JCTC Journal of Chemical Theory and Computation

Article

pubs.acs.org/JCTC

Automatic State Partitioning for Multibody Systems (APM): An Efficient Algorithm for Constructing Markov State Models To Elucidate Conformational Dynamics of Multibody Systems

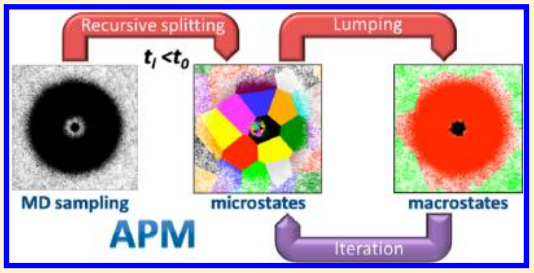
Fu Kit Sheong,[†,‡] Daniel-Adriano Silva,[‡,⊥] Luming Meng,[†,‡] Yutong Zhao,[‡] and Xuhui Huang*[,†,‡,§,∥]

[†]HKUST Shenzhen Research Institute, Nanshan, Shenzhen 518057, China

[‡]Department of Chemistry, [§]Division of Biomedical Engineering, and [∥]Center of Systems Biology and Human Health, School of Science and Institute for Advance Study, The Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong

Supporting Information

ABSTRACT: The conformational dynamics of multibody systems plays crucial roles in many important problems. Markov state models (MSMs) are powerful kinetic network models that can predict long-time-scale dynamics using many short molecular dynamics simulations. Although MSMs have been successfully applied to conformational changes of individual proteins, the analysis of multibody systems is still a challenge because of the complexity of the dynamics that occur on a mixture of drastically different time scales. In this work, we have developed a new algorithm, automatic state partitioning for multibody systems (APM), for constructing MSMs to elucidate the conformational dynamics of multibody systems. The APM algorithm effectively addresses different time scales in the multibody systems by directly incorporating dynamics into geometric clustering when identifying the metastable conformational states. We have applied the APM algorithm to a 2D potential that can mimic a protein−ligand binding system and the aggregation of two hydrophobic particles in water and have shown that it can yield tremendous enhancements in the computational efficiency of MSM construction and the accuracy of the models.



1. INTRODUCTION

Understanding the conformational dynamics of multibody systems is of critical importance in understanding many biological processes such as protein−ligand recognition,[1] protein−protein interactions,[2,3] and peptide aggregation.[4] In addition to biology, insights into multibody dynamics would lay a foundation for understanding many other processes such as self-assembly of nanoparticles[5] and other supramolecular systems.[6,7] Therefore, elucidating the kinetic mechanisms of multibody systems would greatly facilitate protein engineering,[8] structure-based drug design,[9] nanotechnology,[10] and many other important areas in chemistry and biology.

For complex multibody systems, it is difficult to examine the details of their conformational dynamics because doing so requires a detailed understanding of the free energy landscapes of individual molecules and how interactions between different molecules may alter these landscapes. Molecular dynamics (MD) simulations have been shown to be a valuable approach to complement experimental techniques to study the conformational dynamics.[1,11−13] However, many events of interest, such as protein−ligand binding and peptide aggregation, may occur on time scales that are orders of magnitude longer than most atomistic MD simulations, which are tens to hundreds of nanoseconds in length.[14,15] Therefore, it remains challenging to obtain a converged statistical picture of the conformational dynamics of these events directly from simulations.

Markov state models (MSMs)[16−36] are kinetic network models that can model the long-time-scale dynamics from many short MD simulations and thus hold great potential for understanding the conformational dynamics of multibody systems. In an MSM, the conformational space is first divided into a set of metastable states, and the conformational dynamics are simplified as a network of transitions between these states. This projection of dynamics introduces memory effects, where the temporal evolution of the probability for the system to visit a given state depends on its memory kernel.[30] Since the free energy landscapes underlying biological macromolecules often contain many minima (or metastable states) separated by barriers, the conformational dynamics is characterized by a separation of time scales in which intrastate relaxation is faster than interstate transitions. In an MSM, these fast motions are integrated out by coarse-graining in time with a discrete unit of Δt. If Δt is longer than the relaxation time within each metastable state, the model can be considered Markovian. Under this condition, the long-time-scale dynamics can be modeled by a first-order master equation,

Received: August 7, 2014
Published: December 11, 2014

DOI: 10.1021/ct5007168
J. Chem. Theory Comput. 2015, 11, 17−27

$$\mathbf{P}(n\Delta t) = [\mathbf{T}(\Delta t)]^n \mathbf{P}(0) \quad (1)$$

where $\mathbf{P}(n\Delta t)$ is the state population vector, n is the number of steps of propagation, Δt is the lag time, and $\mathbf{T}(\Delta t)$ is the transition probability matrix, which is generated by counting the number of transitions between pairs of states at Δt from MD trajectories.

MSMs have recently been applied successfully to study conformational changes of many single-body systems such as protein folding and RNA folding.[22,23,37−43] In these studies, the "splitting-and-lumping" algorithm is the standard method for determining the metastable states of an MSM.[16,18,22,23] In this method, conformations from MD simulations are first split into a large number of small microstates (e.g., 1000 to 10 000) according to their structural similarity (based on, e.g., the root mean square deviation (RMSD)), where a high degree of structural similarity for conformations within the same microstate is used as an indication of kinetic similarity. Microstate MSMs are especially useful for making quantitative comparisons with experiments but are too complicated for us to gain further mechanistic insight.[44] Microstates are thus often lumped together into macrostates on the basis of their kinetic proximity to form an MSM containing fewer states.[45]

Constructing MSMs for multibody systems is challenging because the intrinsic features of their dynamics occur on a mixture of different time scales. For example, to model protein−ligand binding, one must capture both the conformational changes of the protein and the heterogeneous time scales of the ligand dynamics due to interactions with the protein. In particular, a ligand hardly moves when it is interacting with a protein but diffuses quickly in solution.[12] The uniform geometric clustering at a single resolution used in the splitting-and-lumping algorithm is inadequate to describe both of these regimes. A high-resolution clustering may cause the microstates in the protein−ligand separated region to be too small, while a low-resolution clustering can cause the microstates in the binding region to be too big. Furthermore, ligand binding is often coupled with protein conformational changes that further complicate the system's kinetics. Previously, this issue has often been avoided by studying proteins that do not undergo significant conformational changes upon binding[46] or focusing on the protein−ligand diffusive association process rather than the details of the binding mechanism.[47]

In this work, we have developed a novel algorithm, automatic state partitioning for multibody systems (APM), to deal with the complexity of the multibody systems' dynamics. The key insight of the APM algorithm is to incorporate kinetic information when performing geometric clustering in order to generate microstates with a maximum residence time rather than a uniform size in structural space (such as the RMSD radius). We show that the APM algorithm greatly outperforms the splitting-and-lumping algorithm for both a two-dimensional (2D) potential and a two-particle aggregation system.

2. METHODS

2.1. Splitting-and-Lumping Algorithm for the State Decomposition. In this section, we first review the splitting-and-lumping state decomposition algorithm, which has been automated[18,22] and widely applied to construct MSMs for a single protein or other individual macromolecules. In order to identify the metastable conformational states, this algorithm considers both geometric and kinetic information from MD simulations in two sequential steps: geometric splitting and kinetic lumping.

2.1.1. Geometric Splitting. In this step, geometric clustering is applied to divide the MD conformations into a large number of small clusters or microstates. This clustering is often based on distance metrics such as RMSDs between pairs of protein conformations. The k-centers algorithm[18,48] and its variations[49] are popular algorithms that have been applied for the microstate clustering because of their speed, as an approximate implementation has only a computational cost of $O(kN)$, where k is the number of clusters and N is the total number of conformations. For example, using a GPU parallelized implementation of this algorithm, one can divide 250 000 MD conformations of a 370-residue protein into 4000 clusters within 40 s.[50] Quickly splitting conformations into microstates is especially useful when one deals with large data sets containing millions of conformations from thousands of MD trajectories. It should be noted that the geometric splitting step is essential because it groups conformations from different trajectories into a single microstate, which makes it possible to combine transition events between microstates that occurred in independent MD trajectories into a single model.

