomal RNA kink-turn motif–a flexible molecular hinge. *J. Biomol. Struct. Dyn.* **22**, 183–194.

Rázga, F., Zacharias, M., Réblova, K., Koca, J., and Sponer, J. (2006). RNA kink-turns as molecular elbows: hydration, cation binding, and large-scale dynamics. *Structure* **14**, 825–835.

Scheraga, H.A., Khalili, M., and Liwo, A. (2007). Protein-folding dynamics: Overview of molecular simulation techniques. *Annu. Rev. Phys. Chem.* **58**, 57–83.

Schlick, T., and Peskin, C.S. (1989). Can classical equations simulate quantum-mechanical behavior? A molecular dynamics investigation of a diatomic molecule with a morse potential. *Communications on Pure and Applied Mathematics* **42**, 1141–1163.

Schlick, T., and Peskin, C. (1995). Comment on: The evaluation of LI and LIN for dynamics simulations. *J. Chem. Phys.* **103**, 9888–9889.

Schmidt, J.P., Delp, S.L., Sherman, M.A., Taylor, C.A., Pande, V.S., and Altman, R.B. (2008). The simbios national center: Systems biology in motion. *Proceedings of the IEEE, special issue on Computational System Biology* **96**, 1266–1280.

Schroeder, R., Barta, A., and Semrad, K. (2004). Strategies for RNA folding and assembly. *Nature Reviews Molecular Cell Biology* **5**, 908–919.

Shcherbakova, I., Mitra, S., Laederach, A., and Brenowitz, M. (2008). Energy barriers, pathways, and dynamics during folding of large, multidomain RNAs. *Curr. Opin. Chem. Biol.* **12**, 655–666.

Tucker, B., and Breaker, R. (2005). Riboswitches as versatile gene control elements. *Curr. Opin. Struct. Biol.* **15**, 342–348.

Wang, J., Cieplak, P., and Kollman, P.A. (2000). How well does a restrained electrostatic potential (resp) model perform in calculating conformational energies of organic and biological molecules? *J. Comput. Chem.* **21**, 1049–1074.

Woodson, S. (2002). Recent insights on RNA folding mechanisms from catalytic RNA. *Cell. Mol. Life Sci.* **57**, 796–808.

Ying, S., and Lin, S. (2006). Current perspectives in intronic micro RNAs (miRNAs). *Journal of Biomedical Science* **13**, 5–15.

Zaug, A., Grosshans, C., and Cech, T. (1998). Sequence-specific endoribonuclease activity of the tetrahymena ribozyme: enhanced cleavage of certain oligonucleotide substrates that form mismatched ribozyme-substrate complexes. *Biochemistry* **27**, 8924–8931.

Zhang, G., and Schlick, T. (1993). LIN: A new algorithm combining implicit integration and normal mode techniques for molecular dynamics. *Journal of Compuational Chemistry* **14**, 1212–1233.