One can then construct MSMs based on these microstates by assuming that structurally similar conformations are also kinetically close. However, one must always note that when the number of microstates is small, the above assumption may break, since the geometric size of the microstates is too large and internal free energy barriers may exist within a microstate. The resulting model is thus not Markovian. Bowman et al.[44] showed that these systematic errors due to the non-Markovian nature of the model can be reduced by decreasing the geometric size of the microstates. For example, a 10 000-state MSM for the Villin headpiece can make quantitative predictions of various properties that are in reasonable agreement with the original MD simulations and experiments.[44] On the other hand, if one attempts finer partitioning by increasing the number of microstates, the statistical errors in the estimations of transition probabilities between pairs of microstates become more significant because of the decrease in the number of transition events. One therefore faces a trade-off between reducing systematic errors due to non-Markovian behavior (which requires more states) and reducing statistical errors due to conformational sampling (which requires fewer states).

The above dilemma becomes more obvious when dealing with multibody systems. As discussed in the Introduction and other studies,[12] the time scales of the dynamics in different regions of the conformational space can be extremely different in multibody systems. Geometric splitting by a method such as the k-centers algorithm may then produce a set of microstates with a homogeneous cluster radius but heterogeneous time scales of intrastate dynamics. In this situation, the intrastate relaxation of certain microstates that contain internal barriers may even take significantly longer times than interstate transitions between some other microstates. In order to avoid this complication caused by internal barriers, one then needs to choose a very high resolution geometric splitting over the whole conformational space that may result in overly small microstates, especially in the region where the dynamics is fast. We will further elucidate this issue in the Results and Discussion using a 2D potential.

2.1.2. Kinetic Lumping. Microstate MSMs are useful for making quantitative comparisons with experiment[44] but often

DOI: 10.1021/ct5007168
J. Chem. Theory Comput. 2015, 11, 17−27

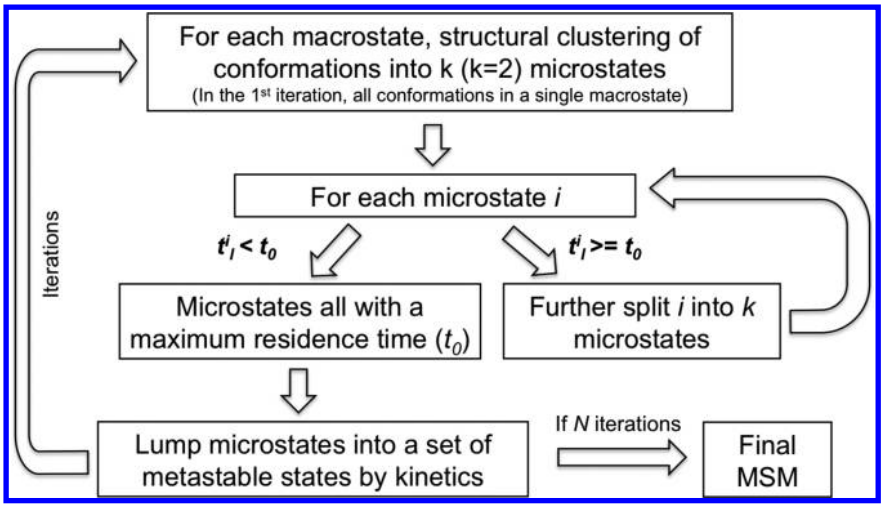


Figure 1. Flowchart of the APM algorithm for constructing MSMs for the multibody systems.

contain too many states for the understanding of mechanisms of conformational changes. To generate a more comprehensible model, one can lump together microstates that can interconvert quickly into the same metastable state to construct a macrostate MSM model. Perron cluster cluster analysis (PCCA)[51] is one of the most commonly used kinetic lumping algorithms. As a spectra-based clustering algorithm, PCCA determines the lumping of states on the basis of the sign structure of the eigenvectors of the transition probability matrix. More recently, other kinetic lumping algorithms such as PCCA+,[52] the Bayesian agglomerative clustering engine (BACE),[53] the hierarchical Nyström expansion graph (HNEG),[21] and the MPP algorithm[54] have also been developed, and their performances have been compared by Bowman et al.[24]

2.2. Automatic State Partitioning for Multibody Systems (APM) Algorithm. To deal with the complexity of the dynamics in multibody systems, we have developed the APM algorithm to incorporate kinetic information when performing geometric clustering in order to generate microstates using an iterative approach.

2.2.1. Overview. The key insight of the APM algorithm is to generate microstates with a maximum residence time rather than a uniform size in structural space (such as a uniform RMSD radius). This is achieved by starting off with a coarse clustering and breaking microstates into smaller ones until they all have residence times below a single upper limit. The residence time of a certain microstate i is measured by relaxation of the transition probability out of this state (i.e., the escape probability) $P(i, t)$:

$$P(i, t) = 1 - \frac{\sum_{j=1}^{k} \sum_{n=0}^{(T_j - t)/s} \delta(m_j(ns) - i)\delta(m_j(ns + t) - i)}{\sum_{j=1}^{k} \sum_{n=0}^{(T_j - t)/s} \delta(m_j(ns) - i)} \quad (2)$$

where $m_j(t)$ indicates which microstate the system is in at time t and T_j is the length of the jth MD trajectory with conformations stored at a time interval of s. By modeling the decay of the probability for the system to stay in a certain microstate using a single exponential, we can use the lifetime (t_l) of the population decay of a certain microstate i, estimated as the time t at which $P(i, t) = 1 - 1/e$, as the indicator of its residence time. If $t_l < t_0$ (where t_0 is the predetermined maximum residence time), the system can relax out of a microstate within t_0.

2.2.2. Detailed Procedure. As shown in Figure 1, we apply a top-down iterative approach to construct MSMs as follows:

(a) Bipartitioning is performed to split a certain state or the whole conformational space (in the first step) into two microstates. In the first step, we randomly pick a conformation and identify another conformation that is furthest away from this one on the basis of a certain distance function. These two conformations are treated as the two centers of microstates, and all of the other conformations are assigned to the closer center. In any later step, a new center conformation is selected as the one furthest from all of the existing centers, and the conformational space is repartitioned following the Voronoi tessellation scheme[55−57] to assign every conformation to its closest center.

(b) For a certain microstate i that is generated, we make use of the escape probability defined in eq 2 for further splitting. If $P(i, t_l) > P(i, t_0)$, which means that the escape probability after 1 unit of lifetime ($P(i, t_l) = 1 - 1/e$ in our single-exponential model) is larger than the observed escape probability at the predefined residence time limit, this microstate is further split as stated in step (a). This procedure is repeated for every newly generated state until all of the microstates have been visited.

In this step, we traverse the microstates in a depth-first manner. One should then expect that conformations in the slow dynamics region (e.g., when a ligand binds to a protein) will undergo multiple rounds of splitting and thus that such regions will contain finer microstates. In this way, we can obtain a partitioning that consist of microstates with variable sizes but a common maximum residence time in dynamics.

(c) Kinetically related microstates are lumped into N metastable macrostates (where N is predetermined) using a spectra-based clustering algorithm. Currently PCCA[51] and a popular implementation of spectral clustering by Ng et al.[58] are available in our code.

(d) The above procedure (steps (a)−(c)) is repeated iteratively. In each iteration, independent geometric clustering following steps (a)−(b) is performed within each macrostate. At the end, all of these microstates are combined into a single set for the kinetic lumping in step (c).

DOI: 10.1021/ct5007168
J. Chem. Theory Comput. 2015, 11, 17−27

(e) An optimization scheme is implemented between iterations. The normalized metastability Q is used as the indicator of the quality of the model. Q is defined as

$$Q = \sum_i \frac{T_{ii}}{N} \tag{3}$$

A new kinetic lumping is accepted and used for the next iteration if $\exp[-(Q_{\text{prev}}^2 - Q_{\text{curr}}^2)] > 0.95$. Alternatively, we have also implemented the Metropolis Monte Carlo scheme to update the iterations.

(f) The resulting MSM is validated using the Chapman–Kolmogorov test.[16,38]

The iteration steps (steps (d) and (e)) are critical because microstates containing internal free energy barriers may appear to have small residence times. A similar strategy has also been applied by Chodera et al.[16] in their automatic state decomposition algorithm to eliminate internal free energy barriers.

The APM algorithm has been implemented using the C++ programming language, and it is available at http://compbio.ust.hk/software/apm.tar.

2.3. Simulation Details. *2.3.1. 2D Potential.* We designed a 2D potential to illustrate the advantage of our APM algorithm. This potential is similar to the Mexican hat potential[59] with three minima and has the following form:

$$V(r) = 30[-N(r, 0, 1) - 3N(r, 10, 5) - N(r, 40, 20)] \tag{4}$$

where

$$r(x, y) = \sqrt{x^2 + y^2} \tag{5}$$

and

$$N(r, \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{r - \mu}{\sigma}\right)^2 \right] \tag{6}$$

This potential contains three energy minima, with the central one much deeper and narrower than the other two (see Figure S1a in the Supporting Information for the potential and Figure 2a for the enlarged central region). The dynamics in the needlelike central minimum are thus significantly slower than the rest. These heterogeneous time scales allow this 2D potential to mimic a two-body system such as protein–ligand binding, with the central minimum representing the bound state.

To sample this 2D potential, we performed 100 *NVT* MD simulations with 20 001 steps using a time step of 0.005 and a saving interval of 100 steps. The velocity-Verlet integrator[60] was adopted with an Andersen thermostat[61] (T = 1, and the coupling constant was 1 per time step). A reflective boundary condition (on both positions and velocities) was set at r = 75 to prevent the particle from diffusing away. As shown in Figure 2b, the points (2 000 100 in total) from our MD simulations widely sampled the conformational space for this simple 2D potential.

2.3.2. Aggregation of Hydrophobic Particles. MD Simulation Details. For the system with two hydrophobic particles in water, we performed ten 100 ns MD simulations starting from the same initial conformation with different initial velocities. The conformation was saved every 0.5 ps, giving a total of 2 million conformations. The interactions between the hydrophobic solute particles and water were set to be purely repulsive through the WCA potential.[62] The parameters for the bonded interactions and van der Waals (vdW) interactions between the two solute particles were taken from the generalized Amber force field (GAFF)[63] for benzene. All of the partial charges on the solute particles were set to zero. A cutoff of 12 Å was used for both vdW and short-range electrostatic interactions. The long-range electrostatic interactions were modeled using the particle-mesh Ewald (PME) method.[64] The two hydrophobic particles were solvated in a cubic water box with 2470 TIP3P[65] waters. In the initial configuration, the two benzene-shaped solute particles were parallel with a separation distance of 6 Å. The simulation system was first minimized with the steepest-gradient algorithm, followed by a 100 ps simulation applying a position restraint potential to the solute molecules. All of the production MD simulations were performed in the *NPT* ensemble (P = 1 bar and T = 300 K). The velocity-rescaling thermostat[66] with a coupling constant of 0.1 ps^{-1} and the Berendsen barostat[67] with a coupling constant of 1.0 ps^{-1} were used for the temperature and pressure couplings, respectively. The time step was set at 2 fs, and the nonbonded pair list was updated every 10 steps. All of the bonds were constrained using the LINCS algorithm.[68]

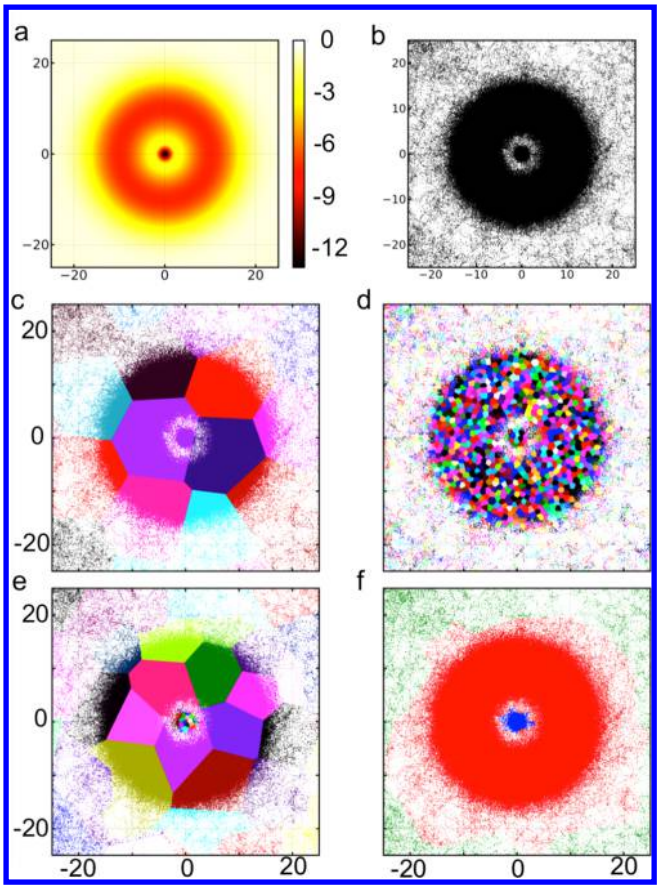


Figure 2. (a) Central region of the 2D potential (see Figure S1a for the full potential). (b) Superimposition of points from MD simulations. (c) A uniform low-resolution clustering with k = 117 produced by the k-centers clustering algorithm (see Figure S1c for the full view). Microstates are shown in different colors. (d) A uniform high-resolution clustering with k = 15 000 (see Figure S1d for the full view). (e) The APM algorithm generates 117 microstates with various sizes (see Figure S1e for the full view). (f) The final three-state macrostate MSM (see Figure S1e). The figures were prepared with matplotlib.[74]

DOI: 10.1021/ct5007168
J. Chem. Theory Comput. 2015, 11, 17–27

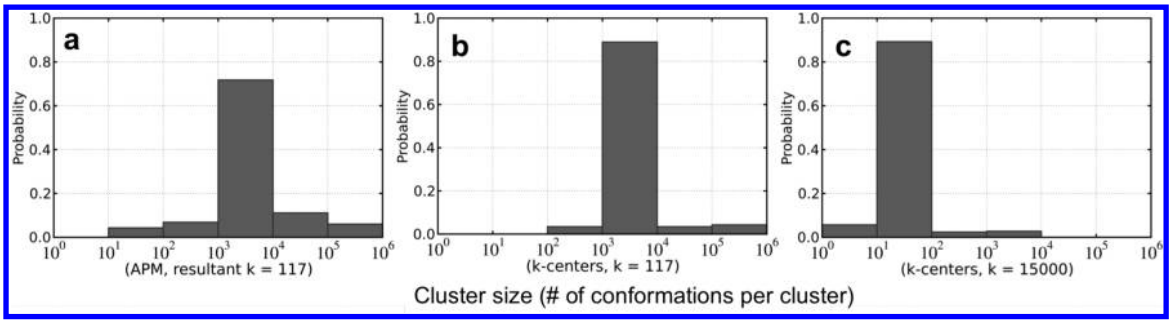


Figure 3. Comparison of the size distributions of clusters generated by the APM algorithm and the *k*-centers clustering algorithm with small k (k = 117) and large k (k = 15 000) for the 2D potential. The cluster size is defined as the number of conformations each cluster contains. The figures were prepared with matplotlib.[74]

The simulations were performed using the GROMACS 4.5.1 software.[69]

State Decomposition Using the APM Algorithm. To apply the APM algorithm, we selected t_0 = 1 ps and performed 200 iterations of clustering. All of the carbon atoms of hydrophobic particles were included in the RMSD calculations during clustering. The normalized metastability Q of the produced MSM (see eq 3) reached 0.925. We set the number of macrostates N to be 5 because there was a stable gap over a wide range of lag times between the fourth- and fifth-slowest implied time scales of the microstate MSM (see Figure 7a). Furthermore, in addition to the separated state, geometrically there should exist four different ways for the two planar benzene-shaped molecules to collapse, resulting in a total of five macrostates (see Figures 8 and 9). Because of the periodic boundary conditions, hydrophobic particles that leave the simulation box on one side can quickly re-enter from the periodic image on the other side, which causes artificially fast kinetics for particles close to the box boundary. To avoid this issue, we included only those MD conformations where the center-of-mass distance between the two hydrophobic particles was less than 19 Å (approximately half of the box size in each dimension). The resultant MSM was further validated by the Chapman−Komogolov test.[16] Finally, we systematically varied t_0 to be 0.5, 1, 2.5, 5, 10, and 25 ps to explore the effect of this parameter on the quality of the MSM constructed from the APM algorithm.

Mutual Information To Compare Different State Decompositions. To quantify the similarity between a pair of state decompositions, we introduced a similarity measurement I based on mutual information between two state partitions. In particular, I between a pair of partitions $f(x)$ and $g(x)$ is defined as

$$I = \frac{\sum_i \sum_j P(f(x) = i,\, g(x) = j) \log \frac{P(f(x) = i, g(x) = j)}{P(f(x) = i)P(g(x) = j)}}{-\sum_i P(g(x) = i) \log P(g(x) = i)} \tag{7}$$

where x represents a particular conformation and i and j are the macrostate indices in the MSM. The denominator corresponds to the entropy of the partition $g(x)$. When two state decompositions are identical, I = 1. When $f(x)$ and $g(x)$ are independent random partitions, I = 0.

We compared the qualities of the state decompositions produced by the APM algorithm and the popular splitting-and-lumping algorithm when different numbers of MD simulations were included as the input data. In particular, we selected the reference state decomposition as the one obtained from the APM algorithm using the whole well-sampled data set (ten 100 ns MD simulations). Since the state decompositions contain hard boundaries, conformations in the transition state region are grouped into a certain metastable state. However, the intrinsic feature of the transition-state conformations (i.e., equal transition probabilities to the initial and final states) makes them difficult to assign to a single macrostate, and their state assignments should be fuzzy. To avoid this ambiguity, we tended not to include those "fuzzy" conformations in our reference state decomposition. To achieve this, we ran the APM algorithm 10 times with different initial seeding conformations and selected only those MD conformations (91.9% of the total conformations) that were always assigned to the same macrostate (indicating that their membership was not "fuzzy") to be included in the reference state decomposition. To test the efficiency of the APM algorithm with the reduced MD data set, we equally divided the MD trajectories into shorter segments and then randomly selected from these segments to construct a reduced data set. For example, 10% and 5% of the original MD data set (ten 100 ns MD simulations) contained 10 segments of 10 ns and 5 ns MD trajectories, respectively. The 1% data set contained a segment of 10 ns MD simulation.

3. RESULTS AND DISCUSSION

In this section, we demonstrate that the APM algorithm outperforms the splitting-and-lumping algorithm using both a 2D potential and a two-particle aggregation system.

3.1. 2D Potential. As elaborated in previous sections, the fundamental hurdle of building MSMs for multibody systems with the splitting-and-lumping algorithm is the existence of dynamics on vastly different time scales when the particles are close together compared with when they are far apart. The 2D potential used in the first test was designed to illustrate this difficulty. There are three regions in the potential surface (see Figure S1a for the full potential and Figure 2a for the enlarged intermediate and inner regions): the outer diffusive region represents the fast dynamics mimicking the situation where the particles in the system are far apart; the intermediate region represents the slower dynamics where different particles start to encounter each other; and the inner needlelike region represents the extremely slow dynamics when the particles are in close contact (or in the correct binding mode). With our MD simulations, we managed to obtain reasonable sampling in all three regions. As shown in Figures 2b and S1b, the density of sampled conformations is much higher in the inner region than in the outer diffusive region.

DOI: 10.1021/ct5007168
J. Chem. Theory Comput. 2015, 11, 17−27

With the splitting-and-lumping algorithm, a uniform geometric clustering on this MD data set regardless of its resolution failed to produce MSMs that can properly identify various metastable minima. In a low-resolution clustering (117 microstates), the inner minimum was mostly clustered together with the intermediate region in the same microstate (see Figures 2c and S1c), and thus, the subsequent kinetic lumping could not distinguish the two minima efficiently. Indeed, this 117-state model failed to separate the inner region and the intermediate region (see Figure S2d), although it successfully distinguished the intermediate region from the outer region (see Figure S2b). On the other hand, in a high-resolution clustering (15 000 microstates), the boundary between the inner minimum and the intermediate region can be clearly identified, but this clustering generates overly small microstates, particularly in the diffusive outer region (see Figure S1d). Indeed, we found that nearly 90% of the microstates contained only 10 to 100 conformations and another 5% of the microstates contained even fewer than 10 conformations under this high-resolution clustering (see Figure 3c). This resulted in very few transition counts out of these microstates and thus introduced large statistical errors in the subsequent kinetic lumping. As shown in Figure S2d, this 15 000-state model mistakenly split the outer region into two parts even though free energy barriers did not exist in this region (see Figure S2d).

The maximum residence time, t_0, is an important input parameter in the APM algorithm that may alter the resolution of the clustering. Therefore, we also examined the impact of t_0 on the quality of the resulting state decompositions. In particular, we systematically varied t_0 over a wide range between 5 and 5000. For this system, we used the exact partitioning based on the analytical model of the potential as the gold standard to examine the quality of state decompositions constructed with different values of t_0. In this gold standard, we divided the potential into three regions with the boundaries set as the local maxima of the potential at r = 2.76 and 26.93. As shown in Figure 5, when t_0 lies between 10 and

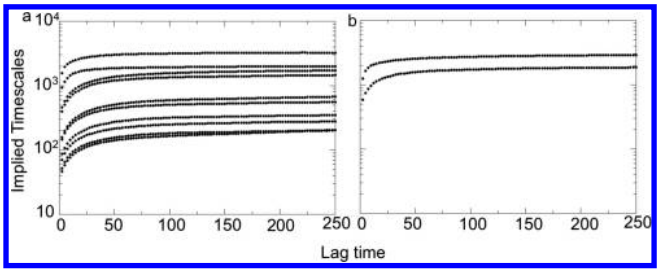


Figure 4. Implied time scale plots of the 2D potential for the MSMs constructed using the APM algorithm containing (a) k = 117 microstates (only the slowest 10 implied time scales are displayed) and (b) k = 3 macrostates.

2500, the APM algorithm can produce state decompositions that are nearly identical to the gold standard, with the mutual information I > 0.95 between the two state partitions (see eq 7 in Methods for the definition of I). This observation indicates that our APM can generate robust results over a wide range of t_0. The poor performance of the APM algorithm at small t_0 (<10) is due to oversplitting of the diffusive region, resulting in large statistical errors. For t_0 > 2500, the APM algorithm also fails to produce accurate state decompositions because t_0 is too large for the algorithm to distinguish the boundaries between

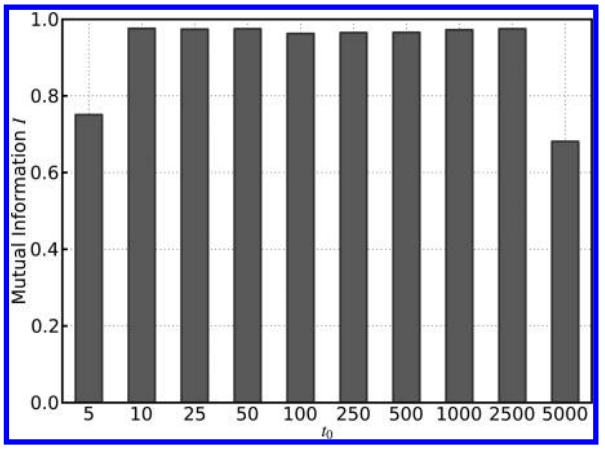


Figure 5. Comparison of mutual information values I for the state decompositions of MSMs generated with different t_0. The reference three-state decomposition is the exact partitioning with the boundaries set as the local maxima of the analytical form of the 2D potential (at r = 2.76 and 26.93). The figure was prepared with matplotlib.[74]

different macrostates. Indeed, among all of the implied time scales describing the inter-macrostate transitions, the shortest one is at ~2000 (see Figure 4B). This value is consistent with the upper bound of t_0, which is related to the maximum intrastate residence time. The above observations indicate that t_0 should be smaller than the shortest implied time scale of the macrostate MSM or the $(N - 1)$th-slowest implied time scale of the microstate MSM (where N is the number of macrostates). In practice, we suggest that t_0 might be chosen to be substantially (e.g., an order of magnitude) smaller than the $(N - 1)$th-slowest implied time scale of the microstate MSM.

3.2. Application to the Aggregation of Two Hydrophobic Particles in Water. We also applied the APM algorithm to construct MSMs for the aggregation of two hydrophobic particles in water. The benzene-shaped hydrophobic particles have purely repulsive interactions with water through the WCA potential,[62] and two of these rigid particles can collapse in water. Similar to the protein−ligand binding system, the dynamics of the hydrophobic particles is greatly slowed down when they are in contact compared with when they are freely diffusing in water. Therefore, when a large number of conformations are superimposed, the collapsed state shows a significantly higher density than the separated state (see Figure 6).

We applied the APM algorithm to construct MSMs for this system from an MD data set containing ten 100 ns *NPT* simulations. Each of these simulations displayed many transitions between the separated and collapsed states (see Figure S3), indicating that we obtained sufficient sampling to study the aggregation of these two hydrophobic particles. After 200 iterations with t_0 = 1 ps, the APM algorithm produced a model containing 126 microstates that can be further lumped into five macrostates (see Methods for details). The slowest implied time scales of the microstate and macrostate MSMs are consistent and occur at around 100 ps (see Figure 7). The coarse-grained macrostate MSM contains one separated state and four collapsed states (see Figure 8). The separated state has a low population (~2%), while all four collapsed states are significantly larger and equally populated (~24%). The four collapsed states correspond to the different orientations the two hydrophobic particles can take when they approach each other.

DOI: 10.1021/ct5007168
J. Chem. Theory Comput. 2015, 11, 17−27

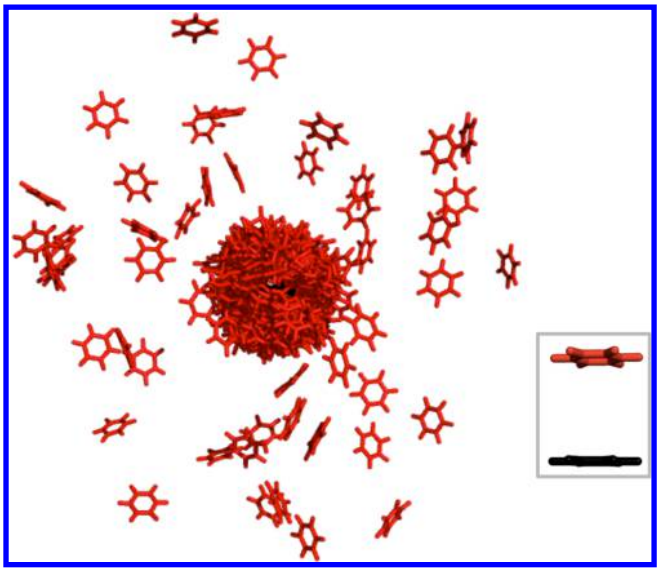


Figure 6. Superimposition of 500 representative conformations for the two-particle system. All of the conformations are aligned to one of the two particles (in black), and the relative positions of the other particle from different conformations (in red) are overlaid. The two particles display fast dynamics when separated but slow dynamics when collapsed. This figure was prepared with PyMOL.[75]

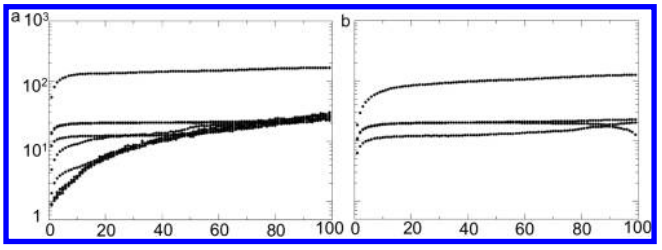


Figure 7. Implied time scale plots of the system of two hydrophobic particles for the MSMs constructed using the APM algorithm containing (a) k = 126 microstates (only the slowest 10 implied time scales are displayed) and (b) k = 5 macrostates.

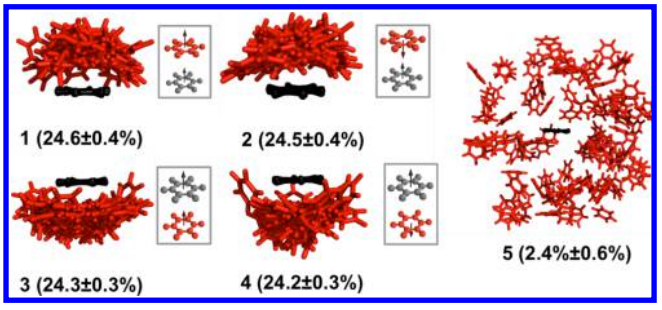


Figure 8. Superimposition of 100 representative conformations from each metastable state. One of the two particles (shown in black) is used to define a reference frame, and the relative positions of the other particle are shown in red. Equilibrium state populations predicted by the MSM are also shown. The figures were prepared with PyMOL[75] and MayaVi.[76]

By projecting the free energy landscape onto a pair of interparticle polar angles, we show in Figure 9 that different collapsed states indeed represent different relative orientations of the two particles. Although these collapsed states are geometrically symmetric, they are kinetically well-separated.

We further validated the macrostate MSM using the Chapman−Kolmogorov test.[16] In particular, we compared the time evolution of the state populations predicted by the propagation of the MSM (see eq 1) with those directly obtained from MD trajectories. For lag times of 50 and 100 ps, the MSM predictions match well with those directly obtained from the MD trajectories (see Figures 10b,c and S4). For a lag time of 10 ps, the MSM predicts faster kinetics for the system to make transitions between the separated state and any of the collapsed states (see the black lines in the top and bottom panels of Figure 10a), while the MSM's prediction for the relaxation from a collapsed state to other collapsed states is reasonable (see the blue, cyan, and red lines in the top panel of Figure 10a). The above observations indicate that MSMs constructed with lag times of 50 and 100 ps are both Markovian, while a lag time of 10 ps is not sufficiently long to make the model Markovian. These results are also consistent with the implied time scale plots shown in Figure 7b. At a lag time of 50 or 100 ps, the implied time scales already reach the plateau indicating that the model is Markovian,[28] while for the lag time of 10 ps the implied time scales are still increasing.

The MSM discussed above was built from a well-sampled MD data set containing numerous transitions between the separated and collapsed states. We also examined the performance of the APM algorithm when less MD data was used as input. In particular, we measured the quality of the state decompositions with less MD data by comparing their values of the mutual information I (see eq 7 and Methods for details) with respect to the reference decomposition generated from the whole MD data set (see Methods for details). When I = 1, the two compared decompositions contain identical macrostates, and greater I indicates greater similarity between the test and reference state partitionings and thus higher quality. As shown in Figure 11, our APM algorithm produced MSMs with I = 0.876 with only 1% of the data set (see Methods for details). In this situation, the APM algorithm still successfully identified all five metastable macrostates. However, the splitting-and-lumping algorithm with a low-resolution microstate clustering (50 microstates and 5 macrostates) failed (I < 0.6) regardless of the amount of data included in the clustering. Further investigations showed that when k = 50, most of the microstates were located in the region where the two particles were separated, and the microstates in the collapsed region were too large (in terms of RMSD radius). Therefore, two of the collapsed states were often mis-lumped into a single state. When a high-resolution microstate clustering was applied (k = 500), the splitting-and-lumping algorithm produced reasonable state decompositions when the amount of data was sufficiently large. For example, when 100% and 10% of the whole data set were included, I values of 0.903 and 0.885 were obtained, respectively. With 5% of the data, the mutual information decreased to 0.812, and it was further reduced to 0.524 with 1% of the data, where the separated region was often split into multiple states. Therefore, the APM algorithm can construct MSMs with high quality using significantly less sampling than the splitting-and-lumping method. This is extremely important when dealing with complex multibody systems where sampling is often limited.

One important input parameter in the APM algorithm is the maximum residence time, t_0. We systematically varied t_0 to examine how this parameter may affect the quality of the MSM constructed from the APM algorithm. As shown in Figure 12, for $t_0 \leq 5$ ps the APM algorithm produced state

DOI: 10.1021/ct5007168
J. Chem. Theory Comput. 2015, 11, 17−27

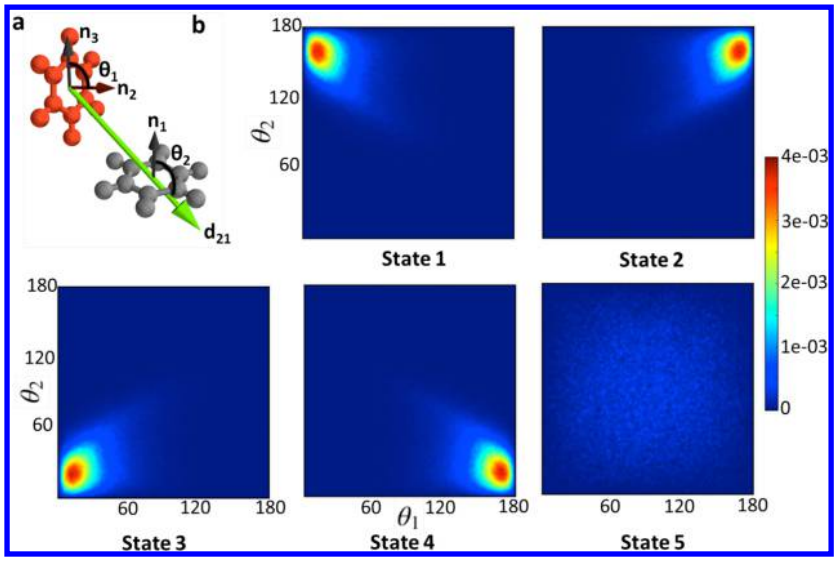


Figure 9. Probability distributions of the two orientation angles θ_1 and θ_2 for MD conformations in different metastable macrostates for the two-particle system. (a) Definitions of θ_1 and θ_2. The vectors $\mathbf{n}_1$ and $\mathbf{n}_2$ correspond to the normals of the black and red hydrophobic particles, respectively. (b) Probability distributions of the two orientation angles θ_1 and θ_2 for macrostates 1−5. The figures were prepared with matplotlib[74] and MayaVi.[76]

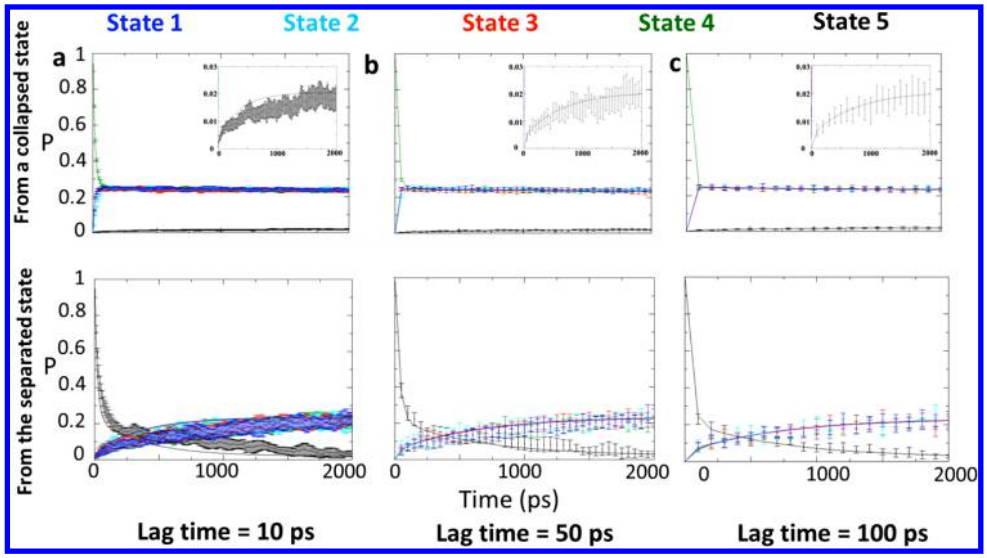


Figure 10. Validation of the MSM using the Chapman−Komogolov test for the two-particle system. Shown are comparisons of the time evolutions of state populations predicted by propagation of the MSM and those directly obtained from MD trajectories when the system starts from one of the collapsed states (top panels) or the separated state (bottom panels). The tests of MSMs with lag times of 10, 50, and 100 ps are shown in (a), (b), and (c), respectively. State populations for the four collapsed states (states 1−4) and the separated state (state 5) are shown in blue, red, cyan, yellow, and black, respectively.

decompositions that were nearly identical to the reference state decomposition ($I > 0.95$). For $t_0 \geq 10$ ps, the quality of the partitioning decreased and two or more collapsed states were combined into a single macrostate. Interestingly, we notice that the fastest implied time scale of the macrostate MSM was 10−20 ps, which corresponds to the transitions between different collapsed states (see Figure 7b). As discussed for the 2D potential system, we suggest the upper bound of t_0 to be the fastest implied time scale of the macrostate MSM. In contrast to the 2D potential system, even when the smallest possible t_0 of 0.5 ps (the interval to save the MD conformations) was used, the lower bound of t_0 was still not reached, indicating that our MD data set was well-sampled.

Even though the above examples all involve multibody systems, the APM algorithm is also very efficient when applied to single-body systems. For example, using the APM algorithm on the alanine dipeptide system, we were able to build a seven-state macrostate MSM from only 97 microstates (see Figure S5c) that could accurately describe the system's conformational dynamics (see the validation of the model in Figure S6). As a comparison, the splitting-and-lumping algorithm required as many as 4000 microstates to produce an MSM with quality similar to that from our APM algorithm (see Figure S5d). It should be noted that the APM algorithm is significantly more beneficial than the splitting-and-lumping algorithm on systems with vastly different time scales of dynamics, which is commonly seen in, but not limited to, multibody systems.

DOI: 10.1021/ct5007168
J. Chem. Theory Comput. 2015, 11, 17−27

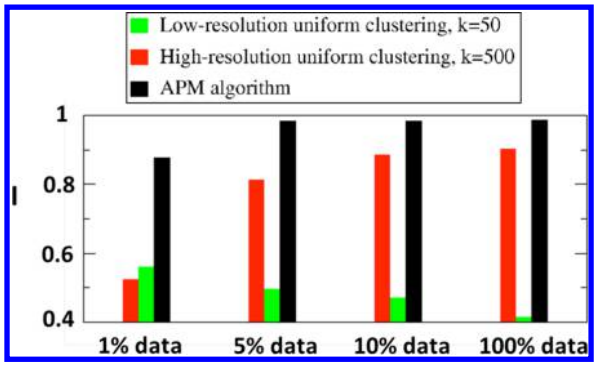


Figure 11. Comparison of values of the mutual information I for MSMs generated by the APM algorithm (black) and the splitting-and-lumping method (red and green). MSMs were constructed using different amounts of MD data. MSMs constructed from 100% of the data using the APM algorithm were validated (see Figure S4) and used as the reference (I = 1).

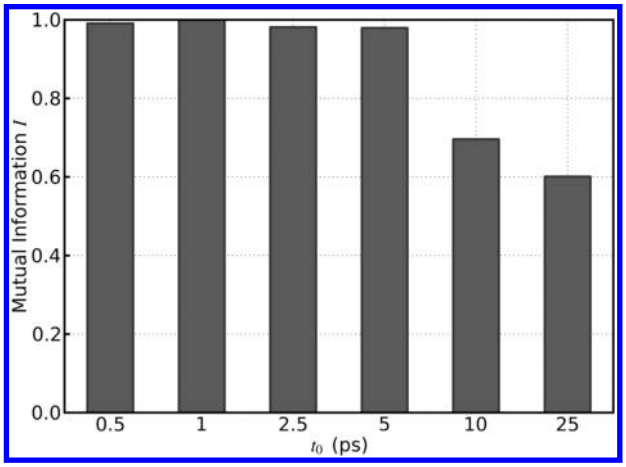


Figure 12. Values of the mutual information I for the state decompositions of MSMs generated with different t_0 for the system of two hydrophobic particles in water. The figure was prepared with matplotlib.[74]

For example, in protein folding, the time scale of the dynamics in unfolded states can be considerably faster than that of the folded state.[70] An alternative way to address this issue is to scale the geometric distance on the basis of the dynamics, such as in the kinetically discriminatory metric learning (KDML) method.[70] However, in the current implementation of KDML, a single scaling factor is associated with each axis (or reaction coordinate), which may not be sufficient to deal with multibody systems such as the 2D potential discussed in this work, where the scaling factors for x, y, and r need to be varied according to the local environment. Perkett and Hagan[71] have developed an approach to construct MSMs for the self-assembly process by focusing on the density distribution rather than the positions of individual subunits. This may not be optimal for certain multibody systems such as protein−ligand binding, where the accurate description of the conformational dynamics of individual particles is important. More recently, Yang and Gao[72] have also constructed MSMs to investigate the kinetics of $A\beta_{37-42}$ aggregation, in which they determined the metastable states on the basis of the physical properties of the system. Finally, we note that our APM algorithm may be combined with the recently developed on-the-fly learning and sampling framework (which has been successfully applied to a rigid protein−ligand binding system) to further improve its efficiency.[73]

4. CONCLUSION

In this work, we have developed the APM algorithm, which can automatically partition the conformational space and build MSMs for multibody systems. By introducing a maximum residence time t_0 for each microstate, we can directly take into account the dynamic information during the geometric state partitioning. Compared with the splitting-and-lumping algorithm, which produces microstates at a uniform spatial resolution, the APM algorithm is able to generate microstates at a mixture of different spatial resolutions adapted to the local conformational dynamics. To impose the upper limit of the microstate residence time, we have implemented a divide-and-conquer scheme to split conformations. In addition, iterations of resplitting and relumping are performed to remove potential internal barriers of microstates. We have shown that the APM algorithm greatly outperforms the popular splitting-and-lumping algorithm for constructing MSMs for a 2D potential as well as the aggregation of two hydrophobic particles. For example, we were able to obtain comparable results using less than a fifth of the sampling when constructing MSMs for the aggregation of two hydrophobic particles in water. We believe that the APM algorithm holds great potential in understanding the mechanisms of many multibody processes such as protein−protein interactions, protein−RNA recognition, peptide aggregation, and hydrophobic collapse. In the future, we plan to build upon the current framework of the APM algorithm to include various new features. For example, instead of predefining the number of macrostates N, we may apply lumping algorithms that do not require a predetermined N, such as HNEG[21] and the MPP algorithm.[54] In addition, since the metastability Q may not be the best indicator of the quality of MSMs, we may apply alternative objective functions in our Monte Carlo step, such as the extent of the Chapman−Kolmogorov test fulfillment or the maximization of implied time scales as suggested in the recent work on applying the variational principle to optimize MSM construction.[34]

■ ASSOCIATED CONTENT

Supporting Information

State decomposition of the full 2D potential, surface area analysis and Kolmogorov test on the hydrophobic particle system, and application and analysis of the alanine dipeptide system. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: xuhuihuang@ust.hk.

Present Address

⊥D.-A.S.: Department of Biochemistry, University of Washington, Seattle, Washington 98195, USA.

Funding

We acknowledge the support from the National Basic Research Program of China (973 Program, 2013CB834703), the National Natural Science Foundation of China (21273188), and the Hong Kong Research Grants Council (ECS 60981, AoE/M-09/12, and T13-607/12R). Computer resources were provided by the National Supercomputing Center of China in

DOI: 10.1021/ct5007168
J. Chem. Theory Comput. 2015, 11, 17−27

Shenzhen. F.K.S. acknowledges support from the Hong Kong Ph.D. Fellowship Scheme 2012/13 (PF11-08816).

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

The authors thank Robert T. McGibbon from Stanford University for useful discussions and constructive comments on the manuscript.

■ REFERENCES

(1) Boehr, D. D.; Nussinov, R.; Wright, P. E. *Nat. Chem. Biol.* **2009**, *5*, 789−796.
(2) Jones, S.; Thornton, J. M. *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 13−20.
(3) Keskin, O.; Gursoy, A.; Ma, B.; Nussinov, R. *Chem. Rev.* **2008**, *108*, 1225−1244.
(4) Straub, J. E.; Thirumalai, D. *Annu. Rev. Phys. Chem.* **2011**, *62*, 437−463.
(5) Grzelczak, M.; Vermant, J.; Furst, E. M.; Liz-Marzán, L. M. *ACS Nano* **2010**, *4*, 3591−3605.
(6) Liu, Y.; Wang, Z.; Zhang, X. *Chem. Soc. Rev.* **2012**, *41*, 5922−5932.
(7) Zhang, J.; Lu, Z.-Y.; Sun, Z.-Y. *Soft Matter* **2013**, *9*, 1947−1954.
(8) Dwyer, M. A.; Hellinga, H. W. *Curr. Opin. Struct. Biol.* **2004**, *14*, 495−504.
(9) Swinney, D. C. *Curr. Opin Drug Discovery Dev.* **2009**, *12*, 31−39.
(10) Park, C.; Yoon, J.; Thomas, E. L. *Polymer* **2003**, *44*, 6725−6760.
(11) Bucher, D.; Grant, B. J.; Markwick, P. R.; McCammon, J. A. *PLoS Comput. Biol.* **2011**, *7*, No. e1002034.
(12) Silva, D.-A.; Bowman, G. R.; Sosa-Peinado, A.; Huang, X. *PLoS Comput. Biol.* **2011**, *7*, No. e1002054.
(13) Meuwly, M.; Cui, Q. *Chem. Phys.* **2012**, *396*, 1−2.
(14) Tang, C.; Schwieters, C. D.; Clore, G. M. *Nature* **2007**, *449*, 1078−1082.
(15) Miller, D. M.; Olson, J. S.; Pflugrath, J. W.; Quiocho, F. A. *J. Biol. Chem.* **1983**, *258*, 13665−13672.
(16) Chodera, J. D.; Singhal, N.; Pande, V. S.; Dill, K. A.; Swope, W. C. *J. Chem. Phys.* **2007**, *126*, No. 155101.
(17) Noé, F.; Fischer, S. *Curr. Opin. Struct. Biol.* **2008**, *18*, 154−162.
(18) Bowman, G. R.; Huang, X.; Pande, V. S. *Methods* **2009**, *49*, 197−201.
(19) Buchete, N.-V.; Hummer, G. *J. Phys. Chem. B* **2008**, *112*, 6057−6069.
(20) Zheng, W.; Andrec, M.; Gallicchio, E.; Levy, R. M. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 15340−15345.
(21) Yao, Y.; Cui, R. Z.; Bowman, G. R.; Silva, D.-A.; Sun, J.; Huang, X. *J. Chem. Phys.* **2013**, *138*, No. 174106.
(22) Huang, X.; Bowman, G. R.; Bacallado, S.; Pande, V. S. *Proc. Natl. Acad. Sci. U.S.A.* **2009**, *106*, 19765−19769.
(23) Huang, X.; Yao, Y.; Bowman, G. R.; Sun, J.; Guibas, L. J.; Carlsson, G. E.; Pande, V. S. In *Pacific Symposium on Biocomputing*; World Scientific: Singapore, 2010; Vol. *15*, pp 228−239.
(24) Bowman, G. R.; Meng, L.; Huang, X. *J. Chem. Phys.* **2013**, *139*, No. 121905.
(25) Voelz, V. A.; Bowman, G. R.; Beauchamp, K.; Pande, V. S. *J. Am. Chem. Soc.* **2010**, *132*, 1526−1528.
(26) Noé, F.; Horenko, I.; Schütte, C.; Smith, J. C. *J. Chem. Phys.* **2007**, *126*, No. 155102.
(27) Schütte, C.; Fischer, A.; Huisinga, W.; Deuflhard, P. *J. Comput. Phys.* **1999**, *151*, 146−168.
(28) Swope, W. C.; Pitera, J. W.; Suits, F. *J. Phys. Chem. B* **2004**, *108*, 6571−6581.
(29) Prinz, J.-H.; Wu, H.; Sarich, M.; Keller, B.; Senne, M.; Held, M.; Chodera, J. D.; Schütte, C.; Noé, F. *J. Chem. Phys.* **2011**, *134*, No. 174105.
(30) Zwanzig, R. *J. Stat. Phys.* **1983**, *30*, 255−262.
(31) Vitalis, A.; Caflisch, A. *J. Chem. Theory Comput.* **2012**, *8*, 1108−1120.
(32) Keller, B.; Hünenberger, P.; van Gunsteren, W. F. *J. Chem. Theory Comput.* **2011**, *7*, 1032−1044.
(33) Pan, A. C.; Roux, B. *J. Chem. Phys.* **2008**, *129*, No. 064107.
(34) Nüske, F.; Keller, B. G.; Pérez-Hernández, G.; Mey, A. S. J. S.; Noé, F. *J. Chem. Theory Comput.* **2014**, *10*, 1739−1752.
(35) Schwantes, C. R.; Pande, V. S. *J. Chem. Theory Comput.* **2013**, *9*, 2000−2009.
(36) Pérez-Hernández, G.; Paul, F.; Giorgino, T.; Fabritiis, G. D.; Noé, F. *J. Chem. Phys.* **2013**, *139*, No. 015102.
(37) Da, L.-T.; Wang, D.; Huang, X. *J. Am. Chem. Soc.* **2012**, *134*, 2399−2406.
(38) Noé, F.; Schütte, C.; Vanden-Eijnden, E.; Reich, L.; Weikl, T. R. *Proc. Natl. Acad. Sci. U.S.A.* **2009**, *106*, 19011−19016.
(39) Morcos, F.; Chatterjee, S.; McClendon, C. L.; Brenner, P. R.; López-Rendón, R.; Zintsmaster, J.; Ercsey-Ravasz, M.; Sweet, C. R.; Jacobson, M. P.; Peng, J. W.; Izaguirre, J. A. *PLoS Comput. Biol.* **2010**, *6*, No. e1001015.
(40) Bowman, G. R.; Voelz, V. A.; Pande, V. S. *Curr. Opin. Struct. Biol.* **2011**, *21*, 4−11.
(41) Chodera, J. D.; Noé, F. *Curr. Opin. Struct. Biol.* **2014**, *25*, 135−144.
(42) Razavi, A. M.; Wuest, W. M.; Voelz, V. A. *J. Chem. Inf. Model.* **2014**, *54*, 1425−1432.
(43) Zhuang, W.; Cui, R. Z.; Silva, D.-A.; Huang, X. *J. Phys. Chem. B* **2011**, *115*, 5415−5424.
(44) Bowman, G. R.; Beauchamp, K. A.; Boxer, G.; Pande, V. S. *J. Chem. Phys.* **2009**, *131*, No. 124101.
(45) Pande, V. S.; Beauchamp, K.; Bowman, G. R. *Methods* **2010**, *52*, 99−105.
(46) Buch, I.; Giorgino, T.; Fabritiis, G. D. *Proc. Natl. Acad. Sci. U.S.A.* **2011**, *108*, 10184−10189.
(47) Held, M.; Metzner, P.; Prinz, J.-H.; Noé, F. *Biophys. J.* **2011**, *100*, 701−710.
(48) Hochbaum, D. S.; Shmoys, D. B. *Math. Oper. Res.* **1985**, *10*, 180−184.
(49) Beauchamp, K. A.; Bowman, G. R.; Lane, T. J.; Maibaum, L.; Haque, I. S.; Pande, V. S. *J. Chem. Theory Comput.* **2011**, *7*, 3412−3419.
(50) Zhao, Y.; Sheong, F. K.; Sun, J.; Sander, P.; Huang, X. *J. Comput. Chem.* **2013**, *34*, 95−104.
(51) Deuflhard, P.; Huisinga, W.; Fischer, A.; Schütte, C. *Linear Algebra Appl.* **2000**, *315*, 39−59.
(52) Weber, M.; Kube, S. In *Computational Life Sciences*; Berthold, M. R., Glen, R., Diederichs, K., Kohlbacher, O., Fischer, I., Eds.; Springer: Berlin, 2005; pp 57−66.
(53) Bowman, G. R. *J. Chem. Phys.* **2012**, *137*, No. 134111.
(54) Jain, A.; Stock, G. *J. Chem. Theory Comput.* **2012**, *8*, 3810−3819.
(55) Dirichlet, G. L. *J. Reine Angew. Math.* **1850**, *40*, 209−227.
(56) Voronoi, G. *J. Reine Angew. Math.* **1907**, *133*, 97−178.
(57) Aurenhammer, F. *ACM Comput. Surv.* **1991**, *23*, 345−405.
(58) Ng, A. Y.; Jordan, M. I.; Weiss, Y. In *Advances in Neural Information Processing Systems*; Jordan, M. I., LeCun, Y., Solla, S. A., Eds.; MIT Press: Cambridge, MA, 2001; pp 849−856.
(59) Gfeller, D.; Rios, P. D. L.; Caflisch, A.; Rao, F. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 1817−1822.
(60) Swope, W. C.; Andersen, H. C.; Berens, P. H.; Wilson, K. R. *J. Chem. Phys.* **1982**, *76*, 637−649.
(61) Andersen, H. C. *J. Chem. Phys.* **1980**, *72*, 2384−2393.
(62) Weeks, J. D.; Chandler, D.; Andersen, H. C. *J. Chem. Phys.* **1971**, *54*, 5237−5247.
(63) Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. *J. Comput. Chem.* **2004**, *25*, 1157−1174.
(64) Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. *J. Chem. Phys.* **1995**, *103*, 8577−8593.
(65) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926−935.

DOI: 10.1021/ct5007168
J. Chem. Theory Comput. 2015, 11, 17−27

(66) Bussi, G.; Donadio, D.; Parrinello, M. *J. Chem. Phys.* **2007**, *126*, No. 014101.
(67) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R. *J. Chem. Phys.* **1984**, *81*, 3684−3690.
(68) Hess, B.; Bekker, H.; Berendsen, H. J. C.; Fraaije, J. G. E. M. *J. Comput. Chem.* **1997**, *18*, 1463−1472.
(69) Hess, B.; Kutzner, C.; van der Spoel, D.; Lindahl, E. *J. Chem. Theory Comput.* **2008**, *4*, 435−447.
(70) McGibbon, R. T.; Pande, V. S. *J. Chem. Theory Comput.* **2013**, *9*, 2900−2906.
(71) Perkett, M. R.; Hagan, M. F. *J. Chem. Phys.* **2014**, *140*, No. 214101.
(72) Yang, Y. I.; Gao, Y. Q. *J. Phys. Chem. B* **2014**, DOI: 10.1021/jp502169b.
(73) Doerr, S.; De Fabritiis, G. *J. Chem. Theory Comput.* **2014**, *10*, 2064−2069.
(74) Hunter, J. D. *Comput. Sci. Eng.* **2007**, *9*, 90−95.
(75) *PyMOL*; Schrodinger, LLC: New York, 2010.
(76) Ramachandran, P.; Varoquaux, G. *Comput. Sci. Eng.* **2011**, *13*, 40−51.

DOI: 10.1021/ct5007168
J. Chem. Theory Comput. 2015, 11, 17−27